Towards a Cognitive Based Framework for Evaluating Information Visualisation Tools as tools for Learning

Kalliopi Kontiza*
University College London
Department of Information Studies

Angeliki Antoniou†
University of Peloponnese
Department of Informatics and Telecommunications

Antonis Bikakis‡
University College London
Department of Information Studies

Rob Miller§
University College London
Department of Information Studies

Abstract

Information Visualisation (InfoVis) is defined as the interactive visual representation of abstract data. We view the end-users' interaction with InfoVis tools as a learning experience which is made up of a set of highly demanding cognitive activities. These activities assist the end-user in making sense and gaining knowledge of the represented domain. There is a consensus in the InfoVis community about the importance of user evaluation studies in measuring the effectiveness of InfoVis tools. However, usability studies, which are most commonly used in this domain, are not appropriate for capturing the end-user cognitive processes that occur during the learning experience. To address this issue this paper discusses work in progress in developing a cognitive based framework for evaluating the effectiveness of InfoVis tools on the user learning process, by employing metrics for measuring the peoples ability to remember, understand, analyse, and derive insights from the visualised data, applied in the Cultural Heritage domain.

Index Terms: H.1.2 [Models and Principles]: User/Machine Systems—Human information processing; H.5.2 [User Interfaces]: evaluation/methodology—; H.5.3 [Group and Organization Interfaces]: Web-based interaction—;

1 Introduction

Cultural Heritage (CH) collections contain notable works of art and cultural heritage objects with associated knowledge and metadata. Users can make better sense of such data when it is organised and structured. Effective visual representations can enhance the audiences' physical and online experience as well as the understanding of a CH collection, allowing the users to browse the visualised data and understand its inherent meaning. This is, after all, one of the main aims of Information Visualisation (InfoVis). In order to develop effective InfoVis tools, two important aspects that should be considered are: the technology for creating the visual representations of data; and the perceptual and cognitive capabilities as well as the information-seeking needs and characteristics of their potential users.

In InfoVis, the focus is not on the raw data itself but on the information it conveys. It is important to note that information and data are not equivalent. As Spence indicates, when looking at the visually represented data, interesting information which is derived from the data, is revealed [26]. This exploration of the visually represented data on an InfoVis tool may result in gaining higher levels of knowledge at the semantic level. However, before such knowledge can be gained, users must interact with the interface at a syntactic level through a set of visual tasks [12]. A deep understanding of the role of interactive visualisation in human cognitive activities has been an important issue in the research of InfoVis. Liu discusses the classical perception-cognition-action model in Human Computer Interaction, one of the few theoretical frameworks which relates some of the aspects of human cognition in an integrative account where cognition is more an emergent property of interaction [16]. Is it possible to propose an equivalent formulation that can better inform our understanding of the relevant issues in InfoVis?

While deriving insight is one of the main purposes of InfoVis [1, 21], measuring how much people gain insight(s) from visualisation is still a challenge [18]. This research investigates the effectiveness of InfoVis tools on user's understanding of CH data in terms of learning. To this end, selected InfoVis tools will be evaluated to examine whether they assist the user in understanding and gaining insight into the visualised data, while the user performs specific tasks afforded by the InfoVis tool interface. The aim of this investigation is to demonstrate the correspondence -if any- between the levels of users learned knowledge and understanding and the user's interaction with CH data through InfoVis tools.

The main research question this research attempts to address is: *Do current InfoVis tools that use visual interfaces to display event-based CH structured data assist users in understanding and learning about the temporal, spatial, contextual and conceptual connections between CH artefacts, their creators and associated events?* Our end goal is to develop a theoretical framework for guiding the design and evaluation of InfoVis tools applied to the CH domain, which will allow us to capture and evaluate the cognitive operations of users perception while the user performs specific actions and learning tasks to manipulate the visualised information. Such tools will allow the users to explore and interact with cultural data with only minimal additional cognitive load.

The rest of this work in progress paper is structured as follows. Section 2 provides an overview of the multiple facets of the evaluation of InfoVis tools from the user perspective. Section 3 presents briefly the main components of the proposed cognitive - based framework. Section 4 describes the two small-scaled pilots we have conducted in preparation for the main experimental study of this research and, finally, section 5 discusses the initial results of this preliminary investigation.

