Masses of Visualizations: An Analysis of Usage Patterns on Many Eyes and Tableau Public

Category: Research

Abstract—We analyze usage patterns of Many Eyes and Tableau Public, two popular Web-based visualization systems. Between them, they have been used by tens of thousands of authors to create hundreds of thousands of visualizations, yet we know very little about who uses them and how. While visualization is common in news media, and data visualizations are consumed by millions of readers and viewers, it is not obvious that many people will want to or be able to create their own visualizations.

Our study sheds light on this topic through a detailed comparison of the authors and the data views they create. We look at the sizes of the datasets used, the types and complexities of data views, and user retention between the systems. There are considerable differences in the types of views created, many of them caused by design choices the systems are based on. We also consider features only present in one of the systems that allow for more complex multi-view visualizations and the use of multiple datasets to gauge the potential for more sophisticated uses of these systems. The data from these systems indicate the general public has considerable capacity to author rich visualizations of data.

1 Introduction

Web-based visualization sites have been available for a few years. These systems let people upload data, create visualizations, and share them. The most prominent, Many Eyes, started in early 2007, amid considerable attention from both the academic community and the media. There is little information, however, how and how much these systems are actually used, whether use is increasing, etc. Early studies of Many Eyes [5] indicated a significant uptake, as well as collaboration between users; but there have been no follow-up studies on usage, nor have there been comparable studies of other web-based or web-centric visualization systems.

Shortly before Many Eyes, in December 2006, Swivel.com was launched. Swivel was much simpler and less academically ambitious than Many Eyes, but run as a start-up rather than an experiment. It shut down in summer 2010, casting doubt on whether there was a market for web-based data visualization.

Yet there is clearly broad interest in visualization. *The New York Times*, *The Washington Post*, *The Guardian*, and other news media are not only increasingly using data visualization as part of news stories, but also experimenting with more sophisticated types of visualization.

Data is also prevalent in today's news stories, be it big data, personal tracking data, data related to current affairs and controversies, etc. It is also increasingly available online, due to open data initiatives and crowd-sourcing efforts like Wikipedia and OpenStreetMap, from government agencies and non-governmental organizations, etc. A study by Google researchers in 2008 found 154 million tables containing relational data on the web [4], and the number has undoubtedly skyrocketed since then.

As a consequence, we expect many users on the web to want to use visualization to make sense of and discuss data. But is this happening? And how are they using the systems available to them?

To answer these and other questions, we decided to study Many Eyes and Tableau Public. Many Eyes is the oldest system that is still around, and is also the only one that has been studied by the visualization community so far. Tableau Public allows users to download and use the Tableau Desktop application for free (with some limitations, see Section 2.3). While Tableau provides more powerful features than Many Eyes, it also presents a much more complex user interface and requires more experimentation and learning.

Our goal is to gain an understanding of the size of these systems, their users' behavior and loyalty, and to test whether users stick to simple and standard chart types, or whether they are able to use what we consider advanced features (such as non-standard views, multiple views, and interaction). We hope that our results are not only of academic interest in the current state of these systems, but will provide useful guidance for further developments of existing systems as well as any future ones.

1.1 Research Questions

Our study is focused on three topics: users, visualization and the data sets they are based on, and advanced features.

We lack demographic and other information about users, but we can analyze their behavior as far as publishing visualizations and data sets are concerned. In particular, we ask the following questions (Section 3.1):

- *How is the number of authors growing over time?*
- *How productive are users?*
- *How well do both systems retain authors?*

Turning to the core, visualizations and data sets, we ask a number of questions about the data people use and the visualizations they create (Section 3.2):

- *How many rows of data do people work with?*
- *What dimensionality does the data have?*
- *How many of the data dimensions are used in visualizations?*
- *What types of data do people want to visualize?*
- *What visualization techniques do they use?*

Finally, we consider advanced features that include multiple data sets and/or multiple views, either as small multiples or as multi-view dashboards (Sections 3.3 and 3.4):

- *Do users use multiple data sets when they can?*
- *How do they join/blend data sets?*
- *Do users use multiple views when they can?*
- *Can users construct interaction links between multiple views?*

1.2 Background and Method

Data was collected from Many Eyes and Tableau Public up to December 31, 2012 (Table 1). For Many Eyes, the data thus spans 24 quarters or six years: Q1/2007 through Q4/2012. For Tableau Public, the collected data spans 12 quarters (three years): Q1/2010 through Q4/2012.

The data contains 46,048 Many Eyes user accounts and 24,563 accounts from Tableau Public (only counting users who have published at least one visualization or data set).

For Tableau Public, each workbook specifies the data sources analyzed (including all of the schema metadata), the types of visualizations produced, and all of the specific VizQL definitions [12] that produce each visualization. For Many Eyes, that data was collected from the visualization types used as well as a heuristics-based classification of data into data types.

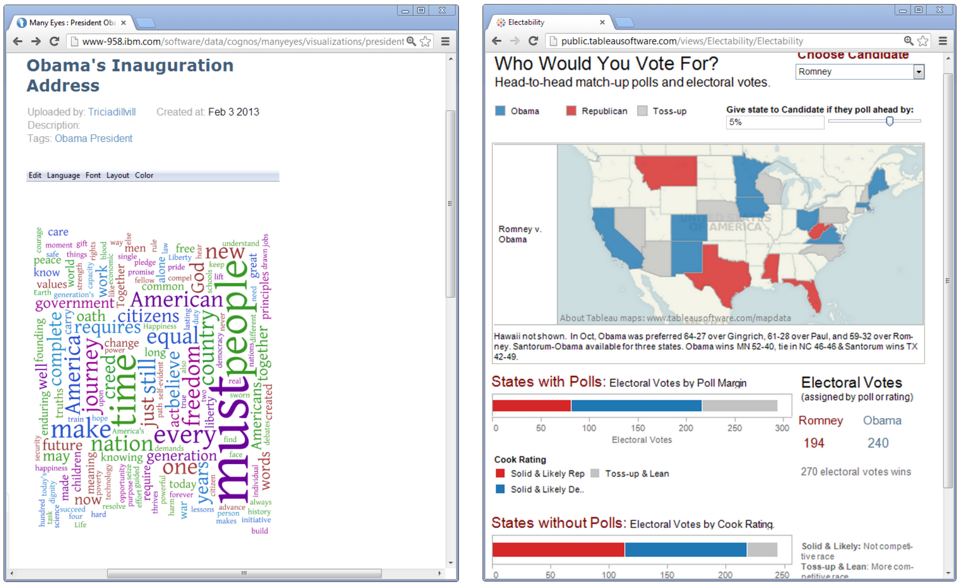

Fig. 1. A common visualization type on Many Eyes is the word cloud (left); complex multi-view dashboards are popular on Tableau Public (right).

