

A Closer Look at Note Taking in the Co-located Collaborative Visual Analytics Process

Narges Mahyar, Ali Sarvghad, and Melanie Tory

University of Victoria



Figure 1. Examples of note taking activities during our observational study. Sometimes note takers are disconnected from group activities.

ABSTRACT

This paper highlights the important role that record-keeping (i.e. taking notes and saving charts) plays in collaborative data analysis within the business domain. The discussion of record-keeping is based on observations from a user study in which co-located teams worked on collaborative visual analytics tasks using large interactive wall and tabletop displays. Part of our findings is a collaborative data analysis framework that encompasses note taking as one of the main activities. We observed that record-keeping was a critical activity within the analysis process. Based on our observations, we characterize notes according to their content, scope, and usage, and describe how they fit into a process of collaborative data analysis. We then discuss implications for the design of collaborative visual analytics tools.

KEYWORDS: note taking, recording, collaboration, tabletop, wall display, history, provenance.

INDEX TERMS: H.5.2 User Interfaces: evaluation; H.5.3 Group and Organization Interfaces: CSCW.

1 INTRODUCTION

We discuss the significance of record-keeping activities during collaborative visual analytics on interactive surfaces. The need to support note taking and chart saving activities arose from observations during a user study that we conducted to examine collaborative visual analytics in the business domain.

Visual analytics tools help users in the business domain to interactively explore relationships and trends in large datasets.

e-mail: {nmahyar, asarv, mtory}@cs.uvic.ca

Business intelligence tools have become commonplace for single users working on desktop machines. By contrast, business users often need to work together to solve problems and make decisions, particularly when each user has unique expertise or responsibilities. However, only a few collaborative visualization tools have been developed (e.g. [3, 9, 31, 32]). Many of these systems, such as ManyEyes [32], were designed to allow casual users to explore and share data on the web.

We focus on co-located work by small groups of known collaborators, as illustrated in Figure 1. We hope to enhance decision making by designing software that facilitates collaborative data analysis. Using such tools on large screen wall or tabletop displays is promising since they are known to support collaborative work. Our work offers suggestions about how to effectively design tools to support group analytics activities and record-keeping around large screens. Previous research has proposed record-keeping and provenance techniques for individual analysts, as well as large screen visualization tools for co-located groups. However, these two topics have not been brought together to design provenance tools specifically for group work. This emerged as a critical need during our observational study, prompting us to analyze note taking and other record-keeping activities in depth in order to establish design requirements. Preliminary results of our study were presented in a workshop paper [14]. Here we contribute a complete description of the study, a much more detailed presentation of the findings, and an in-depth analysis of record-keeping requirements for co-located data analysis.

In this paper, the term *visual representation* refers to artifacts that display data, *visualization* refers to the process of creating and editing visual representations, and *visual analytics* refers to the larger process of using visual representations and other sources of information to form insight and make decisions. Our findings indicate that record-keeping is a pivotal activity that is carried out throughout a data analysis session. We propose a categorization of notes based on their content, scope and usage, and discuss how record-keeping fits into the visual analytics

process. We then discuss potential ways in which record-keeping activities could be integrated into collaborative visualization tools and present some design recommendations.

2 METHODOLOGY

We designed a user study to examine the process of co-located visual analytics in a business context. The study was exploratory in nature rather than designed to test a specific hypothesis. Our goal was to better understand collaborative activities and challenges that might suggest improvements for collaborative analytics tools. Groups of users answered focused business questions and participated in a competitive business scenario. Their work was supported by visualizations of sales data on large wall and tabletop displays.

2.1 Task

Each group completed two tasks, both using a sample e-fashion dataset from Explorer [24]. It contained information about sales of garments in eight states of the United States for three consecutive years. It consisted of 9 columns and 3273 rows of data.

Task 1 consisted of six focused questions designed in a way that users could learn important features of the visualization software. An example question was, “How does the 2003 margin compare to previous years?” Task 1 was intended primarily to help users become familiar with the system. In Task 2, participants were asked to assume the roles of three managers (representing different states) and together determine a marketing budget for the next year. They were told that rationale for the budget should be based on information within the data set.

We received advice from business professionals and faculty members in designing our tasks. Because our participants did not own the data and were not familiar with it, we were concerned about their engagement. Hence, we decided to familiarize them with focused questions first, and then made the scenario competitive to engage all group members in the analysis process.

2.2 Participants

Twenty-seven student participants took part in our study, divided into nine groups of three. To simulate common work situations, all the group members were required to know each other. To mitigate the possible impacts of using students, we mainly selected participants (7 out of 9 groups) who were familiar with the business domain (advanced BCom or MBA students). Participants of the other two groups were computer science graduate students. All users had experience with some kind of data analysis software such as Microsoft Excel.

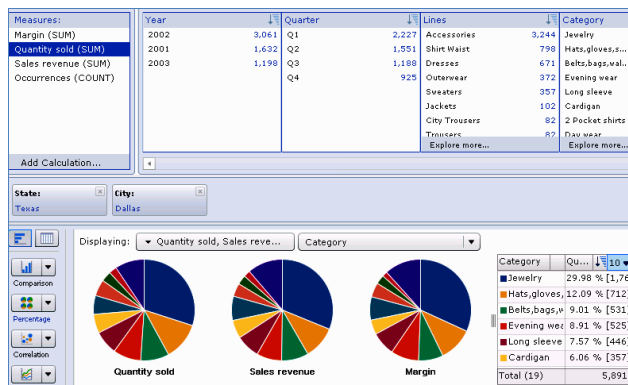


Figure 2. Partial screen shot of Explorer, depicting a comparison chart that visualizes margin, quantity sold, and sales revenue over category, filtered based on state (Texas) and city (Dallas).