2 Dimensions of User Engagement with InfoVis tools

A variety of InfoVis classification schemes have been proposed, striving for a better understanding of the InfoVis space [6]. The classification models that have been proposed so far reveal the multiple facets and perspectives of InfoVis, focusing either on the InfoVis process, its application, or its utility [2]. Any attempts to incorporate and integrate human cognition as one of the fundamental components or parts of the InfoVis, have initiated a modern approach to InfoVis research [22]. Psychology and related behavioral sciences have examined reasoning and other thought processes for decades, through classical scientific processes of reduction, laboratory testing, and scientific induction. One reason that much of this research has, as yet, been unused in the construction of interactive visualisations is the lack of a broad theory of human reasoning with sufficient scope

*e-mail: k.kontiza.12@ucl.ac.uk
†e-mail: angelant@uop.gr
‡e-mail: a.bikakis@ucl.ac.uk
§e-mail: r.s.miller@ucl.ac.uk

and predictive validity to impact the design and evaluation of those applications.

The User's Individual Differences Cognitive factors such as spatial and verbal ability as well as working memory capacity, vary substantially between individuals, and can affect reasoning in many different ways. Spatial and perceptual abilities in particular have been shown to affect how well users can perform several different tasks in a visualisation system. Perceptual abilities include basic visual proficiencies such as scanning speed and visual memory capacity. There has been no consensus on how to characterise individual differences in visual reasoning. One approach is to abstract individual differences in conceptualised measures such as visual memory, learning style and spatial cognitive ability using some standardised tests [8]. Given that both mental acts and physical acts can be analysed, we cannot only pin down individual differences in terms of visual thinking skills, but also specify the properties of the visualisation design. By contrasting how different users approach the same task differently, we can analyse whether the success or failure of task completion is attributable to the user or the properties of design [17]. Ziemkiewicz *et al.* discuss how cognitive abilities and personality factors affect the use of visualisation [28]. For example, Conati and McLaren [8] found that perceptual speed, which measures the speed at which a person can compare two figures, correlates with a user's accuracy at information retrieval tasks in one of two visualisation systems.

The Learning process There are different taxonomies-frameworks that have been proposed for measuring the user engagement in the context of the development of technologies for visualisation and learning; Naps' *et al.* [20] proposed framework has been mostly used for pedagogy purposes and the authors argue that no matter how well visualisation technology is designed, it has little educational value unless it engages learners in an active learning activity. Their engagement taxonomy defines six different forms of learner engagement with visualisation technology: 1) no viewing, 2) viewing, 3) responding, 4) changing, 5) constructing, and 6) presenting while Myller *et al.* [19] extend this taxonomy for software visualisation and collaborative learning.

The Level of Tasks In order to evaluate InfoVis tools, knowledge should be targeted at both the syntactic and semantic levels through a set of low and high level tasks. The syntactic level low-level tasks of the InfoVis user experience, evaluate whether the user understands the syntax of the visual language. The high-level tasks on the other hand correspond to the tasks used to evaluate the visualisation at the semantic level. They are tightly coupled with the visualised domain, unlike the low-level tasks. Evaluating the visualisation using the high -level tasks assists in determining whether the visualisation design corresponds to the requirements of the tool. Plaisant [23] stresses that empirical evaluation studies of InfoVis tools generally include only low-level tasks and that the rest of the tasks are rarely covered, while InfoVis tools are often evaluated in controlled studies that use benchmark tasks [7, 23]. Faisal *et al.* [12] point out that before knowledge can be gained at the semantic level, users must interact with the interface at a syntactic level through a set of visual tasks, such as identifying individual entities, categorising entities, identifying clusters, etc. from which domain independent visual tasks are identified such as identify, locate, rank, generalise, correlate, etc.