System	Start Date	# Visualizations	# Workbooks	# Datasets	Users
Many Eyes	January 1, 2007	149,395 (3.2/user)	n/a	358,880 (7.8/user)	46,048
Tableau Public	February 10, 2010	269,609 (11/user)	73,404 (3/user)	107,596 (4.4/user)	24,563

Table 1. Summary of the collected data from Many Eyes and Tableau Public, from each system's inception until December 31, 2012.

2 RELATED WORK

In 2007 and 2008, *social data visualization* and *visualization for the masses* were assumed to play a big part in the future. Even though there is an ever-increasing number of small start-ups who are providing simple visualization tools on the web today, academic interest in the topic seems to have died down somewhat, at least in visualization.

2.1 Visualization for the Masses

In the wake of highly successful "web 2.0" websites like YouTube, the idea of socially-driven visualization websites [7] spawned a number of experiments, both from large, established enterprises (like IBM's Many Eyes) and small start-ups (like the now-defunct Swivel). Similar to YouTube, the goal was to enable anybody to create and publish visualizations, embed them in blogs, comment, and collaborate by sharing data and creating various views of it.

Collaboration was a driving force behind Many Eyes and also similar systems like the short-lived experiment sense.us [6]. Later work extended the idea of commenting to more structured collaboration for sense-making [19].

Massive online systems lend themselves to experiments, first and foremost about their core functionality and purpose, collaboration and communication [5, 16]. But other experiments included a large study of users laying out small graphs to inform the design of graph drawing algorithms [13].

Crowd-sourcing and citizen science experiments have a longer history in the sciences, such as with the Pathfinder experiment [8]. Both sites studied in this paper rely on users organically creating content, though this has been called into question recently; a more goal-oriented approach creates more and higher-quality responses [18].

2.2 Many Eyes

Many Eyes [2, 17] is a Web-based visualization service that allows users to upload datasets and create visualizations. Unlike Tableau Public, all visualizations are created and published directly through a Web browser. The site was launched in early 2007 by IBM's *Visual Communication Lab* as the first online service that provided ways to not only create static charts, but interactive visualizations that could easily be embedded in blogs and other websites. While both systems share many of the same view types (*i.e.,* bar, line, text, pie, area, scatter, and maps), Many Eyes includes a number of unique techniques that are not available in any other software. In particular, Many Eyes' text views – including word clouds [15], phrase nets [14], and word trees – let users experiment with text data in ways that are still unmatched in most other visualization tools and services.

In addition to the visualization tools themselves, Many Eyes also pioneered the notion of social visualization. The typical Web 2.0 feature of leaving a comment has the added twist that it also contains a live thumbnail reflecting the configuration of the visualization the user was looking at when writing the comment. Users can also browse existing data sets and visualizations and create new ones from what others have uploaded. Many users would thus benefit from the work of somebody scraping or otherwise collecting data.

Other than the original papers by the people behind Many Eyes [5], we are aware of only one other study that looked into visualization activity on the site. *The Guardian* published an informal analysis of Many Eyes in April 2012 [1]. They studied the provenance of the data sources, and reported that the US Census Bureau was one of the most widely used sources. They also presented the most common topic tags for visualizations, most active users, and the number of data sets uploaded per user.

2.3 Tableau Public

Tableau Public [3] is a Web-based visualization platform that launched in February of 2010 (Figure 1(right)). In contrast to Many Eyes, visualization views are created in (currently Windows-only) Desktop client and then published to the web. Tableau Public is a variation of the commercial Tableau Desktop, with the following restrictions: visualizations are limited to 100,000 rows of data, accounts are limited to 50 MB of storage, and content can only be saved by publishing to the Web-facing Tableau Public servers.

Similar to Many Eyes, all content published on Tableau Public can be downloaded by anybody, including the data and the workbook containing all visualization definitions. Visualizations can also be embedded on other websites or shared through social media or email. Tableau Public, however, was designed to have a low "author to consumer" ratio whereas Many Eyes focused more on collaboration and conversation between author and viewer. As of late February 2013, Tableau Public visualizations have been viewed over 100 million times.

Tableau Public allows for more flexibility in the creation of visualization, though it lacks some of the visualization types that Many Eyes has (in particular ones for text visualization). Interaction is generally richer, with control over mouse-over tooltips, selection, etc. It is also possible to build multiple-view *dashboards* that can have actions between the views to filter or highlight data based on user interaction.

2.4 Terminology

Many Eyes and Tableau Public differ in their terminology and the way data and visualizations are organized.

Many Eyes treats data sets and visualizations as independent units: users can publish data without creating visualizations, and create visualizations from existing data already on the site (their own or others') without the need to upload first.

Tableau Public, on the other hand, packages data and visualization definitions into *workbooks*. Workbooks typically contain multiple *worksheets* that each contain one type of visualization. Worksheets can be combined into multi-view *dashboards* that can also include interaction (highlighting, filtering) between the individual views.

Workbooks can include multiple data sets, and individual visualizations can be created with a single data source or by joining multiple data sources together; the latter is referred to as *blending* [10]. The data blending feature lets users join data on the fly from multiple heterogeneous sources without having write a query or specify a data schema with dependencies. A user authors a visualization by selecting the columns from an initial (*primary*) data source which establishes the context for subsequent blending operations in that visualization. Data blending happens when the user drags in fields from a different data source, known as a *secondary* data source. Additionally, the visualization can be further modified by, for example, adding more secondary data sources or drilling down to finer-grained details.

3 Analysis Results

We present the key results of our analysis, organized around our four core questions about overall workload and author behavior, single-dataset analytics, multi-dataset analytics, and and multi-view visualizations.

3.1 Author Behavior

In the following, authors are defined as users who have published at least one data set or visualization. Due to the nature of the data collection, users who never publish anything are not included.

3.1.1 New Authors

To better understand the long-term behavior of authors on these systems, in this section we first answer the question, *what is the growth of new authors until the end of 2012?* Since its inception in January 2007, Many Eyes, has grown to over 46,000 authors who have published over 358,000 data sources and more than 149,000 visualizations. For Tableau Public, its user-base includes 24,500 authors who have contributed over 73,000 workbooks, 107,500 datasets, and 269,000 visualizations (Table 1).