2.3 Apparatus and Software

Identical rear-projected Smart DVIT (digital vision touch) screens were used, one in a wall configuration and the other in a tabletop. Both had a size of 61.2” x 34.4” (70” diagonal) and four HD projectors to create a total resolution of 3840 x 2160.

During our pilot studies we noticed that participants took notes on the margin or back of the instruction sheet; therefore, we decided to provide them with pens. We put pens on a table nearby and informed participants that they were there if needed.

We used “Explorer” [24] (Figure 2) as our visual analytics tool. Explorer allows users to interactively browse data, including selecting variables, filtering, and creating different types of charts. The tool was maximized to fill the screen. The software supported a single input but each user had their own stylus so that they did not have to share a stylus to interact with the system. Note that Explorer was developed as a single user application. We therefore expected some problems when using it collaboratively. We hoped that observing these problems would suggest changes that would better support group work. Using existing software enabled us to conduct preliminary requirements analysis without first designing a collaborative system.

2.4 Procedure

We began with a 10-15 minute introduction to Explorer. Participants then spent approximately 30 minutes on task 1 and 40 minutes on task 2. We offered an optional five-minute break between the two tasks. After task 2, participants spent approximately 10 minutes to summarize and write down their results. We asked participants to create a report of their results at the end of task 2 to justify their decisions. Following the computer-based tasks, we conducted an open-ended interview.

Four groups used a tabletop display, four used a wall display and one used both. This gave us an opportunity to observe and obtain users’ feedback on a variety of display configurations. Participants were allowed to arrange themselves freely around the displays, but generally had to stand to interact with them. Chairs were available near the tabletop and two large sofas were available near the wall display where they could sit if desired.

2.5 Data Analysis

We gathered data in the form of recorded videos, interviews with participants, participants’ notes, screen logs and observations made by a live observer. In total ~630 minutes of video and screen logs were captured (~ 70 per session), plus approximately 20 minute interviews per session.

We created two forms and filled them with information while reviewing each group’s video and screen log side by side. In the first form, we recorded events (i.e. what each group member was doing), roles (i.e. note taker, software controller or observer), time stamp, remarks and issues from the video and in the second form chart types, values mapped to axes, filters and time stamps were captured from screen logs. Information in these two forms helped us to analyze participants’ behavior and interactions both with each other and the software. We also extensively studied the content of users’ notes to figure out when and why users took notes and for what purposes they used them.

Characterizations of collaborative processes activities and notes were refined during iterations of our analysis. Cumulatively, we spent nearly 2 months on analyzing gathered information. Interview material was used to support and explain observations from the recorded material.

3 FINDINGS

Given that we ran an exploratory study, we did not have any particular hypotheses. However, we had anticipated problems

such as incorrect software orientation and inadequate awareness of other users' work. We had planned to concentrate on these challenges, but based on our observations we found note-taking a more interesting and yet less discussed obstacle to investigate.

3.1 Participants' Collaboration and Use of Software

Based on our observations, group members were actively engaged in the analysis process. Their analysis activities mainly consisted of mapping and filtering data for new charts and having discussions about them. At any given time only one of the group members was controlling the software, but they all participated in the cognitive process of analyzing the data. At times, users took turns to obtain information needed individually.

Analyzing users' positions confirmed Tang et al.'s [30] results. Participants positioned themselves close to application controls and areas containing information such as a legend. At any time, the user who was closest to the widgets controlled the application; participants changed positions to allow one another to interact with the system. One of the impacts of the software not being designed for large screens was non-equal interaction. For instance, usually one person had to stand at a far corner to see a chart's legend and read it to the group. This likely impacted the groups' work style (i.e. led to closely coupled work with little parallelism) though we cannot be sure of the magnitude and significance of these effects. Since the software layout dictated positioning, we did not examine position data in further depth.

Typically one user assumed the role of note taker for the group and other users assumed the role of data analysts. We observed this phenomenon in eight groups. In one group (group 7), all three participants actively took notes. Unlike role divisions observed in prior research [11], this assignment of roles was usually not discussed explicitly. When we asked some groups about how they came up with their task division, they said it was based on their knowledge of each other's abilities. Role assignments typically remained the same throughout the work session, but in one instance the note taker changed part way through. In some sessions where one person was in charge of note taking, other members also took notes for themselves separately even though it meant that they had to stop working. For instance, participants of group 7 completely stopped working seven times in Task 2 because they were all taking personal notes. This clearly demonstrates that participants need to take notes individually and separately from the group.

3.2 Phases and Activities

We noticed a similar analytics process among all of the groups that we observed. We characterized groups' actions at two levels: high level phases and low level activities, as shown in Figure 3. Because this characterization is grounded by the particular data that we collected, we cannot generalize the phases and activities to other data analysis situations. However, similarity to other frameworks [1, 11, 16, 21] suggests that many aspects of this process probably occur outside the context of our study.

