MASTERARBEIT | MASTER'S THESIS

Titel | Title

Understanding Visual Complexity of Data Visualizations

verfasst von | submitted by

Sara Hoxhaj

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 921
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Informatik
Betreut von | Supervisor:	Univ.-Prof. Torsten Möller PhD
Mitbetreut von | Co-Supervisor:	Laura Koesten MSc Ph.D.

Acknowledgements

I would like to thank everyone who supported me throughout this thesis. Special thanks go to my two supervisors, Torsten and Laura, for their guidance and feedback at every stage of this work. I would also like to thank everyone in the Research Group of Visualization and Data Analysis, who always took the time to support me and try out my surveys and pilot tests along the way. Many thanks go to Anne-Flore Cabouat, Petra Isenberg, and Tobias Isenberg, whose work inspired this thesis and who were kind enough to share their knowledge and experience with me.
I would like to thank my mom and dad for making this opportunity possible and for always being there for me. My thanks also go to Reni and Joana, who were by my side from the first day I arrived in Vienna and through every step that followed. Thank you to my friends and to Meviola, who was always there whenever I felt homesick or needed a break from work. Finally, a heartfelt thank you goes to Nicolò, with whom I shared this experience and who made every day of it a little more special.

Abstract

Data visualizations play an important role in how people understand information in everyday life. One of the factors that influences how a visualization is perceived and understood is its visual complexity. Visual complexity has been studied across many fields, but no universally accepted definition exists for how it is perceived specifically in the context of data visualizations. Because of this, measuring perceived visual complexity remains an open challenge in the research community.

This thesis takes an exploratory approach to address this gap. Rather than predefining the concept, the goal is to identify a set of items that can be used to capture perceived visual complexity from the viewer's perspective, by drawing on both existing research and direct input from domain experts. To do this, the work followed three main steps: term identification, item generation, and item validation. In the term identification step, a pool of 951 unique terms was collected through a literature review of IEEE VIS publications and an expert survey involving 50 visualization researchers. These terms were then clustered and refined into 49 candidate items. In the validation step, a second survey was shared with 45 domain experts, who rated each item based on how relevant they considered it for assessing perceived visual complexity. Cumulative Link Mixed Models (CLMMs) were used to analyze the results, leading to a final set of 16 items.

The top-rated items focused on interpretability, understandability, and comprehension, suggesting that cognitive effort plays an important role in how visual complexity is experienced. The remaining items cover aspects such as organization, pattern detection, emotional response, and design quality, confirming that perceived visual complexity is not a one-dimensional concept but is shaped by a combination of cognitive, emotional, and visual factors. The 16 items identified in this thesis provide a foundation that future researchers can build on, with the most immediate next steps being cognitive interviews with lay users and a larger validation study involving general audiences.

Kurzfassung

Datenvisualisierungen spielen eine wichtige Rolle dabei, wie Menschen Informationen im Alltag verstehen. Einer der Faktoren, die beeinflussen, wie eine Visualisierung wahrgenommen und verstanden wird, ist ihre visuelle Komplexität. Visuelle Komplexität wurde bereits in vielen Fachgebieten untersucht, doch es gibt keine allgemein anerkannte Definition dafür, wie sie speziell im Kontext von Datenvisualisierungen wahrgenommen wird. Aus diesem Grund bleibt die Messung der wahrgenommenen visuellen Komplexität in der Forschungsgemeinschaft eine offene Herausforderung.

Diese Arbeit verfolgt einen explorativen Ansatz, um diese Lücke zu schlieSSen. Anstatt das Konzept vorab zu definieren, besteht das Ziel darin, eine Reihe von Items zu identifizieren, mit denen sich die wahrgenommene visuelle Komplexität aus der Perspektive des Betrachters erfassen lässt, wobei sowohl auf bestehende Forschungsergebnisse als auch auf direkte Beiträge von Fachexperten zurückgegriffen wird. Zu diesem Zweck umfasste die Arbeit drei Hauptschritte: Begriffsidentifizierung, Itemgenerierung und Itemvalidierung. Im Schritt der Begriffsidentifizierung wurde ein Pool von 951 eindeutigen Begriffen durch eine Literaturrecherche in IEEE VIS-Publikationen und eine Expertenbefragung unter 50 Visualisierungsforschern zusammengestellt. Diese Begriffe wurden anschlieSSend geclustert und zu 49 Kandidatenitems verfeinert. Im Validierungsschritt wurde eine zweite Umfrage an 45 Fachexperten verteilt, die jedes Item danach bewerteten, wie relevant sie es für die Beurteilung der wahrgenommenen visuellen Komplexität erachteten. Zur Analyse der Ergebnisse wurden Cumulative Link Mixed Models (CLMMs) verwendet, was zu einem endgültigen Satz von 16 Items führte.

Die am höchsten bewerteten Items konzentrierten sich auf Interpretierbarkeit, Verständlichkeit und Erfassbarkeit, was darauf hindeutet, dass der kognitive Aufwand eine wichtige Rolle dabei spielt, wie visuelle Komplexität wahrgenommen wird. Die übrigen Items decken Aspekte wie Organisation, Mustererkennung, emotionale Reaktion und Designqualität ab und bestätigen, dass die wahrgenommene visuelle Komplexität kein eindimensionales Konzept ist, sondern durch eine Kombination aus kognitiven, emotionalen und visuellen Faktoren geprägt wird. Die 16 in dieser Arbeit identifizierten Items bilden eine Grundlage, auf der zukünftige Forscher aufbauen können, wobei die unmittelbar nächsten Schritte kognitive Interviews mit Laien und eine gröSSere Validierungsstudie mit einer allgemeinen Zielgruppe sind.

Contents

List of Tables

List of Figures

1. Motivation

The importance and impact of data visualizations in our everyday lives are widely recognized. They are not only an essential tool for communicating information across many fields but also a powerful factor in how people interpret, respond to, and make decisions based on the data they represent. By recognizing their significance, we can dive deeper to explore the factors that influence how visualizations communicate information effectively. One of these aspects is visual complexity, which is the main focus of this thesis. Visual complexity (VC) has been studied for many years across fields such as arts [17], psychology [13], web and user interfaces [40] [12], architecture [34], and data visualization. Definitions of visual complexity vary between fields and studies, often depending on the specific domain of investigation. Even though authors inherently face significant challenges in defining such a broad and abstract term with precision, their explanations generally align with the same underlying idea. One commonly used definition describes VC as the amount of detail or intricacy present in an image [1, 8, 15, 18, 23]. Another perspective frames VC as a form of visual uncertainty, representing the difficulty humans have in interpreting visual representations correctly [14]. Additionally, some authors define it as the combination of different features (color, shape, and orientation) and its variations (e.g., color: red, green, white) [33]. While these definitions come from different fields and often reflect the specific context in which they were developed, no universally accepted definition of visual complexity in data visualizations exists. Rather than predefining the concept, this thesis takes an exploratory approach by consulting domain experts to better understand how visual complexity is perceived and experienced in the context of data visualizations. Because there is no precise definition of VC, it is also difficult to identify what influences it and how it should be measured. As a consequence, a gap exists in the research field, leaving room for studies that systematically examine the factors affecting VC and how it shapes viewers perception and understanding of data visualizations. Addressing this gap is important, as VC can influence how effectively information is communicated, how easily patterns are recognized, and how decisions are made based on the data represented. Therefore, it is essential to investigate the key factors that contribute to visual complexity and their impact on how people perceive and interpret visual information.

1.1. Research Significance

In today's world, data visualizations are everywhere, from news articles and business intelligence dashboards to scientific publications and other public documents. Their presence in significant areas of communication and information demonstrates that visual design decisions are not only important but also influence the way users understand and interpret a data visualization. Even details that might initially appear insignificant, such as color choices or labeling, can accumulate and contribute to the overall visual complexity of a visualization, which ultimately affects how clearly the intended message reaches the viewer [16] [57]. Although charts and design choices are meant to convey data, users understanding can vary greatly depending on their visual complexity [32].

Overly complex visualizations can also affect the viewer by inducing feelings of overwhelm, reducing clarity, or even leading to misinterpretation of the information presented [42, 54]. Conversely, visualizations that are too simple tend to omit important details or fail to engage the user [2]. Finding a balance between simplicity and necessary complexity is therefore crucial to ensure that visualizations are both clear and informative, allowing users to accurately interpret the data while also remaining engaged. This issue is particularly important in domains where precise interpretation of the data is crucial. For example, visualizations of patient data in healthcare systems must convey complex information quickly and accurately. Unnecessary visual clutter or increased cognitive load can negatively affect the decision-making process or even lead to misinterpretations [4]. Similarly, professionals in finance and other business fields rely on dashboards to visualize their data. In this context, excessive VC can obscure important patterns and make decision-making less reliable [27].

Despite the work conducted thus far, a clear understanding of what drives perceived visual complexity in data visualizations is still lacking. The design of a visualization involves a large range of choices, from color and layout to the number of elements and their arrangement, all of which can influence how complex a visualization feels to the viewer. Studying VC is not just about aesthetics or design choices, but about improving how information is communicated and ensuring that it is accessible. Only by identifying the main factors contributing to perceived complexity can researchers provide design guidelines for creating data visualizations that are both engaging and comprehensible.

1.2. Visual Complexity versus Visualization Complexity

In a data visualization, many factors contribute to overall functionality and usability. While some factors are easy to distinguish, others are interconnected and difficult to consider independently. One example of this is *visual complexity* and *visualization complexity*, two terms that are closely related but refer to distinct concepts. Across several studies, the distinction between these two terms is barely noticeable. While the definitions of *VC* are formulated specifically with *visual complexity* in mind, those of *visualization complexity* often refer more generally to 'complexity'. For instance, Guo et al. [23] mention that the definition of complexity should be very close to the number of elements in the visual stimulus, their order of placement, and the difficulty in description. On the other hand, according to Zhu [58], a comprehensive discussion of *data visualization complexity* is still missing. Although he notes that 'complexity' is a general term without a precise definition, he proposes the following explanation: "the difficulty of creating, comprehending, and using that data visualization". Another perspective is provided by

Gärtner et al. [21] who conceptualize *visualization complexity* in terms of the cognitive operations required for users to complete tasks. Rather than expressing complexity as a single number, they propose describing it as a function of various variables, such as the number of items to compare or the task performed. Finally, as a concluding example, Ellemose and Elmqvist [15] distinguish between complexity, *visual complexity*, and *visualization complexity*. They define these concepts as follows:

- **Complexity:** A characteristic that arises when a system contains diverse components whose behaviors are interconnected, such that understanding one component may provide insight into others.
- **Visual complexity:** The level of intricacy present in a visual stimulus, influenced by factors such as the quantity of visual elements, their variation in appearance, their density and spatial organization, the diversity of patterns, and the degree of symmetry or irregularity within the scene.
- **Visualization complexity:** The complexity inherent in a data visualization's structure and information design, shaped by aspects such as the number of represented variables, the visual encoding techniques used, the amount of visual marks or glyphs, the layering of information, the scales employed, the variety of visual channels, and the relationships among visual elements.

While complexity describes a general structural property, *visual complexity* focuses on the objective visual appearance of a stimulus, and *visualization complexity* focuses on how the data within it is structured and represented. Because these concepts are closely related and often overlap, it is not always easy to study one without considering the other. Based on this, we take an exploratory approach to understanding how *visual complexity* is experienced by viewers of data visualizations, and work toward identifying a set of items that can be used to measure it. To guide this work, we address the following research questions:

1. Which terms and concepts do domain experts and the existing literature relate with perceived visual complexity in data visualizations?
2. How can these terms be refined and validated into a set of items suitable for measuring perceived visual complexity?
3. Which of these items are considered most relevant by domain experts for capturing the perceived visual complexity of a data visualization?

By addressing these questions, this thesis makes three main contributions:

1. A curated pool of 951 terms collected from a systematic literature review of IEEE VIS publications and a survey with 49 visualization experts.
2. A set of 49 scale items built from these terms.
3. A validated list of 16 items ranked by domain experts as the most relevant for measuring the perceived visual complexity.

Together, these contributions provide a foundation for future researchers and the data visualization community to build upon, with the goal of eventually developing a fully validated scale for perceived visual complexity.

2. Research Background

Data visualizations and the design choices that shape them play a crucial role in transforming complex datasets into meaningful and accessible visual representations. Over the past decades, researchers have studied how people perceive and process visual information, and how properties such as visual complexity influence understanding [19]. Together, these elements inform designs that balance simplicity, complexity, and clarity in data communication, though what that balance looks like in practice depends on the target audience [46]. More specifically, understanding VC is essential because it directly influences how effectively viewers interpret and extract insights from data visualizations. While significant work has been conducted in the field, very few studies have focused on visual complexity. Prior research has instead emphasized aspects such as graphical perception [9], cognitive load [27], and the evaluation of specific visualization types [26,43,52], leaving visual complexity largely understudied. This is likely because of the abstract and imprecise definition of VC. However, this presents an opportunity for us to dive deep into the topic and understand how VC manifests in data visualizations. This chapter gives an overview of research relevant to data visualization and visual complexity. It follows a bottom-up approach, starting from general definitions and key concepts of complexity before narrowing toward its role in data visualization. Finally, it reviews studies and methods used to measure and analyze visual complexity, providing a foundation for the present study.

2.1. Understanding complexity

Before exploring how visual complexity is explained in the context of data visualization, we found it useful to first consider the general meaning of the term *complex*. Dictionary definitions offer a useful starting point for identifying the main attributes associated with the concept and add another perspective to visual complexity.

According to the *Cambridge English Dictionary* [6], the adjective **complex** has several related meanings:

- "involving a lot of different but related parts"
- "difficult to understand or find an answer because of having many different parts"
- "having many parts related to each other in ways that may be difficult to understand"

Meanwhile, the *Collins English Dictionary* [10] defines it as:

- "made up of various interconnected parts; composite"

- “intricate or involved”
- “consisting of two or more related parts”
- “so complicated or intricate as to be hard to understand or deal with”

On a similar note, *Merriam-Webster Dictionary* [39] offers these interpretations of the adjective **complex**:

- “a whole made up of complicated or interrelated parts”
- “a group of obviously related units of which the degree and nature of the relationship is imperfectly known”
- “composed of two or more parts”
- “hard to separate, analyze, or solve”

Overall, these dictionary definitions share a common understanding of what **complex** generally refers to: something composed of multiple interconnected parts, often so intricate that it is difficult to fully understand. This information provides a conceptual foundation that can be adapted to better understand visual complexity in data visualizations, where similar definitions are relevant.

2.1.1. Visual Complexity: Literature Review

The concept of visual complexity has been examined since the early 20th century by Snodgrass and Vanderwart [47] and continues to be a topic of active research. Throughout this period, numerous definitions have been provided to help adapt the concept across diverse fields, including art, web design, and psychology. Snodgrass and Vanderwart, both with backgrounds in psychology, were among the first to address the concept of VC in 1980. From a cognitive psychology perspective, they described complexity as the amount of detail or the intricacy of the lines in the picture. They also identified aspects of VC that can be easily adapted to thinking about the complexity of data visualizations. When applied to data visualizations, this thinking becomes particularly relevant. One interpretation they offered is that complexity may result from the artists design choices rather than being an inherent property of the object itself. This highlights how design decisions directly influence the perceived VC of a visualization. The properties of the object being presented were also considered in their analysis. In other words, VC arises from both the intrinsic properties of the data and the conventions or design choices used to represent it. Many subsequent studies have used the definition proposed by Snodgrass and Vanderwart, making it one of the most commonly used explanations of visual complexity.

Their approach has been applied in numerous works, such as Allen and other related research [8, 15, 18, 23]. However, Ellemose and Elmquist [15] extend their view by considering VC not only as a structural property but also as a measure of the difficulty involved in understanding a visualization. Furthermore, they describe VC as an objective measure of the intricacy of any visual stimulus, quantifiable in terms of its fundamental compositional characteristics. These include the number of discrete visual elements, their variability, the density of these elements, their spatial distribution and arrangement, the diversity of patterns, and the level of symmetry or irregularity present. Similarly, Zhu [58] defines VC as the difficulty of creating, comprehending, and using that data visualization. Lazaro et al. [33] consider VC as the combination of different features (color, shape, and orientation) and its variations (e.g., color: red, green, white). In terms of visual search and target detection in a visualization, they explain that VC has been most closely associated with concepts related to quantity, with the assumption that larger numeric values correspond to a higher degree of complexity. Lastly, the work of Kim and Lombardino [31] provides insightful definitions of complexity from a slightly different lens. Even though their work focuses on comparing graphs and text, they state that there are two types of complexity that are particularly salient to understanding graphing: form complexity (content) and task complexity. Related to our topic, it is useful to describe their explanation of form complexity, which refers to the number of components that need to be processed simultaneously in working memory to solve a problem.

Allen [1] makes a distinction between VC as an objective property of an image and visual perception as the subjective processing of interpreting that information. He notes that the way the viewer receives and makes sense of the visual information plays an equally important role. Additionally, Harper et al. [24] similarly argue that VC is not purely an objective characteristic but is also shaped by the viewer's own perception. They suggest that two people looking at the same image may experience its complexity differently. This subjective experience directly reflects the cognitive effort that the image demands from the viewer.

Overall, the studies discussed above reveal that, despite the variety of definitions, the core idea of visual complexity centers on the quantity, arrangement, and diversity of visual elements that influence perception and comprehension. In essence, VC emerges as a multifaceted construct shaped by both objective visual characteristics and subjective cognitive interpretations.

2.2. Cognitive Load

Similar to visual complexity, cognitive load is widely applied across various domains such as education, humancomputer interaction (HCI), and data visualization. It not only helps us understand how people process and manage information but also guides us in making better design choices and building more effective systems. Even though cognitive load is not directly studied in this thesis, understanding it remains important: as visual complexity increases, so does the cognitive effort required to process a visualization. Understanding this relationship is essential for interpreting what makes a visualization complex and why that matters for the viewer. This section provides an overview of the key ideas behind cognitive load and its relevance to data visualization.

2.2.1. What is Cognitive Load?

To begin discussing cognitive load, it is important to first understand its meaning. John Sweller, a well-known researcher in educational psychology, first proposed the Cognitive Load Theory in the 1980s. From his highly significant work in the field, three main concepts are particularly relevant to this thesis [49]:

- **Intrinsic Cognitive Load:** describes the inherent difficulty of the material being learned. This load depends on both the complexity of the information and the learner's prior knowledge. As expertise increases, learners are able to group related pieces of information into larger mental structures (schemas), which reduces the effective cognitive demand.

- **Extraneous Cognitive Load:** arises from the way information is presented rather than from the content itself. Ineffective instructional design, such as redundant explanations or poorly organized material, can force learners to process unnecessary information and thereby interfere with learning.

- **Element Interactivity:** describes the extent to which multiple pieces of information must be considered together in working memory. When many elements depend on one another, the task places greater demands on working memory and becomes more cognitively challenging.

Gerjets and Scheiter [22] discuss another type of cognitive load known as **germane cognitive load**. This refers to the mental effort that learners invest in processes directly related to learning. According to Sweller et al. [51], when a task does not fully use a learners cognitive resources, the remaining capacity can be devoted to useful activities such as schema construction (i.e., the process of forming and organizing knowledge structures in long-term memory that help learners recognize patterns and solve problems more efficiently). These processes also increase overall cognitive load, but unlike other types, germane cognitive load contributes to learning rather than interfering with it. It results from helpful cognitive activities such as elaboration, abstraction, comparison, and inference, which are encouraged by the instructional design. As Gerjets and Scheiter explain, these deeper mental processes further support schema construction by engaging higher-level thinking beyond simply holding information in working memory.

Later, in 2024, Sweller [50] explains that Cognitive Load Theory is an instructional theory based on five main assumptions. He states that people process two types of information: biologically primary information, which we acquire naturally, and biologically secondary information, which requires conscious learning. The theory also points out that working memory has a limited capacity, while long-term memory stores knowledge that can be reused when needed. Learning occurs through the accumulation of information in long-term memory, which leads to differences in expertise. Finally, Sweller notes that learners with different levels of prior knowledge benefit from different instructional methods, a concept known as the expertise reversal effect. Additionally, Paas [41] explains that cognitive load is a multidimensional concept, with two components (mental load and mental effort) that can be distinguished. Mental load is imposed by instructional parameters (e.g., task structure, information sequence), and mental effort refers to the capacity allocated to these demands. Changes to instructional design that increase mental load will only work if learners are motivated and put effort into the task. Therefore, the amount of effort they invest is used as a measure of cognitive load.

2.2.2. Cognitive Load in Data Visualization

The concept of cognitive load has been studied across various domains, including data visualization. The work done over the years provides a foundation for us not only to make better design choices but also to understand how users perceive, interpret, and interact with visual information. Related to this, Sweller [49] has explained, using his theory, that for better learning outcomes, materials should be presented in alignment with cognitive processes. Limiting the amount of visual complexity in an image not only reduces the cognitive load but also increases learner efficacy. This can be done by combining related textual and visual components, eliminating redundancy, and simplifying complex tasks to ensure that learners can focus on meaningful processing. Furthermore, Sweller highlights that cognitive load does not depend solely on visual complexity but also on the learner's prior knowledge. Novice users learn more effectively from clear and simplified visuals, while more experienced users find it easier to process and decode more complex visual representations. For that reason, effective data visualization design should balance intrinsic and extraneous cognitive load, adding complexity only as users become more comfortable with the content. In another study, Snodgrass and Vanderwart [47] demonstrate that visual complexity directly affects the speed of cognitive processes and the efficiency of recognition, naming, and categorization tasks. During their experiment, they found out that complex visuals require more mental effort to interpret, supporting the idea from cognitive load theory that minimizing unnecessary visual details can improve information processing and learning efficiency. Before moving on to the next study, it is important to first explain Mayer's Cognitive Theory of Multimedia Learning [38]. In this work, he explains how people learn from words and pictures by outlining three assumptions:

- **Dual channels:** Mayer proposes that people process visual and verbal information through separate cognitive pathways. In the context of the theory, early processing depends on sensory input, whereas later processing is influenced by presentation mode. For instance, written text is first perceived visually before being mentally transformed into verbal or auditory representations.
- **Limited capacity:** The theory assumes that each cognitive channel can only handle a restricted amount of information at a given moment. Consequently, presenting excessive visual or auditory material may overload the learners processing capacity.
- **Active processing:** Learning is viewed as an active cognitive process in which individuals focus on relevant information, organize it into coherent mental structures, and connect it with prior knowledge. Mayer further argues that learners actively

select, structure, and integrate incoming words and images within the limits of each processing channel.