3 Building the Cognitive-based Framework

We are developing a cognitive-based framework which will enable us to evaluate the effectiveness of a set of selected InfoVis tools applied to digital CH collections, on learning. During the evaluation process we will attempt to measure the effectiveness of InfoVis tools on the user learning process while the users perform specific information seeking tasks. This cognitive-based evaluation framework

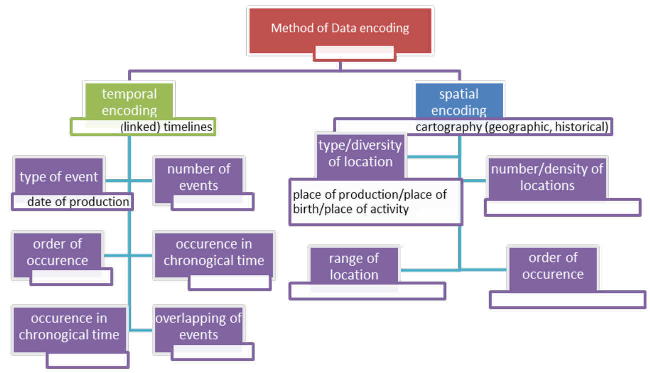

Figure 1: The temporal and spatial characteristics of the CH datasets.

builds on the combination of Shneiderman's Information Seeking Mantra guidelines [24] and Bloom's Revised Learning Taxonomy (RBT) [3, 15]. It identifies the dimensions of *learning* in a museum environment, the dimensions of the visually represented *data* according to their chosen method of data encoding (spatial and temporal) and the dimensions of the visual *tasks* that the InfoVis tools support.

Shneiderman argued that the type of InfoVis needs depends on both the type of data and the demands of the users. He bundled the type taxonomy that he had formerly introduced, with seven specific tasks that users could perform on the data [2]. This hierarchy of *visualisation tasks* in the form of a task taxonomy became known as "the visual information-seeking mantra" [10], by summarising the design philosophy of modern InfoVis systems to the statement "Overview first, zoom and filter, then details on-demand" [24]. Shneiderman proposed a type by task taxonomy (TTT) on InfoVis: the tasks of the TTT are task-domain information actions that users wish to perform and that an InfoVis system needs to support. When Craft *et al.* [9] reviewed the current at the time InfoVis literature, Shneiderman's 'Mantra' was one of the most frequently cited guidelines to InfoVis design. The results of their investigation advocate for the empirical validation of the 'Mantra' and research towards a holistic methodology of InfoVis design to address it [25].

Bloom's initial Taxonomy (BT) [4] is a multi-tiered hierarchical model of classifying thinking according to six cognitive levels of complexity. The model focuses on what processes are involved in cognitive learning and the importance of those processes in learning. Bloom's taxonomy has been applied to several domains such as the formal evaluation of the 'Townsville GeoKnowledge product' [11] where it was considered appropriate for developing geographical maps related tasks for the formal evaluation of the prototype cartographic product. It also inspired Mahyar's *et al.* [18] five-level taxonomy for user engagement, which describes the degree of user engagement with visualisations that increases as the user performs higher-level cognitive tasks from low to high engagement.

In our cognitive-based evaluation framework we address the (visual) task dimension by adopting and defining the 'Mantra' guidelines as follows: **A. Overview**: gain an overview of the entire dataset, **B. Details on demand**: select an item and get details without changing the initial representation/view of the dataset, **C. Zoom and Filter**: zoom in/out on the items of the dataset, **D. Relate**: view relationships among the items of the dataset, **E. History**: keep a history of actions performed on the dataset and **F. Extract**: allow extraction/reuse of the dataset. To address the learning dimension, we applied and defined the RBT categories in the CH domain as follows: **1.0 Remember**: to retrieve, recognise and recall discrete pieces of information like facts, a sequence of events or a step by step process, **2.0 Understand**: to classify pieces of information into groups, to compare and contrast pieces of information, **3.0 Apply**: to carry out or use pieces of information of a procedure in a given situation, **4.0 Analyse**: to differentiate between similar pieces of information, to place pieces of information in the order they occur, to break down prices of information into its component parts, **5.0 Evaluate**: to assess pieces of information and conclude to an outcome, to draw a conclusion, to make a hypothesis and **6.0 Create**: to combine different pieces of information and create new information. Finally to address the (visually represented) data dimension, we identified the temporal and spatial characteristics of the CH datasets in order to assess whether this data encoding structure can be supported by the identified InfoVis tools (Figure 1).