We identified four high-level phases: **problem definition**, **visualization**, **analysis**, and **dissemination**. As shown in Figure 3, there are activities common to all phases such as record-keeping and validation, and activities unique to each phase. We explain each phase and their exclusive activities below. Common activities are explained separately.

Our findings confirmed what Isenberg et al. [11] stated about the non-linear temporal order of activities. In our study, we observed that visualization, analysis, and dissemination occurred in a variety of orders. Visualization and analysis phases were strongly interrelated and quite often, participants moved back and

forth between these two phases. With a much lower frequency, participants in a dissemination phase sometimes returned to previous phases (e.g., to create a chart to include in their report).

3.2.1 Phase I: Problem Definition

Users always started by building a common understanding. For example, they parsed the written description of the problem to build a shared understanding, or posed a new question to be answered. Having a consensus on what problem they were solving was the first step in working collaboratively towards a solution.

3.2.2 Phase II: Visualization

We use the term *visualization* to describe a group of activities resulting in a visual artifact (i.e. a chart). Visualization consisted of mapping data dimensions to visual attributes, filtering data, and creating a visual artifact. For instance, in order to reveal the trend of sales revenue in 2003, participants discovered that they needed to examine values of sales revenue for all the quarters of 2003. Then they mapped "measure" to sales revenue and "dimension" to quarter. Next they chose a "correlation" chart to see the trend of sales revenue for 2001 to 2003. Finally, they applied filtering so that only 2003 data were shown.

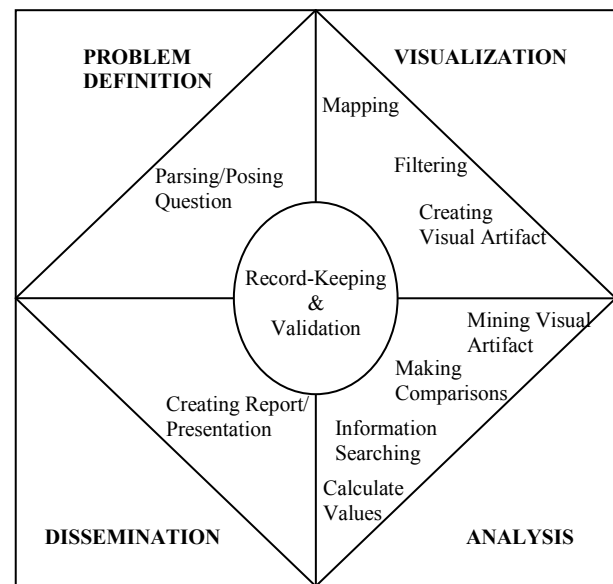


Figure 3. Activities we observed within problem definition, visualization, analysis, and dissemination phases of collaborative visual analytics.

3.2.3 Phase III: Analysis

Analysis was a complex phase that included activities such as:

- Examining visual artifacts,
- Making comparisons by referring to historical information such as notes or saved visual artifacts,
- Calculating derived values through mathematical or statistical operations, and
- Gathering information from external resources (i.e. accumulating information that is not available in the dataset, perhaps from the internet or another person).

The most common activity in this phase was examining charts. Participants worked together to extract information from the chart. In order to achieve this, they made comparisons, calculations or searched for information through external resources. For instance, based on a chart depicting sales revenue for four quarters of 2003 for California, one group decided that they needed to create

similar charts for 2002 and 2001; then they calculated sales revenue averages for all three years and compared values. They repeated the same activities for Texas and New York to make a decision on budget allocation.

Usually the product of this phase was a decision, an answer to a problem, or a hypothesis. In this phase, collaborators often carried out substantial discussion and negotiation to reach a consensus.

3.2.4 Phase IV: Dissemination

In the dissemination phase, participants used products of the analysis phase to generate a semi-formal report of their results. Reporting and presenting are very common activities in business, such as presenting results to the Chief Executive Officer. We observed that while participants were preparing a report they went back to previous phases. This usually happened when they were validating report material or to provide extra content for the report, such as a chart or value.

3.2.5 Validation

Validation activities occurred throughout the entire process, and were concerned with ensuring the correctness of results and a common understanding. In the problem definition phase, participants verified a common understanding of the problem by asking each other questions. In the visualization phase, they verified the correctness of a chart by double-checking filtering and mapping of variables. In the analysis phase, they validated the acceptability of a budget allocation by re-examining charts. In the dissemination phase, participants checked the content of their final report to ensure they were presenting correct material.

3.2.6 Record-Keeping

Record-keeping refers to saving any type of information for further referral and use. In our study this information took the form of charts saved by participants or notes that were written down. Participants took notes in the first two phases to define the strategy, saved values and charts during the analysis and dissemination phases, and referred to their notes and saved charts to facilitate analysis and report writing in the dissemination phase.

3.3 Record-Keeping Strategies

Our observations of record-keeping strategies showed that groups could be divided according to two main approaches: Five groups relied heavily on taking notes and saved only a few charts. Two groups saved many charts and took few notes. Two other groups recorded nearly equal numbers of charts and notes. We only focus on the two extreme approaches. Table 1 shows the number of note taking and chart saving actions by each group. We provide statistics only for task 2, because task 1 did not require users to create a report and therefore very few charts were saved.

