Empowering Sensemaking in the Web's Emerging Visualization Ecosystem

Evan M. Peck
Bucknell University
Lewisburg, PA USA
evan.peck@bucknell.edu

Lane Harrison
Worcester Polytechnic Institute
Worcester, MA USA
ltharrision@wpi.edu

CHI'18 Extended Abstracts, April 21–26, 2018, Montréal, Canada.
ACM ISBN 978-1-4503-4656-6/17/05
http://dx.doi.org/10.1145/3027063.3053142

Abstract
The field of visualization aims to empower people to explore and reason about data through visual encodings and interaction. However, visualization techniques and tools have now moved beyond applications in scientific and analytical settings to form a diverse and vibrant visualization ecosystem on the web. We discuss challenges that this emerging visualization ecosystem raises for supporting sensemaking with data, as well as the new opportunities it brings for visualization design and evaluation.

Author Keywords
information visualization; search; web design; provenance

ACM Classification Keywords
H.5.3 [Information interfaces and presentation (e.g., HCI)]: Group and Organization Interfaces

Introduction
The landscape of visualization has changed. In past decades, innovations in data visualization often targeted scientific and analytical applications. For example, scientists working with climate simulations benefited from advancements in visualization toolkits and interaction techniques. More recently, however, a visualization theory and practice has also shifted to the web, creating a vibrant ecosystem of visualizations that are available to the greater population.

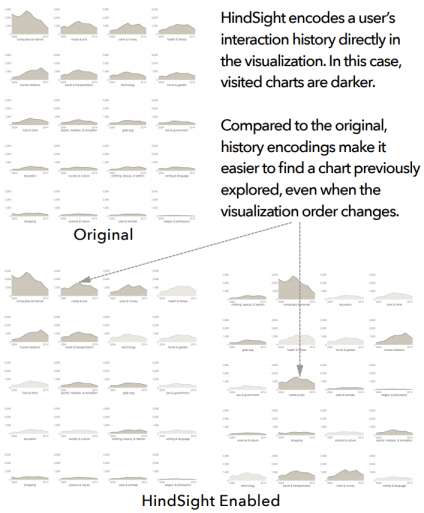

Figure 1: Much like blue hyperlinks being colored purple after a user visits them, HindSight directly encodes personal interaction history in the visualization a user is exploring [1].

Advances in visualization toolkits are contributing to this shift in the visualization landscape. Toolkits like d3, protovis, and Processing, for example, have made it possible to add visualizations to any webpage or web-based application. Similarly, platforms like Tableau Public have lowered the technical barriers to creating web-based visualizations. As these tools continue to be shaped by both research and developer communities, the trend of widely disseminated, novel web interfaces to data will persist.

People are another driver in this emerging visualization ecosystem. Organizations such as *The New York Times* and *ProPublica* have found value in visualization as a means for disseminating data-driven journalism. Hobbyists have also begun to develop web-based visualizations to engage with their communities of common interest. The popular $/r/dataisbeautiful$ page of Reddit (with over 12 million subscribers) includes users posting original visualizations of their own on topics spanning dog breed choices, coffee tasting, and personal financial budgets, and more. Given that The New York Times has over 2.5 million digital subscribers itself, the diversity and number of people engaging with interactive visualizations is larger than ever before.

We characterize this shift in the visualization landscape as the emergence of a *visualization ecosystem* on the web. To borrow from the biological definition of an ecosystem, we observe that researchers, creators, and consumers of visualizations are increasingly behaving as members of an ecosystem, using the web as an environment and various feedback mechanisms such as social media as means for communication. These observations share parallels with longstanding models such as Pirolli and Card's work on information foraging [4], which can be used to further characterize peoples' behavior in this visualization ecosystem.

A growing visualization ecosystem brings with it challenges to support sensemaking. For example, the pace at which visualizations are being developed and disseminated leaves little time for evaluation and reflection on the *effectiveness* of the design choices made. In this paper, we discuss a series of work in which we have addressed sensemaking challenges in the web's visualization ecosystem both *within* individual visualizations, and *between* visualizations.

New challenges also yield new opportunities for visualization research. We believe that describing the visualization landscape as an ecosystem on the web, coupled with demonstrations of new opportunities to support sensemaking within and between visualizations, will lead us to new ways of thinking that could inform future research initiatives in visualization.