Using the method of the 'desired outcome statements', the proposed framework and its coding scheme (assigned letters to the dimension of the (visual) task supported by the InfoVis tool and numbers to the resulting levels of learning activities), allowed us to map the user evaluation tasks with the proposed evaluation framework in a systematic way. The method followed is demonstrated in the following section (subsection 'User Evaluation Tasks) with an example of one of the user evaluation tasks that we developed for one of the Pilot experiments.

4 Piloting the Main Experiment

The Pilots design and procedure So far, we have conducted two small scale pilots in preparation for the main experimental study of this research. Their aim was to explore and improve the process of data collection of the main study, to identify potential problems and to investigate viable solutions to improve the main data collection methodically. The participants of P1 were Master students of the Department of Information Studies at UCL and the participants of P2 were CH and IT professionals all recruited on a voluntary basis, the questionnaires were anonymised and demographic data was not collected. The information delivery method (InfoVis Tool(s) and the equivalent text descriptions) was used as an independent variable in both Pilots, while the personality trait was used as an additional independent variable in P2. The dependent variable in both Pilots was the number of questions answered correctly.

Both Pilots were implemented following a between subjects design approach: in P1 the participants of group A-TimelineTool were instructed to use the interactive browser-based Timeline tool to answer the task-based questionnaire. The participants of group B-TimeText were distributed a text presenting the events of the dataset in chronological order and were instructed to use these textual descriptions (booklet) to answer it. In P2, we employed the same design but experimented with two interactive browser-based InfoVis tools instead of one, which resulted in four user groups; A-TimelineTool, A-MapTool, B-TimeText and B-MapText. No further instructions were provided to the participants regarding the functionalities of the InfoVis tools.

In P1, all participants (n=10) were provided with the same version of the task-based questionnaire and the sessions were time-monitored (10 mins per session). In P2 (n=20), we enhanced the experimental design by adding the control of the visual/non-visual users; a pre-questionnaire was introduced to help us identify the user's cognitive style based on the MBTI standardised personality types [1]. Also, the task-based questions were given in a randomised sequence and were independent from each other in order to avoid the possibility that the participants would build their knowledge gradually to answer more complicated or cognitive demanding questions. The sessions were also time-monitored (15 mins per session, 5 for the pre-questionnaire and 10 for the interaction with the InfoVis tools or the textual descriptions). The task-based questions were mapped to our proposed evaluation framework using the method of the 'desired outcome statements' (see the following subsection 'User Evaluation Tasks'). While in P1 we used ten questions to cover the majority of the categories and the possible combinations of the evaluation framework coding scheme, in P2 we decided to limit the number of the task-based questions to eight and focus on the three lower levels of the user's cognitive process, namely Remember, Understanding and Applying.

[1] https://www.myersbriggs.org/my-mbti-personality-type/

In terms of content and data sample preparation, P1 focused on the temporal aspect of CH dataset and the associated events, while P2 addressed both the temporal and spatial aspects of the dataset. For P1, three representative artists were chosen, whose works are currently displayed at the National Gallery Collection[2]; *Charles Francois Daubigny, Camille Pissarro* and *Claude Monet*. From the related artist's webpages, information related to the event of a painting production was collected such as the date, the location, the title and a thumbnail of the artist's painting available to download on a CC-BY-NC-ND license. For P2, three representative paintings from the highlights of the UCL Art Museum [3] were chosen together with their respective narratives that provided contextual information available to reuse on a CC BY-NC-SA 3.0 licence; 'The Four Founders of University College, Lord Brougham, Jeremy Bentham, Thomas Campbell and Henry Crabb Robinson', 1922, *Tonks, Henry*, 'Under Milkwood', 1954, *Rego, Paula* and 'Female Figure Lying on Her Back', 1912, *Carrington, Dora*. Content from the related entries of Wikipedia articles was used for both Pilots on a complementary basis.

