Exploring the Representation of Cyber-Risk Data Through Sketching

Thomas Miller* Lancaster University
Miriam Sturdee† University of St Andrews
Daniel Prince‡ Lancaster University

Abstract

Dealing with complex information regarding cyber security risks is increasingly important as attacks rise in frequency. Visualisation techniques are used to support decision-making and insight. However, the use of visualisations across different stakeholders and cyber security risk data is not well explored. This work presents an exploratory study in which participants use sketching to represent cyber-risk data. We critically discuss the method and our results demonstrate the usefulness of the method to identify new, diverse visualisation approaches, as well as the richness of stakeholder visualisation conceptualisation.

Index Terms: Cyber-Risk, Sketching, Visualisation, Insight

1 Introduction

To ensure an efficacious risk management process, stakeholders are required to communicate and consult to reduce risks [6]. Challenges arise when communicating risk across different stakeholders and levels of technical granularity. If risks are not communicated correctly, they can result in an operational shutdown, financial loss, damaged equipment, and even the loss of human life [45].

Cyber risk visualisation is vital for sharing risk information and supporting decision-making among stakeholders [22]. However, stakeholders are not involved in developing industry-standard techniques [14], leading to uncertainty about its suitability for different stakeholders. This uncertainty can lead to poor decision-making and increased cyber risks.

By examining how stakeholders depict cyber-risk data, we can compare their representations with existing visualisation techniques. This allows for a more accurate evaluation of industry's visualisations and identification of shortcomings. These insights help enhance risk communication between technical and non-technical stakeholders. The study presented in this paper utilises sketching as an approach, enabling the rapid collection of spontaneous visual representations from multiple cyber-security stakeholders. Academic studies have previously employed sketching to enhance and inform visualisation technique design [36, 41, 43].

In this paper, we explore how stakeholders represent cyber-risk data and how it compare to industry visualisation techniques. We conducted a sketching workshop with 41 participants to gather relevant sketches for comparison. To compare these sketches with industry visualisations, we utilise and expand upon the representation continuum presented by Walny et al. [44]. Participants also provided textual reflections on their sketching experience, which were analysed using Reflexive Thematic Analysis [8] to extract key themes and insights for a deeper analysis.

2 The Process, Visualisation, and Mental Models

Cyber risk assessments extensively use visualisation methods to support decision-making and insight development. This section grounds the reader in the core concepts and underpins the work in cyber risk mental models, user-centred design, and the use of sketching alongside current classification methods for visualisation techniques.

2.1 Cyber Security Risk Assessment

The ISO/IEC 27005 standard [6] defines Risk Assessments as being composed of *Identify*, *Analyse* and *Evaluate* stages to assess an organisation's current or potential risks. The risk assessment output is then used to inform the selection of controls to manage the identified risks.

Risk Identification articulates events that can cause loss and understanding where, how, and why potential losses can occur. This stage is the backbone of the risk assessment process and gathers data for the; identification of assets, identification of threats, identification of existing controls, identification of vulnerabilities, and identification of consequences [6].

Risk Analysis is "the process to comprehend the nature of risk and to determine the level of risk" [6]. A qualitative and quantitative analysis methodology, drawing on subject and objective data is often ascribed. ISO 27005 suggests that qualitative analysis is commonly used to obtain a subjective baseline of risk levels throughout an organisation by obtaining rating evaluations of likelihood and consequences to categorise properties of risks. However, there is a growing trend towards the adoption of quantitative methods [21]. Here, numerical data drawn from various sources (historical cyber attack type frequency and impacts on the organisation or shared risk information between organisations) is used to model an understand risks. Commonly the expected values (arithmetic mean) for the likelihood and impact are used to calculate the Annualised Loss Expectancy, which is used for evaluation purposes [6]. However, Hubbard et al. [21] and the Factor Analysis of Information Risk [16] propose the use of more comprehensive data sets as the product of the analysis for use in Risk Evaluation.

Risk Evaluation is the focus of visualisation techniques. In this stage, risk analysis data are compared against a defined risk evaluation criteria and the risk acceptance criteria identified at the start of the assessment process. Visualisations of the data are essential at this stage to support decision-making and insight generation. The output of this stage is a list of prioritised risks to be dealt with during the risk treatment phase. The work of Pan & Tomlinson [35] indicates that current academic literature does not fully explore risk evaluation and states the field lacks academic work, and concludes there is a gap in the research regarding the selection of risk criteria and the development of risk comparison techniques which are fairer, suitable, and accurate.

[*] e-mail: t.miller@lancaster.ac.uk
[†] e-mail: ms535@st-andrews.ac.uk
[‡] e-mail: d.prince@lancaster.ac.uk

2.2 Existing Visualisation Techniques for Risk Evaluation

Prior to the study, we collated industry-referenced risk visualisation techniques used in the evaluation stage of the risk assessment process. Starting with the NCSC Risk Assessment Guidance (RAG) [32] and the NCSC Cyber Assessment Framework (CAF) [33], through snowball sampling, we identified twenty-six possible risk assessment guidance documents published after RAG and CAF (published in 2016 and 2019 respectively) including documents not referenced by the NCSC CAF and RAG publications. Of this group of documents, only twenty-one were openly accessible, and only eight of this group referenced a form of risk visualisation for evaluation. In total, these eight publications identified thirty-nine risk evaluation visualisation techniques. These techniques are listed and analysed deeper in 5.4. A key insight from the guidance publication review was the lack of details on correctly using the identified visualisation techniques. This included limited information on the correct data to be used, the implementation process, and the appropriate use of the visualisation technique. The suspicion therefore, is that this would lead to incorrect visualisation technique application resulting in poor outcomes from risk assessments.

It is noteworthy that the ISO/IEC 31010 [22] series of standards is an outlier here. The standards document provides access to forty-two risk management tools and techniques, of which 18 are visualisation techniques. ISO/IEC 31010 provides a clear summary of the tools and techniques while clearly articulating where and how they can be implemented within the risk assessment process [22].

2.3 Mental Models and User-Centred design for Cyber Risk Visualisation

Jones et al. describe mental models as a “cognitive representation of external reality” [23]. The work of Kang et al. identifies that, regarding expert and non-expert users, non-expert users have a shallower mental model due to lack of experience [24]. This is something to be aware of when considering the development of visualisation techniques using sketching. Although this is the case, mental models can still positively influence cyber-security visualisation [5, 7, 17, 18, 23, 34, 37, 40]. Implementing mental models of expert and non-expert users can improve the communication and development of visualisation techniques [1]. This can result in a better understanding of risk for an array of users and, in turn, provide better risk communication, consultation, monitoring, and reviewing of risks.

The practice of user-centred design places users’ needs and wants at the core of the design process [11]. This practice enables the evaluation of newly developed visualisation techniques and to see if already existing visualisation techniques correctly cater to users [30]. Multiple cyber security visualisation papers recommend the implementation of user-centred design methods during the early stages of the design process, but none recommend it to evaluate existing visualisation techniques and, therefore, improve the development of new ones [4, 15, 19, 28, 39, 42].

