Your Place or Mine?
Visualization as a Community Component

Catalina M. Danis[1], Fernanda B. Viegas[2], Martin Wattenberg[2] & Jesse Kriss[2]

[1]IBM TJ Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598, USA
danis@us.ibm.com

[2]IBM TJ Watson Research Center
One Rogers Street
Cambridge, MA 02142, USA
{viegasf, mwatten, jesse.kriss} @us.ibm.com

ABSTRACT

Many Eyes is a web site that provides collaborative visualization services, allowing users to upload data sets, visualize them, and comment on each other's visualizations. This paper describes a first interview-based study of Many Eyes users, which sheds light on user motivation for creating public visualizations. Users talked about data for many reasons, from scientific research to political advocacy to hobbies. One consistent theme across these different scenarios is the use of visualizations in communication and collaborative practices. Collaboration and conversation, however, often took place outside the site, leaving no traces on Many Eyes itself. In other words, despite spurring significant social activity, Many Eyes is not so much an online community as a "community component" which users insert into pre-existing online social systems.

Author Keywords

Visualization, communication, distributed community, social data analysis.

ACM Classification Keywords

H5.2. Information interfaces and presentation: User Interfaces. Graphical user interfaces.

INTRODUCTION

The Many Eyes website was conceived as a platform for creating and discussing visualizations. The site, launched in January 2007, allows the public to upload data sets and display them in a variety of interactive visual formats [18]. Design goals were to enable the use of visualization technology by lay users, and to facilitate communication around the visualizations via tools for annotation, sharing and discussion. [9, 10].

In the first eight months after launch, thousands of visualizations were created on the site. In some cases

people spontaneously wrote about how and why they were using the site [19]. Yet much of the activity remained mysterious. One day we might see a burst of activity relating to citations in the PubMed [11] database. Another day a social network diagram of characters on a TV show would appear without comment. And frequently we witnessed data sets and visualizations with odd, code-like titles and labels.

The fragmentary nature of the communication surrounding visualizations was a puzzle, especially given the presence of many features on Many Eyes that allow users to tag, collect, and discuss their work with others. Was this simply "drive-by" testing by casual users? Or was there a broader phenomenon we were missing? Independent blog entries by Many Eyes users suggested that a number of users were incorporating the visualizations they created into other online forums and communities [19]. Unfortunately, the traces of this kind of activity were not enough to let us move beyond speculation.

To better understand the behavior and motivations surrounding visualization creation, we contacted some of our users and interviewed them about how and why they had worked with Many Eyes. These conversations revealed context for much of the activity we had seen. The PubMed visualizations, for instance, were part of a complex collaboration between researchers in two countries and reviewers of one of their papers. The TV show network turned out to be a piece of evidence in a debate taking place at a different web site. The obscure data labels, in several cases, were part of a deliberate audience management strategy for a public website. This new context led us to modify our ideas on how data may be discussed and shared online. It also revealed a mode of usage of Many Eyes that we had not suspected: as a community "component" that could be inserted into other social contexts rather than being bound to the Many Eyes site. Indeed discussion of visualizations was distributed across a variety of online and offline venues.

ABOUT MANY EYES

The Many Eyes site provides "visualization as a service." As described in [18], the service has three main parts. Users may upload data to the site, create and publish interactive visualizations of that data, and discuss their visualizations

and data in various ways. An essential aspect is that all artifacts on the site—data, visualizations, user comments—are visible to the public.

The site design aims at making sophisticated visualization techniques broadly accessible. Uploading data is accomplished by pasting into a text field on a form. One creates a visualization of data by applying one of fifteen templates, ranging from bar charts to treemaps to network diagrams. Configuring these templates is a semi-automatic process. Once a visualization has been published, it is given a unique URL, and anyone on the web can view and interact with it. (Figure 1).

The visualizations may be browsed in several ways: users may see a thumbnail view (Figure 2) of recent or highly rated visualizations, they may search by keywords, or they may look for visualizations created by particular individuals or related to the "topic hubs" described below. Visitors can
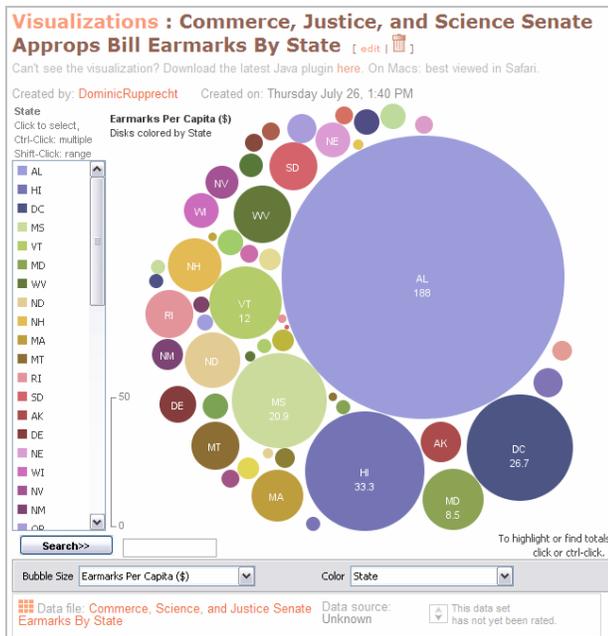

Figure 1: A visualization on Many Eyes

leave comments; each comment has a unique URL which restores the interaction state of the visualization at the time of commenting. (This provides the "mutual accessibility" which is necessary for online discussions of "things"[4].) Users who wish to discuss a visualization outside of Many Eyes can click a "share this" button, which brings up a snippet of HTML that can be pasted into a web page to yield a static image of the visualization that links back to the full interactive version.

