CommentSpace: Structured Support for Collaborative Visual Analysis

Wesley Willett, Jeffrey Heer, Maneesh Agrawala, Joe Hellerstein
{willettw,maneesh,hellerstein}@cs.berkeley.edu, jheer@cs.stanford.edu

ABSTRACT
Collaborative visual analysis tools can enhance sensemaking by facilitating social interpretation and parallelization of effort. These systems enable distributed exploration and evidence gathering, allowing many users to pool their effort as they discuss and analyze the data. We explore how adding lightweight tag and link structure to comments can aid this analysis process. We present CommentSpace, a collaborative system in which analysts comment on visualizations and websites and then use tags and links to organize findings and identify others' contributions. In a pair of studies comparing CommentSpace to a system without support for tags and links, we find that a small, fixed vocabulary of tags (*question, hypothesis, to-do*) and links (*evidence-for, evidence-against*) helps analysts more consistently and accurately classify evidence and establish common ground. We also find that managing and incentivizing participation is important for analysts to progress from exploratory analysis to deeper analytical tasks. Finally, we demonstrate that tags and links can help teams complete evidence gathering and synthesis tasks and that organizing comments using tags and links improves analytic results.

Author Keywords
Information visualization, asynchronous collaboration, social data analysis, tagging

INTRODUCTION
Sensemaking is not only a perceptual and cognitive activity, but also a social one; group interpretation and deliberation are essential components of the analysis process. As analysts collaborate, they contribute their own contextual knowledge and extend the work of others [19, 29, 26]. Such collaboration distributes the effort required to examine large data sets and helps analysts develop a shared interpretation of the data. Collaborative sensemaking tools support group exploration and evidence gathering tasks by helping users build on one another's findings and pool their efforts to collectively organize and synthesize them.

Web-based collaborative visual analysis systems – including sense.us [19], Many Eyes [31], and DecisionSite Posters [27] – facilitate such collaboration by allowing analysts to link freeform text comments and graphic annotations to specific views or states of an interactive visualization. However, these systems have largely focused on using comments to share questions and observations in exploratory analysis, while ignoring more complex analytical tasks such as gathering evidence, organizing findings, weighing alternatives, and synthesizing results. They provide only basic tools for navigating and organizing the comments, either via bookmark trails [19] or general-purpose tags/topic hubs [28, 31]. As the number of comments grow, making sense of them can become a daunting task. Interested researchers or late-joining collaborators must read through lengthy discussion streams and manually synthesize results.

We present CommentSpace, a collaborative visual analysis system that enables analysts to annotate visualizations and apply two additional kinds of structure: (1) *tags* that consist of descriptive text attached to comments or views; and (2) *links* that denote relationships between two comments or between a comment and a specific visualization state or view. The resulting structure can help analysts navigate, organize, and synthesize the comments, and move beyond exploration to more complex analytical tasks.

We focus on tags and links that support hypothesis generation and evidence gathering. These have emerged as common tasks in content analyses of previous systems [19, 30] and are prevalent in the sciences as well as in intelligence and business analytics. Specifically we examine how a small, fixed vocabulary of tags (*question, hypothesis, to-do*) and links (*evidence-for, evidence-against*) can help analysts gather and organize new evidence, identify important findings made by others, and synthesize their findings. For example, an analyst may tag a comment as a *question* or a *to-do*, indicating a point of interest or contention. Another analyst might then respond by posting a *hypothesis*, to which other analysts might link additional comments or views, specifying *evidence-for* or *evidence-against* relationships. Visualizing such structure within threaded discussions (Figure 1b) can help analysts identify related comments and views and then connect them into coherent arguments and narratives. Tags and links also make it easier to locate comments that are relevant to particular analysis tasks. For instance, a new analyst might filter the comments by the *question* tag to see a list of unanswered questions and check if she can contribute answers based on her own expertise. Analysts can also use tags and links to organize existing comments and gather scattered

Submitted for review to CHI 2011.

