

Bangor University

DOCTOR OF PHILOSOPHY

Analysis and synthesis of critical design-thinking for data visualisation designers and learners

Alnjar, Hanan

Award date:
2017

Awarding institution:
Bangor University

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



PRIFYSGOL
BANGOR
UNIVERSITY

School of Computer Science
College of Physical & Applied Sciences

**Analysis and synthesis of critical
design-thinking for data visualisation
designers and learners**

Hanan R Alnjar

Submitted in partial satisfaction of the requirements for the
Degree of Doctor of Philosophy
in Computer Science

Supervisor Prof. Jonathan C Roberts

August 2017

Statement of Originality

The work presented in this thesis/dissertation is entirely from the studies of the individual student, except where otherwise stated. Where derivations are presented and the origin of the work is either wholly or in part from other sources, then full reference is given to the original author. This work has not been presented previously for any degree, nor is it at present under consideration by any other degree awarding body.

Student:

Hanan R Alnjar

Statement of Availability

I hereby acknowledge the availability of any part of this thesis/dissertation for viewing, photocopying or incorporation into future studies, providing that full reference is given to the origins of any information contained herein. I further give permission for a copy of this work to be deposited with the Bangor University Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorised for use by Bangor University and where necessary have gained the required permissions for the use of third party material. I acknowledge that Bangor University may make the title and a summary of this thesis/dissertation freely available.

Student:

Hanan R Alnjar

Acknowledgements

Firstly, I thank my God and Ahlulbayt for continuing favours, support and help for conducting this work and in my life in general. I would like to express my sincere gratitude to my supervisor Prof. Jonathan Roberts for his continuous support throughout my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me over all the time of research and in writing of this thesis. Besides my supervisor, I would like to thank my examiners: Dr. Simon Attfield and Dr. Sa'ad Mansoor for their insightful comments and encouragement. My sincere thanks also go to Dr. Panagiotis Ritsos who offered helpful advice and comments throughout my study. I am grateful to my fellow PhD students in computer science for their support and their helpful comments, especially help with the implementation. Also, I would acknowledge the students of the ICP3036 Computer Science module, at the critical analysis workshop, for allowing their work to be used in my research. Without them I could not have performed my research analysis.

I thank my husband, Nadir for offering his thoughtful, skills and experience in computer applications. I cannot forget his continuing support from him throughout my PhD. I could not have done this without his patience, kindness and enthusiasm. I also thank my family: my brothers, sisters and my adorable nieces and nephews, for supporting me spiritually throughout the whole of my PhD.

I do not know how I can express enough, all my thanks and appreciation to my mother; I am always asking god to extend her life. Without her encouragement and love, I would not have followed my interests and finished my PhD work.

I dedicate this dissertation to my mother. This is very least thing that I can offer her with full of love and gratefulness.

Hanan 2017.

Abstract

Designers of data-visualisation tools think deeply about their designs, and constantly question their own judgements and design decisions. They make these judgements to ascertain how to improve their ideas, such to make something suitable and fit-for-purpose. But self-reflection is often difficult, especially learners often find it difficult to critically reflect upon their own work. Therefore, there is a need to guide learners to perform appropriate critical reflections of their work, and develop skills to make better judgements. There are many visualisation computer tools and programming libraries that help users create visualisation systems, however, there are few tools or techniques to help them systematically critique or evaluate their creations, to ascertain what is good or what is bad in their designs. Learners who wish to create data-visualisation tools are missing structures and guidelines that will aid them critique their visualisations. Such critical analysis could be achieved by creating an appropriate computing tool and using metrics and heuristics to perform the judgement, or by human judgement helped by a written guide. Subsequently, this research explores structures to help humans perform better critical evaluations. First, the dissertation uses a traditional research methodology to investigate metrics in visualisation, to explore related work and investigate how metrics are used in computers to perform judgements. We design a framework that describes how and where metrics are used in the visualisation design process. Second, the focus turns to investigate how humans think and make critical judgement on designs, especially visualisation designs. We undertake an observational study where participants critique a range of objects and designs. An in-depth analysis of this observation-study is performed; through analysis and markup of this data, the work is analysed, and themes are extracted. We used a thematic analysis approach to extract these categories. We use these categories to develop our critical analysis system. Third, we followed an iterative approach to engineer our critical-evaluation system. The output system is the Critical Design Sheet that was created after much refinement and adaption by holding several think-aloud sessions to detect design problems and refine to

an effective model. Fourth, an evaluation process is performed to evaluate the usability of the Critical Design Sheet with users. Testing of the reliability and learnability of the tool was achieved by the analysis of user usage data over two different cohorts of students (PhD's and undergraduate). Fifth, an online implementation of the Critical Design Sheet was developed, which was briefly evaluated to discover if participants were satisfied with the computer version of the critical analysis system. These five parts represent the five contributions of the dissertation, respectively: the metric framework, critical thinking workshop and its analysis, the design of the Critical Design Sheet, the evaluation of the method, and finally the prototype online system. In conclusion, this dissertation provides learners and practitioners with a technique (the CDS) that has been proven to help students successfully critique their visual designs and make decision on their creations. The CDS works by breaking down the visual design into individual categories making it easier for practitioners to critique their work.

Contents

1	Introduction	1
1.1	Motivation and Vision	4
1.2	Hypothesis and Research questions	6
1.3	Aims and Objectives	6
1.4	Scope of research	8
1.5	Research methodology and structure of dissertation	8
1.6	Contributions	12
2	Related work	14
2.1	An introduction to metrics and opportunities for visualisation	16
2.2	Review of metrics in the visualisation literature	20
2.2.1	Decision making and directed thinking	24
2.3	The role of metrics in visualisation design and working towards a visualisation metrology	25
2.4	Proposed metrics framework for visualisation design process	28
2.5	Evaluation methods for visualisation and direct related work to the CDS	36
2.6	Summary	43
3	Critical Thinking	44
3.1	Fundamentals in critical thinking – definitions and concepts	44
3.2	Critical thinking models in the academic curriculum	49
3.3	Thinking with visualisation	52
3.3.1	Creativity and design-thinking strategy	52
3.3.2	Abstract models	56
3.3.3	Reflections on visualisation stages design	57
3.4	Summary	64
4	Design of The Critique Design Sheet	66
4.1	The vision	67
4.2	Workshop to explore design critiquing aspects	71
4.2.1	Participants	71
4.2.2	Workshop preparation and tasks	72
4.2.3	The workshop tasks analysis results: T1 & T2	76
4.2.4	The workshop tasks analysis results: T3 to T8	83
4.3	Design process for the Critical Design Sheet	86

4.3.1	First prototype design—decisions and influences	87
4.3.2	Design of the six principal categories	90
4.3.3	Pilot studies (talk-aloud evaluation)	94
4.3.4	Design of first-impression words, grading & star plot	100
4.4	The final Critical Design Sheet prototype	106
4.5	Using the CDS	108
4.6	Summary	110
5	The evaluation of CDS	112
5.1	Vision of the evaluation	113
5.2	Usability test preparation & Data gathering	117
5.2.1	Experiment 1 procedure & Materials (group1)	119
5.2.2	Experiment 2 procedure & Materials (group2)	121
5.3	Data analysis & Results	124
5.3.1	Observational analysis of CDS completion (good and bad practice)	126
5.3.2	Test Reliability	128
5.3.3	Run T-Test statistic	132
5.3.4	Data analysis of the “First Impression words”	137
5.3.5	Learnability test	140
5.3.6	Qualitative evaluation – a Case study of its use	142
5.4	Summary	149
6	Implementation of CDS online version	151
6.1	Vision for an online CDS version	151
6.2	Web technology features and services	152
6.3	CDS online tool constructions	154
6.3.1	Home Page	156
6.3.2	CDS Interface web-page	157
6.3.3	Results page	159
6.3.4	MyRecords page	161
6.3.5	Server-side of CDS implementation	163
6.4	Evaluation of CDS online tool	165
6.4.1	Method & Results	166
6.5	Discussion and limitations of the online CDS system	167
6.6	Summary	170
7	Conclusions and Future work	171
7.1	Limitations	175
7.2	Future Work	177
7.3	Conclusion	179
	References	180

List of Figures

1	
2	
3	
1.1	Research methodology. A diagram of the whole dissertation showing the four main parts (ABCD)	10
2.1	Visualisation journals distributions display that the IEEE TVCG journal has the maximum number of contributions in our research literature. . .	15
2.2	Towards a visualisation metrology. Think about the creative Design process, Purpose of the metric, its Properties; the different inputs within the Dataflow model, and different Types of the output (Actuators). . . .	27
2.3	The Whole System Model of Metrics (WSMM) for visualisation. This model is used to decide where the metrics apply to visualisation; e.g., whether they apply as part of the algorithm (such as incrementally adapting the result) or as a post-creation process, to evaluate how well the visualisation could be perceived. This model works as a conceptual model, it can help to categorise metrics and help the developer consider what type of metrics they may use.	29
2.4	Using (FdS) methodology as example to explain how metrics (qualitative/quantitative) can be performed and processed by the framework. Each stage is responsible to perform specific task which refers to specific metric type.	35
3.1	Bloom's levels of learning re-produced by the author from [59]. Bloom's taxonomy shows the cognitive processes that a learner moves from the foundation Knowledge level up to Application and ultimately, Synthesis and Evaluation, the original notation and concepts belongs to [20]. The taxonomy is also explained and links are provided for even more useful resources, http://www.odu.edu/content/dam/odu/col-dept/teaching-learning/docs/blooms-taxonomy-handout.pdf	50
3.2	The conceptual framework of design-thinking is characterised as an iterative process and it composed of three main activities, (1) Understand, (2) Explore and (3) Materialise. The image has been created by the author and the concepts taken from Nielsen Norman Group by (Jakob Nielsen and Don Norman), https://www.nngroup.com	54

3.3	Typical visualisation pipeline illustrates perceiving and understanding context as parts of generating visualisation. We added categories to highlight the required contexts, the image has taken from [36].	59
3.4	Typical visualisation pipeline stages. Synthesis ideas are taken from[163] the diagram created by the author.	64
4.1	Methodology of design Critical Design Sheet chapter.	67
4.2	Photograph of participants at the critical-analysis workshop, which was held to elicit vocabulary associated with critical analysis and to understand processes and stages of performing a critique.	72
4.3	Workshop task 3. (A) Minard’s map 1869 of Napoleon’s Russian Campaign 1812. Two dimensions map with six types of data. An example of a flow map. It is published in the public domain and cited in Edward Tufte’s book [156] p.40. (B) New production graphic of Minard’s map programmed with object-oriented graphic display system integrated with IML client-server extension of the SAS Interactive Matrix Language[61]	76
4.4	Workshop task 3: (c) Vincent Van Gogh’s sunflowers painting published in Wiki public domain. (D) An example of children’s drawing. The figure showed C and D painting, the participants should perform the possible comparison between these two paintings of low fidelity and sophisticated artist work during workshop time.	76
4.5	Workshop task 8: Bar chart graph created in excel by author. Participants were asked to critique this graph.	77
4.6	Workshop task 8:Multi-dimensional parallel coordinates visualisation shows nutrients content per 100g of food. The data published on USDA of Agricultural Research Service and the visualisation on food/nutrient-content-parallel-coordinates. The image from https://visual.ly/community/interactive-graphic/food/nutrient-content-parallel-coordinates	77
4.7	Workshop task 8: radar graph created in excel by the author. Open source data have used to create this visualisation taken from Global Mobile Consumer Survey(2014). The chart highlight iPhone’s users next phone chosen elements. Participants were asked to critique this graph.	78
4.8	Affinity diagram for extracting words of participant’s definitions. Each card includes the descriptive words that were mentioned in individual definition for the participant.	79
4.9	The refining outcome of words in figure 4.8 that is categorised under five ideas. Each idea has a title and detailed information of the words that reflects the idea.	80

4.10	The results of Task 2. Two word clouds of descriptive words associated with <i>critical analysis</i> are shown, sketched by representatives from both groups.	82
4.11	Descriptive adjectives sketched by group 1 as answers to task 3 (critiquing the six artefacts).	84
4.12	Analysis process for task 3 and task 4 to perform task 5 by writing explanations about similarities and differences for the given objects under four categories.	85
4.13	The critical thinking model, created and sketched by group 2, demonstrating four stages that simulates the four stages of Polya's [120] problem-solving model: (i) understand the problem, (ii) devise a plan, (iii) carry out the plan and (iv) examine that you have the right answer. .	86
4.14	Critical thinking model created by group 1, they performed task 7 as they attempted to build a visualisation critique process.	86
4.15	One of the principle parts of the Critical Design Sheet (CDS) are the six main categories of self-reflection: user's perception, environment, interface and movement, components, design and visual marks. These help the appraiser to make more focused critical judgements on the design.	92
4.16	(A)The first prototype of the Critical Design Sheet (CDS).	99
4.17	(B) The second version of Critical Design Sheet prototype.	100
4.18	Sentiment scores distilled form SENTIWORDNET3.0 for the 20 impression words.	102
4.19	The changes on star plot graph between the initial design and final developed design. Appraisers use this graph to represent the questions data and that gives them an initial decision on their design and how its looks.	106
4.20	The layout of the seven sections of the Critical Design Sheet (CDS). . .	107
4.21	The final Critical Design Sheet (CDS). A) Choose five "first impression" words. B) score the categories. C) Sum the values from the categories count the quantities of -2, -1 and 1 and 2s. D) Place the value of the sum on the scale. Add information about the user and visualisation to the sheet. E) Complete the star plot to show the breakdown of the parts. F) Reflect on the results, and what to improve.	109
4.22	insight vision of using CDS in the visualisation design environment. The diagram explains two main stages in visualisation construction, the design stage and the implementation stage and the role of CDS in each stage. .	110
5.1	The descriptive sheet that given for students as supportive analysis document helps them in critique design task.	118

5.2	Scenario 1 as experiment material: bubble chart visualisation created in Gapminder tool. (Image from the Gapminder tool, https://www.gapminder.org/ issued under the Creative Commons Attribution 4.0 International license).	122
5.3	Scenario 2 as experiment material: a pie chart created in Excel by the author displayed on magazine.	122
5.4	Scenario 3 as experiment material: a treemap D3.js graph displayed on tablet. Permission to use the image has been granted by the creator Anthony Ina (Viking Lab).	123
5.5	Scenario 4 as experiment material: a network graph displayed on newspaper. Permission to use image has been granted by the creator Cameron Booth.	123
5.6	Scenario 5 as experiment material: a time series graph displayed on computer screen. Image reproduced from the Dygraph “Range Selector” demonstrator, a D3.js library, licensed under MIT free to use, Copyright ©2009 Dan Vanderkam.	124
5.7	Scenario 6 as experiment material: a bar chart displayed on phone. Permission to use the image has been granted by Tom Gonzalez (Managing Director) of BrightPoint Consulting Inc. to use their Vizuly Column Chart image.	124
5.8	The final Critical Design Sheet (CDS). A) Choose five “first impression” words. B) score the categories. C) Sum the values from the categories count the quantities of -2, -1 and 1 and 2s. D) Place the value of the sum on the scale. Add information about the user and visualisation to the sheet. E) Complete the star plot to show the breakdown of the parts. F) Reflect on the results, and what to improve.	125
5.9	BP #1: Two samples of bad work practice. Participants selected just negative choices of over 30 questions of section (B). The star plot graph represents users selected values clearly.	128
5.10	BP #1: Two samples of bad work practice. Participants selected just positive choices over 30 questions of section (B) star plot graph representing users’ selected values clearly.	129
5.11	BP #2: Examples show bad work practice. Participants didn’t select the main group question [1,6,11,16,21,26] in section (B) star plot graph displays a discrete drawing line connection.	129
5.12	BP #3: Such users did not complete all CDS sections which reflect bad work practice. Two examples show missing sections: Left section (A) missing. Right section (F) missing.	130

5.13	Samples of users' good work practice. The left figure illustrates the complete critique practice user filled all the required sections with self-reflections on how improve that design. Right figure shows normal mistakes can be done by users through a critiquing task that actually considers good work practice. The score calculation mistake in section (c), change the selected choice in section (A) & (B) and user's hanging minds, all these correspond good practice which reflects their better understanding towards design.	130
5.14	The results of non-parametric independent samples test for group 1 and group2 (both sample size=10)	135
5.15	The mean values differences between users groups G1 and G2. The network graph shows a large mean difference and the bubble chart shows the lowest mean difference.	136
5.16	Categories data analysis for two users groups G1 and G2.	137
5.17	Parametric independent samples t-test calculation for new sample users of group2 with same group1 users. The test shows the significance p -value > 0.05 for all 6 hypotheses.	138
5.18	Word-cloud visualisation shows an analysis of impression words (section A). It shows such similarities between users' groups selections Pie, Bubble, and Bars charts shows the users' consistency in the use of most words, apart from spectacular, fulfilling and beautiful.	139
5.19	Analysis of word usage from section A) the first impression words, which shows a broad use of most words, apart from spectacular, fulfilling and beautiful. The graph is created in Tableau 10.2 software	140
5.20	The Five Design-Sheets of the student Abdul, who looked at the midpoint algorithm.	143
5.21	A single screenshot of a frame from the final animation of student Ali, who investigated how to create an explanation visualisation of the midpoint pixel algorithm.	144
5.22	The CDS sheets of Ali, who investigated how to create an explanation visualisation of the midpoint pixel algorithm.	144
5.23	Student critiques, showing how they have improved the visualisation, and have graded the CDS appropriately. It is clear that the left evaluation value before implementation is worse than that on the right (after implementation).146	
5.24	Student critique's work shows the improving of evaluation value before implementation (on left) and after implementation (on right)	148
6.1	The CDS online tool process. It describes the pipeline steps that users need to follow in order to complete their critique task, it starts from a login page and ends with storing critique work after getting results. . .	155
6.2	Home page for CDS tool website.	157
6.3	Login page for CDS tool website.	157

6.4	Registration page for new users to create an account for the CDS tool website.	157
6.5	The CDS online tool interface includes different components. HTML canvas, tables, buttons and radio buttons that are used to construct webpages.	159
6.6	CDS result page interface mapping includes different frames and canvases that can be filled with data after pressing the blue result button	160
6.7	CDS interface page illustrates an example of the user's critique work (it shows Paria). All sections are filled with required data and ready to save data then show results.	162
6.8	CDS result page interface mapping includes different frames and canvases that can be filled with data after pressing the blue result button.	163
6.9	MyRecords webpage displays users' data which is organised as records on the HTML table. The records displayed the information of user 'Paria' who made several evaluation works on specific dates, with data invoked from the MySQL database.	163
6.10	SUS usability test questionnaire to evaluate CDS online tool.	168

List of Tables

2.1	The table shows quantitative and perceptual visualisation metrics that applied on different stages of visualisation design process A, B, C, D or E. Several methods either mentioned that measures several aspects like SUS and EUQ	32
4.1	The User Experience Questionnaire set of questions, with the 7-Likert scale and alternating positive and negative rows, Laugwitz et al. [93]. . .	90
4.2	USER'S PERCEPTION GROUP group measures. With these questions users take a holistic view of the whole tool/visualisation or system that being create. users will need to project their minds into how they understand that design and it's purpose.	94
4.3	ENVIRONMENT group measures. Describes the design environment , users need to think of the appropriateness of where the visualisation will be used, the circumstances of its use and how a user would operate it or interact with it in that situation. It is all about appropriateness for the environment.	95
4.4	INTERFACE LAYOUT group measures. Reflects the general principals of visualisation layout design in terms of organisation or task suitability. . .	95
4.5	COMPONENTS group measures. Components are the units of design. They are often placed within frames or windows in an interface. Example components include bar chart, timeline, treemap, scatterplot, etc. They are unique and identifiable parts of the visualisation.	96
4.6	DESIGN group measures. Reflects user's thoughts on design principles, there are many good design principles to follow (e.g., see Munzner [107], Ware [164]), think CARP: Contrast (also, figure/ground), <i>align</i> items in a same way, <i>repeat</i> items, or styles throughout and be consistent. Put together things that are similar (or need to be compared).	96
4.7	VISUAL MARKS group measures. Questions that help users to think about the graphical marks that are used on the page, and how they are put together to make the visualisation. Are the right marks used, in the right places, with the right attributes?	97
4.8	The First Impression words, that were finalised. The user is meant to circle five words that best describe their visualisation to critique.	101

4.9	Various examples express the impression sentiments and word meanings. Users use these examples to help them reflect and make a critical analysis of their visualisation/design.	104
4.10	The Impression sentiments words classification based on positive, negative and neutral characteristic.	105
5.1	The internal-correlation of 30 users (N=30) users completing the categories on six different visualisations.	132
5.2	Independent samples t-test results that shows p -value at significance level $\alpha = 0.05$ and associative mean and SD for nominated visualisations (n=6) comparisons.	133
5.3	The First Impression words, that were finalised. The user is meant to circle five words that best describe their visualisation to critique.	137
5.4	Users time completion data to perform critique six different visualisations. Time taken for group 1 (N=10) and group 2 (N=20).	141

Chapter 1

Introduction

There is growing interest in the workplace for the need to develop more design thinking and problem solving, and critical thinking skills in the workforce. Implicitly this request puts pressure on teaching and academic institutions to teach and train the next generation of workers, to have the appropriate skills in design thinking, problem solving and critical thinking, necessary for the workplace. Indeed, this is not a new “call”. Back in 1988 Carnevale et al. [30] conducted a comprehensive evaluation of skills necessary for the workplace, and reported that “*employers want employees who can think on their feet (problem-solving) and who can come up with innovative solutions when needed (creative thinking)*” [30], then in 2013 they updated their findings to cover the modern digital age, writing:

*“Skills that process information and require sophisticated cost-benefit analyses such as critical thinking, complex problem-solving, judgement and decision-making are also highly valued. Critical thinking is a skill that is often touted by employers as a necessary requirement for success in many occupations. O*Net data confirm this assertion. It is very important in close to 20% of all jobs. Moreover, 96% of all occupations consider critical thinking to be either very important or extremely important to that job.”* Carnevale and Smith [31]

But what is design thinking? Razzouk [127] defined design thinking as “*an analytic and creative process that engages a person in opportunities to experiment, create and prototype models, gather feedback, and redesign*”. Workers develop these skills in their workplace and within their education. In fact, a broad range of academic environments do have a positive influence in developing appropriate skills necessary for designers.

It is given that students, throughout their education, are required to read, think and analyse reasons logically to solve complex problems. These key skills are all useful for designers. But, there is also a need to develop more specific “design thinking” skills in students. In fact, educators or teachers should support students in developing their skills for today’s workplace, especially design thinking, system thinking and teamwork skills [135].

One creative subject is visualisation. In fact, visualisation is an wholly suitable subject to investigate, because to achieve good visualisation outputs developers need to understand data, make critical decisions on what data is important to display, understand good design principles, ideate many alternative solutions, and make critical decision on those visualisation depictions. It is a subject that draws upon a wide set of skills. It is also well understood that visualisation is becoming more common in the workplace. This can be readily demonstrated by the rise of visualisation companies (such as Tableau, Qlik, Periscope, and the development of analytics tools in SAS, Excel, Google etc.) Therefore, visualisation is a subject that can be used to develop many broad skills in an individual, and it is essential in visualisation research and teaching focuses on developing users cognitive skills and design skills to help users create better visualisations.

To complement design thinking for visualisation design, we need users to employ *critical thought*. Critical thinkers seek to find alternative perspectives that they analyse open-mindedly and avoid hasty decisions [116]. If users are unable to assess the products they create and produce, and ascertain what to change, then they will not be able to improve their solutions. Critical thinking needs to 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. It expresses our human ability to make judgements and reason findings and ideas. Facione writes [53]:

“The ideal critical thinker is habitually inquisitive, well informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgements, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and their

circumstances of inquiry permit. Thus, educating good critical thinkers means working toward this ideal.” Facione [53].

However, this “ideal” is hard to achieve. It is difficult to be open-minded and unbiased, it is challenge to know how to organise the information appropriately to make the right decisions. Critical thinking and solving problems is not only a set of processes, but it also requires a more philosophical view to get the developer to think about wider issues, such as empathising with the role of an end-user of the system.

For data visualisation there are several places where the developer needs to think critically. When developers initiate several alternative designs they need to employ critical thought to decide which idea is best. When they complete the final design plan, they need to critically reflect over the design, such to work out how to build it and decide if it is fit-for-purpose. When the developer has built a prototype, they need to critically reflect upon it to work out how improve it.

As developers critique their work they are doing so against baselines. Perhaps they have seen a “good visualisation” by a famous researcher; they thus compare their work against this baseline. Perhaps they have an idea in their mind of what is good or bad. But on the whole, they are making decisions and comparing their work against other people’s work. In addition, they probably have an idea of “value”. Their judgements are valued, and placed on a scale. We can consider that they have an imaginary scale: at the top of the scale is the expert design, created by the famous researcher. Whereas at the bottom is a really bad design. They judge their design against this scale. To achieve this they may have an imaginary metric. A value scale that they create as a judgement of critical thought.

Generally, to achieve a successful visual design analysis outcome developers need to use a rich and diverse set of metrics. The benefit of using metrics in critical analysis is to gain insight design problems and know from where improvements start. Basically, *metrics* are an essential aid for the user to derive judgement on things that need measuring. Metrics could be values that are created by algorithms or formulas. Or could be imaginary (only created in the imagination of the user). Metrics can be presented as

numbers; these numbers represent some property and then can be used to measure or compare that property.

1.1 Motivation and Vision

The current situation is that, while there are methods to help people design visualisations there is little guidance, in the literature, over how to critique them. It is true that much “best practice” is described in the literature. For instance, teachers readily know not to use the rainbow colourmap, and there are many publications that explain why this is the case. Both Rogowitz and Treinish [133], and Richard and Elder [116] discuss issues with the rainbow colourmap, while Borkin et al. [23] look at an application using the rainbow colourmap and how its use could hide the right result in heart disease diagnosis. But while all this research is unquestionably worthwhile, and benefits society, it does not really help the learner. The learner not only has to read many papers, but they have to understand these challenges and then work out how to apply them to their work. What is required is a structure or process that will help organise the thoughts of the learner (or developer) to help them make a well-structured critique of a design.

Therefore the vision of the research is to investigate critical analysis, and develop something that will help learners structure their critical analysis.

In academic education, especially when teaching visualisation teachers need to help learners perform self-reflections on their own creations. Students need to understand what specific aspect of their design needs to be improved. They need to self-evaluate and investigate what is the part of design that is good and what part is not. This critical evaluation is not only suitable for visualisation.

Let us consider an “essay grading scenario”. In this scenario we can consider how writers (the students creating the essays) can critique or judge their own work, and decide if they are good or bad, and how to improve them. We could evaluate how someone does make a critical analysis of a written text. In this evaluation we may consider the different possible measures to evaluate language, cohesion, linking between paragraphs,

the flow of story and other things required to understand and evaluate the essays, etc. We could also consider the teacher (who grades the essay). How do they grade it? How they judge what overall score to give the work? And why he/she gave the essay this value? And especially how is the teacher managing to keep to a consistent grading scheme, and how can several markers look at the work to make the same (hopefully) grade outcome?

One solution could be to give the teacher a checklist. Academics already do this. They create a “marking scheme” for the work. This is a set of guidelines to help teachers grade the work. It is a set of criteria to help guide the teachers, and help them be consistent. It is a structure that is usually made of many small parts. It is a set of metrics to help them in the evaluation to make an appropriate judgement.

The “grading essay” scenario can be used as a metaphor. It is sensible to conceive that a similar structure would be useful in visualisation design. What the “marking scheme” provides for the teacher is an objective structure, it turns an activity (that is basically subjective) into something more objective. An essay grader, who is consistent and grades similarly to other markers, needs to move away from being subjective and using merely their personal opinions and gut instincts, into someone who is consistent and uses objective measures. So if researchers are to do the same for data-visualisation design, they need to be guided to move from being *subjective* on the data-visualisation design, to people who are *objective* in their thinking and methodological in their approach.

Furthermore, in the essay scenario we notice that most of these objective decisions output a metric value that are quantifiable and measurable. In other words, the outcome of the process is a grade. Something that allows all students to be ranked. In a similar way, it would be possible to imagine that critiques on data-visualisations could be ranked. And potentially outputs (visualisations) could be compared and ranked. Much like ranking students’ essays.

Therefore, from these motivations, it is possible to consider some important requirements that should be investigated in the dissertation. First, whatever the outcome looks like, it must have a clear *structure*. It must be something that can help a user critique their work in a systematic way. Second, it must be something that turns *subjective* decision

making into an *objective* task. Third, it will need to incorporate some kind of metrics. Values that can be evaluated, that will help to grade the work.

1.2 Hypothesis and Research questions

These discussions, and the essay critique scenario, start to highlight a set of important concepts that should be explored in the dissertation, including *critical thinking*, *metrics* in data visualisation, and how users make *critical decisions*. The following hypothesis is proposed:

Critical analysis is difficult. Users benefit from following structure. A structured approach, with a clear set of individual processes, will not only provide the user with a structure, but will enable their (often) subjective decisions to be divided into a specific set of objective criteria.

This subsequently raises several research questions: How do users critically evaluate visual designs? How do users reflect on their own designs (self-reflect)? Can we create an explicit structure for users to follow as self-evaluation tool to critique visual design? What metrics are used in visualisation, and where are metrics used in visualisation? We also can consider more general questions, such as what are other researchers discussing about critical thinking?

1.3 Aims and Objectives

The aim of this research is to discover visualisation methods that use metrics for better perform critical analysis of visualisation design. It will investigate how metrics can apply to different stages of visualisation. Also the research will explore evaluation methods for visualisations, and to what type of evaluation that help users to make self reflection on their creation in order to analyse and judge on their designs. The dissertation will investigate what useful structures can be provided to help users critique their work. This is then demonstrated through creating a Critical Design Sheet method

(CDS). Both *quantitative* and *qualitative* evaluations are used in creation and validation of our proposed critique design sheet (CDS) tool to critique visual design. The proposed method consists of six sets of metrics relating with visualisation design criteria to help learners critically analyse their designs.

These aims will be achieved through the following measurable objectives.

- Obj 1 Explore the breadth of the subject of “critical thinking” and “critical analysis”. Understand where it is used, and the concepts that are being discussed in the literature. Especially investigate the principles of critical thinking in regard to teaching and learning. Including how learners make critical decisions on their work.
- Obj 2 Investigate the role of metrics in visualisation and how the metrics are used in data-visualisation research.
- Obj 3 Discover how people perform a critical analysis. What methods do people employ? Perform quantitative/scientific statistical analysis of people’s critical thinking.
- Obj 4 Design a method(s) that could help developers make better design decisions on their creations and enhance the self reflection cues that add more rigorous thinking to their design analysis. Apply the method with users, in a situation that will require them to use it.
- Obj 5 Analyse the designed solution for its *usability* through people using the method and enhance the design as required through iterative design processes;
- Obj 6 Implement the solution into something that can be used by a wide range of people, to share and discuss the quality of different visualisations. Perform a usability analysis of this tool.

1.4 Scope of research

This thesis focuses on outcome of generating processes and techniques to aid developers of data-visualisations to critique and make judgement on their designs. Therefore, some parts of this work will be broad. Especially the subject of *critical thinking* is a topic that is not restricted to the visualisation domain and consequently the research contained within this dissertation will be likewise broad. Whereas, in other cases the dissertation will need to focus on very specific detail. For example, when discussing metrics in data-visualisation, it will be only the data-visualisation literature that will be explored. Where the research will need to focus on the main journals and publication venues of the visualisation community, such as IEEE Transactions on Visualisation and Computer Graphics, Eurographics Association's Computer Graphics Forum, IEEE Computer Graphics and Applications, as well as conferences such as EuroVis, IEEE VIS (particularly InfoVis and VAST), Information Visualisation, ACM CHI, and workshops such as BELIV, Pedagogy of Data Visualisation Workshop.

1.5 Research methodology and structure of dissertation

Within each chapter specific methodologies and processes will be followed that are appropriate for that type of outcome. These will be detailed as required. However the research, that is presented in this dissertation, follows two general principles. First a traditional research, discover and reporting methodology will be used, and second an iterative approach to design and development will be followed.

The traditional research methodology. There are certainly many related papers to read. Therefore, journal articles, conference papers, white papers and books will be searched, read, information summarised and then classified. For this work, several digital libraries and online resources will be referenced. These include Google.Scholar, ACM Digital Library, IEEE Xplore Digital Library, and of course the general search engines (usually Google) will be used to search the materials. This research strategy is demonstrated in the chapter on critical

thinking (Chapter 3) and the main related work chapter (Chapter 2 where different metrics will be explored and evaluated). This methodology is principally used in the first few chapters of the dissertation.

Iterative approach to design and development. It is clear that, at the start of the research, it will be impossible to get everything right the first time! The iterative approach will enable the research to start off vague, test the ideas with users, and then improve the designs such to strengthen the results. Consequently, all the outcomes of this dissertation, will have gone through many iterations. This is important because, through such an approach the research will have been tested with users and refined. The dissertation therefore includes descriptions of this process, and describes how the designs have been improved. This is exemplified in several chapters. For example, Chapter 4 explains that much thought was made into critical thinking, then a workshop was held to ask real users for their views on “critical thinking” and then analysis was made of the workshop results, before initiating the first design solution. Then that design solution was tested with users, changed and improved, and then tested again with users. This was all before giving the solution to a wider set of users (in our case, to a set of approximately 50 students on the third-year Computer Graphics and Visualisation module). This demonstrates the iterative nature of our approach. This type of iteration was used in the second, third and fourth part of the dissertation.

The dissertation is divided into four stages, as shown in the diagram in Figure 1.1. The first stage mainly uses research methodology, and includes Background and Related work, with the last stages using the iterative approach.

Stage A – Chapters 2 and 3 The first stage (A) contains background and the literature review (related work) in Chapter 2, and critical thinking is covered in Chapter 3, as follows:

- **Related Work (Chapter 2).** Research is achieved to investigate different types of metrics and how they are used in visualisation. We classify where metrics can be applied throughout the whole visualisation design process.

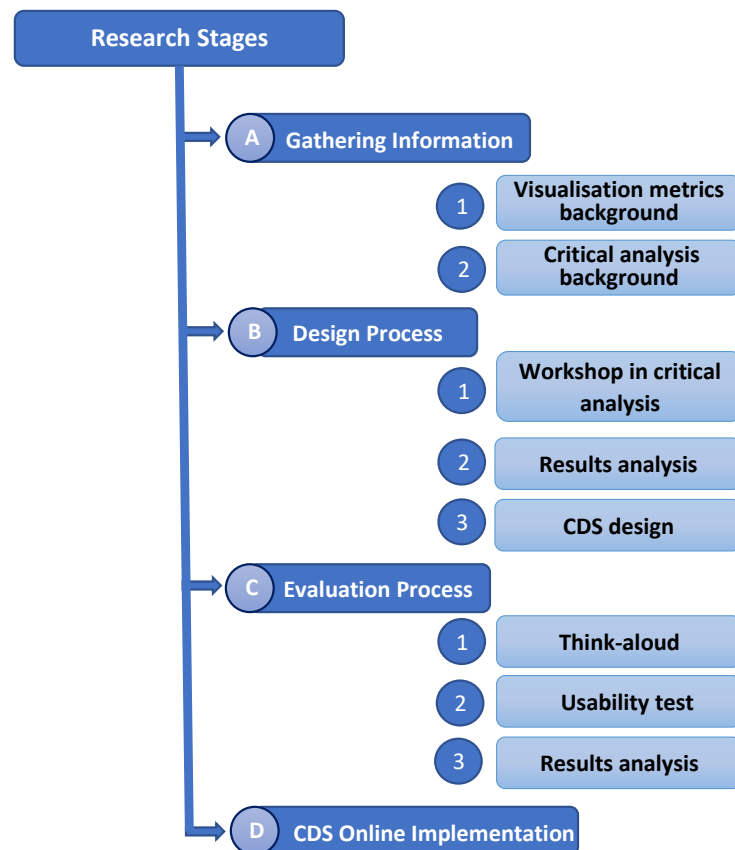


Figure 1.1: Research methodology. A diagram of the whole dissertation showing the four main parts (ABCD)

The output from this research is a new model for thinking about the roles of metrics in visualisation.

- **Critical thinking, (Chapter 3).** Related work about critical thinking is gathered. The content focuses on critical analysis, and critical thinking in different fields of study. This includes asking questions of “how people make judgements on things and what dimensions they used to elicit that judgement”.

The outputs of this investigation directly feed into the structure of the Critical Design Sheet (CDS). The critical thinking chapter provides the backbone of the categories and structure of the CDS. The CDS is designed and developed in the following parts of the dissertation (see Chapter 4 and evaluated in Chapter 5).

Stage B. – Chapter 4 This stage explains the design process, starting with the workshop analysis, and then developing different versions of the CDS. The idea of creating a paper-based version was decided through this process. It was a decision taken shortly after the workshop, and from reflecting on the workshop results. Chapter 4 discusses the whole workshop to design process, and covers the following information:

- **The workshop:** planning and preparation then conducting critical analysis workshop. Details of participants' background, demographic information etc. are explained in this chapter. The aim is to investigate users' perception and reflections on visualisation. Conducting the workshop enabled us to achieve the hypothesis of how users think on design, which us frames ideas about different strategies that users make as they perform a critical analysis of their visual design.
- **Results analysis:** this part provides an in-depth analysis of the outcome of the workshop. The aim was to explore the workshop data and provide quantitative analysis that can be used to help develop the Critical Design Sheet, and to support the design decisions that were made.
- **Design CDS method:** we design the Critical Design Sheet (CDS) method which guides users to reflect on different aspects on visualisation design. The sheet has six parts: metadata, first-impression words, a self-evaluation section of thirty quick questions (covering from usability and considerations on the environment where the tool will be used, to the visual marks that create the visualisation), a grading scale, a star-plot overview and a reflection section.

Stage C. – Chapter 5 This stage explains how the CDS was evaluated. The work is divided into three parts:

- **Think aloud sessions:** where individual participants were asked to think-aloud and explain what thoughts they were making as they work through the stages of the CDS. Each user was asked to critically analyse (the same) visualisation designs using CDS, and think “out loud” when completing all

sections on a sheet. This type of evaluation is important, because it helps us incrementally improve the design. In fact, after each think-aloud session the CDS was improved, and then another think-aloud session was held to see if more improvements were required. Full details of this process are included in the chapter.

- **Usability evaluation study:** A full usability study of the method was made. This included organising several sessions where Postgraduate students critiqued six visualisations. Each were timed, and then the data was gathered. In another session third-year computer science students were given the same task. Full details of this process are included in the chapter.
- **Statistical analysis:** Finally, a comprehensive statistical analysis was achieved on the results from the participants. This was performed to analyse and evaluate effects such as learnability, reliability and consistency.

Stage D. – Chapter 6 The final stage describes a computer program implementation of CDS. An online CDS website was created that mimics the paper-based version. The aim of this website implementation was to create an easier way to record the content of the CDS and automatically calculate the figures. The prototype was evaluated to explore its usability.

1.6 Contributions

The dissertation makes five novel research contributions, as follows:

1. **A framework describing the role of metrics in visualisation.** While researchers have presented and used many metrics, no one has looked how they are used in visualisation. This dissertation presents a new framework that explains where metrics are used in visualisation (throughout the whole process). Additionally this framework is used to classify the literature. A journal of this work is in preparation, with a draft title of “The Role of metrics in visualisation”.

2. **Workshop results and analysis of critical thinking.** The second contribution is the in-depth study and workshop of how people critically think. The dissertation contributes the results and analysis of a workshop investigating how participants critically analyse objects and images. The contribution is the underpinning knowledge and foundational work for critical thinking in visualisation design.
3. **The Critical Design Sheet.** No other researchers have developed a technique to help people critique their visual designs. This work could have significant impact, because it is a useful tool that can be easily deployed to students and can be used by companies, researchers or developers. A journal publication is in preparation, with a draft title of “*Critical Design Sheet: A framework to aid inspection and self-reflection of designs and visualisations*”.
4. **An in-depth analysis and usability study on users of the CDS** demonstrating that the tool is usable and is self-consistent.
5. **An online system implementing the CDS.** A prototype CDS online implementation has been created. A usability evaluation has been performed on the tool.

Chapter 2

Related work

This chapter is going to explore and investigate the types of metrics that have been used in visualisation tools and methods. The first part investigates the types of metrics and their role in visualisation design. This research directly feeds into the second part, where we describe the proposed metric framework, where the goal is to develop a metrology of visualisation. The third part then looks at the broader issue of evaluating visualisations. This is useful because the third part includes automatic evaluation (using algorithms and metrics), empirically (assessed by real users), formally (using formulas and usability measures) and informally (based on the subjective skill of the evaluator). The wider issue of evaluation is relevant to the dissertation, because it places the CDS framework in context with related work.

Part 1. Metrics

- Section 2.1: a brief introduction to metrics and their use in visualisation design.
- Section 2.2: reviews the types of visual metrics used in visualisation studies.
- Section 2.3: the role of metrics in visualisation design, and how it could be possible to work towards a metrology approach in visualisation, especially in the design process.

Part 2. Metric framework

- Section 2.4: the proposed framework explaining where metrics can be used in different parts of the visualisation-design process.

Part 3. Evaluation in visualisation

- Section 2.5: overview on evaluation methods and techniques in visualisation, as a motivation for the CDS and places the CDS in context with related work.
- Section 2.6: chapter summary.

A research methodology is used in this chapter, where digital libraries and search tools are used to examine Related Work. Many sources were investigated, in particular the main visualisation journals of IEEE Transactions on visualisation and graphics, Computer Graphics Forum, IEEE Computer Graphics and Applications. The work also looked specifically at the visualisation conferences of IEEE InfoVis, IEEE VAST, IEEE Scientific Visualisation, Eurographics EuroVis and the conference of ACM SIGCHI on Human Factors in Computing Systems. Information Visualisation, AVI (Advanced Visual Interface) and workshops such as Beyond Time and Errors (BELIV). Figure 2.1 shows the distributions of main visualisation journals that contributes towards building our research literature.

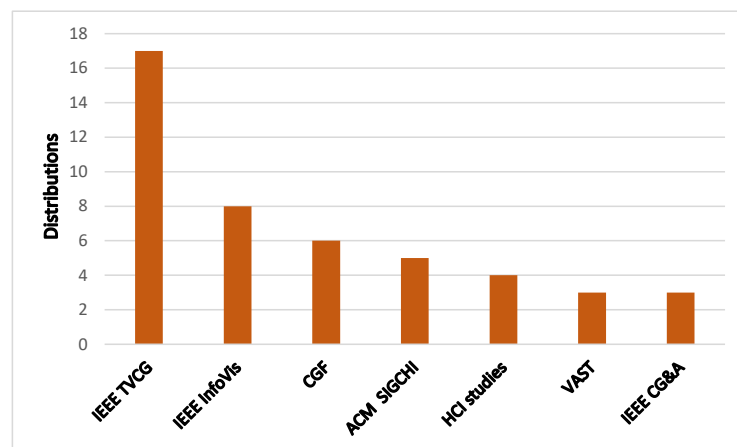


Figure 2.1: Visualisation journals distributions display that the IEEE TVCG journal has the maximum number of contributions in our research literature.

2.1 An introduction to metrics and opportunities for visualisation

A metric is something that gives a value that can then be measured. The Oxford English Dictionary says that a *metric* is “a system or standard of measurement”. In fact the word itself has its origins from the French words *métrique*, and from the word *mètre*, an English metre. But metrics are broader than merely 100 centimetres and it is a term commonly used to explain any form of measurement. The wider field of metrology, defined by the International Bureau of Weights and Measures¹ (BIPM) states that “*Metrology is the science of measurement, embracing both experimental and theoretical determinations at any level of uncertainty in any field of science and technology*”.

There is huge potential for users to use metrics in visualisation tasks such as exploration, by showing alternative statistical versions, automatic decision making for dimension reduction, and even for visualisation design to help users make better design decisions. Metrics could help users perform better and guide them to salient pieces of information. If guidance could be given to the user in the form of a standard set of criteria for “good designs” then developers would in fact create better designs. If there were agreed standards then visualisations could be compared, and everyone would agree that a visualisation was good (or not). It would also mean that people could be guided on how to improve their visualisation designs.

At this stage, it is worth making a counter argument. That, metrics are going to be impossible to create. This argument is much like the *nature* versus *nurture* argument in skills as humans grow from babies to adults. In other words, how much do humans learn, and how many skills are innate in the human makeup. The equivalent argument within the visualisation field concerns the discussion if visualisation is a *craft* or whether it is more about *science*. The main question, that is discussed in the visualisation community, is “how much is visualisation a craft, and how much is it based on theories?” In other words, how much is the design of a visualisation down to a designer’s instinct, their experience and the skills they have learned. Or how much is down to the knowledge that

¹Definition of metrology retrieved from Internet Archive, August 2017, <http://www.bipm.org/en/convention/wmd/2004/>

is accumulative and derivative from scientific theories? Where the theories represent aspects of the process that can be measured and quantified?

Norman [112] advocated a new concept to apply such scientific themes to the crafts. Norman introduced a cognitive engineering process that tries to apply what is known in science to design and machine construction prototypes. Norman's cognitive engineering concept is to explore then understand the issues and to show how to make better decisions if possible. His original central issues, about how people interact with machines, help address the principles behind human action and human performance which are relevant to the development of engineering principles of design. His cognitive engineering concept could be applied to a wider set of challenges. Indeed, the same principles can be used to build something, to help users question their own thoughts and reflect on their work, not just use the concept to describe user's interactions with machines which is used for designing support systems.

It is possible to see these two different standpoints by the titles of books from two different visualisation researchers and authors. On the one hand Schneiderman and Bederson titled their book "The Craft of Information Visualisation" [144], while Ware called his book "Perception for Design" [164] and wrote "This [Ware's] book is about the science of visualisation, as opposed to the craft or art of visualisation".

However these two standpoints can live together happily; they are arguing over different aspects. Ware agrees that it is difficult to scientifically measure aspects of pictures, writing in his book he says "*the issue of how pictures, and especially line drawings, are able to unambiguously represent things is still not fully understood. Clearly, a portrait is a pattern of marks on a page; in a physical sense, it is utterly unlike the flesh-and-blood person it depicts. The most probable explanation is that at some stage in visual processing, the pictorial outline of an object and the object itself excite similar neural processes*" [164]. In other words, our mind fills in the gaps. Because we are human, and all have different experiences, have different learnt knowledge etc., some things can be learnt in the design of visualisation, while others are innate. Subsequently there is a "middle ground".

This centre space, is an engineering viewpoint, which assumes that visualisation is a craft, influenced by human experience, accumulation and sharing of techniques and lessons learnt, which is based on scientific measurement. Consequently, some things in the design of visualisations can be measured: they are scientific, more nature oriented. Whereas other aspects of design are crafted: they are based on experiences, and knowledge and can be learnt and improved. We can imagine, on the craft side thinking about “personal experience and knowledge”, but when we move to become more scientific we can replace them by “design principles”. Likewise change a practice of “invent and test” by the practice of “specify then implement”.

The discussion about nature versus nurture, sensory versus arbitrary etc. are all relevant to the dissertation, because if the goal is to look to metrology and metrics, then it would seem that the direction of the dissertation is travelling in, is towards principals-driven design underpinning cognitive engineering, and away from craft. However, this is not the aim. The aim is to use science to help craft be better. The aim is to create a tool that will help people produce better visualisations.

Essentially, by considering *nature* versus *nurture* and visualisation design as *engineering*, we are considering visualisation in terms of “cognitive engineering”. Cognitive engineering as a field developed after the Three Mile Island in 1979, which demonstrated that the interfaces (in this case with Nuclear power plants) need to consider a wide range of states, human control, and enable users to perform proper and appropriate actions [45]. The design of the system must consider the goals of the user, the limitations of the environment, and any limitations of the systems involved, and be suitable for day-to-day users, rather than “ideal” users in a lab. It must be something that is suitable to be used in “the wild”. First, best design is iterative, and allows early versions to be trailed and tested. Second, cognitive design needs to allow users to critique the systems, such to assure that effective systems are created. Third, as people learn how to develop systems effectively they will accumulate knowledge and understanding of “best practice”, which means thinking about, and considering, alternatives. These requirements push cognitive design towards being an advanced engineering discipline. Furthermore, Dowell and Long [45] assumed that such conceptions of cognitive engineering principles would enable the formulation of cognitive design problems, improve the efficacy of cognitive design and

cognitive design would be improved through design exemplars related to the re-use of knowledge.

Because we are assuming that some aspects of visualisation design can be measured, we can now start to ask “what can be measured in information visualisation and what cannot be measured?” This idea is specifically addressed in Section 2.4, where we place the metrics on different parts of the visualisation-design pipeline. However, this question can be sub-divided into several similar questions. For instance, how much of visualisation design is based on “sensory” or “arbitrary” symbols? This discussion is really one of *semiotics*. How much is created by the developer arbitrarily and how much does visualisation rely on the human sensory system, such to understand the graphics. This question is explored by the domain of *semiotics*. Bertin’s book “Semiology of Graphics” [16] is the classic book that covers this information for the data visualisation community. Humans can understand colours, patterns, shapes, and use their knowledge of the physics of the world to understand of all the sensory input that the brain receives from all the body sensors (including eyes, ears, taste, touch etc.) [164].

Visual metrics have a potential to reduce the workload for the user. Let us imagine a situation where a huge quantity of data is displayed to the user. Metrics could be used to analyse the data, work out the spread of values and display to the user the significant ones. The outcome of the metric would be to reduce the quantity of data that a user needs to view. This would have the positive desired effect that the user could spend their time wisely only on focusing on the pertinent data points. It would support the user to perform a task quicker, and do things that would be difficult or time-consuming to perform.

Additionally, we believe that metrics can be useful in the breadth of the visualisation design process (not only recommending visual depictions to the user). In the data visualisation domain various metrics have been proposed and applied at different visualisation stages. For example, we can consider metrics being applied on *input* data, on the *filter* process, on the *mapping* process or on the output stage. In this case, data metrics would be needed to analyse the data along with other metrics to measure the effectiveness of different visual encoding. Personal visualisation is another area that is growing fast and can benefit from metrics. Because humans have different goals

and tasks, the system would use metrics to suggest what type of visualisation would be suitable for you to see. Maybe the user has a preference for scatter plots over line charts; then the smart visualisation tool will concentrate on the results in that type of display. Metrics would be required that calculate values of personal preferences and then match them to appropriate visualisations.

2.2 Review of metrics in the visualisation literature

Researchers have created and used a variety of metrics to help users perform better at a variety of tasks. One well-known researcher, who investigated various metrics, is Edward Tufte [156], who presented three useful metrics: Lie factor, Data Density, and Data-ink-ratio. The Lie Factor is the relation between the size of the effect in the graphic domain in comparison to the size of the effect in the data domain. It would be especially noticeable in graphics, where the data is projected into three-dimensions, or a linear effect in the data is displayed by a non-linear graphical projection. In fact, the lie factor measures the amount of distortion in the graphic. Data Density evaluates the quantity of the data points displayed in comparison to its size, while the Data-ink-ratio explains the ratio of data ink to the total ink. Tufte explains several principles to help developers create better visualisations, including maximising the data-ink ratio, removing non-data ink and reducing any redundant data-ink.