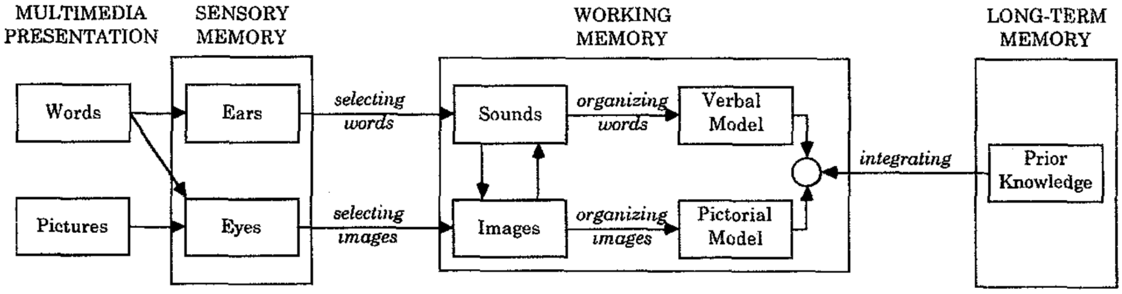


Figure 2.1.: *A cognitive theory of multimedia learning.* [38]

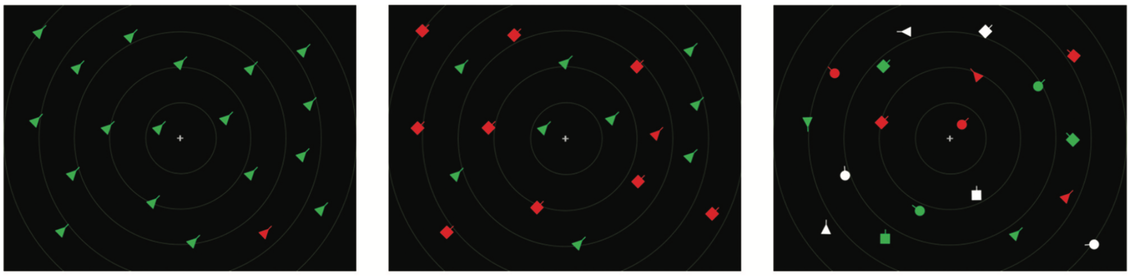
Building on this framework, Allen [1] studied the relationship between cognitive load and visual complexity, mainly based on the Cognitive Load Theory and the Cognitive Theory of Multimedia Learning. He explains that removing extraneous visual information, such as background clutter or irrelevant details, reduces the number of elements in an image and their interactivity. Therefore, both visual complexity and extraneous cognitive load are reduced. This, in turn, frees up cognitive resources needed to process and integrate meaningful information. However, removing too much information can make the image unclear and increase cognitive load. Additionally, from the study, the author concluded that a Goldilocks effect [30] is possible, focusing on visual information. This concept concerns the idea of having *just the right amount*, a term used in cognitive science and developmental psychology. In Allen's work, it was found that medium-visual-complexity images resulted in better cognitive load management and learning outcomes than both high- and low-visual-complexity images. As a result, this suggests the presence of a Goldilocks effect, where an optimal level of visual complexity balances necessary detail with manageable cognitive load. To summarize, he notes that designers should carefully balance visual complexity in educational materials. Removing extraneous details can help, but essential information should be preserved to avoid increasing cognitive load.

2.3. Clutter

Another important factor affecting the way users understand and interpret data visualizations is clutter. Together with cognitive load, it has been studied both jointly and independently and has consistently been shown to be closely related to complexity. Rosenholtz et al. [44] describe clutter as a condition where the amount or arrangement of visual elements negatively affects a person's ability to perform a task. Excess and/or disorganized display items can cause crowding [48], masking [35], decreased object recognition performance due to occlusion, and impaired visual search performance. However, adding more items to the visualization can also improve performance in cases where the items are low-entropy, meaning the appearance of one item can be easily predicted from its neighbors. As explained by Rosenholtz et al. in the paper cited above, it can be easier to spot a trend in data when there are more points or to notice a grouping of items with similar characteristics. Aside from this, they introduce the Feature Congestion measure, which is based on the analogy that the level of visual clutter in a display or scene is related to the ease (or difficulty) of adding a new item that reliably draws attention. A key advantage of this measure is that it can break down clutter, showing which specific elements in the image contribute most to it. Related to this, recent work by Chu et al. [8] investigates what makes visualization images visually complex by systematically combining human perceptual data with objective, imagebased metrics. In their largescale study, they asked hundreds of participants to rate the visual complexity of various visualization images and compared these ratings with twelve computational metrics, including edges, corners, colors, clutter, and object features. Their results show that simple features, such as the number of corners and colors, as well as more complex measures, such as feature congestion, strongly influence how complex a visualization appears. For certain visualization types, edge density also serves as a strong explanatory factor. Additionally, they also observe a nonlinear relationship between text annotations and perceived complexity: adding moderate amounts of text can initially reduce complexity, but excessive text eventually increases it again. This study provides a clear, data-based understanding of how different image features affect the complexity of a visualization, showing how computational measures and human perception are connected and how features can work together to influence complexity. Similarly, Lazaro et al. [33] have studied how VC and clutter can significantly affect visual search and target detection. They define clutter as a state in which the presence, appearance, or organization of excess items causes confusion and reduces task performance. In relation to this, they explained that decluttering methods, such as dimming, dotting, reducing size, or removal, are designed to make irrelevant or less important objects less visually prominent, or even eliminate them, thereby reducing visual competition for attention. For example, Figure 2.2 shows three cockpit display templates illustrating the low, medium, and high visual complexity levels from the 60 templates used in the experiment conducted by Lazaro et al. [33]. They were designed to test how different levels of visual complexity affect the identification and interpretation of targets in aircraft displays. The variations in color, shape, and orientation convey critical information, such as target identity, type, and direction, for cockpit decisions. The experiment's results were not far from what the authors had anticipated; task performance and perceptions were strongly affected by visual complexity levels. Accuracy decreased with increasing visual complexity, and response times were 101% longer at higher visual complexity than at lower visual complexity.

2.4. Measuring Visual Complexity

Beyond the foundational concepts we discussed so far, researchers have developed more concrete methods to measure visual complexity in data visualizations. These methods



(a) Low Visual Complexity (b) Medium Visual Complexity (c) High Visual Complexity

Figure 2.2.: *Designed cockpit displays with three different levels of visual complexity (low, medium, high)* [33]

aim to connect objective visual properties with how complex a visualization feels to users. They differ in both theory and computation, but recent research highlights two main approaches: one analyzes how visual elements are arranged and related to one another using formal rules and models, and the other measures complexity by treating visual patterns as signals and applying time-series methods to estimate their unpredictability.

Duffy et al. [14] use Allens interval algebra, originally developed to describe time intervals, to measure visual complexity in cluster-based visualizations such as scatter plots and parallel coordinates. In this approach, complexity is understood as visual uncertainty, arising when the positions of clusters are unclear or ambiguous. The method examines all possible ways two clusters can relate to each othersuch as overlapping, touching, or being separatealong the x- and y-axes. These many possibilities are then simplified into a smaller set of basic patterns (24 for scatter plots, 35 for parallel coordinates). Each pattern gets a complexity score based on features like overlapping areas, split shapes, or touching edges. Fully overlapping clusters get the highest complexity, while fully separated clusters get the lowest. This provides a clear, rule-based way to measure the ambiguity of relationships between clusters, and it can be extended to multiple clusters by averaging the scores across all cluster pairs.

On the other hand, Ryan et al. [45] study line chart complexity using ideas from information theory that measure regularity and unpredictability. One specific measure, Pixel Approximate Entropy (PAE), calculates how "disordered" or noisy a line chart appears on screen. PAE is based on approximate entropy, a statistic from time-series analysis that measures how likely patterns of a given length are to repeat. Applied to line charts, smooth and predictable lines (like a straight trend) get low PAE, while jagged, noisy lines get high PAE (Figure 2.3). This measure is designed to match how people perceive complexity when glancing at a chart quickly. The method works by creating charts with varying levels of synthetic noise, measuring their PAE, and confirming, through crowdsourced experiments, that higher PAE corresponds to higher perceived complexity and lower accuracy on tasks such as detecting changes or identifying shapes. Overall, this approach provides a practical, data-driven way to quantify the difficulty of interpreting unpredictable line charts.

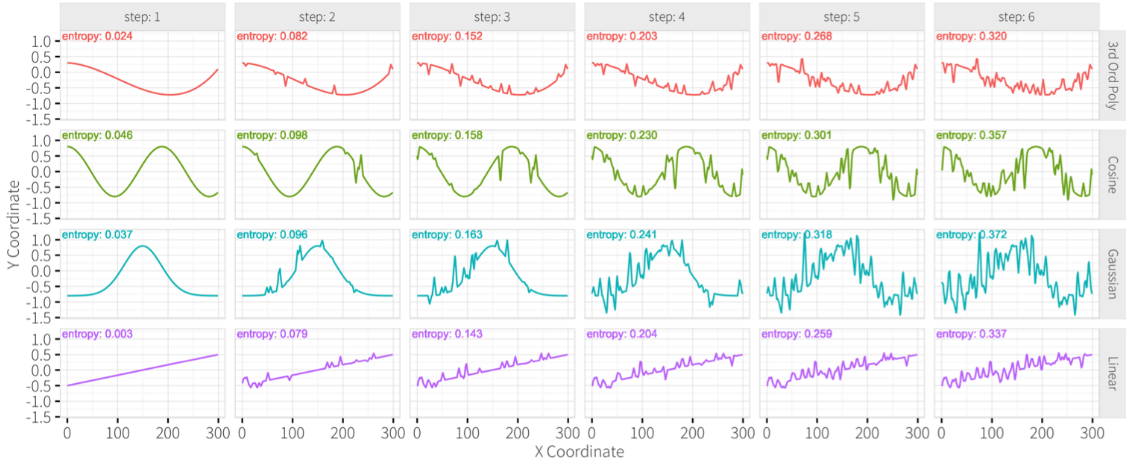


Figure 2.3.: *Visualization of four line charts (3rd-order polynomial, cosine, Gaussian, and linear) with increasing noise from left to right, showing approximate entropy as a measure of perceptual complexity.* [45]

Building on these methods, Ellemose and Elmqvist [15] developed a more general framework for measuring visual complexity across different types of visualizations. In a crowdsourced study with 289 participants rating 640 visualizations, they found that four common computer vision metrics (Shannon entropy, meaningful image complexity, feature congestion, and subband entropy) had very weak correlations with human judgments of complexity. This shows that low-level pixel-based measures, while sometimes useful, do not capture the meaningful patterns and design elements that shape how people perceive chart complexity. To better predict perceived complexity, they ran two experiments. First, they manually coded 32 features for each visualization, including chart type, use of 3D, number of data series, presence of legends, and layout, and used multilinear regression, which explained 49 percent of the variance in human ratings. Second, they tested a vision-enabled large language model, GPT 4o mini. The LLM successfully extracted most features and produced complexity scores that were strongly aligned with human ratings, demonstrating a scalable, effective alternative to manual coding. These approaches show a shift from specialized, rule-based metrics to more general, AI-assisted methods. While Duffy et al. [14] and Ryan et al. [45] focus on measures designed for specific types of charts, Ellmose and Elmqvist [15] show that assessing complexity in a way that matches human perception requires understanding the overall design of a chart, which modern LLMs now make more feasible. This new set of tools not only improves how we measure complexity but also helps create visualizations that are easier for people to understand.

3. Methodology

This chapter provides a detailed explanation of the steps involved in developing a scale to measure perceived visual complexity in data visualizations—from the initial collection of terms to the validation of the final set of items. The approach followed was similar to that of Cabouat et al. [5] and He et al. [25], who developed scales to measure perceived readability and the aesthetic quality of data visualizations, respectively. We also consulted with the authors of Cabouat et al. [5] through several meetings to clarify certain aspects of their methodology and ensure that our approach closely aligned with theirs.

3.1. Term identification

The first step in our process was generating a pool of terms to assess perceived visual complexity in data visualizations. There are two main approaches for identifying suitable terms for scale items: deductive and inductive [3]. The deductive approach systematically reviews existing literature to identify terms used in related studies, while the inductive approach gathers new terms directly from domain experts through surveys and open-ended questions. In this project, we applied both methods and created 5 different pools of terms, as shown in Table 3.1.

3.1.1. Deductive method

We conducted a literature review on IEEE VIS publications from 1990 to 2024, as well as IEEE TVCG and IEEE CG&A articles published at IEEE VIS [28]. The dataset contains metadata for 3754 papers, downloaded in batches by querying their DOIs in Zotero [11]. Then, by using a Python script, we found 280 papers in which both words *complex* and *Likert* were present. From these papers, we derived two sets of terms: those used in the study questionnaires and those extracted from reported participant comments. Each paper was reviewed individually, and we recorded terms used by the authors in Likert-scale questionnaires or surveys, as well as terms appearing in participant feedback. We retained only terms related to visual complexity and excluded those referring to tasks, task completion, overall system properties, or the interactivity of the visualization system. Terms related to our study subject were also excluded, such as *visual complexity*, *complex*, etc. Including these would not provide any meaningful insight, as our goal is to identify the factors and dimensions that contribute to the perceived visual complexity rather than the concept itself. Additionally, terms like *complex* are too broad and general to be useful as scale items on their own. More information about the deductive method can be found in section A.1 of the appendix.

method	corpus	sources	unique terms
deductive *literature review: 166 publications*	questionnaires in studies (Pool 1)	121 studies	227
	reported participants comments (Pool 2)	133 studies	380
inductive *expert survey*	statement collection (Pool 3)	50 experts	224
	term collection (Pool 4)		166
	adjective collection (Pool 5)		66
Sum of collected unique terms			**951**

Table 3.1.: Summary of terms collection.

Terms from study questionnaires (Pool 1) For this pool, we focused on terms used by the authors in Likert-scale questionnaires or surveys. We also included terms from questions addressing visual complexity in the methods, results, or discussion sections. In total, we collected 347 terms, of which 227 were unique, from 121 publications. Additional information is included in section A.1 of the appendix.

Terms from participant comments (Pool 2) For this set, we extracted terms from participant comments and other forms of feedback. Similar to Pool 1, we included only terms related to visual complexity that appeared across various sections of the papers. In total, we obtained 553 terms, of which 379 were unique, from 133 publications. Further details can be found in section A.1 of the appendix.

3.1.2. Inductive method

To incorporate expert input alongside our literature review, we conducted a survey, which we also pre-registered (AsPredicted).

Participants: We invited 148 visualization experts to participate in our survey via direct email. Participants were selected based on our knowledge of their work, and in consultation with the authors of Cabouat et al. [5]. We did not compensate participants for taking part in the survey and waited approximately 5 weeks to collect responses. During this time, 50 submissions were received. Of these, 49 were valid, and one was excluded due to not providing any valid responses to the survey questions, instead using the response fields to cite previous research and express methodological concerns about the use of Likert scales for measuring abstract concepts such as visual complexity.

Participants demographic details, including age, gender, and experience, are provided in Table 3.2.

Characteristic	Category	Nr./%
Age Group	18-24	1 (2%)
	25-34	10 (20%)
	35-44	17 (34%)
	45-54	11 (22%)
	54-64	11 (22%)
Gender	Male	31 (62%)
	Female	16 (32%)
	Non Binary	1 (2%)
	Prefer no answer	2 (4%)
Years of Experience	Mean	18.41
	Range	538

Table 3.2.: Demographic Characteristics of Survey Participants (N = 50)

Procedure: At the beginning of the survey, we provided a brief description of the aim of our project. Participants were then asked to review the consent form and complete background questions regarding their gender, age group, and years of experience. In the next step, we presented 15 data visualizations representing different chart types, each including several elements that may influence the visual complexity of a data visualization (see Figure A.1 in the appendix). These images served as reference points to support experts in reasoning about visual complexity and to provide concrete examples for discussion. All visualizations were drawn from the VIS30K dataset [7]. Experts were then asked to briefly describe their understanding of visual complexity. Specifically, they were required to provide at least one statement by completing the sentence: A data visualization is visually complex when ____. We started the survey with this open-ended question to gather their honest opinions without guiding or influencing their answers.

In the next step, we collected additional statements that were later used to develop the scale terms. We displayed the same set of images and clarified the study's scope. Specifically, we emphasized that our interest lies in participants perceptions of the visual complexity of a single, static data visualization, considering only its visual appearance. In a similar way as before, experts were asked to provide at least three statements to complete the following sentences: This visualization (is/has/) ____. and The visual complexity of this visualization makes me feel ____. The first one will give us terms in the appropriate format for the scale. The second sentence was included to collect adjectives describing participants emotional responses to the visual complexity of a data visualization. This approach lets us both find descriptive terms for the scale and explore how visual complexity affects viewers emotions and reactions. Experts were also given the opportunity to provide additional comments after submitting their suggested items. Supplementary information can be found in Section A.2 of the appendix.

Results: We received 50 responses in total, but one submission was not considered because the expert did not provide any correct answers to the questions; instead, they cited previous research. Using the collected data, we created three additional pools, one for each question in the survey.

Terms from the experts definitions (Pool 3): For this pool, we reviewed the experts' sentences describing their understanding of visual complexity. From each sentence, we identified the key words and recorded them as terms, which could later be used to develop the scale. With this process, we collected 236 terms, of which 224 were unique. See Table A.4 in the appendix for additional details.

Terms from the experts statements (Pool 4): This pool was created from the statements experts provided to complete the sentence: This visualization (is/has/) ____. Since the answers were already in the desired format, we only corrected minor typos and separated each statement into individual entries. In total, we collected 196 terms, of which 166 were unique. Additional details are provided in Table A.5 in the appendix.

Terms from the experts adjectives (Pool 5): Similar to the previous pool, this one was created with the adjectives provided by the experts for completing the sentence: The visual complexity of this visualization makes me feel ____. Again, the terms were already in the appropriate format for the scale, so no additional processing was needed. In total, 189 terms were collected, of which only 66 were unique. Many terms were closely related, often simple variations or synonyms of each other. Details are reported in Table A.6 in the appendix.

3.2. Item generation

The purpose of this step was to generate items that would later be used in the scale. As also explained by Cabouat et al. [5], an item is a combination of a statement or question and its accompanying answer options. To start, we combined all five pools and retained only unique terms, yielding a list of 951 distinct terms. In the sections below, we will explain in detail our approach and how we went from this initial list to the final items.

3.2.1. Conceptual Clustering of Terms

As mentioned earlier, our list of unique terms was very long, making it difficult to analyze directly. For this reason, we needed to determine a method to reduce the number of terms. First, we calculated the *Pool Count* for each term, which is simply the number of pools (out of five) in which the term appears. This does not reflect how many times the term appears in each pool, but rather how widely it is "covered" across them. Next, we sorted the terms by *Pool Count* from highest to lowest and used a bottom-up approach to create *Groups of Related Terms*. Starting with terms at the bottom of the list (with lower *Pool Counts*), we checked whether they were related in some way to any of the terms higher up. If so, we clustered them together.

These groups were created not only with simple morphological variations of a term, but also with terms that refer to the same concept, especially in the context of data visualization. We also tried using clustering methods such as K-Means and X-Means for this process. Nevertheless, the results were not satisfactory. The embedding model and clustering algorithms were unable to capture relationships between terms from the users perspective. In particular, some terms can belong to more than one *Group of Terms*, whereas clustering methods assign each term to a single cluster. For this reason, we found that a manual grouping process was more appropriate and better reflected the way users relate these terms. Each group has a representative term that served as a reference point for evaluating whether lower-ranked terms were related to higher-ranked terms. Additionally, every cluster consists of terms that have a smaller or equal *Pool Count* relative to its representative term. A term could also belong to more than one cluster if it is conceptually related to multiple groups, reflecting how users might interpret or use it.

Semantic Matching of Terms

At the end of this process, we were left with a set of terms that seemed not to be related to any of the existing groups. All these leftover terms had a *Pool Count* of 1 or 2. To avoid leaving them unclustered, we attempted to assign them to the existing *Group of Terms* based on their semantic meaning. This was done using all-MiniLM-L6-v2, a pre-trained sentence embedding model. Based on all the clusters we had up to that point, we tried to match these leftover terms to the most semantically similar existing cluster using cosine similarity scores from the embeddings. This value determines whether the term belongs to an existing cluster, should form a new cluster, or even be excluded. After reviewing the results, we only assigned terms to existing clusters when the match seemed meaningful, leaving the others unassigned.

Pool Coverage

We performed a step to further refine our large pool of terms. Following a similar approach to Cabouat et al. [5], for each *Group of Related Terms*, we identified in which of the five pools the terms in the group appeared. For example, if a group contained multiple terms present in different pools, the groups overall coverage reflected all the pools where at least one term appeared. We retained only those clusters that appeared in at least two different pools, with at least one term from the deductive method (Pool 1 and Pool 2) and at least one from the inductive method (Pool 3, Pool 4, and Pool 5). This ensured that each *Group of Related Terms* reflected both perspectives: those of the users and the experts. As a result, we excluded 8 clusters as their terms were present only in Pool 1 and Pool 2.

Outcome

After completing the last two steps, some terms remained unassigned to any cluster. We decided to exclude them from our big pool for two main reasons. First, they did not clearly belong to any existing cluster, and the suggestions from the semantic matching step were not meaningful. Second, all of these terms had a *Pool Count* of 1, indicating they had very low representation in the overall pool. As a result, 106 terms were excluded from our big pool. In addition, we removed a cluster whose primary term was "complex", as all terms in the cluster contained either "complex" or "complexity". Including it would have been redundant and would not have provided any meaningful insights, as our goal is to identify the factors that contribute to visual complexity. At the end of this process, we have:

- 43 clusters & 43 representative terms
- 8 clusters excluded due to low pool coverage
- 106 terms excluded for not matching any cluster
- 1 cluster excluded due to redundancy

3.2.2. Selection of Primary Terms

In this step, we identified the *primary terms* that would serve as the foundation for constructing the scale items. Because the *Groups of Related Terms* were relatively large, we needed a clear and systematic way to select one term from each cluster. To guide this selection, we discussed the options and followed the suggestion provided by Cabouat et al. [5]. For each term in our big pool, we calculated its weighted frequency, which indicates how common it was across all data pools. This was done by first normalizing its occurrence within each pool by the total number of terms in that pool, and then averaging these values across all pools. Expressed as a percentage, the weighted frequency allows us to compare terms fairly, without letting larger pools dominate the results. The weighted frequency for a term t is calculated as :

$$\mathrm{WF}(t) = \left(\frac{1}{P} \sum_{p=1}^{P} \frac{c_{t,p}}{S_p} \right) \times 100$$

where:

- P is the total number of pools
- $c_{t,p}$ is the count of term t in pool p
- S_p is the total number of terms in pool p

This formula shares some conceptual similarity with TF-IDF, a well-known measure in information retrieval that also normalizes term frequency across multiple documents. However, the two differ in an important way: while TF-IDF penalizes terms that appear frequently across many sources by applying an inverse document frequency component, our weighted frequency does not. In this context, a term appearing consistently across multiple pools is considered more representative and reliable, not less. The goal here is not to identify unique or distinctive terms, but rather those that are most broadly shared across the different sources and perspectives collected in this work. Following this approach, a term that appears consistently across multiple pools will receive a higher weighted frequency than a term that appears many times in just one large pool. After calculating the weighted frequency for all terms, we selected, for each *Group of Related Terms*, the term with the highest value as its representative. This term was identified as the *primary term.* During this process, we encountered a few minor issues, which we had anticipated in our initial discussion. First, because some terms were assigned to more than one cluster, they appeared as *primary terms* for multiple *Groups of Related Terms.* Second, some clusters had multiple *primary terms* because several terms within the cluster had the same weighted frequency. This situation occurred mainly in clusters with a *Pool Count* of 1. In both cases, we decided not to make any changes and postponed the decision to the next phase of item generation.

3.2.3. Constructing Scale Items

At this stage, we had identified our set of *primary terms*, which would serve as keywords for constructing the scale items. For simplicity, we will also refer to these items as statements throughout this section to avoid unnecessary repetition. A key decision in writing the statements was determining their format. Following the suggestion of Cabouat et al. [5], we decided that each item should begin with either *"I find"* or *"I can".* In their study, they found that phrasing items in the first person encourages participants to provide answers that reflect their own experiences and perspectives. Without this phrasing, users tend to respond more generally, considering how others might answer rather than expressing their personal opinion. We decided to keep the same format while also adding the option *"I feel"*, since adjectives had been collected in the expert survey. From this point, we proceeded to form the items using the primary terms.