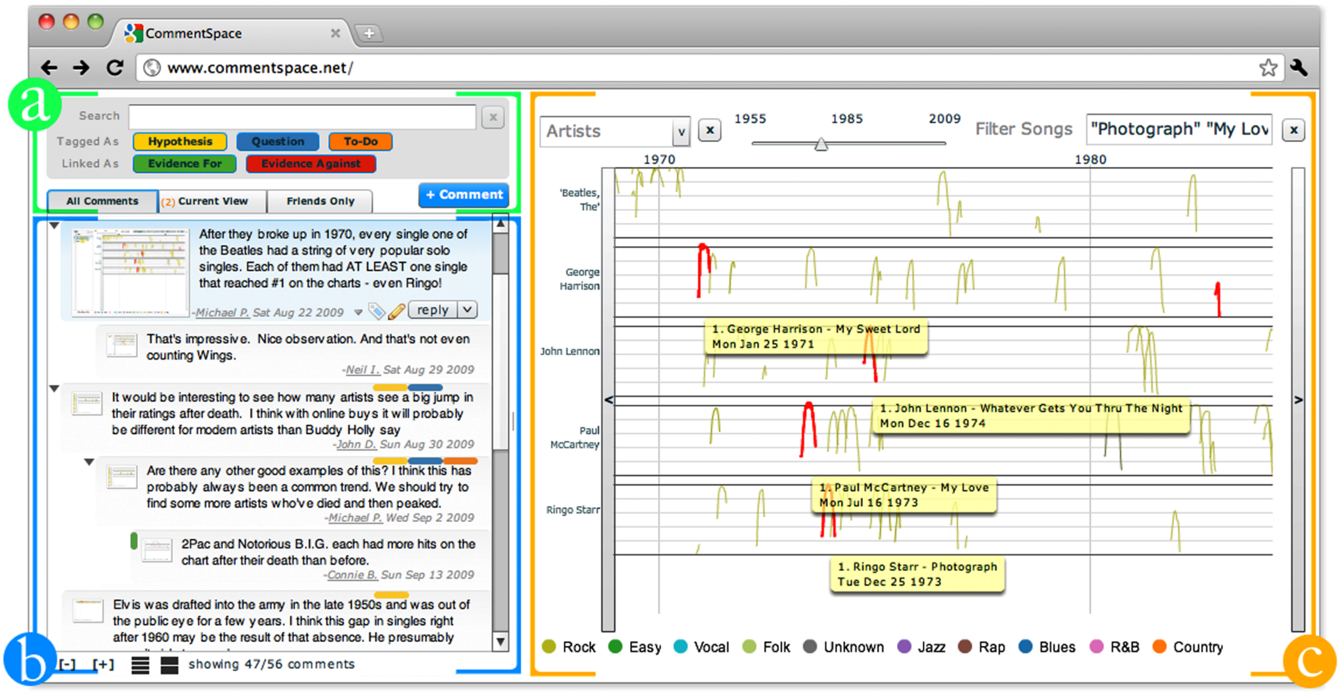


Figure 1. CommentSpace provides a threaded discussion area with search and filtering controls (a, b) alongside an interactive visualization of data from the Billboard Hot 100 chart (c). This chart shows all of the top 100 hits between 1964 and 1980 by members of the Beatles. Color-coded bars on comments indicate tags and links (e.g. *hypothesis, evidence-for*, etc.) and provide lightweight structure that facilitates analysis.

evidence for or against a hypothesis in one location. Such structured organization can help them weigh competing evidence and synthesize related comments.

We designed CommentSpace as a modular software component for authoring, structuring, and navigating text comments. CommentSpace can run in conjunction with any interactive visualization system or website that treats each view of the data as a discrete state. The system must produce a vector of state parameters for each view it generates and be able to render a view from a given state vector. Thus, the state vector serves as a bookmark for returning to a view and for linking a view to comments. Using this mechanism, CommentSpace supports discussions that span a variety of websites and visualization systems.

To better understand how the tag and link structure impacts analysts as they identify, organize, and synthesize evidence, we conducted a pair of user studies and a live deployment in which we compared CommentSpace to a less structured visual analysis tool similar to sense.us [19]. These studies indicate that tags and links help analysts more consistently and accurately classify evidence and establish common ground. We find that users reply to more existing discussions when tags and links are present, suggesting that tag structure encourages analysts to build on existing findings and generate more organized sets of comments. We also demonstrate the importance of managing participation and incentives to help users progress from exploration to deeper analytical tasks. Finally, we show that a team of analysts who use tags and links in a more complex organization and synthesis task produce longer and more detailed analytic results than analysts who do not.

RELATED WORK

Recent years have witnessed a rising interest in social and collaborative technologies, largely driven by increased use of the web as a medium for social interaction. In the area of information visualization, this interest has led to research systems [19, 21], commercial applications [13, 27, 28], and public websites [10, 31] for collaborative visual analysis. Their goal is to enable groups to collectively make sense of data in activities such as ad-hoc exploration, coordinated analysis, dissemination, and follow-up verification.

Just as theories of perception guide the design of visual encoding techniques, we look to theories of social interaction to guide the design of collaborative visual analysis tools [18]. For example, Clark & Brennan's [9] research on *common ground* – the shared understanding needed for successful communication – implies that collaborators are more effective when they can refer to a shared visual environment to ground each other's actions and comments [4, 14]. This observation has led designers of collaborative analysis systems to support synchronous view sharing [1] and asynchronous sharing and reference through bookmarking and graphical annotation of visualization states [13, 19, 20, 31]. In this paper, we investigate additional asynchronous collaboration mechanisms to support visual analysis among teams.

In the context of asynchronous collaboration, work is often broken down into units that can be worked on in parallel. In such situations, collaborators need mechanisms to maintain *awareness* [6, 12] of each other's actions and to *synthesize* individual contributions [2]. In collaborative visual analysis, synthesis often means integrating comments and annotations associated with particular visualization states or data subsets. To reduce the cost of integration, recent systems have provided keyword search of collected comments and tagging of datasets with arbitrary keyword labels [19, 25, 28]. Additional approaches include the creation of "topic hubs" [31] for organizing analyses around topical themes. These systems simplify the process of finding commentary relevant to a topic of interest. To facilitate more consistent results, contributions may also be made more formal; tag vocabularies can be (partially) standardized to provide a shared lexicon for important features of the comments, e.g., to note the presence of a hypothesis or action item [11, 17]. Our approach is similar, in that it uses tags for categorizing comments, but adds a lightweight linking model for organizing comments and visualizations.

A different approach is to use a shared editing (wiki) model rather than a discussion model. For example, the Pathfinder system [21] provides a structured set of "milestones" that can be inserted into wiki text to help scaffold analysis tasks. GeoTime Stories uses a single text story that contains links to specific visualization states as a means to share analysis stories [13]. Many Eyes now also features a "wikified" service that enables visualizations to be embedded in wiki text [22]. These systems integrate contributions via shared editing and the model remains largely informal: contributions can be arbitrary in nature and analysts perform the integration manually in the text. Our system allows analysts to integrate comments without changing their content by authoring semantically meaningful links between them.

Researchers have also investigated highly formalized schemes for integrating analytic work. Argumentation systems [17, 23] typically model hypotheses and evidence in a network structure but provide rigid constraints on the forms of input that analysts can make. These formal models can support computational aggregation and inference, but reduce expressivity and make it more difficult to contribute. Some systems incorporate similar schemes in a more lightweight fashion: for example, the Analyst's Sandbox [33] allows analysts to tag observations as evidence for or against a hypothesis using direct manipulation gestures. Tree Trellis and Table Trellis [8] support aggregation and comparison of linked free-text claims, but are intended largely for introspecting existing sets of claims rather than supporting ongoing analysis. Evidence matrices are a similar approach motivated by the theory of Alternative Competing Hypotheses [3]. Multiple hypotheses constitute the rows of the matrix, while collected evidence constitutes the columns. Similar to argumentation structures, the cells of the matrix are populated with scores representing the degree to which the evidence confirms or disputes the hypothesis. Such formal systems may lead to premature commitment since they can force analysts to think synthetically from the start rather than building on exploratory analysis. In contrast, CommentSpace provides a more lightweight model in which analysts can categorize and connect contributions in an ad hoc fashion, supporting both information foraging and synthesis [26].

