What Design Methods do DataVis Practitioners Know and Use?

Paul Parsons
Purdue University
parsonsp@purdue.edu

Ya-Hsin Hung
Purdue University
hung17@purdue.edu

Ali Baigelenov
Purdue University
abaigele@purdue.edu

Connor Schrank
Purdue University
cschran@purdue.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.
© 2020 Copyright is held by the owner/author(s).
ACM ISBN 978-1-4503-6819-3/20/04 ...$15.00.
http://dx.doi.org/10.1145/3334480.3383048

Abstract
Data visualization as a profession has been growing rapidly in recent years. Although some initiatives are in place to increase engagement between the academic and practitioner communities, we currently do not have a good understanding of how practitioners do their design work, including what methods, approaches, and principles they know and use in their everyday practice. We present a subset of results of a survey in which 87 DataVis practitioners identified their familiarity with popular design methods and the frequency with which they use them in their own work. We also discuss follow-up work to develop a deeper understanding of practitioners' perspectives on design methods.

Author Keywords
Data visualization; Survey; Design practice.

CCS Concepts
•**Human-centered computing** → **Visualization;** *HCI theory, concepts and models;* **Visualization theory, concepts and paradigms;**

Introduction
Data visualization as a profession has seen significant growth in recent years [14]. Within the past year alone, the practitioner-led Data Visualization Society [2] was created and quickly reached 10,000 members [15]. Two of the

largest software platforms for data visualization, Tableau and Looker, were acquired in multi-billion dollar acquisitions by large tech companies, signaling the importance of data visualization within the industry [11]. Along with this growth, attempts are being made to more closely connect the academic and practitioner communities, through various workshops and events (e.g., [1, 4]).

In response to this growth, it is important to develop an understanding of how practitioners are engaging in their design work, both for conducting research studies and training future DataVis designers. To date, however, little research has investigated the everyday practices of DataVis practitioners. As part of a broader research effort to better understand how DataVis practitioners engage with formal design knowledge (e.g., theories, frameworks, principles, methods), we conducted a survey to answer two specific research questions: (1) how *familiar* are practitioners with forms of design knowledge that are common in the academic literature? and (2) how *frequently* do practitioners use these forms of design knowledge?

Although our survey investigated design knowledge quite broadly, including theories, models, frameworks, taxonomies, methods, and principles, in this paper we present only a subset of results, focusing on methods and principles that practitioners know and use. The other parts of the survey require more analysis and follow-up with respondents, and will be presented in a larger report in the future. The results of this survey provide insight into the current state of DataVis practice, highlighting methods and principles that are important for practitioners. This research can help both the academic and practitioner communities better align expectations about the transfer of knowledge between them, and open a space for continued dialogue about how these communities can benefit one another.

Understanding Design Practice

In many application-oriented fields, such as InfoVis and HCI, there is a desire for academic research to influence design practice. Although there has sometimes been an assumption that practitioners will simply apply research findings, several studies indicate otherwise [7, 16, 17]. When practitioners do make use of research findings, typically only certain forms of design knowledge are used [17], and they are often appropriated in ways that were not articulated by the researchers in their original formulation [8].

Within the HCI literature, there is a growing body of scholarship that examines design from a practice-driven perspective (e.g., [7, 8, 16, 17]). Although this style of research has not yet become prevalent in the visualization community, a number of initiatives have been undertaken by visualization researchers to make research more accessible to practitioners. These include the long-running *Data Stories* podcast [5], the recently created blog, *Multiple Views: Visualization Research Explained* [3], and various venues where visualization research is presented and discussed (e.g., [1, 4]). While such initiatives fill an important need and are certainly valuable, other approaches are needed that study DataVis design practice and present it back to the research community. The work presented here is an initial attempt at doing so, by investigating practitioners' knowledge and use of design methods in their everyday practice.