As mentioned in Section 3.1.2, Pool 3 contains all data collected from the experts definitions of visual complexity. In addition to the extracted keywords, we retained their full responses as a reference when formulating the statements. These responses provided important contextual information that supported the development of the items and helped ensure that their wording remained faithful to the experts intended meanings. Following this, we ensured that each *primary term* was used in the same context in which it had been used by the experts in their responses. However, not every *primary term* was found in Pool 3. In such cases, we checked whether any term from the same cluster appeared in Pool 3. If a term from the cluster was found, the item was constructed based on a concept consistent with the experts definitions. If no such term was found, the item was constructed solely based on the *primary term.* Sometimes using the *primary term* directly was difficult because of the item format we needed to follow. This was particularly true for nouns such as *familiarity* or *clarity.* Since the main context of the statement remained unchanged, and to simplify the formulation, we replaced the noun with the corresponding verb or adjective form when available. In practice, this often meant using the second-highest term, which was usually the verb or adjective version of the same concept. For *Groups of Related Terms* containing more than one *primary term*, we generated items for each term and left the decision of removing any items to the next phase. At the end of this process, we obtained **43** clusters, **51** primary terms, and **49** unique statements. More information about the *primary terms* and the generated statements can be found in Table A.7 in the appendix.

3.3. Item validation

As also explained by Boateng et al. [3], once the scale statements are generated, it is important to ensure they actually measure the concept we intend. In their work, they describe this step as *content validity* and state that it should involve subject-matter experts (SMEs) to evaluate the statements. In our case, domain experts not only have a thorough understanding of visualization concepts but also employ the final measurement instrument to evaluate how visualizations are perceived by lay users. For this reason, we conducted a study that was also pre-registered (AsPredicted).

Participants: We invited 145 visualization experts to participate in our survey via email. The participants were drawn from the same group that had completed the first survey. This was done in line with the work of Cabouat et al. [5] to ensure consistency across the study's different stages and to enable participants to evaluate the revised items based on their prior involvement and familiarity with the construct. This time, we also did not compensate them for taking part in the survey and waited approximately 4 weeks to collect responses. During this time, 45 submissions were received, all of which were valid. Participants demographic characteristics, including age, gender, and experience, are reported in Table 3.3.

Characteristic	Category	Nr./%
Age Group	18-24	0 (0%)
	25-34	7 (15%)
	35-44	11 (24%)
	45-54	14 (31%)
	54-64	8 (17%)
	65 +	5 (11%)
Gender	Male	30 (66%)
	Female	13 (28%)
	Diverse	1 (2%)
	Without gender entry	1 (2%)
	I prefer not to answer	0 (0%)
Years of Experience	Mean	21.44
	Range	3-40

Table 3.3.: Demographic Characteristics of Survey Participants (N = 45)

Procedure: When participants began the survey, they were first presented with a consent form that briefly described the tasks they would complete in the following steps. They were then asked to provide background information, including gender, age group, and years of experience. In the next step, participants were shown six data visualizations that served solely as stimuli for the survey. These images were selected from those included in the first expert survey, with the addition of one black-and-white visualization to account for the effect of color on perceived visual complexity [8] (see Figure A.2 in the appendix). Participants were then instructed to rate our items according to how relevant they considered each for assessing a participants perception of the visualization's visual complexity. Specifically, they were asked to imagine participants using these items to evaluate the visual complexity of a visualization. Our statements were displayed in a five-page table, with response options ranging from "not at all relevant" (1) to "very relevant"(5). Similar to the first expert survey, we clarified that our goal was not to assess the overall system's complexity, but rather to focus on the static view of a single visualization. As also explained in Subsection 3.2.3, we included only unique items in the survey. However, statements with a similar context were intentionally retained to allow comparison of their scoring results and to evaluate their effectiveness. Experts were again given the opportunity to leave comments or suggestions related to the survey.

Results: We received 45 responses and did not exclude any, as all of them were valid. Before describing our approach to analyzing this data, we would like to clarify some points that were raised by experts in the survey's comment section (see Section A.4.1 of the appendix). One common concern was the use of both positive and negative statements, which sometimes caused confusion. As mentioned earlier, this was intentional and based on input from experts in the first phase of the project and the wording they used. The statements were not changed to maintain consistency and to evaluate them as originally provided by the experts and the literature review. Additionally, some mentioned that certain terms (e.g., unnecessary complexity, occludes information, poorly designed) used in the statements were unclear, making it difficult to rate the items. A similar reasoning applies here: the wording was kept as provided, and during the evaluation phase, we will identify which statements are more or less suitable for inclusion in the scale.

Aside from the survey comments, we also received feedback from one of the experts via email (see Section A.4.1 of the appendix). The main issue reported there was related to not clearly distinguishing between *perceived visual complexity* and *high perceived visual complexity*. Even though this was not done intentionally, it is worth clarifying that the items are not designed to measure only high perceived visual complexity. Rather, they are intended to capture perceived visual complexity across the full spectrum, from low to high, meaning that both simple and complex visualizations can be evaluated using the same set of items.

For better structure and organization, we describe the data analysis process in the following section.

3.3.1. Reverse Coding

In our survey, both negatively and positively worded items were included, as outlined above. García-Fernández et al. [20] and Vigil-Colet et al. [56] describe several ways to deal with this situation. In our case, the items matched the experts' original wording, so we decided to make the necessary change after collecting the data. This decision is in agreement with the approach of PREVis [5]. Aside from the methods discussed by García-Fernández et al. [20], we chose reverse coding, as it was the most suitable for our data analysis. The idea behind this process is to ensure that all Likert scale scores point in the same direction. This means that a high score should always be associated with "good" quality, and a low score with "bad" quality. If reverse coding is not performed, this consistency is not guaranteed: a high score (5) would indicate a positive evaluation for positively worded items, but the same high score would indicate a negative evaluation for negatively worded items. The first step was to identify which of our 49 items were negatively worded so that their scores could be reversed. An example of this process is shown in Table 3.4. Next, while iterating over the experts' answers, we reversed the scoring of all negatively worded items. We also kept the original values throughout the process to check that everything was done correctly and that the reversed scores were properly aligned. To further examine the collected data, we continue with a more detailed explanation in the following section.

original	1	2	3	4	5
reversed	5	4	3	2	1

Table 3.4.: Reverse of Likert scale scoring

3.3.2. Cumulative Link Mixed Models (CLMMs)

To analyze our data, we needed to decide on a method that was adequate for our data and the way it was collected. The purpose of this step is to identify which items have received the highest scores, thereby determining the most relevant for evaluating the visual complexity of a data visualization. A simple average value of the scores for each item would not be correct, as data retrieved from a Likert scale requires multiple aspects to be taken into consideration, such as the ordinal nature of the responses, the unequal distances between scale points, and the variability between individual participants. For this reason, we based our decision on the findings of Syiem and Velloso [55]. As they explain in the paper, ordinal data carry no metric information. Therefore, the distance between ordinal levels may not be equal, and the labels are just categorical with known ordering. This means that the levels on a Likert scale, such as 1 to 5, can be easily replaced with any other set of ordered categories. One common way to analyze data collected from a Likert scale is to combine the responses to individual items by summing or averaging them before performing further analysis. However, this treats the data as metric and is an approach that has been criticized by several researchers [29, 36, 37]. Approaches such as averaging ordinal data assume equal intervals between scale points, which can lead to distorted estimates that do not reflect the true differences between responses. As a result, the underlying relationships between items may be misinterpreted, potentially leading to incorrect conclusions.

For this reason, Syiem and Velloso [55] suggest using cumulative link models (CLMs) or cumulative link mixed models (CLMMs) when appropriate, as they are specifically designed to handle ordinal data. These methods treat ordinal responses as ordered categories rather than numerical values, while still leveraging the inherent ranking among them. By doing so, CLMMs avoid imposing unrealistic assumptions about equal distances between scale points and instead model the probability of responses across categories, leading to more accurate and reliable insights. A similar point is made by Taylor et al. [53], who argue that CLMMs are a more appropriate method for analyzing ordinal data. They do not treat responses as numbers with equal distances and also include random effects, which makes them suitable for real experimental data. With this in mind, we chose to use CLMMs, as they are the most suitable method for our data. Unlike CLMs, CLMMs include both fixed and random effects. Fixed effects represent the main variables of interest, which in this case are the items, while random effects account for variability across groups, in this case, the experts. This makes CLMMs especially useful when responses vary across individuals, enabling more accurate analysis of ordinal data. They are also more suitable than CLMs for larger datasets with many participants, such as our 49 experts, because they can properly account for this variability.

We first implemented the code by following the guidelines in a technical note at *useR!*. We ran the analysis in RStudio, and the full code is provided in Section A.4.3 of the appendix. The fixed effect in our study are the items and the random effect, the experts who participated in the survey. The CLMMs are fitted in R as shown in Figure 3.1, while accounting for these effects, the rating score given to the item, and the full dataset.

```
model <- clmm(rating ~ item + (1 | userID), data = df)
```

Figure 3.1.: *Fitting the CLMMs model in R*

Next, pairwise comparisons are computed using *emmeans* (*documentation*). This results in 2352 comparisons between all items, where each item is compared with every other item except itself. At this step, the Bonferroni correction is applied to reduce the risk of false positives, which can be high when running many comparisons simultaneously (see Figure 3.2).

```
results <- emmeans(model, pairwise ~ item, adjust = "bonferroni")
```

Figure 3.2.: *Pairwise comparison with Bonferroni correction*

Understanding the output

To explain how the model works, we will provide an example of the output below. The main idea is that the model not only considers the observed scores, but also accounts for how participants use the rating scale. Some participants may consistently give higher or lower scores than others, and this individual response tendency is taken into account. In addition, to identify the most relevant items for evaluating visual complexity, the model compares items and estimates the probability that one item will receive higher ratings than another. In Figure 3.3 we present only one of these pairwise comparisons.

contrast	estimate	SE	df	z.ratio	p.value
q1 - q4	-1.649832686	0.3940567235	Inf	-4.186789839	0.03327228254

Figure 3.3.: *Example output of CLMMs*

- The column *contrast* shows the items being compared. In this example, we are comparing item 1 with item 4.
- The column *estimate* represents the difference in log-odds between items 1 and 4. It shows how the likelihood of receiving a higher rating changes when comparing item 1 to item 4. A positive value means item 1 is more likely to receive higher ratings than item 4, while a negative value means the opposite. Note that a higher rating indicates a better fit for evaluating visual complexity. In this example, a value of **-1.649** means that item 1 has been generally rated lower than item 4. Unlike probabilities, which are limited to values between 0 and 1, log-odds can range from $-\infty$ to $+\infty$, which allows them to express both small and large differences.
- The column *SE* stands for the standard error, and it indicates how precisely the estimate has been measured. It reflects the variability of the estimate across the data, meaning how much the estimate would be expected to change if the analysis were repeated with a different sample. The standard error does not have a fixed range; it can range from 0 to $+\infty$, with 0 indicating perfect precision. A small SE value suggests that the estimate is stable and reliable.
- The column *df* stands for degrees of freedom. In many statistical models, this value reflects the amount of independent information available and which statistical distribution to use. In our case, the *emmeans* package uses *Inf* (infinity), which means it assumes a large sample size and uses a normal (z) distribution rather than a t-distribution to calculate the p-value.
- The column *z.ratio* stands for the z-statistic and is obtained by dividing the estimate by its standard error. It expresses how large the estimated difference is relative to its own uncertainty. The further the z-ratio is from 0, in either direction, the stronger the evidence that the observed difference is not due to chance. In the provided example, a z.ratio of **-4.187** suggests that the difference between item 1 and item 4 is unlikely to be a random result.
- The last column, *p.value*, represents the probability of observing the estimated difference purely by chance. A p-value < 0.05 indicates that the result is statistically significant and is unlikely to have occurred by chance. However, a value of 1 indicates that the observed difference is entirely attributable to chance, providing no evidence of a real difference. In our case, a value of **0.033** indicates that the observed difference is unlikely to be due to chance alone and is statistically significant.

Based on this summary, when interpreting the results from the CLMMs, the two key values to consider are the *estimate* and the *p-value*.

3.3.3. Analyzing the data

This section outlines the steps for interpreting the CLMM output and drawing conclusions from the data. Given the large number of pairwise comparisons, the dataset consisted of 2352 rows, making a direct interpretation of the output impractical. For this reason, the analysis was carried out in multiple steps to identify which items are most relevant for evaluating the visual complexity of data visualizations.

Step 1: For the first step, we filtered the pairwise comparisons based on their p-values. As outlined above, a p-value greater than 0.05 indicates that the result is not statistically significant, meaning that the observed difference is likely due to chance and provides no meaningful evidence for our study. In some cases, we also observed p-values of 1, indicating the absence of any real effect. For this reason, we applied a strict threshold and discarded all pairwise comparisons with a p-value greater than 0.05. As a result, 1782 out of 2352 pairwise comparisons were removed, corresponding to approximately 75% of the total.

Step 2: After thresholding, we were left with 570 pairwise comparisons with a p-value smaller than 0.05. For each item, we then computed the mean of the log-odds estimates across all its statistically significant comparisons, along with the number of times it was ranked higher or lower relative to its counterpart. It is important to note that this mean is computed on the continuous log-odds estimates produced by the CLMM model, not on the raw ordinal Likert ratings. Averaging at this stage is therefore appropriate, as these estimates are continuous numerical values and do not carry the same constraints that apply to ordinal data. An example of this process is provided in Figure 3.4. These calculations were performed only on the comparisons retained from Step 1.

main	compared to	estimate	p.value
q22	q12	1.764949933	0.009252161848
q22	q20	1.716125933	0.02080834879
q22	q3	1.650200326	0.02995628521
q22	q5	-1.663185862	0.0330680641
q22	q6	-1.872888381	0.00275039994

item	avg estimate	count higher than	count lower than
q22	0.3190403896	3	2

Figure 3.4.: *Computed statistics for pairwise comparisons*

Step 3: Having completed the first two steps, we have a list of items with their average estimate, along with the number of times each item was ranked higher or lower than the item it was compared to. No item had a p-value below 0.05 in all of its pairwise comparisons, which explains why 49 items remained after filtering. To identify which of these items were considered most relevant for evaluating the visual complexity of a data visualization, we sorted the list from highest to lowest average estimate. In doing so, we found that 31 out of 49 items had a negative average estimate, meaning that in the comparisons that passed the previous filtering steps, these items were generally ranked lower than the items they were compared to. Accordingly, their *count lower than* value is higher than their *count higher than* value (as shown in Figure 3.4). Therefore, these 31 items were excluded from further analysis. More detailed information can be found in Section A.4.4 of the appendix. For the remaining 18 items, we proceeded with additional analysis. To make the results clearer, we calculated the percentage of times the items were ranked higher than the items they were compared to, and their weighted estimate value. The outcome is shown in Table 3.5, and the full items with their indexes can be found in Table 3.6. The interpretation of this result, including the color coding in the table below and the subsequent steps taken to arrive at the final set of items, will be discussed in the following chapter.

Table 3.5.: Analysis of the 18 items remained from the previous steps

Item	**Avg Estimate**	**Ranked High**	**Ranked Low**	**% Higher**	**Weighted Estimate**
q6	2.695	35	0	72.92	1.965
q5	2.548	33	0	68.75	1.751
q25	2.472	30	0	62.5	1.545
q36	2.258	22	0	45.83	1.035
q16	2.198	19	0	39.58	0.870
q14	2.179	21	0	43.75	0.953
q17	2.130	18	0	37.5	0.799
q38	1.993	14	0	29.17	0.581
q44	1.913	13	0	27.08	0.518
q31	1.857	13	0	27.08	0.503
q41	1.846	13	0	27.08	0.500
q47	1.825	11	0	22.92	0.418
q42	1.819	7	0	14.58	0.265
q4	1.811	11	0	22.92	0.415
q49	1.449	9	1	18.75	0.272
q39	1.369	8	1	16.67	0.228
q22	0.319	3	2	6.25	0.020
q30	0.296	3	2	6.25	0.019

Table 3.6.: Items and their indexes

Index	Item
q6	I find this visualization easy to interpret.
q5	I find this visualization easy to understand.
q25	I find the visualization easy to comprehend.
q36	I find the visualization well organized.
q16	I find this visualization easy to use.
q14	I feel irritated by this visualization.
q17	I find this visualization intuitive.
q38	I find the points in the visualization easy to distinguish.
q44	I find the visualization has poor pattern visibility.
q31	I feel stressed by the visualization.
q41	I can detect patterns in the visualization.
q47	I find the visualization is poorly designed.
q42	I find the visualization is missing important details.
q4	I find this visualization familiar.
q49	I find that the vis. needs additional graphics to communicate effectively.
q39	I cannot distinguish between colors.
q22	I find the visual encoding is poor and violates good practice.
q30	I find the visualization has overlapping information.

4. Findings

In this chapter, we present the analysis and results of the expert survey data collected during the Item Validation step (Section 3.3). This process includes all the necessary steps to make sense of the data and arrive at a conclusion. First, we will provide an overview of the reverse coding process and examine whether negatively worded items are rated higher than positively worded items. Then, we will present the results of the CLMMs analysis, discussing which items received the highest ratings and what this tells us about how experts perceive and evaluate visual complexity in data visualizations.

4.1. Insights into Reverse Coding

As described in Section 3.3.1, we performed reverse coding on our negatively worded items, as this is a common approach when both positive and negative items are present in a Likert scale. Additionally, this allows us to analyze whether there are any differences in how these items are rated by the participants. García-Fernández et al. [20] have explained that there is a tendency for negatively worded items to receive higher scores than positively worded ones. They recognized this trend while analyzing data from the Grit-S scale, but also noted that it can vary depending on the questionnaire, context, and scale type. We further examined this aspect, and more detailed information for each item is available in Table A.9 of the appendix. To get an overview of the scoring, we calculated the average for each item mainly to identify whether there were any visible differences in how positively and negatively worded items were rated. It is important to note that this is not the same as using averages to analyze the Likert scale data and draw conclusions about the final results, which as also explained in the previous chapter, would not be appropriate. Here, the average is used only as a rough exploratory tool to compare the two groups of items at a surface level, before the more rigorous CLMM analysis is applied. Before calculating the averages, the scores of the negatively worded items were first reversed. From our analysis, we were not able to recognize any significant difference between the scores of the positive and negative items. However, the negatively worded items have a group mean of **2.53**, while the positively worded ones have a group mean of **3.18**, which is moderately higher. It is important to note that this result may vary not only with the scale used but also with the type of reverse method applied.

4.2. Analysis of the CLMMs Result

This section presents the results of the CLMMs analysis, which concludes the validation process described in the previous chapter. The 18 items derived from this process are discussed in detail below, while additional information about all 49 items and the corresponding calculations can be found in Section A.4.4 of the appendix. As shown in Table 3.5, we created four main groups using our final 18 items. Those highlighted in green represent the two items with the highest overall ratings, with average estimates ranging from 2.6 to 2.7. Out of the 48 possible pairwise comparisons, q6 and q5 were ranked higher in nearly 70% of cases.

The group highlighted in orange contains items with average estimates ranging from 1.9 to 2.5. These items were generally ranked higher than 23% to 63% of the other items included in the pairwise comparisons. Within this group, however, item 42 appears to behave differently from the others and is therefore highlighted in light red. Although its average estimate is comparable to the rest of the orange group, its values for *% Higher*, *Ranked High*, and *Ranked Low* differ noticeably. While the other items in this group were ranked higher at least 11 times, item 42 received higher ratings only 7 times. Even though its *Ranked Low* value is 0, the overall pattern of results makes this item stand out from the others. For this reason, it was flagged as a potential outlier and will be reconsidered in subsequent analysis steps.

The group highlighted in blue contains items with average estimates close to 1 and *% Higher* values ranging from 17% to 19%. Finally, the gray group contains items with average estimates close to 0. In particular, items q22 and q30 show an almost equal number of *Ranked High* and *Ranked Low* values. After q30, the remaining items begin to show negative average estimates and are more frequently ranked lower than higher in the pairwise comparisons. To visualize these results, we provide a stacked bar chart in Figure 4.1.

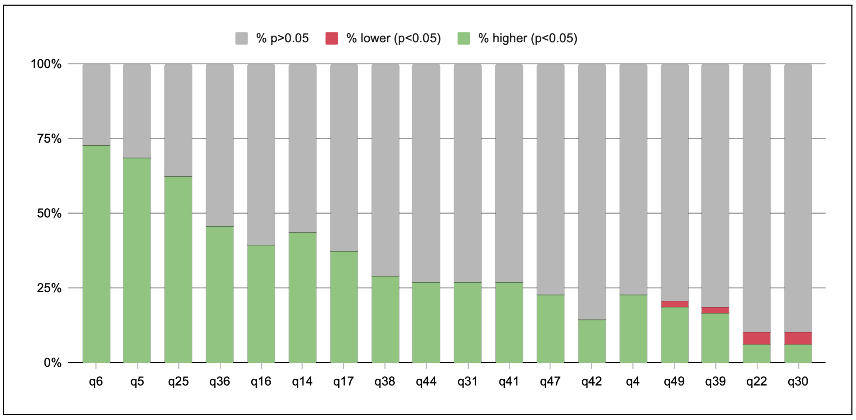


Figure 4.1.: *Pairwise Comparison Results for the 18 Retained Items*

As shown in the image above, each bar represents one of the 18 items and illustrates the distribution of its pairwise comparisons across three categories. The gray segment indicates comparisons that did not reach statistical significance ($p > 0.05$) and were therefore excluded from the analysis. The red segment corresponds to statistically significant comparisons ($p < 0.05$) in which the item was ranked lower than its counterpart, while the green segment reflects those in which the item was ranked higher. For example, consider item 39 (*"I cannot distinguish between colors."*), which was compared with 48 other items. Of these, 39 comparisons were excluded from our analysis because their p-values were greater than 0.05. Among the remaining 9 comparisons, item q39 was ranked higher than its counterpart in 8 cases and lower in only 1. With this in mind, it becomes clear not only which items performed best overall, but also how q42 stands out from the other items in its group (q25q4), as its bar shows a noticeably smaller green segment than the others.

To further validate our results, we examined the rating distributions for each item. Overall, the distributions were found to be consistent with the groupings and results discussed in the previous section. Items with higher average estimates, such as q6 and q5, showed a clear tendency towards higher ratings, while items with lower average estimates received more scattered or lower scores. This pattern supports the conclusion that the final set of 18 items accurately reflects the outcome of the validation process. Due to space constraints, the full table with rating distributions is provided in subsection A.4.5 of the appendix. Building on these findings, we conducted one final iteration on the items. Since the goal of this thesis is to identify the items most relevant for describing the perceived visual complexity of a data visualization, we wanted to ensure that the 18 items not only captured this concept effectively but were also clearly understandable to a general audience. Since the items were formulated based on terms and feedback from experts, some may not be suitable for use in surveys or experiments involving lay people. For this reason, we decided to remove item 22 (*"I find the visual encoding is poor and violates good practice."*) and item 30 (*"I find the visualization has overlapping information."*) from the final set. This decision was based on several reasons:

- Their average estimates are the lowest among all 18 items, both being close to 0.
- They were ranked higher and lower in their pairwise comparisons almost an equal number of times, providing no meaningful evidence.
- Nearly 90% of their pairwise comparisons did not reach statistical significance (p-value$>$0.05) and were therefore excluded from the analysis.
- Finally, the wording of these two items may be too technical for a general audience, and the concepts they aim to capture are already covered more effectively by the remaining 16 items.

