

Harry Potter and the Meat-Filled Freezer: A Case Study of Spontaneous Usage of Visualization Tools

Fernanda B. Viégas, Martin Wattenberg, Matt McKeon, Frank van Ham, Jesse Kriss
IBM T. J. Watson Research Center

Abstract

This paper is a report on early user activity in Many Eyes, a public web site where users may upload data, create visualizations, and carry on discussions. Since the site launched, users have uploaded data and created graphics on everything from DNA microarray data to co-occurrences of names in the New Testament to personal gift-giving networks. Our results show that in addition to traditional data analysis, Many Eyes is used for goals ranging from journalism and advocacy to personal expression and social interaction. We propose several implications of this usage for visualization designers and contend that these findings suggest a role for visualization as an expressive medium.

1 INTRODUCTION

The biggest words on the visualization are *chicken, pork, pulled, steak, and thighs*. *Beef, salmon, and tenderloin* also stand out in large type. On closer inspection, you realize that *pastrami, prosciutto, and turkey* are buried there too. Looking at the title, you see that this is a visualization of the contents of John's freezer—only the meat, of course. You don't know John and you may never meet him but, at some unexpected level, you can relate to the contents of his freezer.



Figure 1: Tag cloud created on Many Eyes

This tag cloud was created by a users of Many Eyes, a website where people may upload data, create visualizations, and carry on discussions. The site provides tools for collecting, visualizing, and discussing data. (The system design is described fully in [18].) In this paper we examine early user activity on the Many Eyes site with an eye toward answering two main questions. The first is focused: Do usage patterns corroborate or

contradict the hypotheses advanced in [15][20][6] that visualizations can catalyze social activity? The second question is open-ended: When powerful visualization technology is made broadly available, how do people spontaneously apply it? Do they engage solely in the visual analytic tasks that are conventionally associated with complex information displays, or do they use visualizations for other purposes?

The first question, regarding social activity, is prompted by results that indicate visualization may have a catalytic effect on communication between users. The evidence for this hypothesis, however, is still incomplete. In the case of PostHistory [15] and the NameVoyager [20], social interaction was observed around visualizations, but in each case it was an unexpected side effect of a system designed for other purposes. In the case of Sense.us [6], lab studies and limited deployments provided further evidence, but data on extended public usage was not available.

The second question, on how people spontaneously apply visualization when given free rein, arises because—for many viewers of Many Eyes—the site represents the first time they have had the ability to make their own visualizations. For the most part, information visualization research projects are rarely available to a large population of users. Typically researchers build applications with specific data sets in mind and it is not uncommon for the creators of a visualization system to be its main users. Even when user studies are carried out, the setting is usually controlled and the tasks are often pre-determined. Artificial tasks and environments make for repeatable experimental results, but possibly at the cost of general validity. It is hard to observe visualization usage “in the wild” and it may be even harder to find a large body of users to study.

To address these issues, proposals have been made for a “case study” approach to visualization evaluation [6]. Learning how users employ applications seems essential for progress in the field of information visualization. Studying the ways in which everyday usage differs from the intended goals of the tools we build might open new areas of research for the infovis community. For example, visual systems are typically built on the assumption that data exploration and analysis are the main tasks people want to accomplish when using a visualization application. How true

is this belief? It is hard to say, given that we do not have a large body of accounts on how users utilize visualization tools.

Because Many Eyes is a public online system, we have the rare opportunity to see how a visualization application fares with thousands of users. Since our target audience is regular Web users (anyone who knows how to browse the Web pages, and has some sort of interest in data), Many Eyes is able to attract a wealth of visitors ranging from novices, who may never have played with interactive visualizations before, to data experts such as scientists and journalists.

To examine the activity of this set of users we employed a range of methodologies, from standard empirical methods—content coding and statistical analysis—to less conventional approaches such as using search engines to find blog posts that referred to Many Eyes. We studied site activity in all its different manifestations: data sets uploaded, visualizations created, and comments contributed.

The results suggest a strong social component of Many Eyes usage, ranging from playful behaviour (using visualization as games) to socially conscious activity (visualization for advocacy and solace) to turning visualizations into a mechanism of self-identity (data mirrors). In answer to our second question, we find that data exploration and analysis are not the only goals when users view and create visualizations. Indeed, a data set that describes freezer contents is just as characteristic as a table of numbers on global warming.

The next section presents an introduction to Many Eyes and an overview of related work. We then introduce case studies that illustrate the ways in which Many Eyes visualizations are used on and off the site. The last sections describe content coding results and an analysis of the limitations of this study. We end with a discussion of the implications of our findings.

2 BACKGROUND AND RELATED WORK

2.1 About the Many Eyes site

Many Eyes is described in detail in [18]; here we only provide a summary of the aspects important for this paper. The site launched in January of 2007, is freely available to the public, and provides “visualization as a service.” Users can upload data to a shared public repository, and then apply any of fourteen basic display techniques—ranging from pie charts to treemaps—to those data sets to create new visualizations.