Method

Our aim in creating the survey was to capture common forms of design knowledge, while keeping it short enough to increase the likelihood of complete responses. Our guiding heuristic was that respondents should be able to complete it in under 20 minutes. Thus, there was a trade-off between being exhaustive in what was included, and keeping the survey at a reasonable length. In the end, we identified 55 specific forms of design knowledge that we categorized as follows: (1) methods and approaches; (2) concepts, principles, laws, and guidelines; (3) theories; and (4) taxonomies and models. In this paper, we report the results of the first 2 categories, and simply refer to both as design methods for ease of reading. Here we define design methods broadly, encompassing formal and informal approaches, principles, and concepts, covering any conceptual or practical tool that designers use in their everyday practice. We briefly discuss the method for the entire survey below, then present only the specified subset of results.

We asked about *familiarity* and *frequency of use* of each of the forms of design knowledge. Responses were given using a 5-point likert-style scale (for familiarity: not at all-1, slightly-2, somewhat-3, moderately-4, extremely-5; for frequency: never-1, rarely-2, sometimes-3, often-4, always-5). We also solicited open-ended responses for each section, although we do not report on them here. Answers to closed-ended questions were required, and answers to open-ended questions were optional. After answering all questions, respondents could input their name and email to be entered into a draw for a $45 gift card with a 1/10 chance of winning. Finally, respondents could also input their name and email to be considered for a follow-up interview [1].

Identifying Relevant Survey Items

To identify relevant survey items, we conducted a literature search using an exhaustive combination of keywords in the form of "information visualization" or "data visualization" and combinations of "theory", "model", "taxonomy", "principle", "law", "concept", "method", or "approach". We also searched for relevant terms such as "design process", "design theory", and "design model". To ensure that we considered a broad range of items, we also consulted popular books on design methods and principles (e.g., [9, 10, 13]).

We initially identified 99 candidate items to include in the survey (see supplemental materials for details). Our guiding heuristic was for 20 minutes being needed to complete the survey. Because demographic questions and short open-ended questions were also being asked, we aimed for 15 minutes for the closed-ended questions that involved these items. With practical experience suggesting 10 to 20 seconds to answer each question, we aimed for roughly 50 items to be included in the final version.

For each item, we searched the literature to determine if it was commonly used in an visualization context. We identified the number of relevant papers in which each item appeared. Following this procedure, we removed items from the list that did not appear often. For example, Hick's Law was in our initial list of candidate items, yet when we performed a literature search using the combination of keywords above, we found fewer than 50 papers in which Hick's Law was discussed in the context of visualization. As another example, searching for cognitive load yielded hundreds of relevant papers, thus it was kept in the final list of items.

It is difficult to be objectively precise with such a method, as each use of a search term must be judged with respect to its relevance. For instance, at the time this work was carried out, a search on Google Scholar for “card sorting” AND “information visualization” yielded 352 papers, yet not all were relevant. For instance, some papers simply mention these two phrases, yet they are not central ideas in the paper. Thus we had to filter through the results to determine their relevance to the visualization literature. To do so, we determined (1) if the papers were published in visualization or

[1]Survey materials and results can be found at https://osf.io/9ubrd/?view_only=542d6fcb93f84a22bf52ae80d98dcd0a

Gender	
Female	33
Male	52
Other	1
Not specified	1

Age	
18 - 25 years	3
26 - 35 years	34
36 - 45 years	33
46 - 55 years	11
> 55 years	6

Degree	
High school	3
Bachelor	26
Masters	39
Doctorate	17
Other	2

Experience	
0 - 2 years	9
2 - 4 years	14
5 - 7 years	20
8 - 10 years	15
> 10 years	29

Table 1: Demographic information of respondents.

HCI venues (e.g., VIS, CHI, EuroVis, IEEE TVCG), and (2) whether the target item was discussed in the context of visualization design or was simply mentioned in an unrelated manner. Our assessment of these two criteria was based on our judgment as experienced researchers in the field, and is necessarily somewhat subjective. However, while this method has limitations, we are confident that it allowed us to narrow down the most relevant items and exclude the less common ones.