Figure 2 shows the cumulative quarterly growth of new authors on Many Eyes and Tableau Public. Both systems are growing steadily and at a slightly increasing rate of change (≈7,600 new accounts per system per year). Many Eyes has been available for three years longer than Tableau Public, and has grown to 46,048 accounts over the course of 24 quarters. By the end of its 12th quarter (Q4 2012), Tableau Public has grown to 24,563 unique user accounts. In contrast, after 12 quarters, Many Eyes had 13,962 users. These systems thus have moderate numbers of users today, but their popularity is continuing to grow by thousands each year.

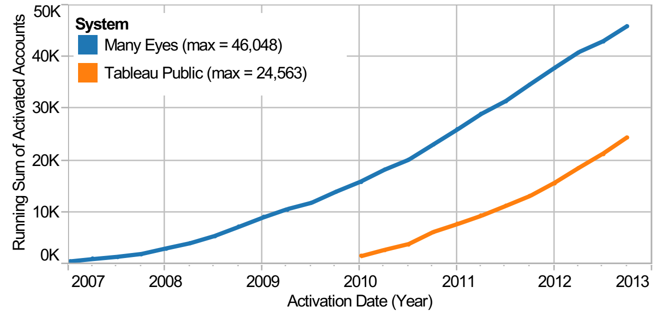

Fig. 2. Many Eyes and Tableau Public activated user accounts. This graph shows the running count of active user accounts since the inception of Many Eyes (Q1, 2007) and Tableau Public (Q1, 2010). By the end of Q4, 2012 (total time span of 6 years for Many Eyes and 3 years for Tableau Public), Many Eyes had 46,048 user accounts and Tableau Public had 24,563.

3.1.2 Author Productivity

How productive are authors in publishing content? Two types of content can be published to Many Eyes and Tableau Public: data sources and visualizations. Figure 3 shows the Probability Distribution Functions (PDFs) of the number of published data sources (top) and visualizations (bottom) per author on Many Eyes and Tableau Public. The publication trends for authors on both systems are quite similar:

	Number of Data Sources Published				
	1	≤2	≤3	≤4	≤5
Many Eyes	44%	65%	76%	83%	86%
Tabeau Public	45%	63%	73%	79%	83%

	Number of Visualizations Published				
	1	≤2	≤3	≤4	≤5
Many Eyes	52%	72%	82%	87%	90%
Tabeau Public	53%	71%	80%	85%	88%

Overall these statistics reveal that most authors on Many Eyes and Tableau Public publish only a few data sources and visualizations. Furthermore, the long tail starting around 14 data sources shows that while these most prolific authors are the minority (with 2–5% representation), their contributions are quite varied (ranging up to 555 data sources for Many Eyes and 715 for Tableau Public). Similarly, for visualizations published, there is a long tail of the remaining top 10% contributing authors. The number of workbooks (visualizations for Many Eyes) published for these authors range from six to 553 for Tableau Public and five to 13,284 for Many Eyes.

The latter number is an outlier caused by the fact that Many Eyes allows users to create visualizations without logging in, so the most prolific user on that system is *Anonymous*. The second-most productive user has only created just over 1,000 visualizations, with the number rapidly decreasing in a typical power-law distribution from there. Among the top 20 users, we find only one member of the Many Eyes team. Tableau Public does not have the notion of anonymous users and thus also does not have a clear outlier like Many Eyes. It does have more active participation from its own employees though, with four of the top 20 contributors being Tableau employees.

Given these results, we categorize Many Eyes and Tableau Public authors into three main groups based on their publication activity:

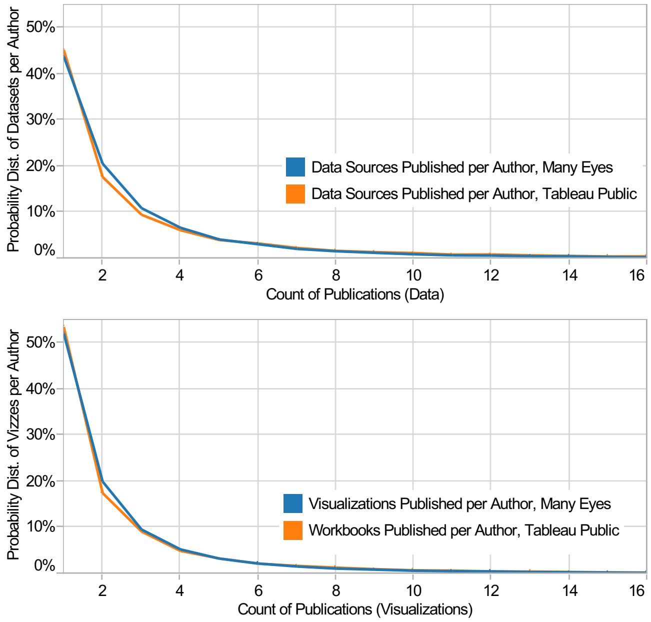

Fig. 3. PDFs of Many Eyes and Tableau Public per author publications of data (top) and visualizations (bottom). Most authors on both systems have one or a small number of data sources and content, while a small number are prolific. The x-axis was cropped for readability, there are authors with hundreds of visualizations and data sets.

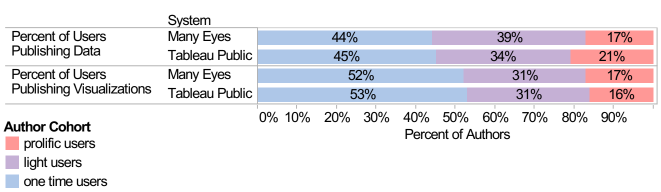

Fig. 4. Percent of authors who publish data and visualizations on Many Eyes and Tableau Public (grouped by author cohort). Authors are divided into three main cohorts: *one-time* users publish one dataset or visualization *light* users publish two to four; *prolific* users publish five or more data sets or visualizations.

1) *one-time* users publish one dataset or visualization; 2) *light* users publish two to four; 3) *prolific* users publish five or more data sets or visualizations (Figure 4). The user-base on Many Eyes and Tableau Public is dominated by one-time and light users.

3.1.3 Author Retention and Churn

How well do both systems retain authors? Figure 6 presents the trends of author retention and churn. We group users into cohorts based on the quarter in which they published their first visualization (workbook on Tableau Public) and track their publications. In comparing the activities of authors on these two systems, we trace each cohort of authors for their first 12 quarters (three years).