Because the site is aimed at lay users as well as experts, it contains tutorials as well as basic "how-to" instructions on the mechanics of uploading data and creating visualizations. There is an essay for each visualization type describing its strengths and weaknesses, and which types of data it may

be appropriate for. The site also includes notes on data "etiquette and style," encouraging users to include metadata and context.
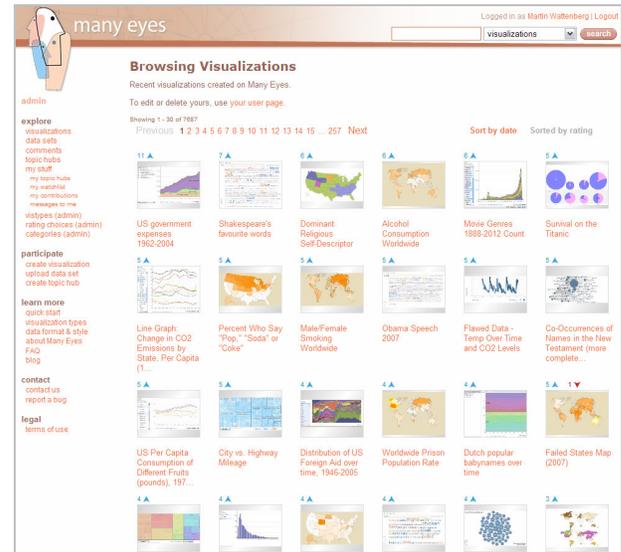

Figure 2: Users may browse visualization thumbnails

Community Features

In addition to the commenting mechanisms, a set of "community" features was launched in June 2007. (Many of these are standard features on other web-based community sites [9, 10].) Since these came after the writing of [18, 19], we describe them here. First, users may create "topic hubs," which are portal pages that may be devoted to a particular theme. A topic hub has a membership list, which any user may join, a set of associated visualizations and data sets, which can be added by any member, and a discussion area. Themes for topic hubs range from the general, such as "Irish Life," to the very particular, such as college classes that use Many Eyes for assignments.

Users also can communicate with one another and monitor the site in a variety of ways. Each registered user has his or her own publicly visible "profile page." This page has an area where one may describe oneself, as well as a "watchlist" of visualizations and data sets, which lets one easily track comments to artifacts of interest. In addition, the profile page contains a list of topic hubs that one has joined, along with indications of recent activity. Similar to traditional social networking sites, users may leave messages on each other's profile pages (although there is no "friends" list). Finally, any registered user may give an up or down rating to a visualization. Aggregate ratings are displayed prominently on browse and search pages.

RELATED WORK
There have been several investigations of how people share visualizations, including preliminary reports on Many Eyes

itself. We describe these now as necessary background for understanding the site and its design.

Information visualization has long been recognized as an effective way to analyze data [2]. In addition to helping individuals make sense of complex information, researchers have noted that visualization is often used collaboratively [16]. Several existing systems explicitly support asynchronous collaboration around visualizations. The DEVise system (an early research prototype) [8] and Spotfire's DecisionSite Posters [13] (a commercially available product) both allow users to engage in web-based discussions around information displays. Each one allows users to create visualizations, bookmark particular interaction states, and leave comments on those states. Both systems were aimed at small sets of expert users in closed (intranet-style) environments. In other words, usage was highly work-focused, and collaboration occurred in a relatively private setting. In contrast, Many Eyes targets the open, massive nature of the Web, where strangers may come together to collaborate in a more spontaneous manner.

There has been recent interest in more public, "social" styles of data analysis. For example, recent sites such as Swivel [14] and Data360 [5] allow public uploading and conversation around data. Neither of these sites, however, offers visualizations that go beyond standard business graphics such as pie and bar charts. In addition, unlike the interactive visualizations on Many Eyes, these sites present only static diagrams.

There are indications that more advanced visualizations may have a special role in collaborative data analysis. In studies of the PostHistory email visualization project, for instance, investigators observed how a graphic depiction of email spurred storytelling and social activity [16]. The NameVoyager, for example, is an interactive web page that displays a graph of historical baby name trends. After it was published, it became a touchstone in a set of thousands of distributed conversations across the web [20].

A research prototype developed in our lab, sense.us [6], features sophisticated annotation and commenting around interactive visualizations. Unlike Many Eyes however, sense.us does not let users create their own visualizations or upload their own data. The system was deployed within the IBM intranet and a controlled lab experiment provided some evidence for the social potential of data visualization.

The Many Eyes site itself was described in more technical detail in [18]. The report in [19] contained usage-based observation of early behavior on the site, including a discussion of blog posts about Many Eyes. Our method in this paper was to seek out and interview a key set of users. This revealed activity which was not observable by an analysis of usage on the site.