In the end, we selected 55 of the 99 items to keep. To provide some organization to the items, we iteratively sorted them into four categories: methods and approaches (16 items); principles, methods, laws, and guidelines (16 items); theories (12 items); and taxonomies and models (11 items). It is important to note that some of the items could fall into more than one category—e.g., mental models and cognitive load could be considered as theories rather than concepts or principles—so we made a judgment about the most appropriate fit for DataVis designers. For instance, cognitive load is commonly viewed as a theory in educational psychology, yet for DataVis practitioners it is more likely to be used as a design guideline (e.g., "avoid unnecessary cognitive load"). Additionally, to ensure that respondents could effectively identify and not misunderstand the items in the survey, we implemented tool-tips that displayed a short description or definition of each item. An external link to a website or paper was also provided so participants could find more information if desired.

Pilot Testing

After categorizing the items, we composed an initial survey draft and conducted a pilot study with 10 participants. Participants consisted of a mix of professors and students, all having experience in HCI and design, and some having expertise in DataVis or survey methodology. After multiple iterations, we addressed the relevant issues stated by the pilot testers and finalized the survey.

Survey Distribution

The survey was hosted on Qualtrics. We distributed the survey through multiple platforms, including: social media (Twitter, Reddit, LinkedIn), the DataVis Society's Slack workspace, the InfoVis email list, and our personal networks. To mitigate sampling bias, we also searched widely online for practicing professionals and agencies to contact, ultimately contacting more than 200 individuals and more than 30 agencies directly.

Results

After discarding incomplete responses, we had 87 participants. Table 1 shows their demographic breakdown. 66 self-reported to have acquired their data visualization expertise by being self-taught, 31 learned from university or college, 20 from an online course (e.g., MOOC), and 6 from a bootcamp (in person). The training demographic question allowed participants to select multiple options, resulting in an array of data that included multiple ways of being trained (e.g., self-taught and online course). 49 participants were located in the US, with the rest being distributed across Europe, Asia, and South America.

Likert-Style Responses

For an initial analysis of the subset of results we are focusing on here, we calculated the average values of both familiarity and frequency for each item. These results are displayed in Figure 1, where familiarity and frequency are plotted on scatterplots. Although this is only one view of the data, it still yields valuable insights. Future work will present a more robust analysis of the data, including different visualizations of the results.

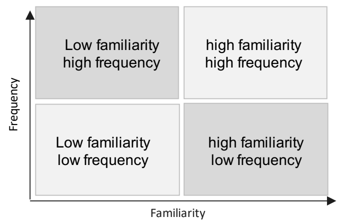

Figure 2: Interpretation of the four regions in each scatterplot.