Figure 6(b) shows that by the end of the 12th quarter, Many Eyes has 2,100 actively publishing authors and Tableau Public has over 4,000 active accounts. Additionally, over 6,800 new visualizations are published on Many Eyes in its 12th quarter, (12,000 on Tableau Public), as shown in Figure 6(a). Even though the total number of users and visualizations is greater in Tableau Public than Many Eyes, both systems show a strikingly similar pattern in terms of workload distribution between new and returning users in Figure 6(c). In the

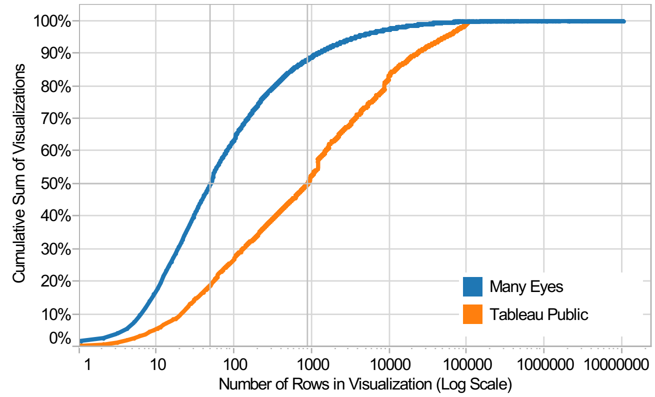

Fig. 5. CDF of number of rows in visualizations on Many Eyes and Tableau Public. 50% of datasets on Many Eyes are 50 rows or less, while that number is just under 900 for Tableau Public.

graph all author accounts are new in the first quarter and thus not returning users. This corresponds to the authors who joined Many Eyes and Tableau Public in the first quarter that the systems were deployed on the Web (*i.e.,* the 2007 Q1 cohort on Many Eyes and 2010 Q1 on Tableau Public). During the second quarter, only 12% of these authors on Many Eyes and 24% on Tableau Public returned to publish. For Many Eyes, we see that by the last (12th) quarter, only 15% are returning authors; Tableau Public, however, has a retention of 37%. Furthermore, these returning authors contributed 25% of the published visualizations that quarter on Many Eyes (51% for Tableau Public). Overall, both systems exhibit the trend of high author turn-over. Looking at the percent of actively publishing accounts by returning authors for each quarter in Figure 6(c), Many Eyes averages 17% and Tableau Public averages 31% (not taking the first quarter into account).

Low retention after initial use is common for free, Web-based services. According to a 2009 Nielsen report [9] only 40% of Twitter users returned to use the site after the first month. However, other websites like MySpace and Facebook achieved retention rates closer to 60%. This result was measured for these other three systems at the same point in their respective user growth curves.

3.1.4 Discussion

Overall, the results thus show a continued growth in users but a low retention rate of these users. The overwhelming majority of users are either "one-time users" or "light" users. A few direct implications of these results are that (1) online visual analytics systems today have a user-base primarily comprised of users with little to no experience. At the same time, (2) while attracting new users to these systems is not a problem (Figure 2), retaining them beyond their first visualization appears to be a critical challenge. There must be other more fundamental causes (perhaps relating to usability or the fact that users tend to not be regular visualization creators) that lead users to abandon the site.

3.2 Single-Dataset Analytics

A potential limitation when using online systems are dataset sizes, because both Many Eyes and Tableau Public impose restrictions on the amount of data that can be used. We wanted to find out what sizes of data users work with, and whether they run into the size limits.

3.2.1 Data Set Size (Number of Rows)

Today's online visual analytics systems are designed for small data. Most of these systems put a bound on the size of datasets that can be processed. On Many Eyes, data sizes are limited to 5MB, while on Tableau Public, each user gets a 50MB account and a visualization can operate on at most 100,000 rows.

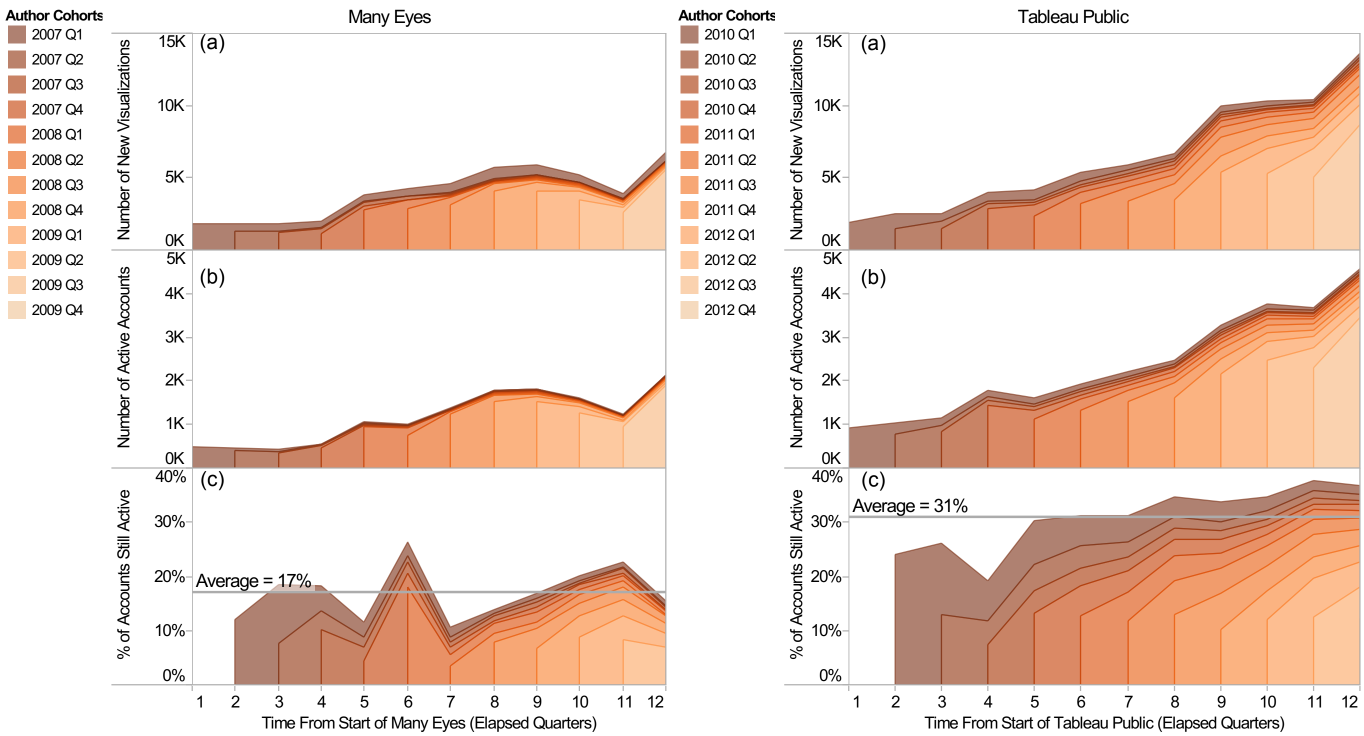

Fig. 6. Many Eyes (left) and Tableau Public (right) author cohorts for the first 12 quarters (3 years). Authors are grouped into cohorts based on the quarter in which they published their first visualization or workbook. Every cohort contains a set of prolific authors, which suggests that (a) overall usage will grow over time as (b) more people use these systems. Tableau Public also has a high rate of prolific authors and a higher rate of author retention than Many Eyes, as shown in (c), which suggests that users value the richer visualizations.