The site is designed for social interaction as well as visualization. Each data set and visualization has an associated discussion area where users can place comments. When a user places a comment on a visualization, a link is kept to the state of the display at that moment; a user reading that comment can click on the link to

restore the visualization to the state it was in when a viewer saw it. The mechanism is similar to those used in Spotfire DecisionSite Posters [12] and DEVise [7].

A distinctive aspect of Many Eyes is that it has been designed to exist as part of the Web ecosystem. Users can place images and links to the site on their blogs, with a special button providing the necessary HTML code. Feeds in RSS format are available to alert viewers to new data, visualizations, or comments of interest. As a result, a significant amount of activity related to Many Eyes occurs on other web sites.

2.2 Related work

Several streams of investigation are relevant to the current work. In the past decade much attention has been paid to the area of social visualization: the display of social data for social purposes [15]. By this definition, some of the practices that have spontaneously arisen on Many Eyes may be considered examples of social visualization. Nonetheless, Many Eyes as a whole differs from conventional social visualization systems in two respects. Social visualizations have traditionally been aimed specifically at data from a particular system (e.g. Usenet or chats) rather than providing generic visualization components, and they often are meant to serve as a conduit for communication or impression formation rather than spurring discussion in a separate forum or context.

A second related area is a set of “communication-minded” visualizations [17] that have been reported and created in recent years. Reports on systems such as the NameVoyager, Vizster, and Sense.us have explored how visualizations can be a spur to social, playful analysis of data. Many Eyes was designed in part to provide a large-scale testbed of these ideas.

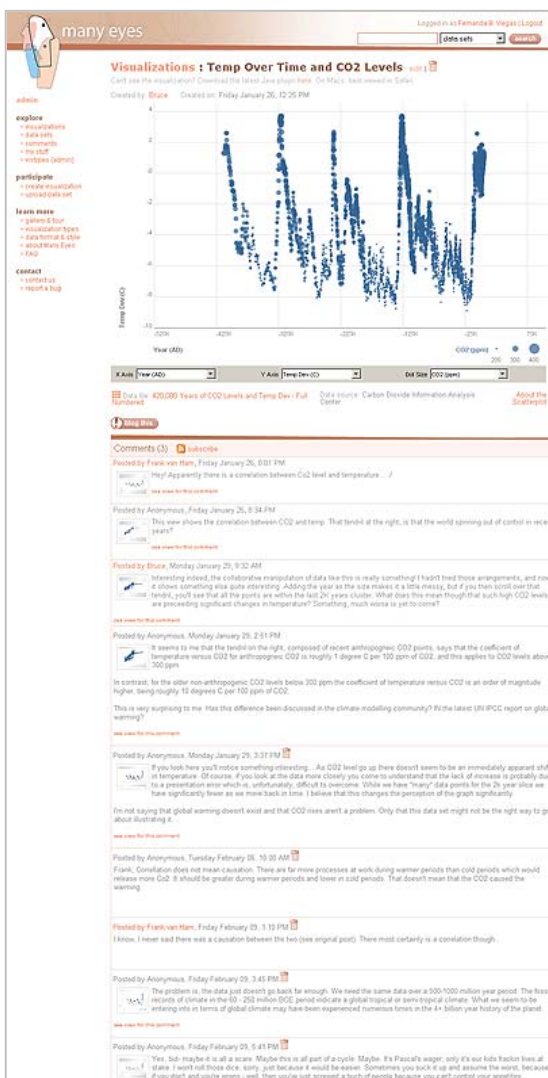
In addition to serving as a testbed for the social aspects of visualization, Many Eyes provides a platform to examine broader questions about visualization usage. Several authors have addressed the question of what people want from visualization. One approach, e.g. [1], has been to ask potential users what they would want to know about various data sets. Another approach is described by [9] in which Plaisant notes that case studies of tools in realistic settings form the least common type of study—yet it is potentially one of the most useful. The current paper presents such a study with a wide base of users on the Internet.

Finally, several commercial web sites aimed at data sharing and charting have recently appeared. Swivel [13], Data360 [3], and Dataplace [4] all provide web-based repositories where users can upload data and make simple charts. All are aimed at group exploration of data, though none provides the kind of sophisticated visualizations seen on Many Eyes. No studies have been published on usage patterns of these sites.

3 QUALITATIVE ACCOUNT

Users create visualizations on Many Eyes for a diverse set of reasons. Aside from simple testing of the site, we have seen examples that range from scientific research to artistic expression. In this section we touch on the major themes that have emerged, with in-depth descriptions of examples of each. The qualitative descriptions are meant to be miniature case studies that provide a sense of the richness of site activity. In several cases, we have included additional information gleaned from blog entries that discuss particular visualizations. In Section 4 we augment these descriptions by quantitative analysis.