These principles need to be interpreted by the designer. Increasing the data-ink ratio is useful and this can be achieved by either increasing the quantity of points or decreasing the size of the graphic. But if the graphic was shrunk to a very small size, while data-density improves, it would be impossible to understand the data in the visualisation was very small. Likewise removing all chartjunk is not necessarily suitable. So-called chartjunk can be beautiful, it can enhance the appearance of the information and make it more attractive, and help users remember the visualisation [22], [164]. Brath [26] expanded Tufte's density metrics to a proposed cognitive complexity overload for 2D and 3D graphical scenes, identifiable points percentage and occlusion percentage. In fact, Brath discards Tufte's assumptions of "less geometry is better". With a cognitive complexity measure, for instance, if a visualisation has 50 or more dimensions with

weak correlations it is difficult to realise a design that will show all the data effectively. Any result would consume higher cognitive efforts. He suggests that a user should redesign the visualisation and place all data items that correlate to the same dimension in the same area to generate independent multiple views of visual components. Such a display would be better than one highly complex visualisation that showed all the results in one display. The aim should be to refine geometrical redundancy for high dimensional visualisations that result in better and more understandable data representations that have a lower cognitive overload.

High dimensional data visualisations are difficult to create, and metrics can help to represent understandable patterns for visual information. Bertini et al. [19] presented an automatic quality metric pipeline for the analysis of visualisation techniques especially for high dimensional data. This processing pipeline can apply metrics both on data space and image space, Bertini proposed clustering metrics to capture a grouping of data on data or image space. The correlation of metric works on data space captures elements that behave differently on large amount of data (e.g., outliers). This includes a complex pattern metric which recognises items that are difficult to detect on data. The feature preservation metric focuses on image quality which compares between features in the data and representations in visualisation in context to preserve as many interest features as possible. These quality metrics support the exploration of high-dimensional data sets.

The reduction of clutter is another goal, that researchers usually try to minimise, especially with high dimensional data sets. Alex Peng [118] presented a clutter reduction metric based on reordering dimensions on a parallel coordinates visualisation technique for car data sets. The goal is to reduce clutter to generate better views to understand and discover the structure of displayed data. He experimented to apply a clutter measure on distinct visualisation techniques (scatter plot matrix and stacking visualisation) to obtain less clutter displayed using a reordering dimensions feature. Bertini [18] used a similar metric for clutter reduction, but it was not for high dimensional data sets, rather his research focused on reducing visual clutter on two-dimensional scatter plots. His aim was to minimise overplotted pixels by using a sampling density technique for a data set of (160,000) data items. While these researchers used different

techniques to minimise clutter, their goals remained the same: to present better views of the data.

Metrics can be used to reduce dimensions for high dimensional data sets. These data sets might consist of an enormous quantity of non-important variables, to hide these variables from final rendering view considers ideal feature. Johansson et al. [80] displayed hundreds of variables in one structure and considered it a major problem in high dimensional data set and that it might be a better solution to apply a user defined metric. This enables an individual to explore and visualise multivariate data sets effectively. Systematic dimensions reduction with the weighting function of Johansson enables for an investigation between the number of variables to keep and the number of variables to lose. This provides an automatic ordering for important variables with user controls for a flexible dimensional reduction technique. This proposed metric helps to obtain the best view structure for ordered variables in a high dimensional data set.

It is important in information visualisation to provide a clear picture and scene for a rendered structure. Various visualisation techniques use screen-space metrics in order to diagnose visual problems regarding overlapping points (in, for example, scatter plots) and congestion of polylines (in parallel coordinate plots). These problems occur more frequently in large data sets, where it is hard to find a clear view for data structure. Paragnostoc [43], Scagnostoc [167] and Pixnostics [43] techniques involve different metrics that help to mitigate these problems by providing users with several multiple views of the rendered structure ranked by the user's preference. A screen space metric with these techniques has limitations, users should have knowledge of the screen size and quality of display to calculate these variables when the display is re-sized. These techniques support a better understanding for information, for example, using a Scagnostics approach in weather data helps navigators to show changes for barometric pressure versus air temperature [158], this approach improves performance to reveal hidden structure in large data sets [42].

Various methods are proposed to measure complexity, yet most of them point represent a subjective view in complexity. Duffy et al. [47] proposed a model to measure cluster-based complexity in parallel coordinates and scatter plots to help in quantifying estimation of the degree of difficulty and to interpret representations correctly. His

proposal measure was based on a symmetric isomorphic for cluster topology [46], the cluster topological properties include the existence of overlapping regions, splitting regions and meeting edges, while Rosenholtz [134] defined complexity as the amount of visual clutter based on congestion features on an image space. So, measuring clutter as a metric has many directions and implementations in visualisation. Several techniques have been mentioned for clutter reduction in this review. In cartogram visualisation, complexity relates to the efficiency of map rendering and complexity of regions. Alam et al. [2] measured a rectangular cartogram in terms of maximum and average polygonal complexity of the map regions. However, detecting the actual degree of complexity or uncertainty is an ill-defined² problem and suffer from a lack of tools in visualisation applications to precisely quantify overlap and occlusion. Cartogram visualisation is a kind of map presents data as geographical distribution, which needs to represent the numerical value for concerned area. To control the representation of data that requires an relative metric to preserve the shape of the extent value. There is set of quantitative metrics proposed to measure the given desired data value that would be represented on the map and estimates the readability of the final representation (see Tobler et al. [153]), and Alam [2] also proposed some cartogram metrics.

Graph drawing uses metrics to layout graphs. Graph aesthetics not only implies “how beautiful the graph” appears, but affects readability. So aesthetics is also a measure of how “easy” it is for a human to read and understand the graph with respect to the continuity of the paths which is the main condition in drawing graphs. Helen Purchase [123] suggested computational metrics which can be defined as objective metrics in graph drawing. These are used to aesthetically measure different graphs, looking to maximise symmetry and minimise edge crossing and edge bends. Purchase’s proposed metrics corresponds to objective computational metrics for graph assessments to some extent. Researchers use aesthetic measurements to make comparisons between different visualisations and versions of graphs. Various aesthetic metrics exist, such as the amount of symmetry and bends of edge crossings in graphs.

Ware [165] considered a human cognitive model, to make judgements and assess the aesthetics of different graph layouts. Ware proposed a cognitive cost measure, which

²An ill-defined question is one that does not have a specific, clear and certain answer. For example, 2 plus 2 is 4; it is well-defined, known by five years old, whereas “draw a picture” is ill-defined.”

assesses the how the graph lines are short, straight and on continuous paths. If graphs involve these metrics, then the resultant graph is considered to have “good characteristics” and help users to perceive that the graph has a high readability. He conducted a series of empirical studies to assess the effectiveness of cognitive cost measures and the results revealed that shortened paths with minimum edge crossing between two nodes, can be perceived as a low cognitive cost, and therefore represented a good aesthetic measure for graph layout. In the same direction, mental effort metrics hold similar aspects but they measure the effectiveness of the graph visualisation in terms of cognitive load [72]. Evaluating the performance of the graph requires that the developer uses effective metrics. Likewise, this is necessary to make judgements regarding visualisation quality, for instance Paas et al. [114] stated that using a subjective rating scale is suitable to gain reliable results, when evaluating the amount of mental effort that is required. This evaluation could be either performed by a numerical rating scale, or employing verbal labels such as “very, very low mental effort” to “very, very high mental effort”.

2.2.1 Decision making and directed thinking

To realise the importance of creating a data visualisation model with the presence of metrics we need to investigate the design process and how metrics could be integrated within this process. Metrics can help users make decisions (and especially better decisions). Therefore there is a need to consider the decision making process, and how metrics can be integrated into such systems.

The decision-making process employs rational thinking to consider alternatives and decide on the best action [131]. It requires the developer to use different types of thinking. These include directed-thinking because there is an ultimate goal to display some data; indirect-thinking, because there are several ill-defined problems to be solved, each with unclear goals; and creative-thinking, where developers use best-practices to design something new [63]. This problem-solving activity thus continues until a satisfactory solution is created. Developers work through different design ideas and potential solutions, before deciding on a method that they think is going to be the most effective. It therefore relies upon human judgement to make a rational consideration of alternative paths and different situations, to decide on the cause of action. It is important

to note that many of the decisions that are made are *judgements* and that creativity in visualisation is an ill-defined problem, where there is no one answer (but multiple, equally equivalent solutions, exist).

The visualisation creative design process relies on explicit and tacit knowledge. It depends on many factors, including the users' knowledge, understanding of the data, understanding of the visual encoding systems (retinal variables [164]), the environment of the user and where the visualisation tool will be used. Some decisions are small and have little impact, while others can be game-changing. For example, choosing the font to display the text is probably a small decision that has little impact. It can be easily changed in the software. Whereas deciding on the number of prototypes to build, is a small decision that has a huge impact, because each prototype needs a lot of money and time to create. Decisions are also made at different times. Questions such as “what is the best visualisation style for my data?” are asked at the start of the process, when the user is faced with some data. Whereas “is my visualisation effective?” is a decision that is made after the design process has been made. However, to use metrics effectively in visualisation we should consider the detail of their potential use. Through such an examination, our aim is to help new users understand the possibilities for incorporating metrics with their visualisation design process; help developers create better codes by knowing what to consider and how best to apply different metrics; and pave the way for researchers to create new metrics for visualisation.

We have mentioned such notable quantitative metrics for visualisation in this review. Which they enhanced quality and performance of visualisation techniques. The next section will discuss the influence of visual metrics on the visualisation design process.

2.3 The role of metrics in visualisation design and working towards a visualisation metrology

Metrology is the science of measurements. It includes theoretical and practical aspects of measurements and its application to different domains. Such measures are commonly well known in the physical sciences. Measuring the attributes of objects is subjected to

some standard measures (e.g., magnitude, dimensions and uncertainty). For example, colours can be measured based on the wavelength of lights (objective) and they can be described by “names” (subjective). The perception of these colours can depend upon many things, including the properties of the materials, colour of the light source and the user themselves. Each measurement has units which refer to the object type (e.g., time, weight and distance). Metrics can be assigned to a single value or to a range of values. So, visualisation designers, as with other fields, can quantify different aspects of the visual scene by metrics. From this viewpoint visualisation metrics can be defined as the metrics that are calculated to measure the attributed and capture the properties of visualisation to extract the meaningful information of data. This also implies that visual metrics can be classified into quantitative and qualitative measures, and considered by a variety of visual attributes from visual design.

To consider the components of data visualisation metrics, we look at how artificial intelligence breaks down decision-making into parts. In fact, intelligent agents enact on a particular environment (in our case this is the data-visualisation creative process). These agents *perceptualise* their environment into internal forms (metrics, timings, decision-trees etc.) and then take *action*. The useful intelligent agent mnemonic PEAS helps us consider the *Performance* metric itself, the *Environment* where the agent is applied, *Actuators* or output of the agent, and *Sensors* as inputs to the environment.

In our model we consider the Sensors as inputs to different stages of the dataflow model (**Dataflow**), we name the actuators **Metric types** as we focus on the type of the metrics that are created. We consider the Environment part of PEAS to be the whole process, and we label this as the creative **Design** process. In addition, we add two further parts: **Purpose** and **Properties** of the metrics.

The components are presented for this metrology as a schematic diagram, shown in Figure 2.2. It draws on our knowledge of best-practice of designing visualisation [131], our experience with creating and using metrics for visualisation and ideas of decision-making and the field of artificial intelligence. The five parts are as follows:

1. metrics in the **Design** process,
2. the **Purpose** of the metric and its quantities,

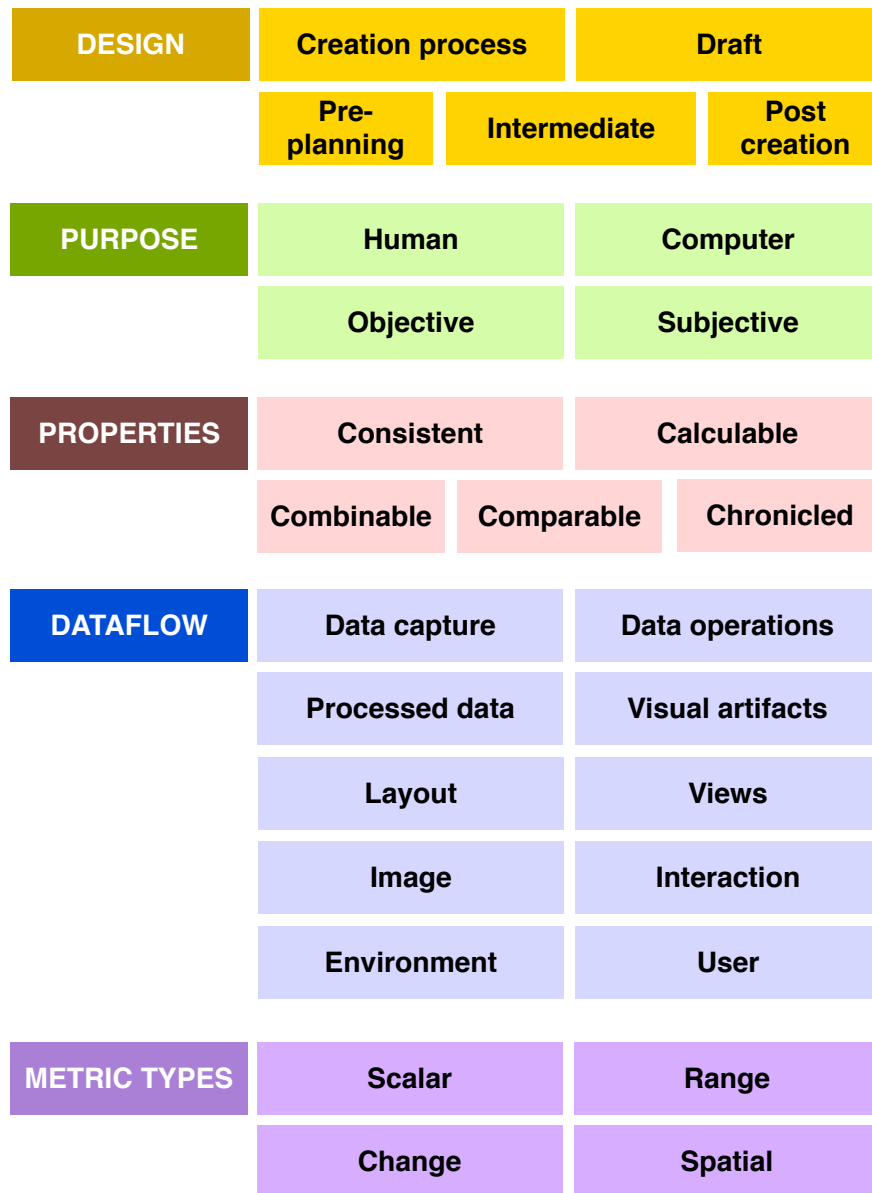


Figure 2.2: Towards a visualisation metrology. Think about the creative Design process, Purpose of the metric, its Properties; the different inputs within the Dataflow model, and different Types of the output (Actuators).

3. the **Properties** of the metric,
4. characterisation of **Dataflow** – the different inputs, that are used to create different metrics,
5. characterisation of **Metric Types** – the types of output of our metric functions.

2.4 Proposed metrics framework for visualisation design process

One of the aims in this research is to see the effect of metrics on visualisation design process stages and create new metrics if required in order to develop the design process. Therefore, to create a holistic model of metrics requires consideration of where metrics can be applied. This needs to create a sketched plan of the visualisation, to explain the process of analysing data, to develop a code, to evaluate a visualisation. Therefore, it is essential to create a whole system model of metrics for visualisation (WSMM). A system that considers every aspect of the visualisation design, build and evaluation process.

To achieve the WSMM, we divide the parts into five manageable stages. Each of the stages has a specific purpose. Figure 2.3 shows metrics model stages: (A) pre-planning, (B) visualisation creation, (C) draft, (D) intermediate and (E) post-creation metrics. This model has been created after much thought and deliberation. In an earlier model parts B and D were the same module, but we realised that it did not fit all scenarios, therefore they were separated into finer parts.

It is a conceptual model that explains where the metrics are used, and describes different places in the engineering of visualisation, where metrics can be applied. It also demonstrates that metrics can be part of an iterative system, where (for instance) the metric can be used to make refinements of the algorithm that terminates when the goal is achieved. Consequently, visualisation designers, as with other fields, can classify different aspects of the visual scene by metrics. From this viewpoint visualisation metrics can be defined as the metrics that are calculated to measure the attributes and capture the properties of visualisation to extract the meaningful information of data. This also implies that visual metrics can be classified into quantitative and qualitative measures, and considered by a variety of visual attributes from visual design.

To see how WSMM can be applied, let us consider a scenario where a developer is creating a visualisation. This engineer will follow a similar set of decisions. The developer would create several prototypes. Metrics could help to plan the prototype

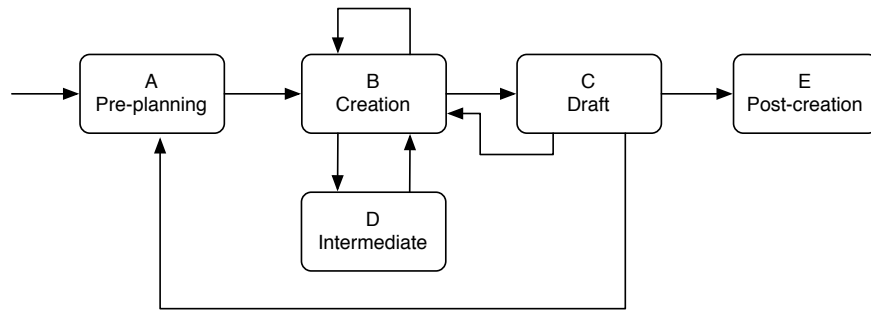


Figure 2.3: The Whole System Model of Metrics (WSMM) for visualisation. This model is used to decide where the metrics apply to visualisation; e.g., whether they apply as part of the algorithm (such as incrementally adapting the result) or as a post-creation process, to evaluate how well the visualisation could be perceived. This model works as a conceptual model, it can help to categorise metrics and help the developer consider what type of metrics they may use.

(maybe evaluating different programming languages). Other metrics could be used as part of the layout algorithm (such as calculating edge crossings) by generating a draft result (D) and then ascertaining whether it has met the goals. Metrics can help the developer ascertain whether the Layout and positions (say) of the visual artefact follows best practice (C). While finally metrics could be used to evaluate the effectiveness of the completed visualisation (post creation).

This model can be used for other creative processes. For instance, sketching different possible designs is a good way to plan how to visualise the data. Following the Five Design-Sheet (FdS) methodology [131] developers make initial sketches of potential ideas, then develop three (or so) principle designs, followed by a final realisation design. Metrics can be applied to this visualisation design process. Metrics could be applied to the pre-planning stage (part A), e.g., the quantity of designs or the neatness of the sketches. The metrics could help the user decide whether they have enough designs. The middle three designs could also be helped by metrics (part B), while the final realisation sheet depicts the final design idea in part C. Users could critique the final design and allocate a quality metric. The developer would then need to decide whether they accept the design (goto E), re-adjust the full concept (and goto D) (ascertain how to precisely re-adjust and then goto B), or start from scratch (goto A).

- **Pre-planning.** In the planning stage metrics can help developers analyse the setup and demands. This could be to ascertain how to best display the information; work out the task of the user and the goals of the system. Sketching, drafting and wireframeing are all parts of the pre-planning stage. Metrics at that stage

could be applied as qualitative or quantitative forms. For example, in the FdS methodology [131] developers make initial sketches of potential ideas to explore and discover alternative solutions. They start with forming potential ideas that will mature through the act of sketching and thinking until they get many design possibilities and solutions. They are selecting or defining three potential ideas from many generated ideas that will develop into the principle designs. This work can refer to user-defined metrics of selecting the quantity of designs and even the quality which is probably the neatness of the sketches (see figure 2.4. Otherwise, metrics in the pre-planning stage could calculate data quality, errors, missing data etc. Database management system (DBMS) is a software for managing data where queries can apply to offer flexible processing to information. However, many examples can be addressed about DBMS related with types of operations performed for preparing data (filtering and cleaning records or tables), since the focus of the framework is on preparing information particularly with the visual design process. In that case, in visualisation exploration in large data sets is considered a challenge. There might be huge erroneous values and complex relationships which are difficult to explore and generate views [130]. There is a need to prepare data by cleaning it at this stage. Data mining techniques might give great benefits to help users in this direction before rendering visualisation.

- **Creation process.** In this part metrics may mainly be used within the core algorithm where metrics values are calculated and quantified. The creation process is where the data flow model fits. Many of the algorithms that use metrics within the core algorithms are incremental or recursive. For instance, in graph drawing the simulated annealing algorithm would be classified as B, but metrics that evaluate the intermediate results are at D. The iterative operation is needed to verify the results by applying metrics values in order to improve the visualisation process. Users can also examine the intermediate results and decide when to stop. In fact, often another metric is used to stop these incremental algorithms. (Note, we classify such metrics that halt the iteration at C). Table 2.1 shows the utility of metrics on parts (B) and (D) of visualisation techniques that are used to enhance the performance of visualisations. For instance, arranging the similar dimensions behaviour in multi-dimensional data sets is crucial for the effectiveness of a large number of visualisation techniques such as parallel coordinates or

scatterplots. The similarity measures determine the partial or global similarity of dimensions that involve working on prerequisites for finding the optimal one or two dimensional arrangements [7]. The visualisation algorithm creates and calculates similarity measures in part (B) of design process framework in a recursive process then identifies the optimal one or two dimensions using an arrangement algorithm (such as a simulation annealing or ant-colony optimisation) part (D) intermediate process until the technique gets good correlations that can be identified easily and the significant amount of nearly lines between the corresponding axes indicates a similar behaviour, for example the stock exchange database, the technique visualised similarity in clusters of stock prices over the same period of time through 20 years. WSMM can classify the use of metrics over the visualisation design process, subsequently we can follow FdS methodology and going with metrics pipeline. After prompting the best three ideas, users will generate 3 sheets in that creation part which allow for the three ideas to be explored and create many details, figure 2.4 illustrates FdS methodology over metric framework. Thinking in creating storyboard visualisations for the potential expanding of each design detail is quite similar to how a visualisation algorithm has automated perform tasks.

- **Draft.** This part explains metrics that provide judgements over the completion of the creation process. Do you do a re-design (if so, go to B), or do you need to consider your goals at the start of the process (if so, go to A). This part provides feedback over the created visualisation. Intermediate result part. This defines metrics that are used to adapt the main outcome of the algorithm. Continuously with creation design with FdS, this stage is meant to apply metrics on the draft of intended visualisation, when users create sheet 5 (realisation sheet) they must decide on which design of three ideas should be taken forward. It is similar to applying a conceptual filter metric to select ideas to be developed further and add enough information to judge that a tool fulfils the goals of the design. This stage of the model is mostly similar to situation awareness as the evaluation area in the visual analytic environment [139], where the outcome methods examine the outcome and check the performance [122]. Was the correct decision made in that stage of the process? Automatic examinations or operators should be able to have decisions to complete the process with an appropriate set of evaluating

metrics. For instance, it may be possible to create a rendering of 100 different parameterisations at the same time. Then a quality metric is used to calculate which is the best one to display. Here the iteration is part of the algorithm. This is different to C, where iteration is at a smaller scale. Many of these algorithms are visually led, i.e., an appropriate metric is created to mimic human eyesight. Many of these metrics speed up the process; where it could be possible to use a human, it is just more efficient, accurate and better to create an algorithm to do it, and it is important for enhancing the creation of good visualisation designs.

- **Post creation.** In this part metrics are applied when the creative process has finished. This could include user-evaluation, or longevity studies. In some cases the outcome of systems or algorithms need to be re-designed, in other cases the company wishes to create the next generation of tools. For example, in software engineering domain sets of metrics applied after the designing and creation stages of modelling new software. That is called the test stage to validate the usability of that system, which could include measuring errors or system usage metrics. The result of this process may be to go back to one of the previous stages and re-design the visualisation. Such evaluating systems such as SUS, UEQ can be used in this stage which already includes a set of measures for evaluation purposes. For example, after users generate their designs with FdS they need an evaluation structure that helps them evaluate different aspects in the design. The categories of metrics for the CDS method can fit with part (E) post creation of the system design process. In this case, metrics mostly related with the development of visualisation design can apply. Figure 2.4 shows the how metrics of these evaluation systems can be applied on the last stage of the framework.

Table 2.1: The table shows quantitative and perceptual visualisation metrics that applied on different stages of visualisation design process A, B, C, D or E. Several methods either mentioned that measures several aspects like SUS and EUQ

Metric / Technique	A	B	C	D	E	Description
Lie factor			●			Measures the effective displayed graphical area in 2D graphs [156]
Ink ratio			●			
Occlusion percentage		●				Measures the occluded points in 3D image [26]
Identifiable points		●				Measures the number of visible data points
Over-plotting		●	●			Measures the density of collision points for given area in pixels [18]

Complexity measure		●				Simplistic measure counting geometric primitives, triangles, cell intersection on isosurfaces [32]
Clutter reduction		●	●			measure to reduce clutter in Pc, SP matrices, star glyph techniques in context of reordering dimensions [118]
Clutter reduction		●				Measures the outlier in context of occlusion in PC using sampling technique [50]
Cluster complexity		●				Measure for estimating visual complexity in PC, SP in the context of cluster visualisation [46]
Correlation measure		●				Measures the correlations, outliers between adjacent axes in PC using user's defined weight function to select dimension numbers [80].
Outlier detection		●				Measures cluster quality for every low dimensions projection [80]
Cluster detection		●	●			
Edge crossing & orthogonality			●			Geometric measures for assessing aesthetic quality of graph drawing [123]
Symmetry			●			
Memory load			●			Cognitive cost measure counts time consuming for continuing paths of node-link graph [165]
Class density		●	●			Measures the density of SP clusters data to gain low overlapped and separate classes [150]
Line crossing		●				Measures the quality of PC visualisation to optimise screen space view with respect to screen resolution for dimensional datasets [43]
Angle crossing		●				
Overplotting		●	●			
Cartographic distortion		●				Quantitative measures to assess the correctness of cartogram representation with input statistical values [2]
Outlier detection		●	●			Measure the distortions on PC intersection polylines on screen space view as abstracted data structure [79]
JND			●	●		(Just Noticeable Difference) perceptual measure to assess the correlations in SP in terms of how it's precision and accuracy [128]
Code lines					●	Measure the software maintenance adaptive efforts [69]
Cluster similarity		●	●			measures global/partial similarity between adjacent dimensions in PC by rearrange similar ones to enhance clustering [7]
Class density		●	●			These measures used to measure the quality of rendered data structure [149]

Histogram density		●	●			in context of correlations and clusters separation quality for PC, SP in high dimensional datasets [149]
Similarity measure			●	●		
Overlap measure		●				
Cluster density			●	●		Used to rank visualisation based on cluster properties on RadVis, SP, pixel-oriented visualisations [3]
Class consistency			●	●		Measures the consistency in clusters data variables in SP of multi-dimensional data [146]
SUS					●	(System Usability Scale) Measures systems usability in context of user's satisfactions as subjective rating scale [29]
UEQ					●	(User Experience Questionnaire) Measures users experience of interactive products [93] and [http://www.ueq-online.org]
CDS	●			●	●	(Critical Design Sheet) Measures users reflections on visualisation design as quantitative evaluation value
FDS	●	●	●			(Five Design Sheet) methodology uses for sketching possible design ideas. Users apply different qualitative measures to select best solutions that takes design forward [131]

PC = Parallel Coordinate, SP = Scatter Plots

Table 2.1 addresses various quantitative metrics in visualisation. We investigate visualisation literature carefully to find visualisation techniques that use quantitative metrics to support and enhance the outcome of these techniques. It is important to create a holistic view of how proposed evaluation methods are suited to the metrics model. The table presents where the visual metrics can fit in a visualisation design process, and a description about the context of what technique has been used and what visualisation has developed after using that metric.

Potentially, developers of visualisation tools can use the metrics in different parts of their code, such that they create higher performing visualisation tools. It is beneficial, when developers plan to create a visual tool in software, that they divide tasks into smaller functions, and follow professional coding processes. This is a divide-and-conquer approach. Consequently, developers should consider (in turn) how metrics can be used to prepare data, how they can be used to best organise data, then create classes to manipulate data and then visualise the output of the tool. At each stage, the developers

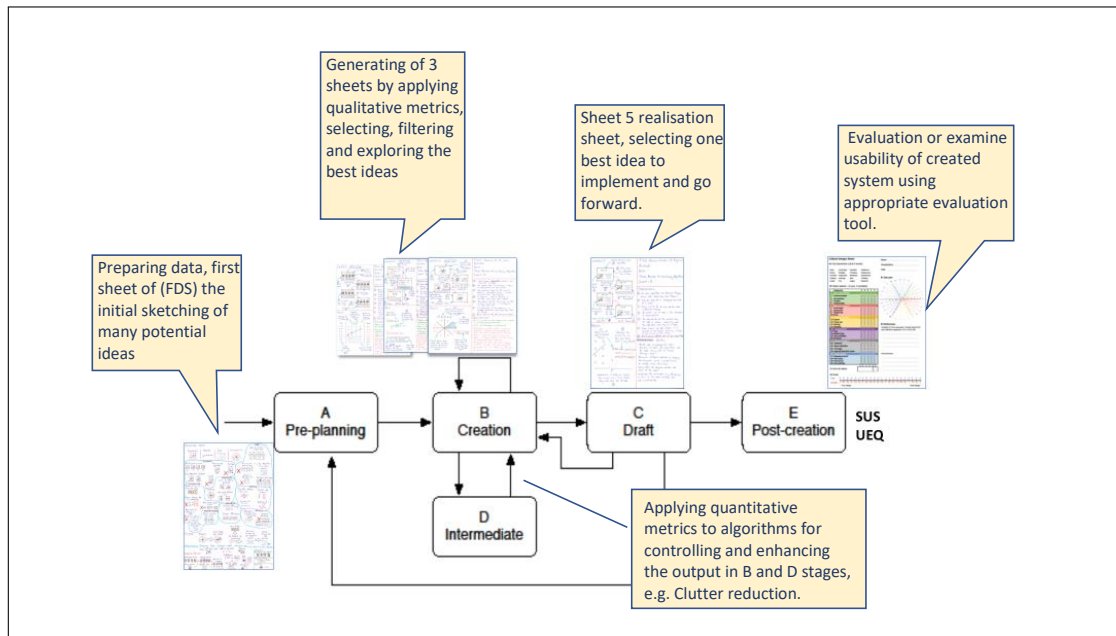


Figure 2.4: Using (FdS) methodology as example to explain how metrics (qualitative/quantitative) can be performed and processed by the framework. Each stage is responsible to perform specific task which refers to specific metric type.

need to think critically. They need to carefully consider how they can improve the code, and improve the performance of their tools. They can achieve that by testing, and refining the code, evaluating how the code runs under different configurations and situations, and applying appropriate metrics. These metrics then can be used to help measure the performance, and provide benchmarks that improvements can be adjudicated against. Conceptually, this model can offer users a systematic approach towards creating a visualisation tool then evaluating it with the use of visual metrics. The descriptions of pipeline stages and examples in the metric table show various examples of how metrics are used in visualisations to enhance understanding and performance of visualisation.

Our proposed method (the Critical Design Sheet) can fit with the pre-planning stage to make a grounded evaluation process by gathering information to ensure this context fits with a specific visualisation tool for evaluation. This fits well with the model by Isenberg et al. [74]. In addition, the CDS method can also be considered to be suitable within a post-creation design process. Where the user is evaluating and reflecting upon the completed implementation. In this case, metrics that relate to the design of a visualisation, and how well it fulfils the design specification can be created. For example, strategies to test the efficiency of visualisation tools [145] such as the usability test can apply, to ensure the tool is deployed in a real workplace.

2.5 Evaluation methods for visualisation and direct related work to the CDS

Over the past twenty (or so) years, many researchers in data-visualisation have been calling for a situation that “every paper has to have an evaluation”. For example, Shneiderman and Plaisant write “*We seek to encourage information visualisation researchers to study users doing their own work in the process of achieving their goals*” [145]. However, some researchers and developers have become frustrated over this situation, whereby some of these frustrations are levied towards particular researchers who have merely applied evaluations to their projects such to appease reviewers. For example, Munzner [106] writes “*Positive informal evaluation of a new infovis system by a few of your infovis-expert labmates is not very compelling evidence that a new technique is useful for novices or scientists in other domains.*” Further, there can be limitations with the use of metrics (in some situations) for evaluation, for example, Bertini writes, “*InfoVis is successful when users can gather new, nontrivial insights, a process that takes place over days or months and can rarely be simulated with a series of short tasks. We judge metrics such as task time for completion and number of errors to be insufficient to quantify the success of a system*” [17]. But while there has been much discussion in the conferences as to whether evaluation should always occur, it remains that evaluation, reflection and improvement of that process are highly important strategies; especially to deliver a rigorous scientific outcome.

The goals of this dissertation (see the Introduction, Chapter 1), and indeed the outcome of the Critical Design Sheet (see Chapter 4), fits within this broad evaluation space. However, the subject of evaluation is very wide, even when just considering evaluation for visualisations. Visualisation evaluation is highly influenced by Human Computer Interaction research, which in turn is influenced by research in Psychology and Perceptual sciences, and even other fields such as Ergonomics and the study of humans (such as anthropology). Further, there are many places (of the visualisation design, build, reflect process, e.g., such as Munzner’s nested mode [105]) where evaluation and certainly reflection is necessary to design good visualisations.

Evaluation comes in different forms and different positions of the design/build process, and can be achieved by different actors. On the one hand, it is possible to evaluate a paper-based low-fidelity design (such as created with the Five Design-Sheet model, see Roberts et al. [131]). While on the other hand, it is possible to evaluate the usability of the completed tool with expert users (such as, Tory and Moller [155] who focus on the evaluation of visualisation designs by expert panels). It is also possible to think about *who* is doing the evaluation (the actors). It could be the developer, such as to improve their own design. Or it could be the real user, who is using the tool in a beta test. Or it could be a student, who is evaluating the tool for a graded assessment. Drawing upon work by Lam et al. [92] and Isenberg et al. [75] and especially from Nielson [110] in his heuristic evaluation of interface design, it is clear that there are many ways to evaluate interfaces or designs. These include:

- automatically (using algorithms and metrics),
- empirically (assessed by real users [34]),
- formally (using formulas and usability measures), and
- informally (based on the skill of evaluators).

So far we have focused on the information visualisation domain, but it is useful to shed light on the *visual analytic* domain which covers various interdisciplinary fields of research [83] combining visualisation, human factors and data analysis. In fact, if critical analysis is automated, then we consider it to be a part of the visual analytic field, in particular focusing on analysing visual design.

Visual analytic (VA) is described as “*the science of analytical reasoning facilitated by interactive visual interfaces*” [40]. This science strongly relies on statistics and mathematics as the driving force of the automatic analysis. Thus, there is a need for robust evaluation methods to measure the effectiveness of the automatic visualisation methods and algorithms. Therefore, for evaluating the visual analytic environment, researchers have developed several *metrics* and *methodologies* to measure the progress of their work, also to understand the impact of their work on the users who will work in such environments. For instance, Scholt [139] proposed five evaluating areas: situation

awareness, collaboration, interaction, creativity, and utility, each area includes several metrics and methodologies for assessment [60], [139].

Human Computer Interaction (HCI) is a very important part of any visualisation tool, and interaction is one key to generating successful visualisation tools. So, let us take the interaction as area of assessment in visual analytic environment, it is necessary to evaluate the usability of the interaction to measure the capabilities of the interaction. Zhou and Feiner [170] added additional tasks “identify; encode; emphasise; reveal; generalise and switch” to Wehrend and Lewis’ [166] tasks “locate, identify, distinguish, categorise, cluster, distribute, rank, compare between entities, compare within relations, associate and correlate”. So, it is valuable to ensure the usability of each interaction as well as to evaluate the utility, which means that the various interactions should provide some insight into the information.

Automatic critiquing systems can be considered as part of the visual analytic field, because researchers in this area proposed and implemented different strategies to build systems that make critical analysis. Such evaluation methods like usability evaluation can be integrated with these systems and are embedded in the design work-flow environment as visual design guidelines. For example Voyant created a feedback system [168], [169] that is aimed at non-expert users that can perceive perception-oriented feedback on their designs from a selected audience as a way to help novice users develop their designs.

Computational machine learning models have been developed in the 1990s to generate critiques without the need of human intervention (e.g., [55], [94]). In the last several years, various advanced systems have been proposed for example, creating designs in the design environment with the ability of humans to justify designs if required within the system. These tools help designers to detect and reduce design problems and offer beneficial features from using users thinking and understanding feedback mechanisms which support designers improve their design [56], [97], [98], [113]. The latest systems generate feedback for making effective changes focusing on four main parts in visual design: theme, layout, typeface and colour. The usefulness of these systems varies from one to another depending on the purpose of the critiquing systems and what their purpose is. For example, crowd-sourcing systems might be useful to use as online visual design instructor to help students learn some design guidelines in order improve

designs. In other words, asking their peers online which are good (or bad designs). On the other hand, evaluating the efficiency of design critiquing systems is a challenging task compared with human ability and a comprehensive understanding for visual design. What is needed are specific rules, that could expand and change with time, that would positively affect the development of these systems.

Empirically and informal evaluation techniques are the most related domains for the research of this dissertation. The proposed method fits within the informal classification, which is classified as heuristic methods, where experts judge the compliance of an interface against predefined heuristics [110], [171]. In fact, Nielson's influential work on "10 Usability Heuristics for User Interface Design" [110] are well respected, techniques to help developers design effective interactive systems. Nielson is also a supporter to get the real end-users involved in the evaluation. In fact, Nielson's heuristics are guidelines for experts to follow. This differs from our goals, because we wish to have systems that can be used by non-experts. While there is much benefit in getting experts to collaborate with you [145], [155], this is not always feasible or practical. Especially in a learning environment it is difficult to get experts involved. There are certainly times when it is possible and suitable to collaborate with external companies, and then it is very important to get experts involved in the projects of students. However, when students or developers create visual designs, they need to take responsibility for their own decisions. If a company expert is relied-upon then the student would not learn how to critique their own designs. Also, it would often be too time-consuming to involve an expert for (potentially) hundreds of students on a module or course. So, when the learner is creating the visualisation for an assessment they need to make decisions about their own design, so they need to follow a structure to help them develop their critical thinking skills.

For visual presentations, there is a need for self-critiquing systems with a formal structure. In fact, developers continuously make judgements on designs such as to work out how to improve them. Kosara [89] suggests 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" [90] but does not offer a formal structure. While there are advantages in critiquing work, it is difficult to be fair-minded and unbiased and to put personal opinions aside. Working together

in a group can help to moderate views, Jackson et al. [77] 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 user personally makes judgements on the visualisation (rather than a group), and we provide a formal structure for the critical process. Saraiya et al. [138] also use insight based methods, and provide a structure of eight characteristics: observation, time, domain value, hypothesis, directed versus unexpected insights, correctness, breadth versus depth and categorisation. This latter work is the closest to our method, but we develop a more systematic structure, define a set of questions and use a single sheet that can be completed by users. Our sheet (the Critical Design Sheet, CDS) also delivers a result (score) that can then be used by the user to compare previous critical analysis of visualisations (which can be used to demonstrate that the tool has improved).

Informal styles of evaluation are useful, especially to help focus and organise personal thoughts within the design process and to aid critical thinking. Amar and Stasko [4] walk the user through thinking about rationales and tasks, and Munzner [105] implicitly advocates critical thinking throughout the whole visualisation creation process. In fact, ideas of visual inspection have a long history. Bertin [16] suggested 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 [37] also suggests three perception operations: detection, assemble and estimation. Lohse [96] uses similar ideas to help classify visual representations, and develop a model to understand graphical depictions [95], while inspection techniques were used by Conversy et al. [39] to observe visualisations.

This, and other work, provides guidelines on how to design visualisations, which are detailed in many books (e.g., the visualisation design book by Ware [164], the book by Munzner [107], and the sketching book by Roberts et al. [132]), but while these books provide a rich tapestry of information for the learner it is difficult for students to aggregate this into a single critical review structure. There is a growing interest in evaluating visualisations [6], [34], [75], [92], [119] and [140]. Several researchers have produced classifications on the different types of evaluation strategies. For instance, Plaisant [119] 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.

However, because this method focuses on usability, our Critical Design Sheet does not fit well with this classification. In another classification, Isenberg et al. [75] extended the seven scenarios of Lam et al. [92] and made a comprehensive review of evaluation techniques in visualisation.

Using the categorisation of Isenberg et al. [75] and Lam et al. [92] we could place the Critical Design Sheet in the category of Understanding Environments and Work Practices (UWP), because it would be useful within the design stage. 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. But in fact, while this is useful this has limitations. It is difficult to know where users reflect through scenarios or to make judgements on design, and while our proposed evaluation method (CDS) could fit with the UWP and VDAR classification it is difficult to precisely see where it fits, because our goals differ. One solution could be to extend the UWP or VDAR or add another category to the process. But we feel the framework, described above in this chapter, actually provides a better way to consider where metrics and design-thinking takes place. In fact, considering visualisation in a wider perspective, there is a notable lack of evaluation methods in literature that focus on self-reflection and understanding for design analysis criteria based on peer guide structure to evaluate the visual design in visualisation.

Researchers can use various evaluation methods for their visualisation design study, these include: heuristic evaluations, log analysis, and observations (see Lam et al. [92]). Lam and colleagues study a combination of evaluation methods, and explore how they are used in designing processes and validate our proposed critique design evaluation method. Studies can also focus on qualitative and quantitative aspects. Collecting information through observational studies and contextual interviews are considered common primary methods to set qualitative evaluations. These methods give understanding to the requirements that would be needed to help designers complete their systems and develop the proposed method. Quantitative evaluation methods involve statistical approaches which is part of traditional empirical scientific experimental approaches. These types of methods are known with “laboratory experiments” or

“studies”, which mainly look at the effect between two (or more) visualisations. Later in this dissertation, we show how experimental work has been conducted to investigate the usability of CDS; hypothesis test development to measure the correlations and the consistency of proposed CDS categories items. Other usability evaluation methods such as System Usability Scale (SUS) have been done to evaluate the CDS online version.

It is important to shed light on the usability test for CDS, because in this research we need to find out the usability of the CDS with end-users. As mentioned above, CDS works as an informal evaluation practice and we want to investigate how easy it is to use and the prototype design of the CDS in real situations. In the visualisation literature that focuses on evaluation we can find a many researchers investigating “heuristic usability” as an inspection evaluation method. Nielsen [109], [110] introduced this type of evaluation in information visualisation, he proposed this method as a tool to help users evaluate their interfaces and find usability problems. A set of heuristics usability is provided to measure an important aspect on systems and these heuristics can be applied to any application. Nielsen distilled nine heuristics suited as the basis for practical heuristic evaluation [103]. The nine heuristics corresponded to principle guidelines for any interface design and it provided an interface standard form with a completed detailed guidelines checklist for evaluation.

The method of the CDS, in terms of reflection, is similar to Molich and Nielsen’s [103] work, because users exercise reflection and critical analysis organised in a set of categories. However, their work differs because the categories are different and the CDS structure is designed, not for experts but, for novices and learners. The CDS structure helps users analyse visual design in terms of (1) Users’ perception, (2) Environment, (3) Interface, (4) Components, (5) Design, and (6) Visual marks. In the CDS, users use the six principle categories to give a grade on a Likert scale for each category, when users utilise the Nielsen heuristics the people performing the evaluation write a report, and the authors gave scores on that report regarding the number of problems on the interface as precisely as possible. In contrast, the CDS uses a subjective rating scale to grade the categories. This has several advantages such as being easy to use; inexpensive; can detect small variations in workload; reliable; and provide decent convergent; construct; and discriminate validity [70], [114]. These positive features support the use of the

Likert scale in the CDS, to rate individual design categories. Chapter 5 will address the usability evaluation of CDS in more detail.

2.6 Summary

This chapter focused on a review of various visualisation metrics, and explored how these metrics participate with the development design life cycle. The review of visual metrics encouraged us to create a metrics model which supports in classifying the use of metrics and how to apply it in the visualisation design process. Many types of evaluation methods for visualisation have been mentioned, both qualitative and quantitative evaluation methods need metrics to collaborate with visualisation enhancements and performance. The chapter also investigated evaluation methods for visualisation, and placed the ideas of the CDS in context with related work. The next chapter focuses on critical thinking and how different fields are integrating and using critical thinking within their domains and for different purposes.

Chapter 3

Critical Thinking

Humans are rational thinkers. While many decisions can be taken quickly, many decisions are taken after deep thought. To make decisions, humans collect facts and objectively analyse details. This chapter presents fundamental ideas of critical thinking and critical analysis. The broad goal of the dissertation is to investigate processes to help humans make better design decisions, therefore, there are many processes and ideas surrounding critical thinking that need to be explored. This chapter particularly addresses educational models and strategies that support and increase students' metacognitive (thinking about thinking) skills. Further, this chapter describes the effect of critical thinking concepts on the design process and the role of thinking in design foresight visualisation pipeline stages. These concepts will be explained in following sections:

- Section 3.1: fundamental and general concepts of critical thinking.
- Section 3.2: critical thinking models in academic curriculum.
- Section 3.3.3: reflections on design visualisation stages.

3.1 Fundamentals in critical thinking – definitions and concepts

Recently, it seems there are increasing demands on critical thinking concepts in academic pedagogy. For instance, the book “Critical thinking and education” [101] that was only

published in 2016 has already been cited 1350 times¹. But the topic of critical thought has a long history. Humans have always been thinking and making decisions. In the early days, it would be possible to imagine early humans arguing and working-through disputes through thinking critically about the situation. However, much of the main evidence has its origins with the early and ancient philosophers of Aristotle, Socrates, Plato, etc. Those early philosophers laid the foundations for *dialectic* and *rhetoric* that they considered to be the essential parts in developing thinking into critical decisions. These two strategies reflect key elements for argumentation for making critical decisions which are presented among the audience. Dialectic processes explain how humans inquire and seek understanding for a question and how to answer it creating a systematic thinking based on appropriate proofs to reach a decision. While, rhetoric is presented as process of persuasion that Aristotle used; it engenders logical reasoning through presenting satisfied evidences to help decision makers (the audience) understand and make decisions to each opposing claim. Those type of early attempts in critical thinking helped to construct methodological reasoning processes for solving complex problems and tracing implications in an insightful way.

In the middle era (about the 15th and 16th centuries) many philosophers followed up the early philosophers' rules and lessons. They expanded their uses of critical thinking in different domains: in art, law, religion and even in freedom. One of the notable thinkers in England at that time was Francis Bacon, his book "Advancement Learning" had a great influence on society, particularly regarding scientific methods. Bacon argued that there where huge possibilities to extend scientific knowledge through reasoning and observation of natural events. He emphasised the importance of seeking scientific reasons, and to use a methodical approach (with the aim to avoid scientists misleading themselves!) This method reflected a new turn in the rhetorical and theoretical meaning of science, and introduced practical thought in the form of a methodological framework for traditional critical thinking.

In the same era, the French philosopher Descartes was also influenced with similar scientific tendencies to Bacon. But Descartes' influence was representative of mathematical knowledge. In fact, he is known as the father of analytical geometry, which led to the development of what we now known as the Cartesian coordinate

¹Citation count according to scholar.google.ac.uk in August 2017

system. He employed the principles of methodological scepticism in his philosophy to develop reasoning thinking in a doubted way. By arguing that everything can be doubted and then investigated if it was indeed possible to doubt in order to acquire a firm foundation for genuine knowledge. Descartes added a notable development in Aristotle's foundations of critical thinking, which he discovered it was incomplete and that method of doubt enhanced Aristotle's foundations. Many thinkers followed these rules in a sceptical methodology such as Hobbes and Locke in the 17th century, which contributed in enhancing thinking critically by explaining everything with evidence and reasons in various fields such as politics and economics. Many scientists appeared at that period which reflects the influence of the scientific revolution on the power of critical thought and its tool in society. Newton, Boyle, Freud and other scientists that have made a significant contribution to the modern science.

Later thinkers benefited from early knowledge of critical thinking, in the 1900's and especially in the 1960's Facione [53] writes about the pedagogical thinkers of the 1960's such as Ennis [51] who wrote "A concept of critical thinking", or Passmore [115] who wrote "On teaching to be critical". Later in the early 1990's there was a great development in learning and teaching critical thinking, where Facione [53] writes The Delphi Report, titled "Critical Thinking: A statement of Expert Consensus for the Purposes of Educational Assessment and Instruction". One important aspect of critical thinking is that *critical thinking* is always "thinking about a topic X", without knowing what the subject is, the process of critical thinking is irrelevant [101]. Consequently, it is impossible to teach critical thinking without its application to a subject. People need to locate the right knowledge, know what the challenge is, to weigh up different solutions. Therefore, this dissertation deals with critical thinking for visualisation.

In particular, the goal of the research in this dissertation is to create a device (or a system, in our case this will be the CDS) that can help learners critique their work. Consequently, this chapter needs to address critical thinking in education (see Section 3.2). In the design chapter, the dissertation also creates a workshop which analyses how individuals think about something. Therefore, we need to observe how critical thinking skills are used to make sense about everyday life.

Psychology researchers have created various definitions for *critical thinking*. These definitions come from different fields in which critical thinking has been applied. Furthermore, critical thinking is never universal in individuals; some people are better at critical thinking than other people. This has fuelled the debate whether critical thinking is an acquired skill, and that it has a set of rational rules; or whether it is fundamental human understanding intertwined with domain knowledge. Some definitions:

“Critical Thinking [is] the systematic evaluation or formulation of beliefs, or statements, by rational standards. Critical thinking is systematic because it involves distinct procedures and methods. And it operates according to rational standards in that beliefs are judged by how well they are supported by reasons”, Vaughn [161].

“critical thinking requires the judicious use of scepticism, tempered by experience, such that it is productive of a more satisfactory solution to, or insight into, the problem at hand”, McPeck [101].

There are many skills involved with critical thinking. Halpern [68] in his book underlined a list of characteristics for critical thinking to be a good critical thinker. He described the term “critical” in a positive sense by referring to the ability of persons to ideally evaluate and judge with the goal of providing a useful and accurate feedback that serves to improve the thinking process. However, these skills need continuous practice time and training in a variety of contexts to be improved.

On the same view, Facione [53] defines six core critical thinking skills that critical thinkers possess:

1. Interpretation,
2. Analysis,
3. Evaluation,
4. Inference,
5. Explanation and
6. Self-regulation.

Having these skills means users need to be comprehensive and careful in solving a problem, while keeping their eye on the goal. It is necessary for scientific practitioners to benefit from these concepts to help in solving problems and develop their abilities to help students improve their critical-thinking abilities to use those cognitive skills or strategies which increase the probability of desirable outcomes.