Sketching “as a method of elicitation is effective at capturing mental models and complex ideas that are hard to explain within words.” [40]. The implementation of sketching within the user-centred design process allows for a diverse collection of user mental models. Additionally, it allows us to collate and identify themes and characteristics of the aforementioned sketches to compare against existing visualisation techniques to determine if they are suitable. Sketching has already been used within the cyber security context; McKenna et al. implemented data sketches to gather feedback on different visualisation techniques for the development of a cyber security dashboard [29]. This enabled McKenna et al. to iteratively develop a visualisation technique that catered to a broader audience. Sturdee et al. explored experts’ and non-experts’ understanding of cybersecurity concepts through the process of sketching as a visual aid [40]. This produces a visual dictionary that users, system developers, and researchers can use to understand better and implement cybersecurity concepts.

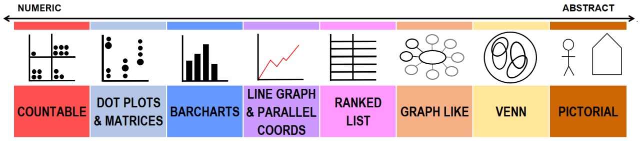


Figure 1: A reproduction of the continuum used in Walny et al. [44]

2.4 Visualisation Classification

The classification of visualisation techniques is also well explored within academia. Keim [25] classifies visualisation techniques based on their dimensions, text/web, hierarchies/graphs, and algorithm/software. These data types are further classified into standard 2D/3D displays, geometrically transformed displays, icon-based displays, dense pixel displays, and stacked displays. Schneiderman [38] also provides the classification of visualisation techniques by identifying data types (one-dimensional, two-dimensional, three-dimensional, multidimensional, tree, and network). Ellis and Dix [13] determine that clutter reduction is required for existing information visualisation techniques. Therefore, they classify the data provided within visualisation techniques to determine if new visualisation techniques are needed. Walny et al. [44] developed a classification approach called ”a representation continuum” with the help “of people with varying degrees of visualisation expertise”. This categorises simple visualisation techniques, such as sketches, based on the range of numerical and abstract data present in the sketch, shown in Figure 1.

Within the domain of cyber-security, Damaševičius et al. [10] provides a comprehensive outline of visual analytical methods and techniques applied to real-time data. Based on the visual variables presented in Bertin [3], Damaševičius et al. compare visual analytics platforms and tools to enable the identification of key challenges for the future research landscape in visual analytics within the cyber security domain. Damaševičius et al. [9] build upon their previous work [10] to identify the strengths and weaknesses of visualisation techniques. Although this academic work is within the cyber-security domain, it focuses on real-time threat intelligence data using dynamic visualisation techniques. In comparison, this paper aims to explore risk assessment visualisation which utilises long-term data. Whilst the work of Damaševičius et al. [9, 10] is rooted in the domain of cyber-security, it focuses on real-time threat intelligence data using dynamic visualisation techniques. In contrast, this paper aims to explore risk assessment visualisation which utilises long-term data [22].

3 Research Method

The methods applied in this study utilise the approach used in Walny et al. [44]. Their work focuses on expert and non-expert data set representation, identifying what participants learned or found interesting about the data. The analysis method produced a representation continuum that orders the visualisation techniques based on the level of abstraction of data representation. This study builds on the work of Walny et al. by extending the data gathering and analysis approaches, in particular refining the approach to constructing the representation continuum. In addition we use the continuum to compare participant visualisation sketches with industry visualisation techniques identified in Section 2.

3.1 Participant Profile

The study had forty-one participants in total, using an opportunistic sampling method drawing on a population of science and technology students studying at Lancaster University. All participants had a bachelor's degree in a computer-related field, such as computer science or software engineering and were all pursuing a masters in cyber-security. All participants had completed a module in cyber-security risk management giving them a strong understanding of risk visualisation techniques and core cyber-risk concepts. Four participants had a previous professional background as cyber-security practitioners ranging from 1-6 years experience. Participants were aged between 21 and 30 years old, with a mix of 13 women, 28 men, and 0 non-binary.

3.2 Ethics

All participants were provided with a consent sheet detailing the purpose of the study, confidentiality, anonymity, and withdrawal measures at the start of the study. Participants were reminded during the study, they were free to withdraw from participation up to any point until two weeks after their participation in the study was completed. After this point, the study data was anonymised, and therefore, it was not possible to remove individual contributions from the data set. Participants were provided with a debriefing sheet and a means of contact if they had any further questions or wanted to provide any additional information after their participation in the study.

3.3 The Workshop - Data Gathering Method

To ensure consistency during the workshop, a script and workflow procedure was used with key activities which were time restricted to ensure progression. Participants had free choice over seating; however, seating arrangements were in table groups rather than individuals being isolated. All the required materials were available to the participants from the outset, including a copy of the workshop flow – see Appendix B. The workshop had three phases and ran for two hours in total.

A synthetic data set using the quantitative methods of Hubbard & Seiersen [21] was generated for use in the study, based on data taken from the Cyber Security Breaches Survey 2023 published by the UK government [20]. This generated a set of risks containing their frequency, minimum, and maximum cost – see Appendix C. The data set was limited to five well-known risk types to enable participants to fully engage with the data due to the limited number and familiarity *(risk 1, risk 2, risk 3, risk 4, risk 5).* This acted as an experimental control to enable comparisons between participant sketches and industry visualisation techniques.

Phase 1 followed the Walny et al. method of data sketching. Participants were asked to explore the data and sketch a representation of their choosing. Participants were assured there was no incorrect approach but to focus on highlighting interesting aspects of the data. Participants were urged to consider helpful concepts such as; connections between different pieces of data, ways to group the data, similarities in the data, differences in the data, interesting patterns, and surprising findings [44].

Phase 2 extended the method of Walny et al. in an attempt to gain a deeper insight into alternative visualisation techniques. Following the initial sketch, participants were asked to provide a further sketch but were not permitted to incorporate the approach from the first sketch. This technique is used in the design thinking process to drive ideation and support a deeper investigation of concepts and novel approaches.

Phase 3 captured the participants' reflections on what they had learnt or found interesting about visualising the risk data set during the session. This was captured in written form.

3.4 Qualitative Data Analysis

Of interest in this research is exploratively understanding the range of approaches used to depict the data and how the sketching approach led to insight. There are numerous methods to analyse user-generated imagery [2, 40, 44] and text [8, 27] as part of qualitative research. Given the exploratory nature of the work, thematic analytical methods were deemed the most appropriate [8].

For the **sketch data**, the research team intended to use template analysis with the identification of a priori themes [26] with the a priori codes and themes drawn from the *representation continuum* identified by Walny et al. [44]. However, we found several issues in the initial application attempts, which required the continuum to be reconsidered prior to its application.

The *representation continuum* (shown in Figure 1) was an attempt by Walny et al. to order the sketched representations from numerical to abstract. However, upon review and when attempting to apply this continuum as part of a template in our analysis, a number of logical inconsistencies were revealed. The conceptualisation of a continuum implies a form of ordinal scale based on some intrinsic properties which define a series of groups that can be ordered from numerical to abstract. In the initial stages of the application of the continuum, it was identified that some numerical representations of data were further up the continuum than expected.