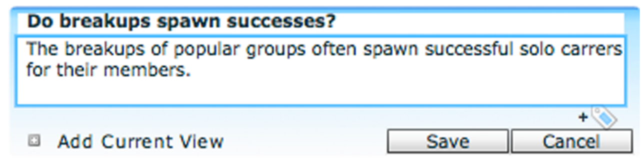
COMMENTSPACE

CommentSpace consists of a threaded, forum-like list of comments along with search and filtering tools (Figure 1a and 1b) paired with an interactive visualization (Figure 1c). The visualization shows data from the Billboard Hot 100 music chart and is based on a design from the New York Times [4]. It depicts the chart rankings of songs by various artists over time. Viewers can observe the rise and fall of individual songs as well as long-term trends in the ranking of artists and genres. They can interactively browse the visualization, hiding and showing artists and filtering to highlight individual songs.

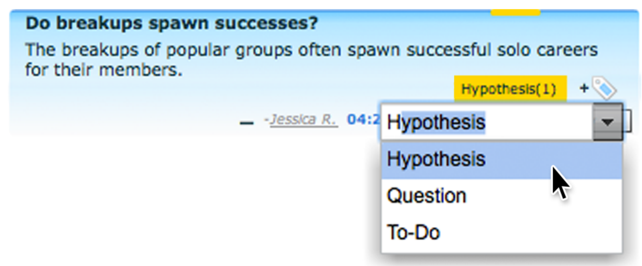
Usage Scenario

To illustrate the use of CommentSpace, we consider a scenario in which a group of analysts are carrying out an analysis task using this visualization.

While reading through existing comments, Jessica wonders if the breakup of popular groups often spawns successful solo careers for their members. She clicks the + *comment* button to post her hypothesis.



She then tags the comment as a *hypothesis* by clicking the blue tagging menu icon on the comment.



Each tag in our vocabulary is associated with a unique color. A yellow tag marker helps analysts visually identify Jessica's hypothesis as they browse and indicates that the comment is a candidate for further evidence or argument. A tally next to the marker (in this case *(1)*) indicates the number of analysts who have applied the same tag to this comment.

CommentSpace also supports links that denote relationships between pairs of comments and between comments and views. Later, a second analyst, Jake, spots Jessica's hypothesis and, intrigued, begins to hunt for supporting evidence. He browses the visualization and builds a view showing the chart success of the former members of California hip-hop group N.W.A. that supports Jessica's claim. He then replies to the original hypothesis, specifying an *evidence-for* relationship, and describes this new view with a comment.

His new observation is threaded into the discussion and appears below the original hypothesis with a small green *evidence-for* link marker to the left of the comment block. Jake adds the current view, so a thumbnail of the current visualization state appears next to the comment. Clicking on this thumbnail loads the view into the visualization panel, allowing users to quickly return to it.

Later, Jessica searches for additional evidence relevant to her hypothesis. Using the search controls at the top of the comment panel (Figure 1a), she filters to show only those comments containing the words "broke up".

By clicking the legend below the search box, she can refine her search further to show, for example, only comments that are flagged as *hypotheses* or *evidence-for*.

Her search finds another observation showing a long string of hits by John, Paul, George, and Ringo after the breakup of the Beatles (also shown in Figure 1). Jessica then drags this observation to her initial comment and links it as *evidence-for* her original hypothesis.

CommentSpace also provides a copy-paste mechanism for linking comments that are distant from one another or visible under different filtering conditions.

The linked comment now appears in the tree below her hypothesis. Unlike standard threaded discussions, such linked comments can appear in multiple places in the comment tree, as the linking makes them part of multiple threads. Thus, the original hypothesis serves as a hub for multiple discussions and observations. Other analysts may reply to it or link in additional comments and views from elsewhere. As the set of comments grows over time, Jessica can quickly return to her original hypothesis comment and filter to see the evidence for and against it. Later, when the analysts begin to organize their findings and synthesize results, they can use tags and links to organize its children into separate chains that contain only the comments that are relevant to their result.

TAGS AND LINKS

CommentSpace introduces a general model in which analysts can tag comments and create links between comments, between visualizations, and between comments and visualizations. Analysts can link comments to multiple visualization states and situate them in not just one, but many threaded discussions. For example, the same comment can appear in both an ongoing discussion and a collection of evidence for a particular claim. When multiple analysts apply the same tag or link to a comment the tally increases, indicating agreement on that classification or relationship.

We focus on exploring the impact of a small, fixed vocabulary of tags and links identified through content analyses in prior collaborative visualization systems [19, 30]. Using a breakdown of the comments generated in these systems as a guide, we selected a minimal set of tags that were common, descriptive, and actionable. The set we selected is tailored towards hypothesis generation and evidence gathering tasks and includes tags for identifying *questions* and *hypotheses* as well as links for indicating *evidence-for* and *evidence-against* a hypothesis. We also include a *to-do* tag for indicating unfinished work. Implicit *reply-to* links are used to maintain the threaded conversation structure and *created-on* relationships are generated between comments and the views they are attached to. We used this small, fixed vocabulary because more flexible free tagging vocabularies can take time to evolve and establish tag meanings [7, 15]. A fixed, task-specific vocabulary also limits analysts' ability to apply tags or links whose meaning is ambiguous or generic and forces them to articulate consistent kinds of structure. Using a fixed vocabulary allowed us to explore the impact of tags and links on particular analysis behaviors without the added complexity of an evolving, community-specific vocabulary.

As in sense.us [19], CommentSpace supports "doubly-linked discussion" whereby authors can follow links between comments and views and only the comments associated with the current view are visible. Doubly-linked discussion can facilitate serendipitous discovery of new comments as users interact with the visualization, but makes it more difficult for discussions to span multiple views. To address this limitation, CommentSpace allows analysts to toggle between a doubly-linked comment panel that shows only comments for the current view and a version that shows all comments. Unlike in sense.us, this master comment list is visible alongside the visualization and users can toggle between the two comment panels using tabs directly above the panel (Figure 1b). This approach encourages discussions that span multiple views and makes it easier to investigate other views without losing track of the current thread.