5. Discussion and Conclusion

This work proposes a first step toward measuring perceived visual complexity in data visualizations. This takes into account how people actually experience and interpret what they see. Rather than developing a fully developed scale, this thesis identifies 16 items to capture perceived visual complexity, providing a foundation grounded in existing research in the data visualization field and direct input from domain experts. Existing approaches to measuring visual complexity tend to focus on technical or algorithmic definitions, like the number of elements on a screen or the density of information. This work does not aim to replace those, but rather to complement them by concentrating on how a visualization feels to the person looking at it. By shifting the focus toward human perception, the 16 items identified here serve as a useful building block for anyone interested in understanding visual complexity from the viewer's perspective. The findings are not the final word on the topic, but rather an invitation for future work to pick up where this thesis leaves off.

5.1. Discussion of results

The most important outcome of this work is the set of 16 items identified through the validation process. The goal is to reflect on what these items represent and what they reveal about how viewers perceive and experience visual complexity in data visualizations. As discussed in the previous chapters, the two items that received the highest overall ratings were q6 (*"I find this visualization easy to interpret."*) and q5 (*"I find this visualization easy to understand."*), followed closely by q25 (*"I find the visualization easy to comprehend."*). Although q25 was placed in a different group based on its statistical results, its average estimate and *% Higher* are both significantly high and comparable to those of the top two. As shown in Table 3.6, these three items revolve around the same core concept: how easily a viewer can make sense of a visualization. This suggests that from the experts' perspective, **interpretability**, **understandability**, and **comprehension** are the dimensions most strongly associated with perceived visual complexity.

These three concepts are directly connected to cognitive effort, which is discussed in chapter 2 as one of the key factors influencing perceived visual complexity. The fact that the top-rated items all relate to how easily a viewer can process and make sense of a visualization confirms that cognitive effort plays an important role in the experience of visual complexity. In other words, a visualization is not simply perceived as complex because it contains many elements or appears poorly designed; it is perceived as complex when it demands significant mental effort to understand. This finding aligns with the existing literature and reinforces the idea that measuring perceived visual complexity cannot rely solely on a chart's visual properties, but must also account for the viewer's cognitive experience.

The remaining 13 items cover a broader range of dimensions that together create a more complete picture of perceived visual complexity, including organization, intuitiveness, pattern detection, visual distinguishability, usability, design quality, and the completeness of the information presented. Additionally, two items capture the viewer's emotional response, addressing feelings of irritation and stress triggered by the visualization. Taken together, all 16 items confirm that perceived visual complexity is not a one-dimensional concept but is shaped by a combination of cognitive, emotional, and visual factors.

5.2. Discussion of methods

The methods used throughout this thesis played an important role in shaping the outcomes presented above. In this section, we reflect on the key methodological choices made and discuss how they influenced the final results.

5.2.1. Reflections on the Clustering Process

As discussed in Section 3.2.1, a manual approach was chosen for clustering the terms collected through the deductive and inductive methods. Both K-Means and X-Means were tested during this process, but the results they produced fell short of the requirements for this work. Hierarchical clustering was also tested as an additional check on how the terms were being grouped, but none of these methods proved sufficient, so manual clustering was kept. As the pool contained 951 distinct terms (Section A.3), different parameters were tested for both K-Means and X-Means. Of the two, X-Means performed somewhat better because it does not require the number of clusters to be defined in advance, making it more suitable for exploring a large, diverse set of terms. The initial results were first examined as a list of grouped terms in a spreadsheet, but this made it difficult to draw any meaningful conclusions. For this reason, the clusters were also visualized in 2D and 3D to provide a clearer picture of how well the terms were separated and how the groups related to one another. The result is shown below in Figure 5.1.

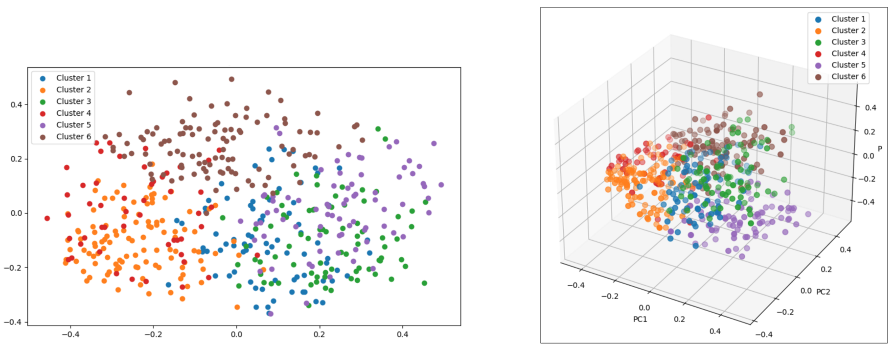


Figure 5.1.: *2D and 3D Visualization of X-Means Clusters*

Even with a visual representation of the clusters, it was difficult to draw any meaningful conclusions. As shown in the figures above, X-Means automatically sets the number of clusters to 6, each distinguished by a different color. However, the 2D visualization clearly shows that the clusters overlap heavily, with no clear separation. Testing different numbers of clusters did not improve the result, suggesting that this was the best the algorithm could achieve with this dataset. While some degree of overlap between clusters is expected, in this case the terms showed no meaningful separation.

The 3D visualization, which enabled interactive rotation and exploration, provided a slightly clearer view of the cluster distribution. However, the overall conclusion remained the same: the clusters were not well separated, and the overlap between them was too significant to draw any useful conclusions. For hierarchical clustering, term similarity was measured using semantic embeddings and cosine distance, with average linkage used to group terms that were most similar. The idea behind this was that terms with similar meanings would end up in the same cluster. The result was a dendrogram showing how all 951 terms were progressively merged into larger groups until forming a single big cluster. Similarly to the result obtained with X-Means, the output was difficult to interpret due to the size of the dendrogram. Beyond this, some terms were clustered in ways that did not seem appropriate in the context of data visualization and the viewer's perspective. For all these reasons, hierarchical clustering was also set aside, and the decision to proceed with manual clustering was confirmed.

The most important reason for choosing manual clustering over an automated method was the need for human judgment when interpreting the terms. Going through each term individually allowed me to recognize that some of them could belong to more than one cluster, depending on how they are understood in the context of data visualization. At the same time, some terms did not fit into any of the existing groups and were excluded from the process entirely. These are decisions that require domain knowledge and human reasoning, which an algorithm working on semantic similarity cannot make.

5.3. Limitations

5.3.1. Sample Size and Diversity

One of the main limitations of this work concerns the participants involved in both surveys. Regarding sample size, 50 experts participated in the first survey and 45 in the second. While these numbers are modest, they are comparable to similar scale development studies in the visualization community, such as Cabouat et al. [5] and He et al. [25], who worked with 29 and 31 experts respectively in their studies, suggesting that our sample size is not unusually small for this type of work. Even though the response rate was reasonable given that participation was voluntary and uncompensated, a larger, more diverse group of visualization experts could have yielded a broader, more representative set of terms and validation results. Additionally, as shown in Table 3.2 and Table 3.3, the gender distribution across both surveys was notably skewed, with male participants accounting for 6266% of respondents in each case. While this likely reflects the broader gender distribution within the data visualization research community, it is still worth acknowledging as a limitation, as the perspectives captured may not equally represent all groups.

Beyond gender and sample size, it is worth noting that both surveys involved only domain experts. This was done intentionally, as expert validation is typically the first step in scale development. However, since the items are ultimately intended to be used with general audiences, testing them with lay users is an important next step. The way experts perceive and evaluate visual complexity may differ from how non-expert viewers experience it, and this will need to be examined in future work.

5.3.2. Data Sources and Subjectivity of the Clustering Process

Another limitation of this work relates to the scope of the data sources used. Since the literature review used the VIS30K dataset, only publications from the IEEE VIS conference were considered for the deductive method. While this provided a large and well-established collection of visualization research, it also means that relevant terms from other venues or journals may have been missed. Studies published outside of IEEE VIS that also investigate visual complexity could have contributed additional terms and perspectives that are not reflected in the final pools. Regarding the manual clustering process, while it proved to be the most suitable approach for this type of data, it inevitably introduces a degree of subjectivity that is difficult to avoid. The way terms were grouped into clusters depended on the researchers' judgment, and a different team might have made different grouping decisions. This means that the final set of clusters, and consequently the items generated from them, could vary depending on who performed the clustering. To reduce this risk, the grouping decisions were discussed and reviewed carefully throughout the process, but it remains a limitation worth keeping in mind when interpreting the results.

5.4. Future work

The 16 items identified in this thesis represent a starting point rather than a finished product, and several directions for future work emerge from the limitations and findings discussed above. The next most immediate step would be to conduct cognitive interviews with lay users to determine whether the 16 items are clearly understood by those without a background in data visualization. These interviews would help identify any wording that might be too technical or unclear for a general audience and guide any changes needed before the items are used more widely. After this, a larger survey on Prolific should be conducted to gather laypeople's input on how relevant the items are. Combining expert input with feedback from general audiences would provide a more complete picture of which items truly capture perceived visual complexity across different types of viewers. Another aspect worth mentioning is the completeness of the 16 items. During an internal exercise in which a few members of the research group independently clustered the items and then compared their results, it was found that one concept was not represented in the final set of items. A possible solution would be to add one more item covering this missing concept, specifically the one with the highest average estimate among those not currently included. However, since confirming this properly would require a more structured process involving a larger group of people, it is left as a direction for future work. More broadly, future work should also consider including research from outside the IEEE VIS community to ensure the items hold up across different real-world contexts.

Bibliography

[1] Christopher Gene Allen. *The effects of visual complexity on cognitive load as influenced by field dependency and spatial ability*. PhD thesis, New York University, 2011.

[2] Scott Bateman, Regan L Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2573–2582, 2010.

[3] Godfred O Boateng, Torsten B Neilands, Edward A Frongillo, Hugo R Melgar-Quiñonez, and Sera L Young. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in public health*, 6:149, 2018.

[4] Michelle A Borkin, Azalea A Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. What makes a visualization memorable? *IEEE transactions on visualization and computer graphics*, 19(12):2306–2315, 2013.

[5] Anne-Flore Cabouat, Tingying He, Petra Isenberg, and Tobias Isenberg. Previs: Perceived readability evaluation for visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 2024.

[6] Cambridge University Press. Complex, n.d. Retrieved October 22, 2025.

[7] Jian Chen, Meng Ling, Rui Li, Petra Isenberg, Tobias Isenberg, Michael Sedlmair, Torsten Möller, Robert S. Laramee, Han-Wei Shen, Katharina Wünsche, and Qiru Wang. Vis30k: A collection of figures and tables from ieee visualization conference publications. *IEEE Transactions on Visualization and Computer Graphics*, 2021.

[8] Mengdi Chu, Zefeng Qiu, Meng Ling, Shuning Jiang, Robert S. Laramee, Michael Sedlmair, and Jian Chen. What makes a visualization complex?, 2025.

[9] William S Cleveland and Robert McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554, 1984.

[10] Collins English Dictionary. Complex, n.d. Retrieved October 22, 2025.

[11] Corporation for Digital Scholarship. Zotero, 2024.

[12] Liqiong Deng and Marshall Scott Poole. Affect in web interfaces: A study of the impacts of web page visual complexity and order1. *Mis Quarterly*, 34(4):711–730, 2010.

[13] Don C Donderi. Visual complexity: a review. *Psychological bulletin*, 132(1):73, 2006.

[14] Brian Duffy, Aritra Dasgupta, Robert Kosara, S Walton, and Min Chen. Measuring visual complexity of cluster-based visualizations. *arXiv preprint arXiv:1302.5824*, 2013.

[15] Johannes Ellemose and Niklas Elmqvist. Eye of the beholder: Towards measuring visualization complexity, 2025.

[16] Stephen Few. Show me the numbers. *Analytics Pres*, 2, 2004.

[17] Alex Forsythe, Marcos Nadal, Noel Sheehy, Camilo J Cela-Conde, and Martin Sawey. Predicting beauty: Fractal dimension and visual complexity in art. *British journal of psychology*, 102(1):49–70, 2011.

[18] Alexandra Forsythe. Visual complexity: Is that all there is? In *International Conference on Engineering Psychology and Cognitive Ergonomics*, pages 158–166. Springer, 2009.

[19] Steven L Franconeri, Lace M Padilla, Priti Shah, Jeffrey M Zacks, and Jessica Hullman. The science of visual data communication: What works. *Psychological Science in the public interest*, 22(3):110–161, 2021.

[20] Jaime García-Fernández, Álvaro Postigo, Marcelino Cuesta, Covadonga González-Nuevo, Álvaro Menéndez-Aller, and Eduardo García-Cueto. To be direct or not: Reversing likert response format items. *The Spanish Journal of Psychology*, 25:e24, 2022.

[21] Johannes Gärtner, Silvia Miksch, and Stefan Carl-McGrath. Vico: a metric for the complexity of information visualizations. In *International conference on theory and application of diagrams*, pages 249–263. Springer, 2002.

[22] Peter Gerjets and Katharina Scheiter. Goal configurations and processing strategies as moderators between instructional design and cognitive load: Evidence from hypertext-based instruction. *Educational psychologist*, 38(1):33–41, 2003.

[23] Xiaoying Guo, Takio Kurita, Chie Muraki Asano, and Akira Asano. Visual complexity assessment of painting images. In *2013 IEEE International Conference on Image Processing*, pages 388–392. IEEE, 2013.

[24] Simon Harper, Eleni Michailidou, and Robert Stevens. Toward a definition of visual complexity as an implicit measure of cognitive load. *ACM Transactions on Applied Perception (TAP)*, 6(2):1–18, 2009.

[25] Tingying He, Petra Isenberg, Raimund Dachselt, and Tobias Isenberg. Beauvis: A validated scale for measuring the aesthetic pleasure of visual representations. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):363–373, 2022.

[26] Lee Howorko, Joyce Maria Boedianto, Ben Daniel, et al. The efficacy of stacked bar charts in supporting single-attribute and overall-attribute comparisons. *Visual Informatics*, 2(3):155–165, 2018.

[27] Weidong Huang, Peter Eades, and Seok-Hee Hong. Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*, 8(3):139–152, 2009.

[28] Petra Isenberg, Florian Heimerl, Steffen Koch, Tobias Isenberg, Panpan Xu, Chad Stolper, Michael Sedlmair, Jian Chen, Torsten Möller, and John Stasko. vispubdata.org: A metadata collection about IEEE visualization (VIS) publications. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2199–2206, September 2017.

[29] Susan Jamieson. Likert scales: How to (ab) use them? *Medical education*, 38(12):1217–1218, 2004.

[30] Celeste Kidd, Steven T Piantadosi, and Richard N Aslin. The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one*, 7(5):e36399, 2012.

[31] Sunjung Kim and Linda J Lombardino. Comparing graphs and text: Effects of complexity and task. *Journal of Eye Movement Research*, 8(3), 2015.

[32] Christian Knoll, Torsten Möller, Kathleen Gregory, and Laura Koesten. The gulf of interpretation: From chart to message and back again. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2025.

[33] May Jorella Lazaro, Yohan Kang, Myung Hwan Yun, and Sungho Kim. The effects of visual complexity and decluttering methods on visual search and target detection in cockpit displays. *International Journal of Human–Computer Interaction*, 37(7):588–600, 2021.

[34] Ju Hyun Lee and Michael J Ostwald. The visual attractiveness of architectural facades: measuring visual complexity and attractive strength in architecture. *Architectural science review*, 66(1):42–52, 2023.

[35] Gordon E Legge and John M Foley. Contrast masking in human vision. *Journal of the optical Society of America*, 70(12):1458–1471, 1980.

[36] Torrin M Liddell and John K Kruschke. Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79:328–348, 2018.

[37] Helen M Marcus-Roberts and Fred S Roberts. Meaningless statistics. *Journal of Educational Statistics*, 12(4):383–394, 1987.

[38] Richard E Mayer. Multimedia learning. In *Psychology of learning and motivation*, volume 41, pages 85–139. Elsevier, 2002.

[39] Merriam-Eebster Dictionary. Complex, n.d. Retrieved October 22, 2025.

[40] Eleni Michailidou, Simon Harper, and Sean Bechhofer. Visual complexity and aesthetic perception of web pages. In *Proceedings of the 26th annual ACM international conference on Design of communication*, pages 215–224, 2008.

[41] Fred GWC Paas. Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. *Journal of educational psychology*, 84(4):429, 1992.

[42] Anshul Vikram Pandey, Anjali Manivannan, Oded Nov, Margaret Satterthwaite, and Enrico Bertini. The persuasive power of data visualization. *IEEE transactions on visualization and computer graphics*, 20(12):2211–2220, 2014.

[43] Wei Peng, Matthew O Ward, and Elke A Rundensteiner. Clutter reduction in multi-dimensional data visualization using dimension reordering. In *IEEE Symposium on Information Visualization*, pages 89–96. IEEE, 2004.

[44] Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano. Measuring visual clutter. *Journal of vision*, 7(2):17–17, 2007.

[45] Gabriel Ryan, Abigail Mosca, Remco Chang, and Eugene Wu. At a glance: Pixel approximate entropy as a measure of line chart complexity. *IEEE transactions on visualization and computer graphics*, 25(1):872–881, 2018.

[46] Regina Schuster, Kathleen Gregory, Torsten Möller, and Laura Koesten. Practitioners' perspectives on designing data visualizations for the general public. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2026.

[47] Joan G Snodgrass and Mary Vanderwart. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory*, 6(2):174, 1980.

[48] James A Stuart and Hermann M Burian. A study of separation difficulty: Its relationship to visual acuity in normal and amblyopic eyes. *American Journal of Ophthalmology*, 53(3):471–477, 1962.

[49] John Sweller. Chapter two - cognitive load theory. volume 55 of *Psychology of Learning and Motivation*, pages 37–76. Academic Press, 2011.

[50] John Sweller. Cognitive load theory and individual differences. *Learning and Individual Differences*, 110:102423, 2024.

[51] John Sweller, Jeroen JG Van Merrienboer, and Fred GWC Paas. Cognitive architecture and instructional design. *Educational psychology review*, 10(3):251–296, 1998.

[52] Justin Talbot, Vidya Setlur, and Anushka Anand. Four experiments on the perception of bar charts. *IEEE transactions on visualization and computer graphics*, 20(12):2152–2160, 2014.

[53] Jack E Taylor, Guillaume A Rousselet, Christoph Scheepers, and Sara C Sereno. Rating norms should be calculated from cumulative link mixed effects models. *Behavior research methods*, 55(5):2175–2196, 2023.

[54] Edward R Tufte and Peter R Graves-Morris. *The visual display of quantitative information*, volume 2. Graphics press Cheshire, CT, 1983.

[55] Brandon Victor Syiem and Eduardo Velloso. Better assumptions, stronger conclusions: The case for ordinal regression in hci. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, pages 1–21, 2026.

[56] Andreu Vigil-Colet, David Navarro-González, and Fabia Morales-Vives. To reverse or to not reverse likert-type items: That is the question. *Psicothema*, 32(1):108–114, 2020.

[57] Colin Ware. *Information visualization: perception for design*. Morgan Kaufmann, 2019.

[58] Ying Zhu. A review of complexity metrics for data visualization. In *2023 27th International Conference Information Visualisation (IV)*, pages 131–135. IEEE, 2023.

A. Appendix

A.1. Deductive method sources

As described in Section 3.1.1, we extracted terms from study questionnaires (Pool 1) and participant comments (Pool 2). Table A.1 lists all 166 sources considered in this work, indicating whether each contributed to Pool 1, Pool 2, or both. Of these, 88 sources were common to both pools. Table A.2 presents the studies and the terms collected from Likert scales and questionnaires used to form Pool 1 (see Section 3.1.1).

Table A.1.: 166 publications from which terms were extracted for Pool 1 and Pool 2.

Authors & Year	**DOI**	**Pool 1**	**Pool 2**
Feyer et al. 2023	10.1109/tvcg.2023.3327402	yes	no
Chen et al. 2023	10.1109/tvcg.2023.3326925	yes	yes
Wang et al. 2023	10.1109/tvcg.2023.3327153	no	yes
Shen et al. 2023	10.1109/tvcg.2023.3327197	no	yes
Piccolotto et al. 2023	10.1109/tvcg.2023.3327203	no	yes
He et al. 2023	10.1109/tvcg.2023.3326941	yes	yes
Alebri et al. 2023	10.1109/tvcg.2023.3326914	yes	yes
Kim et al. 2023	10.1109/tvcg.2023.3327150	no	yes
Morini et al. 2023	10.1109/tvcg.2023.3327185	yes	yes
Lin et al. 2023	10.1109/tvcg.2023.3327170	no	yes
Li et al. 2023	10.1109/tvcg.2023.3326904	yes	no
Xiao et al. 2023	10.1109/tvcg.2023.3326913	yes	no
Shi et al. 2023	10.1109/tvcg.2023.3326522	yes	yes
Feng et al. 2023	10.1109/tvcg.2023.3327168	no	yes
Wen et al. 2023	10.1109/tvcg.2023.3327148	yes	yes
Braun et al. 2023	10.1109/tvcg.2023.3326576	yes	yes
Zhu-Tian et al. 2023	10.1109/tvcg.2023.3326568	yes	yes
Huang et al. 2023	10.1109/tvcg.2023.3326932	yes	yes
Shi et al. 2023	10.1109/tvcg.2023.3327200	no	yes
Gao et al. 2023	10.1109/tvcg.2023.3327353	yes	no
Lin et al. 2023	10.1109/tvcg.2023.3327161	yes	yes
Teng et al. 2023	10.1109/tvcg.2023.3326587	no	yes
Elhamdadi et al. 2023	10.1109/tvcg.2023.3326579	no	yes
Xing et al. 2023	10.1109/tvcg.2023.3326923	no	yes
Deng et al. 2023	10.1109/tvcg.2023.3327162	yes	yes