For example, Walny et al. considered a sketch numeric if numeric data was directly observed in the sketch and "increasingly abstract to the extent to which the data has been manipulated or worked with before being graphically represented in the sketch" [44]. A number of inconsistencies were identified with this approach; for example, a Ranked List was identified as more abstract than bar charts and line graphs, which collate multiple data points. However, a Ranked List provides a numerical representation of data as it is data in a raw form categorised by at least one column [31]. As such, the continuum could not be used 'as is', but its structure – the mechanism to define the ordinal nature of the continuum – needed to be included in the reflexive analysis of the sketched artefacts.

This study produced three sets of visualisations which needed to be analysed - the two from the workshop sketches, and a set retrieved from the industry literature. The categorisation of each visualisation was undertaken by the primary researcher and triangulated with the other authors.

The reflexive development of the representation continuum and the addition of new categories was undertaken iteratively by the three authors to account for individual researcher bias and to ensure objectivity. Image analysis sessions by the primary researcher produced codes of their properties and identified images which did not fit into the existing continuum categories.

These images were used as part of an independent researcher analysis to test the continuum's ordinal structure and whether a new, discrete grouping was necessary. Triangulation between the researchers resolved conflicts and ensured objectivity. This development was guided by comparing uncategorisable participant sketches with a visualisation categories guide developed by Dullaert [12] and enabled the selection of predefined and structured categories used to refine the continuum. The categories and order defining the refined continuum are given in Section 4.

The **textual data** from participants provided a reflection of what they had learnt or found interesting during the session. Reflexive Thematic Analysis [8] was applied to the textual data to extract codes and, in turn, generate themes across participants. The Reflexive Thematic Analysis was split into six phases:

Phase One - Familiarisation with the data set: The primary researcher thoroughly immersed themselves in the data set, reading and re-reading it. During this phase, brief notes were made to capture analytical ideas and insights generated from each data item or the data set as a whole. **Phase Two - Generating initial codes:** Systematically working through the data set, the primary researcher

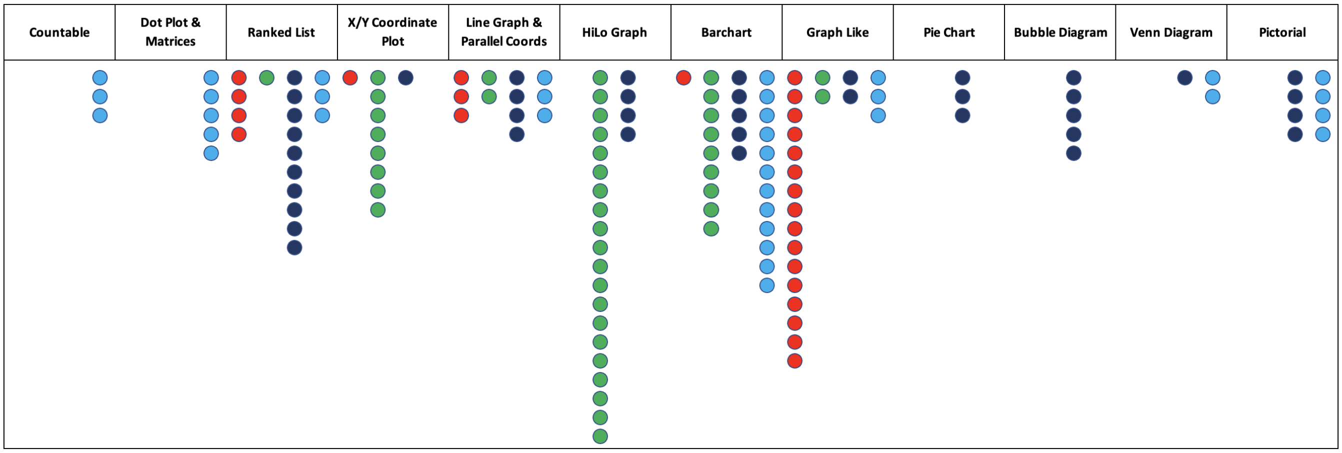


Figure 2: The Refined Representation Continuum. Light-blue dots represent the categories populated on the original Walny et al. continuum. Red dots represent visualisation techniques identified in industry. Green dots represent participants' Phase 1 sketches. Dark-blue dots represent participants' Phase 2 sketches.

identified segments of data that displayed potential interest, meaning, or relevance to the research. Analytically meaningful labels were then assigned to these sections. Afterward, the primary researcher collated the code labels and compiled corresponding segments of data for each code. The final codes can be found in Table 1. **Phase Three - Generating initial themes:** In this phase, the primary researcher grouped codes into clusters based on patterns, connections, and relationships, ultimately forming potential themes that encapsulated vital concepts within the data. **Phase Four - Developing and reviewing themes:** Initial themes underwent a meticulous review and revision process to ensure their accurate representation of the data. The primary researcher took steps to eliminate any potential theme duplication by clarifying the boundaries and scope of each theme. **Phase Five - Refining, defining, and naming themes:** The primary researcher crafted clear and concise descriptions for each theme, while also assigning meaningful and evocative names that captured their essence. The final themes are detailed in Table 2. The iterative process encompassing Phases One to Five was a collaborative effort involving the primary researcher, and the results were shared with other authors for triangulation. This step helped reduce personal subjectivity and bias and served to validate the interpretation of themes. **Phase Six - Writing up:** The identified themes were documented and presented in a sequence based on their significance – see Section 4.5. The themes generated are substantiated with relevant quotes and examples drawn directly from the data set.

4 Results

This section presents the results from the application of the qualitative analysis on the continuum, the sketches, and the textual data. In a conscious effort to uphold objectivity and mitigate individual biases, our research team employed a triangulation approach. This involved comparing and contrasting perspectives on the data, thus validating interpretations and findings. By embracing this method, we explored various perspectives, allowing us to cultivate a more holistic and thorough compilation of results.

4.1 Refined Representation Continuum

This refinement of the continuum identified two categories from Walny et al. that needed to be relocated and identified four new categories. The development of intrinsic properties for each category highlighted the necessity to reconfigure two sections of the continuum. Both Ranked List and Line Graph & Parallel Co-ord were moved to be more numerical representations of data. The categories are ordered from numeric to abstract, as per Walny et al., while taking into consideration the developed intrinsic properties: Countable, Dot Plot & Matrices, Ranked List, X/Y Coordinate Plot, Line Graph & Parallel Co-ords, HiLo Graph, Bar Chart, Graph Like, Pie Chart, Bubble Diagram, Venn Diagram, Pictorial. A more detailed description of the categories can be found within Appendix A.

The new continuum, with examples, is given in Figure 2, along with the data points gathered as part of the study. Red dots indicate data from the industrial visualisation data set, Green dots represent data from the first participant sketch, and dark-blue dots are visualisation data from the second participant sketch. The Walny et al. data set is included as light-blue dots for comparison.

4.2 Industry Representation Continuum

In total, twenty-five industry-referenced visualisation techniques were analysed. The overwhelming majority of the visualisation techniques were graph-like (16), with the remaining techniques of a more numerical nature (9), covering a range of 6 categories. There were four instances of ranked lists identified.