DESIGN DETAILS

CommentSpace is implemented as an Adobe Flash application that can be embedded in web pages containing interactive visualizations or run as an extension for the Firefox web browser. When embedded with a set of visualizations on a site, CommentSpace provides a browser-independent commenting environment that can be tightly coupled with those particular visualizations. Our examples include visualiza-

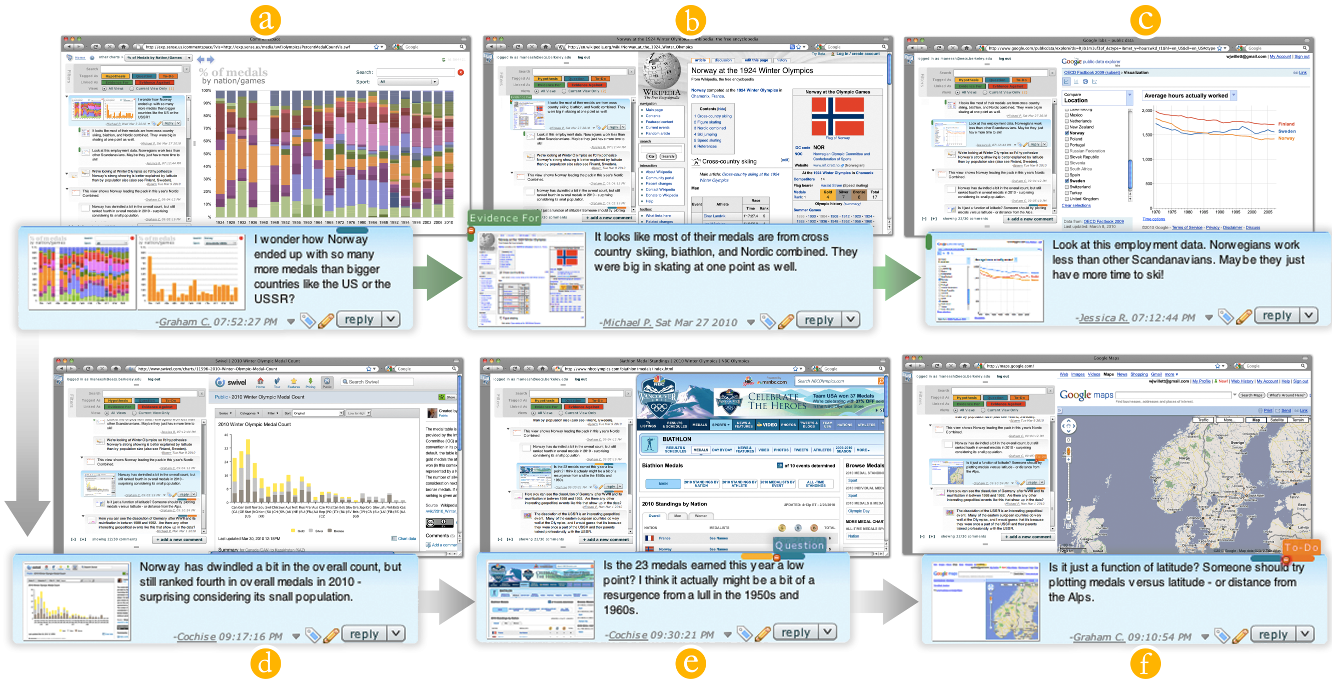


Figure 2. Using the Firefox extension, CommentSpace can facilitate discussion across the web. Here, a discussion begins on (a) a custom Flash visualization of medal counts from the Winter Olympics and incorporates information from (b) Wikipedia, (c) a specific view from Google Public Data Explorer, (d) a chart from swivel.com, (e) an official Olympics webpage, and (f) a view from Google Maps. Replies are shown as grey arrows and *evidence-for* links are illustrated as green arrows.

tions built with the flare toolkit (http://flare.prefuse.org) and Adobe Flex. When used as a Firefox extension, the commenting panel is accessible via a browser sidebar rather than embedded within the page. This version supports linking to and commenting on visualizations as well as *any* view of a web page with a unique URL. Thus, it enables social discussion and evidence gathering across the web and allows collaborators to incorporate information from outside sources in their analyses, as seen in Figure 2. The system is currently deployed alongside a variety of visualizations at *http://www.commentspace.net*.

State Saving and Visualization Support

CommentSpace can be paired with any visualization that implements a simple interface for setting and getting visualization state. The visualization must be able to produce a vector of state parameters for each view it generates, and also render a view from any state vector it produced. These state vectors serve as bookmarks for returning to views or for linking views to comments. Whenever a state change occurs, the visualization must dispatch an event, notifying CommentSpace of the change. Whenever a tag is applied to a comment or a comment is linked to a view, CommentSpace serializes and saves a copy of the state in JavaScript Object Notation (JSON). The CommentSpace web service stores and indexes these state vectors and passes them back to the visualization whenever a state needs to be reloaded.

The browser extension treats URLs as the state vector and thereby makes it possible to link comments to any web page. The extension listens for changes to the current URL (including fragment identifiers - #) and generates a state vector incorporating the URL. This approach is well suited for rich Internet applications like Google Public Data Explorer [16] that provide unique URLs at every visualization state, and makes a compelling argument for designers to build visualizations that provide stateful URLs which update dynamically when the view changes [18]. However, we also include site-specific code to extract state vectors from some useful sites like Google Maps that can generate stateful URLs but don't automatically update the address bar.

Social Sharing and Filtering

As Viégas et al. [30] observed, discussions and continued interactions around visualizations on the web are often more fruitful when they occur within existing communities. To support and encourage analysis within existing groups, CommentSpace also provides several social sharing and filtering tools. Users who log into CommentSpace using a Facebook account can share individual comments and visualization views via their Facebook stream or generate unique URLs to share views by email or IM. They can also filter the comment graph using their Facebook contacts, showing only comments generated by neighbors in their social network.

EVALUATION: STUDIES AND DEPLOYMENT

We conducted two controlled studies and a live deployment to characterize the impact of tags and links in common analysis tasks. In the first study, we tested the impact of tags and links on two specific analysis subtasks: (A) classifying comments left by others and (B) gathering evidence using comments. We also examined usage in a live deployment to assess commenting behavior during exploratory analysis. Finally, we conducted a smaller, qualitative study in which

Figure 3. Versions of the interface seen in the *tag* (left) and *no-tag* (right) conditions. Users in the *tag* condition gain tag filtering controls and see colored tag and link markers on comments.