We believe that note taking and chart saving approaches can be considered and studied as two different strategies for record-keeping for further analytical use. Note that the prevalence of the note taking strategy over chart saving may be an artifact of our experiment since the process of saving charts was rather cumbersome. Because Explorer was not built with record keeping as a focus, chart saving was inconvenient and required users to select menu items and choose between various options. Charts could only be saved as non-interactive images.

Two groups selected chart saving as their main strategy for keeping important information. One of the groups saved all the charts that they created during task 2 and at the end they created a separate word processing document where they put all the charts side by side for further analysis. One of the participants of this group said “I wish we could have all the charts on screen to see

them side by side” which implies that the tool used for analysis should have provided them with this functionality. The other group just saved a number of charts that they thought were more important. At the end they opened charts one by one for further analysis. Note that other groups saved charts as well, but less often and mainly for use in their reports.

Participants reused the saved charts mainly for two purposes. One was for creating a report at the end of the analysis session (7 groups) and the other was for further analysis of data towards end of the analysis session, after creating several charts (2 groups, shaded gray in Table 1). Groups who saved charts for the second purpose saved a larger number of charts. We cannot exactly pinpoint the criteria different groups used to agree on the importance of a chart. Future work is needed to determine if there are any factors that can predict whether a chart is important enough to be saved. One factor that likely impacts the number of charts that are saved is the complexity of charts that a group creates. For instance, users could create a chart where several variables are mapped onto one axis instead of creating separate charts for each variable.

The chart saving strategy suggests that tools should enable users to save important artifacts and reuse them (e.g. as a history). History items may also reduce the number of notes that need to be taken since many findings are already recorded in the data representations. We noticed that the overall amount of note content taken by participants in group 2 was less than the other groups. Based on our analysis of their activities, we attribute this to the fact that group 2 saved a lot of charts as image files. Note that these users still took some personal notes, so the ability to save charts does not eradicate the need to take notes. It seems that even the most sophisticated history mechanism is incomplete if it does not provide users with the ability to take notes.

Table 1. Number of note taking and chart saving actions by each group during task 2. Shaded groups relied heavily on saved charts for analysis.

Group	1	2	3	4	5	6	7	8	9
Notes	8	4	7	7	20	8	11	9	6
Charts	≥4	22	≥8	12	3	≥8	7	2	2

3.4 Characterization of Note Taking Activities

3.4.1 Note’s Content

Based on our analysis of the notes taken by participants, we break down a note’s high-level content into *findings* and *cues*. Findings are recorded results of mathematical or statistical operations (e.g. 27% higher sales in New York, California’s revenue is \$60000), observations (e.g. Men’s wear sales are higher than women’s wear in a graph) and decisions or outcomes of the analysis process (e.g. allocating more budget to Texas).

A cue is anything noted by the user that is not directly extracted from a visual representation. For instance, users in our study wrote their interpretations of the questions in a concise form for themselves, or they drew circles around keywords in tasks’ questions. Cues could be in the form of to-do lists or questions to be asked /answered later on. For instance, one participant who had assumed the role of California’s manager noted “t-shirts” as reminder to look into California’s sales of t-shirts later on. With respect to the visual analytics process described earlier, we observed that findings were mostly recorded during the analysis phase while cues were mostly taken during visualization.

Findings were also sometimes stored as saved charts rather than written notes. We noticed that in task 1 (in which most users were not saving charts), the amount of note taking was higher than in task 2 (in which users were saving charts). Findings were most

commonly recorded in the analysis phase and cues in nearly all phases.

At a lower level, notes typically contained one or more of the following elements: numbers (e.g. data values), drawings (e.g. flags, charts), text (e.g. questions, hypotheses, reminders), and symbols (e.g. %, \$). In addition to ordinary use of symbols (such as \$ for monetary values), participants used symbols to accelerate the note taking process and thereby decrease distraction from the main task. For instance, they used ↑ symbol to indicate the increase of a value such as revenue.

3.4.2 Notes' Scope

Based on the way that notes were shared we divide them into group and personal notes. We consider a note to be a personal note when it is taken for individual use and a group note when the writer intends to share it with the group. Personal notes were not necessarily private; in some cases, they were shared. For instance, during task 2, participants shared personal notes to justify the amount of the budget they were demanding for their state.

We noticed that the nature of the problem influenced the scope of notes. During task 1 (which had a cooperative nature), usually a participant took notes and shared them with other group members. In contrast, in task 2 (which had a competitive nature), participants tended to take notes individually during the analysis phase and then referred to their notes during budget negotiations.

Figure 4 shows a group note. It has been nicely formatted and contains calculated values. This was used as a summary to help decide budget allocation. The tabular data made the analysis task easier by saving important information; it seemed more convenient and efficient to record this information than to revisit previously created charts. The same person who was in charge of note taking also created the final report.

California 29.79% (28.9%)	New York 20.98% (30.6%)	Texas 27.64%
\$ 1.121 million (BM)	\$ 1.129 million	\$ 1.566 million
17769 (OS)	19109	25,000
8935 - sweat + shirts	4736 Sweat T's (←)	11229
2176 - Accessories	3279 Accessories (←)	3612
1967 - shirt waist	2165 Sweaters (←)	3211
\$ 335,000 Sweat + s	\$ 558,000 Sweat T's (←)	\$ 804,000
\$ 113,000 Acces	\$ 182,000 Acces (←)	\$ 187,000
\$ 115,000 shirt waist	\$ 145,000 Sweaters (←)	\$ 161,000
location \$ 29,000	\$ 31,000	\$ 40,000

Figure 4. A note that was taken for group use. It is nicely structured and is comprised of information for all three rivaling participants.