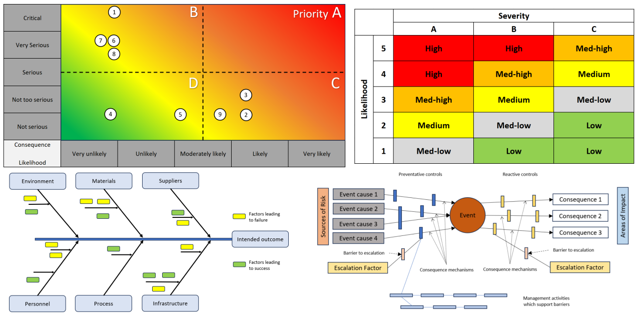


Figure 3: A reproduction of four examples of industry visualisation techniques categorised as Graph Like.

The ranked lists presented risks ranked based on either their rate of occurrence, their consequences, what vulnerability was being exploited, and the risks' overall ranking. The observed Line graphs (3) were used to represent probability distribution functions and cumulative distribution functions. One bar chart was identified, which presented the cost of each risk as a percentage of the annual loss expectancy. One instance of X/Y coordinate plot displayed the frequency and consequence of a set of risks. Additionally, this visualisation technique highlights acceptable and intolerable regions for risks. No techniques were identified in the highly numerical categories of the refined continuum (countable and dot plot) and the highly abstract categories (pie chart, bubble diagram, venn diagram, and pictorial).

4.3 Phase 1: Primary Sketch

For phase one of the workshop, forty-two sketches were collated from the forty-one participants. The most frequently observed was the HiLo Graph (20) covering a range of six categories.

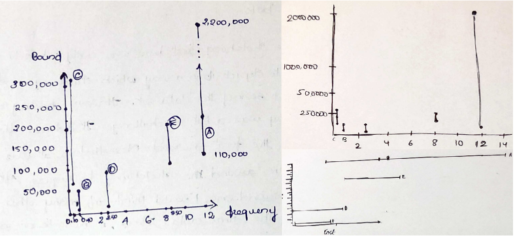
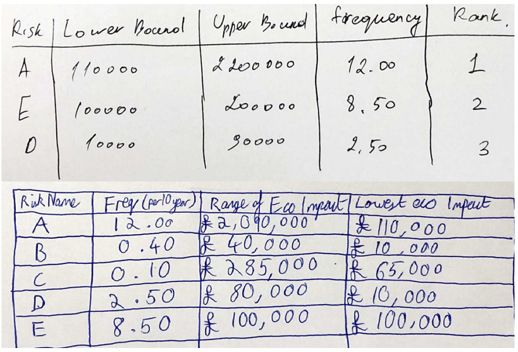


Figure 4: Three examples of the most popular category within participant Phase 1 sketches – HiLo Graphs.

Within the HiLo visualisations (see Figure 4 for examples), the risk data was presented individually based on their minimum to maximum costs. This range was presented as a line on the two-dimensional graph and plotted against the frequency. The vast majority of the remaining sketches were numerically focused (21), and one was abstract. Of these 21 sketches, nine were bar charts; commonly representing two dimensions of data (6), often the risk's name and either their rate of occurrence or cost. The remaining bar chart sketches implement three dimensions of data by increasing the width of the bar chart and the height or by displaying the risk occurrence rate on each bar. Eight X/Y Coordinates were observed within phase one of the study. All of the sketches plotted the risk occurrence rate against the risk's upper or lower bound cost. Three line graphs were sketched by participants, with two of them plotting the risks name against the risks rate of occurrence with a line between the identified points. The final line graph presented a similar approach to X/Y coordinate graphs (a comparison between cost and frequency), but a line was drawn between all identified points. One graph-like representation was observed, which emphasized the frequency of the risk and its minimum and maximum cost per occurrence. From the participants' initial sketches, we identified no techniques in the highest numerical categories (countable and dot plot) and in the highest abstract categories (pie chart, bubble diagram, Venn diagram, and pictorial).

4.4 Phase 2: Secondary Sketch

The second set of sketches from the forty-one participants (where they were not permitted to use the same technique used initially) produced thirty-nine sketches, with 26% of sketches being a ranked list (10 total, see Figure 5 for examples) and covered 10 categories extending all the way to the most abstract level - Pictoral.



Risk	Lower Bocund	Upper Bound	frequency	Rank.
A	110000	2200000	12.00	1
E	100000	200000	8.50	2
D	10000	90000	2.50	3

Risk Name	Freq (per 10 year)	Range of Eco Impact	Lowest eco Impact
A	12.00	£2,090,000	£110,000
B	0.40	£40,000	£10,000
C	0.10	£285,000	£65,000
D	2.50	£80,000	£10,000
E	8.50	£100,000	£100,000

Figure 5: Phase 2 Ranked List examples sketched by participants.

In addition to ranked lists, five bar charts were identified, with an emphasis on presenting the frequency or maximum cost of the risks. Four HiLo graphs sketches were observed. Three of these sketches presented risks based on their minimum to maximum costs and their frequency in a two-dimensional graph. The final HiLo sketch presented the same data but represented as a 'Circular HiLo Graph'(see Figure 6). Four line graphs were sketched by participants with a focus on the frequency of risks for comparison or a comparison of the risk's upper and lower bound costs in combination with their rate of occurrence.

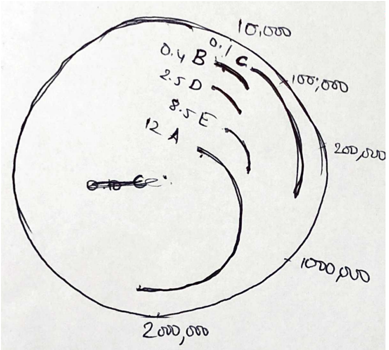


Figure 6: Phase 2 'Circular HiLo' sketch from a participant.

Bubble diagrams (4) and pie charts (3) presented the risks from no numerical data to a list of all values within the provided risk data set. Pictorial (4) representations provided a high-level approximation of the data using the previously outlined techniques, such as stick figures. Two graph-like sketches were identified. One of these sketches provided an overlapped comparison of the range of consequences, whereas the other representation provided no numerical data. One X/Y coordinate plot was observed with a comparison between a risk's rate of occurrence and consequence. Lastly, one Venn diagram was identified, which implemented countable characteristics.

4.5 Phase 3: Participant Reflections

At the end of the workshop, participants shared their insights from sketching the risk data set. Participants wrote between two sentences and one-third of a page. Of the forty-one participants, four did not share their insights. The encoding of the thirty-seven shared reflections produced seventy-two excerpts. These excerpts generated fourteen codes. Table 1 presents the fourteen codes from which themes were identified. Codes are ordered alphabetically, and the coloured cells represent a participant contributing to a certain code. Participants are ordered from left to right by participant number.

Table 1: Codes discussed by each participant with relevant theme numbers.