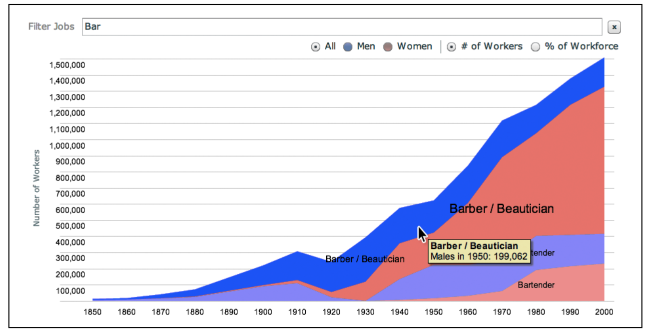


Figure 4. Interactive visualization of occupation data used in tasks A and B. This stack graph shows the size of the U.S. workforce since 1850, broken down by occupation and gender.

teams of analysts used CommentSpace in a complex, multi-stage analysis with exploration, organization, and synthesis phases.

In both studies we compared a version of CommentSpace with tags and links (the *tag* condition) to a version similar to sense.us [19] that provided little support for structuring discussion (the *no-tag* condition). In the *no-tag* condition participants could author new comment threads, reply to existing comments and perform text searches but could not author or view tags and links. In the *tag* condition, participants could add *hypothesis*, *question*, and *to-do* tags along with *evidence-for* and *evidence-against* links. Additionally, *tag* participants could search and filter the comments by their tags and links. Figure 3 shows the commenting interfaces for the two conditions.

Study 1: Tagging and Linking in Analysis Subtasks

We first explored the effect of tags and links on two specific evidence gathering subtasks: (A) classifying comments made by others and (B) authoring comments when gathering evidence.

Method

We recruited 24 paid participants (15 female, 9 male) via mailing lists and a research participation pool. Subjects were university students from a variety of majors. We conducted a between-subjects study in which 12 participants used the *no-tag* interface, while the other 12 used the *tag* interface.

Task A: Identifying and Classifying Comments

Our first task examined how late-joining analysts navigate existing discussions to find comments relevant to a given hypothesis. It also tested whether the presence of tags and links helps users classify those comments more accurately. We anticipated that tags would provide common ground, leading to more consistent categorization of comments, and would make filtering and search more productive. Specifically, we hypothesized that:

H1: Users will identify evidence relevant to a particular claim with greater accuracy when tags and links are present.

H2: Users of a tag-enabled system will use filtering and search tools more extensively to identify relevant evidence.

We gave participants a visualization of U.S. occupation data similar to the one used in sense.us (Figure 4) and a corpus of 181 tagged seed comments drawn from that system [19]. We asked participants to identify as many comments as possible that provided evidence for or against the claim: *Stereotypically male jobs have remained almost entirely male even as women have joined the work force.*

We gave participants 15 minutes to examine and categorize comments that provided evidence for, provided evidence against, or were otherwise related to the claim. Since participants in the *no-tag* condition could not mark comments by tagging them, we asked all participants to write the three-digit identification number of each comment in the appropriate column of a paper worksheet. Subjects were not allowed to add comments, tags, or links during this task.

As a baseline, three of the paper authors (referred to as "experts") also independently coded the comments using the same guidelines as the participants, but with no time limit. Out of 181 comments, the experts identified 9 comments as evidence for the claim, 24 comments as evidence against it, and 19 comments as related but not evidence. The hypothesis was explicitly stated in one comment in the set and, initially, ten comments (both evidence for and against) were linked to it. The total number of comments was large enough that reading every comment individually in the allotted time was difficult.

Results: Classifying Comments

To test hypothesis (H1), we compared the lists of comments classified by each participant in Task A. Because the data are not normally distributed, we report median and median absolute deviation (MAD) and we use the non-parametric Mann-Whitney U-test for significance. Participants classified a similar number of comments in both conditions, (Median=26.5, MAD=4.5) in the *tag* condition and (Median=25, MAD=5) in *no-tag* and there was no significant difference. However, participants in the *tag* condition categorized significantly more ($U=32.5$, $p<0.024$) comments as *evidence-against* (Median=15, MAD=3) than those in *no-tag* (Median=10, MAD=3), showing that tags and links impacted categorization.

To assess the accuracy of users' categorizations, we compared the level of agreement between comment categoriza-

Within Group Agreement

Group	Evidence For	Evidence Against	Related	Unrelated	Average Kappa
(E)xpert	0.586	0.557	0.388	0.839	0.593
(T)ag	0.271	0.405	0.096	0.396	0.292
(N)o-tag	0.271	0.257	0.106	0.344	0.245

Between Group Agreement

Pair	Evidence For	Evidence Against	Related	Unrelated	Average Kappa
E-T	0.395	0.422	0.142	0.452	0.353
E-N	0.380	0.288	0.185	0.428	0.320
T-N	0.282	0.297	0.061	0.368	0.252

Table 1. Average Cohen's kappa values showing within- and between-group agreement for expert, tag, and no tag groups. A kappa of 0 indicates no agreement, while a kappa of 1 indicates perfect agreement.

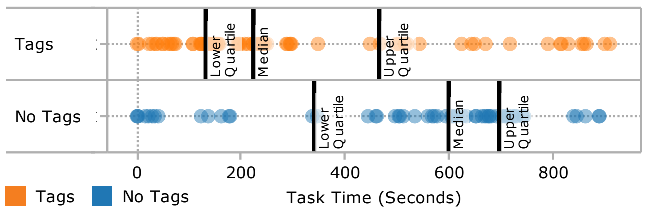


Figure 5. Timing of search and filtering operations in Task 1 (in seconds since the beginning of the task).

tions made by our subjects and those made by the experts. We measured *consistency* (agreement with others in the same condition) and *accuracy* (agreement with the experts) by computing average within- and between-group Cohen's kappa values [24] based on subjects' and experts' categorizations (Table 1). In general, the experts were the most consistent, followed by subjects in the *tag* and then *no-tag* conditions. More importantly, the *tag* group was more accurate - agreeing with the experts more than the *no-tag* group across each of the categories, with the level of agreement on *evidence-against* being the most pronounced. This improvement indicates that tags and links may encourage consensus building and improve shared understanding of comments for late-joining participants.

Results: Filtering and Search
Because they had access to additional tag and link metadata relevant to their task, we hypothesized (H2) that participants in the *tag* condition would filter and search more extensively.