Continued on next page

Authors & Year	DOI	Pool 1	Pool 2
Li et al. 2022	10.1109/tvcg.2022.3209482	yes	yes
Zhao et al. 2022	10.1109/tvcg.2022.3209469	no	yes
He et al. 2022	10.1109/tvcg.2022.3209390	no	yes
Abdelaal et al. 2022	10.1109/tvcg.2022.3209427	yes	yes
Huang et al. 2022	10.1109/tvcg.2022.3209384	yes	yes
Ghai et al. 2022	10.1109/tvcg.2022.3209484	yes	yes
Bertucci et al. 2022	10.1109/tvcg.2022.3209425	no	yes
Sun et al. 2022	10.1109/tvcg.2022.3209428	no	yes
Wang et al. 2022	10.1109/tvcg.2022.3209435	no	yes
Lawonn et al. 2022	10.1109/tvcg.2022.3209374	no	yes
Linhares et al. 2022	10.1109/tvcg.2022.3209477	yes	yes
Tandon et al. 2022	10.1109/tvcg.2022.3209491	yes	no
Shi et al. 2022	10.1109/tvcg.2022.3209434	yes	no
Tyagi et al. 2022	10.1109/tvcg.2022.3209361	yes	no
Tyagi et al. 2022	10.1109/tvcg.2022.3209392	yes	yes
Zhang et al. 2022	10.1109/tvcg.2022.3209440	yes	yes
Chen et al. 2022	10.1109/tvcg.2022.3209385	yes	yes
Chen et al. 2022	10.1109/tvcg.2022.3209497	no	yes
Stokes et al. 2022	10.1109/tvcg.2022.3209383	yes	yes
Shi et al. 2022	10.1109/tvcg.2022.3209486	yes	no
Lin et al. 2022	10.1109/tvcg.2022.3209353	yes	yes
Wang et al. 2022	10.1109/tvcg.2022.3209357	no	yes
Ruan et al. 2022	10.1109/tvcg.2022.3209455	yes	yes
Warchol et al. 2022	10.1109/tvcg.2022.3209378	yes	yes
Soure et al. 2021	10.1109/tvcg.2021.3114822	yes	no
Maher et al. 2021	10.1109/tvcg.2021.3114789	yes	yes
Ying et al. 2021	10.1109/tvcg.2021.3114877	no	yes
Latif et al. 2021	10.1109/tvcg.2021.3114802	yes	yes
Wall et al. 2021	10.1109/tvcg.2021.3114862	yes	no
Narechania et al. 2021	10.1109/tvcg.2021.3114827	yes	yes
Wu et al. 2021	10.1109/tvcg.2021.3114826	yes	yes
Jessup et al. 2021	10.1109/tvcg.2021.3114786	yes	yes
Zytek et al. 2021	10.1109/tvcg.2021.3114864	no	yes
Tang et al. 2021	10.1109/tvcg.2021.3114781	yes	yes
Shi et al. 2020	10.1109/tvcg.2020.3030403	yes	yes
Cashman et al. 2020	10.1109/tvcg.2020.3030443	yes	yes
Chen et al. 2020	10.1109/tvcg.2020.3030411	yes	yes
Chen et al. 2020	10.1109/tvcg.2020.3030338	yes	yes
Wang et al. 2020	10.1109/tvcg.2020.3030433	yes	yes
Lee et al. 2020	10.1109/tvcg.2020.3030435	no	yes
Yuan et al. 2020	10.1109/tvcg.2020.3030432	no	yes
McDonald et al. 2020	10.1109/tvcg.2020.3030363	yes	yes

Continued on next page

Authors & Year	DOI	Pool 1	Pool 2
Baumgartl et al. 2020	10.1109/tvcg.2020.3030437	yes	yes
Choudhry et al. 2020	10.1109/tvcg.2020.3030358	yes	yes
Zhong et al. 2020	10.1109/vast50239.2020.00014	yes	no
Liu et al. 2020	10.1109/tvcg.2020.3030364	yes	yes
Lee et al. 2020	10.1109/tvcg.2020.3030446	yes	yes
Ens et al. 2020	10.1109/tvcg.2020.3030334	no	yes
Eulzer et al. 2020	10.1109/tvcg.2020.3030388	no	yes
Wang et al. 2019	10.1109/tvcg.2019.2934398	yes	yes
Krekhov et al. 2019	10.1109/tvcg.2019.2934370	yes	yes
Filho et al. 2019	10.1109/tvcg.2019.2934415	yes	yes
Wei et al. 2019	10.1109/tvcg.2019.2934208	no	yes
Mumtaz et al. 2019	10.1109/tvcg.2019.2934669	no	yes
Behrisch et al. 2019	10.1109/tvcg.2019.2934300	yes	yes
Yang et al. 2019	10.1109/tvcg.2019.2934557	no	yes
Eulzer et al. 2019	10.1109/tvcg.2019.2934337	no	yes
Kraus et al. 2019	10.1109/tvcg.2019.2934395	yes	yes
Khayat et al. 2019	10.1109/tvcg.2019.2934266	yes	no
Krekhov et al. 2018	10.1109/tvcg.2018.2864498	yes	no
Zhao et al. 2018	10.1109/tvcg.2018.2865020	yes	yes
Blascheck et al. 2018	10.1109/tvcg.2018.2865142	yes	no
Zhao et al. 2018	10.1109/tvcg.2018.2864475	yes	yes
Lu et al. 2018	10.1109/tvcg.2018.2864887	yes	no
Falk et al. 2018	10.1109/tvcg.2018.2864816	no	yes
Tang et al. 2018	10.1109/tvcg.2018.2864899	yes	no
Correll et al. 2018	10.1109/tvcg.2018.2864907	yes	no
Wang et al. 2018	10.1109/tvcg.2018.2865232	yes	yes
Yang et al. 2018	10.1109/tvcg.2018.2865192	yes	yes
Subramonyam et al. 2018	10.1109/tvcg.2018.2865231	yes	yes
Perin et al. 2017	10.1109/tvcg.2017.2743918	yes	no
Bernard et al. 2017	10.1109/tvcg.2017.2744818	no	yes
Dimara et al. 2017	10.1109/tvcg.2017.2745138	yes	yes
Chung et al. 2017	10.1109/vast.2017.8585484	yes	no
Valdez et al. 2017	10.1109/tvcg.2017.2744138	yes	no
Zhao et al. 2017	10.1109/tvcg.2017.2744738	yes	yes
Zhao et al. 2017	10.1109/tvcg.2017.2745279	yes	yes
Arendt et al. 2017	10.1109/vast.2017.8585487	yes	yes
Pienta et al. 2017	10.1109/tvcg.2017.2744898	yes	yes
Yang et al. 2016	10.1109/tvcg.2016.2598472	yes	no
Meuschke et al. 2016	10.1109/tvcg.2016.2598795	no	yes
Kim et al. 2016	10.1109/tvcg.2016.2598620	no	yes
Heimerl et al. 2016	10.1109/vast.2016.7883507	no	yes
Taher et al. 2016	10.1109/tvcg.2016.2598498	yes	yes

Continued on next page

Authors & Year	DOI	Pool 1	Pool 2
Smit et al. 2016	10.1109/tvcg.2016.2598826	yes	yes
Liu et al. 2016	10.1109/vast.2016.7883513	yes	yes
Ren et al. 2016	10.1109/tvcg.2016.2598828	yes	yes
Badam et al. 2016	10.1109/vast.2016.7883506	yes	yes
Felix et al. 2016	10.1109/tvcg.2016.2598447	yes	no
Dimara et al. 2016	10.1109/tvcg.2016.2598594	yes	no
Bach et al. 2016	10.1109/tvcg.2016.2598958	yes	no
Jang et al. 2015	10.1109/tvcg.2015.2468292	yes	yes
Lawonn et al. 2015	10.1109/tvcg.2015.2467961	yes	yes
Albo et al. 2015	10.1109/tvcg.2015.2467322	yes	yes
Raidou et al. 2015	10.1109/tvcg.2015.2467872	yes	yes
Cohé et al. 2015	10.1109/tvcg.2015.2467035	yes	no
Roberts et al. 2015	10.1109/tvcg.2015.2467271	yes	no
Beck et al. 2015	10.1109/tvcg.2015.2467757	yes	yes
Gomez et al. 2014	10.1109/vast.2014.7042482	yes	yes
Waldner et al. 2014	10.1109/tvcg.2014.2346352	yes	yes
Bae et al. 2014	10.1109/tvcg.2014.2346420	yes	yes
Gotz et al. 2014	10.1109/tvcg.2014.2346682	no	yes
Haehn et al. 2014	10.1109/tvcg.2014.2346371	yes	yes
Kondo et al. 2014	10.1109/tvcg.2014.2346250	no	yes
Isaacs et al. 2014	10.1109/tvcg.2014.2346743	yes	yes
Ren et al. 2014	10.1109/tvcg.2014.2346291	yes	no
Zhao et al. 2014	10.1109/vast.2014.7042496	yes	no
Chevalier et al. 2014	10.1109/tvcg.2014.2346424	no	yes
Gramazio et al. 2014	10.1109/tvcg.2014.2346983	no	yes
Tennekes et al. 2014	10.1109/tvcg.2014.2346277	yes	yes
Dwyer et al. 2013	10.1109/tvcg.2013.151	no	yes
Borkin et al. 2013	10.1109/tvcg.2013.155	yes	yes
Zhao et al. 2013	10.1109/tvcg.2013.167	yes	yes
Gratzl et al. 2013	10.1109/tvcg.2013.173	no	yes
Lee et al. 2013	10.1109/tvcg.2013.191	yes	yes
Perin et al. 2013	10.1109/tvcg.2013.192	yes	no
Hajizadeh et al. 2013	10.1109/tvcg.2013.197	yes	yes
Basole et al. 2013	10.1109/tvcg.2013.209	no	yes
Sambasivan et al. 2013	10.1109/tvcg.2013.233	no	yes
Xu et al. 2012	10.1109/tvcg.2012.189	yes	no
Höferlin et al. 2012	10.1109/tvcg.2012.222	yes	yes
Sedlmair et al. 2012	10.1109/tvcg.2012.255	yes	yes
Isenberg et al. 2011	10.1109/tvcg.2011.160	no	yes
Steinberger et al. 2011	10.1109/tvcg.2011.183	yes	yes
Khlebnikov et al. 2011	10.1109/tvcg.2011.184	yes	no
Alper et al. 2011	10.1109/tvcg.2011.186	yes	no

Continued on next page

Authors & Year	DOI	Pool 1	Pool 2
Borkin et al. 2011	10.1109/tvcg.2011.192	yes	yes
Rodgers et al. 2011	10.1109/tvcg.2011.196	yes	yes
Claessen et al. 2011	10.1109/tvcg.2011.201	yes	yes
Alper et al. 2011	10.1109/tvcg.2011.234	yes	yes
Kwon et al. 2011	10.1109/vast.2011.6102435	yes	yes
Tomaszewski et al. 2010	10.1109/vast.2010.5652895	yes	no
Koh et al. 2010	10.1109/tvcg.2010.175	yes	yes
Liu et al. 2010	10.1109/vast.2010.5652910	no	yes
Bruckner et al. 2010	10.1109/tvcg.2010.190	yes	yes
Lee et al. 2010	10.1109/tvcg.2010.194	yes	yes
Wongsuphasawat et al. 2009	10.1109/vast.2009.5332595	no	yes
Kim et al. 2009	10.1109/tvcg.2009.146	no	yes
Robertson et al. 2008	10.1109/tvcg.2008.125	yes	yes
Heer et al. 2007	10.1109/tvcg.2007.70539	yes	yes
Tory et al. 2007	10.1109/tvcg.2007.70596	yes	yes

Table A.2.: 121 publications from which terms were extracted for Pool 1.

Authors & Year	DOI	Terms collected
Li et al. 2022	10.1109/tvcg.2022.3209482	number of lines / fill color / halo / order of elements
Xu et al. 2012	10.1109/tvcg.2012.189	edge type / graph size / edge length / edge crossing / number of nodes / cluttered
Abdelaal et al. 2022	10.1109/tvcg.2022.3209427	visual clutter / network size / network density / number of nodes / visibility of links / overdrawing / proximity of nodes
Alebri et al. 2023	10.1109/tvcg.2023.3326914	background image / icon labels / pictorial bars / icon arrays / enticing / plain
Gomez et al. 2014	10.1109/vast.2014.7042482	large data model / visual density / visualization type / distracting features

Continued on next page

Authors & Year	DOI	Terms collected
Heer et al. 2007	10.1109/tvcg.2007.70539	familiarity / occlusion / number of elements / complicated
Waldner et al. 2014	10.1109/tvcg.2014.2346352	scene complexity / different colors / target detection / annoyance
Bach et al. 2016	10.1109/tvcg.2016.2598958	dense area / spatially close nodes / many links crossing each other / edge clutter / graph size / nr of nodes / graph density / level of nesting / high level of connectivity / clutter-free / confidence
Baumgartl et al. 2020	10.1109/tvcg.2020.3030437	familiarity / useful / size of dataset / number of features / space efficiency
Blascheck et al. 2018	10.1109/tvcg.2018.2865142	familiarity / chart type / data size / confidence
Borkin et al. 2013	10.1109/tvcg.2013.155	mental activity / irritated / stressed / annoyed / relaxed
Braun et al. 2023	10.1109/tvcg.2023.3326576	confidence
Cashman et al. 2020	10.1109/tvcg.2020.3030443	intuitive / mentally demanding / insecurity / stressful / annoyance
Chen et al. 2020	10.1109/tvcg.2020.3030338	ease to use
Chen et al. 2023	10.1109/tvcg.2023.3326925	comprehensive / placement of annotations / graph layout / aesthetics
Choudhry et al. 2020	10.1109/tvcg.2020.3030358	ease of understanding
Claessen et al. 2011	10.1109/tvcg.2011.201	easy to use / easy to understand
Chen et al. 2020	10.1109/tvcg.2020.3030411	easy to understand / easy to compare / aesthetic appearance
Cohé et al. 2015	10.1109/tvcg.2015.2467035	intuitive / text easy to read / symbols easy to read / easy to navigate / feeling lost

Continued on next page

Authors & Year	DOI	Terms collected
Bae et al. 2014	10.1109/tvcg.2014.2346420	link type / helpful
Steinberger et al. 2011	10.1109/tvcg.2011.183	mental demand / background visibility / visual clutter / visual attractiveness
Khlebnikov et al. 2011	10.1109/tvcg.2011.184	dimensionality
Chung et al. 2017	10.1109/vast.2017.8585484	confidence / context visible
Haehn et al. 2014	10.1109/tvcg.2014.2346371	workload
Alper et al. 2011	10.1109/tvcg.2011.186	ease / clutter / enjoyment / confidence / frustration
Dimara et al. 2016	10.1109/tvcg.2016.2598594	confidence
Dimara et al. 2017	10.1109/tvcg.2017.2745138	satisfaction / confidence / easiness
Zhao et al. 2018	10.1109/tvcg.2018.2865020	satisfaction / ease of use / informativeness
Borkin et al. 2011	10.1109/tvcg.2011.192	intuitive
Höferlin et al. 2012	10.1109/tvcg.2012.222	subjective effectiveness / usefulness / comfort / effort
Rodgers et al. 2011	10.1109/tvcg.2011.196	distraction / intrusiveness / relaxation / stress / attractive / beautiful / interesting / engaging
Feyer et al. 2023	10.1109/tvcg.2023.3327402	number of layers / network size / number of edges / number of components / overlapping components
Filho et al. 2019	10.1109/tvcg.2019.2934415	mental workload / effort / temporal demand / frustration / easy to find the required information
Isaacs et al. 2014	10.1109/tvcg.2014.2346743	usefulness
Tomaszewski et al. 2010	10.1109/vast.2010.5652895	understandability / consistency of design / overall usability
Ghai et al. 2022	10.1109/tvcg.2022.3209484	mentally demanding / usability / easy to use / need of support to use the system

Continued on next page

Authors & Year	DOI	Terms collected
Behrisch et al. 2019	10.1109/tvcg.2019.2934300	usability
He et al. 2023	10.1109/tvcg.2023.3326941	aesthetics / readability
Huang et al. 2022	10.1109/tvcg.2022.3209384	easy to understand / easy to use
Huang et al. 2023	10.1109/tvcg.2023.3326932	usefulness / ease of use
Zhao et al. 2018	10.1109/tvcg.2018.2864475	effectiveness / aesthetics
Lu et al. 2018	10.1109/tvcg.2018.2864887	ease of using / helpful feature
Jessup et al. 2021	10.1109/tvcg.2021.3114786	usefulness of individual features
Zhao et al. 2013	10.1109/tvcg.2013.167	easy to use
Khayat et al. 2019	10.1109/tvcg.2019.2934266	perceived usefulness / ease of use
Li et al. 2023	10.1109/tvcg.2023.3326904	familiarity
Kraus et al. 2019	10.1109/tvcg.2019.2934395	difficulty
Krekhov et al. 2018	10.1109/tvcg.2018.2864498	mental demand / effort / frustration
Krekhov et al. 2019	10.1109/tvcg.2019.2934370	mental demand / effort / frustration
Latif et al. 2021	10.1109/tvcg.2021.3114802	usability
Lawonn et al. 2015	10.1109/tvcg.2015.2467961	color coding / useful
Lee et al. 2020	10.1109/tvcg.2020.3030446	familiarity
Lin et al. 2022	10.1109/tvcg.2022.3209353	understandability
Linhares et al. 2022	10.1109/tvcg.2022.3209477	visual scalability / intuitive / easy to use
Liu et al. 2020	10.1109/tvcg.2020.3030364	feature usefulness / intuitive
Correll et al. 2018	10.1109/tvcg.2018.2864907	familiarity
Maher et al. 2021	10.1109/tvcg.2021.3114789	easy to use
Koh et al. 2010	10.1109/tvcg.2010.175	easy to learn / easy to use / fun to use / satisfied
McDonald et al. 2020	10.1109/tvcg.2020.3030363	usability
Shi et al. 2022	10.1109/tvcg.2022.3209434	confidence / mental demand / temporal demand / effort / frustration / cognitive load
Morini et al. 2023	10.1109/tvcg.2023.3327185	visualizations clarity / content clarity
Jang et al. 2015	10.1109/tvcg.2015.2468292	intuitive / easy to grasp / clear / straightforward
Narechania et al. 2021	10.1109/tvcg.2021.3114827	visualization literacy

Continued on next page

Authors & Year	DOI	Terms collected
Shi et al. 2023	10.1109/tvcg.2023.3326522	user confidence / cognitive load / usefulness / user satisfaction / mental demand / effort / frustration
Albo et al. 2015	10.1109/tvcg.2015.2467322	enjoyable visualization / clear to use / understandable / easy to use / easy to learn / mental effort / frustrating
Raidou et al. 2015	10.1109/tvcg.2015.2467872	easy to use / easy to understand / useful
Zhao et al. 2014	10.1109/vast.2014.7042496	useful / usable / intuitive
Smit et al. 2016	10.1109/tvcg.2016.2598826	useful / clearly visualized / clearly visible / helpful contours / color coding / easy to interpret / spatial orientation / visual encoding
Perin et al. 2013	10.1109/tvcg.2013.192	visual encoding / visualization size
Perin et al. 2017	10.1109/tvcg.2017.2743918	effectiveness of visual encoding
Valdez et al. 2017	10.1109/tvcg.2017.2744138	separability
Sedlmair et al. 2012	10.1109/tvcg.2012.255	clear arrangement
Ren et al. 2014	10.1109/tvcg.2014.2346291	expressive / easy to use / easy to understand
Bruckner et al. 2010	10.1109/tvcg.2010.190	easy to use
Chen et al. 2022	10.1109/tvcg.2022.3209385	frustration
Roberts et al. 2015	10.1109/tvcg.2015.2467271	unnecessarily complex / easy to use / need of support / inconsistency / quick to learn / cumbersome to use / confident / need to have more knowledge
Robertson et al. 2008	10.1109/tvcg.2008.125	i enjoyed using this visualization / exciting visualization / screen too cluttered / enjoyable

Continued on next page

Authors & Year	DOI	Terms collected
Ruan et al. 2022	10.1109/tvcg.2022.3209455	visual design is easy to understand / effective design choice / smooth user interaction
Shi et al. 2020	10.1109/tvcg.2020.3030403	understandability / aestheticness / expressiveness / usefulness / ease of use
Lee et al. 2013	10.1109/tvcg.2013.191	ease of learning / ease of use / engagement
Subramonyam et al. 2018	10.1109/tvcg.2018.2865231	ease of use / learnability / intuitiveness / ease of understanding the annotations
Liu et al. 2016	10.1109/vast.2016.7883513	easy to learn / easy to use / easy to identify trends
Soure et al. 2021	10.1109/tvcg.2021.3114822	usefulness of each feature
Lee et al. 2010	10.1109/tvcg.2010.194	ease of use / appearance / readability of the topic terms
Tory et al. 2007	10.1109/tvcg.2007.70596	distracting / interface makes it easy to complete tasks
Ren et al. 2016	10.1109/tvcg.2016.2598828	familiarity
Alper et al. 2011	10.1109/tvcg.2011.234	static visual cues / stereoscopic highlighting
Stokes et al. 2022	10.1109/tvcg.2022.3209383	visual information / textual information / graphical literacy
Hajizadeh et al. 2013	10.1109/tvcg.2013.197	distraction
Shi et al. 2022	10.1109/tvcg.2022.3209486	consistency
Badam et al. 2016	10.1109/vast.2016.7883506	accuracy / intuitiveness / enjoyment
Taher et al. 2016	10.1109/tvcg.2016.2598498	intuitive / easy to interpret / easy to discover themes
Tandon et al. 2022	10.1109/tvcg.2022.3209491	perceived difficulty / spatial visualization assessment / anxiety

Continued on next page

Authors & Year	DOI	Terms collected
Tang et al. 2018	10.1109/tvcg.2018.2864899	aesthetics / legibility / engagement / entertainment / attractiveness / manually-drawn / intuitiveness / ease of use
Tang et al. 2021	10.1109/tvcg.2021.3114781	easy to use / easy to learn / effectiveness
Felix et al. 2016	10.1109/tvcg.2016.2598447	ease of use
Arendt et al. 2017	10.1109/vast.2017.8585487	usefulness of components of the visualization / usefulness of the visualization
Gao et al. 2023	10.1109/tvcg.2023.3327353	usability / easy to use / interface design
Tennekes et al. 2014	10.1109/tvcg.2014.2346277	helpful chart colors / prettiness / interpretation
Tyagi et al. 2022	10.1109/tvcg.2022.3209361	enjoyment / expressiveness / immersion
Tyagi et al. 2022	10.1109/tvcg.2022.3209392	complex / easy to use / inconsistency / easy to learn / cumbersome / high learning curve
Pienta et al. 2017	10.1109/tvcg.2017.2744898	enjoyable
Lin et al. 2023	10.1109/tvcg.2023.3327161	usefulness of each feature
Beck et al. 2015	10.1109/tvcg.2015.2467757	well-explorable relationships / easy to use / self-explaining
Kwon et al. 2011	10.1109/vast.2011.6102435	self-explanatory
Deng et al. 2023	10.1109/tvcg.2023.3327162	facilitates exploration
Wall et al. 2021	10.1109/tvcg.2021.3114862	visualization literacy
Wang et al. 2018	10.1109/tvcg.2018.2865232	readability / aesthetics / attractiveness
Wang et al. 2019	10.1109/tvcg.2019.2934398	easy to understand / aesthetic / expressive
Wang et al. 2020	10.1109/tvcg.2020.3030433	chart familiarity
Warchol et al. 2022	10.1109/tvcg.2022.3209378	usefulness of features
Wen et al. 2023	10.1109/tvcg.2023.3327148	easy to read / placement
Wu et al. 2021	10.1109/tvcg.2021.3114826	usefulness of features
Xiao et al. 2023	10.1109/tvcg.2023.3326913	enjoyment / ease of use
Yang et al. 2016	10.1109/tvcg.2016.2598472	experience with the visualization design

Continued on next page

Authors & Year	DOI	Terms collected
Yang et al. 2018	10.1109/tvcg.2018.2865192	visual design / readability
Zhang et al. 2022	10.1109/tvcg.2022.3209440	visual design convenience / visual design provides enough information / visual design is friendly to green hands
Zhao et al. 2017	10.1109/tvcg.2017.2744738	aesthetics
Zhao et al. 2017	10.1109/tvcg.2017.2745279	perceived usefulness of each feature / layout of the graph / text nodes / text links / comments / references / tags
Zhong et al. 2020	10.1109/vast50239.2020.00014	intuitiveness / ease of use / layout
Zhu-Tian et al. 2023	10.1109/tvcg.2023.3326568	mental load

Table A.3.: 133 publications from which terms were extracted for Pool 2.