Code	T	Participants discussing code																																				
Analysis of data set without using sketches	1										🟩		🟩	🟩		🟩	🟩	🟩					🟩		🟩		🟩	🟩	🟩	🟩				🟩		🟩		🟩
Analysis of data using both sketches	1																														🟩					🟩		
Analysis of phase 1 sketch	1			🟩						🟩		🟩							🟩	🟩	🟩			🟩														
Analysis of phase 2 sketch	1									🟩									🟩		🟩																	
Challenging task	2	🟩	🟩				🟩	🟩								🟩				🟩																		
Comparison of data	3																	🟩				🟩				🟩	🟩	🟩			🟩				🟩			
Comparison of risks	3				🟩										🟩	🟩								🟩		🟩												🟩
Comparison of sketches	3									🟩																												
Findings within the data	4		🟩	🟩									🟩												🟩											🟩	🟩	
Lack of confidence	2			🟩																	🟩																	
Loss of information	2							🟩				🟩					🟩		🟩	🟩																		
Summary of data set	5				🟩	🟩																													🟩			
Summary of sketch 1	5								🟩																							🟩	🟩					
Summary of sketch 2	5								🟩																								🟩				🟩	

The application of reflexive thematic analysis [8] collated the presented codes into five themes: *analysis of data*, *the challenges of representing data*, *a comparison of data or sketches*, *the summarisation of data*, and *findings within the data*, as shown in Table 2.

4.5.1 Theme 1 - Analysis of Data

The most represented theme highlighted from the identified codes was the analysis of data. Surprisingly, most participants prioritised analysing the data set without using any of the sketches (or they failed to specify what sketch they were using). Some participants specified their usage of either the initial sketch or secondary sketch, and two participants highlighted that they used both sketches to analyse the data. Although we have identified a broad approach to the data analysis, participants still all came to similar conclusions around comparing certain risks and identifying which ones to prioritise.

4.5.2 Theme 2 - The Challenges of Representing Data

Participants expressed how difficult it was to represent the data – especially for their second sketch. Participants wrote primarily about the loss of information when representing the allocated data and expressed that this is due to the vast size of values. Although participants were reassured throughout the study that there is no wrong way of approaching the sketching, they still indicated that they wanted to do it “right and perfect” and that this was difficult. Participants also expressed a lack of confidence in the developed sketches, with uncertainties around whether their sketches were coherent or represented the data correctly.

4.5.3 Theme 3 - A Comparison of Data or Sketches

The next most represented theme within the study was the comparison of data or sketches. When comparing the data, participants focused on the lower and upper-bound costs of the established risks, with participants mixed responses around if there was a correlation between the frequency and the size of the bounds. Additionally, participants compared the risks as a whole and prioritised them in order of highest to lowest based upon a combination of their cost and frequency. One participant provided a comparison of the sketches they developed, stating that their initial sketch provided more emphasis on the data set.

4.5.4 Theme 4 - The Summarisation of Data

Some participants summarised the data set they were provided, outlining the names of the columns and the time the data was presented. Two participants provided a simple summary of the initial sketch specifying the data it represents. Finally, three participants summarised their second sketch using the same approach.

4.5.5 Theme 5 - Findings Within the Data

Six participants highlighted their findings from the sketching task. Three of these participants expressed interest in the fact that the data can be visualised in diverse ways, such as aggregating the data or omitting certain components. The remaining participants highlighted the importance of creating a streamlined mapping of data to visual representations to fully understand the data set and its characteristics, stating that this creates a better risk management process.

5 Discussion

This section collates the results presented in Section 4 to provide a comparative discussion surrounding both sets of sketches and the industry visualisation techniques. The summary of this section highlights and answers key research questions.

5.1 Phase 1 vs Phase 2

When comparing Phase 1 and Phase 2 sketches, by the spread of category (green and dark blue dots respectively in Figure 2), Phase 2 presented a broader diversity of visualisation techniques and extended much more into the abstract end of the continuum inhabiting ten categories ranging from Ranked List to Pictorial. Overall, Phase 1 sketches present a more compact and numerically focused representation of the risk data set – in total, six categories from Ranked List to Graph Like. This shows that the method of preventing the use of elements from the first sketch in the second sketch drives diversity of thought and a diverse set of different visualisation techniques. This two-phased method provides an interesting approach for creating and developing a range of diverse visualisation techniques. Although this is the case, their viability within industry is still required to be evaluated.

Of note is the mode of the Phase 2 sketches is in the numerical ranked list category. Interestingly, this is more numerically focused than the mode of the Phase 1 sketches. This comparatively more numerical mode may be due to the risk data being provided to participants in a ranked list format. This ranked list mode may have been driven by participants struggling to represent the data once their initial concepts had been excluded. This can be seen by drawing on the thematic analysis as participants emphasized the challenges of representing data, especially for their second sketch. One unexpected but important finding from this study was how the impact of presenting data to participants in a certain format influences their sketches.

5.2 Industry vs Phase 1

When comparing Phase 1 sketches with industry visualisations based on their placement on the continuum (green and red dots in Figure 2), it is evident that they both cover the same categories, ranging from

Table 2: Identified themes from the sketching study in order of the frequency of occurrence from highest to lowest.

Themes	Frequency of occurrence	Example of participant extract
Analysis of data	28	"The risks can be sorted in both ways: in the amount of impact and the overall threat. While sorted in impacts, Risk B and D shared the same lower bound. Their upper bounds are not higher. A, C, and E can be classified as having high risk. While sorted by overall risks, risk A can be ranked as the biggest risk as it happens the most and drops on a higher amount than the other four risks. Risk E can be classified as the 2nd biggest risk."
Challenges of representing data	16	"I'm not really sure if my graphs are correct and coherent with the data given."
Comparison of data or sketches	14	"With the help of a quantitative approach, we have measured the risk and the risk occurrence. The maximum occurrence of risk A is the highest while Risk B is recorded as the lowest."
Summary of data	8	"The data set has the following classification, such as RISK_NAME, frequency, lower bound, & upper bound. The data has to be compared between the number of risks that have happened with the frequency of the attack. The lower bound and the upper bound are the costs involved in risk remediation."
Findings within the data or sketches	6	"It was interesting to see how we can visualise a particular data set in various ways. Whether it means to find average, grouping, or omitting a particular column."
Total number of excerpts:	72	

Ranked List to Graph Like. This unexpected alignment indicates that participants' initial mental models correspond to the visualisation techniques commonly used in industry. This alignment underscores three important findings: 1) the range of numerical to abstract representations of risk data is suitable for technical stakeholders, 2) Phase 1 sketches can serve as a basis for developing and enhancing industry visualisation techniques due to their appropriate complexity, and 3) the Phase 1 sketches obtained from this study hold relevance and significance. To develop these findings further, future research can explore non-technical stakeholders and investigate whether their initial mental models align with industry visualisations and those of technical stakeholders.

When observing the Phase 1 sketches, we can clearly see that their mode is HiLo graphs. In comparison, the most prevalent industry visualisation category is Graph Like which provides a more abstract representation of risk data. This shows that industry visualisation techniques focus on presenting risks more abstractly to support decision-making. These visualisation techniques are often populated by technical stakeholders and passed upwards within an organisation. This means that technical stakeholders are often required to evaluate a larger array of risks than non-technical stakeholders and provide a list of the most pressing risks. Drawing from the thematic analysis, we can see that participants discussed how they analysed, compared, and summarised the data to present their findings, supporting the need for visualisation techniques for technical stakeholders.