METHOD

Our goal for this study was in-depth understanding of Many Eyes users: why they are interested in Many Eyes and what they use the site for. Rather than a broad survey of limited precision, we carried out individual interviews with 20 users. In addition, we analyzed usage logs as a way of situating our focused analysis in the larger context of the general Many Eyes user and as a guide for future research.

Interviews of bloggers and non-bloggers

This study is focused on two user groups. Soon after the launch of Many Eyes, it became clear that blogs played a key role in the life of the site [19]. Many people wrote posts about their experiences creating and using visualizations on Many Eyes. High-profile blogs drove traffic; in lower-profile blogs we saw tantalizing glimpses of the stories behind usages observed on the site. Recent discussions about "distributed community" [1], along with the results described in [19] led us to view this activity as part of the broader social context around Many Eyes. We therefore decided to interview ten people who had blogged about Many Eyes. While bloggers are certainly not a random sample of site visitors, we wanted a clearer picture of what might be an important usage pattern. We compared them to a sample of non-blogger registered users who were their approximate matches in terms of duration of site use and in the number of data sets uploaded, visualizations created and comments made[1].

Identifying bloggers for interviews was not straightforward. We began by performing a series of searches using Google Blog Search, in July and August of 2007, for the terms "many eyes" and "manyeyes."[2] This produced a vast pool of hits, most of which were irrelevant. We ruled out any post written before June 14 when the new "community features" were added to Many Eyes. We also filtered out results which had nothing to do with the site. Many of the remaining hits turned out to be from computer-generated "spam blogs" or from people who had simply copied other posts; we weeded these out by hand, to the best of our ability, by looking for specific references to the poster's use of the site. This produced a list of 87 bloggers; we then removed blogs not in English, where there was no email address, or where there were other potential problems with contacting a user. We sent email to 18 bloggers from the remaining list, inviting them to participate in a phone interview. We received 12 positive responses (67% response) and interviewed the first 10 we could schedule.

---

[1] We ignored topic hubs and topic hub discussions in our matching of groups since these activities represented less than 1% of total activity on the site.

[2] Google is not the only blog search engine, but sampling on two others—Yahoo and Technorati—indicated that there was no significant difference in the results retrieved.

We selected, at random from a list of 432 registered users who had agreed to be contacted by our company, 16 non-blogger users who matched the usage characteristics of the blogger sample. We contacted them through email since our initial attempt to secure their participation in interviews through messages posted on their profile pages did not succeed. We received a positive response from 10 of the 18 (62% response rate). We interviewed all ten over the course of one month.

Interview Description

We developed an open-ended, structured interview covering four main areas. First, we asked demographic questions (age, occupation) and inquired about data analysis and visualization experience prior to using the Many Eyes tools. Second, we asked about how users discovered the site, their initial motivation for visiting it, and what they had done with the visualization tools. We also asked whether they shared their visualizations with others. Third, we asked them about data validity, their confidence with using the visualization tools and whether they had any concerns about their interpretations of data. We added these questions about data to the interview after we noticed that our first five interview participants brought these issues up repeatedly. Finally, we addressed usability and learnability of the Many Eyes site and solicited additional requirements for functionality. Before the actual user interviews, we adjusted the question sequence and wording after a pilot with two employees of the authors' company.

All but one of the 20 interviews was carried out over the telephone and took 45 to 60 minutes to complete. One interview, with a non-blogger, was carried out through email after several failed attempts to connect over the telephone. The typical interview consisted of two experimenters, with one leading the discussion and the other acting as back-up and note taker. The lead interview role was rotated among the first three authors. The two interviewers were in contact during the session and the lead brought the back-up into the discussion as needed, for example, to respond to questions raised by the interview participant. Participant answers were recorded by the note-taker in a scoring sheet that paralleled the interview questions. The note-taker strove to capture the participant's answers in verbatim form.

Application and Server Logs

We collected two types of log data for the Many Eyes site: application logs to document the activities of users on the site and server logs to report on site traffic. Both logs provide data beginning with the site's launch in January 2007.

The application log contains a record of the creation of artifacts by registered users, including data sets, visualizations, and the creation of topic hubs as well as a record of their social activity, comprised of comments on visualizations or data sets and discussion in a topic hub.

The application log also recorded visualizations created and comments made by anonymous users.

In addition, we collected standard http server logs [7]. Of interest were counts of unique visitors, page requests and visit duration times. Measures of all three items were based on IP addresses and as such are approximate.

RESULTS AND DISCUSSION

Our analysis begins with an overview of site usage, as determined from log files. We then characterize our interviewees, and discuss how they discovered the site and learned to use it. Having established this background, we focus on three topics: First, why did interviewees use the site and how did they extract value from it? Second, what was the nature of their conversations and collaborations around the visualizations? Finally, what concerns did they have specifically about data and visualizations?

A Log-Based Overview of Usage

The analysis of results is focused on the seven month period starting January 23, 2007 through August 14, 2007. Server logs showed a monthly mean of roughly 27,000 unique visitors, 49,000 visits, and 400,000 page requests. The median number of monthly unique visitors was around 20,000, with a peak of over 50,000 in March when the site was written about by a number of high profile bloggers unconnected with the present report. The majority of visitors stayed only briefly, but approximately 10% of visits lasted 30 minutes or more.