The activity logs show more total search and filtering operations by participants in the *tag* condition (Median=10, MAD=6) than the *no-tag* condition (Median=4, MAD=2), but this difference was not significant (U=46.5, $p=0.0749$). However, participants in the *tag* condition were far more likely to search and filter early in the task. On average, more than half the search and filtering operations in the *tag* condition came in the first four minutes of the task, while participants in the *no-tag* condition took until almost the ten minute mark to complete half of their filtering and search operations (Figure 5). Participants using *tags* searched and filtered significantly earlier than participants in the *no-tag* condition (U=2937, $p<0.0005$).

This data provides a possible explanation for the increased level of consistency and accuracy in the *tag* condition. Because subjects in the *tag* condition filtered and searched earlier, they were more likely to find clearly marked pieces of evidence early on. This evidence may have helped calibrate their categorization, making them more likely to mark pieces of evidence for and against the prompt consistently and accurately. Meanwhile, our observations of activity traces indicate that *no-tag* subjects were more likely to scroll sequentially through the list of comments, marking comments as *evidence-for* even if they were only somewhat related.

Task B: Gathering Evidence as Comments
We designed the second task in Study 1 to explore comment authoring in an evidence-gathering task. We instructed participants to spend 20 minutes locating views and generating comments that provided evidence for or against the claim they investigated in Task A. We told subjects that subsequent users would see their comments when attempting to carry out Task A, and encouraged them to organize their comments so that later users could easily find the relevant ones. All participants began the task with the same set of seed comments they had seen in Task A.

We expected that tags would help users identify unanswered questions and other relevant comments more easily, and that they would encourage users to organize their discussions around those comments. Specifically, we hypothesized that:

H3: Users in the *tag* condition will be more likely to reply to existing threads and, in particular, more likely to reply to comments identified as hypotheses or questions.

Results
Participants generated similar numbers of comments in both the *tag* (Median=12, MAD=4) and *no-tag* (Median=12.5, MAD=4) conditions, but those in the *tag* condition generated significantly more replies (Median=7, MAD=3.5) than those in *no-tag* (Median=2, MAD=1.5) (U = 32, $p=0.0226$). Moreover, a chi-square test shows that participants were significantly more likely to reply to existing discussions when tags were present ($\chi^2(1,308)=27.45$, $p<0.001$), confirming hypothesis (H3). These results suggest that tags and links helped *tag* participants identify and build upon interesting observations and encouraged them to organize their findings.

Live Deployments and Exploratory Analysis
We also conducted two, one-month live deployments of CommentSpace to test its social sharing and filtering features. During these deployments, we paired CommentSpace with ten different interactive Flash visualizations (including those shown in Figures 1, 2, 4, and 6) and made them publicly available at *www.commentspace.net*. While tagging and linking were available during most of the deployment and were explained on a help page, we did not specifically instruct users to apply tags and links during their analysis.

Over the course of deployment, the site received about 6,000 page views from over 850 unique visitors. Of those visitors, 180 created an account on the site or logged in using a Facebook ID; 32 of those users left a total of 123 comments. While the number of registered users and comments is relatively small, the ratio of comments per user (0.68) is higher than for Many Eyes (0.31), the only comparable site for which statistics covering a similar time period after launch were readily available [31].

Most of the analytic behavior reflected in these comments was exploratory. Users authored questions and made observations, but few posited hypotheses or responded to prior comments with pieces of related evidence. The lack of evidence gathering behavior was accompanied by a low level of tagging and linking. During our deployments, users with access to tagging and linking tools authored only 5 tags and a single link.

Based on these experiences in the live deployment as well as earlier pilot studies, we suspect that participants in our open-ended exploratory tasks did not have enough incentive to tag or link comments. Because participants in such tasks have no specific reason to revisit their own comments or those of others, they have little motivation to organize or label comments during exploration. This suggests that more specific tasks and incentives are required to facilitate the transition from exploration to more complex modes of analysis.

Study 2: Exploration, Organization and Synthesis

Neither Study 1 nor the live deployment examined how analysts might use tags and links to synthesize new findings and make decisions. In addition we found that users do not have strong incentives to author tags and links during open-ended exploratory analysis. Heer and Agrawala [18] suggest that managing the division of work and providing appropriate incentives are important considerations in designing collaborative visual analysis systems. We designed a second study to investigate these issues.

In Study 2 participants completed a complex three-phase analysis task, consisting of a directed exploration phase, an explicit organization phase in which participants were encouraged to tag and link their comments as evidence for or against specific hypotheses, and a synthesis phase in which they used the organized comments to make decisions and explain them in writing. We managed each phase more explicitly and gave participants greater incentives than in Study 1 or the live deployments. In particular, we gave participants smaller more specific tasks, especially in the organization phase. We explained how team members would benefit from one another's work as a form of social-psychological incentive and we told participants that the best-written synthesis results would receive an extra monetary reward.

We divided participants into two teams; one team worked together using the full, *tag* version of CommentSpace while the other team used the *no-tag* version. Both teams completed the three phases over the course of 5 days. We monitored how the *tag* participants used tags and links during each of tasks, and compared their written synthesis reports against those of the *no-tag* participants.

Method

We recruited 12 paid participants via campus mailing lists and divided them into *tag* and *no-tag* analysis teams, each with 6 members. We asked teams to carry out a series of exploration, organization and synthesis tasks tasks using an interactive visualization (Figure 6) of estimated return on investment for US college students [5]. We offered participants an extra monetary reward for producing the best-written synthesis reports (as judged by a team of experts). Each team shared a comment workspace populated with 70 seed comments drawn from earlier pilot studies.

In the exploration phase, we instructed participants to explore the visualization and existing discussion and leave comments documenting their findings. We encouraged participants to focus on two general areas of inquiry: "*The relationship between graduation rate, the total cost of attendance, and return on investment*" and "*The distribution of schools from each of the university systems in California.*" We gave participants 36 hours to complete the task, and we instructed each participant to leave *at least* 10 comments.

In the organization phase, we instructed participants in the *tag* condition to organize their team's comments. We asked subjects to organize comments by topic, tag them, and link evidence to related hypotheses. To focus the task, we provided two hypotheses as prompts: "*There is a clear correlation between graduation rate, the total cost of attendance, and return on investment*" and "*There are consistent differences in the graduation rates, tuition, and return on investment between the University of California schools, California State schools, and private universities in California*." We instructed the *tag* participants to add links and tags until they were satisfied with the overall organization of the workspace. Because it was not possible to organize content in the *no-tag* condition, we instead asked *no-tag* participants to spend time reviewing the comments left by their team members. Members of each team carried out the task asynchronously over a 24-hour period. During that time they were free to iterate and build upon one another's work.