The User Evaluation Tasks. In order to describe the expected interaction while the user is actively engaged with the InfoVis tool and performs a task, each task was described as a set of 'desired outcome statements', also called learning objectives or outcomes in formal education. Learning outcomes can be defined as:"specific measurable achievements". The learning outcomes are learner-centred, measurable, achievable, and can therefore be assessed. When we create the 'desired outcome statements', we use active verbs to describe the desired results and the intended changes in the user level of understanding/knowledge. Initially we broke down the **'desired outcome statements'** in three parts: **part 'a'** stands for the user activity, **part'b'** stands for behaviour or knowledge, or learning activity being demonstrated and **part 'c'** for the degree of acceptable performance. As a second step, we encoded all the 'desired outcome statements' to the proposed evaluation framework dimensions as the following example demonstrates:

- **A.1.0.a** [*Can you list the name of the Artists, whose works are exhibited in this virtual Art collection?*]

The 'desired outcome statement' that describes the expected user interaction with the assigned InfoVis timeline tool or textual description is: *After viewing the overview of the timeline that is labeled with the names of the Artists it presents/After reading the chronology*, (part a) *the participant identifies and lists* (part b) *the names of the Artists, whose works are exhibited at the Virtual Art collection, using the tool/text* (part c). The code we assigned to this user task is: **A** (Overview) & **1.0** (Remember) & **a** (Factual Knowledge)

The InfoVis Tools. We have identified a list of InfoVis tools to test while piloting the main experiment. The review of the tools was based on: (i) Windhager *et al.*'s systematic review of InfoVis approaches to digital Cultural Heritage (CH) collections [27]; (ii) whether interactiveness is supported by the tool; and (iii) whether further configuration is needed. On a second level, the selected tools were further assessed based on the temporal and spatial characteristics of the CH datasets they could support. For P1, we selected TimeLineCurator, a browser-based, open-source authoring tool that automatically extracts event data from temporal references in unstructured text documents using natural language processing, and which enables the building and display of visually rich, interactive timelines, also known as TimelineJS [4]. TimelineJS is characterised

[2] https://www.nationalgallery.org.uk

[3] https://www.ucl.ac.uk/culture/ucl-art-museum

[4] http://timeline.knightlab.com/

as a peripheral linear chronological timeline that serves as an interactive navigation aid, using panning transitions between successive narrative points [13]. For P2, we selected the equivalent InfoVis map tool [5] from the same suite of storytelling tools to encode and present the spatial characteristics of the dataset. For the temporal data we experimented with the timelines of Timelinestoryteller [6]; a design space, that allowed us to choose from 14 design choices - characterised by three dimensions, i.e. representation, scale, and layout [5] - the timeline design that adequately communicates the temporal characteristics of our dataset.

5 DISCUSSION

We have performed a preliminary investigation running two small scale pilots in preparation for our main experimental study with which we want to test whether an InfoVis Tool (i) assists the user in understanding the temporal, spatial and contextual aspects of CH data and their connections; (ii) enhances the user's understanding in terms of learning while the user performs specific tasks afforded by the InfoVis Tools interface. The aim is to get valuable insight and to ascertain whether information obtained from the user by reading the text is perceived and encoded differently than the case where the same information is obtained through an InfoVis tool.

In more detail, during the two Pilots we tried out the cognitive-based evaluation framework we are developing. This process enabled us to modify any shortcomings in the framework before the main study experiment is carried out. The process of data collection through a questionnaire allowed us to test the method of collecting user input through a standardised questionnaire, decide on changes needed to achieve a clear structure and test its efficacy before applying it to the larger scale study. The users' interaction with this method of data collection provided useful feedback on deciding on the method of the evaluation approach, between an aloud protocol and a non-intrusive user evaluation. Although the Pilots were conducted in a controlled environment with a convenient sample, it assisted us to reflect on the actual conditions and constrains we may be faced up with while conducting the user evaluation in a museum environment with real visitors.

As a final step before conducting the main experimental study, the user evaluation tasks-questions will be examined through a reliability analysis, which will allow us to study the properties of the scales of measurement and the elements that constitute them. We will implement the Cronbach's Alpha internal consistency estimation method [14] based on the scores between each measure/item and the sum of all the others, to determine whether there is a good homogeneity among the user evaluation tasks-questions. For the main experimental study we will also follow standard sampling guidelines and the questionnaire will be enriched with some questions to collect basic demographic data and information about the participants' prior knowledge related to art collections.

REFERENCES

[1] Readings in information visualization: Using vision to think. Morgan Kaufmann, 1st ed., 1999.

[2] Information visualization. vol. 4950 of *Lecture Notes in Computer Science*. Springer, 2008.