Figure 2: Global warming visualization and discussion



3.1 Visual Analytics—global warming and blogs

Three days after the site launched, user Phil¹ uploaded a data set entitled *420,000 Years of CO2 Levels and Temp* and used a scatterplot to display it (Fig. 2). The visualization has generated 20 comments as of this writing. The comments range from political arguments to statistical analysis. The interest in the data set is clearly not just academic, but comes from the context of scientific reports and news coverage about global warming. As Phil, the uploader of the data, commented:

Interesting indeed, the collaborative manipulation of data like this is really something! I hadn't tried those arrangements, and now it shows something else quite interesting. [...] What does this mean though that such high CO2 levels are preceding significant changes in temperature? Something, much worse is yet to come?

A different kind of analytic activity comes from user “Elton”, a PhD student in Linguistics in Germany. (These details are known because he writes about them in a blog.) His research looks at corporate blogs as a new genre with distinctive linguistic characteristics. After stumbling upon Many Eyes, a series of visualizations of his data corpus have become additional evidence for his thesis work. Elton has, thus far, uploaded 43 data sets that track formality measurements on blog entries. He has visualized the majority of these data sets (sometimes using more than one technique for each set of data). The activities of Phil and Elton show that Many Eyes is used for traditional data analysis.

3.1.1 Data Quality and Error Hunting

One of the main concerns in a site like Many Eyes is the possibility of users uploading malicious or incorrect data sets. Because there is no guarantee that content on the site is accurate, the burden of verifying data falls on its users. One theme that runs through many comments is data quality. Often users simply want clarification (are those dollar figures adjusted for inflation?) but in some cases they have spotted genuine errors.

One interesting result is that Many Eyes users are finding that even official data sets are not free from errors. A visualization of twins—triplets and “higher-order”—births by age of mother revealed that American women in their 50s give birth, on average, to 9 twins! The data comes from the National Center for Health Statistics, one of the main governmental health statistics agencies in the U.S. and a respectable source of information.

The incident reveals something visualization experts already know: visual representations can quickly expose data problems. Other problematic data sets from official agencies have surfaced on

¹ All user names have been anonymized for privacy reasons.

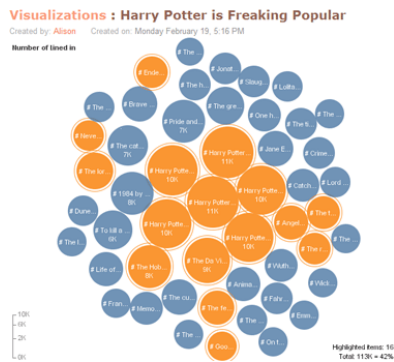


Figure 3: “Harry Potter is Freaking Popular” game. This sequence shows a small series of screenshots of different users’ selections of the books they had read.

Many Eyes. There may be an important role that a public site like this can play in the debate on information available to the public. The name of the site comes from the open-source saying that “many eyes make bugs shallow” and perhaps this adage applies to data as well.

3.2 Sociability – “Harry Potter is Freaking Popular”

One motivation behind Many Eyes was to test the notion that visualization can be a social, playful activity, and in several cases we have seen the emergence of miniature games and lighthearted interaction. Almost a month after the site launched, user Melissa created a bubble chart² (Fig. 3) of the fifty most popular books on LibraryThing, a website that allows users to catalogue their books online. The leading books were all from the Harry Potter series, causing the user to name the visualization *Harry Potter is Freaking Popular*.

The visualizations on Many Eyes all contain a simple pointing mechanism, where a user can highlight a set of items for reference in a comment. One of the authors of this paper used this mechanism to show which books on the list he had read. Soon after, a small game ensued where fourteen other users highlighted the books they had read, often with comments on their own taste in literature. Later, the same highlighting game was played with a visualization entitled *2006 Movies by International Box Office Gross*.

The “what have I read” game required very little from users, due to the built-in highlighting feature. But in other cases users have gone to considerable lengths to perform similar activities. An example is the *Countries I have Visited* game. Users create a new data set with the name of the countries they have visited on one column and the number of times they have been to each one of

these countries on the second column, and upload that data to the site. They then create a world-map where each country is colored based on the number of visits.

3.3 Generating mirrors—personal and collective

The self-revelatory games described in the previous section are part of a broader theme: the creation of personal “mirrors.” It is common for people to use online journals, photo collections, and video sharing sites to reveal aspects of their personal lives online. Therefore, it is not entirely surprising that a trend on Many Eyes is the visualization of personal data.

Users have uploaded data sets that range from collections of personal writing to running and swimming logs to family trees. One user uploaded and visualized a data set of monthly measurements of his weight over the past year. His audience can easily follow the result of a diet that has taken him from his previous 113 kilos (~250 pounds) to a lighter 94 kilos (207 pounds). The user later blogged about his Many Eyes “My Weight” graph, while also reviewing two other Web-based tools that are geared specifically to visually tracking one’s weight throughout the course of a diet: *skinnyr.com* and *traineo.com*.

A different kind of Many Eyes “mirror” depicts not a single person but an entire community. Users have uploaded several data sets with demographics of different virtual worlds. There are a dozen data sets on *Second Life* [11] covering everything from player gender over time to business revenue to the distribution of active users in different countries. Statistics on *Wikipedia* [21] and *World of Warcraft* [22] have also been visualized on the site.