Given these restrictions, we see from the Cumulative Distribution Functions (CDFs) in Figure 5 that the median number of rows in a visualization is low. We also see that there is a significant shift in the curves for these two systems indicating a greater demand for authors on Tableau Public to create visualizations with larger data.

	Number of Rows in Visualizations				
System	≤100	≤1K	≤10K	≤50K	≤100K
Many Eyes	63%	90%	98%	99%	100%
Tabeau Public	28%	53%	84%	95%	100%

Tableau Public also offers a paid tier, Tableau Public Premium, which allows a small number of accounts to go beyond the 100,000 rows limit. These accounts (along with some accounts on Many Eyes) visualize more than an order of magnitude more data, which seems to imply the need for the online visualization of bigger data too.

3.2.2 Data Set Dimensionality

Just like with the number of rows, there is also a large variation in the number of data columns (Figure 8).

	Number of Columns in Data Source				
System	≤2	≤10	≤20	≤100	≤300
Many Eyes	49%	84%	93%	99%	100%
Tabeau Public	2%	28%	52%	90%	99%

Clearly, data sets are not only larger but also contain many more dimensions on Tableau Public than Many Eyes. This is presumably due to the fact that individual visualizations on Many Eyes are typically limited to a small number of dimensions that can be shown at the same time, while Tableau Public allows users to build complex multi-view dashboards. Both systems let the user pick the dimensions to be displayed from the available ones for exploration, however.

3.2.3 View Dimensionality on Tableau Public

Since the data sets uploaded to Tableau Public tend to be multi-dimensional, is it also the case that the visualizations are multi-dimensional? Figure 12 shows the breakdown of data columns used versus available in visualizations with a single data source versus multiple (joined) data sources on Tableau Public (no equivalent information was available for Many Eyes). First, we see that 52% of visualizations with a single data source use at most 3 columns and 90% use at most 6. As expected, the distribution of the columns available is much broader, indicating that there are many more columns available that are not being leveraged by the visualization. For example, 50% of single data sources contain 25 or more columns.

Table 2 shows the breakdown of the most common visualization types used for a given number of columns. The values denoted with a '$\star$' in Table 2 show that a second visualization type was within 5% from the top choice for that given number of columns. For single data sources, we see that the text table is the most common type when there is only one data column present in the visualization. As the number of columns increases, we see a shift in visualization techniques used: bar views become the dominant technique for 2–4 columns and maps are the most popular for 5–8 columns. This behavior is not too surprising since map views have a minimum requirement of two geographic dimensions (*i.e.,* latitude and longitude).

3.2.4 Visualization Types

We study the visualization types most commonly used in both systems. Figure 7 shows the results for Many Eyes (left) and Tableau Public (right). We first focus on the visualizations that are common between the two systems: bar, line, map, pie, and scatter. Our first observation is that bar, map, pie, and area views have the same relative order in both systems. On Tableau Public, for example, there are over three times as many bars (38%) than maps (10%), three times as many maps than pies (3%), and three times as many pies than area views (1%).

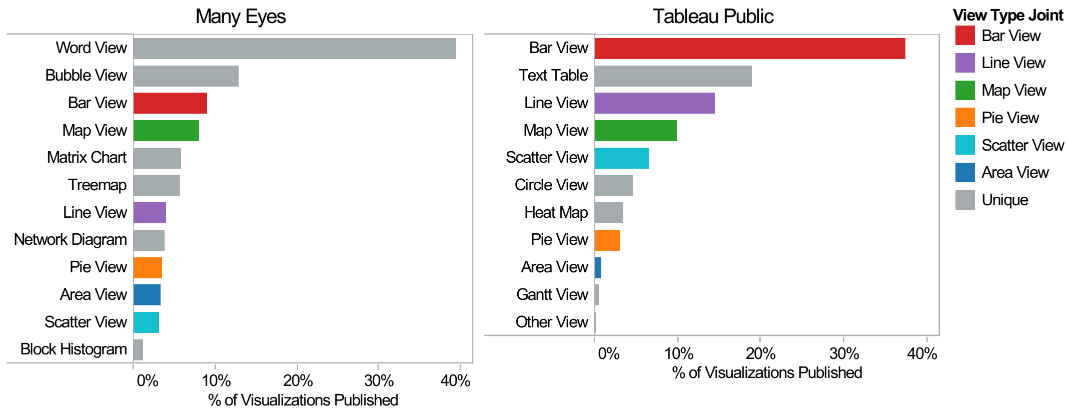

Fig. 7. Common visualization types for single datasets on Many Eyes (left) and Tableau Public (right). Many Eyes is dominated by text visualizations, while bar charts are the most popular type on Tableau Public.

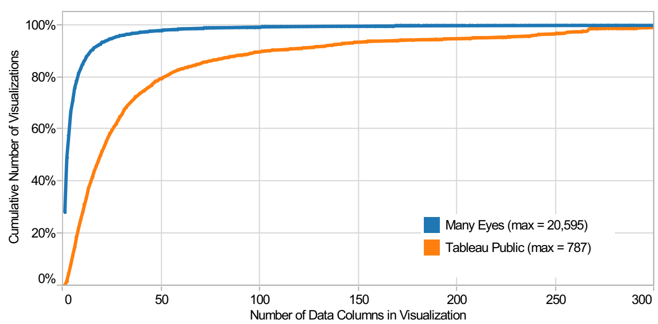

Fig. 8. CDFs of the number of columns in data sets on Many Eyes and Tableau Public. Almost 50% of data sets on Many Eyes have one or two dimensions, while that number is 25 for Tableau Public.

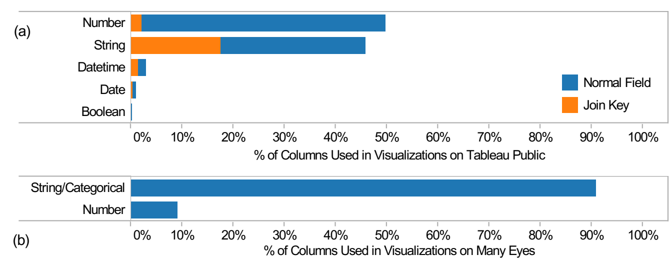

Fig. 9. Data types of columns in visualizations on Tableau Public (a) and Many Eyes (b). Tableau Public is split between numbers and strings, and word-oriented Many Eyes is heavily skewed toward strings.