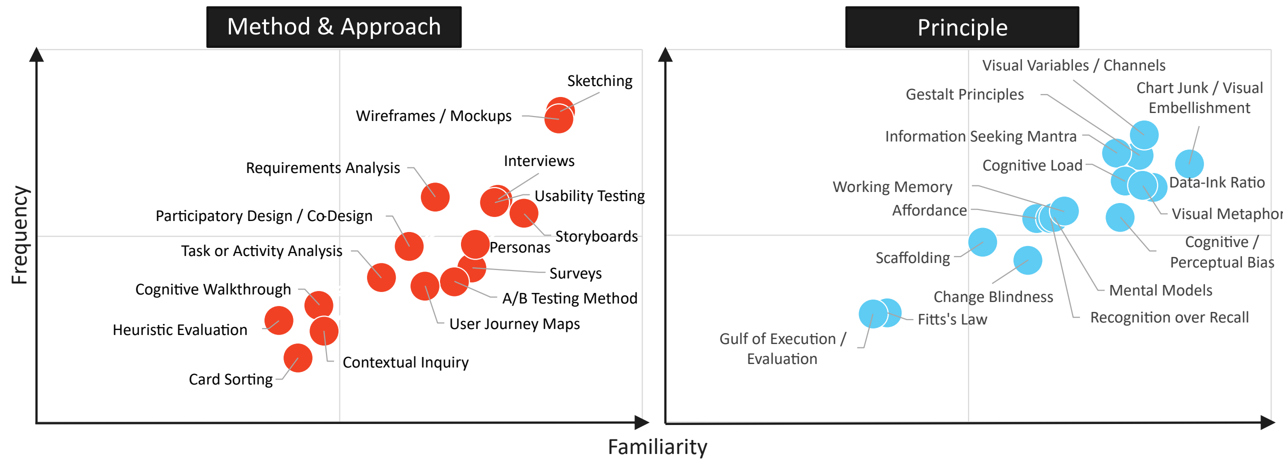

Figure 1: Familiarity vs. frequency of use for design methods and principles

As each item in the survey had 5 possible responses, we scored the responses from 0 to 4. In doing so, each scatterplot can be divided into four dimensions by drawing lines at the mid-points (familiarity=2 and frequency=2 on the x and y axes, respectively). The interpretation of these four dimensions is described in Figure 2. For example, if an item is located in the upper-right quadrant, it suggests that it is both well-known and frequently used; if an item is located in the lower-left quadrant, it suggests that it is not well-known and not used much. Of course, the upper-left quadrant is not likely to be occupied, yet it still forms a logical possibility in which items could be placed.

Below we provide descriptions of each category using a heuristic of a 50% cutoff for the top two options of our scale (moderately/extremely familiar or often/always used), and also describing the least familiar and least used items.

Methods and approaches. At least 50% of respondents report being moderately or extremely *familiar* with the following: requirements analysis, interviews, surveys, usability testing, task or activity analysis, participatory design/co-design, sketching, personas, A/B testing, wireframes/mockups, storyboards, and user journey maps. The most unfamiliar methods are heuristic evaluation, with 28 respondents reporting no familiarity. The second and third are cognitive walkthrough and card sorting, with 22 and 21 reporting no familiarity, respectively. At least 50% of respondents report often or always *using* the following: requirements analysis, sketching, wireframes/mockups, and storyboards. At least 50% of respondents report never using the following: heuristic evaluation and card sorting.

Principles and concepts. At least 50% of respondents report being moderately or extremely *familiar* with the following: cognitive/perceptual bias, visual metaphor, cognitive load, change blindness, visual variables/channels, data-ink ratio, chartjunk/visual embellishment, working memory, mental models, information seeking mantra, recognition over recall, Gestalt principles, and affordance. The most unfamiliar principle is Fitts' Law, with 39 respondents reporting no familiarity. At least 50% of respondents report *using* the following often or always: visual metaphor, cognitive load, visual variables/channels, data-ink ratio, chart junk/visual embellishment, working memory, information seeking mantra, recognition over recall, Gestalt principles, and affordance. At least 50% of respondents report never using Fitts' Law and the gulfs of execution and evaluation.

Discussion and Future Work

Many methods and principles were reported as being highly familiar and frequently used. Sketching and wireframing / mockups were significantly more familiar and used than other methods. This is not surprising, as such methods have been shown to lie at the heart of most design work [12]. The most familiar principle, on average, was chart junk. This result is interesting, although not entirely surprising, as chart junk has been a controversial issue for some years now, and considerable debate has taken place on twitter, where practitioners engage in much of their professional discourse. Other highly familiar and frequently used principles, including chart junk, visual variables, Gestalt principles, and data-ink ratio, have all been popularized in practitioner-oriented resources for many years.

Multiple methods and concepts were reported as being familiar yet not often used. This includes very popular user-centered design methods such as A/B testing, user journey maps, surveys, co-design, and personas. The reason why they are not frequently used is currently unknown, although we plan to clarify in follow-up interviews with respondents. It is possible that some methods are too time- or resource-intensive to use frequently, such as A/B testing and surveys. The familiarity yet limited use of user journey maps may indicate that DataVis practitioners are not thinking holistically about a user's journey with a visualization in the way that UX designers often do. The familiarity yet infrequent use of change blindness may suggest that practitioners do not know how to apply such abstract knowledge in their design work—despite a large body of literature on the relevance of change blindness to visualization. Perhaps more translational work needs to articulate how such concepts can be applied in everyday design situations.