Table 1: Activity on Many Eyes between January 23, 2007 and August 14, 2007 by registered and anonymous users

Activity	Registered	Anonymous
Registration	3.069	N/A
Users who uploaded datasets	1,472	N/A
Datasets	5,347	N/A
Users who created visualizations	972	unknown
Visualizations	3,449	571
Users who commented	227	unknown
Comments	1,268	296

Application logs recorded a total of 3,084 registered users, of which 1,542 (50%) created one or more artifacts on the site. Table 1 shows these events for the registered users as well as for anonymous users. Of the registered users, 48% (*n*=1,472) uploaded one or more datasets to the site and 31.7% (*n*=972) created and saved one or more visualizations on the site. In each case, active users created

on average about three and a half of each of these artifacts. Only 227 or about 7% of registered users made comments.

Characterizing the Interview Participants

As noted above, the two groups of interview participants—bloggers and non-bloggers—were matched by how long they had used the site and by the number of artifacts they created. For bloggers, the mean duration of use was 5.0 months (range: 1-7) while the mean for the non-bloggers was 4.8 months (range: 2-7). Bloggers uploaded an average of 23 datasets (range: 0-56), created 21 visualizations (range: 0-67) and made .5 comments (range: 1-11). The corresponding numbers for non-bloggers are a mean of 22 data sets (range: 4-91), 24 visualizations (range: 4-95) and .33 comments (range: 0-42). On average, the interviewees were far more active than the average site visitor. Our study therefore does not shed light on the experience of site visitors overall—a group that naturally includes a large number of casual visitors—but rather on the ones who derived enough value from the site to return repeatedly.

All participants were selected to be speakers of English in order to facilitate the interview. Fourteen were from the United States, one from Canada and five from Europe. All were male, with ages ranging from 19 to 64 (mean 36). Interviewees reported 10 distinct occupations—including government official, business analyst, and an artist. Six participants were scientists—two researchers and four graduate students.

In general, the study participants were more experienced in data analysis than in visualization. Ten participants, equal numbers of bloggers and non-bloggers, reported advanced knowledge of data analysis methods, several describing themselves as "working with data for a living" or as "data junkies." Of the other half, eight described themselves as having a passing knowledge of data analysis and statistics, often many years prior, while two reported having very little experience. Interviewees had more limited experience with visualization tools. The most commonly used tool was Microsoft Excel followed by statistical packages such at SAS and R. Five users reported having written their own tools and scripts to gather data, indicating a higher level of skill by this sub-group.

Bloggers differed from non-bloggers in terms of where they first heard about the Many Eyes site. Seven of the ten bloggers reported first becoming aware of the site through other bloggers, while none of the users who did not blog did so. Instead, non-bloggers reported various sources, with five finding out about it by "poking around on the internet" or through email and mailing lists. The others heard about the site from colleagues or through paper publications, while two did not remember.

Study limitations due to sampling method

This study is focused on users with a significant interest in Many Eyes, either through very active participation on the site or self-report on their blogs. It does not attempt to understand another important group, the less committed user. There are many interesting questions that therefore remain unanswered, for example, why did 50% of the registered users who uploaded data sets not create or save visualizations. However, the focus on the more active users not only helps us understand the usage of this important group, but also helps us to explore whether such "power" users might serve as "bridges" to help the less data and visualization literate to evolve these skills.

Learnability, Usability

The visualization tools in Many Eyes go beyond familiar business graphics, so we were curious to see how people learned to use them and which aspects of site usage seemed easy or hard. As might be expected from our high-activity sample, most (15) interviewees said the site was easy to use. Nine users cited trial-and-error as one way they learned to use the tools, while twelve mentioned reading the documentation. Several users used glowing terms to describe usability and expressed surprise at how simple it was to create a visualization: In the words of User-P: *"I can't even remember learning it, because it was that easy."* A few interviewees did point out problems, with three mentioning technical glitches—for example, certain characters caused one users' data set to be parsed incorrectly. Users were also split on the ease of bringing in data to the site. Those who did not use spreadsheets discussed the challenges of formatting data into a tab-delimited table, a procedure User-D termed "*laborious.*" Five interviewees had data in spreadsheet format already, however, and these all commented on the ease of the upload procedure—often extremely favorably. As User-H put it, "*what blows me away is being able to paste into that text field.*"

When we asked about missing features, eight interviewees suggested changes to the visualization toolset, from minor adjustments (inclusion of logarithmic scales) to new techniques (e.g., using animation for time-based data). Other requests, described in more detail in later sections, included privacy controls and export features.

User Motivations

The major initial motivations for users coming to the site included a curiosity about the tools and specific needs to visualize data. Participants reported using the visualizations they created on Many Eyes in a variety of ways, some personal and some social. We now discuss these motivations in more detail.

Extracting value: for oneself, with a group, for others

Users who primarily visited the site for personal reasons included people who described themselves as "data junkies." These users were interested in exploring more sophisticated visualization tools than they had worked with before. Some also had specific needs to work with data for

school projects, in order to further their own understanding of their subject matter.