On Many Eyes, bar views (9%) are almost as frequent as maps (8%) and maps are twice as likely to be found than pie (4%) or area views (3%). Overall, the most common visualization type that exists in both systems is the bar view (38% on Tableau Public and 9% on Many Eyes).

The right half of Figure 7 shows that the most common visualization techniques with a single data source on Tableau Public are the bar view (38%), text table (18%), and line view (14%). This result is consistent with Table 2, in which the bar and text table dominate for visualizations containing 1 to 4 columns, and Figure 8 where 86% of visualizations use 5 or fewer columns. For Many Eyes the most common ones are the word view (40%), bubble view (13%) and bar view (9%).

One of the main differences between these systems worth noting is with regard to text data (Figure 7, left). The large number of word views is due to the variety and quality of text visualization views on Many Eyes, most which are not available anywhere else (Tableau Public's text table is just a table, unlike the rich interactive text views on Many Eyes). Similarly, bubble views are attractive but also rather uncommon in visualization and spreadsheet software.

We thus see consistent results for the two most common visualization types used on Tableau Public; the bar view and text table are the most common. Many Eyes, with its stronger focus on text data, has more popular word views than any other type. Hence some types of visualizations are clearly preferred by users over other types of visualizations but there is room for innovative and specialized visualizations.

3.2.5 Data Types

We next consider whether the most appropriate type of visualization depends on the visualized data. For this, we look at the most common visualizations used when viewing attributes with different types. First, in Figure 9(a), we see that `Number` (51%) and `String` (44%) are the most common data types in visualizations of a single dataset on Tableau Public. It is interesting that their use is fairly balanced, while intuition would indicate that numbers might be more common due to the quantitative nature of business analytics. The `Number` data type includes both integers and reals. Finally, we see fewer specialized types such as `Datetime` and `Date`, which indicates that visualizations of time-based data are less prevalent.

Many Eyes, however, has a skewed distribution of `String/Categorical` types. In Figure 9(b), we see that 91% of columns on Many Eyes are of this type. This finding is consistent with the previous one regarding the dominance of text-based visualizations on Many Eyes.

3.2.6 Data Sets per View on Many Eyes

Many Eyes treats data sets and visualizations as entirely different entities, while Tableau Public packages the data into the workbook. The goal of Many Eyes's approach is to share interesting data sets that many users can build visualizations from.

Interestingly, the number of data sets is much larger than the number of visualizations, and the number of data sets per visualization has increased dramatically over time (Figure 10). While there were about 1.2 data sets per visualization in the first quarter of Many Eyes's existence, that number more than doubled to 3.2 during the last quarter of 2012. A sampling of recent data sets shows that many are uploaded

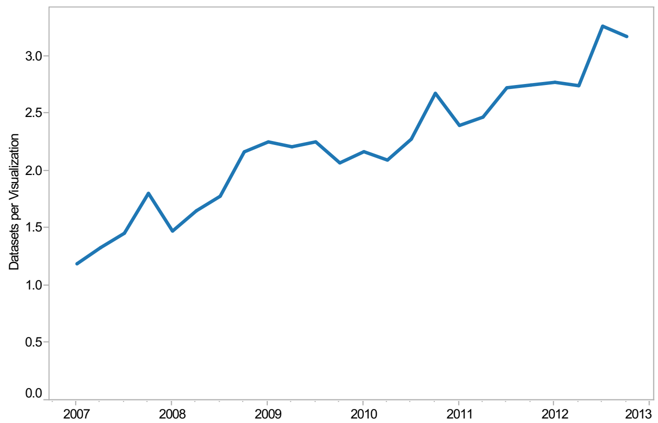

Fig. 10. Number of datasets per visualization on Many Eyes. This number is steadily rising, which means that users are uploading increasing numbers of datasets compared to the number of visualizations created.

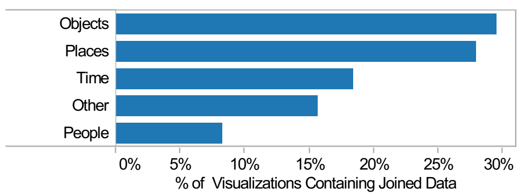

Fig. 11. Common semantic entities of join keys in visualizations with multiple (joined) data sources. Tableau Public authors tend to combine data sets on the same object entities (as in unique identifiers for product codes) or on the same location (as in zip codes on a map).

multiple times, either because of upload errors or because users are not aware that the data is already in the system. It is unclear, however, why that number has increased over time.

3.2.7 Discussion

In summary, most visualizations have modest data sizes, and seem to not be limited by the 100,000 rows restriction, although some users with special privileges visualize datasets with more than one million rows. Additionally, the data sources uploaded to these systems are multi-dimensional. There is thus potential in these systems to support an entirely different class of users with much greater visualization requirements. Furthermore we see that as the number of columns used increases, so does the complexity of the visualization type (*e.g.,* maps require more columns than other types like bar views.) Additionally, visualizations of single datasets tend to use many fewer columns than available. One explanation for this gulf can be drawn from the use of map visualizations in Tableau Public; 62% of such visualizations rely on a Tableau-supplied geocoding database for translating location names into latitude and longitude, since many data sources do not include this necessary context. Finally, visualizations on Tableau Public and Many Eyes contain columns of type `Number` or `String`; the split is very even between these two data types for Tableau Public, and Many Eyes is dominated by `Strings` due to the prevalence of text-based visualizations.

3.3 Integrating Multiple Data Sets

In this section we study the trends in data and visualization on Tableau Public in the context of blending data from multiple data sources. We omit Many Eyes from this section because the platform currently does not support blending data.

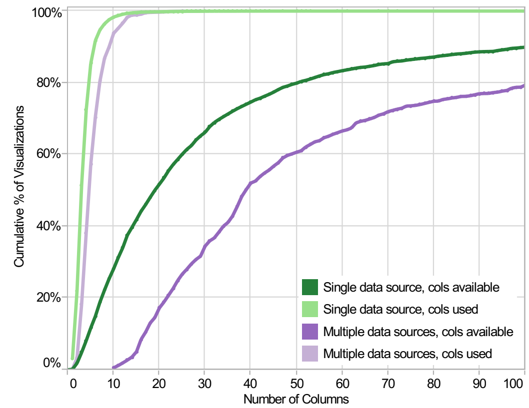

Fig. 12. CDF of the number of columns in visualizations with one vs. multiple (joined) data sources. As expected, the distribution of the columns available is much broader, indicating that there are many more columns available that are not being leveraged by the visualization.