Authors & Year	DOI	Terms collected
Li et al. 2022	10.1109/tvcg.2022.3209482	visual clutter
Abdelaal et al. 2022	10.1109/tvcg.2022.3209427	difficult / hard / overall pattern / dense cluster
Alebri et al. 2023	10.1109/tvcg.2023.3326914	color use / very striking visually / annoying use of color / cluttered by additional text / too complicated / gimmicky visuals / presence of icons / presence of pictorials
Gomez et al. 2014	10.1109/vast.2014.7042482	don't feel confident / difficult to use
Heer et al. 2007	10.1109/tvcg.2007.70539	demanding
Waldner et al. 2014	10.1109/tvcg.2014.2346352	slightly irritating / disturbing
Baumgartl et al. 2020	10.1109/tvcg.2020.3030437	training is needed / background color / highlighting important info / missing annotations / use of color / better differentiation

Continued on next page

Authors & Year	DOI	Terms collected
Bertucci et al. 2022	10.1109/tvcg.2022.3209425	intuitive / not clearly defined boundaries / easy to scan / getting lost / easy to extract what you are looking for
Borkin et al. 2013	10.1109/tvcg.2013.155	mental activity / stressful / confident / easy to navigate / easy to see the data / looks nice / familiarity / easy to understand
Braun et al. 2023	10.1109/tvcg.2023.3326576	cognitive load / choice of color / easy to distinguish / overlapping areas
Cashman et al. 2020	10.1109/tvcg.2020.3030443	intuitive / easy to navigate / straightforward
Chen et al. 2020	10.1109/tvcg.2020.3030338	overused colors / simplicity / intuitive
Chen et al. 2023	10.1109/tvcg.2023.3326925	intuitive / well organized / extensive information / number of annotations / amount of information / easy to understand / overlapping / appropriate distance / annotation placement / reasonable placement / fits normal reading habits
Chevalier et al. 2014	10.1109/tvcg.2014.2346424	crowded / challenging / overwhelmed
Choudhry et al. 2020	10.1109/tvcg.2020.3030358	easily understandable / textual narratives / unfamiliarity / challenging
Claessen et al. 2011	10.1109/tvcg.2011.201	non-intuitive
Chen et al. 2020	10.1109/tvcg.2020.3030411	wasted space / small items
Meuschke et al. 2016	10.1109/tvcg.2016.2598795	color-coding

Continued on next page

Authors & Year	DOI	Terms collected
Bae et al. 2014	10.1109/tvcg.2014.2346420	occlusion / confusing / direction information / dense region / simple / pleasant / not intuitive / missing arrows / easy to recognize
Bernard et al. 2017	10.1109/tvcg.2017.2744818	combination of color / overlays / difficult to distinguish / confused / non-regular shapes / distracted
Steinberger et al. 2011	10.1109/tvcg.2011.183	mentally demanding / clearly visible / highlighted information / visual clutter / visually pleasing / cluttered background / easy to miss elements / cluttered / well visible / easy to follow
Kim et al. 2016	10.1109/tvcg.2016.2598620	intuitive / simple / straightforward
Gotz et al. 2014	10.1109/tvcg.2014.2346682	confusing / disturbing / easy to find correlated events / intuitive / easy to use / very nice visually
Haehn et al. 2014	10.1109/tvcg.2014.2346371	pleasing
Dimara et al. 2017	10.1109/tvcg.2017.2745138	amount of information presented / overwhelming / easy / straightforward
Kondo et al. 2014	10.1109/tvcg.2014.2346250	frustrated / confusing / feeling lost
Heimerl et al. 2016	10.1109/vast.2016.7883507	takes time to mentally process / occlusion / color mapping
Dwyer et al. 2013	10.1109/tvcg.2013.151	cluttered / confidence / cross-boundary edges / nesting / confusing
Elhamdadi et al. 2023	10.1109/tvcg.2023.3326579	clear / easy / colored lines / simple / familiarity

Continued on next page

Authors & Year	DOI	Terms collected
Eulzer et al. 2020	10.1109/tvcg.2020.3030388	intuitive
Zhao et al. 2018	10.1109/tvcg.2018.2865020	axes / easy to handle / easy to locate items / visual clutter / annoying / many data points / crowded / too much information / distracting
Borkin et al. 2011	10.1109/tvcg.2011.192	confusing / symmetric / familiarity / occlusion / need of ordering / orientation / intuitive to 'read' / aesthetically pleasing / color encoding
Höferlin et al. 2012	10.1109/tvcg.2012.222	information overload / difficult to identify objects / too blurry / number of objects / familiarity / color changes / stress / difficult to focus
Rodgers et al. 2011	10.1109/tvcg.2011.196	less detailed / intuitiveness / distracting / feel anxious / too bright / aesthetics / abstract / artistic / dense / boring / crowded / confusing / busy / stressful
Falk et al. 2018	10.1109/tvcg.2018.2864816	tricky to grasp to much information / beautiful image / spending a lot of time / aligns with mental picture
Feng et al. 2023	10.1109/tvcg.2023.3327168	highly detailed / effectively grouped similar images together / reduced cognitive load / highlighting / straightforward
Filho et al. 2019	10.1109/tvcg.2019.2934415	tiring / intuitive
Wongsuphasawat et al. 2009	10.1109/vast.2009.5332595	useful / simple / intuitive / misleading / attractive
Isaacs et al. 2014	10.1109/tvcg.2014.2346743	scattered / hard to scan

Continued on next page

Authors & Year	**DOI**	**Terms collected**
Kim et al. 2009	10.1109/tvcg.2009.146	use of color, highlights, labels and messages / easy to identify missing values
Ghai et al. 2022	10.1109/tvcg.2022.3209484	user friendly / very intuitive / easy to understand
Ying et al. 2021	10.1109/tvcg.2021.3114877	easy / straightforward / easy to follow / intuitive / easy to understand
Gratzl et al. 2013	10.1109/tvcg.2013.173	visually pleasing / easy to understand / intuitive / hard to find the button
Lawonn et al. 2022	10.1109/tvcg.2022.3209374	hard to distinguish between individual gaussians / i can clearly explore and analyze high dimensional data points / extensive / clear
Behrisch et al. 2019	10.1109/tvcg.2019.2934300	useful / easy to understand
He et al. 2022	10.1109/tvcg.2022.3209390	beautiful / nice / attractive / engaging / motivating / interesting / clean / sophisticated / organized / creative / artistic / color harmonious / cluttered
He et al. 2023	10.1109/tvcg.2023.3326941	distinguishable / background color / orientation / icon style / density / outlies / halos / clear / readable / clutter / overwhelming elements / distracting / consistent visual style / visually balanced chart / consistent ink density / overlapping
Huang et al. 2022	10.1109/tvcg.2022.3209384	matches the mental model / intuitive
Huang et al. 2023	10.1109/tvcg.2023.3326932	information overload / feeling overwhelmed

Continued on next page

Authors & Year	DOI	Terms collected
Zhao et al. 2018	10.1109/tvcg.2018.2864475	visually pleasing / intuitive / easy to understand / need time to understand the visual encodings
Isenberg et al. 2011	10.1109/tvcg.2011.160	visual discontinuity / superimposed chart / confusing / demanding too much concentration
Jessup et al. 2021	10.1109/tvcg.2021.3114786	intuitive / overlaying / occlusion / easy to comprehend / easy to use
Zhao et al. 2013	10.1109/tvcg.2013.167	useful / textual labels / familiar / aesthetically pleasing / user-friendly / complicated / difficult to learn / confusing
Kim et al. 2023	10.1109/tvcg.2023.3327150	requires less effort / easier to use
Kraus et al. 2019	10.1109/tvcg.2019.2934395	comprehensive / intuitive
Krekhov et al. 2019	10.1109/tvcg.2019.2934370	easy to sport regions / colored data sets / straightforward / familiar
Latif et al. 2021	10.1109/tvcg.2021.3114802	intuitive / self-explanatory
Lawonn et al. 2015	10.1109/tvcg.2015.2467961	useful / confusing / essential color coding / easy to distinguish between classes / discrete color coding / continuous color coding
Lee et al. 2020	10.1109/tvcg.2020.3030435	annotations / too difficult to see / not enough information / easier to understand / cannot comprehend
Lee et al. 2020	10.1109/tvcg.2020.3030446	challenging / frustration
Lin et al. 2023	10.1109/tvcg.2023.3327170	intuitive
Lin et al. 2022	10.1109/tvcg.2022.3209353	easy to understand / helpful / fun to use
Linhares et al. 2022	10.1109/tvcg.2022.3209477	easy to use / intuitive

Continued on next page

Authors & Year	DOI	Terms collected
Liu et al. 2020	10.1109/tvcg.2020.3030364	clean / well designed / straightforward / easy to understand
Maher et al. 2021	10.1109/tvcg.2021.3114789	not comprehensive / ease of use
Koh et al. 2010	10.1109/tvcg.2010.175	difficult to learn / confusing / less boring / fun / requires too much effort
McDonald et al. 2020	10.1109/tvcg.2020.3030363	less fatiguing / challenging to learn
Morini et al. 2023	10.1109/tvcg.2023.3327185	topic engagement / helpful / motivating / impossible to navigate
Jang et al. 2015	10.1109/tvcg.2015.2468292	useful features / low cognitive load / intuitive
Mumtaz et al. 2019	10.1109/tvcg.2019.2934669	unclear annotations / hard time figuring out the meanings of the metrics / not self-explanatory / distracting information / intuitive / not clear how to use / bad placing / missing legend / missing caption
Narechania et al. 2021	10.1109/tvcg.2021.3114827	intuitive / helpful / confusing / color coding / all information in one space / hard to draw conclusions from / distracting
Liu et al. 2010	10.1109/vast.2010.5652910	color encoding / semantic layout / helpful feature / useful feature / intuitive / difficult to learn / confused / cannot distinguish / does not match mental model
Shi et al. 2023	10.1109/tvcg.2023.3326522	low cognitive load / overwhelmed / intuitive / user-friendly

Continued on next page

Authors & Year	DOI	Terms collected
Albo et al. 2015	10.1109/tvcg.2015.2467322	loaded visualization / confusing / close values / difficult to isolate a value point / pretty visualization / no visual load / color encoding / easy trend identification
Raidou et al. 2015	10.1109/tvcg.2015.2467872	easy to find correlations / overlap in the data / structure detection in the data
Smit et al. 2016	10.1109/tvcg.2016.2598826	clear visualization / intuitive understanding / confused
Piccolotto et al. 2023	10.1109/tvcg.2023.3327203	meaningful and accurate visual encodings / avoid misleading representations / facilitates perceiving relationships in the data / hard to spot
Sedlmair et al. 2012	10.1109/tvcg.2012.255	easily understandable
Bruckner et al. 2010	10.1109/tvcg.2010.190	too subtle highlighting / difficult identifying features
Chen et al. 2022	10.1109/tvcg.2022.3209385	straightforward / irritated / stressful
Robertson et al. 2008	10.1109/tvcg.2008.125	enjoyable / less cluttered / confusing / blurry
Ruan et al. 2022	10.1109/tvcg.2022.3209455	informative
Sambasivan et al. 2013	10.1109/tvcg.2013.233	straightforward / feature helps to distinguish
Shen et al. 2023	10.1109/tvcg.2023.3327197	tiring / visually consistent
Shi et al. 2020	10.1109/tvcg.2020.3030403	nice design / thoughtful design / helpful feature

Continued on next page

Authors & Year	DOI	Terms collected
Lee et al. 2013	10.1109/tvcg.2013.191	challenging / confusing use of color / engaging / interactive / annotations / readability / easier to read / legible / easier to see and read / familiarity / clean graphics / requires effort
Subramonyam et al. 2018	10.1109/tvcg.2018.2865231	labels close to each other
Liu et al. 2016	10.1109/vast.2016.7883513	intuitive
Lee et al. 2010	10.1109/tvcg.2010.194	cluttered
Tory et al. 2007	10.1109/tvcg.2007.70596	hard to distinguish between colors
Ren et al. 2016	10.1109/tvcg.2016.2598828	can scan a lot of information quickly / easy to spot errors / easy to understand / familiar / straightforward / lower learning curve
Alper et al. 2011	10.1109/tvcg.2011.234	helpful
Stokes et al. 2022	10.1109/tvcg.2022.3209383	clear / more understandable / provides better explanation / simple / lack of clutter / busy graph / ugly to look at / unnecessary visual clutter / eye-catching / short and concise titles
Sun et al. 2022	10.1109/tvcg.2022.3209428	intuitive
Hajizadeh et al. 2013	10.1109/tvcg.2013.197	effective visual encoding / occupying a big portion of the screen
Shi et al. 2023	10.1109/tvcg.2023.3327200	easy to understand / simplicity / expressiveness / intuitive / clear
Badam et al. 2016	10.1109/vast.2016.7883506	intuitive / useful feature
Taher et al. 2016	10.1109/tvcg.2016.2598498	intuitive / large dataset / data density / overwhelming / too much to process
Tang et al. 2021	10.1109/tvcg.2021.3114781	intuitive visualization design

Continued on next page

Authors & Year	**DOI**	**Terms collected**
Eulzer et al. 2019	10.1109/tvcg.2019.2934337	less ambiguous / easy to interpret
Teng et al. 2023	10.1109/tvcg.2023.3326587	simple / reduced cognitive load / presents information simply
Gramazio et al. 2014	10.1109/tvcg.2014.2346983	data point hard to find / data point close to the axis / data point surrounded by various different colors
Arendt et al. 2017	10.1109/vast.2017.8585487	confused / not enough context / lacking annotations / intuitive
Tennekes et al. 2014	10.1109/tvcg.2014.2346277	easier to interpret
Tyagi et al. 2022	10.1109/tvcg.2022.3209392	data dimensionality
Basole et al. 2013	10.1109/tvcg.2013.209	intuitive design / easy to use
Ens et al. 2020	10.1109/tvcg.2020.3030334	visually overwhelming / confusing / a lot of dots / a lot of stacks / a lot of colours / a lot of information / familiarity
Pienta et al. 2017	10.1109/tvcg.2017.2744898	easy to quickly compare variables
Lin et al. 2023	10.1109/tvcg.2023.3327161	clearly understandable data / access to 3d visualizations / leads to huge reduction on the time to perform analysis
Beck et al. 2015	10.1109/tvcg.2015.2467757	clear / good amount of information / compact representation / useful keywords / use of word clouds / useful sparklines / features are self-explaining / some features are hidden / display is too packed / display contains too much information

Continued on next page

Authors & Year	DOI	Terms collected
Kwon et al. 2011	10.1109/vast.2011.6102435	difficult to read / difficult to understand / poor presentation / lack of visual cues / unfamiliar view / unclear / visual clutter / entities are clustered / overwhelming / redundant / more spacious / everything is just thrown into a cluster
Deng et al. 2023	10.1109/tvcg.2023.3327162	misleading / intuitive / easy to understand / simple encodings
Wang et al. 2018	10.1109/tvcg.2018.2865232	well-designed / has clear user interface / easy to operate / visually pleasing / easy to read
Wang et al. 2019	10.1109/tvcg.2019.2934398	easy to understand / provides meaningful annotations / helpful short description / the visual design is quite rich / eye-catching
Wang et al. 2020	10.1109/tvcg.2020.3030433	well-illustrated
Wang et al. 2022	10.1109/tvcg.2022.3209357	well-designed
Wang et al. 2023	10.1109/tvcg.2023.3327153	intuitive assessment / overlapping entities / easy to use
Wang et al. 2022	10.1109/tvcg.2022.3209435	straightforward / visually appealing / painful to find useful information
Warchol et al. 2022	10.1109/tvcg.2022.3209378	intuitive / light color shades / difficult to interpret
Wei et al. 2019	10.1109/tvcg.2019.2934208	color coding / easy to judge / clear details

Continued on next page

Authors & Year	DOI	Terms collected
Wen et al. 2023	10.1109/tvcg.2023.3327148	easier to read / provides effective contextual information / aesthetics / clarity / usefulness of visual encoding / visually appealing / easy to use / color coding / highlighting / clear / intuitive / unfamiliarity
Wu et al. 2021	10.1109/tvcg.2021.3114826	intuitive
Xing et al. 2023	10.1109/tvcg.2023.3326923	small font size of axis labels / intuitive data interpretation
Yang et al. 2019	10.1109/tvcg.2019.2934557	familiar / acceptable / easier to perceive / line color / color helpful to distinguish / helpful labels / labels easy to read / colour coding
Yang et al. 2018	10.1109/tvcg.2018.2865192	difficult / visually appealing / difficult when data size increases / intuitive / felt very sparse / easy to find connections / very difficult to use
Yuan et al. 2020	10.1109/tvcg.2020.3030432	use of color
Zhang et al. 2022	10.1109/tvcg.2022.3209440	overwhelming information for first-time users
Zhao et al. 2017	10.1109/tvcg.2017.2744738	visually pleasing / time to understand the visual encoding / too small sectors in pie charts / easy to understand
Zhao et al. 2017	10.1109/tvcg.2017.2745279	arc links / avoid overlap
Zhao et al. 2022	10.1109/tvcg.2022.3209469	simple / exhausting to use / some details are easily overlooked / informative / light white lines / indistinguishable lines
Chen et al. 2022	10.1109/tvcg.2022.3209497	intuitiveness / easy to use

Continued on next page

Authors & Year	DOI	Terms collected
Zhu-Tian et al. 2023	10.1109/tvcg.2023.3326568	occlusion-free / label occlusion / difficult to read the labels / labels overlap a lot / distracting / hard to follow / layout not so appealing visually
Zytek et al. 2021	10.1109/tvcg.2021.3114864	too many factors listed / confusion / hard to interpret

A.2. Inductive method sources

Below, we provide the tables containing all terms collected in Section 3.1.2 and used to construct Pool 3, Pool 4, and Pool 5. Table A.4 lists all statements gathered from the expert survey in which participants described their definitions of visual complexity. For this survey question, participants were required to provide at least one response and could submit additional responses. Consequently, multiple entries may appear under the same user ID. In a similar format are also Table A.5 and Table A.6, for Pool 4 and Pool 5, respectively. The images provided in the survey as references are shown in Figure A.1.

Table A.4.: 75 experts' statements and terms extracted to form Pool 3.

Statement	Extracted terms
12228: "it uses a lot of marks and channels"	a lot of marks / a lot of channels
12228: "information is dense and overlapping"; "the gist is not clear within a few seconds"; "it has many elements and the legend or description is not placed directly beside them"; "color, shape, and size encodings are overloaded"; "there are multiple layers or competing areas of focus"	dense information / overlapping information / gist is not clear quickly / many elements / placement of legend. /placement of description / overloaded color, shape, and size encodings / multiple layers of focus
20419: "It is entangled or incluses many variables (not necessary many variations of a variable)"	entangled variables / many variables
24516: "your eyes have to move over the image multiple times to pull out changes in color, text, orientation, shape, markings in order to understand what the image is trying to tell you."	number of visual elements to scan / multiple visual encodings

Continued on next page

Statement	Extracted terms
24516: "it take a lot of time to understand what information the visualization is trying to convey and to reach a conclusion. "	requires a lot of time to understand / takes a lot of time to reach a conclusion
25066: "there is a large amount of visualize clues (lines, glyphs, etc.) suggesting that there is a large amount of visual information to be inferred by the viewer."	large amount of visual cues / number of lines / number of glyphs / large amount of visual information
27638: "There are lots of marks, organized in ways that are not immediately easy to grasp. "	a lot of marks / non-intuitive organization
30141: "the amount of graphical elements in the visualization is high and their mutual relationship complex "	high number of graphical elements / complex relationship between elements
30141: "the number and arrangement of the graphical elements causes a high cognitive load"	number of graphical elements / arrangement of graphical elements / high cognitive load
30219: "many (e.g., at least 4) visual variables are combined for may different data attributes (color, shape, size, position, text, etc.); its less problematic when encodings somehow correlate (e.g., in scatterplot color correlates to position in a way). "	many visual variables combined for many different data attributes / types of visual variables /
30219: "to many data points are displayed without implying some form of aggregation or simplification (e.g., many lines crossing, very dense scatterplot)"; "too many colors (or I guess other types as encoding, such as glyphs) are used for unorganized data - e.g., in scatterplot I think the use of color is acceptable, as it separates groups, but the word cloud there is no organization, colors are scattered (regardless of the fact that is seems to not have any meaning)"	too many data points without aggregation or simplification / many crossing lines / very dense scatterplot / too many colors / encoding used for unorganized data / colors are scattered
30890: "there is no one obvious global "shape"; that is, it is made up of several smaller visual entities"	not one obvious overall patterns / several small visual entities

Continued on next page

Statement	**Extracted terms**
32296: "higher cognitive effort to understand the visualization type (e.g., visual encoding, usage of color)"	cognitive effort / visualization type / visual encoding / usage of color
36292: "I think of visual complexity in the context of data visualization design as the amount of detail or intricacy in a visualization, characterized by the complexity of underlying data, attributes of the visualization as an image (such as edge density), and aspects of the visualizations design."	amount of detail or intricacy / complex underlying data / attributes of the visualization / edge density / design aspects
36292: "I imagine a visualization will be perceived as more visually complex when 1) the information underlying it is high-dimensional or esoteric, 2) the visual design is unfamiliar or contains many component parts to interpret, or 3) the visualization does not arrange information in a structured way that allows for intuitive interpretation."; "The expertise and background of the viewer likely matters significantly when considering how visually complex a visualization is"	high-dimensional or esoteric information / unfamiliar visual design / many parts to interpret / information not arranged in a structured way / non-intuitive interpretation
41999: "there is no rational sorting/ordering of the information to be able to process them easily."; "there is too much overplotting/cluttering going on."; "I cannot process the information relatively quickly and I have to think and reflect in depth to understand its meaning (and use)."; "not enough details are provided to me (e.g., proper legends, axes info) to understand them (see the color scale in the geographical map on the right)."	information cannot be processes easily / no rational sorting or ordering / overplotting / clutter / cannot process the information quickly / not enough details are provided / missing legend / missing axes info
41999: "data are represented in a more complicated way than necessary (e.g., 3D bar chart)."	overly complicated representation
42165: "several different types of data are displayed at the same time"	displaying different types of data

Continued on next page

Statement	Extracted terms
42165: "understanding the overall message takes careful observation (not glanceable)"; "patterns may occur at different levels of detail"	need of careful observation to understand the overall message / patterns occur at multiple levels of detail
44003: "I cannot quickly grasp which visual channel encodes information."	ambiguity of visual encoding
44003: "even though I understand the visual channels used in the visualization, I cannot quickly understand how they encode the data variation (high or low in data)."; "even though I understand visual channels and how they present data variation, I cannot detect any visual patterns that stand out."; "even if I was able to do all the above, I cannot translate this visual pattern to domain-specific insights."	cannot quickly interpret data variation from visual channels / cannot detect any visual patterns that stand out / cannot translate visual patterns to domain-specific insights
45694: "there are a large number of marks / data points and the marks do not coalesce into simple patterns."	large number of marks / large number of data points / not simple patterns
46329: "it has many visual marks and shapes of varying colors and styles."	many visual marks / many shapes / many colors / many styles
47064: "the information are encoded in ineffective channel for it"s intended the purpose"	ineffective visual channels for encoding information / mismatch between visual channel and purpose
47064: "there are too much information shown"	too much information shown
51657: "number and diversity of visual marks + visual clutter."	number of visual marks / diversity of visual marks / visual clutter

Continued on next page

Statement	Extracted terms
51798: "Preamble (sorry): Obviously, complexity is a complex concept and phenomenon itself: it has many parts, many internal and external relations, itš dynamic, sort of everywhere (both as a phenomenon and concept) and thus to grasp only selectively (as argued from H. Simon to N. Luhmann) and in a contested fashion. While its study has generated hope for a new trans disciplinary field of science, the very complexity / diversity of related efforts has practically also burried the endeavor. I think this has consequences for establishing any "validated scale" or measure of complexity, which can only be achieved by becoming a complex measure / scale itself - or by deliberately embracing selectivity, simplification, and lack of overall consensus."	no terms found