A user on the popular community blog *Metafilter* visualized social network data from that site to Many Eyes. He then posted a short entry about his visualization on the *Metafilter* discussion board:

² Bubble charts are one of the various visualization techniques available on Many Eyes. A bubble chart displays a set of numeric values as circles.

I used IBM's Many Eyes data visualization site to create a network map of my MetaFilter contacts and my contacts' contacts. It got to be pretty big, so I thought I'd post it here.

This post generated a long list of posts from other Metafilter members. Users discussed the visualization itself, the places where each one of them appeared in the graph, the technical challenges of getting this kind of data, and other possible visualization techniques. Here is a small sample of posts:

Wow, even I'm on there. I thought I was only popular on Metachat and in Portland. --posted by Bob

Would make for an interesting meet-up. Everyone seems to have their seating assignments. -- posted by Harry.

I'm either less popular than I give myself credit for. Or I'm a hydrogen molecule. -- posted by Harry

Interesting to see who the most connected people are. Dissappointingly, I only connect to two people. -- posted by Scott.

Here's a tag cloud from the same data. -- posted by Milton.

I feel so....small. -- posted by Mary.

Interesting that I have a lot of people linking to me yet only a few links on the graph. Meaning, I think, that most of those who link to me aren't in this network (monju+I doesn't link to them). Also, that I'm not cool. Also, that I am a failure in social reciprocity since I don't link to most of the people who link to me (if I had, they'd be in this network). -- posted by Ethan.

My dot is larger than I thought it would be. Go me. -- posted by Despina.

I'm trapped in an arm of the Cortex galaxy with a bunch of other nobodys. What a rip off! -- posted by Ntrickster.

Less conspicuous communities have also visualized themselves. NPtech, for instance, is a group of technologists that track information on the web that is tagged "nptech" (nonprofit technologies such as open source software). One of their members used our site to visualize all the tags associated with "nptech" on del.icio.us.com. Soon after the visualization was posted, other Many Eyes users started asking what NPtech stood for. Community members answered the questions by giving a long account of the context for the community and why it was formed in the first place.

3.4 Sending a message

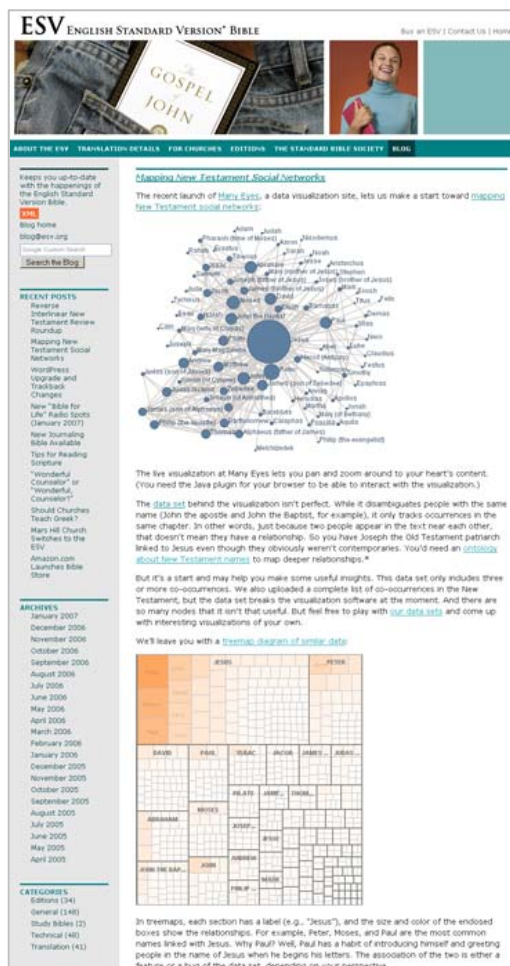
Sometimes visualizations seem to be created as much to communicate a message as to analyze a data set. For example, one user, "Jeremy," visualized the number of people with multiple sclerosis (MS) per country worldwide. Jeremy blogged about this visualization and the entry was picked up a blogger friend of his, who is disabled

with multiple sclerosis. In a heartfelt post, the blogger shows an image of the visualization and says:

Thank you ... for helping us become aware of the fact that there are so many MSers in so many places. We are not alone.

Visitors to the site have created visualizations that advocate positions, that bring solace (as in the MS example), and that illustrate issues of importance to them. One of the distinguishing traits in this class of visualizations is the deep knowledge users usually have of the subject matter they visualize. The intent seems not so much to use visualizations as tools for exploration and discovery as it is to make a statement and to bond around a topic of interest. An interesting example is a series of visualizations relating to the Bible.