3.3.1 Semantic Entities for Data Blending

On Tableau Public, there are 5,532 visualizations that were created by joining multiple data sets. Of these visualizations, we ask *how do authors combine data sets for their analysis?* To answer this question we manually categorized all of the join keys for the 5,532 visualizations (2%) that have blended data to get a sense of the most popular semantic entities. This process entailed inspecting the column name, data type, and data values of each join key. In the case where the column name was in a foreign language, we used Google Translate on the name and (in some cases) values of that column. If we were still unsure, we opened the workbook to inspect the visualization that was associated with that join key. Figure 11 summarizes the semantic entities of the join keys in five different categories: people, places, time, objects, and other. The people category contains any information pertaining to people, including names and demographics. The places category is restricted to geolocations and other identifying characteristics such as zip codes, regions, states, countries, continents, etc. As expected, the time category refers to dates and date times and objects refer to any physical entity that is not a person, place, or time. Objects consist mainly of opaque identifiers like alphanumeric product codes as well as more well-known, descriptive entities such as "university", "department", or "team". Finally, Figure 11 shows that visualizations of multiple data sources tend to join on objects (30%), places (28%), and time (18%).

3.3.2 Number of Data Columns per Visualization

Figure 12 shows the breakdown of data columns used versus available in visualizations with a single data source versus multiple (joined) data sources on Tableau Public. From this CDF, we see that visualizations with columns from multiple (joined) datasets tend to be more complex than those containing columns from a single dataset. For example, 43% of the blended views contain 5 or more columns, while only 15% of views with columns from a single dataset contain 5 or more attributes. Furthermore, we see a familiar trend as with single data sources: there is a sizable gulf between the number of columns used and the number of columns available in the blended data sources. Additionally, Table 2 shows that, like for single data sources, that visualizations containing a single column tend to be text tables (75%). We also see that the bar view dominates for visualizations containing 2–4 columns and map views for 5–7 columns. This finding is consistent with the distribution of visualization types for single data sources. This behavior is not too surprising since map views have a minimum requirement of two geographic dimensions (*i.e.,* latitude and longitude).

Finally, for visualizations of multiple datasets, the maximum avail-

Table 2. Most common visualization types vs. number of columns in the visualization. For 1–6 columns, the same visualization techniques are used for single and multiple (joined) data sets: text, bar, and map indicating that certain visualization types depend greatly on the dimensionality of the underlying data.

Number of Data Sources	Number of Columns in Visualization 1	2	3	4	5	6	7	8	9	10	11	12	13	14
One	Text (68%)	Bar (53%)	Bar (47%)	Bar (32%)	Map* (27%) Bar	Map (32%)	Map* (27%) Bar	Map* (26%) Bar	Bar* (25%) Map	Line (32%)	Map* (24%) Bar	Circle (34%)	Bar (30%)	Bar (20%)
Multiple	Text (75%)	Bar (48%)	Bar (50%)	Bar (41%)	Map* (22%) Bar	Map (35%)	Map (40%)	Text (46%)	Map (49%)	Map (56%)	Scatter* (38%) Map	Scatter* (36%) Bar	Circle (64%)	Map (42%)

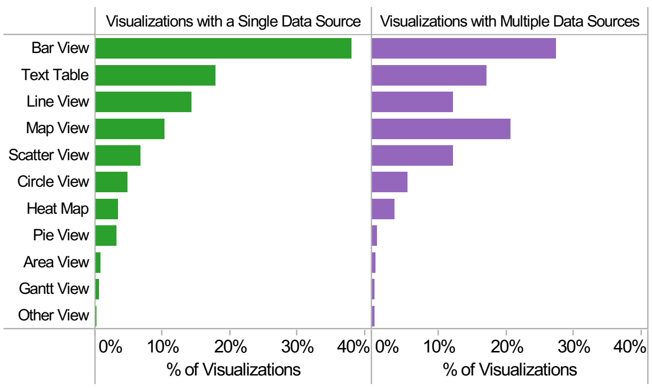

Fig. 13. Common visualization types on Tableau Public for single and multiple (joined) data sources. Single dataset visualizations are more prevalent as bar, text table, and line views. Blended views are found more often as map views and scatter views (which require higher dimensional data).

able columns is 793 and the maximum used columns is 29. Similarly for visualizations of a single dataset, the maximum available columns is 792 and the maximum used columns is 133 (this workbook has a poorly designed dataset that contains one column for each day for 4.5 months).

3.3.3 Data Types in Blended Data

Recall from the previous section that Figure 9(a) shows that `Number` (51%) and `String` (44%) are the most common data types for visualizations with a single data source on Tableau Public. Additionally, the stacked orange bars represent the data types of the join keys, and `String` (18%) and `Number` (3%) types are the most common overall.

3.3.4 View Types for Blended Data

Figure 13 shows that the most common ways to visualize blended data are with a bar view (27%), map view (21%), or text table (17%). Compared to the distribution for single data sets (recall bar views made up 38%, 18% for text tables, and 14% for line views), we see fewer bar views and more maps. This result is consistent with Table 2, in which the text table and map view dominate for visualizations containing 5–10 columns, and Figure 12 where 42% of blended visualizations use 5 or more columns.

Figure 13 also shows the visualization types with higher percentages of blended views: (in order) map views, scatter views, and text tables. Map views are a special case in Tableau Public, because prior to Tableau version 7 (*i.e.,* before January of 2012), filled maps required tricks involving polygon shapes that were placed using blending. This inflates the number of blended views using maps somewhat, though there are also many other use cases where maps can be used as part of blended views. For example, a common blending pattern for maps is to join on a secondary datasource containing detailed latitude/longitude values. Scatter views are generally used for visualizing correlations, and for authors on Tableau Public, this visualization type is useful for showing correlations between measures from two different data sources. Finally, text tables are often used as a trial/debugging tool for checking out the resulting values from the join operation (*e.g.,* how many `Null` values appear?).

3.3.5 Discussion

In summary, data blending occurs primarily by combining multiple attributes about the same uniquely identified entities from different data sources. This type of blending is more common than simply placing multiple entities at the same location or at the same point in time, although the latter two dominate when considered together. This finding is especially interesting for data integration tools. For example, a recent tool provides recommendations of potentially useful data to integrate with a given database [11]. This tool does not consider joining on place or time. It only considers extending semantic entities with additional attributes. With our study, it becomes clear that such a tool would ignore more than half of all blending scenarios. Additionally, blended visualizations tend to be more complex (*i.e.,* use more columns and have more columns available) than unblended ones. However, the distribution of the most common visualization types for a given number of columns is similar for blended visualizations and those using only single datasets. We also see different trends in visualization techniques for those containing blended data versus single data. Blended views tend to be more prevalent in map views and scatter views; these visualization types tend to be more complex (*i.e.,* use more columns).