Critical analysis is another term that is used in similar sentences to critical thinking. Critical analysis plays an essential role in such academic disciplines as philosophy, linguistics and psychology. In teaching and learning, students use the strategies of critical thinking and analysis to analyse discourse, text, scientific systems, phenomena, and even characters. Language processing is a well known field for texts analysing either at a theoretical phase or computerised phase. Analysing text is the process of extracting its features based on several criteria and principles. Charles Bazerman [14] defines text content analysis as “*Content analysis is the identifying, quantifying, and analysing of specific words, phrases, concepts, or other observable semantic data in a text or body of texts with the aim of uncovering some underlying thematic or rhetorical pattern running through these texts*” p.14.

So, this definition gives a comprehensive method for content analysis based on those principles or measures that distinguish patterns in text and for each analysis criteria have many corresponding strategies. Researchers in these fields must break down the text or corpus into individual parts to identify the strategies and highlight conceptual and relational concepts. Rhetorical analysis is a kind of critical reading in which persons might deliver symbolic actions which come from effective communications, and understand how people in specific situations attempt to influence others through language. This kind of persuasive aspect has been used in rhetoric, rhetoricians (Aristotle, Quintilian etc.) developed several terms that they named “canons of rhetoric” which include:

- *inventio* (creation),
- *dispositio* (arrangement),
- *elocutio* (style),

- *memoria* (recollection) and
- *pronuntiatio* (delivery)[141].

These terms could be divided into subsections, such as the Aristotlian terms in invention, of *ethos*, *pathos* and *logos*: detailing, respectively, the trustworthiness and credibility of the orator, how they use emotion to tender support and their logical argument. While now all oral arguments have written, rather than oral rhetoric tradition, these ideas have been refined and they are taught to lower-school students and in Universities. In other words, researchers in other academic disciplines can benefit from text analysis principles and strategies specifically from rhetorical canons in the tradition of Aristotle and interpret these ancient terms into modern tactics structure to help the analyst to devise guidelines for the audience. Transferring this knowledge into visualisation gives support to visual analysts to analyse visual depictions. This study underlies these principles in analysis by breaking the visual depiction into different categories, considering them individually to make an overarching judgement on the design.

3.2 Critical thinking models in the academic curriculum

In this research we are focusing on critical thinking in the scope of academic teaching and learning. This is in terms of helping learners create rationals and reasons around problems and let them reflect on how to solve problems finding solutions to achieve best conclusion. It is essential to address Bloom's Taxonomy [20] as an earlier educational system. His system classifies thinking into six cognitive levels hierarchically. Figure 3.1 shows the pyramid diagram of Blooms taxonomy. The model helps to guide teachers and students to think about different skills; from the low-level skills of retaining knowledge, to the higher (more cognitive) skills of evaluation and creation/synthesis. It has been used as a basis for curriculum analysis, test construction and data summary [9]. In such educational courses Bloom's taxonomy creates a structure (a scaffold) that supports students metacognitively and increases their reflections on what they have learnt. Scaffolding provides support to allow the learner to learn by his/her own self.

Many researchers have adapted and made revisions for the original taxonomy (Bloom's taxonomy) into new terminology taxonomy, all the original sub- categories were replaced with gerunds, and called "cognitive processes". Anderson [5] illustrated the adapted version for the original taxonomy to "Remember, Understand, Apply, Analyse, Evaluate, and Create". Krathwohl [91] presented a combination of knowledge and cognitive processes as taxonomy table to help teachers to classify objectives, activities, and assessments that provides a clear and concise visual representation of a particular course. That table can be used to examine relative emphasis, curriculum alignment and with regards to this examination, teachers can decide where and how to improve the planning of the curriculum and the delivery of instructions. Thus, the cognitive process is helping students to determine the level of their work from self-assessment for their levels of achievement in courses. It is so useful for students to use critical analysis in their curriculum [117], it is considered a successful intellectual skill that students acquire. Further, the logic reasoning elements model [49] helps them to analyse problems through analysing the logic of the subject using fundamental reasoning elements as a guide for solving problems.

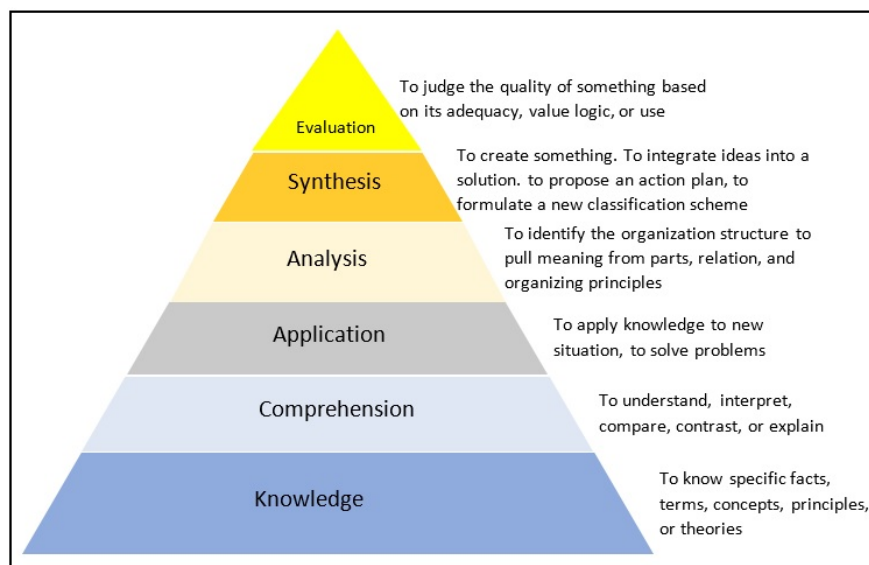


Figure 3.1: Bloom's levels of learning re-produced by the author from [59]. Bloom's taxonomy shows the cognitive processes that a learner moves from the foundation Knowledge level up to Application and ultimately, Synthesis and Evaluation, the original notation and concepts belongs to [20]. The taxonomy is also explained and links are provided for even more useful resources, <http://www.odu.edu/content/dam/odu/col-dept/teaching-learning/docs/blooms-taxonomy-handout.pdf>.

Structural thinking also is considered one of the thinking skills in the Richmond thinking system [129]. Richmond placed seven track skills for the educational thinking system to help the track thinking system framework. He suggested a new learning gestalt

form from three threads: educational process, thinking paradigm, and learning tool. However, he tried to transfer educational learning to a dynamic and successful system of learning by combining these three threads. The Richmond structural thinking skills model reflects that when people use physical converting laws, they must think in units of measures and dimensions to gain a structural presentation and track the solution for the problem. This type of metacognitive considers objective thinking, thinking in measures, factors, dimensions to evaluate decisions and proper judgement, for instance from our life examples thinking about selecting the best hotel, best restaurant, game, etc. What are the dimensions based on that choice?

However, we are trying to address the elements and factors to build critical thinking aspects in as many students as possible. Actually the levels of these skills vary between people, Halpern [67] addressed the difference between higher ordered cognitive skills and lower order cognitive skills in human. For example, in learning mathematics and computational arithmetic methods, there is no consideration for high thinking skills because this kind of thought needs to follow application routes of well-learned rules of basics in mathematical operations to get the right results without concerns about the variance of answers that affects the results, because it always has unique results. In contrast, to make a final decision over, such to compare two pictures and decide over the degree of beauty or the level of colour harmony in them, people do need to use high level cognitive thinking. It is high-level because there is no definite answer; it is subjective. For instance, if a person thinks or asks himself “why is picture A much more beautiful than picture B?” The answer for that question could differ if a person said “the picture A is beautiful” which expresses his/her feeling towards the picture. We can look at this argument from different viewpoints, the first question is asking for the reasons or rationals that help him/her make his final decision (about which one is beautiful). However, measures can be applied to be convinced with the choice, even if that choice is not far from subjectivity with respect to appropriate dimensions of the choice. As example, if a person wants to buy a new computer it is clear that he should specify the purpose and the brand of computer in advance. Then it is important to put in his/her mind several measures in order to select the best one. These measures would include price, specifications (processor, RAM, HDD), and possibly the physical size of the computer. These many aspects of the computer that will help determine the person to make a suitable decision (a decision that they are happy with). Through this thinking

they are critically analysing different aspects of the computers, and judging how they compare in different circumstances.

On the other hand, the saying “this picture is beautiful” probably reflects a person’s subjective feeling. Judging on beauty perhaps does not need the person to realise what dimensions they are considering. Because this kind of judgement is intuitive, it relies on the human’s feelings, and sometimes we don’t know (or really care about) the reasons. However, Martin and White in their book (*The Language of Evaluation*) [100] presented framework mapping types of feeling that interpreted in English text that which traditionally refers to emotion, ethics and aesthetics. They registered emotion as an affect that concerns positive and negative feelings “do we feel happy or sad, confident or anxious” and that are embodied physiologically in our heart from the moment of birth. Anyway, people need to make critical thought and be much more conscious about the decisions when they face complex and ill-defined problems. In fact, using critical thinking models as learning instruments is part of constructing knowledge for students rather than passively receiving it, and it also supports the development of understanding and learning.

3.3 Thinking with visualisation

Focusing on aspects of thinking in visualisation, we see from the framework of the metrics (Chapter 2) that there are different places to evaluate the quality of the visualisation design; consequently, there are likewise several ways to apply critical thinking with visualisation. We specifically turn our attention to first creativity and design thinking (Section 3.3.1), second abstract models (Section 3.3.2) and third reflection (Section 3.3.3).

3.3.1 Creativity and design-thinking strategy

Creative design is the ability to generate something that is new, surprising and valuable [21]. Planning to create a specific visualisation design requires various

skills and concepts like walking through the Five Design-Sheets method [132]. Skills need to be accumulated and design principles and best practices followed. To develop our reflections on creative visual design it is important to emphasise the design process. Design-thinking [127] is one of the interesting areas that play an essential role as a design process. The nature of that process reflects the knowledge of people to produce the work, and the outputs are evaluated to produce knowledge. In fact, *design-thinking* is a concept that can operate in different domains such as science, business and arts that because there is no unique meaning for design thinking and researchers have different theoretical perspectives regards this concept. It can refer to cognitive style which leads to creativity and innovation, others look towards design thinking as general theory of design which designers firmly require professional ways for solving design problems with the diversity of designers' practices. Design thinking acts as a resource in organisations, and focuses on putting the human as the main agent in designing [85]. Design thinking relies on a mix between *thinking* and *knowing* and knowing about the world, and *acting* upon the critical thought. Design thinking is very useful concept in creativity and innovation for visualisation.

Figure 3.2 describes a conceptual framework of the design-thinking process which is characterised to be an iterative process [25]. The iterative cycles, of constructions, can be managed through various activities as a way to develop design to meet the requirements and needs. The cycles include sketching, team discussions, prototyping reviewed by experts or clients and so on [104]. Subsequently, there are three important processes in the design thinking framework, as follows:

Understand. This part is interested with observe user's behaviours and how they think, defining problems and collecting all relevant information and focusing on important ideas.

Explore. This part focuses on discovering creative ideas, finding reasoning to elicit tactical solutions, building the sketched patterns and chance to find alternatives.

Materialise. This is the crucial part in the process, transforming ideas into reality, implementing solutions to test its effects on practical manners.

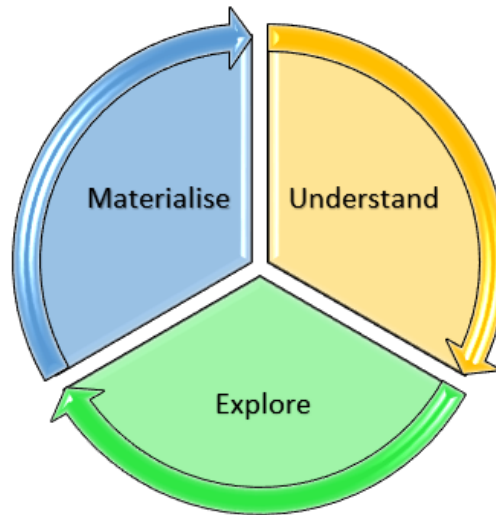


Figure 3.2: The conceptual framework of design-thinking is characterised as an iterative process and it composed of three main activities, (1) Understand, (2) Explore and (3) Materialise. The image has been created by the author and the concepts taken from Nielsen Norman Group by (Jakob Nielsen and Don Norman), <https://www.nngroup.com>

In education, design-thinking is known as an approach to learn the design process. Drilling down further, design thinking needs to have a human-centric approach [125], where designers have to put the needs of the human user at the forefront of their mind, as they consider and contemplate their designs. It is often difficult for a student (developer) to consider the end-user, and how that user would feel and react to the tool that they are designing. But it is very important for the designer to empathise with their end-user. In fact, we place much emphasis in our work by focusing on these aspects of empathy. Our focus, in this research, is to encourage the developer to perform meta-cognition, and take a user-centric view of the design process rather than merely considering how to create a beautiful (aesthetic) design. Developers need to imagine how users would interact with the tool and they need to evaluate the outcomes. However, it is important when thinking about visualisation design to look at visualisation environment requirements, divide this environment into individual *frames* to enhance the understanding of the input and output for each frame [132]. The idea of splitting the concepts into “frames” is similar to the process of critically solving problems especially when we face major problems. The problem solver subsequently needs to divide the problem into individual or smaller challenges. Systematically dividing the visualisation environment into individual frames means looking to visualisation from multiple viewpoints by giving a contextual frame for each manipulated structure. Our hypothesis in the research is to break down the visualisation design into minimum frames or parts (as a forward reference to the CDS,

we include six main categories, see Figure 4.15) that enhance and critically analyse each part to help users evaluate and make decisions on parts of the design. The concept of a “frame”, used here is distilled from Data/Frame Theory. Gary Klein [86] laid out a theory of sense-making that would be useful for the intelligence domain. Each frame can be expressed in a different meaning or forms, such as stories, maps, organisational diagrams, or scripts. These frames enable concepts or ideas to be explored, treated individually, then they can be used in subsequent or parallel processes, which easily extended or adapted if required. The frames can help intelligence experts to consider a scenario with a given set of evidence. They become staging points that can be explored on their own (with their own assumptions) and can be presented to different expert groups.

The idea of “frames” is beneficial because it can save time for presenters. The presenter does not necessarily need to explain the scenario, because the frame instantly brings specific knowledge and implicit details with it. In other words, there is an implied back-story with the frame. However, especially in visualisation this may be problematic. Different visualisation users may have a different understanding of certain terms, or they may have implicitly made assumptions (based on the frame) that are different to their colleague. For example, mapping data to visual retinal variables (size, orientation, colour, texture, transparency, etc.) relies on the user understand how low values are mapped, and how high values are mapped. This is the *polarity* of the mapping. While polarity can be implied, it is not always consistently applied by every user. For another example, visualisation terms such as what “data” is, or “abstraction” or “knowledge” may mean different things. Each one of these terms has its own definition and classification that gives an ambiguity to the representation in the cognitive and computational space [35]. In such cases, there are no agreeable definitions to data, information and knowledge, but such abstracted models define knowledge as compositions of data and information inputs. The Data-Information-Knowledge-Wisdom (DIKW) hierarchy [1] which is known by Ackoff’s model describes a human’s understanding of these terms in cognitive space. Ackoff classified the content of the human mind into five categories:

1. Data,
2. Information,
3. Knowledge,

4. Understanding and
5. Wisdom.

The first four categories manipulate the stored information or what is known the human mind, but the fifth category deals with future vision and when acquiring knowledge principles effectively.

In addition, Ackoff defined *Data* as raw symbols that can exist in any form. Information can be defined as structured data or data that been given meaning by relational connections. *Knowledge* is the useful collection of information and computers can store knowledge as a form of application that acts on data such as simulations or interaction. Understanding refers to the process of manipulating knowledge, even this knowledge is stored in the human mind or synthesised new knowledge from his previous experience [136]. *Wisdom* in Ackoff's model refers to how humans evaluate this understanding, Rowley [136] addressed such arguable points towards fundamental concepts in the wisdom hierarchy, and he also addressed a lack of wisdom hierarchy's definition and application in information systems and knowledge managements. This concept has been neglected by researchers and practitioners in that domain. However, it is not necessary for users to cover all these details or concepts in Ackoff's model, rather they need to know and make sense of the requirements of inputs to the visualisation model.

3.3.2 Abstract models

Abstract models can be known as conceptual models which are forms of composition concepts. The primary objective of a conceptual model is to convey the basic and fundamental functions for the proposed system. People usually use these models to represent the whole system in an abstracted way by which the external world is emulated to some extent. Conceptual models can be used in different disciplines e.g., scientific or non-scientific because it helps in representing concepts and describes the relationships between system entities. However, psychologists consider these models as communication tools even if it could not serve all system requirements. But it can help in conveying and understanding information rather than concepts for users that

have different backgrounds. There are different types of abstract models, each of them represents its purposes such as mental models, mathematical models, scientific models and metaphysical models.

As an example, in cognitive science and psychology, people use mental models to construct their systems. Mental models are considered a way to represent the external reality as internal relationships between system entities. Emerging knowledge between different disciplines helps in transferring the mental model perspective to scientific domains. Mental models are used by different people to help in making decisions and taking actions regarding real life problems. It is a road map in the human mind that is used to interact with the external world. It has advantages such as flexible users can develop mental models with the time for their systems to meet future demands.

In general, psychological, philosophical and linguistics domains which relate to teaching and learning are more frequent in their use of these systems regarding a theoretical base to concepts and principals represented as simulation models. Transferring knowledge and inspirations from these disciplines to visualisation give researchers the potential for creating various visualisation perceptual models that help researchers develop visual modelling systems. The following sections will address typical visualisation models and browse its applications to see how critical analysis and human-centric concepts contribute to the construction of these models.

3.3.3 Reflections on visualisation stages design

Critical thinking skills pave the way for critical analysis in visualisation. In visualisation, users need to use their critical thinking skills and understanding to *analyse*, *evaluate*, *explain* and structure their thinking to be able to solve visual problems. Visualisation has various perspectives which needs to think insight, information visualisation can be defined as

“the study that encompasses forms of data transformations, abstract information and knowledge in the context of interactive representation”,

Ware [164].

Visualisation processes are integrated with human perceptions. Not only does a visualisation engineer need to create the visual designs, but they are using graphics to communicate the data, and it is a human at the end who will use their human senses to understand the information.

The early visualisation models come from the dataflow model used in the Application Visualisation System (AVS) [159] and also the work by Haber and McNabb defining the conceptual model for scientific visualisation systems [66]. Figure 3.3 shows a more detailed version of the visualisation pipeline that interprets the human role as interactions on visualisation stages. Critical thinking aspects clearly reflected on the human-human interaction part which starts from the viewing stage, which focuses on how to make the human eye explore and judge the creation of the visualisation part. The viewing context part is a visualisation process engine which is typically performed by a computer. The second part of the visualisation pipeline is the critical analysis process “Understanding” context which is usually done by the human mind and users can seek how to develop and optimise the visualisation process outcome. Human-human interaction is applying human knowledge and experience to interpret, analyse and evaluate visualisation results, design or visual metaphors [36]. We can consider the second part as a human conceptual on visualisation which reflects the critical thinking concepts on the visual domain.

We propose that critical thinking needs to be applied to all aspects of the visualisation pipeline. In this process, a human is considered to be an integral part especially when interaction is required. Not only choosing the visualisation type, but from loading the data to working out how it is displayed, even to deciding what visual marks to use to display the data, and then discovering if the visualisation is appropriate, effective and usable. It is possible in visualisation to employ Facione’s [53] core critical thinking skills that scientific practitioners can benefit from these intellectual concepts in enhancing and developing the visualisation environment and particularly in typical visualisation design process. Figure 3.3 puts visualisation into two contexts, viewing by eyes “perceiving” context and seeing by thought “understanding” context, that reflects human communications towards visualisation either with his interaction agents or with his cognitive analysis cues. We can demonstrate how a human applies such critical thought

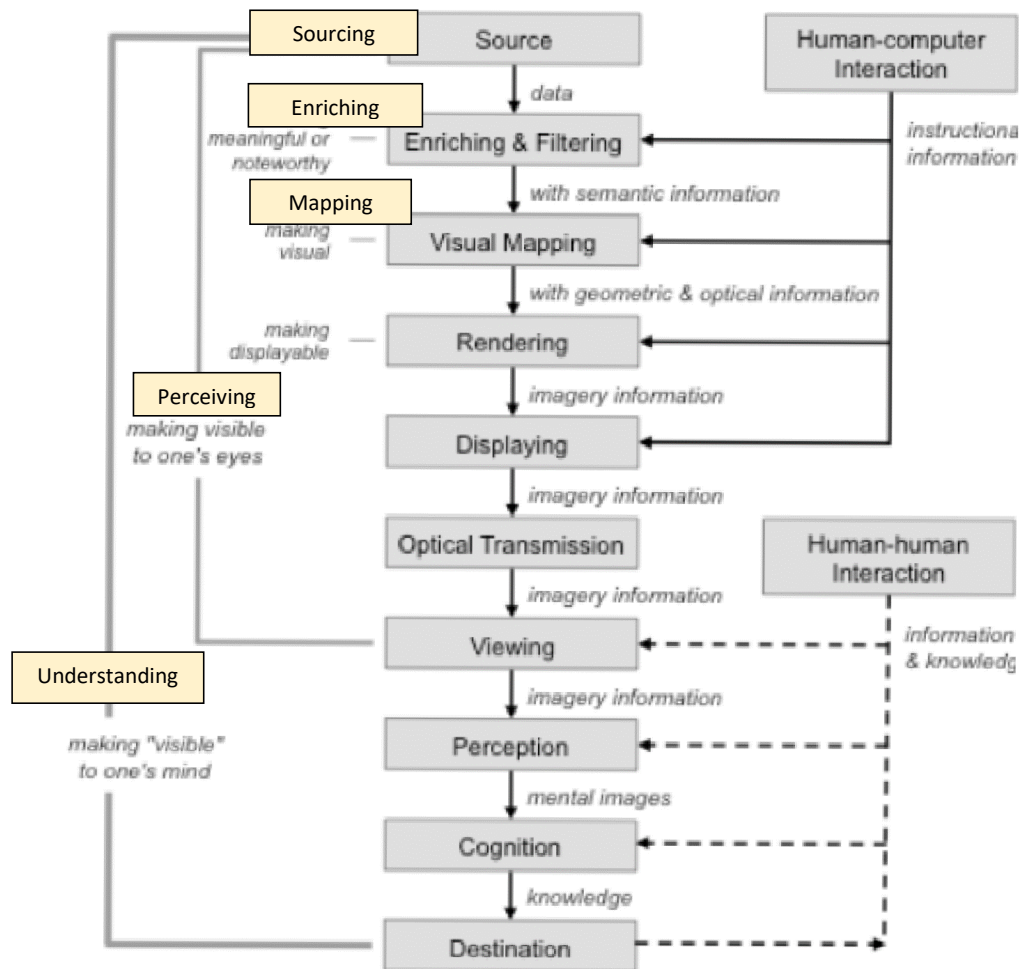


Figure 3.3: Typical visualisation pipeline illustrates perceiving and understanding context as parts of generating visualisation. We added categories to highlight the required contexts, the image has taken from [36].

on enhancing visualisation results or design, theoretically how they are acting the six core critical thinking skills across the visualisation pipeline.

Interpretation: In critical thinking this skill refers to understanding the problem, categorising the information and human ability to deliver the meaning of information to people. In visualisation, the interpretation might reflect the “understanding” context in the visualisation pipeline where information is being displayed where the needs for viewer’s thought processes and cognitive experiences for interpreting received information to understand how it is displayed, and what information intends to convey. Thus, practitioners could critically address what and how computers visualise information and clarify the meaning of challenges. Interpretation helps in conveying visual representations to viewers, that “understanding” part of pipeline support users seeking how to optimise

and promote usefulness and effectiveness of visualisation algorithms overall visualisation phases.

Analysis: Visualisation is the computer-aided seeing of information, while information holds various phenomena in the visualisation process. This skill helps to examine the possible solutions and ideas recognise the requirements, filtering information, categorising data, selecting the appropriate means to visualise data. In other words, the analysis in the visualisation environment probably must act on “percieving” part of the pipeline. To be able to classify important data in the data source (sourcing), quantifying information and structured data in the source are to be included in computational algorithms for filtering (enriching data), visual mapping and rendering (displaying structure). Humans interact with the creation process effectively if they feed the useful information to the process and detect and categorise what is important by breaking down the information into parts to be manipulated. For example, visualising spatial and temporal data on a Coordinate system, requires that the user examines different ideas and selects the best representation, such as in 2D Cartesian coordinates or 1D radial visualisations. They also need to select techniques that helps in exploration and interaction to visualise data such as by a zooming-overview, brushing, etc.. They might need to use colour to show the data changes over time and incorporate other ideas to help with the analysis of that data. All this “consideration” helps to create the best possible visualisation. This metacognition can help deliver a better depiction, whatever the skills of the developers have.

Evaluation: That means the ability to make decision and judgements about the process of visualisation. That could occur at the rendering stage of viewing part of the visualisation pipeline when users display visualisation and interact with the process by questioning themselves conceptually or quantitatively whether the goal is achieved or not. That is considered a temporary judgement in this stage before visualisation transmits into the “understanding” part to be cognitively analysed. Evaluation is deductive in a human’s mind for looking at alternatives and detecting errors either in visualisation results or in the design to put visualisation in the development cycle.

Inference: Critically determine the salient elements needed to draw reasonable conclusions. Introduce in the user's mind the parts of the display, such that they can make correct assumptions about the data. This part fits with the "understanding" part of the visualisation pipeline. Where users can cognitively form knowledge and accumulate experience to draw a conclusion about the results and judge that information relevant to deciding the acceptability. Such cognition helps to formulate (in the mind of the user) alternative solutions, and helps them to judge the validity of each solution, such to choose, accept or reject the different alternatives, and resolve problems, such to make final judgement on the information, based on the proposed evidence.

Explanation: After users have created the visualisations they need to share results and present the outcomes. Users need to describe the procedure or methodology and justify method to themselves or others. It is considered that a user has "good communication skills" if they can give a clear set of reasons for accepting or rejecting a particular hypothesis about a visualisation.

Self-regulation: It is important for practitioners to possess and use this cognitive skill, where users self-reflect and reason and verify both the results produced and the correct procedure used. Users can use deep thinking on the whole visualisation process and examine design problems and errors to recover mistakes and develop them as possible. And that is the critical thinking goal.

Conceptually, users can use these skills to interpret, perceive, understand and enhance the visualisation process. But, practically the general corresponding for typical visualisation process [35], [154], using existing visualisation techniques or programming libraries classes. We can consider the dataflow pipeline as three main processes:

- **Filter & Enhance:** loading the data, filtering the data, organising the sources, filing the data in the database, choosing and selecting the right demonstration data.
- **Mapping:** choosing the mapping and the design, creating the visualisation itself, deciding how the picture is rendered, and displaying it and transmitting it to the right screen.

- **Rendering and displaying:** using the human visual system to understand the visual communication, understanding the items displayed (cognition) and performing any interaction and alteration of the displayed material.

We now consider how critical thinking is applied to each of these three parts of the dataflow pipeline. Even before the user starts doing any visualisation design they will need to think critically about their data. With some projects data is unavailable; perhaps the user does not have permission to use the data (this can happen in the security field), other times the expert has a wish to collect the data, yet does not actually have the ability to capture the necessary information. In this situation, the researcher will need to collect the data, store it and process it. At the very start of the data-visualisation process, the researcher is already thinking critically, and deciding whether the quantity, availability, structure, or permission of information is “fit for purpose”. When the researcher has decided that they do have appropriate data, and that it is available to use, and that they have the right permissions to use it, they can move onto crafting a visualisation of their data.

- **Filter & Enhance:** Referring back to the core thinking skills of Facione [53], the researcher first must **interpret** the data. They must understand what they have, and what they can do with it. It may be required that the developer performs research, to discover how other people have visualised the data. Roberts [132] talks about performing a “visualisation triage”. The idea here is to perform a quick sanity check over the availability of the data, to make sure that the developer does have the required quantity of data, and that there is a need to perform the visualisation. This triage could be performed through a quick Internet search, asking an expert, or the developer themselves may have the right knowledge to make an appropriate decision. For example, the Internet search may present results to support the case that someone else has performed a suitable visualisation on the data, and have published their results and their code. This result could make the *need* of performing the visualisation moot, because the found code could obviously be used to display the data, and you as the developer need not repeat the task! On the other hand, even if a solution has been found on the Internet, it may be still be worthwhile that the developer does the work: perhaps they are a student, and have been asked to perform the task as an assignment.

Looking again to the skills that Facione list, the next skills is to perform a data **analysis**. Developers need to decide what types of data they have, and how the data is distributed and arranged. Perhaps the data is bunched together, perhaps it is distributed evenly across the data range. Critical thinking about the data, and performing an analysis of the data types, helps the user to understand how to map the “retinal variables” effectively.

- **Mapping:** Data mapping requires further critical thought. Mapping uses structured data at the output of the filtering data stage. In this stage, data will be mapped to geometrical shapes or objects with appropriate semiotic symbols (points, lines, shapes, etc.) and attributes (size, colour, position, texture, orientation, transparency). Dimensions must determine in this stage 2D or 3D visualisation and other features that are required to draw visual depictions such as the utilisation of space ratios that might have an effect on displaying data. The outcome of this stage is passed to the next stage, rendering and displaying geometric data on a proper display medium which depends on the visualisation task.

There are always many different strategies and designs that could be used to display the data. Given some imaginary data, developers could display it using length (such as bar charts), size dependent circles (such as a bubble plot), lines on an axis (such as a parallel coordinate plot), objects in a glyph (such as a star plot), etc. Developers need to apply critical thinking to decide which mappings are suitable, which mappings will tell the right story, which mappings are suitable for the end-user, etc. During the design process, developers need to **evaluate** (another Facione skill) each of the solutions to make suitable design decisions.

- **Rendering and displaying:** The final part of the dataflow model is to render the visualised data to the user. There are many critical decisions that need to be made here. Rendering visualisation can be in different forms either simple graphics or complex interfaces based on what data we want to visualise. Users can explore the outcome of this stage to analyse visual model results. The cognitive process stage in the visualisation pipeline defines human perception aspects and how humans interact with all stages of the visualisation design process. The human acts out the essential role in developing a visual design by exploring visual representation construction and viewing by making the displayed information visible to his mind and showing how the information is processed through interaction. A human can

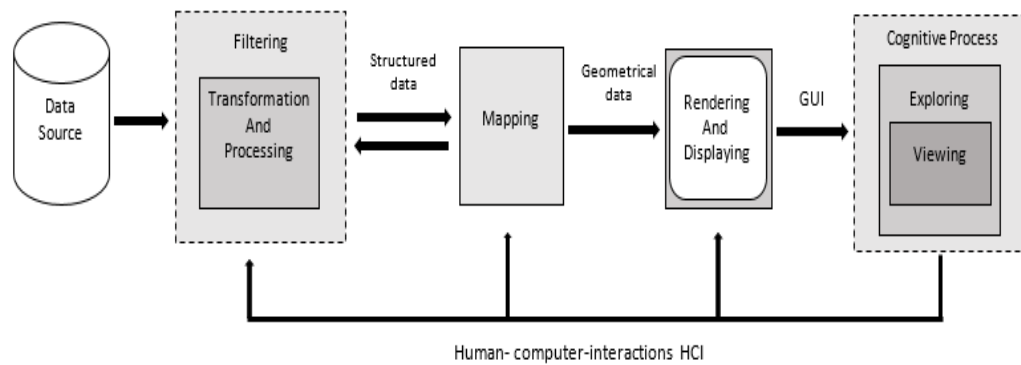


Figure 3.4: Typical visualisation pipeline stages. Synthesis ideas are taken from[163] the diagram created by the author.

interact on all visualisation stages, for instance, they start with deciding what data types should be present in designing, enhancing, filtering and organising data forms, what important mapping scheme decides graphical data structure by determining dimensions, and what proper rendering environment. This kind of mind information processing is applied to the visualisation model based on each stage's requirements; it expresses human ability to make judgements and reason findings and ideas. The cognitive process can include various perceptual activities here in this model exploring overarching aspects for the human mind, and other activities that might be abstract from this aspect. In summary, the visualisation model can provide a rigorous planning and thinking for desiring the visualisation task and performance.

One important aspect is that the developer needs to empathise with the end-user, and imagine how they would use the visualisation tool. In respect of this task, the developer needs to make **inference** and draw conclusions over the situation to put themselves in the shoes of their users.

3.4 Summary

This chapter outlines the critical thinking fundamentals and concepts regarding how users think and plan to solve problems. Using critical thinking skills as a tool in the educational system enhances students' understanding and builds critical thinking skills in them, that leads to improvements in self-assessment for class work. The analytic

domain benefits from ancient rhetorical canons in philosophical studies and uses similar principals in computational fields, which helps in making reflections on ideas with structured manners. This inspiration helps to support and enhance visualisation models' development with more respect to the human role to interact on the visualisation process. Design-thinking contributes in developing the design process for novices and experts. Inspiration has come from the iterative design process which reflect on designing our proposed tool, we insightful followed the design thinking framework in figure 3.2 to formalising and applying a cognitive sense in order to generate something helps users evaluate visual design. We are benefited also from considering humans as the main agent in design thinking approach, even our aim is helping users in critique visualisation rather than helping them in the design creativity. We rely on the creativity in design as a concept to propose features that possibly would be include in our tool, and that is accepted if these features meet with our assumptions and tool purpose for visualisation. Consequently, this thought participates in planning and generating ideas to create the tool, by following the cycle of understand and explore then materialise.

Chapter 4

Design of The Critique Design Sheet

This chapter will demonstrate how the critical design sheet CDS was planned, prepared, and designed as an evaluation tool for visualisation. The chapter focuses on the research question “is there a structure that users can follow to critique visual design and help to make self-reflection on design?” To achieve that goal, the following steps were followed: (Step 1) Performed qualitative research to explore ideas and strategies by questioning users, that can be done through conducting a critical analysis workshop. (Step 2) Collected data and performed the analysis on information from the workshop to summarise the outcome of the workshop. (Step 3) Created an initial structure of CDS after much deliberation and exploring alternative solutions. (Step 4) Incrementally improved the CDS quality through conducting pilot studies to detect design errors and problems of the CDS prototypes. (Step 5) Produced final design sheet to be evaluated by end-users. Figure 4.1 illustrates Critical Design Sheet methodology, the Design involves five main stages. The sections that demonstrate CDS creation and design are presented in:

- section 4.2: Workshop to explore design critiquing aspects (Step 1), and Outcome analysis (Step 2).
- section 4.3: Preliminary design of the Critical Design Sheet (Step 3) and Talk aloud evaluation (Step 4).
- section 4.4: The final Critical Design Sheet prototype (Step 5).



Figure 4.1: Methodology of design Critical Design Sheet chapter.

4.1 The vision

When a developer wants to create a tool, it is important to consider a set of questions about the software development project before starting to design the tool or upgrade of the software. For instance, Fritz and Murphy [62] conducted a pilot study with several developers asking them to summarise the questions they ask themselves when they perform any software development, each question requires the user to think across multiple domains and issues. As a result the researchers determined a list of 78 questions that developers ask about specific people, specific code, progressing of work and other relevant questions which are interested for developer's work. For the full list of questions that Fritz and Murphy discover, refer to their research [62], however to summarise their work, we include seventeen sample questions below, that show the variety of the questions the users collated.

1. Who is working on what?
2. What are they [coworkers] working on right now?
3. What have people been working on?
4. Which code reviews have been assigned to which person?
5. What is the evolution of the code?
6. Who made a particular change and why?
7. What classes has my team been working on?
8. Who has made changes to my classes?
9. How do recently delivered changes affect changes that I am working on?
10. What's the most popular class? [Which class has been changed most?]
11. Which features and functions have been changing?
12. Is progress (changes) being made on plan items?
13. Who caused this build to break? (Who owns the broken tests?)
14. How do test cases relate to packages/classes?

15. Which API has changed (check on web site)?
16. How is the team organised?
17. Who has made comments in defect?

Fritz and others introduced an information fragment model with a prototype tool that helps developers to follow links in a development environment and simplify the answering of these 78 questions. When users understand the answers to these questions they can develop better computing programs. It is clear, just by looking through these questions, that they can be grouped into several key categories. There are two questions that are useful to ask of these 78 questions, for this dissertation. First is “what types of questions are there”, in other words, how can we categorise these questions into a smaller set, such that we can make an indicative list of questions. And second, “how many categories are suitable”.

We need to summarise these 78 questions, because (1) it would be inappropriate for us to address all 78 questions as examples in a research for our model; 78 questions is far too many to get students to critically analyse their visualisation. (2) We also want to have a representative number of questions. In other words, we do not need to have specific questions, but we want more general questions, and certainly it is not necessary to determine 78 questions to develop a tool. (3) We need to have indicative questions that are appropriate for the visualisation domain. Fritz and Murphy are talking about software development and select their questions to show their model properties. (4) Therefore we need to amalgamate questions, and group them together where they are both addressing the same issue, consider the similar part of the computer programming process (e.g., at the start, middle or end of tool building).

To perform this consolidation phase, we used the methodology of “affinity diagramming”. The questions were put on individual sheets of paper (we used large sticky-notes) and then placed them on a large desk, such that each question could be seen, and was not obscured by another sticky-note. The position of the sticky-note was then moved together (or apart) from other sticky-notes, if we thought that they were similar, and that they were addressing the same type of issue, and expressed similar ideas; based on the four-part criteria, detailed in the previous paragraph. We performed this process over the course of two days, and by two people. That is, we performed an original grouping

on the first day, and then returned back to the affinity diagram on the next. This helped us perform careful and consider deliberation on the questions and helped us explore different potential categorisations. Finally, we consolidated the questions into 5 groups. This process helped us to analyse and categorise large transcripts of questions into a smaller set of categories under the above criteria. Then we summarised the different aspects and wording in questions in each group under one main title to consolidate 5 main questions, into the following five questions.

q1. who is going to use the tool?

q2. what is the purpose of the tool?

q3. what kinds of tasks should be performed using this tool?

q4. what is the form that the tool should be?

q5. what do we include to support the user's reflection on work?

The first question (q1) addresses the original vision. The tool is for predominately students or developers. It is meant to be used by someone who may be stuck, and unsure how to improve the design. This clearly fits the learner domain. Developers could use the tool, as they need to be systematic in their approach to improving prototypes. They vision is that they could use it as a way of being consistent across several projects. Like for the developer, an expert visualisation designer may benefit from the tool because they too could compare several designs. But the primary reason is to help learners.

The second question (q2) focuses on the reason for the tool. The reason for the tool is that it will be something that will act as an aid for users to create better visual designs. It is envisaged that the tool would be used as part of a designers toolbox. They create a design, can use the tool to help them improve it, and then they can change their designs appropriately. The vision is that the Critique Design Sheet method is a critical analysis method for evaluating the visual design. The aim of creating this type of evaluating method is to help learners in the visualisation domain to make self reflections on their design after creation and enable their critical thinking skills to be applied on design to analyse their works. This includes human perception and cognition

in the visualisation systems contributing to an effective role or features in designing and evaluation tools [154].

The third, fourth and fifth questions (q3,q4,q5) need much more discovery and investigation. It is unclear at the start of the research for the dissertation who it will be for and the purpose of the tool, but less clear to the main tasks. It is possible to think through some of the roles: to “improve” quality; to “aid” a user to decide on the course of action; to “prompt” to get to the next stage; to help the develop “reflect” on his/her work. But at the start of the research journey, it is unclear how this will be achieved. Lastly, the form of the tool is unclear. How will it be designed? What parts does the tool require? How will a user complete the tool? What cognitive properties should the tool include to help the user’s reflection? There are certainly many questions developers should address and think critically about regarding designing visualisation tools or methods. There are also many unknowns at the start of the research.

Consequently, this chapter focuses on the design and makeup of the CDS. It is clear that we will need to ask users, and explore how they critique objects. We will achieve this through holding an interactive workshop. From this workshop we will draw together the results that will create the underpinning knowledge necessary to create an appropriate tool.

The workshop is required because after the review of the visualisation literature, it is clear that there is little research focusing on critical analysis terminology and critique vocabulary. The vision is to seek a structure or process that users need to follow in order to critique their design; but no researchers have actually achieved research in this direction. Certainly, from the related work chapter it is clear that researchers have focused on evaluation, but while this research is excellent, no other researchers have provided something that can help users progress. No other systematic self-critiquing tools exist.

So, to achieve a suitable design of the CDS, it is imperative to achieve a workshop, to ask others how to critique and to explore the vocabulary of critiquing visualisations. This CDS fits in with other methodologies. The Critical Design Sheet (this tool) for evaluating visual design can be considered a step towards developing techniques that

will aid users in the whole visual design cycle, through making initial designs (using the Five Design-Sheet method [132]) to evaluate the effectiveness of these designs using the Critical Design Sheet (this dissertation), then taking action forwards and the user/developer then implements their design.

The next sections will focus on designing the CDS method prototype through conducting an observational study to investigate the strategies and terminology that would be used to devise the critique design sheet for visualisation.

4.2 Workshop to explore design critiquing aspects

It is evident from our literature research that, while there is a long history of critiquing oral speeches (from the Rhetoricians), they used critical analysis on written texts, lots of work on thinking and problem solving (e.g. [44], [48], [120]). From visualisation literature, there is a prior work correlate with critical analysis concepts such as [77], [89], [90], [138]; there is little research on the terminology of critical analysis, and the processes that people need to go through to critique, or how people critique modern visualisations. At the start of the investigation, we had the vision to develop a critical evaluation sheet, but we were not sure of the vocabulary to use, or how to organise the sheet. In order to establish a foundation for our work, and to understand how people make critical judgements on visual pictures and objects, we held a day-long hands-on workshop to explore critical thinking and critical analysis. The workshop was structured around several activities starting with broader questions on critical analysis, gradually getting more specific to critiquing information visualisations.

4.2.1 Participants

Ten participants took part in the workshop, split into two groups of five. We advertised around the University for participants. We wanted people with different backgrounds and experiences. Four participants were in work (one in marketing, one in finance and two in computing), five were postgraduate students in computer science and one was a

researcher fellow. They were six males and four females with ages between 25-40. Four postgraduate students had taken an information visualisation course, all had created charts in Excel. We show a photograph in 4.2 of the participants of group 1 from the workshop.

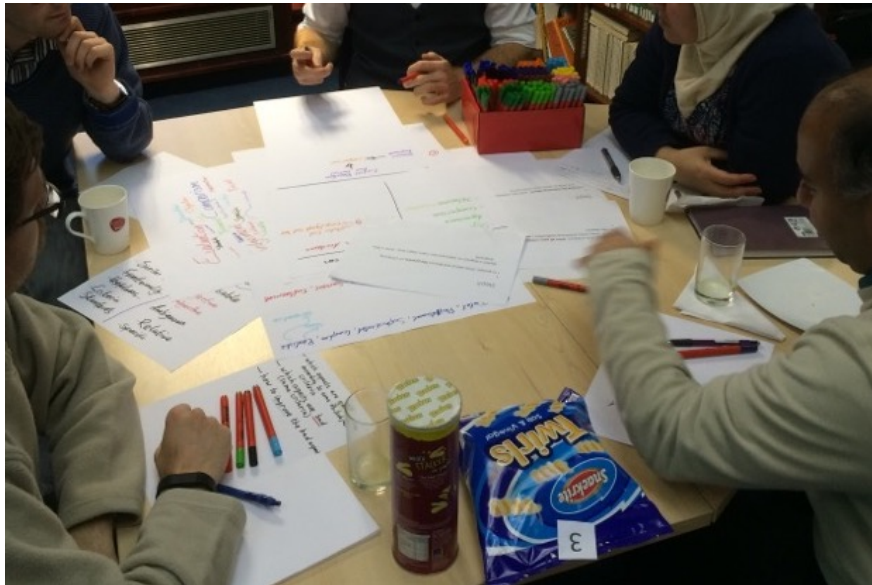


Figure 4.2: Photograph of participants at the critical-analysis workshop, which was held to elicit vocabulary associated with critical analysis and to understand processes and stages of performing a critique.

4.2.2 Workshop preparation and tasks

The principle aims (A#) for conducting and organising the workshop were:

- A1:** To discover the vocabulary of critiquing visualisations, to understand words that people use and definitions of critical analysis.
- A2:** To understand the strategies and structures that people used when they made a critical analysis.

The first aim is required because we needed to be able to place the right descriptive words on the critical analysis sheet. We had a vision that needed to include a list of words on the sheet that people could circle. These would then be used to summarise the visualisation they were critiquing. The second aim is required because we needed to understand how to structure the critical analysis sheet. What sections we would

need, and how does someone perform a critical analysis. The analysis and reflection of the workshop would then influence the design process to create a suitable sheet. To achieve the first aim, it required participants to use tasks that encouraged both individual thought and group discussion. For example, for the vocabulary question, we wanted to understand what an individual person thought about critiquing before being placed in a group, where their thoughts could be influenced by other members of the group. In particular, some people are less inclined to talk in group settings and their voice (and maybe their good idea) is not heard [81].

We planned in advance what tasks to ask the participants. We deliberated about the tasks and prepared the questions and materials in advance of the workshop. We came up with the idea that first we would ask some definitions to try to collect some of the words and vocabulary that the participants would use. Second we had the idea to get the participants to first critique general objects, before specialising on visualisation imagery. The idea of starting general and then specialising into the visualisation domain was a deliberate choice, so as to mitigate priming the participants with the organiser's views or allowing for our personal opinions to influence their thought and judgement of things by ways that we want, then they prepared themselves before coming. That could influence the introduction of the final topic. Indeed, at the start of the day we were careful to introduce the workshop, saying that we wanted to understand how people made judgements over things – we did not mention visualisation and creating a model to help people create better designs. Again, we did not want to influence ideas or decisions.

So as to be able to analyse the results we also needed to record the sessions. With two groups it would be possible to have two leaders take transcripts of the conversations. This is what happened, and the transcripts allowed post evaluation of the conversations. Also, photographs were taken during the session. It could have been possible to record the session in audio, but it was decided to take live transcripts to save time transcribing after the event, and live coding also means that knowledge of the event can be added at that time. For instance, we added notes when people were sarcastic or pointed at things (e.g., “Fred picked up two objects, just to compare them”). In addition, all the notes and sketches made by the participants were collated at the end of the workshop, and analysed.

We set two tasks (T#) to address the first aim (A1):

T1: Individually write down a definition of critical analysis that explains how to critique something. Underline key words in your definition.

T2: In groups, brainstorm words associated with critical analysis and produce a word-cloud. Aim for more than 15 words.

To achieve the second aim (A2), there is a need to analyse strategies of thought. However this is difficult because we are asking participants to engage in metacognition (thinking about thinking). Flavell [57] structures metacognition into four parts: the **person** doing the thinking and their **experience**, the **task** that they are performing and the confidence of the person to achieve the task, the **strategy** they use (perhaps the problem is similar to something they have seen before) and finally the experience (did the person actually find the solution and was it right?). Therefore, it is clear that we need to ask the participants to consider the process that they are following (metacognition), sketch a model of their process (to make their cognitive model explicit), and also to confirm and accept/reject their model (by using their model).

It is also clear that we need to address several potential priming effects: First we could not just ask participants to critique something and then ask them how they did it, without getting them to practice. We need to get them to critique several pictures and objects before asking them to think about their thinking. Second, we know from Flavell [57], Facione [53] and Duncker [48] that the person doing the critique comes with their own experience and knowledge. They may have their own biases and their procedures may be different for familiar objects compared to unfamiliar. Third, it is possible that the type of objects that they are evaluating could elicit different critiquing processes, e.g., critiquing a visual that is focused on art may be different to a visualisation that is more functional, such as a bar chart. Fourth, some people may struggle to perform metacognition and find it difficult to construct a model of their thinking processes. Fifth, group influences may engender different critiquing processes. Sixth, we also wanted to tease out a process that would be suitable across visual depictions and be general and intuitive for users as well. Therefore, to mitigate these potential priming effects we did the following:

1. Gave the participants several artefacts to critique and practice on.
2. Divided the participants into two groups so as to mitigate group biases and overcome any problems of group dynamics in one specific group. Also, if the groups become too large then the views of some participants will not be heard.
3. We encouraged the groups to develop their metacognitive model through discussion, and producing a sketch of their processes.
4. We then gave the participants some tasks, so that they could confirm/deny or adapt their critiquing process model.

The tasks now continue from the previous two tasks, were all done in small groups, were printed on individual pieces of paper and given to the groups when they had finished the previous task.

T3: In turn, critique the following six objects/images.

T4: Reflect on your critiques. Were all your critiques the same?

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 your critiquing process (e.g., you could draw a flow diagram).

T8: Using your process critique in turn the bar chart, parallel coordinate plot and radar plot visualisation.

The objects and images were carefully chosen. We decided on these objects after much discussion, needed a range of artefacts, some that were familiar, unfamiliar, some expensive others cheap, some well designed others not. For task 3 we decided on (i) a packet of luxury potato snack (chips) and (ii) a home-brand version, (iii) a famous painting (reprint of Vincent Van Gogh's sunflowers) and (iv) a child's painting of a sunflower Figure (4.4), (v) Minard's map of Napoleon's Russian Campaign, and (vi) a

modern visualisation presentation of Minard's map figure (4.3). For task 8 we used a bar chart created in Excel (4.5), a detailed parallel coordinate plot of food nutrients (4.6) and a radar plot of statistical information about mobile usage in Europe figure (4.7).