Figure 4: New Testament visualizations

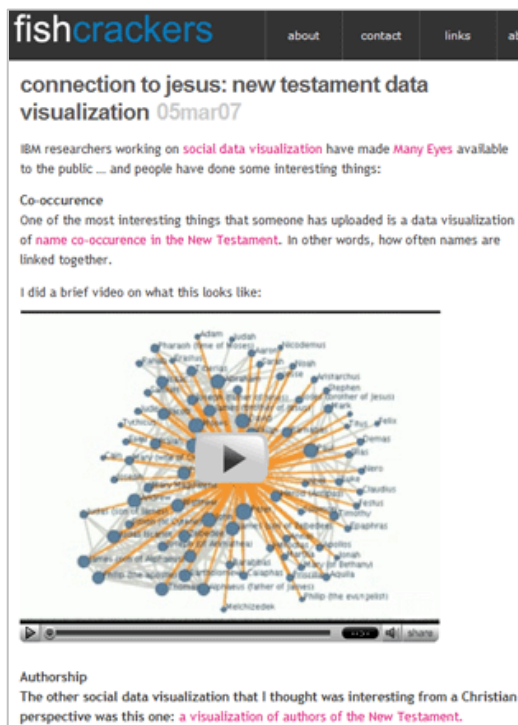


Two days after launch, a user uploaded a data set of name co-occurrences in Biblical verses and visualized it both in a network graph and in a treemap. The user then wrote about it on a

Christian blog that has wide readership (Fig 4). This blog entry received more than a hundred comments and references (many more than other posts we examined on this blog). At first the entry spread over a community of Christian blogs and later reached a wider audience via such high profile blogs as BoingBoing. In fact, a search on technorati.com revealed that there were over 200 links back to the original blog entry (more than twice the number of links back to the well-known New York Times interactive infographic entitled “Faces of the Dead” remembering 3000 fallen US service members during the Iraq war [8]).

The Bible network diagram had a catalyzing effect: it caused the creation of additional visualizations of data about the Bible. One user said on his blog that he was inspired to upload and visualize data on the length and authorship of New Testament documents. Another user made a tag cloud of the first half of the Old Testament. The trend of religious-themed visualizations continues to this day.

Figure 5: Video blog post about Many Eyes



Two aspects to this phenomenon bear mentioning, because they contradict some of our own expectations for the site. First, almost all discussion about the religious visualizations created on Many Eyes is happening offsite. There is a strong community of Christian bloggers that is used to interacting through blog posts and it is no

different when they refer to the visualizations created on Many Eyes. Some users are utilizing the “blog this” button to add images (with links) of the visualizations on their blogs. Other users are more ambitious: one even created a video of himself interacting with the original social network graph of name co-occurrences, looking for nodes that were not directly linked to Jesus (Fig. 5). Given that Many Eyes has features that encourage discussion around Many Eyes visualizations both on and off the site [18], it is interesting that such a high proportion of activity occurred on external domains.

Second, we find it surprising that the most linked visualization on the site—the graph visualization of name co-occurrences—does not reveal any relationships that would be unexpected to knowledgeable Christians. Rather than showing a discovery or novel insight, the display shows viewers what they expect to see: Jesus dominates the graph. He is represented as the biggest dot in the center of the graph with the highest degree of connection to other nodes. The fact that the visualization reinforces a common understanding may help viewers bond with others of the same faith. Perhaps in this context, a fact does not have to be new to be worth communicating.

3.5 Innovative uses of visualizations

Many users came up with creative ways to use visualizations. The fourteen visualization techniques available on Many Eyes were deliberately written to be as simple as possible. Despite this simplicity, we found several instances where our users managed to find unanticipated ways to use these techniques.

One example of such as use is shown in figure 6. The data set relates to a well-known web site, Digg.com, on which users share interesting URLs (termed “digs”) and annotate these URLs with comments. The graph is a hierarchical stacked area chart meant for displaying multiple time series. In this case, however, the creator is showing a “time series” with only two x-axis points, labeled “Digs” and “comments”. The creator has also clicked the “percentage” checkbox at lower right, which normalizes the series on a percentage basis. At first this simply looks like an error—but on close inspection, the visualization turns out to be an interesting way of comparing the number of digs in a topic area to the number of comments. It is apparent, for example, that technology-related topics are discussed less often relative to “world & business” URLs. A brief conversation on the site with the creator confirmed that this effect was intentional.

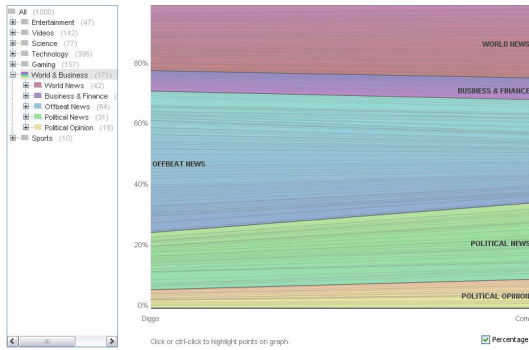


Figure 6: Novel use of Stack Graph for Categories

Several other novel uses of visualization components emerged. One person used the bubble chart to show data on the sizes of the planets, yielding an elegant visual pun since the disks in the visualization resembled planetary silhouettes. The tag cloud was used in a variety of unexpected ways beyond showing word frequencies (or the contents of John’s freezer in Figure 1). One of the most creative was a user who would take two different, related texts (for example *Apocalypse Now* and *Heart of Darkness*), paste them together, and create a tag cloud of the result. As the user explained in a comment:

This is the first in a series of what I'm going to call a 'lit mash' -- two or more texts mashed together and visualized... In the background, I was also thinking of William Burrough's 'cut-up' technique, although this isn't strictly a cut-up ...more of a statistical dicing, I guest.