For example, two of the graduate students stated that the visualizations helped them gain insights in their academic research. User-D, for instance, is a historian who has been studying the former occupations of Canadian Parliament members in the 19th century. After analyzing a treemap he created, he realized that one of his preconceived notions about the political scene at the time was incorrect:

It hadn't occurred to me [...] that administrators had suddenly become very very important over time. And it just popped right out. I'd come in with preconceived notions of politicians being lawyers. The big thing we hadn't accounted for was the bureaucrats and professional politicians were the ones making policy. Suddenly the people who were formerly involved in enacting policy are now the people making policy… The effect of [that] bias… was enormous.

Eight of interviewees started using the Many Eyes tools to communicate with their work or school teams. For example, User-C, a government official in central Europe, incorporated Many Eyes visualizations in a public audit report:

A visualization of the budget data was shown in a press conference and [on a] PowerPoint presentation. The Tag cloud was used for the cover of the State Budget Audit Report for 2006. It shows the frequency of the words in one of the paragraphs of the law which was the focus of the audit.

User-T, a university professor, focused on the shared exploration value of visualizations among students in his research methods courses:

What I really like the visualizations for is not providing the answer but starting the conversation and providing a launching off point… [The students] are much more likely to spark a conversation around a graphic representation than a numeric representation.

We also heard about explicitly pedagogic uses, in which an interviewee created and shared visualizations with the intent of teaching something to a group. While this occurs in peer relationships as well, our best examples come from teacher-student situations. For example, User-O, a middle and high-school teacher working with learning-disabled students, has been testing the impact of Many Eyes visualizations in class:

I showed [the visualization] to the students and they loved it. I could create a map of the world and they could see that red means one thing or green means something else. They liked it because they were able to understand examples, [that they] previously couldn't. [I am using Many Eyes] to find a different learning modality for our students.

Perhaps surprisingly, User-O further described how the tag cloud visualization turned out to be effective for dyslexic children—even though it was entirely made of text. User-T, an educator who teaches research methods courses, also noted the power of visualizations to connect with users who are not comfortable with analyzing data:

Some of this turns into visual literacy appreciation. [My] students are terribly gun shy about quantitative research. [Many Eyes shows] the fun side of numbers and they respond surprisingly well.

These examples highlight the power of visualizations to drive insights for individuals and groups which may be elusive when data is presented in textual form.

Broadening the audience

We heard repeatedly that sharing of data became a focus even when users initially were only interested in exploring the tools for themselves. A typical case was User-M, who said that he "doesn't really create [visualizations] with an audience in mind," but went on to add that he had begun to use his visualizations as evidence in some online debates. Similarly, many users reported aiming initially at a small, bounded set of people and then widening their audience. For example User-C, a retiree, was initially interested in visualizing data on an illness that afflicts a close friend. Over time, as he found interesting views of data (what he termed "little gold nuggets") he started looking for ways to share the data and engage others.

When I put things out on Many Eyes, I try to leave at least a comment that shows the significance of the data or visualization.

Similarly, User-Q is a university administrator who was initially interested in using visualization to navigate across courses in the university's web catalog. He found that the techniques enabled him to get insights about how students aggregate course work for their degrees. Consequently he has been sharing his visualization on his blog.

User-J, a graduate student working with a colleague located in a different continent, originally benefited from the Web accessibility of Many Eyes for a team-based analysis of research data. (These were the visualizations of PubMed data mentioned in the introduction.) Later, he and his teammates found another use when they submitted an article for review in an academic conference:

[We were] collaborating with university in Brazil, so it was good to explore this visualization [with my colleague there]. At the end it was complicated to put [the visualization] in paper because they have 10 or 11 page limit. So that time we put things in Many Eyes so the reviewers could access the [live] visualization.

Thus, while several of the users were initially attracted to Many Eyes only to explore the tools or visualize data for themselves, they frequently found social uses for their visualized data. Similarly, those whose social uses were originally aimed at small teams of peers often found themselves engaging broader audiences.

Conversation and commentary

That so much activity was collaborative leads to something of a mystery. Consider that Many Eyes was designed with the goal of supporting on-site discussion. Each visualizations and data set has its own forum; the June redesign added topic hubs with their own discussion areas;

and users can leave each other messages and create watchlists to monitor other conversations on the site. Thus there was ample opportunity for discussion on-site.

What is puzzling, in light of the heavy collaborative use of the visualizations that users produced, is that the logs show these features were used relatively infrequently. Table 1 showed that the commenting functionality was used little (only 1,268 comments from 3,084 registered users). Despite the portfolio of collaborative functionality, the bulk of the discussion around the visualizations reported by the interviewees occurred off site. Where, how and why did this happen?

Where Discussion Took Place
The general answer to this question is that it took place in social contexts which were defined by the user: in small work teams, public meetings, a user's personal blog, or other web forums in which the user participated. While the audience size and the degree of familiarity between the creator of the visualization and their audience varied, the common thread is that the user targeted his visualizations by defining where the discussion took place.

This pattern — of importing Many Eyes visualizations into an existing debate or presentation — was common. User-I described hanging printouts of visualizations of volunteers' activities on a wall in meetings in order to communicate the importance of the work to trustees and funders of volunteer organizations. User-B, the artist, shared his visualizations of art funding with various email based groups:

An email group in the Northeast of England that has about 150 members. Another email group of 60 ... plus a link on my blog.

