Supporting Team-First Visual Analytics through Group Activity Representations

Sriram Karthik Badam[1,*] Zehua Zeng[1,†] Emily Wall[3,‡] Alex Endert[4,§] Niklas Elmqvist[1,2,¶]

[1]Department of Computer Science, [2]College of Information Studies, University of Maryland, College Park, MD, USA
[3]School of Computer Science, [4]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, USA

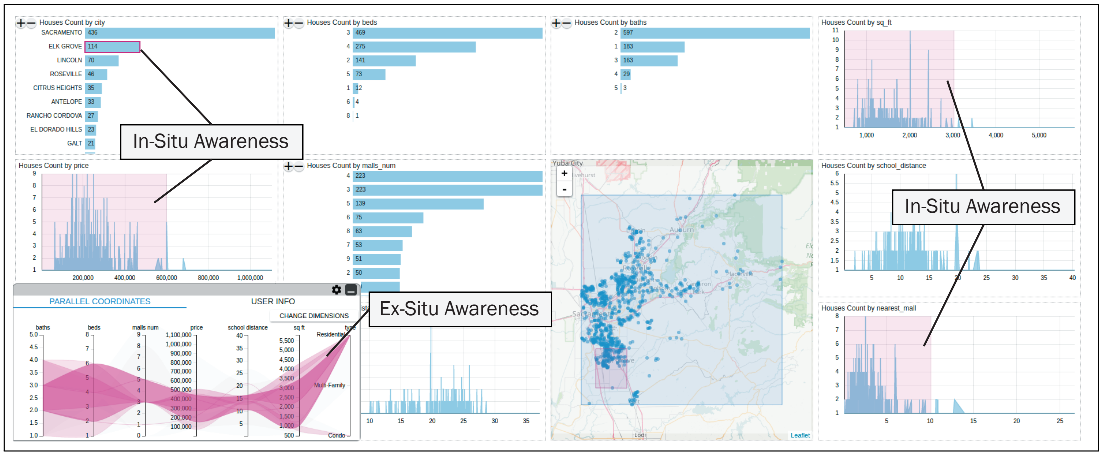

Figure 1: Our InsightsDrive tool providing a dashboard of interactive visualizations for real estate data with brushing-and-linking activated. InsightsDrive follows a team-first design for collaborative visual analytics by combining seamless in-situ awareness (selection shadows in pink) with an ex-situ awareness widget (parallel coordinates) providing coverage information of the collaborators.

Abstract

Collaborative visual analytics (CVA) involves sensemaking activities within teams of analysts based on coordination of work across team members, awareness of team activity, and communication of hypotheses, observations, and insights. We introduce a new type of CVA tools based on the notion of "team-first" visual analytics, where supporting the analytical process and needs of the entire team is the primary focus of the graphical user interface before that of the individual analysts. To this end, we present the design space and guidelines for team-first tools in terms of conveying analyst presence, focus, and activity within the interface. We then introduce InsightsDrive, a CVA tool for multidimensional data, that contains team-first features into the interface through group activity visualizations. This includes (1) *in-situ* representations that show the focus regions of all users integrated in the data visualizations themselves using color-coded selection shadows, as well as (2) *ex-situ* representations showing the data coverage of each analyst using multidimensional visual representations. We conducted two user studies, one with individual analysts to identify the affordances of different visual representations to inform data coverage, and the other to evaluate the performance of our team-first design with exsitu and in-situ awareness for visual analytic tasks. Our results give an understanding of the performance of our team-first features and unravel their advantages for team coordination.

*e-mail: sbadam@cs.umd.edu
†e-mail: zzeng@cs.umd.edu
‡e-mail: emilywall@gatech.edu
§e-mail: endert@gatech.edu
¶e-mail: elm@umd.edu

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical user interfaces (GUI)

1 Introduction

In the midst of increasingly ubiquitous data, collaboration is becoming a necessity for effective data analysis [31]. However, such collaboration adds complexity to the sensemaking process [17]. A major challenge for collaborative sensemaking is providing team members an awareness [10] of the activities of others to coordinate the sensemaking task, avoid interference with each other, and improve the team's collective performance. To answer this challenge, collaborative visual analytics (CVA) tools have explored the concepts of presence, attention, communication [20], coverage [13]—the data being explored by each user—and collaborative brushing [18, 27]. However, many mechanisms for supporting collaboration, including interpreting coverage and communicating observations, are explicit and heavyweight in nature, as they require the analyst to deviate from the actual sensemaking activity. This is because these operations are "analyst-first"—designed to just extend an individual analyst's capabilities to work with a group beyond exploring the data by herself.