Although Phase 1 sketches and industry visualisation techniques span the same categories, there are no industry visualisation techniques that are within the most popular category of Phase 1 (HiLo graph). This disparity highlights the under-representation of visualisation techniques for technical stakeholders. Additionally, this emphasises the opportunity for the development of new visualisation. Promoting the development of visualisation techniques within the aligned range can produce appropriate visualisations for sets of stakeholders within industry.

5.3 Industry vs Phase 2

When comparing Phase 2 sketches with industry visualisation by the coverage of category within the refined representation continuum (dark blue and red dots respectively in Figure 2), Phase 2 sketches populate a broader spectrum of categories. Both continuums start at Ranked Lists, but Phase 2 sketches extend to Pictorial, the highest abstract category within the continuum. These findings coincide with the comparison between Phase 1 and Phase 2 sketches showing that participants' secondary sketches provide a greater diversity of visualisation techniques. Although Phase 2 sketches exceed the range of commonly used visualisation techniques within industry, they could be used to implement a more diverse range of visualisation techniques. Additionally, characteristics developed from "out of scope" techniques may be applied to more traditional visualisation techniques to enable a wider stakeholder catering.

When comparing modes, industry visualisation techniques provide a more abstract representation of cyber risk data, emphasising the use of Graph Like approaches. By comparison, Phase 2 sketches mode (Ranked Lists) provides a heavily numerical data representation. After thoroughly comparing Phase 2 sketches and industry visualisation, the findings aligned with the previously identified challenges of representing data. Thus, this comparison yielded no additional information to provide further insights. One way to provide further insights would be to re-run the study with non-technical stakeholders. This would produce a more insightful comparison.

5.4 Findings

Based on the discussion so far, the results and the identification of industry-focused visualisations, it is possible to address the core questions:

What visualisation techniques are being used within industry: The background section of this paper identifies and collates existing visualisation techniques that are being used within industry. These techniques were imported into the refined continuum and spanned six categories ranging from Ranked List to Graph Like. From reviewing the populated continuum, it can be determined that industry visualisation techniques are weighted to be primarily graph-like. Overall, in regard to the data and the refined continuum, industry visualisation techniques are weighted to provide an abstract representation of data.

How do individuals visualise their perceptions of cyber-security risk data: From the clustering of participant sketches and the development of the refined continuum, it can be clearly seen that cyber-security-focused stakeholders provide an array of ways to visualise cyber-risk data. When applying participants' Phase 1 sketches to the refined continuum, there is an apparent preference towards numerical representations of data. Participants' phase two sketches, where they could not use any sketching techniques they applied in the phase one sketch, populated a broader range spanning ten of the twelve refined continuum categories. This shows a creative diversity from participants and could be used to enhance the features of already existing visualisation techniques within industry. The application of Reflexive Thematic Analysis to participant reflections displayed the challenges of visualising vast ranges of data and the loss of information in doing so. Participants particularly struggled with their phase-two sketches, with 25% of participants sketching the data set exactly as it was presented to them.

How do stakeholder sketches compare with the visualisation techniques used within industry: The comparison of Phase 1 sketches with industry-implemented visualisation highlights how participants' inceptive mental models align with the range of visualisation techniques utilised within industry. Although this shows industry is providing the correct granularity of information for technical stakeholders, disparities start to occur when reviewing the visualisation techniques themselves. Over half of the participants' Phase 1 sketches were HiLo Graphs, a visualisation technique not utilised within industry. This highlights the need to develop and implement new visualisation techniques within industry to cater to a spectrum of stakeholders. The vast variance in modes also raises concerns about the efficacy of currently utilised industry visualisation techniques. Evaluation of them is required to determine if they are appropriate for various stakeholders.

Can these sketches help develop future visualisation techniques or refine existing ones: The answer to the previous research question shows a disparity between industry visualisation and participant sketches. Participant sketches enabled the discovery of visualisation techniques that were not used within industry and highlighted the importance of a user-driven design process for future visualisation techniques. In addition to this, participants provided sketches with slight adaptations to visualisation used within industry. These adaptations can be used to refine existing techniques and enhance their capabilities when conveying risk data. For this to be a rigorous process, existing visualisation techniques within industry must be evaluated. This provides a baseline for comparison when refining these visualisation techniques. Additionally, this enables the implementation of the newly identified visualisation techniques with the ability to compare their performance against currently utilised visualisation.

6 Practical Implications

The findings of this study reveal a distinct disconnect between cyber-security-focused stakeholders and the prevailing visualisation techniques employed in industry. Surprisingly, the most popular visualisation technique among the aforementioned stakeholders is entirely absent from current industry practices. Reducing this misalignment can lead to practical implications that enhance the overall effectiveness of cyber-security operations: **Enhanced Risk Analysis:** Correcting the misalignment of visualisation techniques enables cyber-security analysts to accurately detect and analyze risks. This can help stakeholders identify anomalies, trends, and potential attacks more quickly and effectively. **Improved Decision-Making:** Catered visualisations provide decision-makers with clear insights into the current cyber-security landscape. This leads to better-informed decisions regarding risk prioritization, resource allocation, and incident response strategies, which ultimately strengthens the organisation's overall cyber defense posture. **Effective Communication:** When visualization techniques align with the needs of different teams and stakeholders, communication improves. Clear and intuitive visualisations facilitate effective knowledge sharing, enabling technical and non-technical staff to understand and respond to cyber-security issues more efficiently. **Streamlined Incident Response:** Reducing the misalignment speeds up incident response times and allows for more precise containment and remediation actions, minimizing potential damage.

One approach to reducing misalignment and enhancing the overall effectiveness of cyber-security operations is translating the identified sketches into visualisation techniques that are applicable to industry. This would consist of analysing the sketches from this study further and extracting key concepts, patterns, and design principles. Collaboration with visualisation experts, user experience (UX) designers, and cyber-security experts to refine and expand the initial sketches into detailed design concepts is essential. After the visualisations development, it would undergo usability testing involving cyber-security analysts and pertinent stakeholders. This process aims to collect feedback regarding its effectiveness, user-friendliness, and practicality. Section 2 identifies a lack of documentation surrounding the correct utility of visualisation techniques within industry. To reduce ambiguity and align with ISO 31010, the implementation of user documentation and training materials to enable stakeholders to effectively utilise the visualisations should be developed. This would encompass details about the target audience for the visualization technique, the types of data it can accommodate, and the manner in which this data can be employed to populate the visualization technique. This approach can be applied to enhance existing visualisation techniques with identified characters or support the integration of visualisation techniques that are not utilised in industry – such as the HiLo Graph.

7 Reflection on Study

We employed an opportunistic sampling approach to gather a group of technically inclined participants interested in cybersecurity. The range of participant sketches and industry visualisations suggests that the sample size was appropriate. However, it is important to note that the participant group does not represent the entire population interacting with cyber-risk data, and therefore, the findings may not be generalised but as an initial exploration, provide rich insight which can be capitalised on via further research. Conducting a similar study with technical and non-technical participants that span a wider age range would be beneficial for a more comprehensive comparison and set of results. Nonetheless, the sketches provided by the technical participants served as a valuable dataset for comparing against industry visualisations, addressing core questions, and identifying future research directions.