[3] L. W. Anderson and Benjamin S Bloom. *A taxonomy for learning, teaching, and assessing: a revision of Bloom's Taxonomy of educational objectives / editors, Lorin W. Anderson ... [et al.].* Longman, abridged ed. ed., 2001.

[4] B. S. Bloom. *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain*. Addison-Wesley Longman Ltd, 2nd edition edition ed., 1956.

[5] https://storymap.knightlab.com/

[6] https://timelinestoryteller.com/app/

[5] M. Brehmer, B. Lee, B. Bach, N. H. Riche, and T. Munzner. Timelines revisited: A design space and considerations for expressive storytelling. 23(9):2151–2164, 2017.

[6] S. K. Card. Visualizing retrieved information: A survey. 16(2):63–67, 1996.

[7] C. Chen and M. P. Czerwinski. Empirical evaluation of information visualizations: an introduction. 53(5):631–635, 2000.

[8] C. Conati and H. Maclaren. Exploring the role of individual differences in information visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '08, pp. 199–206. ACM, 2008.

[9] B. Craft and P. Cairns. Beyond guidelines: What can we learn from the visual information seeking mantra? pp. 110–118. IEEE, 2005.

[10] A. Dadzie and M. Rowe. Approaches to visualising linked data: A survey. 2(2):89–124, 2011.

[11] M. Dodge, M. McDerby, and M. Turner, eds. *Geographic visualization: concepts, tools and applications*. Wiley, 2008.

[12] S. Faisal, P. Cairns, and A. Blandford. Challenges of evaluating the information visualisation experience. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 2*, pp. 167–170. British Computer Society, 2007.

[13] J. Fulda, M. Brehmel, and T. Munzner. TimeLineCurator: Interactive authoring of visual timelines from unstructured text. 22(1):300–309, 2016.

[14] S. B. Green, Y. Yang, M. Alt, S. Brinkley, S. Gray, T. Hogan, and N. Cowan. Use of internal consistency coefficients for estimating reliability of experimental task scores. 23(3):750–763, 2016.

[15] D. R. Krathwohl. A revision of bloom's taxonomy: An overview. 41(4):212–218, 2002.

[16] Z. Liu, N. J. Nersessian, and J. T. Stasko. Distributed cognition as a theoretical framework for information visualization. 14(6):1173–1180, 2008.

[17] Z. Liu and J. T. Stasko. Mental models, visual reasoning and interaction in information visualization: A top-down perspective. 16(6):999–1008, 2010.

[18] N. Mahyar, S.-H. Kim, and B. KWON. Towards a taxonomy for evaluating user engagement in information visualization. In *Workshop on Personal Visualization: Exploring Everyday Life*, vol. 3, 2015.

[19] N. Myller, R. Bednarik, E. Sutinen, and M. Ben-Ari. Extending the engagement taxonomy: Software visualization and collaborative learning. 9(1):1–27, 2009.

[20] T. L. Naps, G. R\s sling, V. Almstrum, W. Dann, R. Fleischer, C. Hundhausen, A. Korhonen, L. Malmi, M. McNally, S. Rodger, and others. Exploring the role of visualization and engagement in computer science education. In *ACM Sigcse Bulletin*, vol. 35, pp. 131–152. ACM, 2002.

[21] C. North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, 26(3):6–9, 2006.

[22] R. E. Patterson, L. M. Blaha, G. G. Grinstein, K. K. Liggett, D. E. Kaveney, K. C. Sheldon, P. R. Havig, and J. A. Moore. A human cognition framework for information visualization. 42:42–58, 2014.

[23] C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced visual interfaces*, pp. 109–116. ACM, 2004.

[24] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pp. 336–343. IEEE, 1996.

[25] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pp. 1–7. ACM, 2006.

[26] R. Spence. *Information visualization: design for interaction*. Pearson Prentice Hall, 2. ed ed., 2007.

[27] F. Windhager, P. Federico, E. Mayr, G. Schreder, and M. Smuc. A review of information visualization approaches and interfaces to digital cultural heritage collections. In *Proceedings of the 9th Forum Media Technology 2016*, 2016.

[28] C. Ziemkiewicz, A. Ottley, R. Crouser, K. Chauncey, S. Su, and R. Chang. Understanding visualization by understanding individual users. 32(6):88–94, 2012.