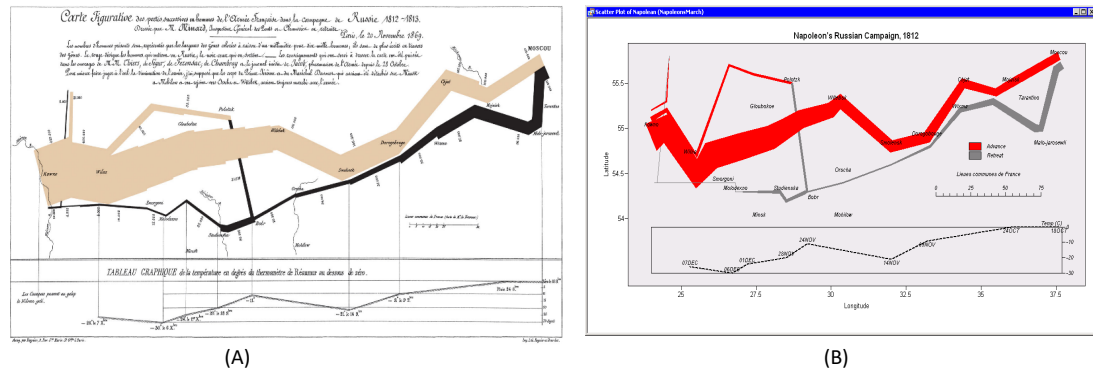


Figure 4.3: Workshop task 3. (A) Minard's map 1869 of Napoleon's Russian Campaign 1812. Two dimensions map with six types of data. An example of a flow map. It is published in the public domain and cited in Edward Tufte's book [156] p.40. (B) New production graphic of Minard's map programmed with object-oriented graphic display system integrated with IML client-server extension of the SAS Interactive Matrix Language[61]

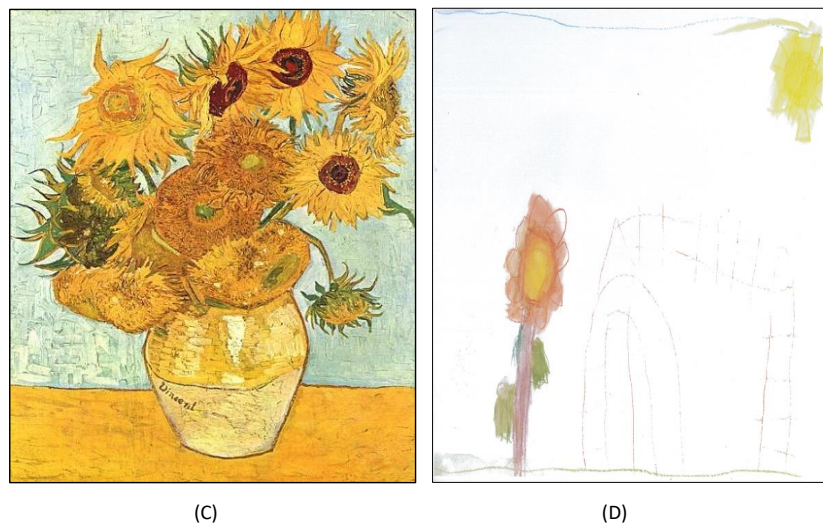
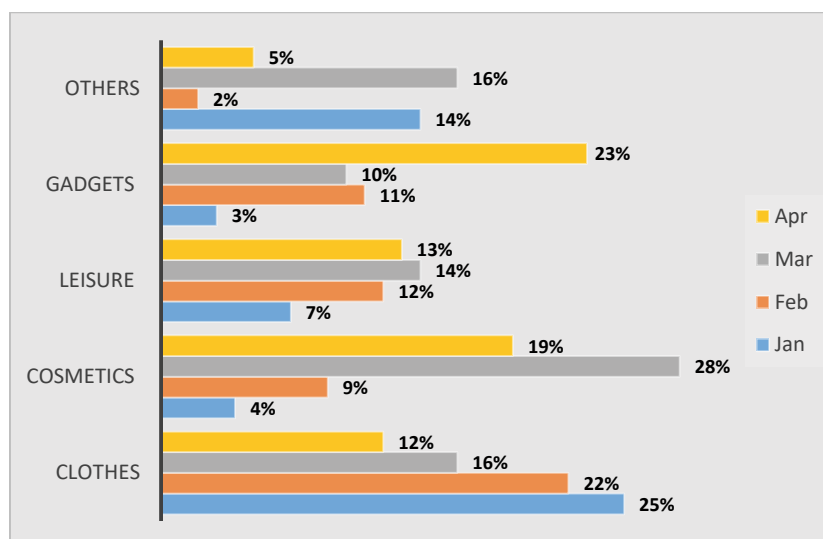


Figure 4.4: Workshop task 3: (c) Vincent Van Gogh's sunflowers painting published in Wiki public domain. (D) An example of children's drawing. The figure showed C and D painting, the participants should perform the possible comparison between these two paintings of low fidelity and sophisticated artist work during workshop time.

4.2.3 The workshop tasks analysis results: T1 & T2

In the workshop ten participants turned up and we split them into two groups of five. We deliberately kept the numbers small, such that we could observe the interactions and



Bars chart

Figure 4.5: Workshop task 8: Bar chart graph created in excel by author. Participants were asked to critique this graph.

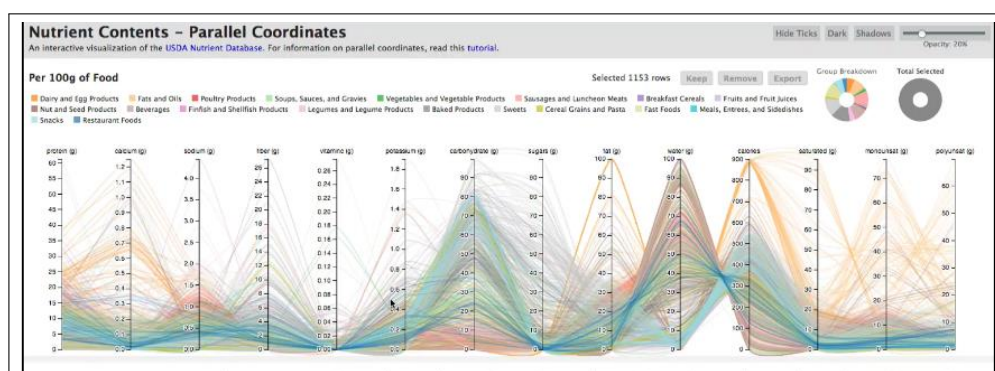


Figure 4.6: Workshop task 8: Multi-dimensional parallel coordinates visualisation shows nutrients content per 100g of food. The data published on USDA of Agricultural Research Service and the visualisation on food/nutrient-content-parallel-coordinates. The image from <https://visual.ly/community/interactive-graphic/food/nutrient-content-parallel-coordinates>.

transcript the discussions. Two observers were used to take notes (one for each group) and record conversations throughout the day.

Task 1 we received ten different definitions, we include nine for illustration:

“To use research, evidence, prior knowledge and logic to examine a claim for reliability, truth and applicability.”

“Critical analysis is the process of making an non biased and objective judgement of something through application of knowledge and research within a given applicable field.”

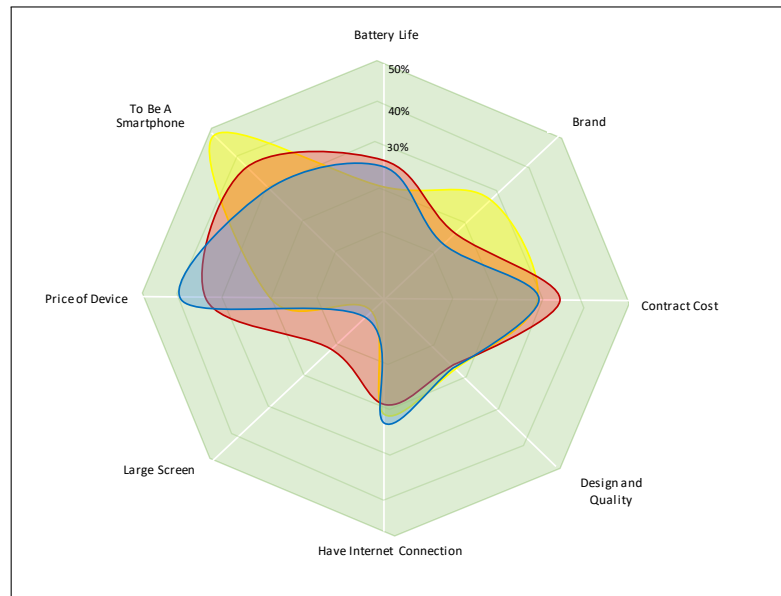


Figure 4.7: Workshop task 8: radar graph created in excel by the author. Open source data have used to create this visualisation taken from Global Mobile Consumer Survey(2014). The chart highlight iPhone’s users next phone chosen elements. Participants were asked to critique this graph.

“To observe and consider all the details of a subject without bias or prejudice and present this information.”

“To look at something with a critical eye. To consider what is being said and the use of the words and language. To weigh up both sides and give a reasoned answer.”

“To critically analyse something means to analyse something thoroughly, without any bias. For example, areas would be highlighted in which it lacks, as well as areas it performs well.”

“Evaluate the situation/event/object through your own eyes and attempt to understand, as much as possible, the multiple ways in which your own past and upbringing may have influenced your subjective opinion on said situation/event/object. Evaluate the situation/event/object through the eyes of your perception of the "opposite" viewpoint and do the same with regards to them. Identify, describe, and attempt to explain both viewpoints and also describe their similarities and differences, and resist the urge to demonise or idealise.”

“The act of studying a subject or situation in order to analyse and judge it before coming to a conclusion or deciding how to proceed. An important skill to have in life and the basis for how one approaches scientific research.”

“A critical analysis is a detailed evaluation of something, such as a performance, a piece of writing, etc. A list of improvements that can be made or errors that should be rectified.”

“To give a subjective view of a piece of writing, film, music etc. Having broken it's parts down and evaluated them. Then write a subjective opinion of the overall content of the parts. For example, film critics will critically analyse movies, and then give a subjective opinion of the movie, in a movie magazine etc. So as to give the prospective viewer, an idea of the what the movie is like, according to their (the film critiques) own opinions.”

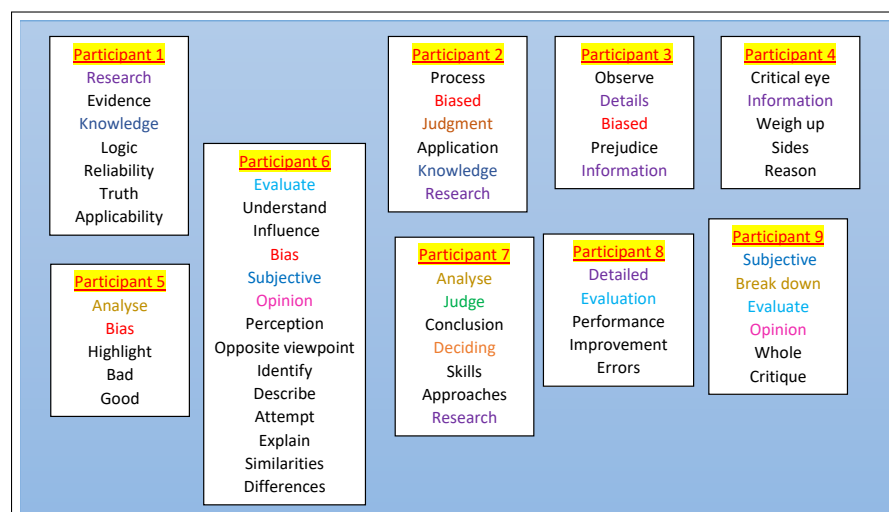


Figure 4.8: Affinity diagram for extracting words of participant's definitions. Each card includes the descriptive words that were mentioned in individual definition for the participant.

Transcripts or texts are the source in the discourse analysis research which presents the rich source materials for a variety of linguistic phenomena to analyse. In fact, many studies addressed different procedures for doing discourse analysis and that might be constructed or based on the nature of the discourse [121]. It is useful to address here what methodology we have used to analyse the participants script's definitions. We used a similar analysis strategy that was described as the thematic analysis [27], by extracting data into individual coded chunks, and a selection of that extraction will be used in final

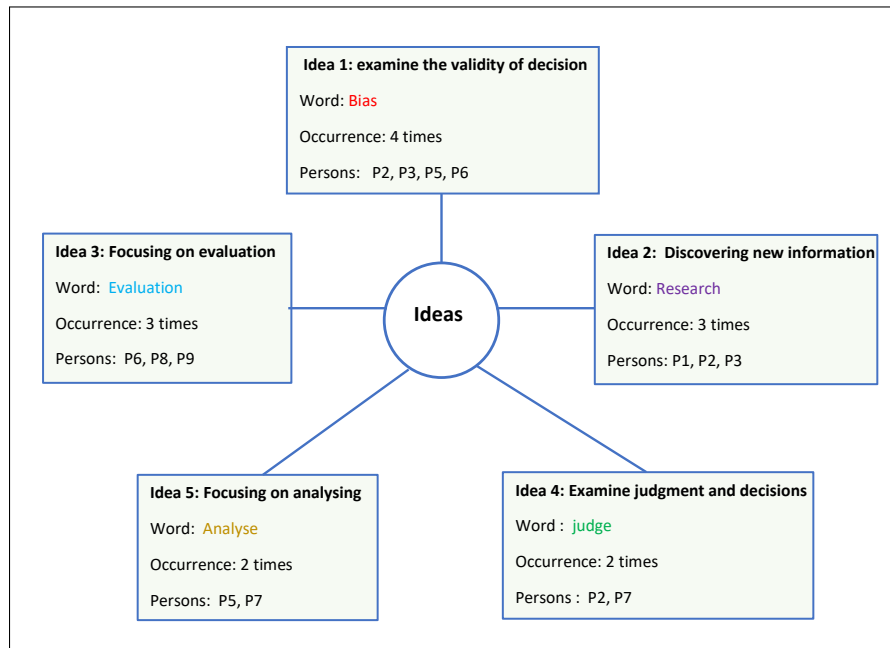


Figure 4.9: The refining outcome of words in figure 4.8 that is categorised under five ideas. Each idea has a title and detailed information of the words that reflects the idea.

analysis. Thematic analysis focuses on identifying, analysing and reporting interesting patterns/themes within data which guide the coding and analysing features in texts.

We benefit from the thematic analysis process with applying classic extracting and coding data in our analysis, in the thematic analysis process data passes through six phases in order to generate an understandable theme. With our analysis process we used similar aspects but not the exact process. Figure 4.8 illustrates the organisation of extracting description words from each definition into individual codes of data. We used affinity diagrams to order the potential words as individual cards. Then we began to find what patterns or features in the words could be used to collate them under similar ideas and reflect similar meanings, we gave the same colour for similar words even the words that hold similar meanings, for example “analyse” with “break down”, or “research” with “information”. We were refining these descriptive words in the second phase into five ideas, to generate new data coding by counting the occurrence of such words in different paragraphs. The new cards of ideas involve detailed information about the selective words that helped us to find such information to analyse them. Also, we were collating the similar words under one title that reflect a close meaning. Figure 4.9 shows the thematic map of the refining words under five ideas. After an analysis of these paragraphs, and the ideas that been contained within them, we will discuss these ideas and place them into four categories, as follows:

1. Four participants focused on idea 1, the idea of serious examination for validity of their decision by referring for the word '*bias*', they emphasised on performing in-depth thinking, such as to make fair and balanced judgements. One wrote "critical analysis is to observe and consider all the details of a subject without bias or prejudice and present this information".
2. Three focused on the idea 2, the idea of discovering information to support knowledge to reach an objective judgement, a critical analysis should be research to use evidence for weighing up the claims, for instance one participant wrote "To use research, evidence, prior knowledge and logic to examine a claim for reliability, truth and applicability" another wrote "[critical analysis is the] process of making a non biased and objective judgement of something through the application of knowledge and research within a given applicable field".
3. Two participants focused on idea 3 of examining the judgement and decision, and 3 of them focused on idea 4 evaluation and judgement, in which they refer to a similar goal. One wrote "critical analysis is the studying a subject or situation in order to analyse and judge it before coming to a conclusion or deciding how to proceed." While another focus on highlighting good versus bad, he wrote "areas would be highlighted in which it lacks, as well as areas it performs well". Which means, "[decide] which aspects are good, according to some objective or subjective criteria, [decide] which aspects are bad (same criteria)".
4. The rest of the ideas also focus on other aspects of critical analysis. Two participants addressed the idea of 'opinion' which reflects the subjective point of view with respect to the opposite viewpoint, one wrote "Evaluate the situation/event/object through the eyes of your perception of the 'opposite' viewpoint and do the same with regards to them."

The analysis of these ideas reflects the different perspectives for critical analysis. That words with the ideas hone our thought to put these words in our design consideration further.

Task 2 got the participants to discuss words as a group. Both groups tackled this task in the same way. One person was selected (by the group) as being the scribe, and

everyone pitched in words. They had a discussion over the words on the page, reflecting to decide if they needed to add (or remove) any words. They discussed which were the most important, and worked through the list of words, starting with the most important, checking them off, and transcribing them to another paper to make the word cloud. We show both word clouds in Figure 4.10.



Figure 4.10: The results of Task 2. Two word clouds of descriptive words associated with *critical analysis* are shown, sketched by representatives from both groups.

Regarding the analysis of this task we used the same procedure to analyse the vocabulary. Our goal is to discover the vocabulary of critiquing visualisation and what words people usually used. Figure 4.10 highlighted those descriptive words which give a strong association for critical analysis in a large font as a high priority order and gradually go down with the words of small font which reflects a weak relation with critical analysis. We looked at the words that the candidates underlined, at the words within the word clouds and from the transcript of their discussions. Words that they underlined were predominantly verbs, such as think, formulate, imagine, improve, analyse. While

these are interesting, we wished to highlight adjectives. Therefore, words from their underlined text were not useful, but we can usefully take the adjectives from their definitions, these included: good, bad, useful, clear, fair. The word cloud provides a useful set of words that describes what a critical analysis is, but again does not provide descriptive adjectives that can be used on a critical design sheet.

To further this analysis we looked through the complete transcript of the day, and all notes that participants made and found the following words: ambiguous, appealing, basic, busy, classic, colourful, confusing, 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.

4.2.4 The workshop tasks analysis results: T3 to T8

The transcript of the whole day, representing a talk-aloud of the conversations within the day, makes interesting reading. Most participants contributed and spoke their minds quite openly. For example, we follow the dialogue of one group answering Task 3, where they compared the child's artwork with the oil painting. Each of the five participants (P) contributed to the conversation:

P1: Let's consider the paintings.

P2: No idea about paintings, how are we going to critique this?

P3: Basic versus advanced. [pointing to each picture] I guess this is basic and this is advanced, some are complex ... but complex can be taken in a different light.

P1: One painting is minimalist. That looks like the artist has been copying something, whereas the kid's work is from their imagination.

P4: Maybe one influenced the other. [They added the word influence to the word cloud and continued discussing other aspects].

P5: 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.

This dialogue is useful because it reassures that the group were thinking appropriately and deeply over the tasks, and confirms ideas of critical thinking from Facione[53].

For **Task 3** the participants used many adjectives, their notes (for the six objects of group 1) are shown in figure4.11.

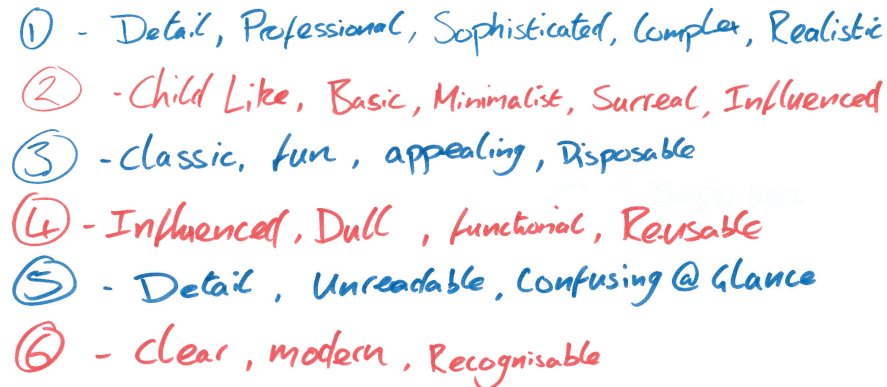
- 
- The image shows six numbered handwritten notes in blue and red ink, each describing a set of adjectives for critiquing an artefact. The notes are as follows:
- ① - Detail, Professional, Sophisticated, Complex, Realistic
 - ② - Child Like, Basic, Minimalist, Surreal, Influenced
 - ③ - classic, fun, appealing, Disposable
 - ④ - Influenced, Dull, functional, Reusable
 - ⑤ - Detail, Unreadable, Confusing @ Glance
 - ⑥ - clear, modern, Recognisable

Figure 4.11: Descriptive adjectives sketched by group 1 as answers to task 3 (critiquing the six artefacts).

For **Task 4** they explained that they needed to put their own emotions aside, and 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 and liked 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 [51], where users need to use logic, understand the criteria and be pragmatic in their approach.

When discussing the similarities and differences (**Task 5**) group 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 ideal of what it could look like. (4) Whether it was suitable for the context (or environment) where it would be used. For Task 5, Group 2 likewise included appearance and comparison, but also described that the initial reaction to the design was very important. Figure 4.12 shows a distillation of ideas to perform task 5.

Finally, addressing the last three tasks (**Tasks T6, T7 and T8**), the two groups sketched a useful diagram expressing the stages of their critique process. In fact, from the observation, both groups managed to sort out how to perform these tasks in same

Similarities	Differences
<ul style="list-style-type: none"> - Style - Appearance - Comparison - Influence → Context 	<ul style="list-style-type: none"> - Emotional Vs Academic - Context - Identity → photos Did → Crisp/graph Did Not

Figure 4.12: Analysis process for task 3 and task 4 to perform task 5 by writing explanations about similarities and differences for the given objects under four categories.

direction but they differed in the practice. For example, group 2 produced a useful four-stage process, they managed to: (1) think about the problem and define the terms of critical analysis, (2) brainstorm ideas and organise their thoughts, (3) produces a balanced review by putting emotions aside, and finally (4) reflect on the analysis to reject or accept the critique. Again this is useful, we can say that this model has similar concepts with the process described by Polya [120] on problem-solving. Polya's model describes a method of how students solve problems based on four phases, in all phases there are many questions to stimulate students thinking in the methodological procedure. So, the sketched model of participants can share similar features with Polya's model, where users should understand the problem first, then devise a plan or state a procedure by thinking what method they should use to perform the ideas or think and seek such similar solved problems to learn how to solve their problems, then carry out the plan and confirm that the right answer was gained by looking back to the start point and examine the solution. Figure 4.13 shows the result of answering Task 6, T7 and T8 for group 2. While group 1 sketched different critique stages, applied critical thinking and focused on evaluating the decisions. They referred to the term discussion and it is applied over each stage of the critique process, they used the peer review part of the process also, which is useful to focus on the human role in the iterative critique process. Figure 4.14 shows the critique process diagram for group 1.

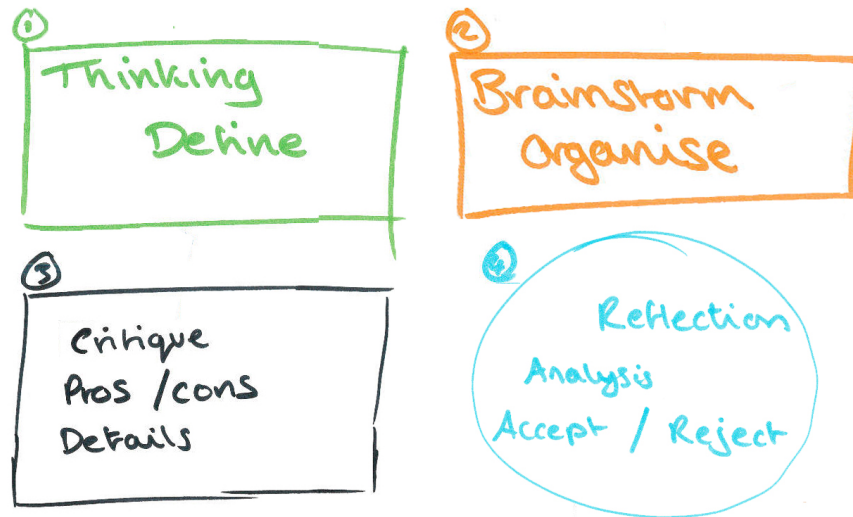


Figure 4.13: The critical thinking model, created and sketched by group 2, demonstrating four stages that simulates the four stages of Polya's [120] problem-solving model: (i) understand the problem, (ii) devise a plan, (iii) carry out the plan and (iv) examine that you have the right answer.

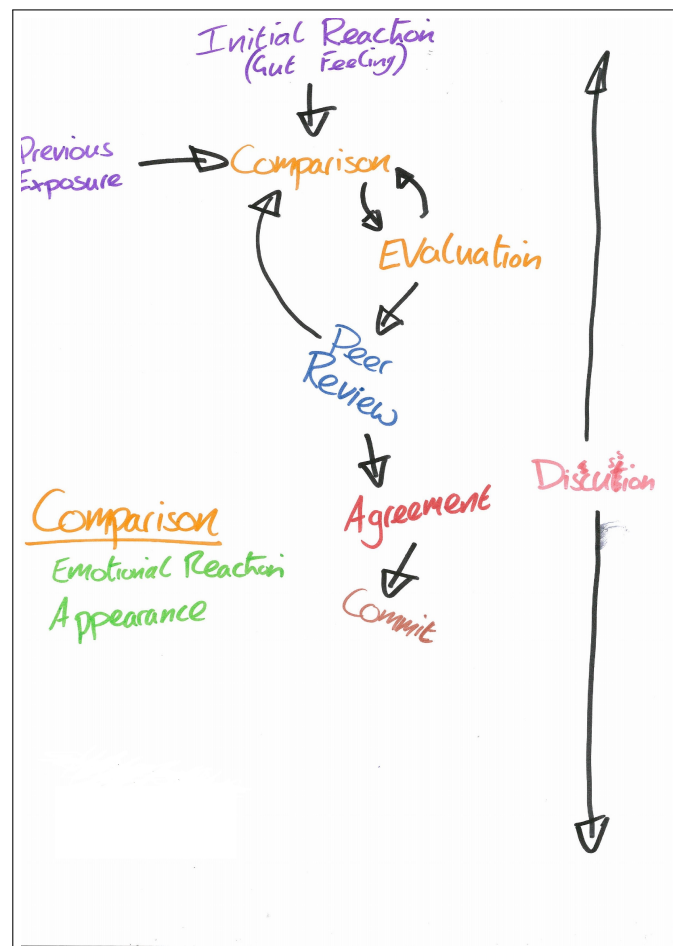


Figure 4.14: Critical thinking model created by group 1, they performed task 7 as they attempted to build a visualisation critique process.

4.3 Design process for the Critical Design Sheet

To develop the final critical analysis sheet we followed an iterative design method. We first reflected on the workshop outcomes and implemented an initial design (this

section), then evaluated our design with several pilot studies, incrementally improving the CDS after each study. We finally evaluated the method statistically and used it with end-users. This section details the design and pilot study.

4.3.1 First prototype design—decisions and influences

The workshop helped us hone our thoughts on critical analysis. Importantly, to learn from workshop lessons that we could organise the critical analysis into sections. The distillation of the observation study, the participants had made classifications of style (and layout), appearance, comparison with other ideas and to an ideal of what it could look like, context and initial reaction. We can see this distillation in figure 4.12 where participants categorised the properties of objects into four categories which was easier for them to make comparisons between objects. Second, we learned that descriptive adjectives are useful. Therefore, we had an idea of putting a series of adjectives on the sheet, for users to select and that clearly can be seen in figure 4.11 which supports our reflective idea. Third, from the workshop we also understood that there was a need to create something that could be used several times on the same design, and help the user see positive changes in their design decisions. We should look to the consistency of design decision each time we evaluate. After thinking in group1 critical thinking model figure 4.14, we can recognise the word “agreement” then “commit” which refers to whether people do approve and accept the evaluation value or not, and then do action based on that agreement. This developed the idea that we would therefore need to include a value or score, to quantify the information on the sheet to help users see if they make the same judgement again and again.

However, the analysis of workshop tasks enhanced and honed our reflections and vision of the proposed tool. From observations, we realised that people did not want to spend too long on their critical analysis, therefore we wished to have something that is simple and quick to evaluate. Inspiration can come from many places. In our case, we looked to the structure of existing questionnaires. In particular, we looked at the System Usability Scale (SUS) questionnaire. Developed by Brooke [29] in the late 1990s. The SUS measures the usability of a tool and consists of 10 questions that participants answer on a Likert scale. It has been widely applied to measure usability in websites, software

tools and visualisations etc. It is popular because it is simple, quick and easy to use, and the score is robust over time, and in fact many developers have analysed the scores over time, such as Bangor et al. in 2008 [12] and 2009 [13]. This inspires us, because likewise we would like to have something that is easy to use, quick and could be adapted for a wide range of usages. For the record, the ten questions of the SUS are as follows:

- 1. I think that I would like to use this system frequently.*
- 2. I found the system unnecessarily complex.*
- 3. I thought the system was easy to use.*
- 4. I think that I would need the support of a technical person to be able to use this system.*
- 5. I found the various functions in this system were well integrated.*
- 6. I thought there was too much inconsistency in this system.*
- 7. I would imagine that most people would learn to use this system very quickly.*
- 8. I found the system very cumbersome to use.*
- 9. I felt very confident using the system.*
- 10. I needed to learn a lot of things before I could get going with this system.*

The first design consisted of ten questions, and like the SUS we alternated between positive and negative wording of the questions. The advantage of this strategy is that it combats response laziness, minimising the results from participants who (say) merely checked Likert position 5 throughout. We developed a list of questions, such as:

- I found the colour map suitable, especially for those with colour deficiencies.

- 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. Especially, we did not believe that ten questions allowed us to perform an in-depth critique of a visualisation design. Second, when we tested these questions, we found it difficult to complete the responses by alternating positive to negative. The wording and the structure was too complex. We discussed this problem with a colleague, who had not been party to 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 to our goal. The SUS questionnaire is set for users, and the results are analysed by the developer after the questions have been completed. But we want the designer/developer to use it for their own work and make decisions on their own work. We wish for the developer to use the Critical Sheet in one sitting and in a short time window.

We therefore looked to other questionnaires. The structure of the User Experience Questionnaire (UEQ) [93] was another questionnaire that inspired our work. The UEQ measures user-experience through a set of six categories: Attractiveness, Efficiency, Perspicuity, Dependability, Stimulation and Novelty. The set of questions are shown in Table 4.1.

Users complete a 5-part Likert scale of 26 questions (see Table 4.1). Evaluating, for instance, if the tool is annoying or enjoyable, creative or dull, valuable or inferior. Like the SUS, the UEQ also alternates the positive and negative questions. However, while the UEQ is gaining popularity it is not as popular or used as widespread as the SUS. This is probably because the UEQ is more time-consuming for the participants, and more complex (and thus time-consuming) for the experimenter. One of the advantages of the UEQ is that the developers have statistically analysed many users; creating weightings of average-user responses [93]. However, this further complicates how the results are calculated, requiring the use of a spreadsheet. We did not want to require users to input

Table 4.1: The User Experience Questionnaire set of questions, with the 7-Likert scale and alternating positive and negative rows, Laugwitz et al. [93].

	1	2	3	4	5	6	7	
annoying	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	enjoyable
not understandable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	understandable
creative	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	dull
easy to learn	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	difficult to learn
valuable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	inferior
boring	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	exciting
not interesting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	interesting
unpredictable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	predictable
fast	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	slow
inventive	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	conventional
obstructive	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	supportive
good	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	bad
complicated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	easy
unlikeable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	pleasing
usual	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	leading edge
unpleasant	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	pleasant
secure	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	not secure
motivating	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	demotivating
meets expectations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	does not meet expectation
inefficient	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	efficient
clear	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	confusing
impractical	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	practical
organised	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	cluttered
attractive	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unattractive
friendly	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	unfriendly
conservative	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	innovative

their values into a spreadsheet. We wanted to have a quick and simple way to calculate the scores.

4.3.2 Design of the six principal categories

As mentioned in the last section, the UEQ measures the user-experience of usability devices properties through six categories of measurements helped to inspire the creation of the six-part structure of the CDS. One of the participants of the workshop pointed out his insight thinking on visualisation and said

“When we are critiquing we need to be clear in 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". Quote from a participant at the workshop.

The ideas which that participant was probably referencing is Shneiderman's mantra of "overview first, zoom and filter, details on demand" [143]. While, this mantra is excellent at summarising the important elements of interacting with information, his mantra is best applied to a specific visualisation for a particular purpose as a design principle, rather than a technique to criticise something. But the quote from the participant, and Shneiderman's mantra, and the structure of the UEQ all suggest that a user should criticise different aspects of the visualisation. That a user would need to think systematically through a series of stages, and evaluate the imagery from a number of different perspectives. In other words, a user would need to look at the work through a series of different *perspectives*.

Therefore, this got us thinking: "what are the different perspectives that a user would need to use?" "What categories are there in visualisation"? This consideration helped us make the following thoughts:

- First, it is clear that a user has to be reading the visualisation. Therefore, the first *perspective* should be focused to the user. This is sensible, much of the Human Computer Interaction literature asks the developer to consider "the user". So, our first perspective is to consider whether the design is suitable for the user.
- Second, the visualisation is used within an environment, whether this is a computer screen or a mobile phone. A visualisation may be suitable for one environment, yet not for another.
- Third, it is clear that there are many facets of a visualisation (such as windows, menus etc.), so thinking in the constructions and the components of this visualisation is there any consistency, are there aspects of the design that can be quantified and evaluated? We could consider questions such as: Is it an easy to understand layout? Is the visual appearance of the layout and menus etc. and suitable for the task? etc.

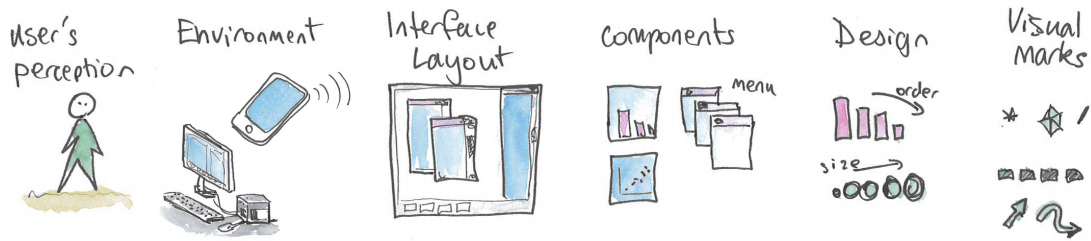


Figure 4.15: One of the principle parts of the Critical Design Sheet (CDS) are the six main categories of self-reflection: user's perception, environment, interface and movement, components, design and visual marks. These help the appraiser to make more focused critical judgements on the design.

We are contemplating these ideas in a hierarchy. At each stage the person doing the critique can then tune their thoughts from rough visual design aspects to finer details. These initial thoughts then led to a more refined set of six stages. In our final design, we start off with the big picture, think about the user, and then drill down into thinking about the individual graphical marks. We show this hierarchy in Figure 4.15. For each level users can ponder a range of questions. We define the categories below:

- **User's perception.** The visualisation is created for someone else to use. The end-user comes with their own skills, abilities and knowledge. We need to critique the design for its suitability for that user assess user's comprehension for that design. Table 4.2 explains set of measures and description for that measures as potential questions that be useful for user's mind to project his understanding on design.
- **Environment.** The environment describes the technology that will display the visualisation. It could be (say) print form, e-magazine, smart phone, tablet, desktop computer or power-wall. The position of the user could change, e.g., on site with a mobile tablet or sat at a desk. We mean by that what is the right place that visualisation will be used or displayed. Table 4.3 shows the possible set of measures that help users to think about the appropriateness of a design environment.
- **Interface layout.** The analysis of interface is dependent upon the visualisation purpose and what tasks that would perform. Number of frames, menus of commands, objects, presentation areas, etc. These elements are important to construct a visualisation interface. Usually, a human interacts with the

visualisation through an interface. In a traditional desktop environment the user inputs commands through a mouse, results are displayed through visuals and sound. The user can move virtual windows on the screen. On a tablet the user swipes with their fingers. In immerse virtual reality the user can input using their body, and see the results in three-dimensions. Table 4.4 explains the possible measures that help users to think and critique design interface.

- **Components.** A component is an identified group within a visualisation. Traditional components include bar charts, scatterplots, menu commands, legends. The components could be displayed in different ways, such as a grid (a matrix of small multiple views) or in a tabbed window. Table 4.5 shows the possible measures that help users to analyse design components critically.
- **Design.** Good design can refer to the organisation of any part of the system. This reflects how aesthetically a design looks. The aesthetics here, not exactly the same artists perspectives, rather than it reflects how graphically the design has been organised. For traditional visualisation and depictions on a desktop computer we would need to consider the balance and use of colours, alignment of items, styling etc. Table 4.6 describes a set of measures with explanations about intending meaning.
- **Visual marks.** A visualisation is made from many graphical marks; lines, polygons, circles to colours and texture [16]. In this category we can consider the layout of the visual marks such to consider if they are too close or overplotted. Other issues can be considered, such as redundant ink or chartjunk. E.g., chartjunk does not necessarily need to be bad, it could make the visualisation more memorable [22]. Table 4.7 shows the possible visual marks measures that help users to think about graphical properties in visual design.

These are intended main categories which considers our measurements that help to analyse and critique visual design. These design categories involved several items that measure specific visual design details. We include a list of questions that appraisers will find useful as they perform their critical analysis.

Table 4.2: USER’S PERCEPTION GROUP group measures. With these questions users take a holistic view of the whole tool/visualisation or system that being create. users will need to project their minds into how they understand that design and it’s purpose.

	Item measures	Descriptions
1	User’s Perception	Would this design be clear to them? Would they be able to understand the visualisation? Is the task clear? Is it understandable? Is it fit for the user? If a user eye-balls the picture would they understand what it was trying to present?
2	Understandable	Do you think the end-user would understand it? Is it clear? While it is difficult to project your mind into the end user, consider whether you understand what is going on.
3	Assumptions	Does the visualisation make excessive assumptions? Perhaps it relies on domain knowledge. Maybe it needs too many skills to use the tool, requiring long training. In the context of the user/task is this suitable? If it to be displayed in the public domain then it should be clear and make few assumptions of the users! If it is for an expert user, then maybe it can require more skills and knowledge to operate, use or understand.
4	Trustful	Would you trust the presentation of data in this visualisation? Does it imbue you with confidence? Would it be something that you would recommend to your friends to use? Tufte [156] presents a lie factor, to estimate whether the graph is biased. While we are not suggesting you perform a lie factor analysis, we are getting you to think whether the visualisation distorts the truth in any way.
5	Useful	Is there a clear use for the visualisation. Is there utility? Is it something that would be of benefit to someone?

4.3.3 Pilot studies (talk-aloud evaluation)

The talk-aloud style of evaluation methods has a wide use in psychology and information systems. The think-aloud technique has been used for different purposes and methodologies, for instance, particularly in Information Systems, it can be used in the development cycle of systems tasks [64], [162], and in usability studies for computer-based systems and interfaces [33], [99]. In this section first, I will focus on a think-aloud protocol which would benefit my research purpose. Second, we used the talk-aloud technique instead, which is quite varied from the think-aloud in the framework, but they are joined for the purpose. Talk-aloud has been used for asking participants about their actions when performing the given task, participants announce their actions orally or verbally [52] rather than their thought in such silent protocols.

Table 4.3: ENVIRONMENT group measures. Describes the design environment , users need to think of the appropriateness of where the visualisation will be used, the circumstances of its use and how a user would operate it or interact with it in that situation. It is all about appropriateness for the environment.

	Item measures	Descriptions
6	Environment	For this category take a holistic view to decide whether the user can perform their required task with the tool/visualisation in the given environment.
7	Technology	Is the right display technology used? It may be paper printout in a magazine or it may be a desktop, whatever technology used, is it suitable for the visualisation type and system used?
8	Interaction and/or animation	Is there the right interaction functionality? Can you do the things you want to do? Is the interaction organised appropriately? Can you scale when you need to zoom in? If there is no interaction (e.g., it is a printed visualisation, and nothing can be altered) is this appropriate? If animation is used (instead of interaction) is this suitable?
9	Output size	Is the size of the output device right? Too small and it may not show all the data (e.g., a mobile phone has a smaller display than a powerwall, but sometimes portable systems are better than big installations).
10	Place	Does the environment (where the visualisation is displayed) give the right vibe? Is it a positive experience? Is there appropriate ambience? Is it suitable for the task and the user? Is the context where the visualisation will be used suitable for the visualisation!

Table 4.4: INTERFACE LAYOUT group measures. Reflects the general principals of visualisation layout design in terms of organisation or task suitability.

	Item measures	Descriptions
11	Interface Layout	Is it suitable interface? If tabbed interface is this good? If cascaded or tiled interfaces are these suitable for the visualisation display?
12	Layout	Is the organisation and relationships between the frames clear? Is it ergonomically designed?
13	Frame size	Are the sizes of the containing windows/frames suitable for the interface and the visualisation task?
14	Spacing	Is there wasted space within the layout of the frames? Is spacing used efficiently? E.g., if lots of wasted images, this may be chartjunk, and may not be suitable.
15	Quantity	Is there the right amount of windows? Too many (may confuse), too few may not be showing all the data.

The think-aloud protocol is a type of qualitative evaluation method which is conducted as part of the design process [148]. Gathering data through qualitative methods has different forms such as field notes, artefacts, video tapes, audio tapes, computer records and logs.

Table 4.5: COMPONENTS group measures. Components are the units of design. They are often placed within frames or windows in an interface. Example components include bar chart, timeline, treemap, scatterplot, etc. They are unique and identifiable parts of the visualisation.

	Item measures	Descriptions
16	Components	Think about the individual parts and the content of each window/frame. Take a holistic view and consider if the components are suitable for the user.
17	Type	Type affords an interpretation. Can we read off maximum and minimum values, or compare values (as required)? Is this type (e.g., bar chart, line graph) correct for the purpose of this presentation?
18	Relationship	Are relationships between different visualisation components clear? Is it clear how one view is related to another? Would a user understand this relationship? Do you understand how each view is related to a legend?
19	Task suitability	Is the choice of the view suitable for the task? E.g., if it is a bar chart – is this suitable or should I use a scatter plot? If I have a pie chart, is this suitable for your purpose?
20	Structure	Is the organisation of the parts suitable (e.g., things close together are easier to compare)? Does it look well planned? Does it demonstrate the data? Is it clear how the individual parts of the visualisation fit together?

Table 4.6: DESIGN group measures. Reflects user's thoughts on design principles, there are many good design principles to follow (e.g., see Munzner [107], Ware [164]), think CARP: Contrast (also, figure/ground), *align* items in a same way, *repeat* items, or styles throughout and be consistent. Put together things that are similar (or need to be compared).

	Items measures	Descriptions
21	Design	Consider the whole design; is it suitable and well designed?
22	Aesthetic	Is it beautiful? Is it well laid out? Does it follow good design principles?
23	Space utilisation	Is there wasted space (maybe the design is on the left hand side of the page — is this good?) Does it cover the page/screen appropriately?
24	Coverage	Does the design show all the required data (does it include all the necessary data)? Does it represent the right quantity of data? Is aggregation correct?
25	Legend/labels	Are there legends, labels etc? Is this context-giving information correct?

The common qualitative methods for collecting data is conducting observation and interview approaches. This helps users focus on their thoughts to perform the tasks which encourage the learnability of users to learn and explore computer systems [160]. Alternatively, it is used to evaluate the usability of websites, for example, Benbunan-Fich [15] conducted an experiment to examine the usability of a commercial website using the think-aloud protocol to detect usability problems, poor navigation and cumbersome

Table 4.7: VISUAL MARKS group measures. Questions that help users to think about the graphical marks that are used on the page, and how they are put together to make the visualisation. Are the right marks used, in the right places, with the right attributes?

	Item measures	Descriptions
26	Visual marks	Are the marks on the page appropriate? Can them be distinguished? If the task is to compare, can the marks be used to compare, etc?
27	Mapping	Does the mapping (transfer function) match with the data? If the data is linear, is a linear Transfer Function used?
28	Mark types	Are the marks appropriate? Is the choice of length, size, texture, transparency (etc.) correct? Is the colour map correct and suitable for the user? Is it accessible?
29	Scale/zoom	Right size of marks? Too big and everything is messy, too small and you cannot see anything. Is the data shown at the right zoom? If it was zoomed in, would this still be fine? Would it make it clearer with a non-linear zoom?
30	Overplotting	Are there too many things too close, and on top of each other?

interactivity. There are many positives that think-aloud technique provides; it leads to thinking more insight to detect usability problems, it enables hearing participants and listens to their suggestions, plans and frustrations. However, the advantages of this technique are, that there is a limitation of using it to some extent. It suffers from a lack of realism due to the presence of the evaluator or experimenter and carefully selects participants to be aware of their thoughts to gain concurrent verbalisation.

In this research, during our design process we performed three pilot studies. We need to test the design problems for initial prototype of the critical design sheet then iterate to improve. Figure 4.16 shows the initial design of the CDS, where the categories of this design are very similar to the UEQ questionnaire system.

The first talk-aloud session was held with one volunteer. They were asked to talk aloud and say their thoughts as they not only approach the CDS but as they use it. Such as not to prime their thoughts and initial reactions, they were told that they were going to perform an evaluation using a questionnaire and then use it. Six visualisations were prepared for the participant to use. This whole session was video taped, and was conducted in a closed room with the volunteer, and two researchers, and the camera.

The first talk-aloud was very important. It became apparent that some of the questions were unclear, and that the wording was confused. What was interesting, though, was

that the participant started to be picky about some minute detail. By focusing on the detail we can conclude that the overarching goals and aim of the sheet were clear, and that further iterations would need to focus on the wording rather than re-designing the complete CDS structure. This first version also contained alternative (positive and negative) questions (much like the SUS and UEQ questionnaires). However, we discovered that this has a huge disadvantage. The talk-aloud volunteer really struggled to calculate all the values. It did not take long to calculate the overall score. In both the SUS and UEQ the calculations are done off-line (not as the participant completes the questionnaire) and also there were achieved by a spreadsheet, not by hand. This is an important difference. We are asking the participants to make the calculation themselves, and expecting them to do it fast. Consequently, after this think-aloud session we need to think again about how we can get the user to calculate the overall score – making it quick and easy to achieve.

We made several considerable changes from the feedback of the first talk-aloud session evaluation. Figure 4.17 shows the second versions of critical design sheet. The following changes were made:

- First, we adjusted such names of the individual questions. For example, we originally had labelled question 10 on the sheet as “*ambience*”, intending that the user would consider whether the visualisation fits within the environment and that it provides the right impression. Yet, after feedback from the second talk-aloud session we changed this to “*place*”.
- The participants also made suggestions to improve the values of the Likert scale. Originally, we had valued the Likert scores 1..5 but after considering the feedback we adjusted the values from -2 to +2, with 0 the mid-point. This also makes it much quicker to add up all the values and calculate the score for the overall grading of the design.
- We changed the star plot, such that the lengths of each parts were the same.
- We also changed the overall structure of the sheet. Initially we had one question per category, and then the remaining two questions per category. The idea was to put the most salient questions first, and then ask the participants to complete

details. We see this in the top-left of Figure 4.16. In the new version, Figure 4.17 we joined all these questions together into one long set of questions.

A second and third talk-aloud session was performed to confirm the details of the questions. For each talk-aloud session we used two postgraduate students (one working on visualisation and the other on virtual reality). We recorded the sessions on video and took notes on their comments. We used six different visualisations (including line-graph, bar chart, parallel coordinate plot) which they worked through, in turn. The sessions lasted 90 minutes, because we wanted to allow them carefully consider each part of the critical design sheet. We also wanted to ascertain whether the questionnaire was easy, quick to complete and comprehensive.

Critique Design Sheet:

Description:

In this sheet a critique method for visualization, user can use this critique design method to find fast judge on his design and the way to improve it. Please answer the following questions in scale from 1 to 5 (1,2 for negative respond 4,5 positive respond and 3 is neutral) and put your selection values on star plot and calculate the average. Follow the structure to give final judgment on your design

Name

Title

Outline categories questions: circle your choice

Human perception	Environment
Q1- Do you find this visualization understandable? <div style="text-align: center;">1 2 3 4 5</div>	Q2- Does design displayed in suitable places? <div style="text-align: center;">1 2 3 4 5</div>
Interface	Components
Q3- Visualization interface simple? <div style="text-align: center;">1 2 3 4 5</div>	Q4- Is the number of component sufficient? <div style="text-align: center;">1 2 3 4 5</div>
Design elements	Retinal variables
Q5- is color suitable with this visualization? <div style="text-align: center;">1 2 3 4 5</div>	Q6- mapping data relevant with retinal variables? <div style="text-align: center;">1 2 3 4 5</div>

Shorten detailed questions: shade your choice of this questionnaire

		1	2	3	4	5
Q1	a Difficult perceive information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	b Unpleasant design	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Q2	c Unsuitable screen size	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	d Need interaction	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Q3	e Conventional design	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	f Need knowledge	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Q4	g Context not fit with purpose	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	h cluttered	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Q5	i Inefficient design space	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	j Unsuitable design elements	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Q6	k Bad marks choice	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	l Unclear data mapping	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Visualization terms (guidance): Threshold value: 3; 1 ≥ average ≤ 5
Average > threshold: good; average < threshold: bad (need to improve);

Visualization results:

Mean value (average)

Weakness categories:

Strength categories:

Visualization judgment:

How to improve:

Figure 4.16: (A)The first prototype of the Critical Design Sheet (CDS).

Critique Design Sheet :



shade your choice for the following questions

	1	2	3	4	5
Q1- User's impression?					
a Understandable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b Assumptions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c Trustful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d Useful/utility	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Q2 - Right environment?					
e Technology	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f Interaction	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g Output size	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h Ambience	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Q3 - Interface layout?					
i Layout	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
j Frame size	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
k Spacing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
l Quantity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Q4 - Components?					
m Type	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
n Relationship	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
o Task suitability	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
p Structure	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Q5 - Aesthetics?					
q Aesthetic	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
r Space utilization	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
s Coverage	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
t Legend/labels/tick marks	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Q6 - Visual marks ?					
u Mapping (match)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
v Mark types	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
w Scale/zoom	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
x Overplotting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Number of choices

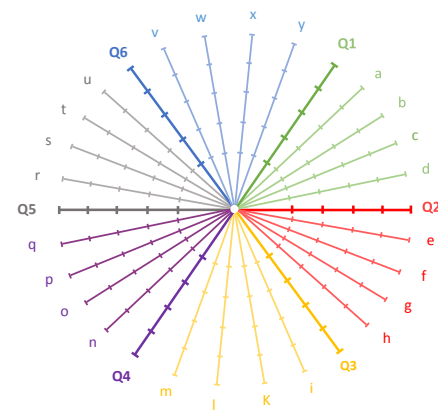
x	x	x	x	x
-2	-1	0	1	2

Result

The Total

User name

User title



Visualization results:

expected average value:

Weak categories: {choices mostly < 3}

Strength categories: {choices mostly > 3}

Visualization judgment:

How to improve:

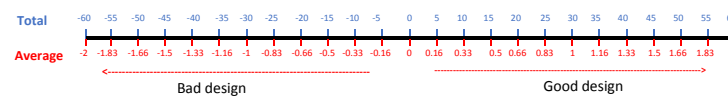


Figure 4.17: (B) The second version of Critical Design Sheet prototype.

4.3.4 Design of first-impression words, grading & star plot

One of our primary goals was to record first-impressions onto the critical design sheet. This idea was also mentioned during one of the talk-aloud sessions when participants suggested including the user's emotion and his/her gut instinct feeling firstly before starting the questionnaire. After an in-depth discussion we included several emotional adjectives to express the user's impression. We can consider this impression as kind of cognitive simulation for the humans instant thought towards objects before looking at objects insightful. To achieve how to design these impression words, we turned to (i) the descriptive adjectives that were summarised from tasks 3 to 8 (see page 83), (ii) the analysis of the workshop transcript (see subsection 4.2.3) and (iii) we added some extra words that we (the authors) had been using in development of the critical sheet. After

discussion and the pilot studies, we finalised a list of 20 words (placed in a 4x5 grid at the top of the sheet, see Figure 4.21), and the words are shown in table 4.8. The order of the words is meant to be random, such as to stop someone merely choosing the first five words.