Finally, in the synthesis phase, we asked all participants to complete a decision-making task using the visualization and the comments generated by their team. We posed two decision-making tasks based on the earlier prompts. In the first, we asked each subject to "*Produce a ranking of the top schools based on the relationship between graduation rate, the total cost of attendance, and return on investment*." In the sec-

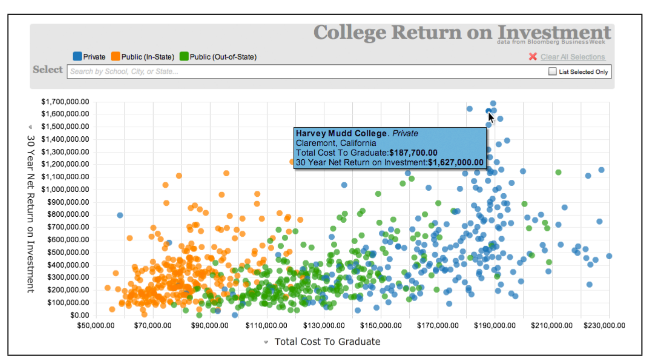


Figure 6. Interactive visualization of college return on investment data used in Study 2. This view plots universities according to their graduation rate and annualized return on investment. Color indicates public (in-state or out-of-state) and private universities.

ond, we asked students to "*Distribute a pool of imaginary funds amongst the public, in-state, and out-of-state schools in California*." We chose these questions to force participants to think critically and construct an argument that built on the exploratory analysis and organization they had completed. We asked participants to provide a short (1-2 paragraph) response to each prompt and to cite the ID numbers of each of the comments that informed their decision. These citations, along with post-study surveys and interviews with select participants, allowed us to connect the synthesis behavior in this phase to the exploration and organization in the earlier phases.

Results

All 12 of our recruits completed the exploration and organization tasks. Of these, ten (6 *tag*, 4 *no-tag*) completed the synthesis task. The two remaining participants dropped out due to scheduling conflicts. We examined all comments generated by the participants and scored them to assess their length, quality, and relevance to topic. We removed one participant in the *tag* condition who produced short, incomplete comments after the task deadlines had expired.

Because of the scope and duration of Study 2, we used a smaller number of participants than in Study 1. Due to the small sample size, most numerical results of this study do not achieve statistical significance. Nevertheless, we believe the qualitative results and feedback from interviews are indicative of real-world usage by teams of analysts.

Exploration. During exploratory analysis, participants in both conditions authored roughly the minimum number of comments (Median=10, MAD=0). Three *tag* subjects applied at least one tag, but no participants tagged heavily, and none authored links – mirroring the results from our live deployment which suggest that organization requires additional motivation.

Organization. In the organization task, the five *tag* participants applied 84 tags and 15 links across 60 of the 138 comments in the workspace. *Tag* participants added the majority of their tags (83%) to comments authored by other users, indicating that they actively considered comments other than their own. There was also very little disagreement when tagging. Two or more users added identical tags to 14 comments, but no two users ever added competing tags or links to the same comment. This result suggests that, even without explicit coordination, users can author tags and links that organize the content without conflicting with one another.

While we also asked participants in the *no-tag* condition to review the comments left by other participants during the second phase, our logs show that *no-tag* participants spent less time in this phase (Median=12 minutes, MAD=6 minutes) than *tag* participants (Median=23 minutes, MAD=13 minutes) and examined fewer comments .

Synthesis. We found that *tag* participants produced longer responses in the synthesis task (Median=3082 total characters, MAD=574) than those in the *no-tag* condition (Median=1480 total characters, MAD=487). To compare the quality of the responses, three independent expert evaluators (one of whom was an author) rank-ordered the anonymized responses from best (1) to worst (9) based on their clarity, consistency, and use of comment citations. The average Spearman's rank correlation coefficient between the evaluators was 0.70, indicating good inter-rater reliability. For each response, we averaged the rankings from all three evaluators to compute an average rank. Comparing the average ranks of all responses, we found that *tag* participants ranked significantly higher (Median=3.83, MAD=0.5) than those in the *no tag* group (Median=6.17, MAD=1) using a Mann-Whitney U test ($U=5.5$, $p<0.0013$). *Tag* participants also cited more comments in their responses (Median=10, MAD=3) than the *no-tag* participants (Median=6, MAD=1). In addition, 79% of the comments cited by *tag* participants had been tagged or linked in the organization step and comments that had been tagged or linked were nearly three times more likely to be cited than those that had not. These results mirror our post-study interviews, which suggest that the organization task helped *tag* participants gain a better understanding of the findings, which they carried over to the synthesis task.

The stronger synthesis responses authored by *tag* participants reflect both their use of tags and link structures during synthesis and the increased awareness of the comments they gained in the organization task. *Tag* participants spent more time in the organization task than their *no-tag* counterparts and visited more comments and views while doing so. However, *tag* participants also cited comments that had been linked together during organization, but had not previously been adjacent to one another, suggesting that they used the tag and link structure directly when generating their result.

DISCUSSION

Our studies demonstrate that tags and links can help participants identify and organize information in a collaborative visual analysis tool. We offer a few concrete takeaways regarding the use of tags and links for collaborative evidence gathering and synthesis tasks:

- Analysts using tags and links were more consistent and more accurate when classifying comments. This result suggests that tags and links are useful when establishing common ground and can help late-joining participants get up to speed in ongoing discussions. We note however, that consensus among analysts is not always desirable and may be symptomatic of groupthink. Competing and divergent interpretations are often desired, in which case tag vocabularies need to be designed to encourage this.

- Analysts using tags and links searched and filtered significantly earlier and classified content more accurately than *no-tag* participants. Tags and links affect how analysts explore and help them calibrate the way they categorize findings. Developers should be careful to select tags and links that encourage desired types of contributions.

- Analysts were significantly more likely to reply to existing discussions when tags were present. This result shows that tags and links encourage contribution and continued discussion and can be used in collaborative visual analysis systems to promote more focused dialog.