In contrast, figure 5 shows notes taken by three participants of a group. It can be clearly seen that the notes are less structured and every participant just took notes of what they found important to themselves (e.g., notes about their own state).

In general, group notes were more carefully organized than personal notes, but this varied depending on the individual's note taking style. Personal notes were not always organized or written legibly or in a way that everybody could understand them at a glance. Sometimes authors of personal notes used abbreviations or symbols that could be interpreted only by the note taker. Possibly they were writing as fast as possible to minimize distraction, since taking notes was not their primary focus.

3.4.3 Notes' Usage

Notes were used for a variety of different purposes, most commonly to further analyze findings and facilitate the problem

solving process (analysis phase), validate or remind the person of something (all phases), and create the final report (dissemination phase). Users referred to notes mainly during the analysis phase. Saved values, calculated percentages, drawn charts and other information helped users to make comparisons and reach decisions. Notes also facilitated the problem solving process by recording the direction and sequence of the steps taken. This could help users to more easily determine the next step. For example, by recording the names of the charts created or values calculated or observed, participants could determine the completeness level of the task (e.g., what and how many more charts were to be created). Figure 6 shows an example of recorded information about steps taken and information gathered. It was filled in gradually as information was found in various charts. The figure shows the completed version, indicating that participants have finished their calculations for all three states.

State	Approximate Population (millions)
California	36.5
New York	19.3
Texas	23.9

Remember that you represent one state. You should try to obtain as much of the budget as you reasonably can for your state, while still cooperating to ensure the success of the company as a whole.

① sweat T-shirt

Texas 42 50% T-shirt
New York 30
California 28

State	Approximate Population (millions)
California	36.5
New York	19.3
Texas	23.9

Remember that you represent one state. You should try to obtain as much of the budget as you reasonably can for your state, while still cooperating to ensure the success of the company as a whole.

sales revenue 2.2 million 2.8 1.7

67

28% → 40%
20%
19% → 08% 26%

State	Approximate Population (millions)
California	36.5
New York	19.3
Texas	23.9

4.2 mil sales revenue 2003

Figure 5. Three personal notes taken by participants in one group. This group did not have a defined note taker.

3.5 Awareness with Respect to Note Taking

We noticed that the manual note taking process impacted awareness among group members. This can be seen in Figure 1, where note takers are disconnected from the other team members. This is not necessarily unproductive, but it is possible that the tool design forced this work style that might not be always desirable. Sometimes participants lost a sense of what others were doing while they were taking notes, and then had to catch up. For example, the person who took the note in figure 4 was assigned the role of note taker. He was sitting most of the time observing others (who were exploring data and creating visualizations). He therefore was unable to work directly with the application a lot of the time. Although this division of roles may not have been

unproductive for the group, it did deprive one group member of the opportunity to participate equally in analysis activities.

high Q3 sales				
New York				
	Q1	Q2	Q3	Q4
o1	555	480	257	372
	684	693	500	886
	747	856	914	634
Texas				
	Q1	Q2	Q3	Q4
o1	759	615	329	496
o2	1014	796	794	1438
o3	1102	1088	1032	962
Cali				
	Q1	Q2	Q3	Q4
o1	519	441	394	349
	650	549	760	842
	729	789	775	698

Figure 6. A participant organized information in a tabular format. This note shows that the group has calculated values for all four quarters and for all three states.

3.6 Wall Display versus Tabletop

In line with the literature [23], our study revealed that a wall display could support larger groups of people and provide a common view for presentations. On the other hand, the interactive tabletop display offers potential for supporting formal and informal collaborative activities, such as planning, designing, and organizing. Most of the groups mentioned that they preferred a wall display for audience-based situations like presentation and a tabletop display for more collaborative situations.

With regards to note taking, we observed that a note taker of a group working on the wall display was more disconnected from the group. This was partly because the note taker usually had to sit or lean on a surface to take notes. In addition, often other users obscured the screen by standing in front of it, so it was difficult for the note taker to keep track of what was happening on the screen. This was a less significant issue with the tabletop because the note taker could stand side by side with the others.

4 RELATED WORK

We now discuss how previous research relates to the findings of our study. First, we present work that addresses the collaborative visual analytics process and requirements. We then discuss related work that specifically highlights the importance of history mechanisms and note taking during the course of analysis.

4.1 Collaborative Visual Analytic Process

While substantial research has been devoted to computer supported cooperative work (CSCW) in general, collaborative visual analytics is still under explored due to its unique challenges. Researchers have identified the need for very flexible tools [11, 22, 26]. This includes flexibility to change ordering of activities, work styles (from closely coupled to independent), role assignments, and workspace organization [10, 11]. We also observed that activities did not occur in any consistent order, confirming that flexibility is critical.