Table 4.8: The First Impression words, that were finalised. The user is meant to circle five words that best describe their visualisation to critique.

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

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, therefore we used scores from the SentiWordNet3.0 lexicon [10]. 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 [147] all scores are between 0 and 1. Using sentiment score helps us to classify these words on rigorous and scientific bases to distinguish between positivism and negativism of the words. We used a combination of adjectives (a), nouns (n) and very few verbs (v) as word types. With multiple sentiment scores, we selected the word's type that is more convincing in context and commonly used between people.

Figure 4.18 shows the table of sentiment words scores with definitions and examples for each word, in the different positions in which these words have been used. For that the word's number refers to the various situations in which these word appears. The high score of a positive means the word has positive meaning, and the opposite is true. Equivalent scores refers to the neutralism or sentiment words that have zero for positive and negative. We selected words type that was closer to our intended purpose. The **Notes** column refers to issues regarding some of the words, as follows:

Note#1. *Complex* word holds a negative meaning found in several dictionaries, such as the online dictionary (Dictionary.com), but in SentiWordNet3.0 we found it had a neutral meaning with score (0). In this case, we can not rely just on SentiWordNet3.0 and we need to use other dictionaries.

Notes	Word	(-ve) score.	(+ve) score	Definition	Example
	Clear#15	0	0.5	easily deciphered, well described	"Readable handwriting."
	Clever#3	0	0.625	showing inventiveness and skill	"a clever gadget."
Note#1	Complex#1				
	Useless#1	0.625	0.125	having no beneficial use or incapable of functioning usefully	"a kitchen full of useless gadgets."
	Confusing#1	0.625	0	causing confusion or disorientation	"a confusing jumble of road signs."
	Reliable#1	0	0.375	worthy of reliance or trust	"a reliable source of information."
	Organized#1,#2	0	0	formed into a structured or coherent whole, methodical and efficient in arrangement or function	"how well organised she is."
	Average#3,#1	0	0	an intermediate scale value regarded as normal or usual	"The average income in New England is below that of the nation."
	Sensible#1	0	0.5	showing reason or sound judgment	"a sensible choice."
	Pointless#1,#2	0.625	0	not having a point especially a sharp point, serving no useful purpose.	"my pencils are all pointless."
Note#2	Moderate#3 a	0	0.5	marked by avoidance of extravagance or extremes.	"moderate in his demands."
	Moderate#1 n	0	0	the person who takes a position in the political centre.	
	Moderate#1 v	0	0	preside over.	"John moderated the discussion."
	Bad#1	0.875	0	which is below standard or expectations, below average in quality or performance.	"take the bad with the good."
	Indifferent#1	0.25	0.25	marked by a lack of interest	"the universe is neither hostile nor friendly."
	Spectacular#3	0.125	0.25	having a quality that thrusts itself into attention	"a spectacular rise in prices."
	Useful#1	0	0	being of use or service	"a useful member of society"
	Beautiful#2	0	0.75	delighting the senses or exciting intellectual or emotional admiration	"a beautiful painting"
	Vague#1	0.375	0.25	not clearly understood or expressed	"their descriptions of human behaviour become vague, dull, and unclear."
	Fair#2	0.125	0.125	not excessive or extreme	"a fairish income"
Note#3	Indistinctive				
Note#4	Fulfill#1v	0	0	put in effect.	"carry out a task"

Figure 4.18: Sentiment scores distilled form SentiWordNet3.0 for the 20 impression words.

Note#2. *Moderate* word comes in different positions or types adjective, noun or verb, the characteristic of moderate mostly refers to neutral but we found it has a positive score as an adjective and this does not agree with our assumption (selecting sentiment descriptive adjective words more than describe action words like verbs).

Note#3 and #4. these words are not found in the SentiWordNet3.0, we have made a search in different lexicons to find classifications and get more details about the characteristics of these words.

We placed all words in a table and recorded an average of all scores in the lexicon for each word. After discussion we pondered simplifying the list to seven positive, seven negative and six neutral words. These words are placed in a random order on the top of the sheet, which users can circle to record their first impression. Table 4.10 shows the classification of the words. In light of this, we include a list of examples that can help the appraiser with how to use these words to state the user's impression in visualisation through their critical analysis processes. Table 4.9 shows the list of examples.

Basically, to act effectively on creating an overall score of the user's critique we deliberated a lot about how one simply achieves this task. On the one hand, having a single value that can be awarded to a visualisation is useful. For example, if a user performs the CDS for a second time the value would allow them to quickly decide whether the visualisation has improved. On the other hand, a single value would need to be created through the calculation of an average. Therefore, a single value could potentially hide a bad design. We can imagine a situation whereby the visualisation is designed with some well-designed parts, while others are poor. A low CDS score would prompt us to address these issues. However, it could have the same score as a design that was critiqued as being unexceptional for all categories, which would probably need a complete re-design.

The situation is much like Anscombe's quartet [8] 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. At the same time we wanted to have a simple and quick way to calculate the CDS score. Therefore we decided to use a simple (weighted) sum and use a star-plot to show the breakdown of the categories. The 5-part Likert scale allows an appraiser to choose a very poor score (-2) to neutral (0) to an excellent score (+2), the scale measures have been inspired from the UEQ system from negative (-2) to positive (+2) scale measures. Thus, the appraisers are simply distinct between poor and excellent value options. We

Table 4.9: Various examples express the impression sentiments and word meanings. Users use these examples to help them reflect and make a critical analysis of their visualisation/design.

Type	Words	Example
Negative	Bad	“I think there are many bad things in this visualisation.”
	Complex	“Too much complexity in this visualisation.”
	Confusing	“I cannot understand some parts of this visualisation. I am confused.”
	Indistinctive	“This visualisation has not got any distinctive points, nothing stands out as being special.”
	Pointless	“I think this visualisation has no purpose, there is no benefit to the user”
	Vague	“This visualisation is vague, I can not interpret what is going on.”
	Useless	“My view is that this visualisation is useless.”
Positive	Clear	“This visualisation is clear I can understand it”
	Clever	“This visualisation shows a high standard level of innovation. [If someone else created it, you may say:] I wish I had created this visualisation.”
	Reliable	“I can trust data and results for this visualisation.”
	Sensible	“This visualisation is reasonable. Things seem to be in their correct place. I cannot fault the visualisation.”
	Spectacular	“The construction of this visualisation showed spectacular work and ingenuity.”
	Fulfilling	“This visualisation fulfils the goals. It is fulfilling.”
	Beautiful	“The whole appearance and look of this visualisation is beautiful.”
Neutral	Fair	“I am neutral towards this visualisation. It an agreeable visualisation, with nothing too bad or nothing too spectacular. ”
	Indifferent	“I am indifferent towards this visualisation neither good nor bad. It is OK.”
	Moderate	“These two visualisations have been rated with moderate scores. This is a reasonable visualisation, nothing outrageous nothing too unreasonable.”
	Organised	“This visualisation has got efficient parts, and the arrangement makes the visualisation suitable. ”
	Average	“My view on this visualisation is average. If I could order the visualisations, this fits somewhere in the centre.”
	Useful	“This visualisation is merely useful for its purpose. It is only suitable.”

can treat the appraisers checked values as a matrix c_{ij} of rows (categories) and columns (Likert values), where i represents rows and j columns, as follows:

Table 4.10: The Impression sentiments words classification based on positive, negative and neutral characteristic.

Negative	Neutral	Positive
Bad	Average	Beautiful
Complex	Fair	Clear
Confusing	Indifferent	Clever
Indistinctive	Moderate	Fulfilling
Pointless	Organised	Reliable
Vague	Useful	Sensible
Useless		Spectacular

$$Score = -2 \sum_{i=1}^{30} c_{i1} - \sum_{i=1}^{30} c_{i2} + \sum_{i=1}^{30} c_{i4} + 2 \sum_{i=1}^{30} c_{i5} \quad (4.1)$$

The calculation gives an idea that it was a positive or negative visualisation score value. In the sheet, a scoring bar also is added, to allow users to quickly place the overall score on a line. In contrast, there is an average value to the opposite of each total value this could make easy and quick average calculations to mitigate the calculation burden on users.

The star plot helps to show the breakdown of the categories. In visualisation, charts and graphs deliver information to users better and faster than other representations for data. Our point of view is to include a star plot graph to represent the questionnaire data which could give a quick assumption to users and decide how positive or negative visualisation average score. We deliberated a lot about the star plot design, and went through several iterations, especially after receiving feedback from the pilot studies. We originally had the star plot with a bigger axis for the six main categories, and then smaller lines for the additional questions for each category. But this was confusing, and after the second talk-aloud sessions and adjusting the number of questions with several CDS parts design, we changed the star-plot to be a numbered axis, where each number matches with the category questions. Figure 4.19 shows the development of the star-plot design.

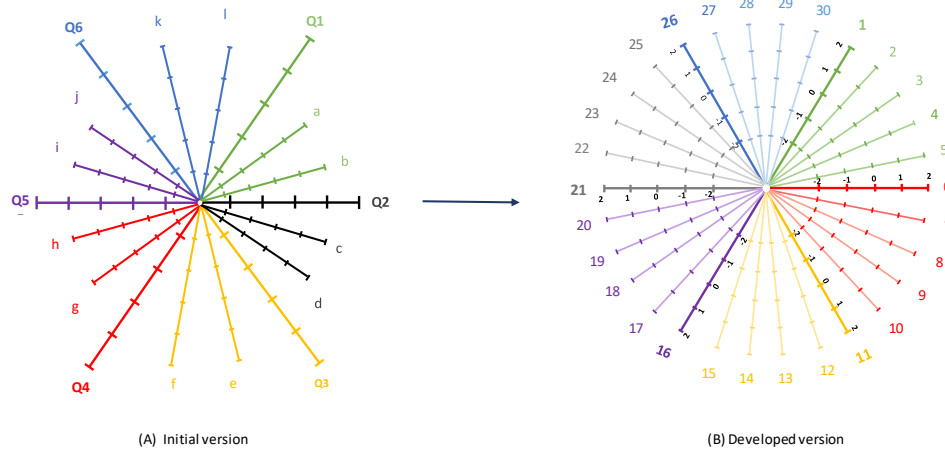


Figure 4.19: The changes on star plot graph between the initial design and final developed design. Appraisers use this graph to represent the questions data and that gives them an initial decision on their design and how its looks.

4.4 The final Critical Design Sheet prototype

The Critical Design Sheet (CDS) consists of seven sections, Figure 4.20 shows the mapping pattern for CDS to understand the over all structure and to see if the parts fitted well and see if sections were organised well enough. It is important to find whether the organisation of sections is fitted or mimic the flow of the user's thought when they want to use CDS in critique tasks. For example, we need to prepare users from the beginning to state their overall impressions or gut instinct feeling towards design and just detect what things are good or bad in that design. Then users will continue their reflections until they become able to make formal decisions on the design.

So, the full sheet design is in Figure 4.21 we decided upon that design after an iterative design process side by side with evaluations and discussions in each stage over the CDS design process. We suggest that the appraiser proceeds as follows:

1. The Information panel is completed; includes date of use, author and a description (or name) of the visualisation being evaluated.
2. Considers the design, and reaches an initial view of the design. Five keywords are circled in section A. The idea is to capture initial thoughts. Reflection over the choice of these words is made in section E. A comparison can be made between

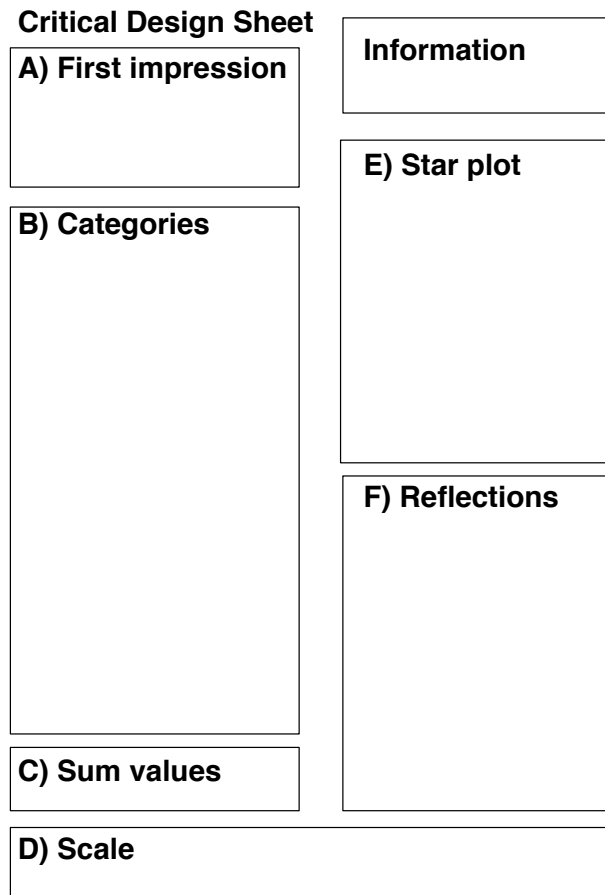


Figure 4.20: The layout of the seven sections of the Critical Design Sheet (CDS).

the initial thoughts (as indicated by the choice of words) and the scores (and sub-categories) as shown by the grade and star plot.

3. The appraiser works through each of the six categories in turn, deciding a Likert value for each category. Each of the six categories has a principle category (a killer question) where the user can take a holistic view on the design of the visualisation in a specific category (these are marked by the categories 1,6,11,16,21 and 26). The words on the sheet (user's perception, understandable, assumptions, trustful etc.) act as prompts to critique a particular topic. A full list of questions is included in Table 4.9.
4. The score is calculated by first summing up the quantity of checked sub-categories in each Likert column and then multiplying them by their weights (see Equation 4.1).
5. The score ranges between +60 (good) and -60 (poor). The scale also indicates an average 5-part Likert value.

6. The star plot is completed. The lower values are positioned in the centre of the plot, with the higher scores on the outside.
7. Finally the assessor reflects on the critical analysis that has been performed. First, they need to contemplate whether the first impression is different to the result. If it is different, why is this so? What aspects would make the first impression different to the objective scoring? The second aspect to consider is the overall grading, both shown on the scale and also in the star plot. Is it a good score? Is it generally good, or are some parts good while other parts not? Thirdly, they need to consider the whole critical analysis and make an action-plan that would improve the design.

4.5 Using the CDS

Now that we have reached or achieved a prototype for CDS it is appropriate before starting to examine the usability of CDS. The next chapter will explore its use and how users benefit from using it. Holistically, the CDS could be used in different applications, such as in the planning stages and when the users generate their initial ideas. Or the CDS could be used to evaluation and investigate the final application, and if it meets with the requirements. It is important to state the purpose of CDS before explaining or using it. Particularly, our aim in this research is to help users follow a structure when they want to evaluate visual designs. The CDS assists users to make insightful thoughts on their designs. The five sections in the CDS prototype helps to guide the user to make various reflective thoughts. Consequently users can generate a final (holistic) yet subjective decision through doing such small objective tasks within the sheet such as selecting initial impression words, giving values on rating scales, calculating total to get average value and finally reflect on overall evaluation process, with state of how to improve the ideas of design.

Holistically, Figure 4.22 illustrates the CDS usage in the visualisation environment. The CDS can act in visualisation design stage by evaluating the design ideas to think insightful on the best ideas that fit with design requirements. The CDS can help users

and using the CDS as a guide to evaluate that design. By using the CDS in this way, the developer can make more objective judgements on the design; deciding whether it would be suitable for the user to use, that it is the right design for the environment that the tool will be used in, etc. It can help the developer make decisions at the design stage of the tool, to either reject the idea or continue to refine and improve the idea, or take the idea into the implementation stage. At the implementation stage of visualisation, the CDS is also useful as an evaluation tool. At this stage, it can help users to judge whether their implementations need be improved. When the final implementation has been made, it could be used by a typical end-user as a way to evaluate if the tool is fit-for-purpose, which will help the developer decide how to improve the tool for next versions.

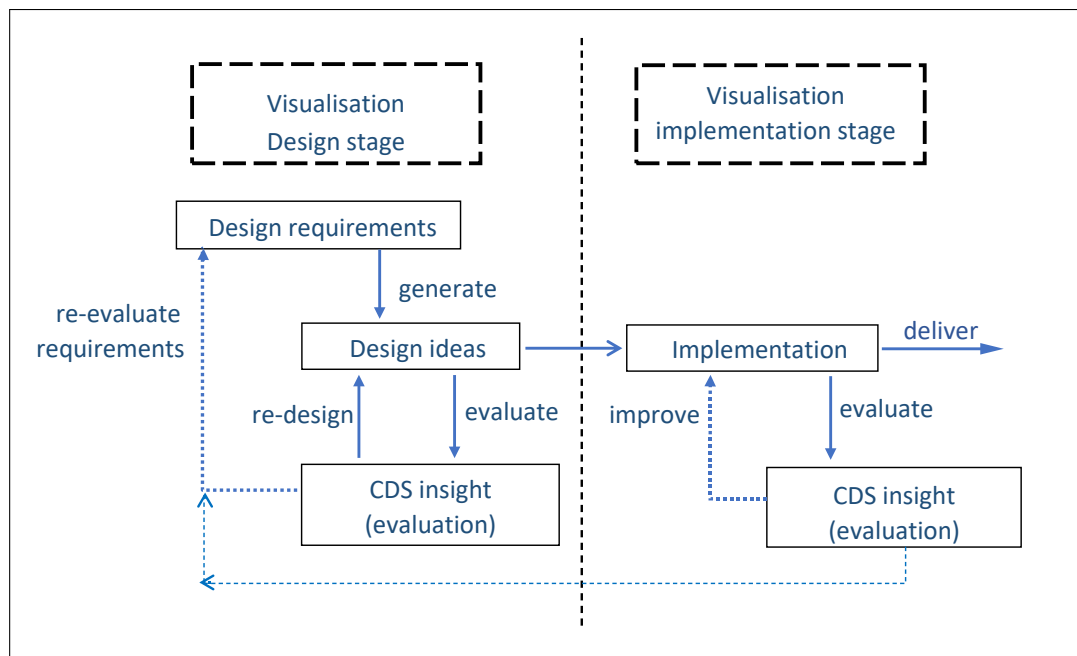


Figure 4.22: insight vision of using CDS in the visualisation design environment. The diagram explains two main stages in visualisation construction, the design stage and the implementation stage and the role of CDS in each stage.

4.6 Summary

This chapter focused on creating a Critical Design Sheet (CDS) method. We moved through various stages of the process to achieve that goal. In that way, we investigated many visualisation models to find an answer for the research question “is there a structure that users can follow to critique visual design and help to make self-reflection

on design?” We started the process with investigating and observation work to find how people think critically towards objects and designs. This workshop gave us data, and helped us to consider the situations where people critique, and what processes they use to achieve critical thinking, and gave us different terminology that then was put into the final Design Sheet. Subsequently, we distilled information from our investigation work, and ideas also from previous knowledge and experience, to create structures that serve the critical analysis domain which is represented in the CDS. The CDS structure enables people to make subjective decisions about small categories, they can then put these subjective decisions together to make a result that is more objective and takes into consideration many aspects of the design. Consequently, the CDS provides a way to structure the assessment of a visualisation and help appraisers think through different aspects of the visualisation design.

The final solution consisted of six component parts: (A) the first impression words, which are based on words from the workshop and early prototype studies, and based on lexicographic words scores. Twenty words are provided to allow users to circle just five. Seven positive and seven negative words with six neutral words. (B) The main part of the worksheet where users judge the visual depiction in a 5-part Likert scale, in six categories of the User’s impression, environment, interface layout, components, aesthetics and visual marks. (C) Users sum the values to get a single combined score that then they can mark on (D) the scale. Users then complete the star plot (E) which enables them to quickly see which category needs to be improved (or possibly the whole visual depiction would need to be re-designed). Finally the user reflects on their evaluation (F) by initially considering the main parts of the CDS and then how to improve the design, and what to do next.

This chapter has spent much time and consideration to find an objective structure for critical analysis, such that users can make judgements on design. The CDS fulfils the design brief, but it is not an easy task for a user to be critical, and so the next part is to decide whether it is an effective structure. The next chapter will investigate the effectiveness of the strategy, and how it could be used by end-users.

Chapter 5

The evaluation of CDS

This chapter presents the evaluation of the Critical Design Sheet (CDS) that was designed in chapter 4. The overarching aim is to evaluate if the CDS is usable. Thus, to achieve this goal we conducted usability evaluation sessions as an observational study on two different cohorts (Group1 and Group2) for gathering data, following in both cases the same process. For our analysis we have two groups: group1 – the PhD students group; and group2 the undergraduate students group as participants to use the CDS, data was be gathered, and then the data was analysed using different techniques. Analysis was performed to investigate the usability of CDS through testing (1) the reliability of the questions, by finding inter-items correlations of categories questions; (2) if users were generally consistent in their judgements (such that we can say “most people agree that a bad visualisation is bad, or a good visualisation is good”. To achieve this, the variance between users was evaluated using a T-Test analysis, for that analysis we have two groups: (group1 – the PhD group; and group2 the student group). The hypothesis is that there is no difference between the grades between these two groups over all six visualisations. (3) The impression words data were also evaluated, to discover how they are used and if people are consistent. Finally, (4) the learnability of the CDS was investigated, through looking at completion time’s of users when completing multiple CDS and compared with different groups of people.

The chapter has broadly three parts: first an introduction to the work sharing the vision of the evaluation, second, an explanation of the cohorts and general tasks given to the user, and third the evaluation of the data; as follows:

Part 1:

- Section 5.1. Vision for the evaluation.

Part 2:

- Section 5.2. Scenario and data-gathering. Explanation of the tasks that were given to the different cohorts, such as to capture the data.

Part 3:

- Section 5.3. Results of users using the CDS (both group1 and group2).
- Section 5.3.1. Observational analysis of completed CDS sheets. An observation of bad and good practice.
- Section 5.3.2. Testing reliability of how users complete the questions. The hypothesis is that most people would agree that a good design is good and answer the CDS accordingly.
- Section 5.3.3. Investigate consistency of judgements (looking at the variance of the data using T-Test).
- Section 5.3.4. Analysis of the impression words.
- Section 5.3.5. Exploration of learnability.
- Section 5.3.6. Qualitative evaluation – what do the participants say?

5.1 Vision of the evaluation

According to the ISO 9241 standard of human-computer interaction (that originally focused on visual display terminals), usability looks at evaluating the “effectiveness, efficiency and satisfaction with specific users” [58]. But what is evaluation really? And how does it apply to visualisation and to the CDS tool?

For visualisation tools and techniques, it is important when creating visualisation tools to discover if they are actually usable. What the developer wants is their tools to be

used, again and again by a user. If the user does not like the tool (or technique), or they do not find that it does what they want it to do, then they will not use it. It will be forgotten. Likewise, with the CDS, we want people to use it. It therefore needs to be something that will work and is effective, and is fit-for-purpose. We have already done some evaluation. In fact, the design of the CDS has gone through an iterative approach – which itself is a good strategy for generating effective visualisations. But now it has been completed, we need to discover if it is usable, and evaluate what real-users really think about it.

There are many approaches to evaluation and in fact different aspects of the CDS that can be evaluated. The visualisation literature (such as [17], [34], [75], [119], [145]), especially the reviews of Insenberg et al. [75] and Lam [92] provide a comprehensive look at different evaluation techniques. This literature is useful, and we look to the literature to find the best and suitable techniques for this evaluation chapter. Indeed the chapter focuses on the evaluation of the CDS that was designed in Chapter 4.

Because different aspects need to be evaluated, we need to use different evaluation techniques. Thus, as researchers we need to choose what to evaluate, and what techniques are suitable for the task [145]. The goal of creating CDS is to help learners and it would be useful for experts too to make self reflections on visual designs that aid the evaluation of the design. Therefore, “what do we want to evaluate and what knowledge would we gain from that evaluation?” and “what evaluation techniques do we use?”

To address the first question, it is valuable to summarise some of the main reasons for the evaluation. Jeng [78] provides a list of useful grounds for evaluation that were used in the analysis of library human-computer-interface tools, and Insenberg et al. [75] and Lam [92] summarise what can be evaluated for visualisation, and also in addition Chapter 2 on related work and Chapter 3 as this dissertation covers many aspects where metrics can be applied, and it follows that it is possible to consider that an evaluation (by a user) can take place at any and all of these places, so we turn the reader to these resources, but it is useful (at this point) to summarise some reasons. After looking at the literature, we realised that our overarching aim is to put CDS in a practical domain to measure ease of use by doing a usability evaluation. Under that aim we will focus on

evaluating individual sections to cover our overarching goal and use different evaluation techniques. Usability covers a wide range of aspects in evaluation and we need to examine the consistency of CDS and ensure the validity of data collection and analysis as well. So, we decided that we should evaluate two important aspects:

- Reliability, and
- Learnability

In fact, many researchers who have developed any type of Human Computer Interface often focus on investigating *usability*. Where researchers are trying to evaluate if the tools are usable. But usability is only one type of evaluation which learnability is part of usability that is performed. Others include: to explore if the tool is *functionally correct, pleasing, easy to learn, fit-for-purpose*, etc. In fact, just focusing on usability, Nielson [108] talks about judging “perceived usability”. In other words, the idea and goal of the evaluation is from the user’s perspective; that the user needs to make a judgement whether the tool is usable in their view. For Nielson, his focus was to evaluate if people could learn the tool (test learnability), and explore the ease with which the user can remember all the necessary parts. Because evaluation is done by humans, to decide whether the tool fulfils the goals, we need to think how a user would react to the displayed tool. In fact, the word “perceived” that Neilson uses is the key to understanding an evaluation. Consequently, the evaluation that takes place is in fact a judgement that someone makes in respect to their own knowledge, views and beliefs. Consequently, we need to ask actual users their views on the CDS.

Usability has long history and is used in various evaluation purposes in academic research [102], [137]. Software usability is well known field of usability applications this is in regards to many stages that the software pipeline passes through. So, it is significant to the software practitioners to be aware and able to determine which method is best suited to every situation in a software project [71], [108]. For the CDS, it is clear that we need to evaluate *usability*, and also we need to look and focus on learnability. It is important that the students (the end-users) learn how to use the CDS system. In fact, every tool requires time for users to learn. The argument is that better tools are quicker and easier to learn. There are many parts that a user needs to learn in the CDS.

While users do not need to remember the categories (they are listed explicitly on the CDS paper) they do need to realise what each of the categories means. Consequently, a user would need to read the crib-sheet (the list of examples in Table 4.9), that describes the meanings of each category and the examples given on the sheet, before they can easily interpret the stages.

To achieve the required evaluation, we also plan to use standard and well-known methods. There are a number of methods to evaluate usability. The techniques that we could choose from, include: formal usability testing, usability inspection, card sort, category membership expectation, focus groups, questionnaires, think aloud, analysis of site usage logs, cognitive walkthrough, heuristic evaluation, claims analysis, paper prototyping, and field study.

In our research, we used a breath of usability tests with CDS to inspect and decide if the CDS sections were internally consistent. The overarching usability measures of CDS is “how ease of use” and “how ease to learn”, we need to investigate from these measures how appraisers use the CDS in critiquing tasks and then find the variance between users critiques. From these techniques heuristics and the questionnaire would provide positive features such as being inexpensive, no need for advanced planning and simple evaluation methods [103]. Usability considers informal assessment methods mostly reviewed by experts and offered advice and not based on quantitative assessments approaches [171].

We can say our evaluation fits within the informal classification, which is classified as heuristic methods, where experts judge the compliance of an interface against predefined heuristics. The CDS method is a type of inspection method which is measured against different important aspects in visual design. So, it is important to get the experts review user’s visualisation work and provide a valuable insight into usability problems [155], this helps to mitigate difficulties when evaluated by end-users. Essentially, usability has different criteria to evaluate systems. However, it is not a panacea, and may not provide a definitive result, or something that can be acted upon [65], but it is still an important approach which empirically can test the reliability and validity of collected data after conducting several experiments and doing an analysis for the outcomes.

In this research we will focus on the use of a “questionnaire” to collect data. Collecting data for a usability study via questionnaires and testing the efficiency, effectiveness and satisfaction is widely used in many purposes and various academic domains such as in psychology and health-care which are aimed to improve health care research [11], [76], [82], [124], or used for assessment of academic digital libraries [78], [84], [152]. Another field, it is used in is enhancing the interface’s visual appearance for example the impact of colours on website design in distinct cultures [41], investigate the user’s interface satisfaction on a different interface style for learning graphical software applications [54], or used in a comparative study between conventional online banking and a tag-based interface to improve user satisfaction of online and mobile banking [126]. Test usability through questionnaires in evaluating opinion visualisation systems helps in development of those systems [142].

Our aim, when we designed the CDS was to create a set of heuristic questions classified into categories, that will cover all aspects of design. Users will use these items to measure visualisation attributes, the design environment and several conceptual aspects. Consequently we can investigate different aspects of the questionnaire and investigate how users have completed the CDS using statistical analysis software (such as SPSS). Statistical analysis, such as to find internal consistencies using Cronbach’s alpha statistic metric and running a t-test analysis on data samples can be used to determine the degree of variance between users. These measures can then be used to determine whether the system is usable, and beneficial, and supports the case that the CDS is usable. So, in the next sections we will explain the methodologies used, and how the experiments were run (and data captured), to achieve our evaluation goals and draw clear conclusions of what we learnt.

5.2 Usability test preparation & Data gathering

We conducted two separate usability evaluation sessions as observational studies on two different cohorts (Group1 and Group2) for gathering data, following in both cases the same process. Group1 included 10 PhD students within the Computer Science department, where half had experience in creating visualisations with JavaScript

Questionnaire analysis sheet:

In this sheet we introduce a method (CDS) to critique visualizations. Users can use the method to judge their design and find ways to improve it. The method consists of 3 main parts; the questionnaire, star plot and reflections. The questionnaire is set of grouped questions the star plot shows the selections to investigate design improvement for; the reflections part provides an evaluation in cognitive way. These questions should be answered after you finish the test.

Users should follow the steps of the method in order from (A) to (F). A simple calculation is needed to total the selections value. Sum up all columns separately, multiply by the weights (-2,-1,1,2) and then sum up the results horizontally to get the total in step (c). Use the scale in step (D) to find the average.

The following table shows the meaning of keywords used in the questionnaire:

Index in CDS	keywords	The intended meaning
2	Understandable	Is it clear? Can you understand what is going on?
3	Assumptions	Does the visualization make excessive assumptions? Does it rely on domain knowledge. In the context of the user/task is this suitable? If it to be displayed in the public domain then it should be clear and make no assumptions!
4	Trustful	Would I trust the presentation of data in this visualization. Does it give me confidence and advice friends to use it? Is it an exemplar visualization?
5	Useful/utility	Is there clear utility? Has the visualization got a purpose? Is that purpose clear?
7	Technology	Is the right display technology used? It may be paper printout in a magazine or it may be a desktop, is this suitable for the visualization type?
8	Interaction	Is there the right interaction functionality? Can you do the things you want to do? E.g., order, organize, scale.
9	Output size	Is the size of the output right - too small and it may not show all the data. Too big and the user may not be able to see all the data!
10	Place	Does the environment (where the visualization is displayed) give the right vibe(sense)? Is it a positive experience? Is it suitable for the task and the user? Is the context where the visualization will be used suitable for the visualization!
12	Layout	Are the layout of the frames clear? Is it suitable interface? If tabbed interface is this good? If cascaded or tiled interfaces are these suitable for the visualization display?
13	Frame size	Are the sizes of the containing windows/frames suitable for the interface and the visualization task?
14	Spacing	Is there wasted space within the layout of the frames? Is spacing used efficiently? E.g., if lots of wasted images, this may be chartjunk, and may not be suitable.
15	Quantity	Is there the right amount of windows? Too many (may confuse), too few may not be showing all the data.
17	Type	Type affords an interpretation, Can we read off maximum and minimum values, or compare values (as required). Is this type (e.g., bar chart, line graph) correct for the purpose of this presentation?
18	Relationship	For (e.g., multiple views). Are relationships between different visualization components clear? Do I understand how one view is related to another? Do I understand how the view is related to a legend?
19	Task suitability	Is the choice of the view suitable for the task? E.g., If it is a bar chart - is this suitable or should I use a scatter plot? If I have a pie chart, is this suitable for my purpose?
20	Structure	Is the structure of the visualization right? Is it suitable? Does it demonstrate the data? Is it clear how the individual parts of the visualization fit together
22	Aesthetic	Is it beautiful? Is it well designed? Does it follow good design principles?
23	Space utilization	Is there wasted space (maybe the design is all to the left hand side - is this good?)
24	Coverage	Does the design show all the required data? Does it represent the right quantity of data? Is aggregation correct?
25	Legend/labels/	Are there legends, labels etc. Is this context-giving information correct?
27	Mapping	Does the mapping (transfer function) match with the data? If the data is linear, is a linear Transfer Function used?
28	Mark types	Are the marks appropriate? Length, size, color choice correct?
29	Scale/zoom	Is the data shown at the right zoom? If it was zoomed in, would this be better? Would it make it clearer with a non-linear zoom?
30	Overplotting	Would it be clearer with a linear zoom? Would it be clearer bigger/smaller scale?

Figure 5.1: The descriptive sheet that given for students as supportive analysis document helps them in critique design task.

software, and all had created graphs in Excel. Group2 consisted of 20 undergraduate students who were studying for a computer science major and were taking a graphics and visualisation unit. All the participants were volunteers and none received compensation for participating.

5.2.1 Experiment 1 procedure & Materials (group1)

We invited PhD candidates within the School of Computer Science to participate in our experiment. 10 participants were recruited to conduct the first usability session. The number is a convenient sample. We wanted to have a long session with these participants, and so too many people would have been difficult to manage and monitor, while too few would make the evaluation less suitable and convincing. We needed to monitor when they had completed each part, and time the sessions. Therefore if we had too many participants it would be much more difficult to time them. An email invitation was sent to the PhD students in Computer Science department to recruit them; they committed to coming by replying to the email and saying ‘Yes’, and then turning up at the allocated date/time.

Materials: We prepared a pack of paper copies for each participant. They received the pack when they arrived. The pack consisted of:

- Consent form.
- Six printed CDS sheets (see Figure 5.8).
- A support set of questions as descriptive sheet (the figure 5.1).
- Six visualisations to critique.

The quantity of six visualisations was chosen, again as a convenient balance. We needed to give the participants a set of different types of visualisations. Too many visualisations would take the participants too long, and they would get tired. Too few and they would not be exposed to enough, and we would not be able to look at learnability.

We started the session by giving an introductory talk. This lasted about 15 minutes. It was given to introduce to the participants the CDS, and to explain the CDS goals and the purpose of doing this task. We also gave explanations about each CDS section, and how to use the CDS to critique a visualisation. While we believe the CDS is understandable with little instruction, most of the time it will be used in an education setting, and some

instruction would be normally given to the students. Therefore, we were telling the participants information that they would be given in a normal situation.

The participants were given the task to complete the CDS sheet on the six given visualisations. The tasks that the participants were asked to perform are as follows:

1. Select 5 impression words out of 20 words. (CDS-section 1)
2. Complete tick 30 questions of Likert scale (-2 to 2) selection for six principles categories. (CDS-section 2)
3. Calculate the average of your selections. (CDS-section 3)
4. Present all questionnaire values on a starplot graph. (CDS-section 4)
5. State your reflections and decisions as comments about your critique between the first impression and the final decision. Write any improvement suggestions if possible. (CDS-section 5)

It is useful to explain more about the pack of papers, the support set of questions is a detailed list of 30 questions with descriptions and examples (see figure 5.1). Which helps appraisers to understand the visualisation terms that are used in CDS six principles categories. The selected visualisations were:

1. a bubble chart created from Gap Minde figure 5.2,
2. a pie chart created in Excel figure 5.3,
3. a treemap figure 5.4,
4. a transport network map figure 5.5,
5. a time series graph figure 5.6, and
6. a bar chart figure 5.7.

These were carefully chosen to include both familiar and (potentially) unfamiliar visualisations. Each had a written background-scenario. The written scenario on each visualisation explains to the participant the *environment* that visualisation is displayed and what *technology* that has been used, some annotation is also added to the visualisation graphics to explain the controlled buttons (*the interface*) and mimicked the real design, so users can virtually interact with the design. For instance, the pie chart visualisation showed annual research-grant income at a University, and would be included in the University's magazine (to express the environment where the visualisation will be used). The scenarios were deliberately kept brief and simple.

Each participant was observed and notes were taken. Each visualisation was numbered. Participants were asked to critique the visualisations in order and write the time they started each CDS at the top of the page. This timing information would allow us to calculate how long it took for them to complete each CDS. Notes were taken in the group, such as to make sure that the timings concurred with those written by the participants. Recording visualisation critique time is important for us to check the performance of the user's critique with the time to test learnability.

We gave participants at least 1 hour to completely critique six visualisations. Unfortunately, some participants did not complete all six visualisations within the hour. Just 2 participants completed the critique of all six visualisations within an hour and eight participants completed just 3 visualisations the rest completed them by their selves out of session time, we asked them too to time themselves when critiquing.

5.2.2 Experiment 2 procedure & Materials (group2)

This second usability session (group2), was conducted on 20 undergraduate students who were studying for a computer science and were taking a graphics and visualisation unit. We needed to have all experiments be consistent, so the same procedure and materials were provided to the students. However we did not provide the consent form, because this exercise was used as part of an assessment. These tasks were given to the student to practice using the CDS; later the students would use the CDS to evaluate their own designs and tools that they were building. This situation not only gave us the

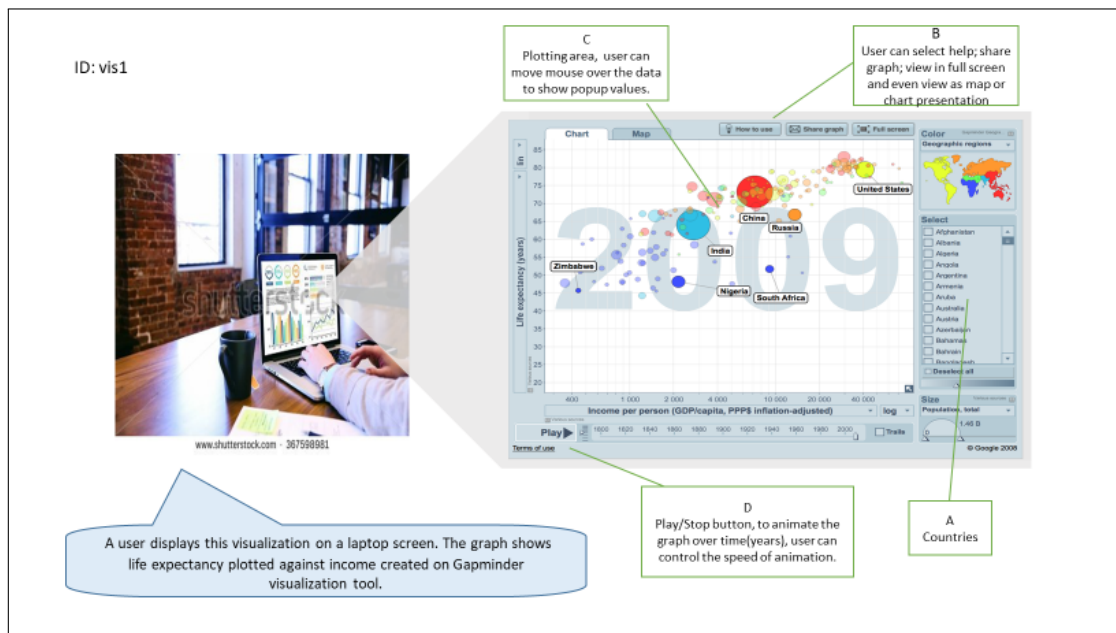


Figure 5.2: Scenario 1 as experiment material: bubble chart visualisation created in Gapminder tool. (Image from the Gapminder tool, <https://www.gapminder.org/> issued under the Creative Commons Attribution 4.0 International license).

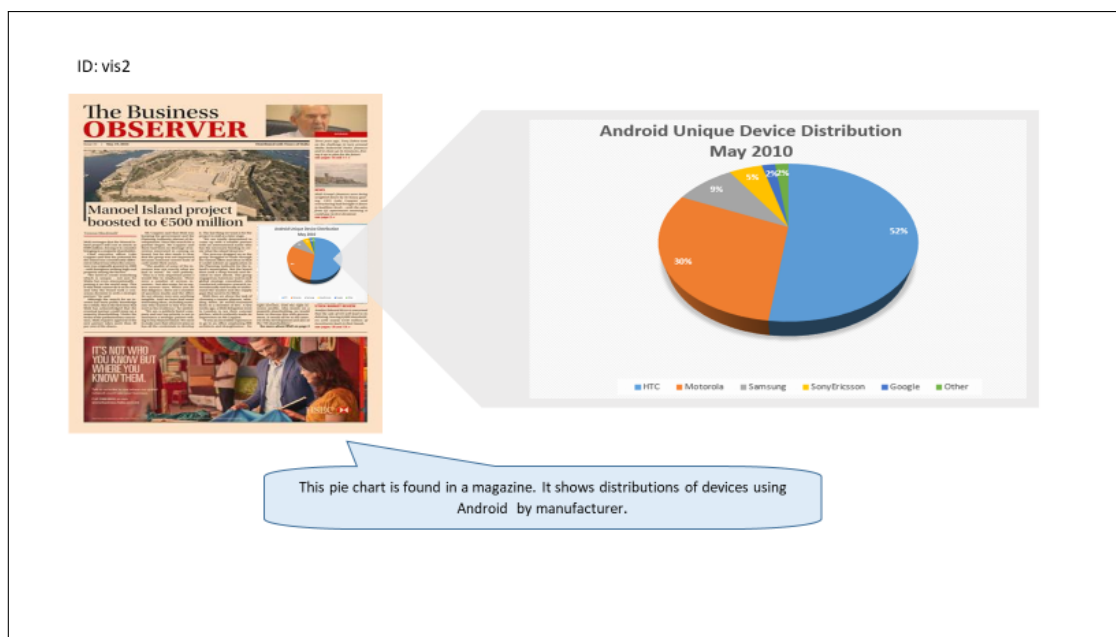


Figure 5.3: Scenario 2 as experiment material: a pie chart created in Excel by the author displayed on magazine.

ability to run the same experiment (this time on 20 undergraduate students), but also to get them to use the CDS for an actual application.

We asked students to time themselves when they did the critical evaluation for each visualisation and record that on the sheet. Most students completed critiquing the six visualisations within an hour, only one student did not complete the full six in the hour.

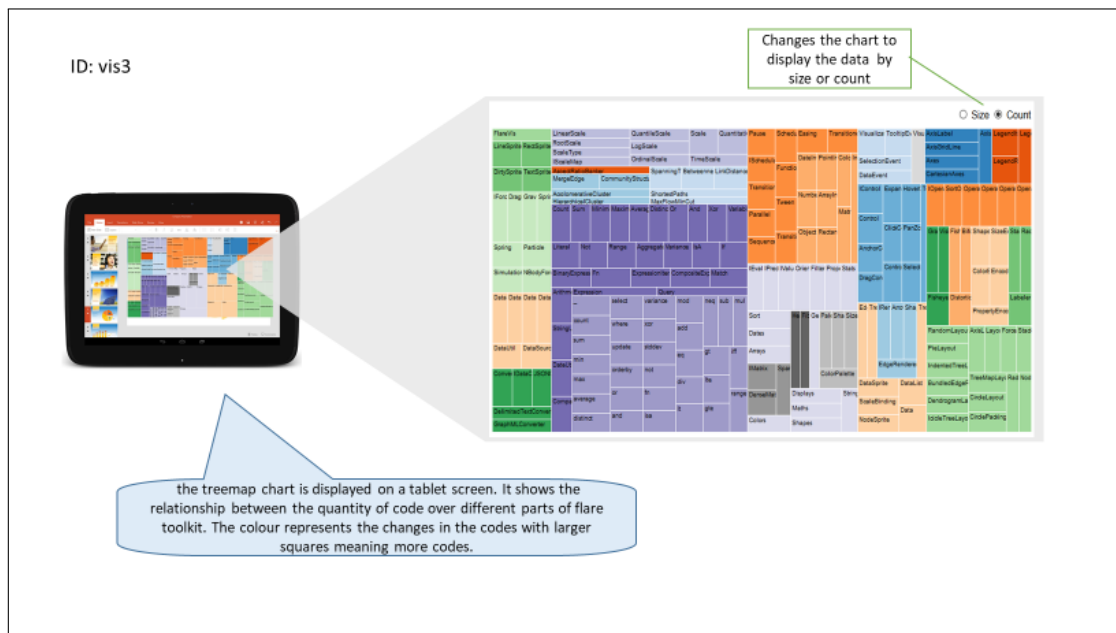


Figure 5.4: Scenario 3 as experiment material: a treemap D3.js graph displayed on tablet. Permission to use the image has been granted by the creator Anthony Ina (Viking Lab).

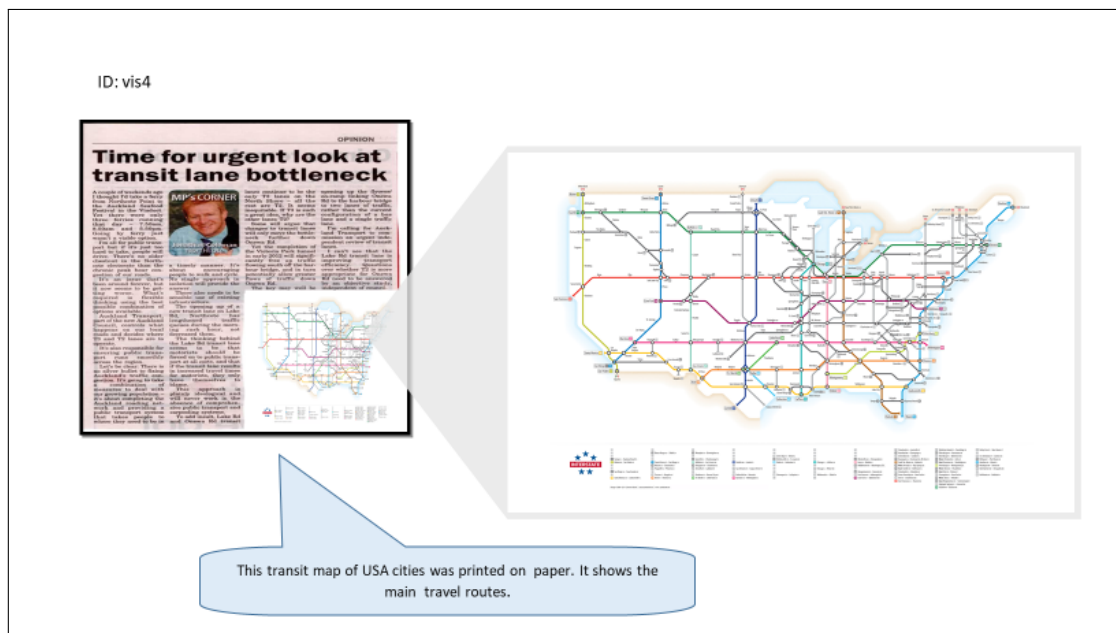


Figure 5.5: Scenario 4 as experiment material: a network graph displayed on newspaper. Permission to use image has been granted by the creator Cameron Booth.

The students of group2 used the CDS two further times: once on the fifth sheet design from their Five Design Sheet (FDS) [132] planning of their visualisation tool, FdS is visualisation module that given for year 3 computer science students to learn how planning and sketching to build a visualisation tool or application. Again, students used CDS when they had built their visualisation tools.

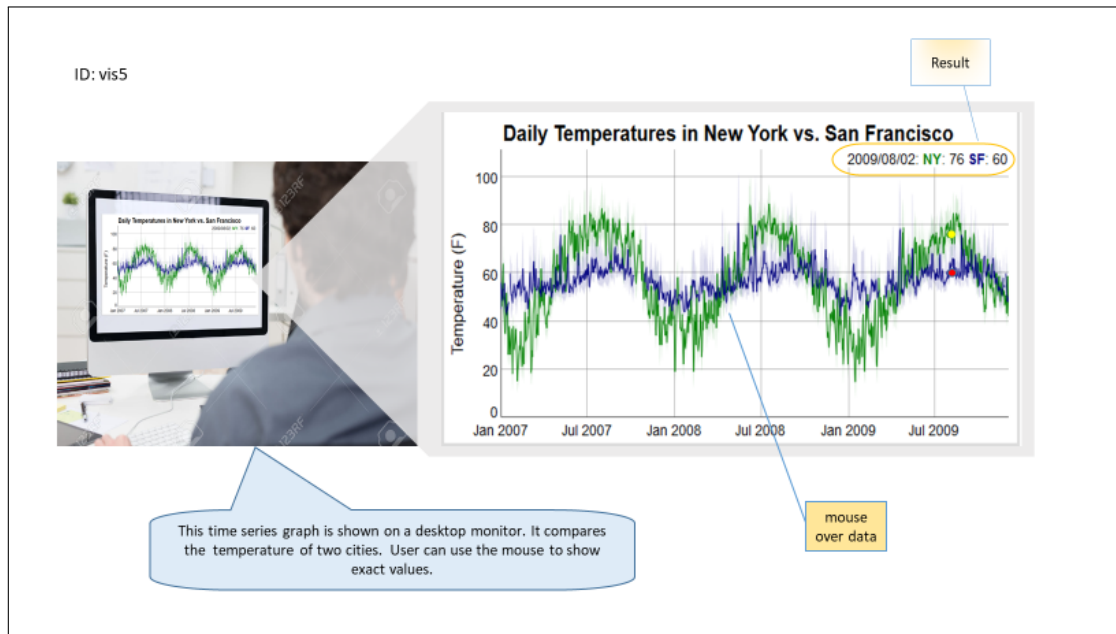


Figure 5.6: Scenario 5 as experiment material: a time series graph displayed on computer screen. Image reproduced from the Dygraph “Range Selector” demonstrator, a D3.js library, licensed under MIT free to use, Copyright ©2009 Dan Vanderkam.



Figure 5.7: Scenario 6 as experiment material: a bar chart displayed on phone. Permission to use the image has been granted by Tom Gonzalez (Managing Director) of BrightPoint Consulting Inc. to use their Vizuly Column Chart image.

5.3 Data analysis & Results

In this section the analysis of data is divided into two different types of data analysis: First, we will analyse the user’s reflections on how they have critiqued their work, as qualitative analysis. We will investigate how users fill the data in each of the CDS

sections. This analysis focuses on sections B, E & F of CDS. For reference in this section, we include the labels and the final CDS sheet (see Figure 5.8).