Continued on next page

Statement	Extracted terms
51798: "Complexity of visualizations: Can be understood as a variable which depends on data- and design-related factors, but also perception- and cognition-related ones. It tends to rise or vary with focused data dimensionality and volume, number of encoding channels / facets / views, display density, navigation / interaction options, but also with (the lack of) contextual information, data / visualization literacy, and prior knowledge."; "Productive simplification: A visualization is visually complex when it asks to much of its user. (Background: Commonly, the complexity of the data / world is the very reason we want to have visualizations providing overview, focus of attention, or the signal within the noise. Visualizations which reproduce too much of the initial problem fail their job.)"; "More general simplification: Complexity can be productively understood as a "problem marker" - we use it to highlight phenomena which are "not easy" to grasp - relative to the (perceptual, cognitive, or scientific) standard tools of a community. If you are in the business of making things accessible / comprehensible / intelligible (for a specific target group), you want to avoid it in general. "; "It makes sense that the dominant belief system in the VIS design community is minimalism (data/ink=>max, chart junk=>min, visual complexity=>min) as it makes sense to provocatively and productively challenge it whenever needed (see a growing body of anti-tuftean literature), as (visual) complexity can be understood as a design factor or material itself to be handled or wielded to all sorts of ends. "	no terms found

Continued on next page

Statement	Extracted terms
53610: "the data to be visualized is complex (size, dimension, ...) and/or its visual representation is complex (density, dimension, familiarity, richness of interactions, ...)."	data complexity / data size / data dimension / complex visual representation / density / dimension / familiarity / richness of interactions
54172: "the visual representations used are confusing and not appropriately chosen for the given data and tasks."; "the designer uses visualization methods that are overly complex for the given data and a simpler and more intuitive solution would be more appropriate."; "there is too much visual clutter."; "is uses too many colors and one cannot distinguish between them."	confusing visual representations / visual representations not appropriately chosen / overly complex visualization methods / too much visual clutter / too many colors / cannot distinguish between colors
54172: "the user cannot quickly and easily understand the main message of the visualization."	difficult to interpret / cannot understand quickly the main message / cannot understand easily the main message
55536: "there are too many primary colors (>7), too many points or lines such that gaps in clusters are no longer visible or structure is hidden, too many variables, visualization paradigm is hard to discern "	to many primary colors / too many data points / too many lines / structure is hidden / gaps in clusters are no longer visible / too many variables / visual paradigm is hard to discern
56041: "it contains a great amount of data points"	a great amount of data points
56041: "it contains multidimensional data"; "it contains multiple levels of granularity"; "it is difficult to perceive and understand"	multidimensional data / multiple levels of granularity / difficult to perceive / difficult to understand
58688: "it employs a large number of independent visual encodings"; "it is visually noisy, with high variability in the color or intensity of pixels, and limited predictability or visual continuity within the visual space of the chart"	large number of independent visual encodings / visually noisy / high variability in color / high variability in intensity of pixels / limited predictability
58688: "it has a large number and/or high density of distinct visual elements"	large number of visual elements / high density of distinct visual elements
59684: "has too few or too many items, difficult to understand (e.g., lack annotation, legend), or cluttered. "	too few or too many items / difficult to understand / lack of annotations / lack of legend / cluttered
61092: "there is little to no white space"	little to no white space

Continued on next page

Statement	Extracted terms
61092: "there is overlapping content"; "it is a 3D projection"	overlapping content / 3d projection
61544: "there are many items (e.g., points, lines) which have very similar visual weight/prominence"	number of items / number of points / number of lines / items with similar visual weight / items with similar visual prominence
63735: "it uses encoding channels that require a lot of mental processing to understand (e.g., 3D, angle, area, ...)"	encoding channels require high mental processing
63735: "it uses a lot of channels to encode data"	number of visual channels
64716: "it is difficult to read"	difficult to read
64716: "it is time-consuming to extract information"	time-consuming to extract information
64800: "the ink rate is high"	high ink rate
65237: "due to design decisions the visual needs higher mental effort to fully grasp. (e.g. the scatterplot is visually simple, while the grouped bar chart is visually medium complex: even though it uses a well known idiom and has not too many visual elements, it still is complex due to its weird grouping with some elements appearing in multiple groups and others not, and the three green shades)"	high mental effort / not too many visual elements / weird grouping of elements / elements appearing in multiple groups / use of three green shades
65237: "there are many visual items to decipher. (e.g. the tornado visualization is visually complex as there are many data points, it has many high contrast edges even though the overarching pattern can be easily seen: the counter-clockwise swirl)"	many visual items to decipher / many data points / many high contrast edges
67023: "there is too many things going on (e.g., too many colors, complex shapes, and too much information) and lacking intuitive means for understanding."	too many things going on / too many colors / complex shapes / too much information / lack of intuitive understanding
67485: "there are a lot of colors and labels."	a lot of colors / a lot of labels

Continued on next page

Statement	Extracted terms
69028: "there appears to be a lot of intricate details plotted on the visualization, including a large number of visual elements or text"	a lot of intricate details plotted / large number of visual elements / amount of text
70671: "1) it has a lot of dimensions 2) it has a lot of visual elements (including chart junk but not necessarily) 3) it has a lot of graphical marks / visualizes a lot of data items 4) shows a difficult to understand relationship"	number of dimensions / number of visual elements / chart junk / a lot of graphical marks / a lot of data items / relationship difficult to understand
71760: "the amount of information is excessive"; "the data contains complex insights"; "visualization results in visual clutter"	excessive amount of information / data contains complex insights / visual clutter
71760: "colors are overly abundant"	overly abundant colors
74185: "its elements, relationships, or patterns are difficult to perceive, interpret, or mentally organize due to high information density, numerous variables, intricate visual encodings, or unclear structure."	elements, relationships, or patterns are difficult to perceive / elements, relationships, or patterns are difficult to interpret / elements, relationships, or patterns are difficult to mentally organize / high information density / numerous variables / intricate visual encodings / unclear structure
74598: "when it utilizes multiple visual channels. when it represents high-dimensional data, with each dimension having multiple variables. when the visualization display involves occlusion and clutter."	multiple visual channels / high-dimensional data / each dimension has multiple variables / occlusion / clutter

Continued on next page

Statement	**Extracted terms**
74766: "Everyone can recognize complexity, but I doubt anyone could verbalize it. Imagine asking "An object is red when..". How would a person respond to that or even put it to words? Any so-called expertise in visualization wouldn't help. These articles have attempted to define complexity: Curious Objects: How Visual Complexity Guides Attention and Engagement - Zekun Sun, Chaz Firestone. Predicting perceived visual complexity of abstract patterns using computational measures: The influence of mirror symmetry on complexity perception - Andreas Gartus, Helmut Leder."	no terms found
81230: "A data visualization is visually complex when it contains a large number of elements, with non-trivial and irregular spatial or connected relationships to one another, encoding a lot of information, that takes time to understand."	a large number of elements / amount of information encoded / takes time to understand / complex spatial or connected relationships
83249: "it uses lots of colors"; "it uses non-traditional encodings"; "it contains lots of text"	a lot of colors / non-traditional encodings / contains a lot of text
83249: "it is overplotted"	overplotted
83936: "there is a richness to the visual depiction that requires time and effort to make sense of."	richness of visual depiction / requires time to understand / requires effort to interpret
83936: "there are many details or high frequency variation that may have meaning that is useful to interpret."; "meaningful pattern and structure in the visual depiction is evident at several scales."	too many details / high frequency variation / evident patterns at several scales

Continued on next page

Statement	Extracted terms
84167: "the core insights available in the visualization require nontrivial cognitive load to ascertain."; "(unnecessary complexity) the amount of visual processing required to make sense of the visualization is disproportional to the depth and volume of insights conveyed. "; "(unnecessary complexity) the amount of visual information impairs one"s abilities to make sense of that information."	core insights require high cognitive load / excess visual information impairs understanding / visual processing demands disproportionate to insights conveyed
84167: "insights can be gained through deep and layered reading. "	insights require deep reading / insights require layered reading
88985: "encoded information exhibits elaborateness and intricacy in terms of quantity and interrelationships"	amount of encoded information / complex interrelationships / elaborateness of data representation / intricacy of visual encoding
89777: "it contains a high frequency (or number) of colors, textures, or shapes or any combination thereof."	high number of colors / high number of textures / high number of shapes
95894: "lots of visual elements distract my attention"	a lot of visual elements / distracting
95894: "visual elements overlap with each other"	overlapping visual elements
96252: "it attempts to show a lot of data and uses a complicated / unfamiliar layout. Hence, large data in a familiar layout can be of low complexity (e.g., a map of a familiar region with thousands of temperature values mapped), but small data can also be complex if the representation is complex, e.g., results in lots of clutter, such as edge crossing. "	a lot of data / complicate layout / unfamiliar layout / familiar layout / complex representation / lots of clutter / edge crossing
96560: "It takes much time to understand the visualization"	takes time to understand the visualization
96560: "the phenomena shown are very unfamiliar to me"; "The visualization is highly cluttered / noisy /messy"; "The visual encoding chosen is poor and violates good practices (messy icons, poor color scales, etc.)"	unfamiliar topics / highly cluttered / noisy visualization / messy visualization / poor visual encoding / visual encoding violates good practices / messy icons / poor color scales
97095: "a non-standard visualization technique / chart type."	non-standard visualization technique / non-standard chart type

Continued on next page

Statement	Extracted terms
97095: "it contains visual clutter."; "it uses 3D for encoding abstract data."; "it contains lots of details without appropriate labels/groupings/annotations."	visual clutter / a lot of details without appropriate labels / a lot of details without appropriate groupings / a lot of details without appropriate annotations / 3d
98968: "There are extraneous elements/features in the display that do not play a role in facilitating interpretation"; "The pattern(s) is(are) difficult to perceive"	extraneous elements in the display that do not play a role in facilitating interpretation / pattern difficult to perceive
98968: "There are multiple, overlapping data series within a single plot"	multiple data series / overlapping data series / multiple data series in a single plot
99658: "it has many rows and columns in case of tabular data, many datasets, many points in time, many crossing lines in case of graph-based visualizations, a high density of visual marks, several areas with high density"	tabular data with many rows and columns / many datasets / many points in time / many crossing lines / a high density of visual marks / several areas with high density

Table A.5.: Terms provided by the experts to form Pool 4.

User ID	Terms
12228	"uses too many different colors, shapes, or sizes at the same time"; "makes the main message clear within a few seconds"; "feels clean and easy to scan"
12228	"requires me to look back and forth between the visual and its legend or description"; "feels crowded with overlapping information"
20419	"I can easily discriminate data points in this visualization"; "I can see the different variables in this visualization "; "The graph is adequate to the type of data shown"
24516	"has too many color changes"; "does not have enough color changes"; "has too much text"
24516	"does not have enough context"; "is easy to find the important feature"
25066	"maximizes visual "signal" while minimizing "visual noise"."; "connects visualization complexity with visual complexity well; I as the viewer am spending my time digesting the "right" areas."; "makes the most of visual complexity to relay its point."
27638	"has lots of details"; "takes time to understand"; "looks daunting"
30141	"easy to understand"; "uncluttered"; "provides the right level of detail"
30219	"is too detailed"; "is confusing"; "has too many visual elements"

Continued on next page

User ID	Terms
30219	"is easily readable"; "is informative"; "is too simple"; "contains too much information"; "is easily understandable"; "is misleading"; "is visually appealing"; "is distracting"; "is accurate"; "is clear"; "has too much going on"
30890	"is based around a single obvious shape"; "uses too many colors"; "presents too much data at once"
32296	"rich in information"; "overwhelming"; "hard to understand"
36292	"contains a lot of detail"; "is difficult to read"; "is highly intricate"
41999	"contains many distinct elements to process."; "has unnecessary complexity that hinders comprehension."; "has a level of detail that helps me understand the data."
41999	"has a clear and organized layout."
42165	"requires careful observation to understand"; "encodes information in an intuitive way"; "is showing too much data at once"
42165	"is cluttered"
44003	"is difficult to understand"; "does not show patterns well"; "(unintentionally) hides the underlying information"
45694	"a complex visual appearance"; "visual complexity"; "looks visually complicated"
46329	"has many visual marks."; "is complex to interpret."; "has a lot going on."
47064	"complex"; "hard to read"; "difficult to understand"
51657	"complex"; "cluttered"; "easy to read"
51798	"is hard to understand in total"; "is asking more than I"m usually willing to invest (as a user type x / y / z) into reading it"; "is perceptually overly challenging"
53610	"displays too many primitives"; "is too sparse to represent the data faithfully"; "can be understood without interaction"
54172	"is visually pleasing"; "uses appropriate color palette(s)"; "gives you the overall understanding of the domain in focus and the main purpose of the visualization"
55536	"confusing"; "difficult to infer the message"; "difficult to interpret"
56041	"contains a lot information"; "requires quite some effort to identify all detail"; "contains several aspects of the data "
56041	"requires some time to read"; "shows several dimensions at once"
58688	"is overly complicated"; "has too much information to be usable"; "is trying to convey too much in the given space"
58688	"is difficult to parse"; "is difficult to read"; "can be understood at a glance"; "has a lot going on"; "will take a lot of time and effort to correctly read"

Continued on next page

User ID	Terms
59684	"include a title so the viewer can grasp the topic quickly to support an "overview from text" and details from graphics drawings style. or the other way around: overview from graphics and details from text. (the former is better I think)"; "whenever possible, add a schematic view. e.g., the magnified view helped reduce vc . "; "bar is simple."
59684	"the scatterplot without axis labels is just a bad drawing. "
61092	"I can read this visualization quickly."; "This visualization is easy to understand."; "The visualization requires like mental effort."
61092	"I can see all the data points on this visualization. "; "I need to be an expert to understand this visualization."; "I can recreate this visualization by drawing/sketching it."
61544	"is busy"; "pulls my attention in many directions"; "has a lot going on"
63735	"uses many channels to encode data"; "requires a lot of mental effort to understand"; "is easy to make sense of"
64716	"clutter"; "clarity"; "readable"
64800	"cluttered"; "overloaded"; "too many visual channels"
64800	"too many components"
65237	"many visual items (marks, labels, axes, views)"; "tidy"; "cluttered"
65237	"too many details competing for my attention"
67023	"hard to read"; "hard to understand"; "hard to know where I should look at"
67485	"This visualization is easy to understand."; "I want to spend time trying to decode the visualization."; "This visualization is off-putting."
69028	"has many visual and textual elements"; "is effortful to read"; "takes time for me to understand"
70671	"visually overwhelming"; "many marks"; "unnecessary visual elements / chart junk"
70671	"a lot of data"
71760	"is easy to read / interpret (perceive)"; "is easy to comprehend "; "is easy to remember "
74185	"This visualization contains many elements that make it appear dense or crowded."; "This visualization has a structure that is difficult to interpret at first glance."; "This visualization displays information in a way that feels intricate and detailed."
74598	"This visualization utilizes multiple visual channels"; "This visualization has significant amounts of occlusion and clutter."; "This visualization is displaying a large number of dimensions each with multiple variables."
74766	"I would not"; "I would not"; "I would not"
81230	"a lot of information"; "is difficult to understand"; "would take time and effort for me to extract insight"

Continued on next page

User ID	Terms
83249	"My eyes don"t know here to look in the visualization."; "Data cases can be individually interpreted."; "The visualization contains the information necessary to decode its meaning."
83936	"requires significant effort to understand"; "has very many meaningful details "; "shows meaningful patterns at many spatial scales"
84167	"conveys a range of information about the data"; "enables a number of different insights into the data"; "would benefit from additional graphical elements to better communicate the data"
84167	"could communicate the same ideas using fewer visual elements"; "offers a clear workflow by which I would analyze the data"
88985	"many interconnected components"; "intricate pathways or relationships"; "multiple layers of abstraction or granularity"
89777	"a high number of colors"; "a high number of shapes"; "a high number of features"
95894	"easy to understand"; "too many elements"; "no engaging focal point"
96252	"complex."; "cluttered."; "difficult to read."
96560	"a mess"; "hard to understand"; "elegant"
96560	"poorly designed"; "visually complex"
97095	"is hard to understand"; "is complex"; "has too much detail"
98968	"appears to show a clear pattern "; "has overlapping visual components "; "is well organized (i.e., perceptual organization)"
99658	"easy to understand"; "easy to interpret"; "aesthetically pleasing"

Table A.6.: Adjectives provided by the experts to form Pool 5.

User ID	Terms
12228	"overwhelmed"; "frustrated"; "distracted"
12228	"curious"; "interested"; "engaged"; "lost"
20419	"confused"; "I can read some pattern or phenomenon in it"; "I need a magnifier "
24516	"anxious"; "confident"; "overwhelmed"
24516	"informed"; "knowledgeable "
25066	"lost (unable to figure out on what to focus)."; "drawn in (captures my attention and provokes me to think more)."; "bored (not really interesting)."
27638	"fatigued"; "intrigued by all the details"; "confused"
30141	"confused"; "overloaded"; "stressed"
30219	"confused"; "informed"; "engaged/interested"
30219	"exhausted"; "distracted"
30890	"confused"; "overwhelmed"; "intrigued"
32296	"overwhelmed"; "stressed"; "time-consuming"
36292	"overwhelmed"; "bored"; "confused"

Continued on next page

User ID	Terms
36292	"understanding"; "clarity"
41999	"confident."; "confused."; "cautious."
41999	"uncertain."; "irritated."; "lost."; "stressed."; "engaged."
42165	"anxious"; "frustrated"; "confused"
42165	"confident"
44003	"puzzled"; "curious"; "overwhelmed"
45694	"overwhelmed"; "interested"; "intimidated"
45694	"engaged"
46329	"unsure"; "overwhelmed"; "confident"
47064	"overwhelmed"; "confident"; "in control"
51657	"confused"; "overwhelmed "; "confident"
51798	"tired"; "frustrated"; "appalled"
51798	"bewildered"; "attracted"; "interested"; "impressed"; "overwhelmed"
53610	"overwhelmed"; "intellectually bored "; "wanting to explore the data"
54172	"confident"; "distracted"; "discouraged to explore it in more detail"
54172	"calm/relax; as it is very visually pleasing"
55536	"inadequate"; "uneducated"; "nauseated"
56041	"challenged"; "exhausted"; "impressed"
56041	"overwhelmed"
58688	"overwhelmed."; "confident I am getting all the relevant information."; "like I am missing important details."
58688	"strained."
59684	"science is interesting (wanting to know the science behind these)"; "images are beautiful "; "3D is powerful"
61092	"confused"; "confident"; "informed"
61544	"distracted"; "confused about what to look at"; "interested in knowing more about the topic"
63735	"stressed"; "empowered"; "smart"
64716	"confused"; "overwhelmed"; "disoriented"
64800	"overwhelmed"; "confused"; "lost"
64800	"disoriented"
65237	"overwhelmed"; "intrigued"; "engaged"
65237	"confused"
67023	"disoriented"; "confused"; "overloaded"
67485	"curious"; "confused"; "dumb"
69028	"overwhelmed"; "that I need to slow down to read it in detail"; "engaged "
70671	"overwhelmed"; "confused"; "impressed"
71760	"overwhelmed"; "confused"; "distracted"
71760	"intrigued"; "mentally fatigued"; "curious"; "cognitively overloaded"; "lost in the details"
74185	"overwhelmed"; "intrigued"; "confused"
74598	"confused"; "overwhelmed"; "frustrated"

Continued on next page

User ID	Terms
74766	"I would not"; "I would not"; "I would not"
81230	"overwhelmed"; "impressed by the complexity of the data"; "informed"
83249	"overwhelmed."; "confused. "; "stressed."
83936	"curious"; "motivated"; "stimulated"
84167	"overwhelmed"; "informed"; "confused"
88985	"overwhelmed"; "confused"; "curious"
89777	"confused"; "overloaded"; "perplexed"
95894	"informed"; "curious"; "confused"
95894	"overwhelmed"
96252	"nervous "; "calm"; "overwhelmed"
96560	"happy"; "confused"; "irritated"
96560	"angry"
97095	"confused"; "dizzy :) "; "engaged"
98968	"confused"; "anxious"; "frustrated"
98968	"intrigued"; "curious"
99658	"informed well"; "satisfied (because of a clear display)"; "stressed (because of difficulties in interpretation)"

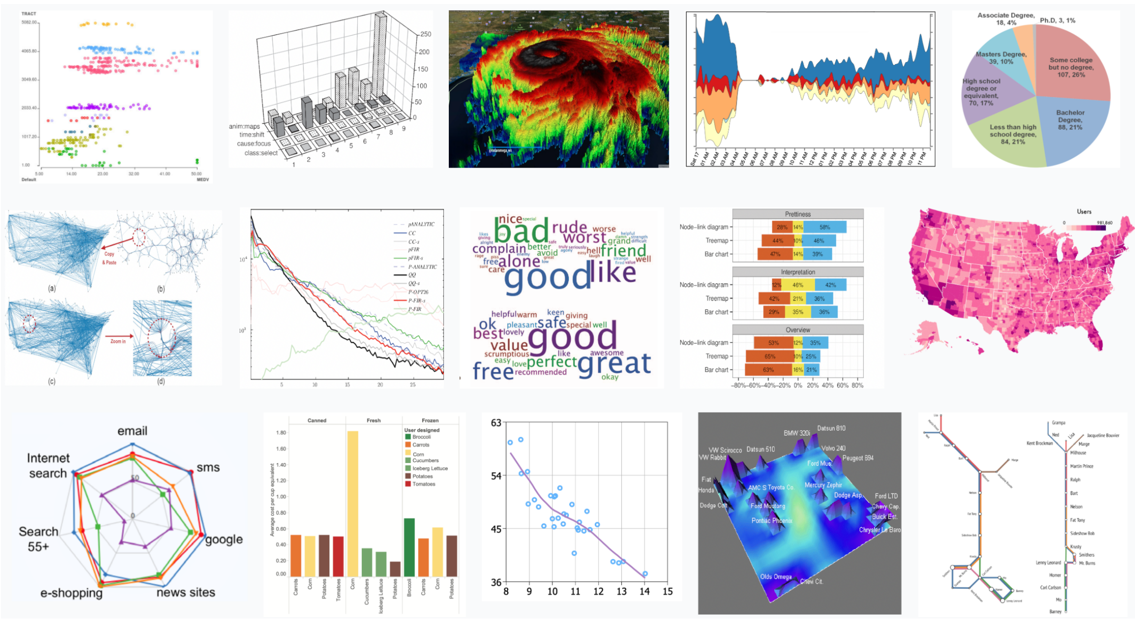


Figure A.1.: 15 images used as a reference point in the expert survey.

A.3. Item generation

Below, we provide more information on the process described in Section 3.2.3. For reasons of space, in Table A.7 we will show only the *primary terms* and the items we generated.

Overall, out of 51 *primary terms*, we count only 49 unique items as duplicates were not considered in the following step of the project. The underlined items are duplicates; only one was kept.