- In our live deployments and pilots studies, analysts did not have enough incentive to tag or link comments during open-ended exploration. Because analysts in such tasks often have no immediate reason to revisit their comments, they have little motivation to author additional structure, even if that structure may be useful later. Developers and managers need to guide participation using explicit tasks and incentives in order to facilitate the shift from exploratory analysis to deeper analytical tasks like organization and synthesis.
- Tagging and linking resulted in better synthesis when conducted as part of an explicit organization task than during emergent exploratory analysis. This result suggests a staged approach to collaborative analysis, wherein users first explore a data set, identifying interesting patterns and outliers, then organize those observations to facilitate deeper analysis. Such behaviors have precedent in Wikipedia, where an entire class of contributors categorize articles written by other editors [32]. The lightweight structure provided by tags and links makes this staging possible.

CONCLUSION

In this paper, we demonstrated that the addition of tags and links to a collaborative visual analysis tool can help analysts identify findings in evidence-gathering tasks and improve synthesis. We presented CommentSpace, a system for collaborative visual analysis that allows analysts to comment on interactive visualizations and supplement their comments with tags and links. Based on our studies and deployments using CommentSpace, we believe that this kind of structured support provides a useful mechanism for organizing and navigating text comments and visualization views, but only when staged and managed effectively.

REFERENCES

1. A. D. Balakrishnan, S. R. Fussell, and S. Kiesler. Do visualizations improve synchronous remote collaboration? In *ACM CHI*, pages 1227–1236, 2008.
2. Y. Benkler. Coase's penguin, or, linux and the nature of the firm. *Yale Law Journal*, 112:369, 2002.
3. D. Billman, G. Convertino, J. Shrager, P. Pirolli, and J. Massar. Collaborative intelligence analysis with cache and its effects on information gathering and cognitive bias. In *HCI Consortium Workshop*, 2006.
4. M. Bloch, S. Carter, J. Corum, A. Cox, and M. Ericson. Jacksons billboard rankings over time (interactive graphic). New York Times interactive graphic, June 2009.
5. What's your college degree worth? Bloomberg Businessweek interactive table, June 2010.
6. J. Carroll, M. Rosson, G. Convertino, and C. Ganoe. Awareness and teamwork in computer-supported collaborations. *Interacting with Computers*, 18(1):21–46, 2006.
7. E. Chi and T. Mytkowicz. Understanding the efficiency of social tagging systems using information theory. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, pages 81–88. ACM New York, NY, USA, 2008.
8. T. Chklovski, V. Ratnakar, and Y. Gil. User interfaces with semi-formal representations: a study of designing argumentation structures. In *Proceedings of the 10th international conference on Intelligent user interfaces*, pages 130–136. ACM, 2005.
9. H. Clark and S. Brennan. Grounding in communication. *Perspectives on socially shared cognition*, 13:127–149, 1991.
10. Data360. http://data360.org.
11. N. Diakopoulos, S. Goldenberg, and I. Essa. Videolyzer: quality analysis of online informational video for bloggers and journalists. In *ACM CHI*, pages 799–808. ACM, 2009.
12. P. Dourish and V. Bellotti. Awareness and coordination in shared workspaces. In *ACM CSCW*, page 114. ACM, 1992.
13. R. Eccles, T. Kapler, R. Harper, and W. Wright. Stories in geotime. *Information Visualization*, 7(1):3–17, 2008.
14. D. Gergle, R. Kraut, and S. Fussell. Language efficiency and visual technology: Minimizing collaborative effort with visual information. *Journal of Language and Social Psychology*, 23(4):491, 2004.
15. S. Golder and B. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198, 2006.
16. Google Public Data Explorer. http://www.google.com/publicdata/home.
17. T. Gordon and N. Karacapilidis. The Zeno argumentation framework. In *Proceedings of the 6th international conference on Artificial intelligence and law*, pages 10–18. ACM, 1997.
18. J. Heer and M. Agrawala. Design considerations for collaborative visual analytics. *Information Visualization*, 7(1):49–62, 2008.
19. J. Heer, F. Viégas, and M. Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Communications of the ACM*, 52(1):87–97, 2009.
20. N. Kong and M. Agrawala. Perceptual interpretation of ink annotations on line charts. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, pages 233–236. ACM, 2009.
21. K. Luther, S. Counts, K. Stecher, A. Hoff, and P. Johns. Pathfinder: an online collaboration environment for citizen scientists. In *ACM CHI*, pages 239–248. ACM, 2009.
22. M. McKeon. Harnessing the Web Information Ecosystem with Wiki-based Visualization Dashboards. *IEEE Transactions on Visualization and Computer Graphics*, pages 1081–1088, 2009.
23. I. Mistrik, B. P. Springer, S. J. B. Shum, S. J. B. Shum, A. M. Selvin, A. M. Selvin, M. Sierhuis, M. Sierhuis, J. Conklin, J. Conklin, C. B. Haley, C. B. Haley, B. Nuseibeh, and B. Nuseibeh. Hypermedia support for argumentation-based rationale. In *15 Years on from gIBIS and QOC. In: Rationale Management in Software Engineering (Eds*, pages 111–132. Springer-Verlag: Berlin, 2006.
24. K. Neuendorf. *The content analysis guidebook*. Sage Publications, Inc, 2002.
25. A. Perer and B. Shneiderman. Systematic yet flexible discovery: guiding domain experts through exploratory data analysis. In *Proc. 13th Intl Conf. on Intelligent User Interfaces*, pages 109–118. Citeseer, 2008.
26. P. Pirolli. *Information foraging theory: Adaptive interaction with information*. Oxford University Press, USA, 2007.
27. TIBCO Spotfire Decision Site. http://spotfire.tibco.com.
28. Tableau Server. http:// tableausoftware.com.
29. J. Thomas and K. Cook. Illuminating the path: The research and development agenda for visual analytics. *IEEE Computer Society*, 2005.
30. F. Viégas, M. Wattenberg, M. McKeon, F. Van Ham, and J. Kriss. Harry potter and the meat-filled freezer: A case study of spontaneous usage of visualization tools. In *Proc. HICSS*. Citeseer, 2008.
31. F. Viégas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, 2007.
32. M. Wattenberg, F. Viégas, and K. Hollenbach. Visualizing activity on wikipedia with chromograms. *Human-Computer Interaction–INTERACT 2007*, pages 272–287, 2007.
33. W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort. The Sandbox for analysis: concepts and methods. In *ACM CHI*, page 810. ACM, 2006.