Several studies have examined how users analyze data to characterize the processes and activities involved (e.g. [4]). More relevant to our work are studies that consider analytic processes of groups, by using software supporting collaborative work [16, 21], or by using paper-based tasks [11, 22]. Findings of previous studies, regardless of whether the tasks were paper-based or software-based, resulted in similar lists of activities. For instance,

Mark et al. [16] identified processes of parsing the question, mapping variables, finding or validating a visual representation, and validating the entire analytical process. Isenberg et al. [11] identified processes of browsing, parsing, discussing collaboration style, establishing task strategy, clarification, selecting, operating and validating. These lists of activities bear strong resemblance to our own characterization. For instance, our first phase “problem definition” has been identified in previous work as “parsing” [11, 16] or “problem interpretation” [21].

In contrast to previous work, our framework captures the whole process of a visual analytics session (as opposed to say, simply the visualization or analysis phase) and breaks each phase down into lower level activities. We believe this two-level structure provides a useful way to think about the analysis process. We also highlight record-keeping as a critical activity during all phases. Although record-keeping has been previously mentioned as a relevant action [1, 7, 28], its importance may have been underrepresented in previous frameworks describing collaborative analytics processes.

4.2 Record-keeping

Many researchers have mentioned advantages of history tools and their importance [6, 18, 19, 20, 23]. History tools enable users to review, revisit and retrieve prior states [28]. They can also be used to create a report or presentation following analysis [6].

The ability to add metadata (annotation, notes) to a visual representation has also been recommended [6, 8, 9] and implemented [2, 7, 13, 29], as it is often difficult for data analysts to remember previous findings and cues and reuse them efficiently [17]. Both textual and graphical annotation of visualizations may be necessary. Heer et al. [6] mentioned the importance of annotations for distributed settings. Moreover, previous research shows that notes “act as a bridge between the analyses executed in the system and [a user’s] cognitive process” [29]. Shrinivasan et al. [29] take annotation a step further by automatically recommending related notes based on the current analysis context. Our findings highlight note taking as a pivotal activity during the course of analysis, emphasizing the importance of including such provenance tools in visual analytics systems.

We emphasize that the vast majority of history / provenance tools have focused on single-user systems. Although previous work has postulated that history tools may be even more important for collaborative work [5, 15], little guidance is available to help build such tools effectively. Extending history mechanisms to represent activities of multiple co-located users is non-trivial due to issues of awareness, disruption, organization, and so on. In a previous workshop paper [25], we hypothesized how history tools might need to change to support multiple users. In the next section, we use evidence from our observational study to propose more specific design guidelines and considerations.

5 DISCUSSION

5.1 A Clear Need for Record-Keeping Support

Our main finding is the importance of recording findings and cues (as notes or saved charts). Although some previous research [7, 8, 9] has suggested allowing annotation of visualizations, our study highlighted the importance of note taking as a critical activity. Taking and using notes was a frequent activity in all phases of the collaborative decision making process. Lack of support for record-keeping had negative consequences such as disruption to workflow and decreased awareness of group activity. The importance and difficulty of record-keeping was somewhat unexpected, as we did not tell participants that they should take notes and expected the major bottleneck to be interaction

challenges with the single-user software. This highlights the need to build explicit record-keeping support into collaborative visualization tools. Recently, some research [13, 29] has demonstrated how this can be done for single users, but work remains to extend this idea to multi-user systems.

5.2 Impact of Task Nature on Note Taking

Our study further illustrated that the nature of the task affects both the process of collaboration and the division of workspace. Task 1, which involved focused questions, encouraged a highly-coupled collaborative style of work, while task 2, which required competition, led to a loosely-coupled collaborative work style. In the interviews, most of our participants said that they would have preferred to explore information for task 2 individually and then later share their results. As a result, notes taken in task 1 were public, while notes taken in task 2 had a combination of public and private scopes. This finding emphasizes the need to support both individual and jointly coupled activities as previously suggested [11, 12, 22, 27]. More importantly for us, it suggests that both group and individual record-keeping is necessary. An effective collaborative analytics system should provide both public and private records that are easy to distinguish, and enable users to seamlessly switch between them.

5.3 Recommendations to Support Note Taking

How to design effective record-keeping functionality for co-located work is not entirely clear. Here we offer some suggestions, which vary depending on the nature of the collaboration, whether or not the record involves data (or is linked to data), and whether the note is taken for group or personal use.

5.3.1 Integration Level for Notes and Saved Artifacts

Should notes be integrated with a history mechanism (i.e. along with saved artifacts and system states) or kept as a separate “notebook”? Our analysis suggests that either answer would be too simplistic. Some notes, especially annotations and other notes of findings, have a clear link to an artifact that helped to form the insight. For example, a user might save a chart of revenue across different states, note that revenue is highest in New York, or write a reminder to later break down the New York revenue data by year and quarter. In these cases, the record is either a particular representation of data itself, or can be linked to one. Retaining this link and enabling the user to return to the artifact and system state would have a clear benefit. However, notes also served as cues (e.g. reminders) or collected together findings from a variety of sources; in these situations a notebook style is more relevant.

As a result, we suggest a hybrid model in which notes can be collected together in notebook pages but parts of a note could link to related artifacts, which might be stored in a chronological history. This would ensure that users could easily refer to source data when reviewing any given note. Artifacts could similarly link to the notes associated with them, and might also be directly annotated with drawings or text. We expect that such functionality would also simplify the task of recording findings. For instance, instead of writing down “Men’s wear has higher sales than women’s wear”, the user could simply circle the male bar in a bar chart showing sales broken down by gender.