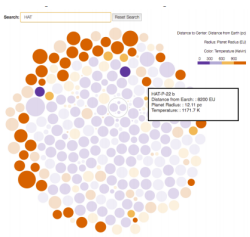

Figure 2: Adding text-based search to visualizations allows users to explore a more diverse range of sensemaking goals [2].

Sensemaking with Visualizations on the Web

One reality of this emerging visualization ecosystem is that sensemaking in visualizations is currently not well supported. Consider, for example, a scatterplot visualization of movie budgets and earnings. Tools like d3 make it easy for the visualization designer to add interactivity, such as allowing users to mouseover points to investigate detailed information about each movie. *Within* this visualization, however, the goals of the user may not align with the affordances of the visualization design- if the user wants to search for a movie they saw recently, they must manually hover over datapoints, one by one, until they find it.

Sensemaking is also seldom supported when moving *between* visualizations- if a user wishes to find other data and visualizations related to this movies visualization, how would they begin? If they wanted to revisit the visualization months later, could the find it? In several recent studies, we have taken steps to address challenges like these, and to explore what it means to support sensemaking in the visualization ecosystem *within* and *between* visualizations.

Sensemaking within Visualizations

When people encounter a visualization of interest, one could argue that they engage in a form of sensemaking *within* the scope of the visualization. Through several studies, we have observed that visualizations on the web often do not include basic functionality for sensemaking that is ubiquitous in other interfaces, such as web browsers. In several experiments, we have demonstrated that adding such functionality has a significant impact on how people engage with visualizations on the web.

Interaction history, for example, is ubiquitous in web contexts and often appears in other areas of computing. A page of web search results, for example, contains an array of blue links that change to a shade of purple after a user has visited them. Many exploration-based video games include a mini-map that shows areas of the world that a player has explored. To bridge this gap in visualization, we encoded users' interaction history in the visualization itself, a technique we call HindSight [1]. Controlled crowdsourced experiments comparing HindSight to original visualizations on the web indicated that users often interacted with significantly more individual data items in the visualization.

Text-based search is also a ubiquitous feature in computing contexts. One can hardly imagine having a word processor or web browser that did not include functionality to search within a page for keywords of interest. Yet visualizations rarely include this functionality, perhaps because the value of search functionality in visualization is unclear. To explore this gap, we adopted several popular visualizations from the web with search functionality [2]. Similar to HindSight, adding search had a significant impact on user exploration behavior- many people made use of the search functionality, and engaged with data-items longer if they had searched for them.

These studies demonstrate that the bar is low for significantly improving user experience with visualizations on the web. More broadly, future visualization research should focus on developing a design space of low-barrier interaction strategies that benefit people without expertise or training.

Sensemaking between Visualizations

In the web's visualization ecosystem, sensemaking occurs not only *within* a data visualization, but also *between* visualizations. In our movies scatterplot example, we highlighted unsupported sensemaking tasks including recall, or finding a visualization after visiting it, and lateral movement, where a user seeks other visualizations related to the one they are currently viewing.

To address affordances related to recall in the visualization ecosystem, we built SightLine, a system that enables personal provenance of visualizations visited on the web [5]. SightLine consists of a browser extension that first detects whether an interactive visualization is on a page, and sends record of a visualization visit to a backend server that generates additional metadata about the visualization and webpage. SightLine provides a web-accessible interface for users to explore their own history of visited visualizations, including dates, thumbnails, and search queries on visualization metadata. In addition to personal provenance, SightLine provides a social signals page, which shows the most-visited visualizations across active users.

The visualization ecosystem on the web also raises questions of barriers to accessibility, namely, can people find visualizations that already exist on the web? To explore this question, we conducted an in-lab study with 15 undergraduate students in STEM majors in which we asked them to locate visualizations on the web related to topics they chose. The results of this study indicate that undergraduate students, who are relatively high-literacy in comparison to the general population, struggled to articulate search queries that resulted in data and visualizations related to their topic of interest. These obstacles potentially undermine peoples' access to valuable data-driven perspectives on topics they are interested in.