This example shows how the simple tools on the Many Eyes site can be used for a complex type of personal expression.

4 CATEGORIES OF CONTENT

The examples of section 3 provide evidence for a rich set of activities catalyzed by Many Eyes. To gather a broader set of data on site usage, we collected a large selection of comments on the site (92, representing 25% of all comments, excluding those from members of our lab) as well as 379 data sets (24% of non-lab uploads) and a selection of 36 external blog entries describing Many Eyes. We then had two researchers code these comments, data sets, and blog entries according to preset rubrics, with discrepancies being reconciled between the two after coding.

4.1 Comments

Our data set included 92 comments that did not come from site developers. To classify these comments, we relied the rubric from [6]. We chose to reuse that scheme since it was developed for the same purpose and so that we could

compare directly with the results of the sense.us experiment. The rubric includes labels for observations, questions, hypotheses, links or references to other views, usage tips, socializing or joking, affirmations of other comments (“Yes, that’s right!”), to-dos for future actions (“Can someone find inflation-adjusted data?”), tests of system functionality, data integrity, and site design. Note that comments are allowed to have multiple labels.

Label	Many Eyes %	Sense.us %
Observation	46.3	80.6
Question	15.8	38.1
Hypothesis	11.6	35.5
Data integrity	9.5	15.7
Socializing	11.6	9
System design	11.6	9
Testing	4.2	5.6
Tips	4.2	4.1
To do	4.2	2.6
Affirmation	13.7	1.5

The results are shown in the table above. As with the sense.us experiments, observations and questions are the most common types of comments, although we see fewer of these in Many Eyes than in sense.us. The levels of socializing and talking about the system were about the same. The only category that is much larger in Many Eyes is affirmations.

These results show that data analysis (in the form of observations, questions, and hypotheses) does occur on Many Eyes, but less intensely than in the experiments of sense.us. One explanation for the overall differences is that much of the sense.us usage took place in groups of people who were academically inclined and already knew each other—thus more time was spent on purposeful analysis, while on Many Eyes, where most users are strangers, affirmations play an important role as introductory chit-chat. Another source of the difference may be in the type of data: the sense.us discussions took place around a static collection of statistics that was selected by the experimenters to be amenable to extended analysis, whereas discussions in Many Eyes occur in the context of a constant flow of incoming data sets, some of them quite simple—such as the most popular books in a library or the top grossing movies in a given year.

4.2 Data sets

The fact that the data on Many Eyes is contributed by users makes comparisons with sense.us more difficult, but raises an interesting question as well. What sorts of data do people want to visualize in practice? To explore this issue, we took a sample of 379 randomly selected data sets from the total of 1,895 on the site as of March 2007. Two of the authors inspected these and jointly decided on a set of labels; they then independently coded the data sets with this rubric. Any discrepancies were reconciled between the two coders. The table below shows the results of this labeling.

Label	Percentage
Society (e.g. demographics)	14.0
Economics	12.7
Obscured or anonymized	12.4
Arts and culture	10.8
Web & new media	10.3
Science	10.0
Test data	9.5
Politics	7.4
Technology	6.6
Personal data (e.g. weight loss)	6.3
Religion	5.8
Foreign language	4.7
Sports	4.2
Health	2.4
History	1.3
Education	1.0
Surveys	0.5
Environment	0.5
No label assigned	0.3

These numbers must be interpreted with caution. Some types of data are more easily available than others: for instance, the U.S. government makes demographic and economic information available for free. Furthermore, the users of Many Eyes at this point are hardly a random sample of web surfers due to factors

ranging from the “early-adopter” phenomenon [10] to the type of publicity that site has received—an article in *Nature* [1], for instance, may have led to a disproportionate number of scientists visiting the site.

Nonetheless, a few patterns are worth noting. One is the diversity of topics. The categories are broad but none accounted for more than 14% of the data sets. A second point is the relatively large sizes of categories, such as arts or religion, that are not common topics for visualization research. A third pattern is the large number of data sets that we judged to be obscured in some way—e.g., a network where the nodes were labeled with six-digit numbers, or schools districts referred to as “District A,” “District B,” and so on. The presence of these obscured data sets suggests a desire for privacy.

4.3 Blogs

Inspired by the web-wide conversations that occurred around the NameVoyager [20], Many Eyes was designed to appeal to bloggers. We view the vast net of blogs as an essential part of the social system in which Many Eyes lives. Much of the infrastructure of Many Eyes—from the stable URLs that capture visualization state to the one-click “blog this” feature—was meant to ensure that Many Eyes could be a full player on the Web.