Such notes could be captured on the shared display within specialized notebook containers that ideally would support both text and diagrams. Furthermore, these could be treated by the system as if they were artifacts such as charts. For example, they could potentially be added to a chronological history in the same way that a chart would be added. This would capture the development of the note over time, making it easier to understand the process that was followed. This might be particularly useful

for helping a novice learn the process that an expert analyst followed, or to help an analyst who is new to the project understand what work was done by previous analysts. Finally, because the number of notes and data artifacts can grow large quite quickly, we believe that searching and filtering both types of objects will be important.

5.3.2 Notes for group versus individual use

Group notes could take the form of a shared history/ notebook, plus shared note containers or papers as described above. To keep track of who did what, they might be spatially organized or colour coded by user. Individual notes present a greater problem because users may wish to keep their notes private or may want to avoid the burden of viewing all other users’ notes. At the same time, individual notes occasionally need to be shared. One possibility is to provide private space within a shared display, if there is sufficient screen real-estate and if the notes are not confidential. Another alternative is to provide each user with a private display such as a tablet or digital paper. These could be linked to the common display to enable sharing. Ordinary paper notes or an unlinked private display are also viable options, but are more difficult to share with several people at once.

5.3.3 Record-keeping for different types of collaboration

We observed tightly coupled work, where a shared history / notebook would probably suffice. For loosely coupled work, participants may need to corroborate and combine the outcomes of their individual work. In this case, it may be better to give each person personal space to work independently, but also allow sharing. Individual notes and history items that could later be merged together could allow each user to track their individual work and then later compare it to the work of others. Note that although they allow private work, individual desktops may not be the best solution here because they make sharing cumbersome.

5.4 Generalizability

Our results are subject to some caveats. We chose to focus on the business domain, so our users were primarily business students. We suspect that collaborative use of visualization tools will be similar for other group decision-making tasks, but it is possible that we observed some peculiarities unique to business. Secondly, we chose to utilize existing visualization software, to ensure that users could work with interactive and customizable representations of data. However, our users’ behaviour may have been influenced by the available technology, especially their closely coupled work style and the tendency of most groups to write notes rather than save charts. Finally, we examined a group size of three. Collaborative processes are likely to differ for pairs of users or much larger groups.

6 CONCLUSION AND FUTURE WORK

We characterized phases and activities involved in collaborative visual analytics for co-located groups. We also identified record-keeping as a process that is intensively used by data analysts. We characterized notes according to whether they were findings or cues, and whether their scope was for personal or group use. We also described how notes were taken and used within four identified phases of data analysis. These analyses enabled us to offer numerous suggestions of how to better support record-keeping activities within visual analytics tools.

Additional studies are required to answer questions about how exactly note taking support should be provided in collaborative visualization systems. For instance, it is still unclear how we can best support both individual and group note taking activities. Another important consideration is the form of input/s to be used

for note taking in a shared display situation. Further studies should also be conducted in other application domains. For instance, in some disciplines, records of decisions need to be kept for legal purposes and may therefore need to be more formal and detailed. Further research is needed to investigate how these diverse needs can be best supported within visual analytics tools.

ACKNOWLEDGMENTS

We thank SAP and the Natural Sciences and Research Council of Canada (NSERC) for funding this research.