Critical Design Sheet

Name
Visualization
Date

A) First impression (circle 5 words)

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

B) Categories

	-2	-1	0	1	2
1 User's perception					
2 Understandable					
3 Assumptions					
4 Trustful					
5 Useful/utility					
6 Environment					
7 Technology					
8 Interaction					
9 Output size					
10 Place					
11 Interface layout					
12 Layout					
13 Frame size					
14 Spacing					
15 Quantity					
16 Components					
17 Type					
18 Relationship					
19 Task suitability					
20 Structure					
21 Design					
22 Aesthetic					
23 Space utilisation					
24 Coverage					
25 Legend/labels					
26 Visual marks					
27 Mapping					
28 Mark types					
29 Scale/zoom					
30 Overplotting					

C) Sum values

	-2	-1	0	1	2
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
Total					

D) Scale

Total: -60 -55 -50 -45 -40 -35 -30 -25 -20 -15 -10 -5 0 5 10 15 20 25 30 35 40 45 50 55 60

Average: -2 -1 0 1 2

Poor design Good design

E) Star plot

F) Reflections

Consider: First impression (A), grades (C,D,E) and individual categories (1,6,11,16,21,26)

Improvements:

Critical Design Sheet

A) First impression

B) Categories

C) Sum values

D) Scale

Information

E) Star plot

F) Reflections

Figure 5.8: The final Critical Design Sheet (CDS). A) Choose five “first impression” words. B) score the categories. C) Sum the values from the categories count the quantities of -2, -1 and 1 and 2s. D) Place the value of the sum on the scale. Add information about the user and visualisation to the sheet. E) Complete the star plot to show the breakdown of the parts. F) Reflect on the results, and what to improve.

Second, statistical data analysis will be performed using the analysis software SPSS 21.0. This analysis focuses on sections A & C of the CDS sheet (the first impression words and the Sum values).

To perform the analysis the results from all participants were aggregated together (N=30) and treated as one cohort. This was done to investigate inter-item consistency (Reliability). Our aim is to see how consistent the use of the sheet across undergraduate and postgraduate levels was. We need to investigate the result for the usability attribute “how easy is it to use” by checking the consistency between user groups of their total scores and decisions. We will collect and analyse quantitative data represented in impression words and questions values to find the variance. We want to investigate the difference between group 1 (the 10 PhD students) and group 2 (the 20 undergraduate students).

5.3.1 Observational analysis of CDS completion (good and bad practice)

It is important to analyse users' work on all CDS sections, and to evaluate how users complete the CDS in a practical way, and how they used CDS in to evaluate visual designs. We analysed and observed participants' behaviour of how they were completing the CDS data for all six visualisations. Using Lam's classification, evaluating how a user understands and performs the task is classified as "Understanding environment and work practice" (UWP), its goal is to evaluate often design implications based on a more holistic understanding of current work-flows and work practices, the conditions of the working environment itself, and potentially current tools in use.

The CDS method fits within the UWP goal, and works towards helping users to understand the design environment and characteristics. Isenberg [73] conducted a similar study as an observational study to present a framework for information analysis in collaborative and individual activities around information visualisations in a non-digital setting.

Therefore, to analyse our collected data we will take samples of users works to show empirically from field observation study good work practice and bad work practice. We listed several notes from our analysis to show samples of bad work practice when users used CDS to make a critical design evaluation. We noticed that, when most users completed the CDS that they completed the sections by shading in the boxes, others marked selections with tick marks and others marked them with cross lines. All these different behaviours to select values. In respect, it does not really matter how they complete the boxes. Here is list of bad work practices (we name them Bad Practice, or BP) that was extracted by analysing the completed sheets:

- **BP #1:** some users were overly critical and not very discerning. Perhaps they were very negative, giving overly negative ratings for most questions. This behaviour reveals that users could not detect what is a really bad category in the design from other ones, see Figure 5.9 for examples. In contrast, some users have selected just positive selections without looking at design insight to detect which category

needs more thinking or redesign Figure 5.10 shows examples on users work practice. This can suggest a few behaviours. First, perhaps the participants were lazy; that they only ticked one box. Second, it could be that the participants could not distinguish between good or bad designs, or perhaps they generally thought that they were terrible or fantastic.

- **BP #2:** some users did not fill the killer questions [1,6,11,16,21,26] on the questions list. This definitely affected the overall score average, and made a gap in the understanding of questions even on the star plot, it showed a discrete graph line connection. Figure 5.11 illustrates examples about that work practice. This could be improved by better training and getting the participants to do this another tie.
- **BP #3:** some users have ignored completing all CDS sections. The CDS sections are designed to help users follow a process and to help designers detect design faults, problems, successes and good practice. This behaviour unfortunately does not give a clear self-evaluation for visualisations, for example missing section A (impression words) means that the participant misses out on expressing their “gut instinct response” towards the design. Section F (reflections), this section is important to be fully completed, as it helps to address the design problems that can occur in the visualisations, and helps users to think how to develop their ideas further. Figure 5.12 shows users’ examples.

In general, most of the users who completed the sheets do demonstrate good practice. And they have low errors and few mistakes. While errors can occur, and we have seen some participants complete the star-chart wrongly (mis-interpreting the low scores and the high scores). But participants easily decided how to overcome this situation by crossing out their mistakes or doing the correct answer in a bolder/thicker line. We summarise good practices (GP) from the completed sheets, below:

- **GP #1:** While the participant made mistakes in calculations of visualisation scores, the users corrected their mistakes and it does not affect the evaluation and critical analysis of their work. Some users selected values and then changed their mind after they looked carefully for design. Again, this is fine. It is a

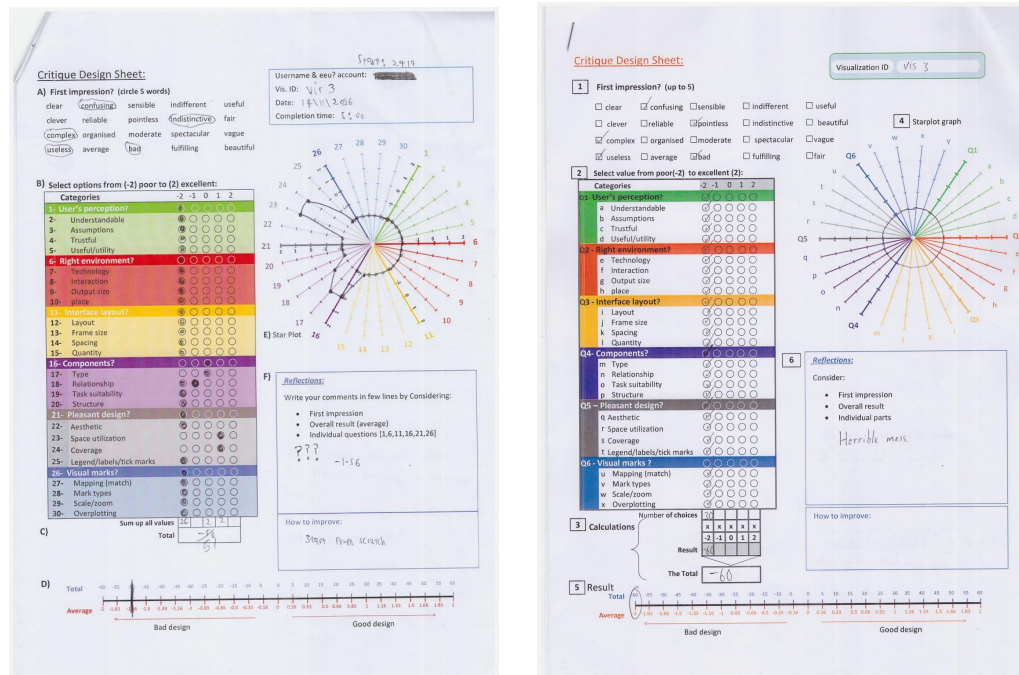


Figure 5.9: BP #1: Two samples of bad work practice. Participants selected just negative choices of over 30 questions of section (B). The star plot graph represents users selected values clearly.

demonstration of good practice and shows good use of critical thinking skills to analyse the design. Figure 5.13 on right side shows sample explain these good work practice.

- **GP #2:** Generally, most users time themselves to complete the tasks and they have done all sections as required. This is considered good practice; where users have quickly completed the forms and used them for the correct purpose. For example, this is shown in Figure 5.13 on the left side.

5.3.2 Test Reliability

Statistically, testing the reliability of the questionnaire means finding the inter-correlations or inter-item consistency for questions cross the users. For the CDS we need to investigate the inter-item consistency for six categories items (30 questions) for each visualisation cross all users in group1 and group2 (N=30).

Our hypothesis is that participants would, in general, complete the forms in a similar way. In other words, that most people would think (and agree) that a visualisation was

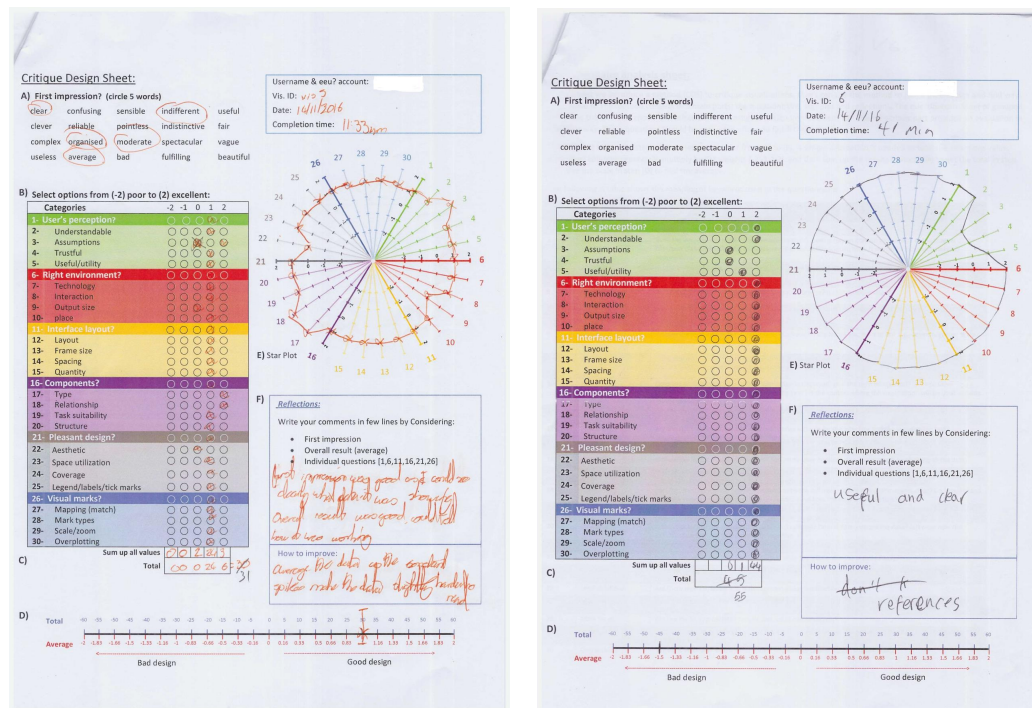


Figure 5.10: BP #1: Two samples of bad work practice. Participants selected just positive choices over 30 questions of section (B) star plot graph representing users' selected values clearly.

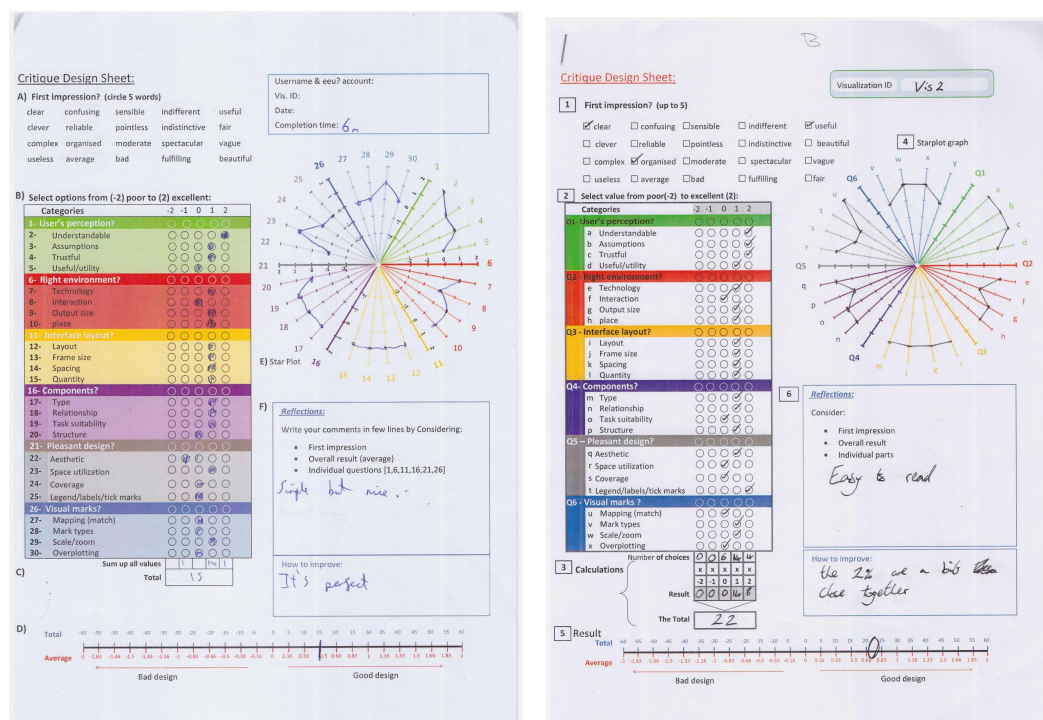


Figure 5.11: BP #2: Examples show bad work practice. Participants didn't select the main group question [1,6,11,16,21,26] in section (B) star plot graph displays a discrete drawing line connection.

good, and another visualisation was bad. If participants do not agree then the correlation is low and $\alpha = 0$, whereas the more items that share co-variance would demonstrate that participants share the same judgement and so agree that the underpinning concept

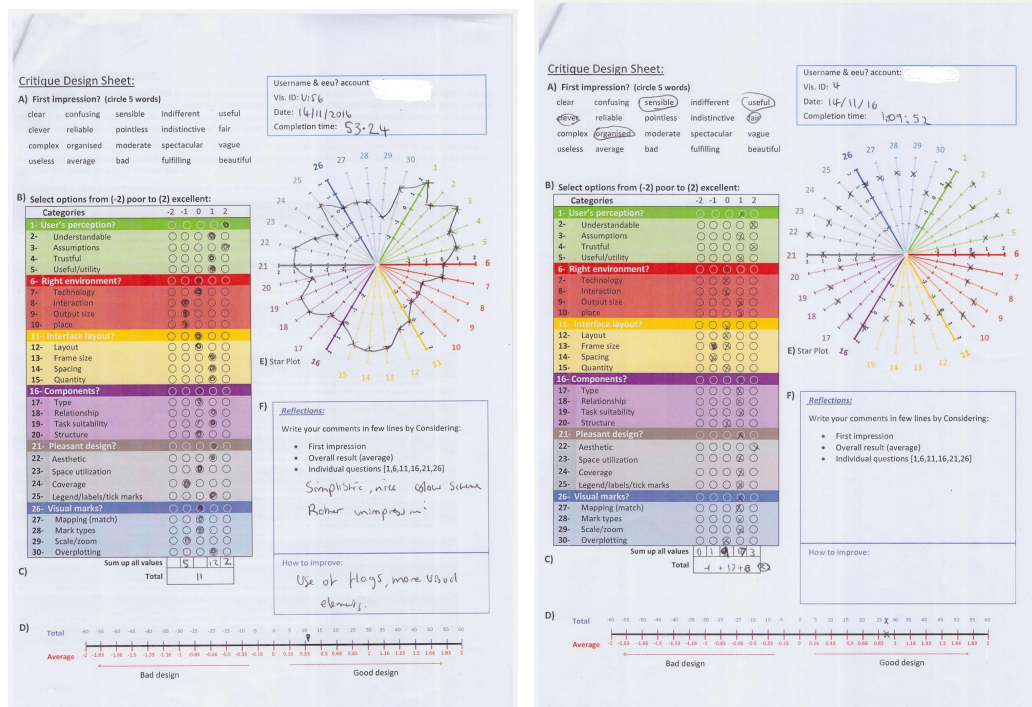


Figure 5.12: BP #3: Such users did not complete all CDS sections which reflect bad work practice. Two examples show missing sections: Left section (A) missing. Right section (F) missing.

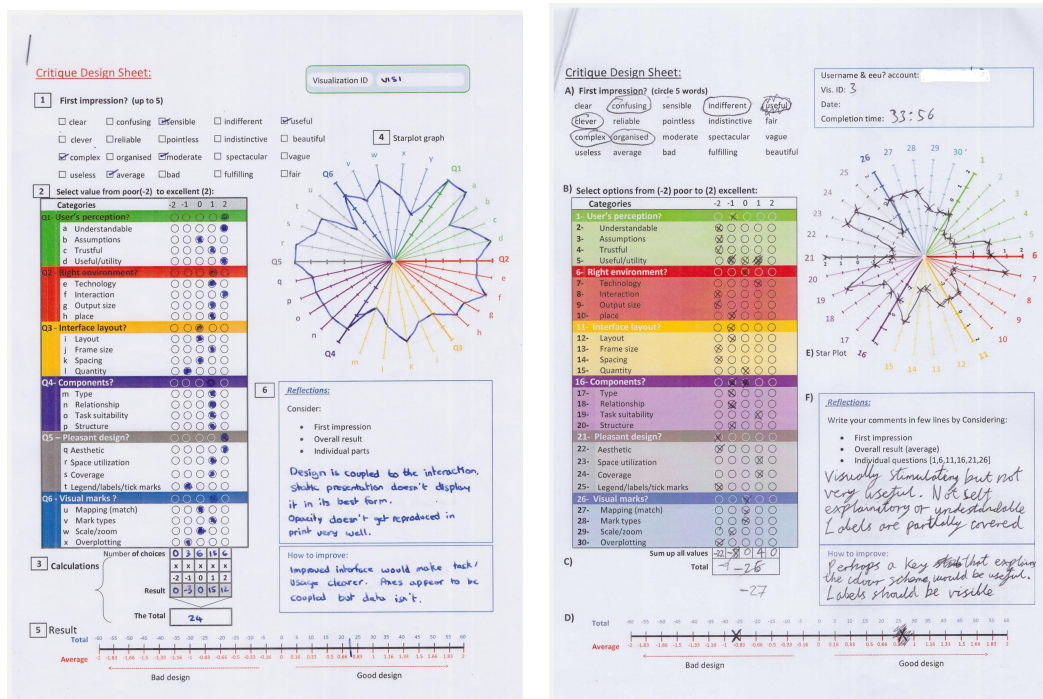


Figure 5.13: Samples of users' good work practice. The left figure illustrates the complete critique practice user filled all the required sections with self-reflections on how improve that design. Right figure shows normal mistakes can be done by users through a critiquing task that actually considers good work practice. The score calculation mistake in section (c), change the selected choice in section (A) & (B) and user's hanging minds, all these correspond good practice which reflects their better understanding towards design.

is similar, as demonstrated with a higher α coefficient. Furthermore, the notion of what makes a “good” α coefficient is dependent upon the type of data that is being evaluated. In most cases, the recommendation is that α value between 0.7 and 0.8 to be “acceptable”, 0.8 to 0.9 to be “good” and > 0.9 excellent.

To achieve the analysis, we collected data for total scores for each visualisation for each user, for both groups. The data was organised into a two-dimensional table in SPSS sheets to analyse them using Cronbach’s alpha statistic. We do note that the score that we are using is a weighted average of the Likert scale values 1 and 2, and a negative weighted score of the lower Likert values -1 and -2. However, this sum represents an average score of subjective values of the overall CDS. Values that are in the centre (0 position of the Likert scale) receive no weighting and are thus not considered within the calculation. For reference, the weighted equation is written in Equation 4.1 on page 105.

Cronbach’s alpha values are shown in Table 5.1, which supports the situation that users (in general) participants did broadly answer in a similar way. Putting aside potential issues with Cronbach’s α [87], [151], where the calculation can be sensitive to the number of scales or α value less than 0.3 which recommends excluding the scale from the calculation, it is clear that this is a positive result.

The results reveal that Cronbach’s alpha values for assessing inter-item correlations of 30 users belongs to 6 visualisations is greater than the value of 0.7 (a score of Cronbach alpha that defines the correlation to be questionable). Therefore, the interpretation of the high α values supports the hypothesis that there is consistency between users’ judgements on the given visualisations. The scores are highly correlated and we can notice the variance between items is very small values. That means that the questionnaire items could be used to measure the same aspects for most graphical designs and help the user to analyse the visual design in the same criteria. And that different visualisations could be graded by different people and then compared. The internal consistency of α value reflects that users received similar aspects about understanding CDS categories questions and they evaluated the design with similar judgements and decision making.

Table 5.1: The internal-correlation of 30 users (N=30) users completing the categories on six different visualisations.

Visualisations	Cronbach's alpha	Variance
Bubble chart	.89	0.080
Pie chart	.91	0.157
Treemap chart	.92	0.130
Network graph	.95	0.041
Time series graph	.92	0.068
Bar chart	.95	0.057

5.3.3 Run T-Test statistic

The next aim is to evaluate how different individuals complete the form for various types of visualisation in the same way. For our analysis we have two groups: group1 – the PhD group; and group2 the student group). The hypothesis is that there is no difference between the grades between these two groups over all six visualisations.

In this section, we will explore this finding using an independent samples t-test. This test compares the means of two independent samples populations to validate if there is a significant difference between these two associative means samples of total score for each visualisation or not. The hypothesis requires a parametric test because it needs at least one dependent variable (test variable) and two independent grouping variables. Running this test helps us to discover the variance between users critical judgements or decisions on visualisations by checking the total score value for each visualisation per user.

We do note that the samples of populations are different in two groups (10 PhDs students for group1, 20 undergraduate students for group2). But, in this test we took the same populations sizes (10 PhDs students for group1 and 10 undergraduate students for group2) to avoid any violation effects on assumptions required for the t-test which might yield an inaccurate p values. We selected 10 users of 20 of those who have completed any non missing data in CDS sheets tasks. Because missing values in the spreadsheet will be treated as 0 values which violated the test results. So, in this test we need to

measure statistically if there is a significant difference in the user's evaluations on six visualisations between the two groups. The dependent variable of the T-Test is the total questionnaire value for each visualisation, the mean values and Standard Deviation (SD) for these two samples populations are calculated. We run this test using a SPSS compare means analysis. Table 5.2 shows the calculated significance ρ values for each visualisation.

Table 5.2: Independent samples t-test results that shows ρ -value at significance level $\alpha = 0.05$ and associative mean and SD for nominated visualisations (n=6) comparisons.

Visualisations	Group	Mean	SD	Sig. $\rho >0.05$	Bonferroni ρ adjustment
Bubble chart	1	102.6	18.4	.564	3.36
	2	108.0	13.2		
Pie chart	1	113.9	20.4	.623	3.72
	2	99.2	16.13		
Treemap chart	1	78.6	21.5	.728	4.32
	2	67.9	17.4		
Network graph	1	111.9	11.4	.007*	0.042*
	2	90.6	25.17		
Time series graph	1	105.2	11.2	.051	0.25
	2	112.8	19.0		
Bar chart	1	102.8	19.2	.53	3.18
	2	122.0	17.4		

* Significance with $\rho < 0.05$.

The output of independent samples T-Test in SPSS provides detailed information about the test variables values. Statistic information is displayed means and SD for each groups and independent sample describes Levene's test for the equality variance section which contains the F value (test statistic) and the important value is the ρ -value. Based on this value the null hypothesis can be rejected or accepted. Table 5.2 shows the calculated significance ρ values for each visualisation.

After analysis the ρ -value for the bubble, pie chart, treemap and bar chart visualisation are greater than the chosen significance level $\alpha = 0.05$, so we do not reject H_0 the null hypothesis that there is no statistically a 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 very close to 0.05, still it's greater with 0.001, yet the null hypothesis is not rejected based on the equal variance between groups.

Since not all the null hypothesis has been rejected, and it would be better in t-test comparisons to assure that statistically all the null hypothesis to be accepted which reflects a non-significant difference between means values. In that case, we can apply an adjacent correction value which is called the Bonferroni correction. This procedure statistically is a way of dealing with a Type I error which inflates when a multiple t-test is done. However, the Bonferroni correction rate is very useful when there are hundreds of pairwise comparisons on sample populations, but we could use this procedure even with a small number of comparisons. When we used the Bonferroni ρ value which is calculated by taking just the alpha level α divided by the number of tests (comparisons) αn , and compare the ρ value with that correction rate, in case of $\rho < \alpha n$, we would reject the null hypothesis for this individual case. When we replaced each ρ with $n * \rho$, and compared these adjusted values with α . Table 5.2 displayed the Bonferroni adjustment values for each ρ , we can see the network graph still has small ρ value, smaller than α and again the null hypothesis is rejected. That means the possibility of increasing this value using this procedure is still unsuccessful. It is important to look at the users' data for the network graph visualisation, we can say the problem with this data is that it is quite non-normally distributed within the sample group. This considerably reduced the power of the test, the high variance in total values between the users group causes violation of equality means assumption which yield to significance ρ less than 0.05. Figure 5.16 explains clearly the big difference between those 2 groups in total values for the network graph which confirms our view that group 2 data does not fully satisfy the requirements of the t-test.

Subsequently, we use a non-parametric t-test to check the normality of data distribution, even if we assumed that the data in the observed sample follows a normal distribution. We run Mann Whitney U independent samples on the same data samples of the parametric t-test on SPSS. Figure 5.14 shows the output results of the test. The results confirm our assumptions with the normality distribution of data, and also shows two rejected hypotheses regarding the significance of mean variance for the network graph and bar chart graph. While in the parametric t-test we gained just one rejected hypothesis, that

Figure 5.14: The results of non-parametric independent samples test for group 1 and group2 (both sample size=10)

Hypothesis Test Summary				
	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Bubble_chart is the same across categories of groups.	Independent-Samples Mann-Whitney U Test	.631 ¹	Retain the null hypothesis.
2	The distribution of Pie_chart is the same across categories of groups.	Independent-Samples Mann-Whitney U Test	.063 ¹	Retain the null hypothesis.
3	The distribution of Treemap_chart is the same across categories of groups.	Independent-Samples Mann-Whitney U Test	.089 ¹	Retain the null hypothesis.
4	The distribution of Network_graph is the same across categories of groups.	Independent-Samples Mann-Whitney U Test	.043 ¹	Reject the null hypothesis.
5	The distribution of TimeSeries_graph is the same across categories of groups.	Independent-Samples Mann-Whitney U Test	.190 ¹	Retain the null hypothesis.
6	The distribution of Bar_chart is the same across categories of groups.	Independent-Samples Mann-Whitney U Test	.023 ¹	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

¹Exact significance is displayed for this test.

means that we will consider whether the first t-test is best statistical result that meets with our hypothesis.

Another way to look at this data is to plot the averages of the answers between group1 and group2, and across each of the six categories (Perception, Environment, Interface, Components, Design, Visual Marks). This is shown in Figure 5.16. Where there is a clear difference between the averages the length of the bars would differ substantially. For instance, merely looking at the Network visualisation between group1 and group2, in the Environment Category (top right of Figure 5.16) it is clear that group1 are positive, and group2 are negative. Meaning that group1 are generally selecting that the Environment category is somewhat suitable; whereas group2 are evaluating that the Environment is totally inappropriate for the Network visualisation.

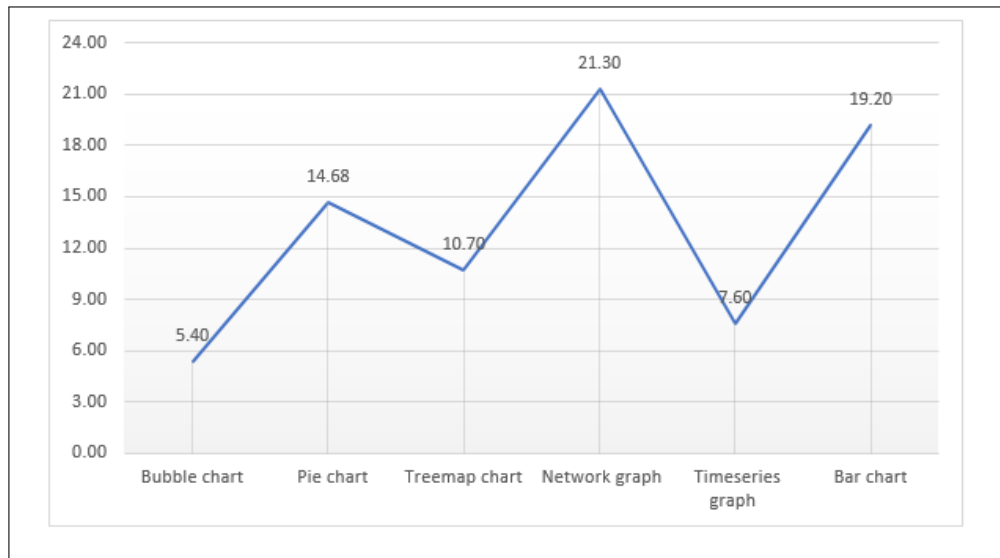


Figure 5.15: The mean values differences between users groups G1 and G2. The network graph shows a large mean difference and the bubble chart shows the lowest mean difference.

Mathematically we can look at the SD for visualisations that have an p -value < 0.05 and have a large mean difference between values (more than 4) from the average values to other visualisations. For the visualisation, again look to Figure 5.16. We can see the data distributions of six visualisations categories for both groups. Simple calculations have been done on categories' raw data on Microsoft Excel files to find out comparisons between the average value to a single category for six visualisations in both groups. Figure 5.15 shows the line graph of the absolute difference in mean values in table 5.2 of two users groups. The network graph has the high difference value from the other visualisations and that is confirmed by Figure 5.16 which indicates that the Network graph has a significant difference for mean values in the CDS categories analysis. Environment, Interface and Visual marks categories presents this difference average between groups. While other visualisations have better consistency on different categories for both groups even when two groups have an unequal size number.

Additionally, to make a t-test powerful and verify requirements, we can apply a new distribution for group2 users' data. We can replace the low users' total values with another users' data that has high total scores. After applying that substitute solution, we get good results. All the hypothesis has been accepted which refers to the equality of variance between users' groups and the goal is met. SPSS tables 5.17 show the results of groups statistics. We can say in this test we achieved the test assumptions of equality hypothesis which means that all null hypothesis are not rejected based on the significance p -value greater than α .

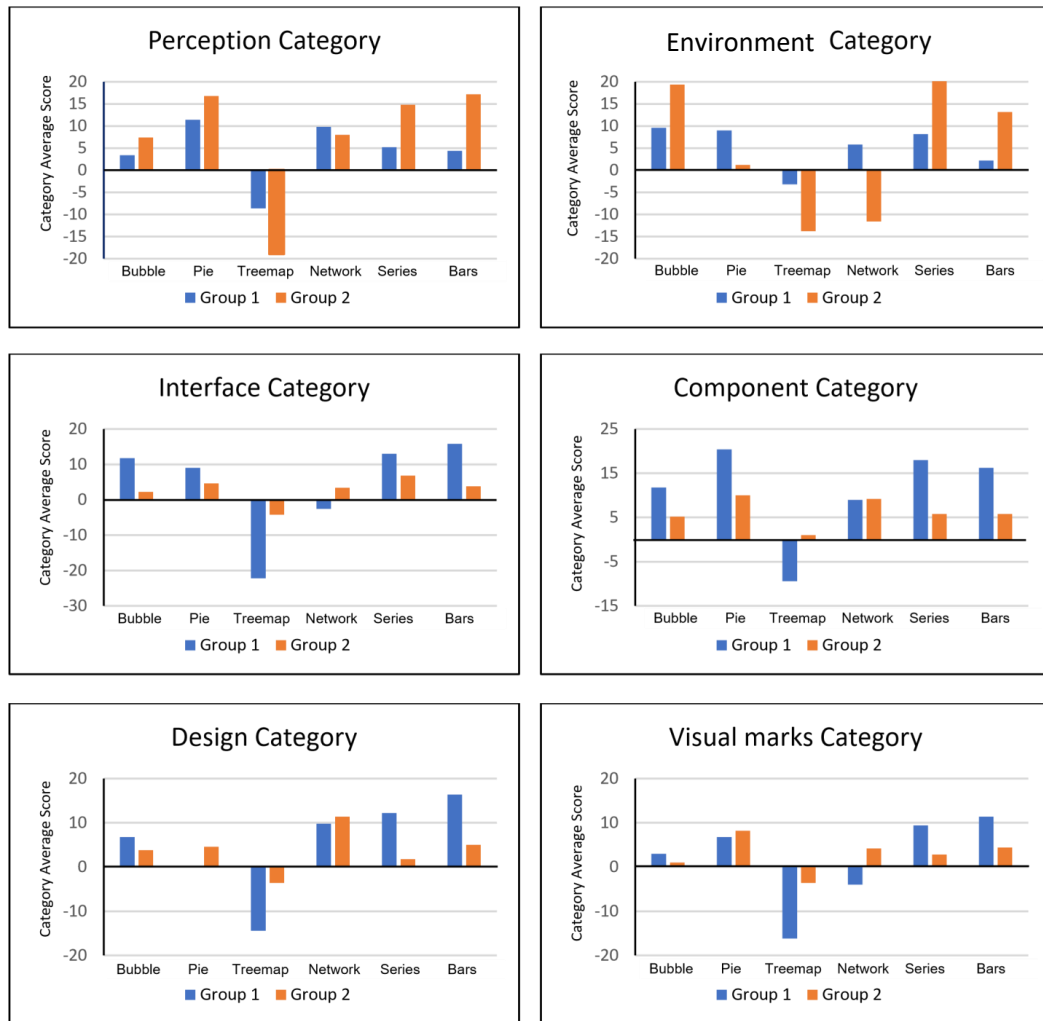


Figure 5.16: Categories data analysis for two users groups G1 and G2.

5.3.4 Data analysis of the “First Impression words”

The next evaluation was to explore how participants used the “First impression words”. The Design chapter 4 describes the twenty words, that participants can circle. For reference these are as shown in Table 5.3.

Table 5.3: The First Impression words, that were finalised. The user is meant to circle five words that best describe their visualisation to critique.

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

To analyse the words we looked at the frequency of word use. This provides an aggregated view of how most people classified the visual depictions.

Group Statistics					
	groups	N	Mean	Std. Deviation	Std. Error Mean
Bubble_chart	1	10	102.60	18.416	5.824
	2	10	106.40	14.968	4.733
Pie_chart	1	10	113.90	20.415	6.456
	2	10	99.60	16.588	5.246
Treemap_graph	1	10	78.60	21.567	6.820
	2	10	68.20	16.491	5.215
Network_graph	1	10	111.90	11.484	3.631
	2	10	107.20	16.335	5.166
Time_series_graph	1	10	105.20	11.282	3.568
	2	10	112.90	19.536	6.178
Bar_chart	1	10	102.80	19.211	6.075
	2	10	115.50	17.335	5.482

Independent Samples Test							
		Levene's Test for Equality of Variances		t-test for Equality of Means			
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference
Bubble_chart	Equal variances assumed	.038	.848	-.506	18	.619	-3.800
	Equal variances not assumed			-.506	17.278	.619	-3.800
Pie_chart	Equal variances assumed	.162	.692	1.719	18	.103	14.300
	Equal variances not assumed			1.719	17.276	.103	14.300
Treemap_graph	Equal variances assumed	.167	.688	1.211	18	.241	10.400
	Equal variances not assumed			1.211	16.843	.242	10.400
Network_graph	Equal variances assumed	1.526	.233	.744	18	.466	4.700
	Equal variances not assumed			.744	16.150	.467	4.700
Time_series_graph	Equal variances assumed	4.375	.051	-1.079	18	.295	-7.700
	Equal variances not assumed			-1.079	14.402	.298	-7.700
Bar_chart	Equal variances assumed	.632	.437	-1.552	18	.138	-12.700
	Equal variances not assumed			-1.552	17.813	.138	-12.700

Figure 5.17: Parametric independent samples t-test calculation for new sample users of group2 with same group1 users. The test shows the significance p -value > 0.05 for all 6 hypotheses.

To achieve the analysis we aggregated the selective impression words across both groups. The impression words data is represented in two different visualisations, Figure 5.18 shows a word-cloud visualisation of the impression data for both groups. We can notice such similarities in both groups selections words even when the groups are not equal in number, pie chart and bubble chart highlight high consistency, most users used similar impression words such as *sensible*, *clear*. With a Treemap chart all users agreed on *complex*, *confusing* as a negative impression on that design. Time series graph

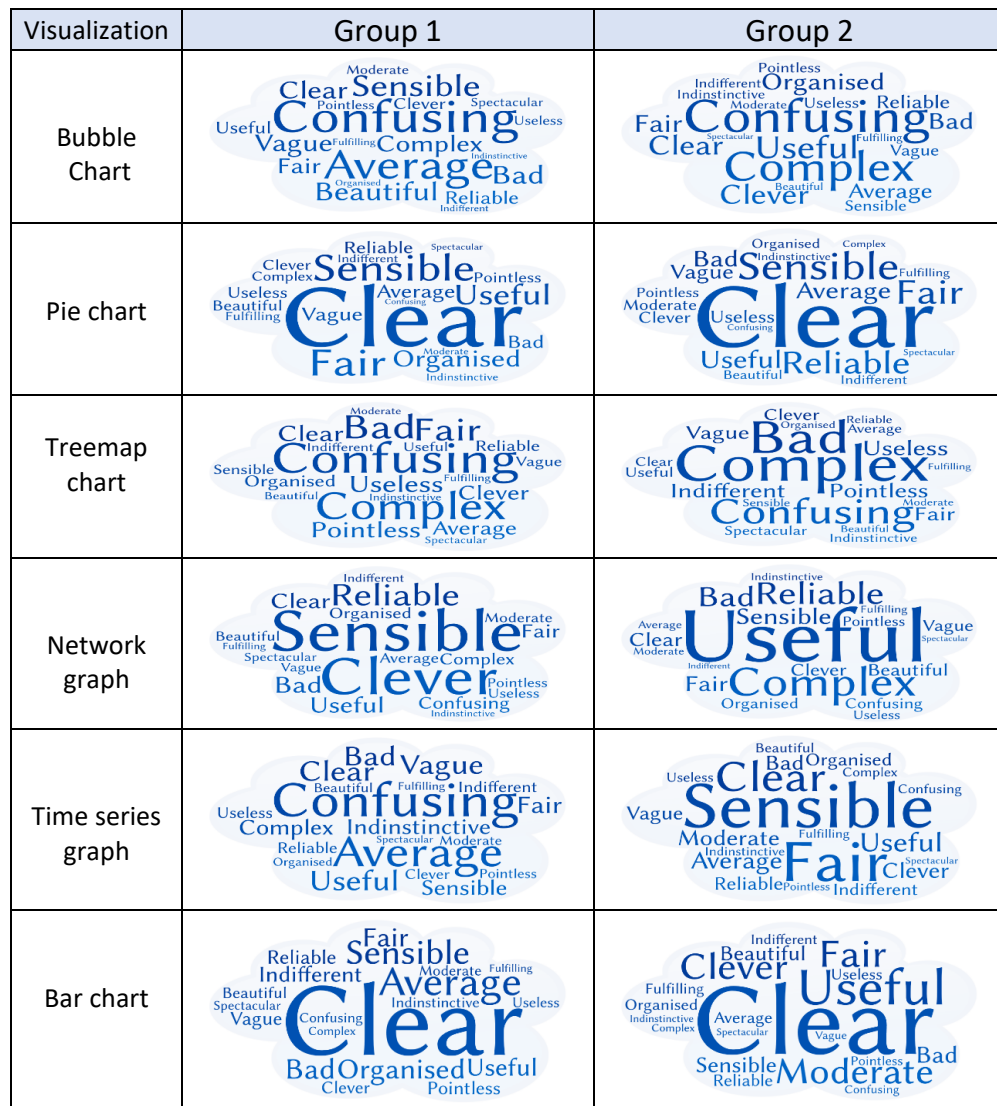


Figure 5.18: Word-cloud visualisation shows an analysis of impression words (section A). It shows such similarities between users' groups selections Pie, Bubble, and Bars charts shows the users' consistency in the use of most words, apart from spectacular, fulfilling and beautiful.

additionally, the network graph showed variance in users selections of impression words, there is less agreement in choosing similar impressions and that is due to a variety of negative, positive and neutral impression words in CDS. Users have free choices to choose what words they need to reflect their emotions.

We need to be more precise when displaying impression words in data. Therefore, we look into more detail, to work out how the participants have selected the First Impression words. Therefore, we use a bubble plot visualisation to show the number of selected words in both groups (group1 and group2), across each plot (Bubble, Pie, Treemap, Network, Time and Bar chart) and for each word classification (Positive, Neutral and

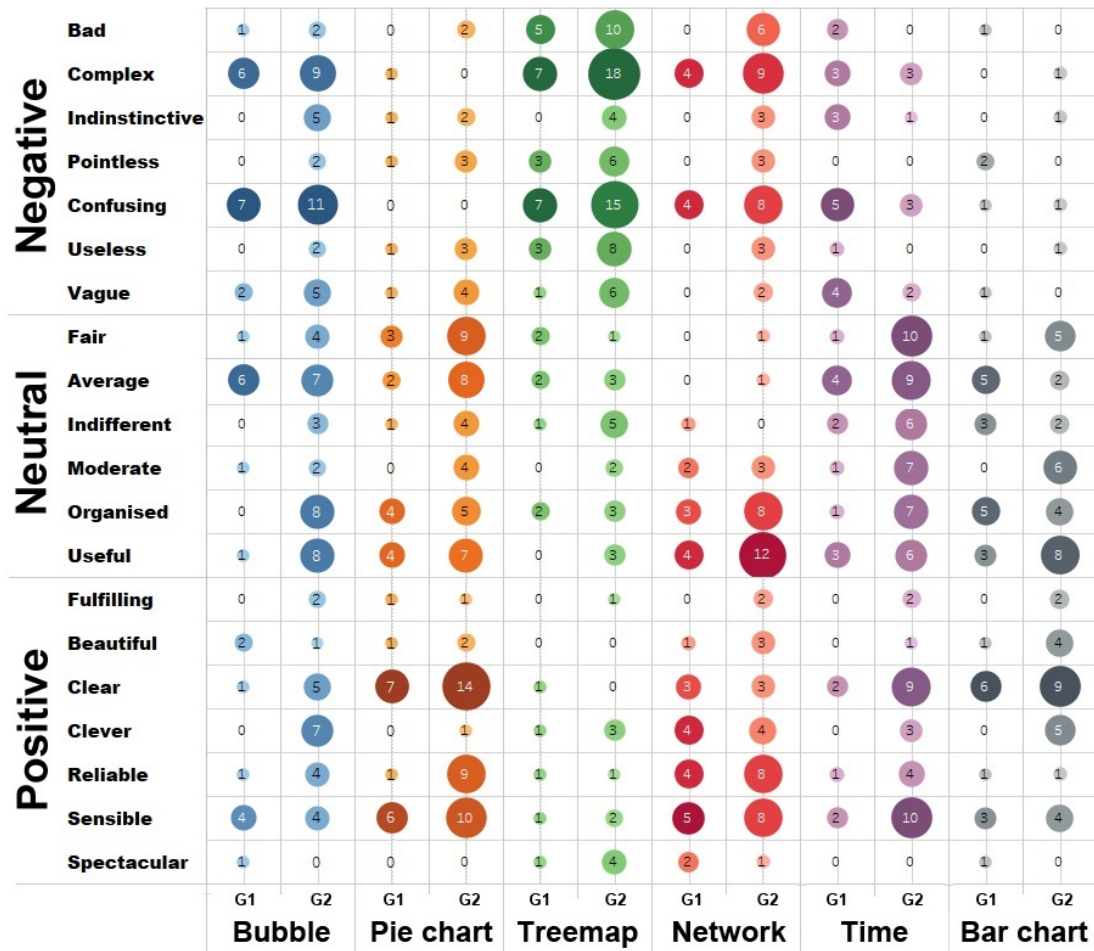


Figure 5.19: Analysis of word usage from section A) the first impression words, which shows a broad use of most words, apart from spectacular, fulfilling and beautiful. The graph is created in Tableau 10.2 software

Negative). This gives us an opportunity to judge the consistency and agreements between groups selections. We show the popularity of the words in Figure 5.19.

While there are some differences between groups, there is overall agreement. It is clear that some words are more popular, such as *confusing*, *complex* and *bad*, which were chosen more frequently than other words. Also, some words such as *spectacular* and *fulfilling* were less used throughout.

5.3.5 Learnability test

Learnability is considered one of the important usability aspects specifically in learning new software, systems, interface design [108], [110] or a website. In this research,

Table 5.4: Users time completion data to perform critique six different visualisations. Time taken for group 1 (N=10) and group 2 (N=20).

Visualisations	G1 time	G2 time
Bubble chart	00:21:09	00:15:06
Pie chart	00:08:07	00:09:08
Treemap chart	00:10:00	00:08:04
Network graph	00:08:04	00:08:00
Time series graph	00:11:00	00:07:00
Bar chart	00:09:06	00:60:03

learnability can inferred from the amount of time that users take to perform one visualisation critique.

We analysed how long the participants took to evaluate each sheet. Table 5.4 shows the average of completion time for each visualisation across users' groups. The hypothesis is that the first few visualisations will take longer than the last few, and that each visualisation should not take too long overall. In other words, that the users will get to know how to use it. We certainly did observe that the first CDS completion took much longer than following occurrences, with each sheet completion getting faster. Participants seem to head towards completing the sheet in 5 to 8 minutes for group 2. Whereas group 1 the completion time of the sheet was from 8 to 11 minutes. The time taken to evaluate the visualisation may be dependent on the complexity of the visualisation itself (or rather the perceived complexity). It may also be dependent on familiarity with the visualisation. Our results demonstrate that it is definitely dependent on the familiarity with the CDS process itself. Consequently, from the results, as shown in Table 5.4 we believe that decrements in the time of critiquing each sheet provides a good indication that participants do get faster and start to realise how to complete the CDS effectively. However, questions about the complexity of visualisations and readability are out of the scope of this research, and are therefore left to other researchers to investigate. Work by other researchers has started to explore aspects of visual literacy (for instance, see work by Boy et al. [24] on assessing visualisation literacy).

5.3.6 Qualitative evaluation – a Case study of its use

Practically, it is important to put CDS into a real situation, to enable users to use the CDS for a real application, rather than the test six visual depictions that were used in the evaluation tasks.

The CDS was used as part of the critical thinking and evaluation stage of the ICP3036 Computer Science module. This is a third year module that focuses on Computer Graphics and Visualisation. Students were tasked to create a visual explanation of a Computer Graphics algorithm. Their visual explanation was meant to be an animation that explains the many steps of an algorithm. A list of 80 algorithms was offered to the students, and each student chose a different algorithm to explore. Some of the popular algorithms included: Marching Cubes Algorithm and Dividing Cubes Algorithm [38] or other graphics algorithms such as Bresenham's line drawing algorithm [28].

Students were asked to create a design of their explanatory animation using the Five Design Sheet (FdS) methodology. In other words, they sketched their ideas on sheets of paper, using five sheets. The first sheet is the ideas sheet. The second, third and fourth sheets developed three (possible) solutions, and finally the fifth sheet is the realisation sheet that they would then implement. The students also created a storyboard of their animation, such as to plan how their solution would explain the algorithm. Finally, they then implemented their design. They used either Processing.org or OpenGL to implement the solution.

The students used the CDS in two places. First, after their sketches, and second after their implementation. In the first place, they reflected on their sketched animation ideas. They used the CDS as a basis to adapt their ideas in the FdS and improve their solutions. They then used the CDS sheet after they had implemented their final code. They then used both of these CDS sheets in their final critical analysis report, within which they were asked to reflect upon the positives and negatives of the design/implementation. It is worth stressing that the students were encouraged to use the structure of the CDS in their critical reflective report.

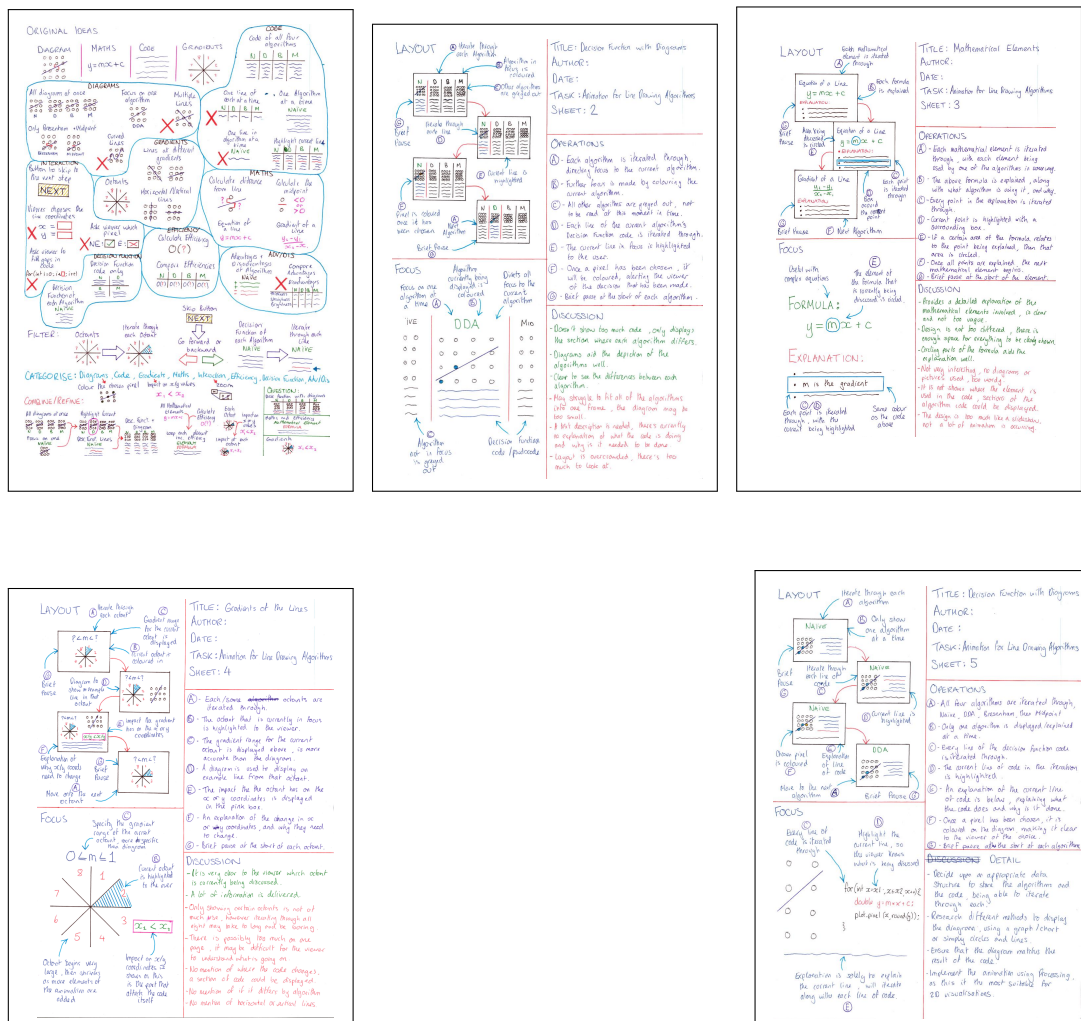


Figure 5.20: The Five Design-Sheets of the student Abdul, who looked at the midpoint algorithm.