Similar to our studies targeting sensemaking barriers *within* visualizations, we find that sensemaking *between* visualizations is even more challenging and unsupported by our current repertoire of techniques and systems.

Challenges in the Visualization Ecosystem

The emerging visualization ecosystem brings with it pressing challenges as well as new opportunities for exploring

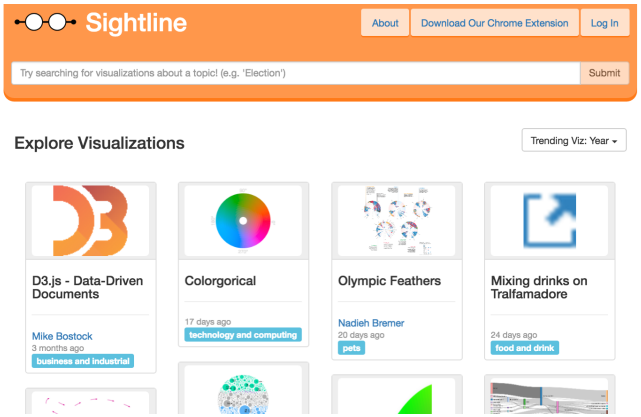

Figure 3: SightLine is an automatic data collection and discovery service that leverages the web's rich visualization ecosystem.

how people engage with and make sense of data through visualizations. While the space of possibilities in this new landscape is wide, we lay out here several themes as possible next steps in supporting sensemaking in the visualization ecosystem.

Refining Task/Goal Taxonomies

Task taxonomies are a cornerstone of visualization design and thinking. By discretizing aspects visualization use, and providing corresponding language that capture the diverse actions that users take while engaging with visualizations, task taxonomies have emerged as critical tools in the visualization design process. Our studies of sensemaking within and between visualizations, however, indicate possibilities such as a) designers of visualizations on the web are not fully making use of existing taxonomies, suggesting a gap in design practice, or b) that existing task taxonomies do not capture the diversity of user goals with visualizations

encountered on the web, suggesting new opportunities for taxonomy inference and evaluation.

Augmenting the Visualization Experience
One unique opportunity of the current visualization landscape is that it is adaptable. Namely, the open nature of web applications makes it possible to create web extensions that augment a users' experience of a visualization. For example, Hullman *et al.* have developed a web extension for translating distances and areas into terms that are more familiar to a given user [3]. Similarly, we developed SightLine to give users a low-barrier means for tracking and re-visiting visualizations they encounter across the web.

Supporting Diverse Backgrounds
Many visualizations techniques have been developed *by* scientists *for* scientists. As advanced visualization techniques propagate throughout the visualization ecosystem, there is significant risk that these techniques do not align with the abilities of end-users. In concrete terms, we discovered a visualization literacy gap in our study of undergraduate STEM students, who struggled to reliably locate data and visualizations on the web. Contrast this with experts in the visualization community, who are more likely know that a query such as *d3 nytimes guns* would return visualizations and data on gun statistics. While there exists some early attempts to define visualization literacy in general terms, precisely how literacy differences manifest in the visualization ecosystem remains an open question.

Conclusion
In this paper, we characterize in the emerging ecosystem of visualizations on the web. We review several studies which target sensemaking both *within* and *between* visualizations on the web. The changing visualization ecosystem brings with it new challenges for visualization design and evaluation, as well as new opportunities for advancing the theory and practice of visualization.

References

[1] Mi Feng, Cheng Deng, Evan M Peck, and Lane Harrison. 2017. HindSight: encouraging exploration through direct encoding of personal interaction history. *IEEE transactions on visualization and computer graphics* 23, 1 (2017), 351–360.

[2] Mi Feng, Cheng Deng, Evan M Peck, and Lane Harrison. 2018. The Effects of Adding Search Functionality to Interactive Visualizations on the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '18)*.

[3] Yea-Seul Kim, Jessica Hullman, and Maneesh Agrawala. 2016. Generating Personalized Spatial Analogies for Distances and Areas. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. 38–48.

[4] Peter Pirolli and Stuart Card. 1995. Information foraging in information access environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*. 51–58.

[5] Jordan Sechler, Lane Harrison, and Evan M Peck. 2017. SightLine: Building on the Web's Visualization Ecosystem. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2049–2055.