Does the site succeed on these terms? Technorati [12], a leading blog search engine, reports that 1,215 different blogs have linked to the site. For comparison, Technorati finds 1,605 direct links to the NameVoyager [20] and 165 to the SmartMoney Map of the Market [19]. The latter sites are well-known visualizations that have been public for years, so the Many Eyes statistic reflects a relatively high level attention among bloggers.

Of the blog entries that mentioned Many Eyes, we found 36 that were written by users who had uploaded data and created their own visualizations. Most other entries either described the system itself or referred to visualizations that others had created. Because we were interested in understanding more about users’ motivations to *create* visualizations, these entries provide an excellent window into user intent. Of course, this self-selected segment of users is unusual in many ways, so we cannot draw conclusions about motivations of the typical user. At the same time, we probably can draw conclusions related to the segment of the population that is motivated to blog about their creations—a critical subset of the Many Eyes audience due to its role in drawing attention to the site.

Following the same process used to label data sets, we categorized the motivations described by the bloggers. The table below shows the results:

Label	Number of posts
Research	4
Personal Expression	4
Journalism	12
Social interaction	3
Education	2
Analysis	10

A few of these terms merit explanation. “Research” means performing some kind of original investigation, as in the prose style analysis described in section 3. Personal expression includes the “mirror” visualizations described above. A “journalism” blog post is one that is designed to communicate a particular factual message. Analysis indicates that the entry included hypotheses about a data set.

At a high level, the blog post categorization reflects a fairly even division between personal topics, journalism, and analysis. To reiterate the point made in section 3, it is interesting to contrast this diversity of purpose with the strong concentration on analysis that seems to drive much of the research in the information visualization field.

5 IMPLICATIONS FOR RESEARCH AND DESIGN

We see two primary implications of this broad user interest in non-traditional goals for visualization. First, some common methods of assessing visualizations may be missing the point. Studies that measure the time it takes users to find outliers or make comparisons do not necessarily predict all aspects of the value provided by visualizations. Instead, researchers may wish to investigate a technique’s ability to catch and hold a viewer’s attention or to spark discussion. One particularly important aspect of visualizations may be the ability to catalyze conversation and storytelling as in the collective mirrors of section 3.3. These concerns do not rule out quantitative methods or laboratory studies—attention and social activity are at least as amenable to quantification as notions such as “insight” or “discovery.” At the same time, it is also natural to look to qualitative methods (case studies, ethnography) as an important source of information. Understanding non-analytic goals may lead to the reassessment of the value of certain aspects of visualizations—perhaps features such as animation, which are not strictly necessary for analytic purposes, may be seen as more important in the context of these broader uses.

A second potential implication is an inversion of the traditional “let the data speak for themselves” view of statistics and visual analytics. In fact, data visualizations may become more broadly valuable when they can be tailored to express a message or point of view. In other

words, when people have something to say, what they need is a display method flexible enough to let them say it.

Several of the examples of section 3 illustrate this idea. The network visualization showing Jesus at the center of the New Testament at first seems to convey an unsurprising point. Yet this point was seen as well worth repeating by many of the bloggers who discussed it. In the YouTube video that one user made of this visualization, the Many Eyes network diagram’s support for node rearrangement and highlighting was used to support a miniature sermon. A second example is the “Harry Potter” game, which shows how users playfully exploited our highlighting mechanism to subvert the analytical purpose of the visualization.

Designers may therefore wish to augment communicative power of their visualizations. Here is one example of how this might work. The network diagram on Many Eyes has been popular, and many people have taken advantage of the fact that its nodes can be rearranged. Could similar capabilities be applied to other visualizations? The Many Eyes bubble chart, for example, follows a traditional type of visualization design. It takes significant CPU cycles to create an optimized space-filling layout of disks. The algorithm works well, and it never occurred to the designers to allow users to change the positions of the circles. But it might actually be a natural way to let people categorize items or show other variables. This change in perspective—going beyond annotation to allow users to make wholesale modifications to visualizations—seems potentially quite general and fruitful.

6 CONCLUSION AND FUTURE WORK

This paper began with two main questions. First, was the social behavior observed around visualization in [15][20][6] a fluke? Second, when lay users are given the tools to create visualizations, what do they use the tools for?

To address the first question, consider that in [15][20][6] social behavior either came serendipitously or in limited, controlled environments, so it was not clear such behavior could be deliberately fostered on the open web with a set of entirely volunteer users. The observations of this paper indicate that we were in fact able to stimulate a wide array of social activity by following the design hypotheses advanced in those early papers. Some differences did appear between what we saw on Many Eyes and what was reported in [6], with Many Eyes showing less pure analytical behavior and slightly more socializing.

While the long-term behavior on the site remains to be seen and analyzed, these preliminary results add some support to the idea that data visualization can form the core of a wide array of social behavior, both sparking conversation and serving as a mirror or camera for social systems. Aside from the need for longer-term studies, there are a variety of related areas for

future work. One key issue is whether the full scale of the internet is necessary for the phenomena we saw. Could a large corporate intranet support the same diversity of data uploaded, for example? What about a single college class?