Some used their visualizations as supporting materials for discussions on other web sites. For example User-M, a psychology researcher, targeted some of his visualizations at other members of a news discussion site in which he participates:

I'm an active contributor on a web site called [XYZ]. *It's pretty much a decentralized news community thing in which the guts of what's going on is commentary on news articles. That develops into full-fledged debates in which data presentation can be useful.* [I create a visualization] *Then I post a link to it within my argument.*

The diagram of the social network of a TV show mentioned in the introduction turned out to be one of this user's exhibits in an argument.

In some cases the visualizations our users created reached enormous audiences. User-D blogged about a visualization he created with a religious message and received attention beyond his usual readership. His description of the result shows he was not expecting such a wide audience:

I didn't anticipate all the traffic that my first visualization got. It got over 40,000 [visits] – which is a month's worth of traffic for me. Also, the direct feedback—from my blog readers to Digg. The

progression was from "people you know" to "people you don't know" to "people who are against you."

Why did so many discussions around Many Eyes visualizations take place outside of Many Eyes? In some cases, users collaborated around data with others in the physical world. In such cases, the appropriate way to share data was by hanging data visualizations on walls of meeting rooms or through government publications. Yet in many other cases, where the sharing was done digitally, the goal was to target a specific audience. For example, User-L—a blogger and linguist who participates in a number of linguistics-related discussion lists—noted:

My goal was to put the data there [in Many Eyes], *get the result, and comment somewhere else. Most of the elaboration has taken place outside Many Eyes in my case.*

This user further referred to the arcane nature of his subject (word distributions in texts) when asked about holding discussions on the Many Eyes site:

I could bring it [into Many Eyes], but I do not assume that the rest of the world is interested. [I] *talk with an in-group. It's an issue of pre-existing community; might not mesh with a new community.*

Other bloggers expressed a sense that they were choosing their audience. User-Q, the university administrator, said he wrote for those "*who play with web technology on the fringes of their universities.*"

Further validation for the value of off-site discussion came when we asked about desired improvements to Many Eyes. Five interviewees wanted the ability to export the visualizations for use outside of Many Eyes, either as a static image or to embed in their own site.

The on-site discussions were not completely ignored, of course. Our logs show that 1,268 comments have been made (excluding comments made by members of our research team). Six of the interviewees reported posting comments to other people's visualizations. However, several of the users who tried to engage others on the Many Eyes site found little response. For example, User-E said about his activity:

I have not gotten a lot of responses from people on the site—but I don't get discouraged, I just keep trying.

When asked whether he had posted comments on the site, User –T, the teacher of research methods, responded:

A little bit ... At first there weren't a whole lot of comments going on, it felt like an empty room. Would want a critical mass of people... they weren't engaging with the data. [I] *wanted more critical comments on the data.*

The contrast between the feeling of an "empty room" on the site, and User-F who received a month's worth of traffic from a Many Eyes-based post is dramatic—and suggests that it is worth rethinking traditional notions of success based on on-site critical mass. We return to this issue in the General Discussion.

Concerns about data—and other eyes

We asked several questions about how users regarded the public nature of data sets they used and the visualizations they created: Did they have any concerns about putting it out on the Many Eyes site? Did they have concerns about their interpretations of data? Were they comfortable with their selection of visualization techniques? The responses varied widely, but two themes emerged from the answers: privacy considerations and data/visualization correctness concerns.

Privacy and Confidentiality

Three users requested a private mode in Many Eyes where they could upload and visualize confidential and proprietary data. This request is not surprising given that data sets people work with on a daily basis are not always public.

One of interviewee's concerns was about privately identifying information, such as names of individuals. One of the ways this issue was addressed was to anonymize data—the source of the code-like data labeling alluded to in the Introduction. For instance, one individual who runs a volunteer center and had uploaded data about the accomplishments of various volunteer organizations replaced the names with pseudonyms. Others did likewise for private marketing data or the reported goals for volunteers in non-profit organizations.

A second method of obtaining some privacy was metadata concealment. Several users gave obscure titles to their data sets and visualizations—for example, one user created a series of treemaps with titles such as "Treemap of t," "Treemap of whlp," "Treemap of 314." Pseudo-private discourse techniques such as these are well-known in the blogging community where bloggers rely on their readers' contextual knowledge when disclosing information publicly. For example, a blogger may use descriptions such as "my sister" or "my friend from work" rather than disclosing people's full names [12]. Readers familiar with the bloggers' social network are able to parse such postings.

These ad hoc methods of access control clearly did not satisfy all users, however. Seven interviewees, when asked about desired future features, wanted to control access to their data and visualizations (e.g., User-G wanted a "*walled-in community*" for proprietary data).

Even though confidential data is a clear motivator for private viewing, it was not the only reason mentioned by interviewees for wanting a private mode on Many Eyes. Thirty-five percent of participants mentioned data validity and visualization choice as strong motivations for private viewing. Users felt that getting a data set to a point where it was publishable took a number of iterations; a process during which they would prefer their data to be kept private. User-X reported being frustrated by the fact that the first time he uploaded a data set to Many Eyes, it was immediately picked up by someone else and visualized even though it was incomplete.

The data set I uploaded at first was not completely accurate; which leads me to my biggest complaint—if you upload a data set then you can't get rid of it; even if it's wrong. So now I upload data sets and very quickly I try out a bunch of visualizations before making anything public.