Such that a reader understands what an FdS appears like, we include the results of a student's answer to the ICP3036 assessment. Let us call this student "Ali", as we have made his details anonymous. Ali chose the Midpoint algorithm. This is a line-drawing algorithm, much like the Bresenham line-drawing algorithm, it takes the equation $y = mx + c$ and changes the floating point values into discrete pixels that can then be plotted on a digitised screen. The results show how Ali has considered different concepts to visualise and explain to the reader. He has also considered alternative solutions (in the FdS sheets 2,3 and 4) and has thought about a solution (fifth sheet of the FdS). Figure 5.20 shows the five sheets of the Five Design-Sheets method. Figure 5.21 shows the final animation that the student achieved. Figure 5.22 shows the two versions of the CDS that they achieved. On the left is the CDS sheet that they completed on their FdS design and on the right is the one they completed after they had finished coding the animation.

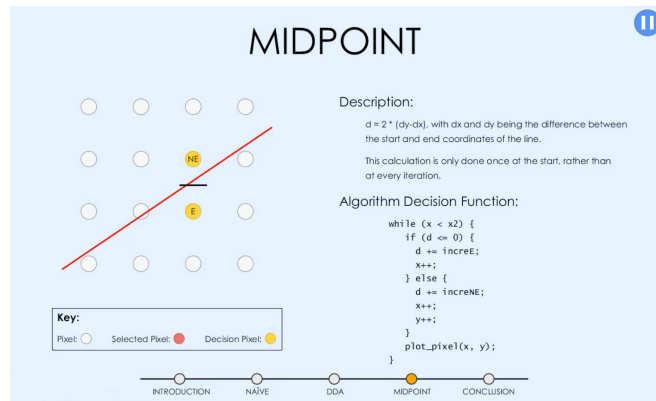


Figure 5.21: A single screenshot of a frame from the final animation of student Ali, who investigated how to create an explanation visualisation of the midpoint pixel algorithm.

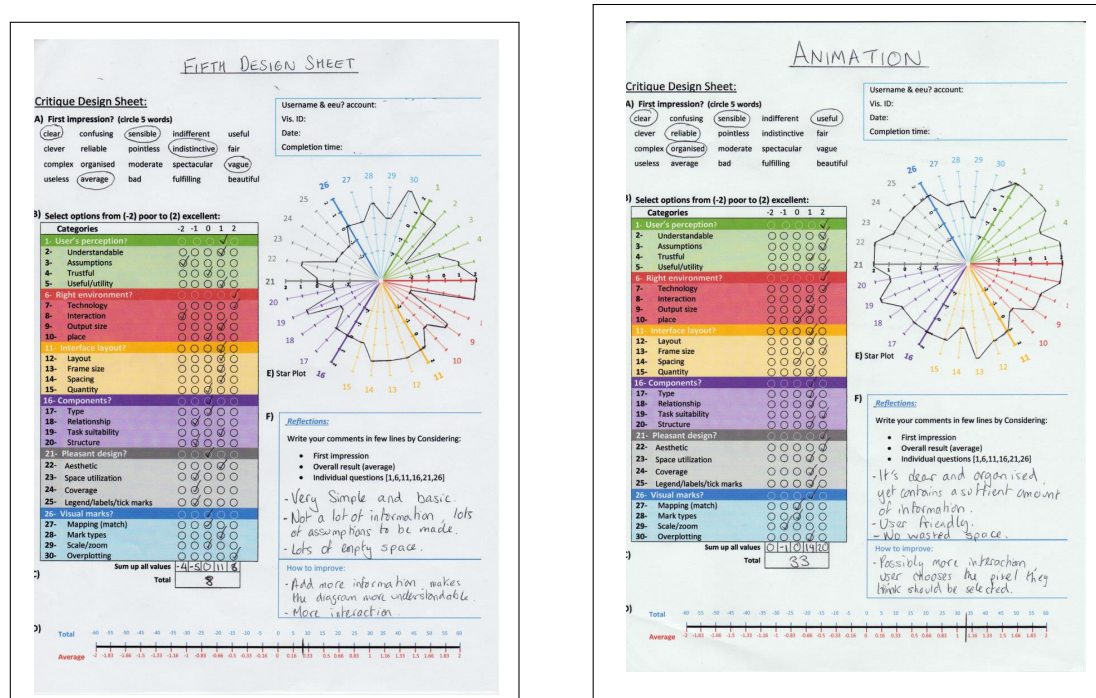


Figure 5.22: The CDS sheets of Ali, who investigated how to create an explanation visualisation of the midpoint pixel algorithm.

For the evaluation of his FdS, Ali writes many interesting comments about his use of the CDS. It demonstrates that the CDS can, and does, act as a prompt. It demonstrates that Ali has thought deeply about his problem, and demonstrates that the CDS is performing (to this student) as we have hoped. We include parts of his critical reflection below, such as to demonstrate this situation:

"To review my final idea, I have completed a Critique Design Sheet, which enabled me to review many different aspects of my idea. My first impressions of this animation idea was that it is very clear and simple, yet not too interesting or exciting... The next step of the Critique Design

Sheet is to score the animation idea based upon a variety of categories. I scored this idea an 8, meaning that I am more positive about this idea than negative.”

“[...] An issue that I had discovered when completing the [CDS] sheet was that my design contains a large amount of assumptions. This animation only explains how the algorithms work, it is assumed that the viewer knows why these algorithms are needed in the first place. To solve this, some kind of introduction is required, giving an overview of line drawing algorithms as a whole.”

“[...] From plotting the scores from the previous steps onto the Star Plot, I discovered which categories my idea is strong in, and which it is not. My designs’ strongest category was Interface Layout, with the weakest being Pleasant Design. I found this very surprising, as I felt that these categories are somewhat similar, yet they differ so greatly in the scoring.”

After Ali has created his implementation, he then reflected on his design, using the CDS. Again, we include some of his words, such as to demonstrate that deep thought has occurred in this student, using the prompt of the CDS.

“To review my final animation as a whole, I once again completed a Critique Design Sheet. The first impressions gained from this animation were relatively similar to that of the idea on the fifth sheet, with it appearing clear and understandable...”

“The score from the categories was also much more positive, with a score of 33. This means that overall I am very pleased with the result, with a lot more positive opinions than negative. The only question that returned a poor answer was the use of zooming. This is not something that I am too worried about as I feel that a zoom feature would not benefit or assist the explanation much at all. The Star Plot also displayed how much happier I am with my final implementation than the original design, with almost all areas increasing in score. There was also not any category in which my

animation struggled too greatly, with the Pleasant Design category issues that I had on the previous CDS being removed.”

It is clear, from this last example, that the CDS does help, especially if students follow the structure of the CDS. To exemplify this situation, we include the reflections of another student, which demonstrates a clear negative opinion of the FdS (realisation sheet) and then an improvement in the final implementation. The CDS of this student’s work is shown in Figure 5.22 on the left. This student writes:

“The final design displayed on the fifth design sheet was confusing, pointless, indistinctive, vague and useless. I filled out the critical design sheet based on the design and it scored very low with an average of -1.33, which is the score for a very bad design. The design was not easily understandable by the user and it was not useful. The technology I was going to use would not have proved effective for the data complexity of the design. The design was very much a statistical representation rather than a graphical explanation. It was not aesthetically pleasing and it did not use the frame well.”.

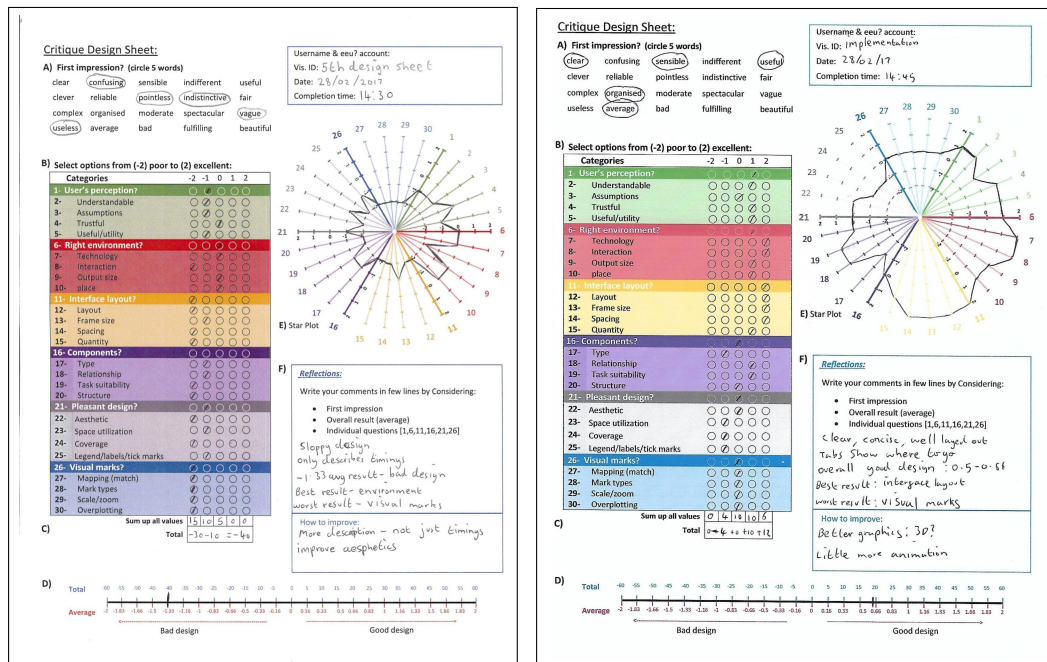


Figure 5.23: Student critiques, showing how they have improved the visualisation, and have graded the CDS appropriately. It is clear that the left evaluation value before implementation is worse than that on the right (after implementation).

Then, the student reflected on his creation. Figure 5.23 on the right shows that after his improvements he is judging his work to be better, with a higher than average value. This student wrote:

“This is why I went on to change the design completely for my final implementation. I again used the critical design sheet to evaluate the implementation. I will break down my analysis in to the sections as they are laid out on the critical design sheet.”

Finally, another student realised that there were flaws in his first sheet (Figure 5.24 left), and said as follows:

“ Before beginning my implementation of the idea, I completed a CDS on the Fifth sheet of the FDS. After completing the CDS, it was made apparent that it was on average a strong idea, as it scored 43/60 on the CDS. However, I had come to the realisation that the idea still had some flaws. ”

Then after tackling some ideas flaws regarding design problems, his evaluation on final implementation become as shown in Figure 5.24 (right).

“The implementation was good, I performed a CDS on the application. The score the application obtained was 53/60, which is better than the previous CDS performed. However, there is still room for improvement such as showing the normal arrows in a better way and making the view frustum a bit clearer. ”.

Additionally, to gain the benefit of our proposed method students were asked to write comments on the positive and negative aspects of the CDS, and how the CDS helped them to critique their own visualisations. Twelve participants of group 2 replied. Here's several samples of students feedback. One student wrote,

“It helps you see on paper what's good and what's bad”.[..]

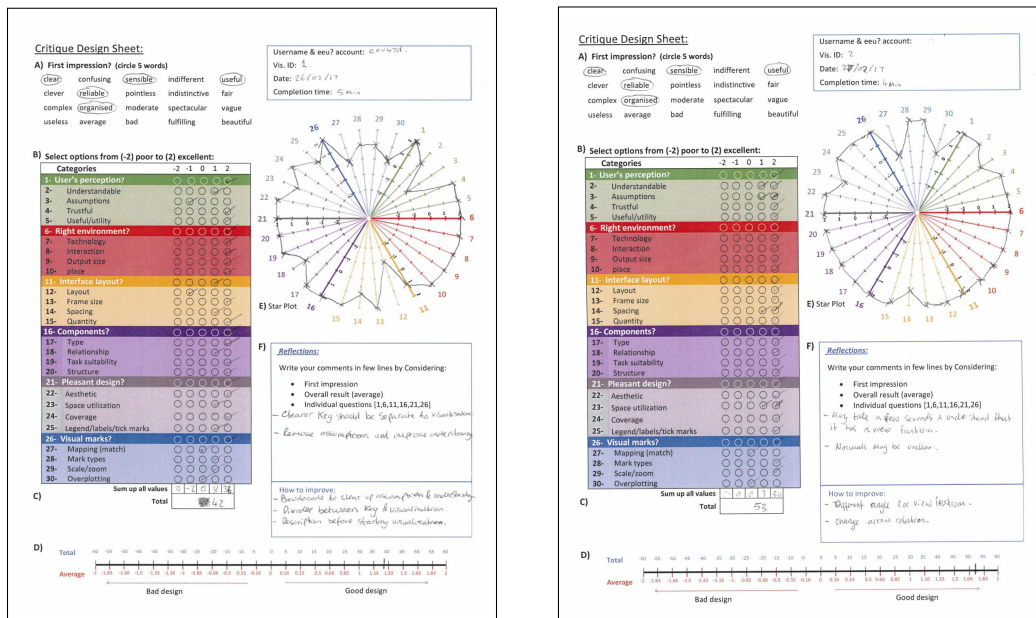


Figure 5.24: Student critique’s work shows the improving of evaluation value before implementation (on left) and after implementation (on right)

“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 the developers clear areas in which they can improve their visualisation on”.

Another said:

“The CDS’s structured approach to design evaluation helped me critique and analyse my design. It provided a very good structured way for me to critique my work on pre-determined evaluation points. It also made me consider some evaluation points I would not have thought of if I were to not use the methodology.”

Another student wrote his reflection on using CDS in his work:

“The critical design sheet greatly helped me compare my original design with my final animation. Originally I described my design as: sensible, organised, moderate, average and vague. The comments associated with sensible and organised collate with the higher scores of Environment, Components and Design. The vagueness is represented in low scores for visual marks. I was happy with the overall score of 20.”

Another wrote:

“The CDS were quite 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.”

5.4 Summary

This chapter evaluated the CDS with a range of measures. We focused on usability and learnability. It is clear from the results that the system (the CDS sheet) is usable, and indeed the CDS has been used with real students, and has been demonstrated to be useful. Especially, their comments about the CDS are positive and helpful, and support the fact that the CDS is a useful tool. We also explored learnability, and conclude that it is certainly true that the CDS takes time to be learnt. However, most participants managed to achieve a suitable speed after doing the CDS three times. Although we did not explicitly invite students to answer about their speed of use of the CDS, no one in the students group² complained in their reflective report that the CDS took too long. We did ask them to express negative views (as well as positive views) and this was something that they did not discuss. Therefore, we conclude that, while the CDS does take time to learn, the learning speed is appropriate for this type of critical analysis.

This evaluation chapter also looked at the use of the words and how people completed the forms. It is clear that different people use slightly different styles (e.g., coloured in blocks of the Likert scale, rather than tick marks) but these differences have no observed effect on the outcome and use of the CDS. In addition, while some participants missed out some of the sections, in earlier trials of the CDS (as listed as Bad Practice in this

chapter), when they completed the final versions of the critical analysis reports for their explanatory animation, no one mis-completed the CDS form.

The overarching question of this chapter as to evaluating “how easy is it to use?” is a subjective view. Our evaluation supports that it can be effectively used, and that participants are generally consistent in their judgements.

Finally, one result from the CDS is the grade. This grade that can be used to express how good or bad the visualisation is judged to be. While these scores are subjective, we have observed (through students using it, and their critical evaluation reports) that the grade is a useful tool for the students’ own purpose. We also acknowledge that the score that is created is still based on subjective values. While our analysis does support the case that most people provide similar judgements, visualisation design is still going to elicit emotional and non-rational responses. It is not our goal to produce a single score for a visualisation, rather to create a tool (the CDS) that works alongside current design methods, and helps users critique their systems.

From analysis of data, we can say CDS meets with usability objectives of testing reliability and learnability and goals are achieved. Over all agreements and consistency of users are reached in the analysis of impression words to some extent. So, we are satisfied with these initial results of CDS and aim to develop a usability evaluation with more numbers of people to be able to fix the usability problems. We addressed two case studies as a practical usability for the CDS tool, they showed how users benefited from CDS in improving their reflections on design and their feedback confirms the usage of CDS on thinking of visual design development.

The CDS is a tool that is used on paper. It was deliberately designed to be used on a sheet of paper, such that it can be given to students and developers to be easily used. However, the next chapter looks at how the CDS idea can be adapted to be used on a computer. The computer-based implementation would allow users to store their results, adapt grades and see how their project has developed, using the CDS as a critical evaluation structure.

Chapter 6

Implementation of CDS online version

The chapter will address several objectives. First, the design and construction of a Critical Design Sheet website as an online tool to support users in their critical evaluation. Second, to present a usability evaluation for the CDS online tool using the SUS (System Usability Scale) evaluation questionnaire. Finally, the chapter presents a discussion, including a discussion of limitations, of the CDS online tool.

This chapter contains three parts:

1. Sections 6.1 and 6.2: Introductory material, where we outline the vision for building a prototype online CDS system.
2. Section 6.3: design and implementation of the CDS online tool.
3. Section 6.4: evaluation of CDS online tool.

6.1 Vision for an online CDS version

Now that the CDS paper version has been implemented, and the groundwork of the CDS explored fully, and that we have demonstrated it to be successful, and that it has been used within teaching and by students. The last vision in this research is to investigate how an electronic version of the CDS could be implemented.

This chapter presents a prototype implementation of the CDS. The vision is not to create a fully-functional system, as most of the dissertation has been focused on the creation of the CDS. However, the aim of this chapter is to look at the feasibility of implementing the CDS as an online tool.

The vision is therefore to create a computer-based version. In fact, we also look to build an online tool. There are many benefits to creating such a system. (1) to be accessible and applicable to a broad sector of learners. (2) Users can create an account to be enable storage of their critique work and retrieve it securely. (3) The website can become a resource for *good* or *bad* visualisations and users have the potential to review their work any time.

6.2 Web technology features and services

One of the key aims of the vision was to implement an online tool. It is worth presenting some brief background on Web Technology, and explain why an online tool was chosen.

Web technology has become an interesting field of study and development for a wide range of scientific and social applications. Either developers create web-based applications to be download on any device and used with the Internet, or create websites opened by a web browser and accessed from any computer. Whereas in the early 2000s most of the tools developed were standalone pieces of software, the move now is to build web-enabled software that runs within a web browser. This dissertation is unable to address all the advantages and describe all the possibilities of web technologies, so we will focus on the important points of using web technology that helps and supports our research.

When developing a visualisation tool, it is good to have it available to a wide set of users. Web technologies has enabled this to happen. For instance, we can see many successes of web visualisation programs being integrated into webpages. We can see this through the success of (for example) Tableau public and D3.org. It is therefore worthwhile for visualisation developers to build their applications as web-based applications or even a website interface, such that it can be accessed by a broad set of end-users.

Therefore, there are many advantages to using web applications or websites: (i) they are easy to access by end-users; (ii) developers have access to many toolkits to help them develop, these are flexible libraries that enable a wide range of styles and services to be easily developed; (iii) they are easy to develop over time and can be easily edited by the developer and updated; (iv) many sectors of users' interactive systems, and users know how to use them; (v) they can be deployed on a range of different devices (the software libraries automatically adapt the screen size and commands for the device that they are being shown on).

Web-based applications and websites construction generally consist of two main parts:

1. Client-side techniques: it calls the front-end User Interface (UI). This side in web technology is responsible for creating and designing interfaces and manipulating the document object model (DOM) for HTML documents. The DOM manipulation environment enabled us to do various tasks and functions such as updating the content of the page, displaying new UI elements, or loading entire pages in response to user gestures and other events. There are many client-side technologies but HTML, cascading style sheets (CSS) and JavaScript are three very popular combinations.
2. Server-side techniques: this technique involves hardware and software server sides, here we mean server-side is the software which is called a web-server. This software is included on several files and protocols that are responsible for controlling how web users access hosted files. The important protocol on the web-server is HTTP (Hypertext Transfer Protocol) which is used by browsers to view web-pages. We use this in constructing a CDS tool as a dynamic web-server which involves a web-server application plus database. It called "dynamic" because the application server updates the hosted files before sending them to the browser via the HTTP server. Server-side scripts provide strong and secure connections between users (client-side) and servers (web database) to manipulate data, and that it's the way to process our request on server. Well known server-side languages are PHP, Python, Ruby.

To build the website there are different software choices that developers can make. WordPress, Weably, Wix and other software help in quickly building websites with providing users with the required components and plugins without the need for too many efforts for creating coding. For those software has advantages and disadvantages, we do not need to make comparisons between them, rather we need to mention the variety of website software builders. In this research we started to use WordPress, however it became apparent that it was difficult to adapt it to our purpose. While it is easy for end-users to build software in the style of other WordPress pages, it is more difficult to setup a new style.

Therefore, we took the decision to use conventional techniques to build the CDS website. We specifically need to implement CDS sections in a dynamic environment and then generate evaluation results combined with the user's reflection. We used HTML, JavaScript and CSS to create the website, HTML provides a DOM environment (document object model) to create a static web-pages interface and JavaScript to add more dynamic function on the objects to enable the interaction on a website. CSS is to add attributes and style to the HTML elements and describes how these elements should be displayed.

6.3 CDS online tool constructions

To develop the CDS tool, there are two main considerations to make. First, what does the tool look like? Is it designed in the same way to the paper-based CDS? Second, how does a user approach the CDS online tool? What processes do they follow?

The main idea was to transfer the paper-based CDS method into online version to be a dynamic website and be applicable to many users. Consequently, early in the design cycle we decided to keep the CDS website simple and so to directly follow the main stages of the CDS. Creating CDS web pages requires handling dynamic content because we need to save and store users' information and their evaluation works on a web database. So, users need to create a personal account and log into the website. They need to do various activities, upload data, manipulate and save their data.

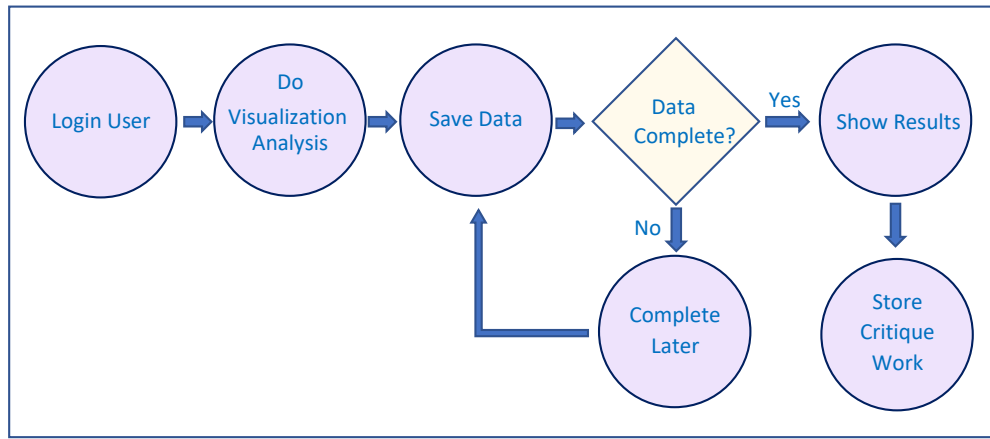


Figure 6.1: The CDS online tool process. It describes the pipeline steps that users need to follow in order to complete their critique task, it starts from a login page and ends with storing critique work after getting results.

After much thought, the process of the tool was decided to follow the five main steps. This is shown in Figure 6.1. First the user logs in to the system. Second, they do the visualisation analysis much like they would with the paper-based system. Third they can save/load the data. Fourth, they are able to come back to previous saved versions. Fifth they can look at the results and then work out how to use the results from the online CDS. In this way the user separates the creation of the scores, from the saving, viewing and retrieving of the grading, to showing the results. The CDS website therefore has the following principal web pages:

- Home page
- Login page
- CDS interface page.
- Result page.
- User's critique history page.
- Feedback page.

In the following sections we will explain the respective parts and describe how the UI was built using client-side and server-side technologies. To construct the online system

the technologies used, were: (1) HTML v.5 (2) JavaScript (3) CSS (4) Apache for XAMPP v3.2.2 web server (5) MySQL (6) PHP (7) Bootstrap. All these technologies are combined together and present a dynamic content website.

6.3.1 Home Page

The home page usually in websites is the start and welcoming page. The content of this page includes several parts: (i) description about CDS goals and purposes. (ii) Menu bar provides users with important tabs and functions that are required to be carried out as tasks. (iii) The foot bar includes information about the website. (v) The contact form enables users to contact website administrator(s). Figure 6.2 shows the home page of the CDS website.

The Menu bar includes several tabs that offer quick access to the required pages of Feedback, Help and LogIn. The feedback page contains the SUS survey that questions the usability of the website design and functionality. It considers CDS website usability evaluation to collect users' opinions to improve the website. The help page includes all the descriptions and how to use the CDS website, the terminology of CDS questions is organised on a table to describe the meaning of the category's questions.

The login page enables users to log into the website by asking users to register and create an account for new ones and just logging in for those that already have accounts. Figures 6.3 and 6.4 show the login and registration pages respectively, the website includes two login pages. The registration page is for new users to register the required information user-name, password and email address. We ask users to fill this information to allow them to be able to see all their visualisations that they have critiqued. We keep the information secure for individual users, that are stored in a database and keep login IDs unique to prevent any mix up with users. The second page is the normal login page with user name and password that provides access to the CDS interface page.

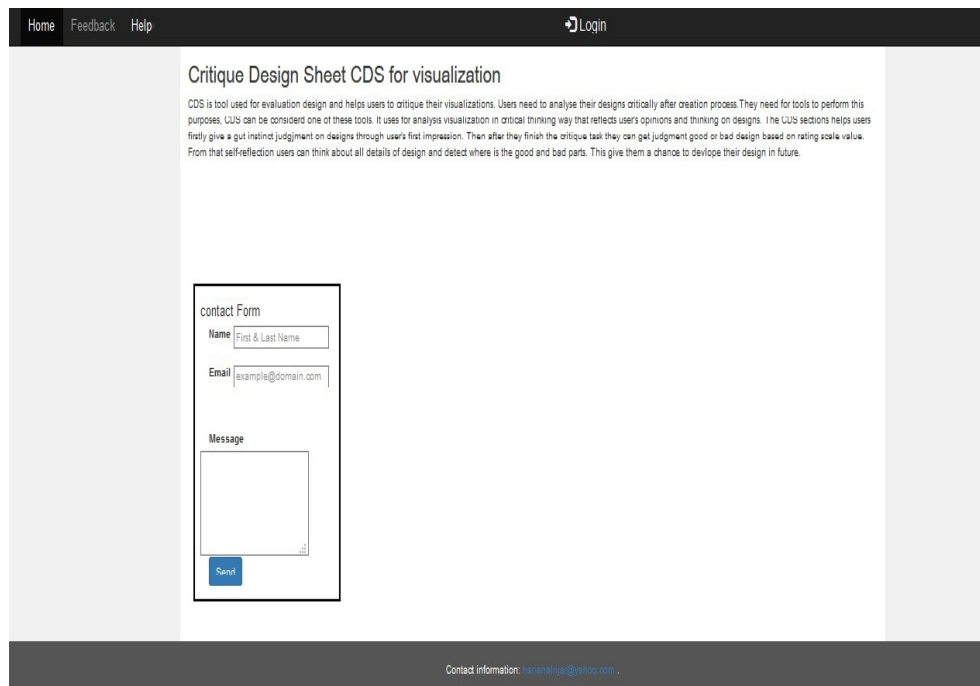


Figure 6.2: Home page for CDS tool website.

User Login ?

Figure 6.3: Login page for CDS tool website.

Registration Form

Please fill all information....

Figure 6.4: Registration page for new users to create an account for the CDS tool website.

6.3.2 CDS Interface web-page

This page is the main page of the website that users will work on to evaluate their design. Basically, it is very similar to the CDS paper version. Obviously, the online version can

add extra (additional functionality) in comparison to the paper-based equivalent. It is possible to automatically calculate the scores, and also the star-plot can be automatically drawn by computer. We therefore add the required functionality to make CDS suitable to web application work practice. Users can do various tasks to navigate between entire pages, upload design images, store data, calculate the average of evaluation and show critique results. The interface consists of components and buttons to perform a design evaluation. Figure 6.5 shows the CDS online interface.

To perform a complete critique design, users should fill all required data like upload images, select 5 impression words and select 30 question values. Additionally, a star plot graph is shown to visualise the choices that users make. The star plot just displays data values on screen, whether users complete all evaluation steps or not. Next users would need to save their results by pressing the Save button. All the data is saved on a database directly on a specific user's account. The green button (show results) then appears, pressing that the button will call the results web page to show evaluation analysis results. However, if users could not complete all the required data of CDS sections for design evaluation, users will not be able to show results at that moment. Data will be saved on the database under a status called "Edit", which will allow users to complete their critical analysis at a later time. Users can return to their records and review different evaluations from the "MyRecords" tab. Thus, users have the choice either to complete design evaluation steps and see the analysis results, the work will be saved as a complete status or can postpone the work for another time and will be saved as edit status.

We used different components to construct this web page to be similar for CDS paper-based versions. A HTML canvas is used to draw a star plot graph using available HTML graphical methods and JQuery (JavaScript library) to make the required functions. JavaScript offers the required functions to interact with DOM components such as uploading images, pop up alerts to inform users with various controls on data and other functional details. Figure 6.5 shows evaluation examples that have been done by one of the participants who attended the usability study of the CDS website.

[Home](#)
[CDS](#)
[Results](#)
[MyRecords](#)
[Feedback](#)
[Help](#)

[Log out?](#)

CDS Interface

To save your work press [Save data](#) when finish

Section1 : Select your file

Browse...

No file selected.

Section2: select 5 Impression words :

☐ Confusing

☐ Average

☐ Useless

☐ Clever

☐ Clear

☐ Indifferent

☐ Reliable

☐ Pointless

☐ Indistinctive

☐ Beautiful

☐ Organized

☐ Complex

☐ Sensible

☐ Useful

☐ Spectacular

☐ Moderate

☐ Bad

☐ Fulfil

☐ Vague

☐ Fair

Section3 : fill up the questionnaire:

key questions	selections				
	poor				good
1- User's perception ?	1	2	3	4	5
2- Understandable	1	2	3	4	5
3- Assumptions	1	2	3	4	5
4- Trustful	1	2	3	4	5
5- Useful	1	2	3	4	5
6- Right environment ?	1	2	3	4	5
7- Interaction	1	2	3	4	5
8- Technology	1	2	3	4	5
9- Place	1	2	3	4	5
10- Output size	1	2	3	4	5
11- Interface layout?	1	2	3	4	5
12- Layout	1	2	3	4	5
13- Frame size	1	2	3	4	5
14- Spacing	1	2	3	4	5
15- Quantity	1	2	3	4	5
16- Components	1	2	3	4	5
17- Type	1	2	3	4	5
18- Relationship	1	2	3	4	5
19- Task suitability	1	2	3	4	5
20- Structure	1	2	3	4	5
21- Pleasant design ?	1	2	3	4	5
22- Aesthetics	1	2	3	4	5
23- Space utilization	1	2	3	4	5
24- Coverage	1	2	3	4	5
25- Legend, Labels, Tick marks	1	2	3	4	5
26- Visual marks	1	2	3	4	5
27- Mapping (match)	1	2	3	4	5
28- Marks types	1	2	3	4	5
29- Scale/zoom	1	2	3	4	5
30- Overplotting	1	2	3	4	5

show starplot

Contact information: hananaingar@yahoo.com

Figure 6.5: The CDS online tool interface includes different components. HTML canvas, tables, buttons and radio buttons that are used to construct webpages.

6.3.3 Results page

This page displays the results of the CDS design critique after being saved and processed.

The results are represented in different forms: (i) star plot categories graphs, (ii)

critique design report card and (iii) the critique conclusion form. Figure 6.7 shows the results page components construction holding data results (i) Starplot categories graph canvas describes the individual category of the CDS questions values Perception, Environment, Interface, Component, Design and Visual marks. We can see from starplot data representation where are the good categories of evaluation and where are the bad ones, then we can decide how to improve the bad categories further. (ii) The critique design report card displays the average value of design (average of 30 questions) and also presents the weak categories and strong categories level. The bars on this window give a percentage for the average value to each category as statistical ratings analysis of CDS categories. In our vision, generating these types of analysis can support a user's understanding and self-reflection about design.

Figure 6.6: CDS result page interface mapping includes different frames and canvases that can be filled with data after pressing the blue result button

(iii) The critique conclusion form interprets perceptually the critical relations between CDS sections. One of the reflections on design is thinking of the relations between the emotions or first impression with the average value of the questionnaire. The relation

can describe the consistency and the fair mindedness of the judgement on visualisation design. From this point, the conclusion draws a fair picture to reflect the user's reflection towards design and that conclusion is generated automatically. Additionally, on this page we put the user's reflection form the same as on a CDS paper-based version. So, users are enabled to record their comments and further reflections to be saved with all data on a database. This is useful when users return to their records of the old critique's works and see the old comments, that could remind them about the critical points on that design.

We will show an example that addresses the evaluation design process. Users can follow the process diagram in figure 6.1 to perform the critique task. Figure 6.7 shows the full example of a user's critiquing, a user's filled CDS data sections, then the user calls the show results page in figure 6.8 which is displayed the evaluation average based on the entire data is 2.6 (the target > 3), and the percentage levels of categories described on the critique design report card. The starplot categories graph shows a very low rating for the Environment and Components questions groups. The critique conclusion helps users to reflect over all results, and help them conclude whether the design is good or bad. The user's impression of data selections was mostly negative. So, the conclusions present the normal association between a negative impression and low evaluation score which (in this case) is interpreted as a bad design.

6.3.4 MyRecords page

Once users have judged the visual depictions and saved the results, they need a way to load and display their information. Figure 6.9 shows the table of records belonging to user X. It highlights the important information such as username, Image name, the value of evaluation, date of critique and status. The status describes the evaluation status (if the evaluation is completed or not). Users can review the completed results under status. This status could be "Complete" or ongoing represented by "Edit". Users can then select the necessary page link to access the required page. Users can load pages to "edit" and then complete. After completion the data will be saved under "complete" status. The information is updated in the Mysql database, and displayed back to the user on the HTML page, the codes are written through PHP, we will explain this mechanism

[Home](#)
[CDS](#)
[Results](#)
[MyRecords](#)
[Feedback](#)
[Help](#)

[Log out?](#)


CDS Interface

To save your work press [Save data](#) when finish

You are logged in **paria**
[Show results](#)

Section1 : Select your file

[Browse...](#) migration-map.png



Section2: select 5 Impression words :

☐ Confusing
☐ Average
☐ Useless
☐ Clever

☐ Clear
☐ Indifferent
☐ Reliable
☐ Pointless

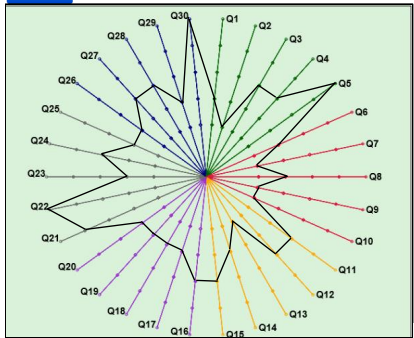
☐ Indistinctive
☐ Beautiful
☒ Organized
☒ Complex

☐ Sensible
☒ Useful
☐ Spectacular
☐ Moderate

☒ Bad
☐ Fulfil
☒ Vague
☐ Fair

Section3 : fill up the questionnaire:

key questions	selections				
	poor				good
1- User's perception ?	1	2	3	4	5
2- Understandability is the design clear? Can you understand what is it?	1	2	3	4	5
3- Assumptions	1	2	3	4	5
4- Trustful	1	2	3	4	5
5- Useful	1	2	3	4	5
6- Right environment ?	1	2	3	4	5
7- Interaction	1	2	3	4	5
8- Technology	1	2	3	4	5
9- Place	1	2	3	4	5
10- Output size	1	2	3	4	5
11- Interface layout?	1	2	3	4	5
12- Layout	1	2	3	4	5
13- Frame size	1	2	3	4	5
14- Spacing	1	2	3	4	5
15- Quantity	1	2	3	4	5
16- Components	1	2	3	4	5
17- Type	1	2	3	4	5
18- Relationship	1	2	3	4	5
19- Task suitability	1	2	3	4	5
20- Structure	1	2	3	4	5
21- Pleasant design ?	1	2	3	4	5
22- Aesthetics	1	2	3	4	5
23- Space utilization	1	2	3	4	5
24- Coverage	1	2	3	4	5
25- Legend, Labels, Tick marks	1	2	3	4	5
26- Visual marks	1	2	3	4	5
27- Mapping (match)	1	2	3	4	5
28- Marks types	1	2	3	4	5
29- Scale/zoom	1	2	3	4	5
30- Overplotting	1	2	3	4	5

[show starplot](#)


Contact information: homonulaga@yahoo.com

Figure 6.7: CDS interface page illustrates an example of the user's critique work (it shows Paria). All sections are filled with required data and ready to save data then show results.

in Section 6.3.5. It is important to mention that the displayed information only belongs to one user: the user holds a unique username to log into their account. They cannot see other people's data or edit/completed results. The table would show all records for a specific username.

Implementation of CDS online version 162

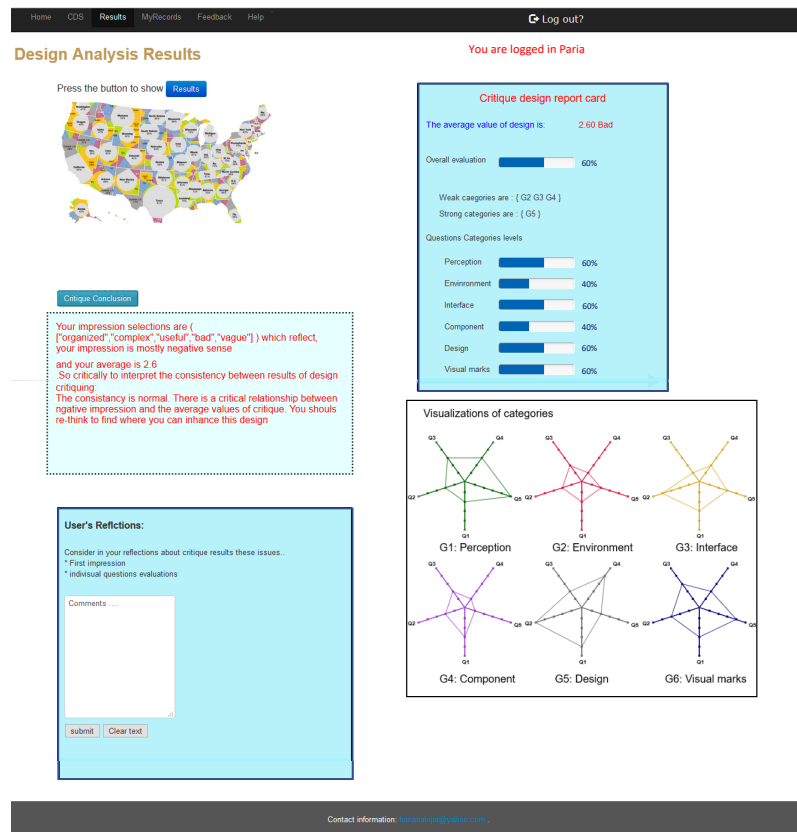


Figure 6.8: CDS result page interface mapping includes different frames and canvases that can be filled with data after pressing the blue result button.

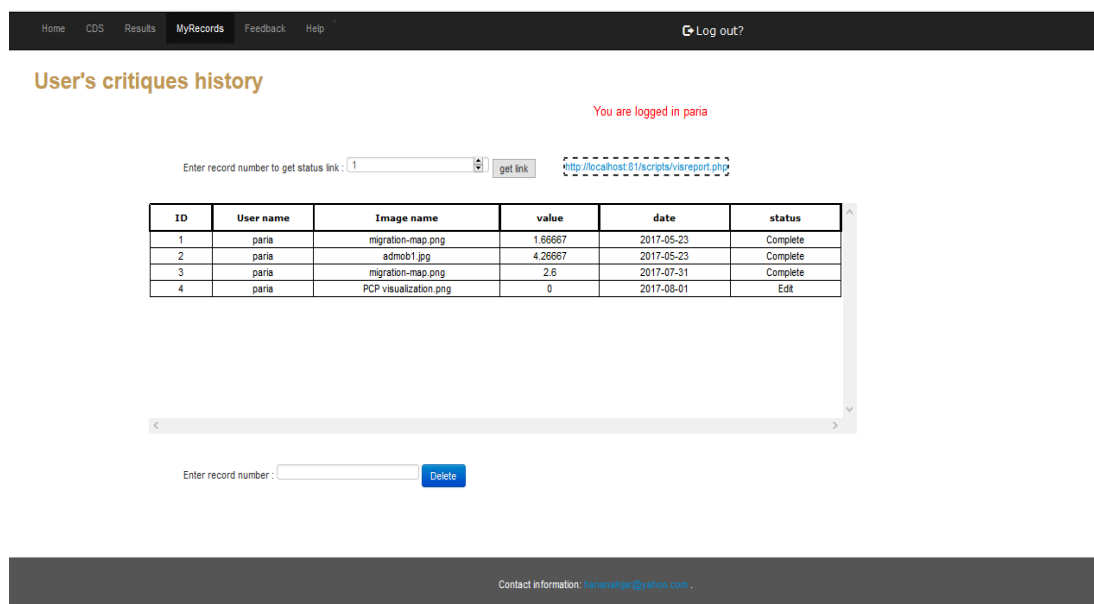


Figure 6.9: MyRecords webpage displays users' data which is organised as records on the HTML table. The records displayed the information of user 'Paria' who made several evaluation works on specific dates, with data invoked from the MySQL database.

6.3.5 Server-side of CDS implementation

In our website we constructed a web server and other scripting technologies locally.

This implementation is only meant to be a prototype, such as to demonstrate the

feasibility and worked through some of the design decisions. The reason for this could be adapted and developed into a fully working system, by a developer. Therefore, for this prototype implementation we used XAMPP Apache v3.2.2 service for Windows as a web server installed on a local PC. Bitnami which is an app library for server software (<https://bitnami.com/stacks>), we installed the required software or modules that needed our work like Mysql and PHP application from specific library site. It is worth mentioning that these open source technologies are popular and have a wide range of tools on web development domains. Because it is an open source, the community can see the original code and developers can write developed features that they want for upgrading. Therefore, by using these technologies, a developer should be easily able to take the ideas forward and make it into a fully working system. Another benefit of these technologies is free to be used by community and no need to buy a licence or purchase upgraded latest versions.

So, the Apache server serves the CDS website by supporting sending browser (client) requests to save the user's login information, image name, question selections and first impression word selections. All this information is transferred to a Mysql database on specific tables that have been created before. PHP is a flexible language and offers unlimited control with the web server. It embeds easily on a HTML document by using open `<?PHP` and close `?>` tags as a label to trigger PHP commands. Browsers are not responsible to parse PHP commands it just interprets the client's side scripting as HTML documents. PHP commands are parsed by web servers and send the request back to the browsers to show outputs [111].

It is useful to talk about types of connectivity between the PHP and Mysql database. PHP offers three types of connections to send and retrieve data with the database. Each one has pros and cons, for example using a `mysql-connect()` function easily handles data and setup a connection to the database. But it is deprecated for PHP 5.4 and upgrade's versions that's why using this function causes security whole in connection. W3Schools website (<https://www.w3schools.com>) describes some features about these connections. Using the PDO (PHP Data Object) and MySQLi extensions (the "i" stands for improved) are important for web application security because they support with PREPARE statement. Another advantage of PDO can work with 12 different databases whereas the MySQLi extension works only with MySQL database. In the CDS website

we construct the connection with the MySQL database using the PDO connection. We used the MySQL server address (the web server) it is most likely be localhost or 127.0.0.1, username, password (this was given for database when installed) and database name.

6.4 Evaluation of CDS online tool

Usability evaluation of websites is considered to be an important stage in the development of websites, so as to get users to be satisfied with the websites [126]. The System Usability Scale has been used by many web developers to evaluate usability on such systems [13], and have been used by university students on different types of systems [88].

In order to test the usability of the CDS website, we have adopted SUS (System Usability Scale) as an evaluation instrument to measure overall users satisfactions. SUS is a questionnaire of 10-items which reflects subjective assessment for system usability, the items yielded on Likert scales which are five choices from 1 strongly disagree to 5 strongly agree. Brooke [29] was the first researcher who presented the SUS questionnaire, he suggested that there's a need for methods with broad general measures that can be used to assess usability across a range of contexts. Then he put a simple small group of questions as a quick-and-dirty questionnaire as a method to evaluate the usability of industrial products. Various researchers have used the SUS to measure effectiveness, sufficiency and users satisfactions of systems or websites. Based on the context that SUS fits with, it considers a highly reliable usability method in its results. Tullis [157] conducted an empirical study to compare usability of websites by using five types of questionnaires, the study revealed that SUS has more reliable results among the other questionnaires over different sizes of samples and concluded that SUS can get reasonably reliable results even on a small size sample (e.g., 12-14) participants. Furthermore, SUS recorded a high robust tool as a usability assessment in extensive research for more than 200 studies having been used many times to evaluate a wide range of interfaces that include Web sites, cell phones, IVR, GUI, hardware, and TV user interfaces [12], [13], the robustness belonged to the flexibility of SUS and the statements actually cover a variety of aspects of system usability.

6.4.1 Method & Results

In this section we will explain how we used the SUS questionnaire to test usability of the CDS website. We conducted the experiment on a small size sample of 7 participants. We made a friendly invitation to colleagues and students within the School; most participants were our colleagues in the Computer science department. Five participants were PhD students, one was a computer science lecturer and the seventh an external contact who had comprehensive computing experience. The ages ranged between 25-45.

The procedure for conducting the experiment was as follows: all participants came into the computing laboratory, and worked on a specific PC with the libraries installed. Each participant was invited to attend the session, and came individually. They were given a short verbal introduction to the CDS, to inform them what they were going to be doing. All participants had prior experience in using CDS, because they had also been a participant at one of the other studies. This was a deliberate decision as it would enable us to spend less time on telling them what to do and how to perform the experiment. Indeed, we feel that this is a sensible situation, as most participants would to receive an introduction to the CDS method, before using it. Participants were asked to use the website to analyse a visual design and to create an account for personal use in future. As well as this, we asked them to navigate the website to explore various webpages to identify available assistance functions. After the participants had created an account they were able to log into the system and perform a critiquing task.

After they had performed the tasks they were directed to complete the System Usability Scale questionnaire [13] (of 10 questions), which has been used before in this dissertation. This SUS questionnaire was located by following a URL linked from the online CDS system. Google forms were used to create a SUS questionnaire template and responses were sent straight to the same form (and recorded in an excel spreadsheet, for later analysis). During the session, notes were taken. Notes were made to evaluate how the participants were using the system, and where they were looking.

We calculated SUS scores (users' responses) to measure the overall users' satisfaction when users were evaluating a visual design on the CDS online version. We followed

the calculations scheme according to Bangor et al. [13], when (and if) the SUS results (the SUS score) is over 85 then the system/product is highly usable, over 70 to 85 it is characterised from good to excellent, a value from about 50 to about 70 shows that the system is acceptable, but it has some usability problems and needs improvement, and finally, a system with a SUS score below 50 is considered unusable and unacceptable.

We calculated the score, with the 7 participants, having an overall users satisfaction level of 78.75/100. Therefore, this score puts the CDS online tool between being: good to excellent. The interpretation of our SUS result means that the CDS online tool can meet with the users' needs about how they analyse the visual design by the aid of a computer to show critique results and save data. Web technologies support users work practice on CDS online tool to make it faster and widely applicable, users can even address their reflections on design and store information on their accounts for future design improvement. In addition, users were satisfied with the overall website presentation, although this experiment was focusing on the overall accessibility and functionality of the CDS rather than assessing website aesthetics. Users gave a holistic point of view about the CDS website presentation. Figure 6.10 shows the SUS questionnaire.

6.5 Discussion and limitations of the online CDS system

This section starts to discuss some of the limitations of this work and study. We focus on (i) CDS online system (website) construction, and (ii) the website evaluation.

Right from the start of this chapter, we expressed that this CDS online tool would be a prototype. The main focus of the dissertation was to create the CDS system and methodology itself, rather than focus on any implementation. However, even with this limitation it is clear that it is feasible to create an online tool. Web technologies were deliberately chosen to implement the prototype. This again was an early decision in the design process. However, it is clear that much improvement could be made to increase the aesthetics in the website of the current implementation. The tools used to develop the website, though, are standard web systems, and it would not take a professional developer too long to create a more aesthetic implementation from the

Critical Design Sheet (CDS) Survey

*Required

Email address *

Name: *

1- I think that I would like to use this system frequently. *

Mark only one choice.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

2- I found the system too complex. *

Mark only one choice.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

3- I thought the system was easy to use. *

Mark only one choice.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

4- I think that I would need the support of an expert person to be able to use this system. *

Mark only one choice.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

5- I found the various sections in this system were well integrated. *

Mark only one choice.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

6- I thought there was too much inconsistency in this system. *

Mark only one choice.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

7- I would imagine that most people would learn to use this system very quickly. *

Mark only one choice.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

8- I found the system very cumbersome and clumsy to use. *

Mark only one choice.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

9- I felt very confident using the system. *

Mark only one choice.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

10- I needed to learn many things before I could get going with this system. *

Mark only one choice.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

Figure 6.10: SUS usability test questionnaire to evaluate CDS online tool.

current one that is presented in this dissertation. The implementation also demonstrates a successful structure. Referring back to Figure 6.1 where we show the five step process.

Indeed these steps seem to work well. Considering anecdotal evidence of observing the participants who evaluated the tool, it was clear that they understood what they needed to do and these stages work well. There are other ways to implement such systems. Another popular way could be to create an App. It would be possible to adapt the current prototype into a system that would work on a mobile tablet or phone. This App could be beneficial to users, and it would be more portable and perhaps users could make critical judgements of visualisations when away from a fixed computer.

The evaluation of the online CDS tool also has its limitations. The use of the SUS is a convenient way to evaluate an interactive tool. And certainly, the results of the SUS are promising; placing the CDS online tool as having “good to excellent usability”. However, we acknowledge that only 7 participants were used for this evaluation. While the SUS has been successfully applied to a small cohort of participants, it still remains that this is a small value. What would be better is to put the CDS website into general use, and use it in a bigger, long-term trial. Perhaps, like the CDS sheet, putting the CDS-online tool into the classroom and get students to use it. This would enable a much larger cohort of people to use the tool, and a much larger set of people to evaluate it. That said, the goal of this work was to investigate and focus on the feasibility of the implementation, and not to create something that can be used by the public without further development. In fact, this feasibility has been very successful. We have created a tool that works, that considers the main stages of how a user would interact with the tool, and investigate what buttons and interactions are required. We have developed database tables to store the data, and have developed a way for users to re-load their results, as well as visualising the results of their critiques. We also highlight that much of this dissertation has focused on creating the CDS in the first place, and that that evaluation is substantial, see Chapter 5. Whereby that evaluation demonstrates that the CDS Sheet is useful, and we have used it in a real-life scenario, by real end-users. Unfortunately, a wider study, for the CDS online tool, is outside the scope of this dissertation, and would take too long to complete within the time given. We can tackle these challenges in future work and find many opportunities to enhance the CDS website in future.