Due to space limitations, participants in the study were grouped around tables rather than working individually, which may have influenced the sketches due to discussions about the data. This impact could be positive, leading to improved visualisation techniques through idea exchange, or it could result in duplicate sketches. To minimize this impact, a two-phased sketching approach was implemented. Figure 2 demonstrates the diversity of sketches achieved with this approach. An unexpected finding was how the Ranked List format of the data influenced participant sketches when struggling to represent it. Further consideration may be necessary regarding the format in which data is presented to participants or restrictions on mirroring its original presentation format.

To mitigate personal bias or interpretation in the analysis of the qualitative data, the sketches, and the results, the research team employed a collaborative approach. All team members reviewed and discussed changes made to the continuum, the generated codes, the themes derived from the reflexive thematic analysis, and the results of the study.

Due to the limited scope of exploratory studies, it may not be possible to address broader research questions or extract definitive answers. To address this limitation, the researchers in this paper proposed a specific and focused research question in the Introduction. This question was further divided into four sub-questions and answered in Section 5.4.

8 Conclusion & Future Work

This work expands on the research of Walny et al. by conducting a qualitative analysis of participant sketches and written reflections. Participants created two sets of sketches: one without any restrictions and another set where they were instructed not to use techniques from the first set. Additionally, a third set of visualisations was generated by identifying twenty-five industry visualisation techniques used for risk evaluation. Initial analysis of participant sketches revealed logical inconsistencies within the continuum proposed by Walny et al. As a result, the continuum was refined to better represent a logical progression from highly numerical to highly abstract categories. New categories were developed based on the discovery of visualisation techniques from participants' sketches. The three sets of sketches were then used to populate the continuum, facilitating a comparison. Furthermore, Reflexive Thematic Analysis was employed to identify key themes from participants' reflections.

The findings of this study highlight the differences between the representations of cyber-risk data by industry and technical stakeholders. Participants' Phase 1 sketches aligned with the same range as industry visualisation techniques but introduced a greater emphasis on numerical representations, specifically HiLo graphs, which were not found in industry guidance. Phase 2 sketches provided a diverse range of visualisations that can be leveraged to enhance existing techniques within the industry. Additionally, the Reflexive Thematic Analysis revealed key themes related to data analysis, challenges in representing data, data comparison, data summarisation, and notable findings within the dataset and sketches.

This work identifies important research directions for the future; 1) The evaluation of existing industry visualisation techniques to assess their effectiveness, 2) Applying participant sketches to industry, which depends on the completion of (1) to determine if they outperform current techniques, 3) Expanding the study to include multiple stakeholder groups and age range for a more representative comparison, and 4) Applying this study to other domains to provide a better holistic understanding and validate the findings of this study. This can facilitate the development and implementation of tailored visualisation techniques that cater to the diverse needs of both technical and non-technical stakeholders.

References

[1] F. Asgharpour, D. Liu, and L. J. Camp. Mental models of security risks. In *International conference on financial cryptography and data security*, pp. 367–377. Springer, 2007.

[2] M. Banks. *Using visual data in qualitative research*. SAGE Publications Ltd, London, 2nd edition edition. ed., 2018.

[3] J. Bertin. *Graphische semiologie: diagramme, netze, karten*. Walter de Gruyter, 2010.

[4] D. M. Best, A. Endert, and D. Kidwell. 7 key challenges for visualization in cyber network defense. In *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, pp. 33–40, 2014.

[5] C. Bravo-Lillo, L. F. Cranor, J. Downs, and S. Komanduri. Bridging the gap in computer security warnings: A mental model approach. *IEEE Security & Privacy*, 9(2):18–26, 2011. doi: 10.1109/MSP.2010.198

[6] BSI. ISO/IEC 27005: Security techniques - information security risk management, 2011. Last Accessed: 26-10-2021.

[7] L. J. Camp. Mental models of privacy and security. *IEEE Technology and society magazine*, 28(3):37–46, 2009.

[8] V. Clarke, V. Braun, and N. Hayfield. *Thematic Analysis*. 2021.

[9] R. Damasevicius, J. Toldinas, A. Venckauskas, S. Grigaliunas, and N. Morkevicius. Technical threat intelligence analytics: What and how to visualize for analytic process. In *2020 24th International Conference Electronics*, pp. 1–4. IEEE, 2020.

[10] R. Damaševičius, J. Toldinas, A. Venčkauskas, Š. Grigaliūnas, N. Morkevičius, and V. Jukavičius. Visual analytics for cyber security domain: State-of-the-art and challenges. In *Information and Software Technologies: 25th International Conference, ICIST 2019, Vilnius, Lithuania, October 10–12, 2019, Proceedings 25*, pp. 256–270. Springer, 2019.

[11] I. DIS. 9241-210: 2010. ergonomics of human system interaction-part 210: Human-centred design for interactive systems. *International Standardization Organization (ISO). Switzerland*, 2009.

[12] M. Dullaert. Chart guide. `https://chart.guide/`, 2023. Last Accessed: 20-06-2023.

[13] G. Ellis and A. Dix. A taxonomy of clutter reduction for information visualisation. *IEEE transactions on visualization and computer graphics*, 13(6):1216–1223, 2007.

[14] M. J. Eppler and M. Aeschimann. Envisioning risk: A systematic framework for risk visualization in risk management and communication. *Institute for Corporate Communication (ICA) Working Paper*, 5:2008, 2008.

[15] R. F. Erbacher. Visualization design for immediate high-level situational assessment. In *Proceedings of the ninth international symposium on visualization for cyber security*, pp. 17–24, 2012.

[16] J. Freund and J. Jones. *Measuring and managing information risk: a FAIR approach*. Butterworth-Heinemann, 2014.

[17] B. Friedman, D. Hurley, D. C. Howe, H. Nissenbaum, and E. Felten. Users' conceptions of risks and harms on the web: a comparative study. In *CHI'02 extended abstracts on Human factors in computing systems*, pp. 614–615, 2002.

[18] S. Furman, M. F. Theofanos, Y.-Y. Choong, and B. Stanton. Basing cybersecurity training on user perceptions. *IEEE Security & Privacy*, 10(2):40–49, 2011.

[19] J. R. Goodall, A. A. Ozok, W. G. Lutters, P. Rheingans, and A. Komlodi. A user-centered approach to visualizing network traffic for intrusion detection. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems*, pp. 1403–1406, 2005.

[20] U. Government. Cyber security breaches survey 2023. `https://shorturl.at/lxyPV`, 2023. Last Accessed: 20-06-2023.

[21] D. W. Hubbard and R. Seiersen. *How to measure anything in cybersecurity risk*. John Wiley & Sons, 2016.

[22] ISO. ISO 31000: Risk management - guidelines. `https://bit.ly/32gzCHO`, 2018. Last Accessed: 01-06-2023.

[23] N. A. Jones, H. Ross, T. Lynam, P. Perez, and A. Leitch. Mental models: an interdisciplinary synthesis of theory and methods. *Ecology and Society*, 16(1), 2011.

[24] R. Kang, L. Dabbish, N. Fruchter, and S. Kiesler. {"My} data just goes {Everywhere:"} user mental models of the internet and implications for privacy and security. In *Eleventh Symposium on Usable Privacy and Security (SOUPS 2015)*, pp. 39–52, 2015.