Data validity and visualization choice

Those using publicly available data sets felt comfortable with its accuracy, assuming that if the data sets were from reputable sources, they would be correct. Thus, it was considered important to acknowledge the source of data. (Overall, 62% of data sets on the site are accompanied by source information.) User-M's sentiment was typical:

It's a concern of mine but I try not to work with data that is not up to a minimal standard. Citing the source is important.

User T, a university professor, was clear about the importance of citing sources:

Unless there's a link to the source, I never trust data that I see on the site.

Interviewees became more concerned about the validity of data sets for which they could claim some responsibility, as in data one derives from publicly available sources. For example, user-S discussed his concerns with drawing conclusions from a data set he created from a public source:

If you draw the wrong conclusion, then it's easy for someone to go back to your raw data and find the error.

As we noted, many of the users we interviewed were experienced with data analysis or visualization. The more sophisticated ones raised subtler questions about the validity of data and the meaning of the relationships that visualizations seemed to suggest. For example, asked whether he had concerns about the validity of his historical data about the occupations of members of the Canadian Parliament, User-D, commented:

That data is correct in terms of how people self-identify [their occupations]*, but whether that's true to life is a whole different story.*

GENERAL DISCUSSION

These results resolve some of the minor mysteries reported in the Introduction. They also shed light on the larger question of how a site that has so much activity may also feel in certain ways like an empty room. In this section we discuss implications of our study for discussions around data as well as other forms of online communication.

The value of visualization

The tools provided on Many Eyes for creating visualizations appeared to provide genuine value to our interview participants. This replicates previous findings about the value that visualizations provide users [2]. In some cases, participants credited visualizations with promoting new insights into data they had been working with for extended periods. Others commented on the power of visualizations to engage people and cause them to ask

questions rather than simply look for answers. Still others credited the visual representations of data with making previously inaccessible domains of knowledge accessible. Perhaps the best evidence of value is that our interview participants kept using the technology; their pattern was to return when they had additional data to explore.

The interviews also revealed that Many Eyes visualizations are making their way into users' communicative practices. That 18 out of 20 interviewees shared their visualizations suggests communication is highly relevant to the application of visualizations in everyday life. Interviewees reported sharing their creations with groups ranging from small research teams, to loosely coupled stakeholders organized around some topic, to larger audiences of strangers accessed through standard paper-based publications or web audiences reached through blogs. From science to education and politics, users of the site are co-opting visualization techniques and embedding them in a series of daily communication practices. Many Eyes visualizations are helping users insert data into everyday discourse.

Also, our users described issues related to data correctness, the selection of appropriate visualization techniques, and data interpretation. Often these concerns related to the fact that Many Eyes data and visualizations are public. The request for a private view mode was motivated not only by a desire to upload private data but also by the wish to publish data "when it's ready." A second set of data-related considerations are typified by the comment of User-S that he was more concerned with accuracy when using Many Eyes, since a reader could go back to the original data to disprove his claims. Of course, any debater should be concerned about being refuted or the quality of evidence— but visualizations may sharpen awareness of these issues.

Many Eyes visualizations as a community "component"
Although many visualizations on the site succeeded in sparking communication, very little of this communication took place on the site itself. Few of our interviewees and less than 7% of the registered users left comments on the site. One might hypothesize that users who only shared their visualizations with people they know–work teams, research groups or members of the same organization–may have been reluctant to engage in discussions with strangers on a public site. But how about our sample of bloggers? Recall that one blogger told us of receiving over 40,000 hits to one of the entries in which he used a visualizations he created on the site. So, why did he not develop discussions on the Many Eyes site?

We suggest that since Many Eyes provides persistent URLs to visualizations for easy linking, it can become a kind of *community component*, playing a small role in many other communities rather than being a full-fledged community in itself. Indeed, while one interviewee described Many Eyes as lacking "critical mass" in its comment forums, it seems

to be a successful part of many other social circles. Some videos from YouTube (www.youtube.com) serve the same function, being embedded on sites all over the web instead of having been generated by a cohesive community that exists only on YouTube. Indeed, the ubiquitous availability of web widgets[3]—ranging from miniature photo albums to quizzes and quick online games—has changed the shape and boundaries of online communities. Baym observes that online communities are starting to spread themselves across multiple online spaces [1] and that it is not enough to study a particular web-site to understand a given online community anymore. While the traditional views of online community have emphasized concepts such as audience critical mass, or lurkers, it may be that in the "web 2.0" era, new concepts are helpful for "community component" sites. Perhaps instead of requiring a critical mass of on-site participation, it is enough to achieve "critical connectivity," that is, a sufficient number of links from the outside web that potential users are likely to find the site.

We also speculate that a blogger's decision to hold discussions at his place, not ours, may be related to the notion of common ground [3] and the notion of audience management observed in other social phenomena, including blogs [15]. The important thing for several of the bloggers we interviewed was to reach pre-existing audiences—e.g., people interested in religion, linguistics or web technologies. For those who inhabit several online spaces, it may be easier and more fruitful to bring in artifacts from the outside to existing "places" rather than to join a whole new community and develop an audience for their work.

CONCLUSION AND FUTURE RESEARCH
The main research goal of Many Eyes is to test the hypothesis that asynchronous collaboration around visualizations can provide value to users. A second goal is to explore the ways users might appropriate this technology, with the hope of uncovering new uses for visualization. Our study shed light on both these areas.