6.6 Summary

This chapter presented the development of the CDS online system. It presented how the tool was created and described the different technologies that were used to construct it. While there could be different ways to implement the tool (such as using website-building software like WordPress, or creating an App) the use of Web technologies have enabled a prototype tool to be created. The developed software could be readily taken by a developer, enhanced, adapted and made into a more commercial system. The compatibility of web techniques offered a great opportunity to facilitate a system where the information can be easily stored and retrieved, using a remote database. Also the chapter described the evaluation of CDS website using a System Usability Scale and the results were positive, demonstrating that the participants felt that the system was usable; that users had a “good” level of satisfaction with the system. Finally, the chapter discussed some of the limitations with this work, and started to discuss ideas of future work. Indeed, the next chapter concludes the dissertation by drawing all the ideas and concepts together, and discussing, in more depth, some of the successes and limitations of the research.

Chapter 7

Conclusions and Future work

In the introduction we wrote that “it is difficult to be open-minded and unbiased, it is challenge to know how to organise the information appropriately to make the right decisions.” The vision of the dissertation was to investigate structures that would help users to make better critical judgements. We wrote the following vision:

Therefore the vision of the research is to investigate critical analysis, and develop something that will help learners structure their critical analysis.

In fact, the hypothesis was written to explore this idea of providing a structure for people to follow. The hypothesis was written as follows:

Critical analysis is difficult. Users benefit from following structure. A structured approach, with a clear set of individual processes, will not only provide the user with a structure, but will enable their (often) subjective decisions to be divided into a specific set of objective criteria.

In summary, and from a holistic point of view, we have achieved this hypothesis and have created an objective structure that does help users better structure their thoughts. The dissertation has drawn from prior research, and investigated other peoples’ research in the topic. The research focused on Critical thinking concepts and how can be applied on every part of visual design process. Our solution was to move from thinking about one big problem, to think about individual parts. So, by applying critical thinking on every part of problem helps the user to better reflect, and find better solutions to their problems.

At the start of the dissertation six objectives were proposed. These range from looking at the literature, to evaluating the final tool. It is useful to list again the research objectives that mentioned in the introduction chapter (Chapter 1 with the objectives listed on Page 6), and investigate whether they have been met.

Obj 1 Explore the breadth of the subject of “critical thinking” and “critical analysis”. Understand where it is used, and the concepts that are being discussed in the literature. Especially investigate the principles of critical thinking in regard to teaching and learning. Including how learners make critical decisions on their work.

The objective was achieved and discussed in Chapter 3. The work explored critical thinking and critical analysis principles and concepts. Specifically, we have learnt different concepts and lessons in academic teaching and learning. For example, students need to use critical thinking in their education; especially in Computer Science students need to create a program and then think about their creation. This helps them reflect on their assignments. The well-known Bloom’s taxonomy is a useful structure enabling students to understand different skills necessary in higher education. Critical thinking was also explored in the workshop that was held, and explained in Chapter 4. Definitions of critical analysis were created by the participants, and the words that were analysed were added the First Impression words of the Critical Design Sheet (CDS).

Obj 2 Investigate the role of metrics in visualisation and how the metrics are used in data-visualisation research.

This objective has been achieved; it is described in the literature review chapter 2. A framework (and a diagram) were created after much deliberation and idea iteration. The framework then was used to organise the many visualisation metrics in the literature, including quantitative metrics and heuristic metrics.

The metrics framework model allows users to explore where their metrics will be used, in comparison to other metrics. The proposed metric model gives us insightful thought into different visualisation metrics and how they are used within

the visualisation design process. The chapter also presented a broad picture of the importance of metrics.

- Obj 3 Discover how people perform a critical analysis. What methods to people employ? Perform quantitative/scientific statistical analysis of people's critical thinking.

The dissertation has met with that aim by conducting a critical analysis workshop to explore how people use their reflections to critique objects, see Chapter 4. The workshop helped us identify users thinking towards critiquing visualisations and found opportunities critical aspects to create critical analysis scheme or structure. Several steps were performed: (1) ideas and strategies of how people perform critical analysis were explored by questioning users at the workshop. (2) Data was collected, and information summarised from the outcome of the workshop. The workshop helped to develop the words for the First Impression, the structure of the overall CDS, and most of the questions in the CDS.

- Obj 4 Design a method(s) that could help developers make better design decisions on their creations and enhance the self reflection cues that add more rigorous thinking to their design analysis. Apply the method with users, in a situation that will require them to use it.

This objective is achieved in Chapter 4. The dissertation has created a Critical Design Sheet method for visualisation. This research first explored different methodologies to achieve the goal. We first used the workshop to create the underpinning data that then helped to organise the parts of the CDS. We carefully analysed workshop outcome, and discovered different definitions and the divide and conquer approach. In the workshop we looked at different domains and investigated how users critique different products and images. Inspiration came from many sources, especially we looked to the SUS and UEQ evaluation systems. From these we discussed alternative design ideas, and eventually created the Critical Design Sheet. The CDS was developed through an iterative process, where the required concepts were developed and then assessed and evaluated by users in several talk-aloud sessions. The final CDS divides the analysis of visual design principles into six categories user's perception, environment, Interface,

components, design and visual marks in order to cover multiple considerations on visualisation design attributes.

Obj 5 Analyse the designed solution for its *usability* through people using the method; Enhance the design as required through iterative design processes;

The dissertation has met this objective by testing the usability of CDS, see Chapter 5. We evaluated the CDS with samples of students by conducting two usability studies. Empirically, we have successfully analysed the results qualitatively as evaluation users practices with CDS potentials to evaluate users reflections, and statistically we analysed CDS results which the results confirmed our assumption of reliability in getting promising high values of Cronbach's alpha measure of internal consistency was applied to the results, which shows that the internal questionnaire items are well correlated to give closely rating values between users groups. Additionally, we used CDS in teaching to see how the CDS could be used in a real situation, with real students. We have received positive comments about the use of CDS as tool to help self-reflection, and help people reflect on their designs for their assessment. These results all support the fact that the CDS is a useful tool. Surprisingly, students' positive feedback and supportive results have encouraged us to expand and put the method into future assessments. CDS can be used for many different types of visualisation tools, systems of output. So we have tried to keep the terminology general and inclusive.

Obj 6 Implement the solution into something that can be used by a wide range of people, to share and discuss the quality of different visualisations. Perform an usability analysis of this tool.

To achieve this objective we created the CDS online version, see Chapter 6. Web technology offers various benefits to serve easy spreading and sharing information for wide range of users. The created computer-based version of CDS in fact supports (i) the system to be accessible and applicable for broad sector of learners, (ii) users can create account to be enable store their critique work and retrieve them securely, and (iii) the website can become a resource for *good* or *bad*

visualisations and users have the potential to review their works any time. We are successfully evaluate the CDS online version using SUS usability system with score value 78.75 which is explained to be have “good” and suitable usability.

7.1 Limitations

There are different challenges that this research has faced, and while it is clear that there are many successes, there are also limitations with the research. Broadly, we follow the order of the dissertation.

Metrics. In the related work chapter we discussed many different metrics. While the review is comprehensive and includes hundreds of references, there would always be metrics that could be discovered and added to the review. One of the challenges, however, is that it is clear that these metrics have been created by different authors for a variety of purposes. Consequently they have been published in different journals, conferences and locations. Also, many of the publications are unclear how the metrics are created. In other words, the papers do not go into enough detail to understand how the metric is calculated and how they are used. This situation is changing, and hopefully it will become easier to understand how these algorithms are created and work. This is because many researchers are calling for researchers to publish their data, algorithms and techniques in a way that would be more easily taken and used by other researchers. In fact, research funders are starting to make it a requirement of holding the grant money that the data and algorithms created with the money are also published.

Design of the CDS. The dissertation has created a successful CDS design. We have evaluated the design and have evaluated that it is successful. However there are always adaptations and improvements that could be made. One of the challenges with the current CDS is that it is difficult to do quickly. It takes someone about five minutes to perform a CDS. While this is not too long, and indeed it is time well spent, because it is beneficial to reflect on the design of a visualisation; so actually the five minute direction is not too long. But, a quicker version could be

possible. One way to achieve this would be to use the “killer questions” on the CDS. This would mean that the first question of each of the six categories was used. Users would lose out on doing an in-depth evaluation; but would gain a quicker evaluation. They would be losing quality, but gaining more time. Also, the star-plot takes time to complete; so here, it could be possible to remove this plot, and just use the values in the Likert scale as a guide to what is critiqued to be good or bad. However, the disadvantage in making all this change, is that the person is not putting so much critical thought into their evaluation. In one respect, it is more important that the user takes time to reflect on their design (and the CDS really forces that user to spend more time on the evaluation). The CDS provides the excuse to spend time evaluating the work. Consequently, by removing all these detailed elements, and making the CDS simpler, the user would lose out on thinking about the designs in a deep way.

Another question that could be asked of the CDS is: “is the CDS a general structure that could be used for any visualisation?” The dissertation did not answer this question. However, the goal, and the setup of the workshop was designed to make the CDS a generalisable structure. In fact, the evaluation did use several different scenarios across a variety of domains and visualisation styles. Therefore the design of the CDS was in fact done in a way to make it general. However to answer this question fully a long-term study would be required that tackled a range of visualisation types. Unfortunately a long-term evaluation is impossible to achieve within the duration of a PhD. When we look at the applications that the students performed, then they were applied to many animations, interactive tools, and used by 35 students across the full ICP3036 cohort. It was unfortunate that we only captured data of 20 of those students, because only 20 students fully completed the initial training and evaluation exercises. But, that said, 20 students is still a substantial user group. Therefore we conclude that the CDS is suitable for a wide range of evaluation types.

Usability evaluation. The usability evaluation demonstrates that the CDS was usable.

In fact, the iterative design strategy has offered insight into what users think about the CDS. It was excellent to have a wide range of users in the development of the CDS, from the initial 10 workshop participants, to the talk-aloud group, and

then to the 10 PhD students (group1) and 20 undergraduate students (group2). Therefore overall 30 “students” (at both undergraduate and postgraduate levels) used the tool. This is a substantial quantity of participants. However, how many participants is enough to make the evaluation successful? It is clear that there could always be more people to evaluate a tool, but with 30 we have had a successful result. As mentioned above, it would be beneficial to make a long-term study. It is good that the CDS will be used in the course next year, and the technique has already been used by a third-year project student in their project. Where he said, “it was useful to apply the technique to my project, and help me structure my thoughts”.

Implementation. It is excellent to have an online version of the CDS. There are many excellent aspects of the online CDS tool. It definitely is usable (as shown in the evaluation) but there is much functionality that could be improved with this prototype. The aim of this work was to create a prototype, it was not to deliver a fully functional system. It has achieved its goals. However, the CDS website could be better implemented, more interaction added, using for instance, modern JavaScript’s libraries even semantic bootstrap components to enhance the presentation of webpages. We have spent long time in the CDS design process and that effect on the CDS implementation side. Moreover, there are limitations to the usability evaluation; it was achieved by only 7 participants, which while demonstrates that the tool is usable, it would have been good to get more users in the evaluation. But even with 7 participants it is clear that the idea works, and the results have highlighted where improvements could be made, such as improving the login and retrieval, and the visualisation of the results could be improved and look better.

7.2 Future Work

There is much future work that can be achieved. To investigate how the ideas can be taken forward, we consider (1) how the current work could be improved to help the user

perform the CDS more efficiently and better, and (2) how the work could be done by the computer, rather than doing it by hand.

Help the user perform the CDS better. If we think about products created by manufacturing companies (such as crisps, chocolate bars, fridges or even computers), then they are delivered and used by the community and then adapted. Such consumer products are then improved, they may be downsized, adapted for a new market etc. We can consider the CDS to be a product, and therefore we can consider how to change and improve the CDS in the future. So what does adapting, improving or downsizing mean for the CDS. We can think about downsizing, where the CDS is made shorter. In fact, one of the challenges (as explained in the section on Limitations, Sec 7.1) is that users take about 5 minutes to complete the form. In the future it could be possible to change the CDS such that it is much quicker to use. One way to achieve this could be to remove most of the 30 questions and make them only 6 (i.e., use the killer questions). However it may be better to adapt the 30 questions and change them to a new lighter version of questions. There are obviously disadvantages to a quicker question sheet (as discussed above) but re-thinking the questions could help the users perform the critical analysis quicker. Potentially the CDS could be used for other critical evaluation tasks. So the question we could ask is “could it be applied to the design of other objects?” Indeed, it could be possible to adapt it and apply it to critique other objects. Perhaps some of the terminology would need to be adapted; for instance, we could apply it to building a Lego model. Someone could think if the design is suitable for the user, or whether it looks pleasing, etc. Certainly some of the categories would need to be interpreted, but theoretically it could be possible to apply the CDS to different fields.

The online version is definitely something else that could be developed. Changes could be made to the website to improve the webpages aesthetics and add more interaction to make the tool more effective, and perhaps run on mobile and tablet screen sizes. The long term vision here could be to create a CDS tool repository to enable users share their work and show examples of good and bad visualisations. It could be a place where developers and learners meet, and discuss and share ideas and different designs, such to create a developers community.

Make the computer do the CDS. Rather than getting the user to critique the design, a long term goal is to get the computer to perform the evaluation. To move the CDS to be an automated critiquing system. In fact, potentially the same structure could be used, where the computer uses different metrics to evaluate the parts. Indeed we are now a complete circle. We started off the dissertation thinking about metrics and how to evaluate the designs, and now we are thinking how the ideas could be automated using metrics. Generally, there are several possibilities that could be added, such as to develop a querying system to mimic the human response of the gut instinct, or first impression. Perhaps a range of metrics would need to be created to achieve this. This could be achieved using computer-vision algorithms; perhaps they could learn detect what is bad or good parts on design, using such as automatic feedback system to generate critique report.

7.3 Conclusion

We started our research with the vision “our goal is to create structure for critiquing helps learners to use and structure their critical analysis”. Much effort has been applied to achieve this goal, starting with the investigation of critical thinking, human-centric visualisation systems and evaluation systems. We believe that we achieved a structure that suitably helps users to make self reflection on their visual design. The CDS has been evaluated and proven to achieve a successful usability score value in terms of reliability and learnability, which give us opportunity to develop and promote the method in future work. Finally, the CDS has been used in a real situation, with real students, which evidenced that users benefited from using the CDS structure, and that it helped them to reflect on their designs and improve their design solutions.

References

- [1] R. L. Ackoff, 'From data to wisdom', *Journal of Applied Systems Analysis*, vol. 16, no. 1, pp. 3–9, 1989 (p. 55).
- [2] M. Alam, S. G. Kobourov and S. Veeramoni, 'Quantitative measures for cartogram generation techniques', in *Computer Graphics Forum (CGF)*, Wiley Online Library, vol. 34, 2015, pp. 351–360 (pp. 23, 33).
- [3] G. Albuquerque, M. Eisemann, D. J. Lehmann, H. Theisel and M. Magnor, 'Improving the visual analysis of high-dimensional datasets using quality measures', in *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, IEEE, 2010, pp. 19–26 (p. 34).
- [4] R. Amar and J. Stasko, 'A knowledge task-based framework for design and evaluation of information visualizations', in *Proceedings of the 2004 IEEE Symposium on Information Visualization*, 2004, pp. 143–150. doi: 10.1109/INFVIS.2004.10 (p. 40).
- [5] L. W. Anderson, D. R. Krathwohl, P. Airasian, K. Cruikshank, R. Mayer, P. Pintrich, J. Raths and M. Wittrock, 'A taxonomy for learning, teaching and assessing: A revision of bloom's taxonomy: Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving in small groups.', *New York. Longman Publishing. Artz, AF, & Armour-Thomas, E.(1992). Cognition and Instruction*, vol. 9, no. 2, pp. 137–175, 2001 (p. 50).
- [6] K. Andrews, 'Evaluating information visualisations', in *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, ser. BELIV '06, Venice, Italy: ACM, 2006, pp. 1–5, ISBN: 1-59593-562-2. doi: 10.1145/1168149.1168151 (p. 40).
- [7] M. Ankerst, S. Berchtold and D. A. Keim, 'Similarity clustering of dimensions for an enhanced visualization of multidimensional data', in *Information Visualization, 1998. Proceedings. IEEE Symposium on*, IEEE, 1998, pp. 52–60 (pp. 31, 33).

- [8] F. J. Anscombe, 'Graphs in statistical analysis', *The American Statistician*, vol. 27, no. 1, pp. 17–21, 1973 (p. 103).
- [9] N. Athanassiou, J. M. McNett and C. Harvey, 'Critical thinking in the management classroom: Bloom's taxonomy as a learning tool', *Journal of Management Education*, vol. 27, no. 5, pp. 533–555, 2003 (p. 49).
- [10] S. Baccianella, A. Esuli and F. Sebastiani, 'Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining.', in *LREC*, vol. 10, 2010, pp. 2200–2204 (p. 101).
- [11] T. Bakas and V. Champion, 'Development and psychometric testing of the bakas caregiving outcomes scale', *Nursing research*, vol. 48, no. 5, pp. 250–259, 1999 (p. 117).
- [12] A. Bangor, P. T. Kortum and J. T. Miller, 'An empirical evaluation of the system usability scale', *Intl. Journal of Human-Computer Interaction (HCI)*, vol. 24, no. 6, pp. 574–594, 2008 (pp. 88, 165).
- [13] A. Bangor, P. Kortum and J. Miller, 'Determining what individual sus scores mean: Adding an adjective rating scale', *Journal of usability studies*, vol. 4, no. 3, pp. 114–123, 2009 (pp. 88, 165–167).
- [14] C. Bazerman and P. Prior, Eds., *What writing does and how it does it: An introduction to analyzing texts and textual practices*. Routledge, 2003 (p. 48).
- [15] R. Benbunan-Fich, 'Using protocol analysis to evaluate the usability of a commercial web site', *Information & Management*, vol. 39, no. 2, pp. 151–163, 2001 (p. 96).
- [16] J. Bertin, *Semiology of graphics*. University of Wisconsin Press, 1983 (pp. 19, 40, 93).
- [17] E. Bertini, C. Plaisant and G. Santucci, 'Beliv'06: Beyond time and errors; novel evaluation methods for information visualization', *interactions*, vol. 14, no. 3, pp. 59–60, May 2007, ISSN: 1072-5520. DOI: 10.1145/1242421.1242460 (pp. 36, 114).
- [18] E. Bertini and G. Santucci, 'Quality metrics for 2d scatterplot graphics: Automatically reducing visual clutter', in *International Symposium on Smart Graphics*, Springer, 2004, pp. 77–89 (pp. 21, 32).
- [19] E. Bertini, A. Tatu and D. Keim, 'Quality metrics in high-dimensional data visualization: An overview and systematization', *IEEE Transactions on*

- Visualization and Computer Graphics (TVCG)*, vol. 17, no. 12, pp. 2203–2212, 2011 (p. 21).
- [20] B. S. Bloom, C. of College and U. Examiners, *Taxonomy of educational objectives*. Longmans, Green New York, 1964, vol. 2 (pp. 49, 50).
 - [21] M. A. Boden, ‘Creativity in a nutshell’, *Think*, vol. 5, no. 15, pp. 83–96, 2007 (p. 52).
 - [22] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva and H. Pfister, ‘What makes a visualization memorable?’, *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 19, no. 12, pp. 2306–2315, 2013 (pp. 20, 93).
 - [23] M. Borkin, K. Gajos, A. Peters, D. Mitsouras, S. Melchionna, F. Rybicki, C. Feldman and H. Pfister, ‘Evaluation of artery visualizations for heart disease diagnosis’, *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2479–2488, 2011 (p. 4).
 - [24] J. Boy, R. A. Rensink, E. Bertini and J.-D. Fekete, ‘A principled way of assessing visualization literacy’, *IEEE transactions on visualization and computer graphics (TVCG)*, vol. 20, no. 12, pp. 1963–1972, 2014 (p. 141).
 - [25] D. Braha and Y. Reich, ‘Topological structures for modeling engineering design processes’, *Research in Engineering Design*, vol. 14, no. 4, pp. 185–199, 2003 (p. 53).
 - [26] R. Brath, ‘Metrics for effective information visualization’, in *Information Visualization, INFOVIS 1997. Proceedings., IEEE Symposium on*, IEEE, 1997, pp. 108–111 (pp. 20, 32).
 - [27] V. Braun and V. Clarke, ‘Using thematic analysis in psychology’, *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006 (p. 79).
 - [28] J. E. Bresenham, ‘Algorithm for computer control of a digital plotter’, *IBM Systems journal*, vol. 4, no. 1, pp. 25–30, 1965 (p. 142).
 - [29] J. Brooke, ‘SUS – a quick and dirty usability scale’, *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996 (pp. 34, 87, 165).
 - [30] A. P. Carnevale *et al.*, ‘Workplace basics: The skills employers want.’, *Training and Development Journal*, vol. 42, no. 10, pp. 22–26, 1988 (p. 1).
 - [31] A. P. Carnevale and N. Smith, ‘Workplace basics: The skills employees need and employers want’, *Human Resource Development International*, vol. 16, no. 5, pp. 491–501, 2013. DOI: 10.1080/13678868.2013.821267 (p. 1).

- [32] H. Carr, D. Brian and D. Brian, 'On histograms and isosurface statistics', *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 12, no. 5, 2006 (p. 33).
- [33] J. M. Carroll, R. L. Mack, C. H. Lewis, N. L. Grischkowsky and S. R. Robertson, 'Exploring exploring a word processor', *Human Computer Interaction*, vol. 1, no. 3, pp. 283–307, 1985 (p. 94).
- [34] C. Chen and M. P. Czerwinski, 'Empirical evaluation of information visualizations: An introduction', *International Journal of Human-Computer Studies*, vol. 53, no. 5, pp. 631–635, 2000. doi: doi.org/10.1006/ijhc.2000.0421 (pp. 37, 40, 114).
- [35] M. Chen, D. Ebert, H. Hagen, R. S. Laramée, R. Van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann and D. Silver, 'Data, information, and knowledge in visualisation', *IEEE Computer Graphics and Applications (CG& Application)*, vol. 29, no. 1, 2009 (pp. 55, 61).
- [36] M. Chen and L. Floridi, 'An analysis of information visualisation', *Synthese*, vol. 190, no. 16, pp. 3421–3438, 2013 (pp. 58, 59).
- [37] W. S. Cleveland, 'A model for studying display methods of statistical graphics', *Journal of Computational and Graphical Statistics*, vol. 2, no. 4, pp. 323–343, 1993. doi: 10.1080/10618600.1993.10474616 (p. 40).
- [38] H. E. Cline, W. E. Lorensen, S. Ludke, C. R. Crawford and B. C. Teeter, 'Two algorithms for the three-dimensional reconstruction of tomograms', *Medical physics*, vol. 15, no. 3, pp. 320–327, 1988 (p. 142).
- [39] S. Conversy, S. Chatty and C. Hurter, 'Visual scanning as a reference framework for interactive representation design', *Information Visualization*, vol. 10, no. 3, pp. 196–211, 2011 (p. 40).
- [40] K. A. Cook and J. J. Thomas, 'Illuminating the path: The research and development agenda for visual analytics', 2005 (p. 37).
- [41] D. Cyr, M. Head and H. Larios, 'Colour appeal in website design within and across cultures: A multi-method evaluation', *International journal of Human-Computer Studies*, vol. 68, no. 1, pp. 1–21, 2010 (p. 117).
- [42] T. N. Dang and L. Wilkinson, 'Transforming scagnostics to reveal hidden features', *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 20, no. 12, pp. 1624–1632, 2014 (p. 22).

- [43] A. Dasgupta and R. Kosara, ‘Pargnostics: Screen-space metrics for parallel coordinates’, *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 16, no. 6, pp. 1017–1026, 2010 (pp. 22, 33).
- [44] J. Dewey, *How we think*. D. C. Heath & Co., Boston, New York, Chicago, 1910 (p. 71).
- [45] J. Dowell and J. Long, ‘Target paper: Conception of the cognitive engineering design problem’, *Ergonomics*, vol. 41, no. 2, pp. 126–139, 1998 (p. 18).
- [46] B. Duffy, H. Carr and T. Möller, ‘Integrating isosurface statistics and histograms’, *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 19, no. 2, pp. 263–277, 2013 (pp. 23, 33).
- [47] B. Duffy, A. Dasgupta, R. Kosara, S. Walton and M. Chen, ‘Measuring visual complexity of cluster-based visualisations’, *arXiv preprint arXiv:1302.5824*, 2013 (p. 22).
- [48] K. Duncker and L. S. Lees, ‘On problem-solving.’, *Psychological monographs*, vol. 58, no. 5, p. i, 1945 (pp. 71, 74).
- [49] L. Elder and R. Paul, *Analytic thinking: How to take thinking apart and what to look for when you do*. Dillon Beach. CA: Foundation for Critical Thinking, 2003 (p. 50).
- [50] G. Ellis and A. Dix, ‘Enabling automatic clutter reduction in parallel coordinate plots’, *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 12, no. 5, pp. 717–724, 2006 (p. 33).
- [51] R. H. Ennis, ‘A concept of critical thinking.’, *Harvard educational review*, 1962 (pp. 46, 84).
- [52] K. A. Ericsson and H. A. Simon, *Protocol analysis*. MIT press Cambridge, MA, 1993 (p. 94).
- [53] P. A. Facione, ‘Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. research findings and recommendations.’, *ERIC*, vol. ED315423, 1990 (pp. 2, 3, 46, 47, 58, 62, 74, 84).
- [54] A. Feizi and C. Y. Wong, ‘Usability of user interface styles for learning a graphical software application’, in *Computer & Information Science (ICCIS), 2012 International Conference on*, IEEE, vol. 2, 2012, pp. 1089–1094 (p. 117).

- [55] G. Fischer, R. McCall and A. Morch, 'Design environments for constructive and argumentative design', *ACM SIGCHI Bulletin*, vol. 20, no. SI, pp. 269–275, 1989 (p. 38).
- [56] G. Fischer, K. Nakakoji, J. Ostwald, G. Stahl and T. Sumner, 'Embedding critics in design environments', *The knowledge engineering review*, vol. 8, no. 4, pp. 285–307, 1993 (p. 38).
- [57] J. H. Flavell, 'Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry.', *American psychologist*, vol. 34, no. 10, p. 906, 1979 (p. 74).
- [58] C. Flavián, M. Guinalíu and R. Gurrea, 'The role played by perceived usability, satisfaction and consumer trust on website loyalty', *Information & Management*, vol. 43, no. 1, pp. 1–14, Jan. 2006, ISSN: 0378-7206. DOI: 10.1016/j.im.2005.01.002 (p. 113).
- [59] M. Forehand, 'Bloom's taxonomy', *Emerging perspectives on learning, teaching, and technology*, vol. 41, p. 47, 2010 (p. 50).
- [60] C. M. Freitas, P. R. Luzzardi, R. A. Cava, M. Winckler, M. S. Pimenta and L. P. Nedel, 'On evaluating information visualization techniques', in *Proceedings of the working conference on Advanced Visual Interfaces (AVI)*, ACM, 2002, pp. 373–374 (p. 38).
- [61] M. Friendly, 'Re-visions of minard', *Statistical Computing & Statistical Graphics Newsletter*, vol. 11, no. 1, p. 1, 2000 (p. 76).
- [62] T. Fritz and G. C. Murphy, 'Using information fragments to answer the questions developers ask', in *Proceedings of the 32Nd ACM/IEEE International Conference on Software Engineering - Volume 1*, ser. ICSE '10, Cape Town, South Africa: ACM, 2010, pp. 175–184, ISBN: 978-1-60558-719-6. DOI: 10.1145/1806799.1806828 (p. 67).
- [63] K. J. Gilhooly, *Thinking – Directed, Undirected and Creative*. Academic Press, 1996 (p. 24).
- [64] R. L. Glass, I. Vessey and S. A. Conger, 'Software tasks: Intellectual or clerical?', *Information & Management*, vol. 23, no. 4, pp. 183–191, 1992 (p. 94).
- [65] S. Greenberg and B. Buxton, 'Usability evaluation considered harmful (some of the time)', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 2008, pp. 111–120 (p. 116).

- [66] R. B. Haber and D. A. McNabb, 'Visualisation idioms: A conceptual model for scientific visualisation systems', *Visualisation in Scientific Computing*, vol. 74, p. 93, 1990 (p. 58).
- [67] D. F. Halpern, 'Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring.', *American psychologist*, vol. 53, no. 4, p. 449, 1998 (p. 51).
- [68] ———, *Thought and knowledge: An introduction to critical thinking*. Routledge, 2002 (p. 47).
- [69] J. H. Hayes, S. C. Patel and L. Zhao, 'A metrics-based software maintenance effort model', in *2011 15th European Conference on Software Maintenance and Reengineering*, IEEE Computer Society, 2004, pp. 254–254 (p. 33).
- [70] K. C. Hendy, K. M. Hamilton and L. N. Landry, 'Measuring subjective workload: When is one scale better than many?', *Human Factors*, vol. 35, no. 4, pp. 579–601, 1993 (p. 42).
- [71] A. Holzinger, 'Usability engineering methods for software developers', *Communications of the ACM*, vol. 48, no. 1, pp. 71–74, 2005 (p. 115).
- [72] W. Huang, P. Eades and S.-H. Hong, 'Measuring effectiveness of graph visualizations: A cognitive load perspective', *Information Visualization*, vol. 8, no. 3, pp. 139–152, 2009 (p. 24).
- [73] P. Isenberg, A. Tang and S. Carpendale, 'An exploratory study of visual information analysis', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2008, pp. 1217–1226 (p. 126).
- [74] P. Isenberg, T. Zuk, C. Collins and S. Carpendale, 'Grounded evaluation of information visualizations', in *Proceedings of the 2008 Workshop on 'BELIV' BEyond time and errors: novel evaluation methods for Information Visualisation*, ACM, 2008, p. 6 (p. 35).
- [75] T. Isenberg, P. Isenberg, J. Chen, M. Sedlmair and T. Möller, 'A systematic review on the practice of evaluating visualization', *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 19, no. 12, pp. 2818–2827, Dec. 2013, ISSN: 1077-2626. DOI: 10.1109/TVCG.2013.126 (pp. 37, 40, 41, 114).
- [76] B. Jack and A. M. Clarke, 'The purpose and use of questionnaires in research.', *Professional nurse (London, England)*, vol. 14, no. 3, pp. 176–179, 1998 (p. 117).

- [77] 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 *Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors-Novel Evaluation Methods for Visualisation*, ACM, 2012, p. 4 (pp. 40, 71).
- [78] J. Jeng, 'Usability assessment of academic digital libraries: Effectiveness, efficiency, satisfaction, and learnability', *Libri*, vol. 55, no. 2-3, pp. 96–121, 2005 (pp. 114, 117).
- [79] J. Johansson and M. Cooper, 'A screen space quality method for data abstraction', in *Computer Graphics Forum (CGF)*, Wiley Online Library, vol. 27, 2008, pp. 1039–1046 (p. 33).
- [80] S. Johansson and J. Johansson, 'Interactive dimensionality reduction through user-defined combinations of quality metrics', *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 15, no. 6, pp. 993–1000, 2009 (pp. 22, 33).
- [81] S. Johnson, *Where good ideas come from: The natural history of innovation*. Penguin UK, 2010 (p. 73).
- [82] M. C. Jones and D. W. Johnston, 'Distress, stress and coping in first-year student nurses', *Journal of Advanced Nursing*, vol. 26, no. 3, pp. 475–482, 1997 (p. 117).
- [83] D. A. Keim, F. Mansmann, J. Schneidewind and H. Ziegler, 'Challenges in visual data analysis', in *Information Visualization. INFOVIS 2006. Tenth International Conference on*, IEEE, 2006, pp. 9–16 (p. 37).
- [84] R. Kengeri, C. D. Seals, H. D. Harley, H. P. Reddy and E. A. Fox, 'Usability study of digital libraries: Acm, ieee-cs, ncstrl, ndltd', *International Journal on Digital Libraries*, vol. 2, no. 2, pp. 157–169, 1999 (p. 117).
- [85] L. Kimbell, 'Rethinking design thinking: Part i', *Design and Culture*, vol. 3, pp. 285–306, 2011. DOI: 10.2752/175470811X13071166525216 (p. 53).
- [86] G. Klein, B. Moon and R. Hoffman, 'Making sense of sensemaking 2: A macrocognitive model', *IEEE Intelligent Systems*, vol. 21, no. 5, pp. 88–92, Sep. 2006, ISSN: 1541-1672. DOI: 10.1109/MIS.2006.100 (p. 55).
- [87] P. Kline, *Handbook of psychological testing*. Routledge, 2013 (p. 131).
- [88] O. Konstantina, T. Nikolaos and K. Christos, 'Perceived usability evaluation of learning management systems: Empirical evaluation of the system usability

- scale', *The International Review of Research in Open and Distributed Learning*, vol. 16, no. 2, 2015, ISSN: 1492-3831 (p. 165).
- [89] R. Kosara, 'Visualization criticism-the missing link between information visualization and art', in *Information Visualization, 2007. IV'07. 11th International Conference*, IEEE, 2007, pp. 631–636 (pp. 39, 71).
 - [90] R. Kosara, F. Drury, L. E. Holmquist and D. H. Laidlaw, 'Visualization criticism', *IEEE Computer Graphics and Applications (CG& Application)*, vol. 28, no. 3, 2008 (pp. 39, 71).
 - [91] D. R. Krathwohl, 'A revision of bloom's taxonomy: An overview', *Theory Into Practice*, vol. 41, no. 4, pp. 212–218, 2002, ISSN: 00405841, 15430421 (p. 50).
 - [92] H. Lam, E. Bertini, P. Isenberg, C. Plaisant and S. Carpendale, 'Empirical studies in information visualization: Seven scenarios', *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 18, no. 9, pp. 1520–1536, Sep. 2012, ISSN: 1077-2626. DOI: 10.1109/TVCG.2011.279 (pp. 37, 40, 41, 114).
 - [93] B. Laugwitz, T. Held and M. Schrepp, 'Construction and evaluation of a user experience questionnaire', in *Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society on HCI and Usability for Education and Work*, ser. USAB '08, Graz, Austria: Springer-Verlag, 2008, pp. 63–76, ISBN: 978-3-540-89349-3. DOI: 10.1007/978-3-540-89350-9_6 (pp. 34, 89, 90).
 - [94] A. C. Lemke and G. Fischer, 'A cooperative problem solving system for user interface design', in *Proceedings of the Eighth National Conference on Artificial Intelligence - Volume 1*, ser. AAAI'90, Boston, Massachusetts: AAAI Press, 1990, pp. 479–484, ISBN: 0-262-51057-X (p. 38).
 - [95] G. L. Lohse, 'A cognitive model for understanding graphical perception', *Human-Computer Interaction (HCI)*, vol. 8, no. 4, pp. 353–388, 1993 (p. 40).
 - [96] J. Lohse, H. Rueter, K. Biolsi and N. Walker, 'Classifying visual knowledge representations: A foundation for visualization research', in *Visualization, 1990. Visualization'90., Proceedings of the First IEEE Conference on*, IEEE, 1990, pp. 131–138 (p. 40).
 - [97] K. Luther, A. Pavel, W. Wu, J.-I. Tolentino, M. Agrawala, B. Hartmann and S. P. Dow, 'Crowdcrit: Crowdsourcing and aggregating visual design critique', in *Proceedings of the companion publication of the 17th ACM conference on*

- Computer Supported Cooperative Work & Social Computing*, ACM, 2014, pp. 21–24 (p. 38).
- [98] K. Luther, J.-L. Tolentino, W. Wu, A. Pavel, B. P. Bailey, M. Agrawala, B. Hartmann and S. P. Dow, ‘Structuring, aggregating, and evaluating crowdsourced design critique’, in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ACM, 2015, pp. 473–485 (p. 38).
 - [99] R. L. Mack, C. H. Lewis and J. M. Carroll, ‘Learning to use word processors: Problems and prospects’, *ACM Transactions on Information Systems (TOIS)*, vol. 1, no. 3, pp. 254–271, 1983 (p. 94).
 - [100] J. R. Martin and P. R. White, *The language of evaluation*. Springer, 2003, vol. 2 (p. 52).
 - [101] J. E. McPeck, *Critical thinking and education*. Routledge, 2016 (pp. 44, 46, 47).
 - [102] R. B. Miller, *Human ease of use criteria and their tradeoffs*. IBM, Systems Development Division, Poughkeepsie Lab, 1971 (p. 115).
 - [103] R. Molich and J. Nielsen, ‘Improving a human-computer dialogue: What designers know about traditional interface design’, *Communications of the ACM*, vol. 33, no. 3, 1990 (pp. 42, 116).
 - [104] T. P. Moran and J. M. Carroll, *Design rationale: Concepts, techniques, and use*. L. Erlbaum Associates Inc., 1996 (p. 53).
 - [105] T. Munzner, ‘A nested model for visualisation design and validation’, *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 15, no. 6, pp. 921–928, Nov. 2009, ISSN: 1077-2626. DOI: 10.1109/TVCG.2009.111 (pp. 36, 40).
 - [106] T. Munzner, ‘Process and pitfalls in writing information visualization research papers’, in *Information Visualization: Human-Centered Issues and Perspectives*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 134–153, ISBN: 978-3-540-70956-5. DOI: 10.1007/978-3-540-70956-5_6 (p. 36).
 - [107] ———, *Visualization Analysis and Design*. A K Peters/CRC Proess, 2014, ISBN: 9781466508910 (pp. 40, 96).
 - [108] J. Nielsen, *Usability engineering*. Elsevier, 1994 (pp. 115, 140).
 - [109] ———, ‘Usability inspection methods’, in *Conference companion on Human Factors in Computing Systems*, ACM, 1995, pp. 377–378 (p. 42).

- [110] J. Nielsen and R. Molich, 'Heuristic evaluation of user interfaces', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 1990, pp. 249–256 (pp. 37, 39, 42, 140).
- [111] R. Nixon, *Learning PHP, MySQL, JavaScript, and CSS: A step-by-step guide to creating dynamic websites*. " O'Reilly Media, Inc.", 2012 (p. 164).
- [112] D. A. Norman, 'Cognitive engineering', *User centered system design*, vol. 31, p. 61, 1986 (p. 17).
- [113] Y. Oh, 'Toward a theory of design critiquing', PhD thesis, Carnegie Mellon University, 2010 (p. 38).
- [114] F. Paas, J. E. Tuovinen, H. Tabbers and P. W. Van Gerven, 'Cognitive load measurement as a means to advance cognitive load theory', *Educational psychologist*, vol. 38, no. 1, pp. 63–71, 2003 (pp. 24, 42).
- [115] J. Passmore, 'On teaching to be critical', in. Routledge and Kegan Paul London, 1967, ch. Chapter 12, The concept of education, pp. 192–212 (p. 46).
- [116] R. Paul and L. Elder, *The miniature guide to critical thinking: concepts and tools*. Foundation for Critical Thinking, 2004 (pp. 2, 4).
- [117] R. Paul and L. Elder, 'Critical thinking', *The Foundation for Critical Thinking*, 2008 (p. 50).
- [118] W. Peng, M. O. Ward and E. A. Rundensteiner, 'Clutter reduction in multi-dimensional data visualization using dimension reordering', in *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, IEEE, 2004, pp. 89–96 (pp. 21, 33).
- [119] C. Plaisant, 'The challenge of information visualization evaluation', in *Proceedings of the Working Conference on Advanced Visual Interfaces*, ser. AVI '04, Gallipoli, Italy: ACM, 2004, pp. 109–116, ISBN: 1-58113-867-9. DOI: 10.1145/989863.989880 (pp. 40, 114).
- [120] G. Pólya, *How to solve it: A new aspect of mathematical method*. Princeton university press, 1973, republished 2014 (pp. 71, 85, 86).
- [121] J. Potter and D. Edwards, 'Discourse analysis', in *Introducing Psychological Research*, Springer, 1996, pp. 419–425 (p. 79).
- [122] A. R. Pritchett and R. J. Hansman, 'Use of testable responses for performance-based measurement of situation awareness', *Situation awareness analysis and measurement*, pp. 189–209, 2000 (p. 31).

- [123] H. C. Purchase, 'Metrics for graph drawing aesthetics', *Journal of Visual Languages & Computing*, vol. 13, no. 5, pp. 501–516, 2002 (pp. 23, 33).
- [124] J. Rattray, M. Johnston and J. Wildsmith, 'The intensive care experience: Development of the ice questionnaire', *Journal of Advanced Nursing*, vol. 47, no. 1, pp. 64–73, 2004 (p. 117).
- [125] I. Rauth, E. Köppen, B. Jobst and C. Meinel, 'Design thinking: An educational model towards creative confidence', in *DS 66-2: Proceedings of the 1st International Conference on Design Creativity (ICDC 2010)*, 2010 (p. 54).
- [126] R. Ravendran, I. MacColl and M. Docherty, 'Usability evaluation of a tag-based interface', *Journal of usability Studies*, vol. 7, no. 4, pp. 143–160, 2012 (pp. 117, 165).
- [127] R. Razzouk and V. Shute, 'What is design thinking and why is it important?', *Review of Educational Research*, vol. 82, no. 3, pp. 330–348, 2012 (pp. 1, 53).
- [128] R. A. Rensink and G. Baldridge, 'The perception of correlation in scatterplots', in *Computer Graphics Forum (CGF)*, Wiley Online Library, vol. 29, 2010, pp. 1203–1210 (p. 33).
- [129] B. Richmond, 'Systems thinking: Critical thinking skills for the 1990s and beyond', *System dynamics review*, vol. 9, no. 2, pp. 113–133, 1993 (p. 50).
- [130] J. C. Roberts, 'State of the art: Coordinated & multiple views in exploratory visualization', in *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on*, IEEE, 2007, pp. 61–71 (p. 30).
- [131] J. C. Roberts, C. Headleand and P. Ritsos, 'Sketching designs using the five design-sheet methodology', *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 22, no. 1, pp. 419–428, Jan. 2016, ISSN: 1077-2626 (pp. 24, 26, 29, 30, 34, 37).
- [132] J. C. Roberts, C. J. Headleand and P. D. Ritsos, *Five Design-Sheets – Creative design and sketching in Computing and Visualization*. SpringerNature, 2017, To appear June 2017, ISBN: 978-3-319-55626-0. DOI: 10.1007/978-3-319-55627-7 (pp. 40, 53, 54, 62, 71, 123).
- [133] B. E. Rogowitz and L. A. Treinish, 'Data visualization: The end of the rainbow', *IEEE spectrum*, vol. 35, no. 12, pp. 52–59, 1998 (p. 4).
- [134] R. Rosenholtz, Y. Li and L. Nakano, 'Measuring visual clutter', *Journal of vision*, vol. 7, no. 2, pp. 17–17, 2007 (p. 23).

- [135] A. J. Rotherham and D. T. Willingham, '21st-century skills', *American Educator*, pp. 17–20, 2010, (Spring) (p. 2).
- [136] J. Rowley, 'The wisdom hierarchy: Representations of the dikw hierarchy', *Journal of Information Science*, vol. 33, no. 2, pp. 163–180, 2007. DOI: 10.1177/0165551506070706 (p. 56).
- [137] J. Rubin and D. Chisnell, *Handbook of usability testing: how to plan, design and conduct effective tests*. John Wiley & Sons, 2008 (p. 115).
- [138] P. Saraiya, C. North and K. Duca, 'An insight-based methodology for evaluating bioinformatics visualizations', *IEEE Transactions on Visualization and Computer Graphics(TVCG)*, vol. 11, no. 4, pp. 443–456, 2005 (pp. 40, 71).
- [139] J. Scholtz, 'Beyond usability: Evaluation aspects of visual analytic environments', in *IEEE Visual Analytics Science and Technology, (VAST) 2006 Symposium On*, IEEE, 2006, pp. 145–150 (pp. 31, 37, 38).
- [140] M. Sedlmair, M. Meyer and T. Munzner, 'Design study methodology: Reflections from the trenches and the stacks', *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 18, no. 12, pp. 2431–2440, Dec. 2012, ISSN: 1077-2626. DOI: 10.1109/TVCG.2012.213 (p. 40).
- [141] J. Selzer, 'Rhetorical analysis: Understanding how texts persuade readers', in *Chapter 10: What writing does and how it does it: An introduction to analyzing texts and textual practices*, C. Bazerman and P. Prior, Eds., Routledge, 2004, pp. 279–307 (p. 49).
- [142] A. Shamim, V. Balakrishnan and M. Tahir, 'Evaluation of opinion visualization techniques', *Information visualisation*, vol. 14, no. 4, pp. 339–358, 2015 (p. 117).
- [143] B. Shneiderman, 'The eyes have it: A task by data type taxonomy for information visualizations', in *Visual Languages, 1996. Proceedings., IEEE Symposium on*, IEEE, 1996, pp. 336–343 (p. 91).
- [144] B. Shneiderman and B. B. Bederson, *The Craft of Information Visualization: Readings and Reflections*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003, ISBN: 1558609156 (p. 17).
- [145] B. Shneiderman and C. Plaisant, 'Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies', in *Proceedings of the 2006 AVI workshop on 'BELIV' BEyond time and errors: novel evaluation methods for information visualization*, ACM, 2006, pp. 1–7 (pp. 35, 36, 39, 114).

- [146] M. Sips, B. Neubert, J. P. Lewis and P. Hanrahan, 'Selecting good views of high-dimensional data using class consistency', in *Computer Graphics Forum (CGF)*, Wiley Online Library, vol. 28, 2009, pp. 831–838 (p. 34).
- [147] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, 'Lexicon-based methods for sentiment analysis', *Computational Linguistics*, vol. 37, no. 2, pp. 267–307, 2011 (p. 101).
- [148] J. C. Tang, 'Findings from observational studies of collaborative work', *International Journal of Man-machine studies*, vol. 34, no. 2, pp. 143–160, 1991 (p. 95).
- [149] A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnork and D. Keim, 'Combining automated analysis and visualisation techniques for effective exploration of high-dimensional data', in *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, IEEE, 2009, pp. 59–66 (pp. 33, 34).
- [150] A. Tatu, P. Bak, E. Bertini, D. Keim and J. Schneidewind, 'Visual quality metrics and human perception: An initial study on 2d projections of large multidimensional data', in *Proceedings of the International Conference on Advanced Visual Interfaces (AVI)*, ACM, 2010, pp. 49–56 (p. 33).
- [151] M. Tavakol and R. Dennick, 'Making sense of cronbach's alpha', *International journal of medical education*, vol. 2, p. 53, 2011 (p. 131).
- [152] R. L. Thomas, *Elements of performance and satisfaction as indicators of the usability of digital spatial interfaces for information-seeking: Implications for ISLA*. University of Southern California, 1998 (p. 117).
- [153] W. Tobler, 'Thirty five years of computer cartograms', *ANNALS of the Association of American Geographers*, vol. 94, no. 1, pp. 58–73, 2004 (p. 23).
- [154] M. Tory and T. Moller, 'Human factors in visualization research', *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 10, no. 1, pp. 72–84, 2004 (pp. 61, 70).
- [155] ———, 'Evaluating visualizations: Do expert reviews work', *IEEE computer graphics and applications (CG& Application)*, vol. 25, no. 5, pp. 8–11, 2005 (pp. 37, 39, 116).
- [156] E. Tufte, *The visual display of quantitative information: Cheshire, ct*, 1983 (pp. 20, 32, 76, 94).

- [157] T. S. Tullis and J. N. Stetson, 'A comparison of questionnaires for assessing website usability', in *Proceedings of the Usability Professionals Association (UPA) 2004 Conference*, Jun. 2004, pp. 7–11 (p. 165).
- [158] S. o. M. S. University of Maine. (). Weather data set., [Online]. Available: <http://gyre.umeoce.maine.edu/buoyhome.php>. (p. 22).
- [159] C. Upson, T. Faulhaber, D. Kamins, D. Laidlaw, D. Schlegel, J. Vroom, R. Gurwitz and A. Van Dam, 'The application visualisation system: A computational environment for scientific visualisation', *IEEE Computer Graphics and Applications (CG&Application)*, vol. 9, no. 4, pp. 30–42, 1989 (p. 58).
- [160] H. Van Oostendorp and S. De Mull, 'Learning by exploration: Thinking aloud while exploring an information system', *Instructional Science*, vol. 27, no. 3-4, pp. 269–284, 1999 (p. 96).
- [161] L. Vaughn, *The Power of Critical Thinking: Effective Reasoning About Ordinary and Extraordinary Claims*. Oxford University Press, 2008 (p. 47).
- [162] I. Vessey and S. Conger, 'Learning to specify information requirements: The relationship between application and methodology', *Journal of Management Information Systems*, vol. 10, no. 2, pp. 177–201, 1993 (p. 94).
- [163] C. Ware, *Information Visualization: Perception for Design*, ser. Interactive Technologies. Elsevier Science, 2004, ISBN: 9780080478494 (p. 64).
- [164] ———, *Information Visualization: Perception for Design*, 3rd ed. Amsterdam: Elsevier (Morgan Kaufmann), 2012, ISBN: 978-0-12-381464-7 (pp. 17, 19, 20, 25, 40, 57, 96).
- [165] C. Ware, H. Purchase, L. Colpoys and M. McGill, 'Cognitive measurements of graph aesthetics', *Information Visualization*, vol. 1, no. 2, pp. 103–110, 2002 (pp. 23, 33).
- [166] S. Wehrend and C. Lewis, 'A problem-oriented classification of visualization techniques', in *Proceedings of the 1st Conference on Visualization'90*, IEEE Computer Society Press, 1990, pp. 139–143 (p. 38).
- [167] L. Wilkinson, A. Anand and R. L. Grossman, 'Graph-theoretic scagnostics.', in *INFOVIS*, vol. 5, 2005, p. 21 (p. 22).
- [168] A. Xu, S.-W. Huang and B. Bailey, 'Voyant: Generating structured feedback on visual designs using a crowd of non-experts', in *Proceedings of the 17th ACM*

- Conference on Computer Supported Cooperative Work & Social Computing*, ACM, 2014, pp. 1433–1444 (p. 38).
- [169] A. Xu, H. Rao, S. P. Dow and B. P. Bailey, ‘A classroom study of using crowd feedback in the iterative design process’, in *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, ACM, 2015, pp. 1637–1648 (p. 38).
- [170] M. X. Zhou and S. K. Feiner, ‘Visual task characterization for automated visual discourse synthesis’, in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, ACM Press/Addison-Wesley Publishing Co., 1998, pp. 392–399 (p. 38).
- [171] T. Zuk, L. Schlesier, P. Neumann, M. S. Hancock and S. Carpendale, ‘Heuristics for information visualization evaluation’, in *Proceedings of the 2006 AVI ‘BELIV’ workshop on BEyond time and errors: novel evaluation methods for information visualization*, ACM, 2006, pp. 1–6 (pp. 39, 116).