[25] D. A. Keim. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):1–8, 2002.

[26] N. King. Template analysis. 1998.

[27] K. Krippendorff. *Content analysis : an introduction to its methodology*. SAGE Publications, Inc., Los Angeles, CA, 4th edition. ed., 2019.

[28] J. Landstorfer, I. Herrmann, J.-E. Stange, M. Dörk, and R. Wettach. Weaving a carpet from log entries: A network security visualization built with co-creation. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 73–82. IEEE, 2014.

[29] S. McKenna, D. Staheli, C. Fulcher, and M. Meyer. Bubblenet: A cyber security dashboard for visualizing patterns. In *Computer Graphics Forum*, vol. 35, pp. 281–290. Wiley Online Library, 2016.

[30] S. Miksch and W. Aigner. A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Comput-*

ers & Graphics, 38:286–290, 2014.

[31] P. Mylavarapu, A. Yalcin, X. Gregg, and N. Elmqvist. Ranked-list visualization: A graphical perception study. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[32] NCSC. NCSC risk assessment guidance. `https://www.ncsc.gov.uk/collection/risk-management`, 2016. Last Accessed: 12-05-2023.

[33] NCSC. NCSC CAF guidance. `https://www.ncsc.gov.uk/collection/caf`, 2019. Last Accessed: 10-07-2023.

[34] M. Oates, Y. Ahmadullah, A. Marsh, C. Swoopes, S. Zhang, R. Balebako, and L. F. Cranor. Turtles, locks, and bathrooms: Understanding mental models of privacy through illustration. *Proceedings on Privacy Enhancing Technologies*, 2018(4):5–32, 2018.

[35] L. Pan and A. Tomlinson. A systematic review of information security risk assessment. *International Journal of Safety and Security Engineering*, 6(2):270–281, 2016.

[36] P. Reid, F. Hallett-Hook, B. Plimmer, and H. Purchase. Applying layout algorithms to hand-drawn graphs. In *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*, pp. 203–206, 2007.

[37] D. Ricketts and D. Lockton. Mental landscapes: Externalizing mental models through metaphors. *Interactions*, 26(2):86–90, 2019.

[38] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *The Craft of Information Visualization*, pp. 364–371. Elsevier Inc, 2003.

[39] J. Stoll, D. McColgin, M. Gregory, V. Crow, and W. K. Edwards. Adapting personas for use in security visualization design. In *VizSEC 2007: Proceedings of the Workshop on Visualization for Computer Security*, pp. 39–52. Springer, 2008.

[40] M. Sturdee, L. Thornton, B. Wimalasiri, and S. Patil. A visual exploration of cybersecurity concepts. In *Creativity and Cognition*, pp. 1–10, 2021.

[41] F. Van Ham and B. Rogowitz. Perceptual organization in user-generated graph layouts. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1333–1339, 2008.

[42] M. Wagner, W. Aigner, A. Rind, H. Dornhackl, K. Kadletz, R. Luh, and P. Tavolato. Problem characterization and abstraction for visual analytics in behavior-based malware pattern analysis. In *Proceedings of the Eleventh Workshop on Visualization for Cyber Security*, pp. 9–16, 2014.

[43] J. Walny, S. Carpendale, N. H. Riche, G. Venolia, and P. Fawcett. Visual thinking in action: Visualizations as used on whiteboards. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2508–2517, 2011.

[44] J. Walny, S. Huron, and S. Carpendale. An exploratory study of data sketching for visual representation. In *Computer Graphics Forum*, vol. 34, pp. 231–240. Wiley Online Library, 2015.

[45] E. Zio. Challenges in the vulnerability and risk analysis of critical infrastructures. *Reliability Engineering & System Safety*, 152:137–150, 2016.

Appendices

A Description of Continuum Sections

- **Countable:** numerical values represented as tallies or multiple points. Each value can be visually observed, meaning this sketching method provides exact values.

- **Dot Plot and Matrices:** represents numerical values as a group of data points plotted on a simple scale. This approach is considered suitable for small to moderate sizes of data. When being used, it must conserve numerical information and provide exact values.

- **Ranked List:** displays numerical and written values as an ordered list of a given data set. This can be ordered based on any column.

- **X/Y Coordinate Plot:** single points plotted on a two-dimensional graph with all points the same size and provides exact values for each point.

- **Line Graph & Parallel Co-ords:** line graphs display information as a series of data points plotted with lines between them and is often represented on a two-dimensional graph. Parallel co-ord plot creates multivariate plots of the data with axis values corresponding to values that the data holds e.g. the frequency of a risk, lower bound, and upper bound. Provides exact values.

- **HiLo Graph:** numerical values represented as a line between two points. These two points are exact values for the lower bound and upper bound of a range of data. The lower bound is not required to start at zero.

- **Bar Chart:** multiple bars on an x-axis starting from zero the appropriate y-axis value, this can be represented vice versa. Provides an aggregation of data rather than individual values.

- **Graph Like:** high-level points that allows an abstract comparison between different data points but lacks accuracy. Exact values can be used to increase numerical accuracy.

- **Pie Chart:** uses slices to represent numerical proportions of a data set commonly with no values provided. Unable to represent negative values

- **Bubble Diagram:** numerical values are represented as circles where the size represents the value and colour represents the severity of said value. Unable to represent negative values and can provide up to three dimensions of data visualisation.

- **Venn Diagram:** provides a logical relationship between data using closed curves drawn on a plane. The curves overlap to show relationships between data, and the areas of the curves are proportional to the number of elements contained within.

- **Pictorial:** a storytelling representation of data using icons such as stick figures and line drawings. Provides a high-level approximation of a data set.

B Workshop Sheet

In this workshop you will explore data retrieved from a data breach investigation. This workshop will look into how this data can be represented and how we interpret it.

B.1 Sketching Task

Using the data set that has been provided, represent the given data on a blank sheet of paper that you have been provided. How you do that is completely up to you. **There is no wrong way to do this.** What is required of this task is to think about what might be interesting about the data, and to draw the data as you explore it. The data set is quite large, and you may not have time to draw the entire this, it is okay to pick only the parts of the data that you find interesting. **This must be done individually.** It may be helpful to think about the following:

- Connections between different pieces of data
- Ways to group the data
- Similarities in the data
- Differences in the data
- Interesting patterns
- Surprising findings

B.2 Sketching Task 2

Now that you have completed one sketch, do another sketching using none of the techniques you used to draw the first one.

B.3 Findings Task

Using a separate piece of paper, please write down what you learned or found interesting about this data during the session (once again there are no wrong answers). This will be used after the study as part of a qualitative analysis to identify common trends. Please try to write this in your best handwriting.

B.4 Finishing Up

Thank you for participating in the study, if you have any questions regarding the study or think of additional information you would like to contribute please do not hesitate to contact me at t.miller@lancaster.ac.uk.

C Data Set

risk_name	frequency	lower_bound	upper_bound
A	12.00	£110,000	£2,200,000
B	0.40	£10,000	£50,000
C	0.10	£65,000	£350,000
D	2.50	£10,000	£90,000
E	8.50	£100,000	£200,000