Our observations have also provided an initial answer to our second main question. It appears that lay users want to exploit visualization for an extremely diverse set of activities. A significant amount of analysis and sensemaking did occur on the site: this fact is supported both by particular examples such as a graduate student performing analysis for his thesis, and by the overall content coding that shows many examples of observations, hypotheses and questions. At the same time, examples such as the “litmash” show that completely different types of behavior occur as well. The examples we provided are backed up by the content coding results, which show that many non-analytic activities are common.

Again, longer-term observation is clearly warranted. Some of the behavior we have seen could be due to exuberant testing of “cool” technology, and may taper off with time. Furthermore, in exploring both questions an important next step is to talk with our users directly. A natural future study would reach out to users and openly ask about motivations for and reactions to using Many Eyes.

Despite these caveats, it is worth contemplating the fact that users are reinventing visualization technology for unexpected social and personal purposes. A historical precedent suggests some implications: In the early days of the telephone, businessmen formed the primary market for telephone companies [1]. The telephone was promoted as a replacement to the telegraph, allowing business messages to be sent more easily. Before the 1920s, residential sales efforts emphasized the “business” of the household and ways the telephone could help the affluent household manager accomplish her tasks. A “Telephone at Christmas” ad campaign recommended the telephone as an aid in holiday preparations, not as a means for giving season’s greetings. In fact, companies discouraged clients from socializing over the phone, reasoning that these high-tech machines should be used for “serious” purposes. Not until the late 1920s did softer themes appear, linking the telephone to sociability. Social interaction, an industry driver today, was ignored and even resisted by the industry in its early history.

Could a similar irony befall the field of information visualization? Historically researchers have emphasized data exploration, research, and sensemaking; the current excitement around visual analytics continues this tradition. In the past decade, however, we have seen academic explorations of other sorts of applications, ranging from social visualization to ambient information displays. While it is too early to draw firm conclusions from user behavior on Many Eyes,

one might speculate that in the future such “non-analytic visualizations” will be as much of an industry driver as traditional scientific applications.

7 REFERENCES

- [1] Robert Amar, James Eagan, and John Stasko, *Low-Level Components of Analytic Activity in Information Visualization*, In Proc. of IEEE InfoVis '05, pp. 111-117, 2005.
- [2] Butler, D. *Data sharing: the next generation*. In Nature 446, 10-11, published online 28 February 2007. <http://www.nature.com/nature/journal/v446/n7131/full/446010b.html>, retrieved 03-30-2007
- [3] Data360: <http://www.data360.org>, retrieved 03-30-2007
- [4] DataPlace: <http://www.dataplace.org>, retrieved 03-30-2007
- [5] Fisher, C. *America Calling: A Social History of the Telephone to 1940*. University of California Press; Reprint edition, 1994.
- [6] Heer, J., Viégas, F.B., & Wattenberg, M. *Voyagers and Voyeurs: Supporting Asynchronous Collaborative Information Visualization*. In Proc. of CHI, 2007.
- [7] Livny, M., Ramakrishnan, R., Beyer, K., Chen, G., Donjerkovic, D., Lawande, S., Myllymaki, J., & Wenger, K. *DEVis: Integrated Querying and Visual Exploration of Large Datasets*. In Proc. of ACM SIGMOD, May, 1997.
- [8] New York Times: Faces of the Dead. http://www.nytimes.com/ref/us/20061228_3000FACES_TAB1.html, retrieved 03-30-2007.
- [9] Plaisant, C. The Challenge of Information Visualization Evaluation. In Proc. of AVI 2004.
- [10] Rogers, E. Diffusion of innovations. Free Press, 1995.
- [11] Second Life: <http://secondlife.com>, retrieved 03-30-2007
- [12] Spotfire DecisionSite Posters: http://www.spotfire.com/products/decisionsite_posters.cfm, retrieved 03-30-2007
- [13] Swivel: <http://www.swivel.com>, retrieved 03-30-2007
- [14] Technorati: <http://technorati.com>, retrieved 03-30-2007
- [15] Viégas, F., boyd, d., Nguyen, D., Potter, J. & Donath, J. *Digital Artifacts for Remembering and Storytelling: PostHistory and Social Network Fragments*. In Proc. of HICSS-37, 2004.
- [16] Viégas, F., & Smith, M. Newsgroup Crowds and Authorlines: Visualizing the Activity of Individuals in Conversational Cyberspaces. Proc. of 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. of IEEE InfoVis 2007.
- [19] Wattenberg, M. Visualizing the Stock Market. In Proc. of CHI 1999.
- [20] Wattenberg, M. *Baby Names, Visualization, and Social Data Analysis*. In Proc of InfoVis 2005.
- [21] Wikipedia: <http://www.wikipedia.org>, retrieved 03-30-2007
- [22] World of Warcraft: <http://www.worldofwarcraft.com>, retrieved 03-30-2007