Arguably all of the methods and principles that were reported as low familiarity and low frequency of use are very common in the HCI literature and are well known in the interaction design community. For instance, card sorting, heuristic evaluation, cognitive walkthroughs, contextual inquiry, Fitts' Law, and Norman's gulfs of execution and evaluation have been very well established in the user-centered design literature for multiple decades, and they commonly appear in practitioner-oriented books and other resources. In fact, multiple surveys indicate that heuristic evaluation is consistently one of the most popular and commonly used methods among user-centered designers [18]. This result suggests that our respondents are largely not trained in or familiar with human-centered design practice, despite the fact that heuristic evaluation has been discussed in the visualization literature for some time now [6, 19].

In follow-up work, we plan to interview a subset of respondents to provide more clarity on reasons behind their responses. We will also perform further analyses of the data presented here in addition to the larger set of survey items.

REFERENCES

[1] 2018. Information+ Conference. (2018). Retrieved January 5, 2020 from http://informationplusconference.com/2018

[2] 2019. Data Visualization Society. (2019). Retrieved January 5, 2020 from www.datavisualizationsociety.com

[3] 2019. Multiple Views: Visualization Research Explained. (2019). Retrieved January 5, 2020 from https://medium.com/multiple-views-visualization-research-explained

[4] 2019. VisInPractice @ VIS 2019. (2019). Retrieved January 5, 2020 from https://visinpractice.github.io

[5] Enrico Bertini and Mortiz Stefaner. 2020. Data Stories. (2020). Retrieved January 5, 2020 from https://datastori.es/

[6] Sheelagh Carpendale. 2008. Evaluating information visualizations. In *Information Visualization*, Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and North Chris (Eds.). Springer, 19–45.

[7] Elizabeth Goodman, Erik Stolterman, and Ron Wakkary. 2011. Understanding interaction design practices. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, Vancouver, BC, Canada, 1061. DOI:http://dx.doi.org/10.1145/1978942.1979100

[8] Colin M. Gray, Erik Stolterman, and Martin A. Siegel. 2014. Reprioritizing the relationship between HCI research and practice: bubble-up and trickle-down effects. In *Proceedings of the 2014 conference on Designing interactive systems - DIS '14*. ACM Press, Vancouver, BC, Canada, 725–734. DOI: http://dx.doi.org/10.1145/2598510.2598595

[9] Bruce Hanington and Bella Martin. *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions* (58480th edition ed.). Rockport Publishers.

[10] Jeff Johnson. *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Rules*. Morgan Kaufmann. DOI:http://dx.doi.org/10.1016/B978-0-12-407914-4.00004-X

[11] Dan Kopf. 2019. The multibillion-dollar sales of Tableau and Looker are a coming of age for data visualization. (Jun 2019). https://qz.com/1640415/acquisitions-of-tableau-and-looker-show-coming-of-age-for-dataviz/.

[12] Bryan Lawson. 2012. *What designers know*. Routledge.

[13] William Lidwell, Kritina Holden, and Jill Butler. *Universal Principles of Design*. Rockport Publishers.

[14] Elijah Meeks. 2019a. 2019 Annual Data Visualization Survey Results. (Aug 2019). https://medium.com/nightingale/2019-annual-data-visualization-survey-results-334d3523073f.

[15] Elijah Meeks. 2019b. 2019 Was the Year Data Visualization Hit the Mainstream. (Dec 2019). https://medium.com/nightingale/2019-was-the-year-data-visualization-hit-the-mainstream-d97685856ec.

[16] Yvonne Rogers. 2004. New theoretical approaches for human-computer interaction. *Annual review of information science and technology* 38, 1 (2004), 87–143.

[17] Erik Stolterman. 2008. The Nature of Design Practice and Implications for Interaction Design Research. *International Journal of Design* 2, 1 (2008), 55–65. DOI: http://dx.doi.org/10.1016/j.phymed.2007.09.005

[18] Karel Vredenburg, Ji-Ye Mao, Paul W Smith, and Tom Carey. 2002. A survey of user-centered design practice. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 471–478.

[19] Torre Zuk, Lothar Schlesier, Petra Neumann, Mark S Hancock, and Sheelagh Carpendale. 2006. Heuristics for information visualization evaluation. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*. ACM, 1–6.