3.4 Multi-View Visualizations

In contrast to Many Eyes, Tableau Public supports visualizations with multiple views, including small multiples as well as multi-view dashboards. These are typically used for creating more complex displays of multi-dimensional data than would be possible with a single view. These are clearly advanced features, so we were wondering how often they are actually used by authors.

3.4.1 Small Multiples

With small multiple views, the user can compare a quantitative measure across the members of a (categorical) dimension. Figures 6 and 13 are examples of small multiple views. To create a small multiples view, the user has to move a dimension field onto a shelf that already contains a dimension or measure. There is no built-in mechanism to suggest how to do this, and Tableau's *Show Me* feature also does not contain small multiples as an option.

We found that 39% of the visualizations published on Tableau Public are small multiple views.

3.4.2 Multi-View Dashboards

Dashboards are a complementary technique to small multiple views, providing multiple coordinated displays of data. In contrast to small multiples, where the individual views are identical, dashboards can contain any combination of different visualizations. While adding fields to a single data view tends to make the view more complex and harder to work with, multiple coordinated displays can help split such views into separate displays, which makes it easier to follow.

Each display is explicitly linked to the other views and this allows users to simultaneously explore multiple dimensions of a data source. For example, a link can be defined between views to filter or highlight the members of a dimension that are common to the views. This technique can also be used to explore data from multiple, heterogeneous data sources.

Overall, we found that a majority (74%) of Tableau Public visualizations are featured on a dashboard. Of these dashboard views, we found that 62% are actually true multi-view displays (the rest were using the dashboard for special formatting features of a single visualization), and 42% of those contain actions between views (highlighting, filtering, etc.).

3.4.3 Discussion

The number of both small multiple views and dashboards on Tableau Public was surprising to us. Both require a fairly sophisticated user, or at least considerable experimentation for users without training (which is the vast majority of Tableau Public users). But motivated users with burning questions will not only learn to use tools to be able to answer them, but also explore and make use of advanced features.

This result clearly demonstrates the need for, and usefulness of, advanced features like multi-view visualizations in online visualization tools, which are currently missing from the majority of them.

4 Conclusions

In this paper, we studied four primary dimensions of two popular online visual analytics systems: (1) what types of users are leveraging these systems and what are their workloads, (2) what are the trends for doing visual analysis over a single-dataset, (3) how do users analyze data joined from multiple sources, and (4) how often are advanced visualization techniques such as multi-view displays utilized.

We found that such systems today need to effectively support primarily novice users with small datasets. These findings also point to the lack of online, visual analytics tools that would better support users with larger datasets and more sustained data analysis, visualization, and sharing needs.

We also discovered that most visualizations of single-dataset (and multi-datasets) tend to use far fewer columns than available. Since both systems have a large repository of potentially useful data sets, we need tools that can help connect users to other good quality data to aid them in their analysis, especially in the case where additional context is needed in order to take advantage of columns that would otherwise go unused due to their opacity.

The use of advanced features like data blending, multiple views, and actions should be encouraging to the visualization community: when turned into usable tools, these features get picked up by users even when they are provided with little guidance, but are self-motivated to answer questions about data.

References

[1] The Guardian Datablog. `http://gu.com/p/2nntt/tw`, 2011.

[2] Many Eyes. `http://many-eyes.com/`, 2012.

[3] Tableau Public. `http://www.tableaupublic.com/`, 2012.

[4] M. J. Cafarella, A. Halevy, Z. D. Wang, E. Wu, and Y. Zhang. Webtables: Exploring the power of tables on the web. In *Proceedings VLDB*, pages 538–549, 2008.

[5] C. Danis., F. Viegas, and M. W. andJ. Kriss. Your place or mine?: visualization as a community component. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, pages 275–284, 2008.

[6] J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and Voyeurs: Supporting Asynchronous Collaborative Information Visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1029–1038. ACM Press, 2007.

[7] R. Kosara, B. Fitzgerald, H. Rosling, W. Sack, and F. B. Viégas. Panel: The Impact of Social Data Visualization. In *IEEE Visualization Conference Compendium*, pages 128–130. IEEE CS Press, 2007.

[8] K. Luther, S. Counts, K. B. Stecher, A. Hoff, and P. Johns. Pathfinder: an online collaboration environment for citizen scientists. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 239–248, New York, NY, USA, 2009. ACM.

[9] D. Martin. Twitter Quitters Post Roadblock to Long-Term Growth. `http://blog.nielsen.com/nielsenwire/online_mobile/twitter-quitters-post-roadblock-to-long-term-growth/`, 2009.

[10] K. Morton, R. Bunker, J. Mackinlay, R. Morton, and C. Stolte. Dynamic Workload Driven Data Integration in Tableau. In *Proceedings of the ACM SIGMOD Conference*, pages 807–816, 2012.

[11] A. D. Sarma, L. Fang, N. Gupta, A. Halevy, H. Lee, F. Wu, R. Xin, and C. Yu. Finding Related Tables. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, SIGMOD '12, pages 817–828, 2012.

[12] C. Stolte, D. Tang, and P. Hanrahan. Polaris: a system for query, analysis, and visualization of multidimensional databases. *IEEE Transaction on Visualization and Computer Graphics*, 8(1):52–65, 2002.

[13] F. van Ham and B. Rogowitz. Perceptual organization in user-generated graph layouts. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1333–1339, 2008.

[14] F. van Ham, M. Wattenberg, and F. B. Viegas. Mapping text with phrase nets. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1169–1176, 2009.

[15] F. B. Viegas, M. Wattenberg, and J. Feinberg. Participatory visualization with wordle. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1137–1144, 2009.

[16] F. B. Viégas, M. Wattenberg, M. Mckeon, F. V. Ham, and J. Kriss. Harry Potter and the Meat-Filled Freezer: A Case Study of Spontaneous Usage of Visualization Tools. In *Hawaii International Conference on System Sciences*, pages 159–168, 2008.

[17] F. B. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. Many eyes: A site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, 2007.

[18] W. Willett, J. Heer, and M. Agrawala. Strategies for crowdsourcing social data analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 227–236, New York, NY, USA, 2012. ACM.

[19] W. Willett, J. Heer, J. M. Hellerstein, and M. Agrawala. CommentSpace: Structured Support for Collaborative Visual Analysis. In *CHI*, pages 3131–3140, 2011.