Table A.7.: Generated items and their primary terms

Primary term	**Item**
distracting	I find many visual elements distracting.
occludes information	I find the visualization occludes information.
cluttered	I find the visualization cluttered.
familiar	I find this visualization familiar.
easy to understand	I find this visualization easy to understand.
easy to interpret	I find this visualization easy to interpret.
clear	I find that the structure is unclear.
confident	I feel confident about this visualization.
difficult to read	I find this visualization difficult to read.
confused	I feel confused by the visualization.
overwhelmed	I feel overwhelmed by the visualization.
curious	I feel curious about this visualization.
confident	I feel confident about this visualization.
irritated	I feel irritated by this visualization.
a lot of information	I find there is too much information in this visualization.
easy to use	I find this visualization easy to use.
intuitive	I find this visualization intuitive.
visually pleasing	I find this visualization visually pleasing.
cognitively overloaded	I feel that this visualization causes cognitive overload.
mental demand	I feel the mental demand of this visualization is high.
color coding	I find the color coding overly complex.
visual encoding	I find the visual encoding is poor and violates good practice.
attractive	I find this visualization attractive.
unnecessary complexity	I find the visualization has unnecessary complexity.
easy to comprehend	I find the visualization easy to comprehend.
confused	I feel confused by the visualization.
overloaded	I find the visualization overloaded.
data size	I find the large number of data items makes the visualization too complex.
"dense	I find the visualization dense. I feel the visualization is dense.

Continued on next page

Primary term	Item
" informed	I feel informed by the visualization.
overlapping information	I find the visualization has overlapping information.
overwhelmed	I feel overwhelmed by the visualization.
stressed	I feel stressed by the visualization.
visualization type	I find the visualization type requires higher cognitive effort.
engaged	I feel engaged with the visualization.
too many colors	I find the visualization uses too many colors.
not enough context	I find the visualization provides not enough context.
well organized	I find the visualization well organized.
takes time to understand	I find this visualization takes time to understand.
easy to distinguish points	I find the points in the visualization easy to distinguish.
cannot distinguish between colors	I cannot distinguish between colors.
several dimensions at once	I find the visualization displays several dimensions at once.
i can detect patterns	I can detect patterns in the visualization.
missing important details	I find the visualization is missing important details.
effort	I find the visualization requires effort to understand.
poor pattern visibility	I find the visualization has poor pattern visibility.
all data points visible	I find all data points are visible in the visualization.
variables visible	I find the variables in the visualization visible.
poorly designed	I find the visualization is poorly designed.
complex visual appearance	I find the visualization has a complex visual appearance.
additional graphics help communication	I find that the visualization needs additional graphics to communicate effectively.

A.4. Item validation

In this section, we will give more details about the next step of the project. As noted in Section 3.3, the survey shared with the experts included six visualizations that served as stimuli. These visualizations are presented below in Figure A.2.

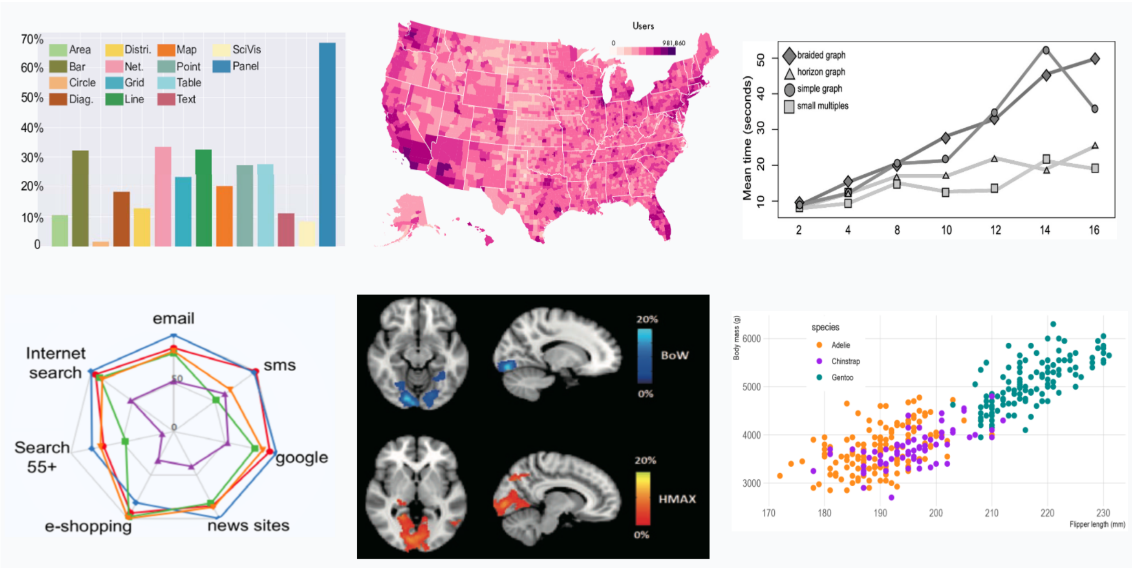


Figure A.2.: 6 images used as stimuli in the expert survey.

A.4.1. Reported comments

Below, we present the comments provided by experts at the end of our survey. Out of 45 participants, only 18 of them left a comment. In Section 3.3, we addressed those remarks that may require further clarification or that were not clearly explained in the survey prompt. However, these represent only a subset of the total comments received and are presented in Table A.8.

Table A.8.: Comments provided by the experts

User ID	Comment
60724	"Just wanted to give you some feedback on the study design, although it's likely too late to do anything about it, but: I was confused with the statements in the table, because some are positive (e.g., familiar, confident) and some are negative (e.g., cluttered, difficult to read). It makes the whole thing confusing because my brain can't help but see these sentences with a direction (positive/negative). And I actually feel like this is influencing my responses. I would solve this by having both the negative and the positive wording in each sentence (e.g., I find / don't find many visual elements distracting, or I find it easy/hard to use). Not the prettiest, but would address this issue (for me)."
59470	"the star plot missed a legend. the bar chart has too many categories, and the choropleth map needs more contrast"

Continued on next page

User ID	Comment
70168	"I found the questions somehow overlapping each other a bit while completing the survey. I think it might be better if the questions are organized into several themes and the participants are informed beforehand about what they're expecting to see. That might make me more confident in my answers."
94073	"An interesting paper on this topic: M. Chu et al., What Makes a Visualization Image Complex?,"" in IEEE Transactions on Visualization and Computer Graphics, vol. 32, no. 1, pp. 374-384, Jan. 2026, doi: 10.1109/TVCG.2025.3633827."
69034	"I completed this by imagining a person making the statement and me evaluating whether it indicates a problem of visual complexity. I hope I did it right! I found your description above a bit confusing."
45127	"As I'm looking at these it seems to me like there's a tension between phrases which use technical terms but might be quite helpful and those which are less precise but I'd expect a broader audience to understand – which I imagine is part of what you are studying! I'm also wondering if there may be different groups of phrases which are more or less relevant depending on whether complexity is desirable or not."
82144	"I am not sure whether I answered correctly, as it is hard to not distinguish between positive and negative influences on perceived complexity. It is unclear why sentences that include the term *complex* would not automatically be in the *very relevant* category?"
15569	"*I find the large number of data items makes the visualization too complex.* - strange as complexity is specifically mentioned. I answered *I find the vis has a large number of data items.*"
56082	"Also *I find the visualization has a complex visual appearance.*"
43947	"I found it difficult to give *correct* answers, I think it would be easier if I had to provide statements about concrete cases. And, qualifications can be triggered by many aspects; Take the fourth statement here *I feel irritated by this visualization*: My irritation can be triggered by the visualization being too simple (why use a visualization to show 1 number) or too complex (this is a complete mess) or ignoring perception (nobody told you that yellow on white is a bad idea), where is the effing legend, etc. Meanwhile, it is an interesting topic, and I am curious what will come out of your research!"

Continued on next page

User ID	Comment
10306	"It is difficult to answer these questions without seeing concrete image examples—potentially all of the listed factors may play a role to a certain degree. Evaluating the degree is somewhat challenging. For example, a bar chart can be *poorly designed* yet still easy to understand, and therefore not necessarily visually complexity. In contrast, a poorly designed parallel coordinates may be far more complex. Because of this, making an absolute judgment about each individual dimension is somewhat challenging. The conclusion seems to depend on what I am to see. What does *unnecessary complexity* mean on page 3? Are *occludes information* and *overlapping information* referring to similar concepts? *poorly designed* seems to operate at a higher level than the other concepts on the list. This is an interesting experiment. Thanks for inviting me."
69942	"The terms containing complexity feel like they might lead to a circular definition. Best of luck with the study!"
45306	"As no definition of *complexity* is given, it is not so easy to assess the statements."
70979	"Depends on the medium and context. Mobile vs laptop /on-the-go. What is the difference between *mental demand* and *cognitive overload*?"
23597	"You ask *How relevant do you think the following terms are for describing the perceived visual complexity of a visualization?*, but then you present entire sentences... (I think that there's actually a mismatch between the question that you're asking and the rows for which you wish the survey participants' relevance judgment). Furthermore, I do not find this survey distinguishing carefully enough between *high perceived visual complexity* and *perceived visual complexity* – the perceived visual complexity of a visualization can be low (or high [or ...]). A statement like *This visualization clear* could be relevant for a appropriately large perceived visual complexity."
12205	"There's a distinction here between good complexity and bad complexity that some of your items start to get at. I answered the questions based primarily on the concept of "bad" complexity."
37437	"Interpreting whether *complexity* is just one side of the spectrum or the name of the whole spectrum (i.e. including simplicity) made handling this survey rather complex - I went with the latter interpretation."
12244	"The rating was difficult to do because they can be applied to both highly complex visualizations and simple visualizations, and I have to give a high rating for expressing both complexity and simplicity. It would have been easier to separate these two questions in positive / negative bins. "

The feedback received by email is presented below, while preserving the experts anonymity.

" I've completed the survey, but I fear that it may have some problems. In

the following, I share two points that came to my attention:
– You ask *How relevant do you think the following terms are for describing the perceived visual complexity of a visualization?*, but then you present entire sentences... (I think that there's actually a mismatch between the question that you're asking and the rows for which you wish the survey participants' relevance judgment).
– Furthermore, I do not find this survey distinguishing carefully enough between *high perceived visual complexity* and *perceived visual complexity* – the perceived visual complexity of a visualization can be low (or high [or ...]). A statement like *This visualization clear* could be relevant for a appropriately large perceived visual complexity. "

A.4.2. Negatively worded items

In Section 3.3.1, we explain how we reverse-coded the negatively worded items. The results are shown in Table A.9. Items marked with an **N** in the Context column are the negatively worded ones. The list is ordered from highest to lowest average score and rounded to three decimal places.

Table A.9.: Comparison of scores for negatively vs positively worded items

Item	Context	Average Score
I find this visualization easy to interpret.	**P**	4.134
I find this visualization easy to understand.	**P**	3.978
I find the visualization easy to comprehend.	**P**	3.912
I find the visualization well organized.	**P**	3.689
I feel irritated by this visualization.	**N**	3.578
I find this visualization intuitive.	**P**	3.556
I find this visualization easy to use.	**P**	3.534
I find the points in the visualization easy to distinguish.	**P**	3.356
I find the visualization has poor pattern visibility.	**N**	3.267
I feel stressed by the visualization.	**N**	3.223
I can detect patterns in the visualization.	**P**	3.223
I find this visualization familiar.	**P**	3.178
I find the visualization is poorly designed.	**N**	3.178
I find the visualization is missing important details.	**N**	3.112
I find that the visualization needs additional graphics to communicate effectively.	**N**	3.089
I cannot distinguish between colors.	**N**	3.067
I find the visual encoding is poor and violates good practice.	**N**	2.978
I find the visualization provides not enough context.	**N**	2.978
I find the visualization has overlapping information.	**N**	2.934

Continued on next page

Item	Context	Average Score
I find the variables in the visualization visible.	**P**	2.867
I feel confident about this visualization.	**P**	2.823
I feel informed by the visualization.	**P**	2.823
I feel engaged with the visualization.	**P**	2.689
I find the visualization uses too many colors.	**N**	2.689
I find this visualization visually pleasing.	**P**	2.667
I find the visualization dense.	**N**	2.645
I find the visualization displays several dimensions at once.	**N**	2.645
I feel the main idea is not clear within the first few seconds.	**N**	2.623
I feel curious about this visualization.	**P**	2.578
I find all data points are visible in the visualization.	**P**	2.534
I find that the structure is unclear.	**N**	2.512
I find this visualization attractive.	**P**	2.467
I find the visualization occludes information.	**N**	2.423
I find the color coding overly complex.	**N**	2.4
I feel confused by the visualization.	**N**	2.356
I find the visualization type requires higher cognitive effort.	**N**	2.267
I find the large number of data items makes the visualization too complex.	**N**	2.223
I find this visualization difficult to read.	**N**	2.156
I find the visualization requires effort to understand.	**N**	2.134
I find many visual elements distracting.	**N**	2.112
I find this visualization takes time to understand.	**N**	2.112
I find there is too much information in this visualization.	**N**	2.067
I find the visualization overloaded.	**N**	2.067
I find the visualization has a complex visual appearance.	**N**	2.067
I feel that this visualization causes cognitive overload.	**N**	2.045
I find the visualization has unnecessary complexity.	**N**	2.023
I find the visualization cluttered.	**N**	1.978
I feel the mental demand of this visualization is high.	**N**	1.978
I feel overwhelmed by the visualization.	**N**	1.912

A.4.3. CLMMs Implementation

The code used to analyze the collected data with CLMMs is provided below, and the analysis was conducted in RStudio.

```r
library(ordinal)
library(emmeans)

# load data
setwd("...") # path to working directory
df <- read.csv("testData.csv") # file with collected data

# prepare variables
df$rating <- factor(df$rating, levels = 1:5, ordered = TRUE)
df$item   <- factor(df$item, levels = paste0("q", 1:49)) # numeric order q1 to q49
df$userID <- factor(df$userID)

# fit the CLMM
model <- clmm(rating ~ item + (1 | userID), data = df)

# pairwise comparisons
results <- emmeans(model, pairwise ~ item, adjust = "bonferroni")
pairwise_results <- as.data.frame(results$contrasts)
write.csv(pairwise_results, "pairwise_results.csv", row.names = FALSE)

# get the reversed pairs
reversed <- pairwise_results
reversed$contrast <- sub("(\\w+) - (\\w+)", "\\2 - \\1", reversed$contrast)
reversed$estimate <- -reversed$estimate
reversed$z.ratio  <- -reversed$z.ratio

# combine original and reversed
both_directions <- rbind(pairwise_results, reversed)
both_directions <- both_directions[order(both_directions$contrast), ]
write.csv(both_directions,
          "pairwise_results_both_directions.csv",
          row.names = FALSE)
```

Listing A.1: CLMMs Implementation

A.4.4. Overview of the 49 items after the filtering steps

The following table presents the 49 items retained after the first two filtering steps, sorted from the highest to the lowest average estimate (see Table A.10). A separate index table (Table A.11) is also provided to match each item to its corresponding index.

Table A.10.: Average estimate and ranking counts for the 49 items

Item	Avg Estimate	Ranked Higher	Ranked Lower
q6	2.69	35	0
q5	2.55	33	0
q25	2.47	30	0
q36	2.26	22	0
q16	2.20	19	0
q14	2.18	21	0
q17	2.13	18	0
q38	1.99	14	0
q44	1.91	13	0
q31	1.86	13	0
q41	1.85	13	0
q47	1.83	11	0
q42	1.82	7	0
q4	1.80	11	0
q49	1.45	9	1
q39	1.37	8	1
q22	0.32	3	2
q30	0.30	3	2
q35	-0.64	1	2
q46	-0.98	1	3
q9	-1.93	0	3
q29	-1.94	0	3
q23	-2.02	0	7
q7	-2.039	0	6
q18	-2.05	0	4
q45	-2.07	0	5
q2	-2.07	0	7
q21	-2.08	0	7
q33	-2.12	0	3
q28	-2.13	0	3
q13	-2.15	0	5
q34	-2.15	0	3
q40	-2.15	0	3
q8	-2.17	0	3
q1	-2.17	0	14
q32	-2.18	0	8
q15	-2.18	0	15
q43	-2.20	0	11
q10	-2.20	0	13

Continued on next page

Item	Avg Estimate	Ranked Higher	Ranked Lower
q11	-2.20	0	7
q26	-2.21	0	16
q27	-2.21	0	11
q37	-2.21	0	13
q48	-2.22	0	16
q19	-2.30	0	16
q3	-2.31	0	18
q24	-2.31	0	16
q12	-2.34	0	20
q20	-2.37	0	18

Table A.11.: Items and their indexes

Index	Item
q1	I find many visual elements distracting.
q2	I find the visualization occludes information.
q3	I find the visualization cluttered.
q4	I find this visualization familiar.
q5	I find this visualization easy to understand.
q6	I find this visualization easy to interpret.
q7	I find that the structure is unclear.
q8	I feel the main idea is not clear within the first few seconds.
q9	I feel confident about this visualization.
q10	I find this visualization difficult to read.
q11	I feel confused by the visualization.
q12	I feel overwhelmed by the visualization.
q13	I feel curious about this visualization.
q14	I feel irritated by this visualization.
q15	I find there is too much information in this visualization.
q16	I find this visualization easy to use.
q17	I find this visualization intuitive.
q18	I find this visualization visually pleasing.
q19	I feel that this visualization causes cognitive overload.
q20	I feel the mental demand of this visualization is high.
q21	I find the color coding overly complex.
q22	I find the visual encoding is poor and violates good practice.
q23	I find this visualization attractive.
q24	I find the visualization has unnecessary complexity.
q25	I find the visualization easy to comprehend.
q26	I find the visualization overloaded.
q27	I find the large number of data items makes the visualization too complex.

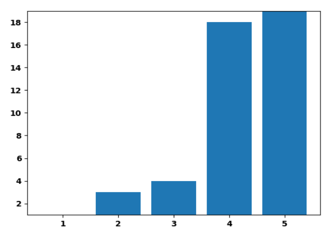
Continued on next page

Index	Item
q28	I find the visualization dense.
q29	I feel informed by the visualization.
q30	I find the visualization has overlapping information.
q31	I feel stressed by the visualization.
q32	I find the visualization type requires higher cognitive effort.
q33	I feel engaged with the visualization.
q34	I find the visualization uses too many colors.
q35	I find the visualization provides not enough context.
q36	I find the visualization well organized.
q37	I find this visualization takes time to understand.
q38	I find the points in the visualization easy to distinguish.
q39	I cannot distinguish between colors.
q40	I find the visualization displays several dimensions at once.
q41	I can detect patterns in the visualization.
q42	I find the visualization is missing important details.
q43	I find the visualization requires effort to understand.
q44	I find the visualization has poor pattern visibility.
q45	I find all data points are visible in the visualization.
q46	I find the variables in the visualization visible.
q47	I find the visualization is poorly designed.
q48	I find the visualization has a complex visual appearance.
q49	I find that the visualization needs additional graphics to communicate effectively.

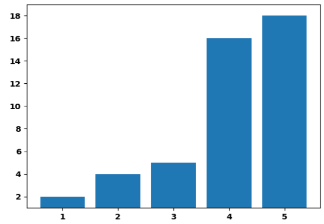
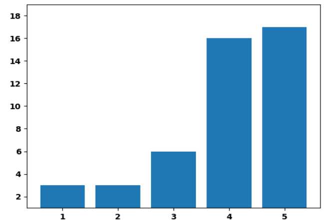
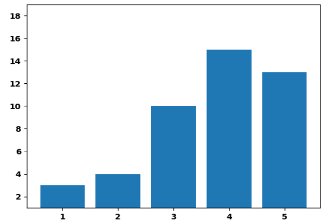
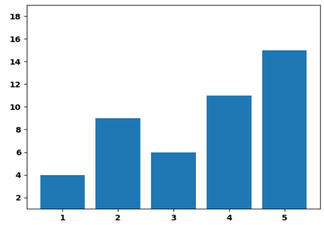
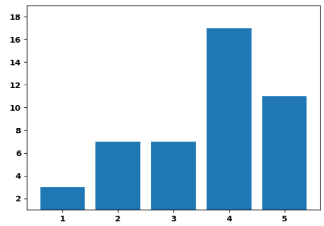
A.4.5. Rating distribution of the 18 items

This subsection presents the rating distributions for the final 18 items, providing an additional check on the results obtained in section 4.2.

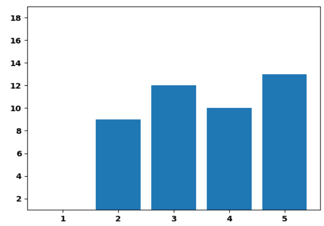
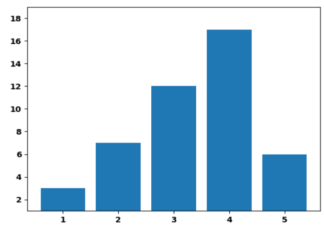
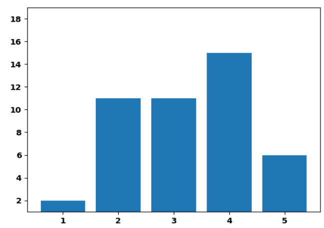
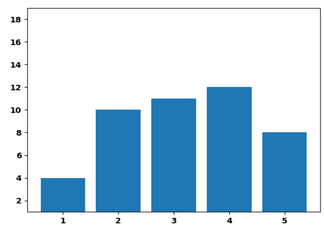
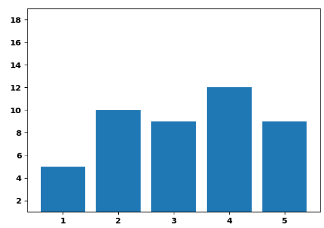
Table A.12.: Rating distribution of the 18 items and their average estimate value

Item	Avg Estimate	Score Dist.
q6	2.69	

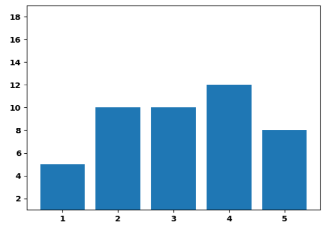
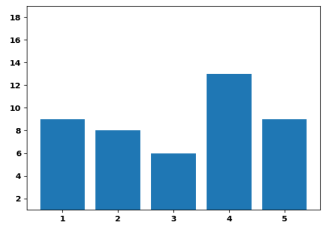
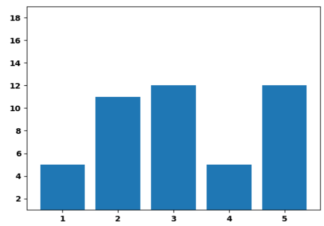
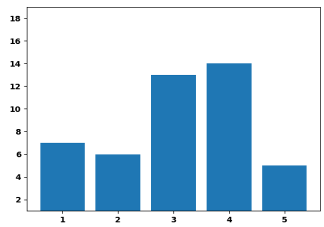
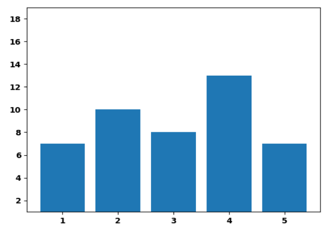
Continued on next page

Item	Avg Estimate	Score Dist.
q5	2.55	
q25	2.47	
q36	2.26	
q16	2.20	
q14	2.18	

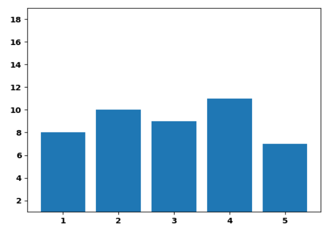
Continued on next page

Item	Avg Estimate	Score Dist.
q17	2.13	
q38	1.99	
q44	1.91	
q31	1.86	
q41	1.85	

Continued on next page

Item	Avg Estimate	Score Dist.
q47	1.83	
q42	1.82	
q4	1.81	
q49	1.45	
q39	1.37	

Continued on next page

Item	Avg Estimate	Score Dist.
q22	0.32	
q30	0.30