The patterns of usage we have observed on Many Eyes seem to corroborate our hypotheses about the collaborative nature of much activity around visualizations. Even interviewees who initially used Many Eyes for individual analysis reported that they eventually engaged in collaboration around Many Eyes. Perhaps not surprisingly for a public site, issues of privacy and audience management were frequent concerns. Less expected were some of the ways that our interviewees handled these issues. In particular, a high proportion of social activity related took place outside of the site: on blogs, in face-to-face meetings, or in email. These modes of use suggest that Many Eyes became a community component, sparking social activity in other arenas.

---

[3] A widget is a portable piece of code that can be installed and executed within any HTML web page by an end user.

These findings point to themes for future research and design. One question is whether the idea of a community component is applicable beyond data visualization. Given the massive popularity of "widgets" such as Facebook applications or online quizzes that spark discussion, it seems possible that such components may play an important role in the web ecosystem.

In addition, our interviews suggest several research directions related specifically to data. While a natural way to handle privacy concerns would be to provide flexible access control for Many Eyes, that might nullify some of the aims of the site in encouraging open, public participation around data. An opposite, yet complementary, approach would be to make data handling even more public by supporting incremental data collection from many users. Such a feature might prompt Many Eyes users to the site as the nexus of collective data aggregation.

Finally, our interviewees have appropriated the Many Eyes technology in ways that we would not have anticipated. An example is the teacher who used Many Eyes in a class for disabled students: helping accessibility for cognitive impairments was not on our minds as we built the site—and the idea that a tag cloud could be helpful to dyslexic children is counterintuitive. Investigating ways that visualizations can help people with cognitive disabilities seems like a promising avenue of research.

ACKNOWLEDGEMENTS

REFERENCES

1. Baym, N. K. The new shape of online community: The example of Swedish independent music fandom. First Monday, 12, 8 (2007).

2. Card, S., McKinlay, J., and Shneiderman, B., (Eds.), (1999). Introduction, *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufmann.

3. Clark, H. H., and Brennan, S. A. (1991). Grounding in communication. In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.). *Perspectives on socially shared cognition*. Washington: APA Books, pp. 127-149.

4. Churchill, E. and Erickson, T. Introduction to this Special Issue on Talking about Things in Mediated Conversations. *Human Computer Interaction*, Vol. 18 Nos. 1&2, pp. 1-12. Lawrence Erlbaum Associates, 2003.

5. Data360: http://www.data360.org, retrieved 03-30-2007

6. Heer, J., Viégas, F.B., & Wattenberg, M. *Voyagers and Voyeurs: Supporting Asynchronous Collaborative Information Visualization*. In *Proc. CHI 2007*, ACM Press (2007), 1029-1038.

7. http://httpd.apache.org/docs/1.3/logs.html/.

8. Livny, M., Ramakrishnan, R., Beyer, K., Chen, G., Donjerkovic, D., Lawande, S., Myllymaki, J., & Wenger, K. *DEVise: Integrated Querying and Visual Exploration of Large Datasets*. In *Proc. ACM SIGMOD*, ACM Press (1997), 301-312.

9. Mynatt, E., Adler, A., Ito, M., and O'Day, V. (1997) Design for Network Communities. In *Proc. CHI 1997*, ACM Press (1997), 22-27.

10. Preece, J. *Online Communities: Designing Usability and Supporting Sociability*. New York: Wiley & Co., 2000.

11. PubMed. http://www.ncbi.nlm.nih.gov/sites/entrez.

12. Qian, H., & Scott, C. R. (2007). Anonymity and self-disclosure on weblogs. *Journal of Computer-Mediated Communication*, 12(4), (2007), 1428-1451.

13. Spotfire DecisionSite Posters: http://www.spotfire.com/products/ decisionsite _posters.cfm, retrieved 03-30-2007

14. Swivel: http://www.swivel.com, retrieved 03-30-2007

15. Viégas, F.B,. Bloggers' Expectations of Privacy and Accountability: An Initial Survey. *Journal of Computer-Mediated Communication,* 10 (3), 2005, article 12.

16. Viégas, F.B.., boyd, d., Nguyen, D., Potter, J. & Donath, J. *Digital Artifacts for Remembering and Storytelling: PostHistory and Social Network Fragments*. In *Proc. HICSS-37*, 2004.

17. Viégas, F.B. & Wattenberg, M. *Communication-Minded Visualization: A Call to Action*. IBM Systems Journal. 45(4), 2006.

18. Viégas, F.B., Wattenberg, M., van Ham, F., Kriss, J., & McKeon, M. *Many Eyes: A Site for Visualization at Internet Scale*. In *Proc. IEEE Information Visualization (InfoVis)*, 2007.

19. Viégas, F.B., Wattenberg, M., van Ham, F., McKeon M & Kriss, J. *Harry Potter and the Meat-Filled Freezer: A Case Study of Spontaneous Visualization*. In *Proc. HICSS-41*, 2008.

20. Wattenberg, M. *Baby Names, Visualization, and Social Data Analysis*. In *Proc Information Visualization, (InfoVis),* 2005.

21. Wikipedia: http://www.wikipedia.org, retrieved 03-30-2007.