REFERENCES

- [1] S. Card, J. D. Mackinlay, and B. Shneiderman, Readings In Information Visualization: Using Vision To Think. *Morgan Kaufman* publishers, Inc., San Francisco, USA, 1999.
- [2] S. E. Ellis, and D.P. Groth. A collaborative annotation system for data visualization. *Proc. Working Conference on Advanced Visual interfaces*, AVI '04. ACM, New York, NY, pp. 411-414, 2004.
- [3] C. Forlines, A. Esenther, C. Shen, D. Wigdor, and K. Ryall. Multi-User, Multi-Display Interaction with a Single-User, Single-Display Geospatial Application, *Proc. ACM Symp. User interface Software and Technology*, ACM, New York, NY, pp. 273-276, 2006.
- [4] D. Gotz and M.X. Zhou. Characterizing users' visual analytic activity for insight provenance. *Information Visualization*, vol. 8, Palgrave Macmillan, pp. 42-55, 2009.
- [5] J. Heer, F. van Ham, S. Carpendale, C. Weaver, and P. Isenberg. Creation and collaboration: engaging new audiences for information visualization. In A. Kerren et al., eds. *Information Visualization, LNCS 4950*, Springer Berlin, pp. 92-133, 2008.
- [6] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala. Graphical Histories for Visualization: Supporting Analysis. Communication, and Evaluation, *IEEE Trans. Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1189-1196, Nov. 2008.
- [7] J. Heer, F.B.Viegas, and M. Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Commun. ACM*, pp. 87-97, 2009.
- [8] T. Isenberg, P. Neumann, S. Carpendale, S. Nix, and S. Greenberg. Interactive Annotations on Large, High-Resolution Information Displays. *Proc. 2006 Confe. Compendium of IEEE VIS*, IEEE Computer Society, Los Alamitos, CA, pp. 124-125, 2006.
- [9] P. Isenberg, and S. Carpendale, "Interactive Tree Comparison for Co-located Collaborative Information Visualization", *IEEE TVCG*, vol. 13, no. 6, pp. 1232-1239, 2007.
- [10] P. Isenberg and D. Fisher. "Collaborative Brushing and Linking for Co-located Visual Analytics of Document Collections", *Proc. EuroVis, Computer Graphics Forum*, vol. 28, no. 3, pp. 1031-1038, June 2009.
- [11] P. Isenberg, A. Tang, and S. Carpendale. An Exploratory Study of Visual Information Analysis. *Proc. CHI 2008*, ACM, New York, NY, pp. 1217-1226, 2008.
- [12] T. Jankun-Kelly, K.-L. Ma, and M. Gertz. A Model and Framework for Visualization Exploration. *IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 2, pp. 357-369, 2007.
- [13] N. Kadivar, V. Chen, D. Dunsmuir, E. Lee, C. Qian, J. Dill, C. Shaw, and R. Woodbury. Capturing and supporting the analysis process. *Proc. IEEE VAST 2009*, pp. 131-138, 2009.
- [14] N. Mahyar, A. Sarvghad, and M. Tory, Roles of Notes in Co-located Collaborative Visualization, *Proc. Workshop on Collaborative Visualization on Interactive Surfaces (CoVIS 2009, October 11, 2009, Atlantic City, USA)*, pp.13-16.
- [15] G. Mark, K. Carpenter, and A. Kobsa. A Model of Synchronous Collaborative Information Visualization. *Proc. 7th Intl Conf. Information Visualization*, IV. IEEE Computer Society, Washington, DC, 373, 2003.
- [16] G. Mark, and A. Kobsa. The Effects of Collaboration and System Transparency on CIVE Usage: An Empirical Study and Model. *Presence: Teleoper. Virtual Environ*, vol. 14, no. 1, MIT Press, Cambridge, MA, pp. 60-80, 2005.
- [17] C.C. Marshall. *Work Practices Study: Analysts and Notetaking*. Unpublished report, May 1990.
- [18] C. Meng, M. Yasue, A. Imamiya, and X. Mao. Visualizing Histories for Selective Undo and Redo, *Proc. 3rd Asian Pacific Computer and Human Interaction*, APCHI. IEEE Computer Society, Washington, DC, 459, 1998.
- [19] K. O'Hara, and A. Sellen. A comparison of reading paper and on-line documents. *Proc. CHI*, ACM, New York, NY, pp. 335-342, 1997.
- [20] T. Oinn, M. Addis, J. Ferris, D. Marvin, M. Senger, M. Greenwood, T. Carver, K. Glover., M.R. Pocock, A. Wipat, and P. Li. Taverna: a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics 2004*, vol. 20, pp. 3045-3054, 2004.
- [21] K.S. Park, A. Kapoor, and J. Leigh. Lessons Learned from Employing Multiple Perspectives In a Collaborative Virtual Environment for Visualizing Scientific Data. *Proc. CVE 2000*, ACM, New York, NY, pp. 73-82, 2000.
- [22] A. Robinson. Collaborative Synthesis of Visual Analytic Results. *Proc. IEEE VAST 2008*, pp. 67-74, 2008.
- [23] Y. Rogers and S. Lindley. Collaborating around large interactive displays: Which way is best to meet?. *Interacting with Computers*, vol. 16, no. 6, pp. 1133-1152, 2004.
- [24] SAP Business Objects, Explorer in the cloud, <https://create.ondemand.com/explorer>.
- [25] A. Sarvghad, N. Mahyar, and M. Tory, History Tools for Collaborative Visualization, *Proc. Workshop on Collaborative Visualization on Interactive Surfaces (CoVIS 2009, October 11, 2009, Atlantic City, USA)*, pp.21-23.
- [26] S. D. Scott, K. D. Grant, and R. L. Mandryk. System guidelines for co-located collaborative work on a tabletop display. *Proc. ECSCW'03*, K. Kuutti et al., Eds. ECSCW. Kluwer Academic Publishers, Norwell, MA, pp. 159-178, 2003.
- [27] S.D. Scott, M.S.T. Carpendale, and K.M. Inkpen. Territoriality in collaborative tabletop workspaces. *Proc. CSCW*, ACM, New York, NY, pp. 294-303, 2004.
- [28] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *Proc. IEEE Visual Languages*, IEEE Computer Society, Washington, DC, 1996.
- [29] Y.B. Shrinivasan, D. Gotz, J. Lu. Connecting the Dots in Visual Analysis. *IEEE VAST*, pp. 123-130, October 2009.
- [30] A. Tang, M. Tory, B. Po, P. Neumann, S. Carpendale, Collaborative Coupling over Tabletop Displays. *Proc. the Conference on Human Factors in Computing Systems (CHI'06)*, ACM Press, New York, pp. 1181-1290, 2006.
- [31] F. Vernier, N. Lesh, C. Shen. Visualization Techniques for Circular Tabletop Interfaces. *Proc. Advanced Visual Interfaces*, ACM, New York, NY, pp. 257-265, 2002.
- [32] F.B. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. ManyEyes: a Site for Visualization at Internet Scale. *Proc. IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1121-1128, 2007.
- [33] J.S. Vitter. US&R: A New Framework for Redoing, *SIGPLAN Not.* vol. 19, no. 5, ACM, NY, USA, pp. 168-176, 1984.