In this paper, we explore alternative CVA tool designs that are inherently "*team-first*"[1] where the visual interface considers the needs of the team as a whole and seamlessly feeds the group activity without significantly deviating the users from their tasks. For this purpose, we present the design space for capturing group activity and providing group awareness, and discuss guidelines for creating team-first tools in terms of integrating presence, attention, coverage, and communication aspects into the visual interface. We then present InsightsDrive, a prototype CVA tool (Figure 1), that instantiates this design space for collaborative multidimensional

[1]Compare this to mobile-first web design where the goal is to ensure that a website works on mobile devices first, and computer screens afterwards.

data analysis by automatically capturing the interactions performed by the team and seamlessly feeding them back to the individuals through group activity representations. This includes visualizations for (1) in-situ awareness that show users' selections as color-coded shadows in the visualizations in the VA interface, and (2) ex-situ awareness that show users' *data coverage* information using a scatterplot or parallel coordinates plot embedded in a separate interface widget. InsightsDrive focuses specifically on synchronous collaborative activity within small distributed groups with a flat hierarchy and captures the latest interactions of the users in these awareness visualizations. Beyond these features, InsightsDrive acts like any other VA tool by providing multiple data visualizations to interactively explore a dataset to develop insights from the data.

Presenting group activity by taking the best advantage of both in-situ and ex-situ representations can help the analyst directly and effortlessly adapt to both the interactions and coverage of her collaborators. For this reason, InsightsDrive can be an effective team-first CVA tool. We evaluate our work through two user studies. The first user study involved a qualitative assessment of ex-situ awareness to understand the affordances of different visual representations for group activity. It revealed the tradeoffs of parallel coordinates and scatterplot representations for capturing data coverage. The second user study involved a quantitative comparison of our InsightsDrive tool, whose design integrates in-situ and ex-situ awareness, for a decision-making task against an analyst-first tool, which uses ex-situ representations to just extend the analyst's capabilities on a VA interface. Results from this study showed that the combination of both awareness types led to faster decision making within teams due to better coordination (observed from participant feedback). It also revealed the effectiveness of these group activity representations to support the team-first approach.

2 BACKGROUND

Our basic idea in this paper is to describe an alternative perspective for the creation of CVA tools. Rather than creating tools that extend individual analysts' capabilities, we propose the notion of team-first analytics, wherein the design of the tool focuses on supporting the needs of the team as a whole. Here we review the background on computer-supported cooperative work, collaborative visual analytics, and group awareness.

2.1 Awareness and Presence

The field of computer-supported cooperative work (CSCW) focuses primarily on the theory, design, and practice of software used concurrently by multiple users [3]. While the scope of CSCW spans decades and disciplines, here we focus on *coordination*: mechanisms that facilitate the collaborative process on a meta level without directly contributing to the collaborative task [30]. Such coordination is vital to ensure efficient collaboration, particularly as the number of collaborators grows. Collaborative editors such as Google Docs provide coordination mechanisms such as chat, comments, shared highlighting, suggestions, and revision histories.

One of the key aspects of efficient coordination is to establish *common ground*, or "mutual knowledge, mutual beliefs, and mutual assumptions" [7] about the shared task. Achieving and maintaining such grounding in communication requires *group awareness*: an up-to-date understanding of the interactions of other collaborators in the shared space [10]. This is particularly important in remote collaborative sessions since such settings lack familiar physical awareness cues. Several approaches in general HCI and CSCW focus on providing group awareness, including techniques such as the use of "radar" overviews of the shared space [12] and showing "ghost" arms of remote collaborators on a tabletop display [29].

Presence is a special form of group awareness, where the spatial proximity of a user to an object conveys an interest in that object. This effect is intrinsic to the physical world, but is more elusive in digital settings; for this reason, presence and proximity are commonly used in 3D virtual environments. Nevertheless, the concept can be used to great effect in standard desktop applications. For example, Laufer et al. [21] created a synchronous collaboration extension to the Prezi presentation tool where the locations of avatars represent the current focus of each collaborator on the canvas.

2.2 Collaboration in Visual Analytics

Collaborative visual analytics can be succinctly defined as the shared use of visual analytics software by multiple users, and has been named one of the grand challenges of the field [31]. The value proposition for this practice is simple: involving multiple analysts generally improves the analytical outcomes in terms of time, quality, or both. As a case in point, Mark et al. [23] discovered significant improvement for collaborative visualization compared to single-analyst usage, and Balakrishnan et al. [5] similarly point to significant performance gains when analysts used a shared visual representation. However, while collaborative VA and visualization has many similarities with CSCW and groupware, it also has its own distinct set of challenges [17], including its typically expert analyst audience, its focus on sensemaking rather than productivity, and its long-term, multi-stage, and multi-representation workflow. This means that existing CSCW techniques cannot be applied indiscriminately; Isenberg et al. [17] survey the similarities and differences between visualization and CSCW.

Collaboration is often classified by space (co-located or distributed) and time (synchronous or asynchronous) [3]. The most common setting is asynchronous and distributed. Asynchronous social data analysis [15] was best captured in IBM's now-defunct ManyEyes [33] website, but these ideas live on in commercial tools such as Tableau Public.[2] Co-located and synchronous settings are also common. Here multiple analysts work together on an analytical task in the same room. Many of the visual analytics systems for co-located collaboration have been guided by work by Robinson [26] as well as Isenberg et al. [19], which both study the behavior of individuals as well as groups in co-located paper-based analysis. The simplest approach is simply to connect multiple laptops and devices in the same; VisPorter [6] and PolyChrome [2] are examples of frameworks to enable this. Such frameworks allow for building co-located collaborative environments using specialized hardware that enable multiple users to interact simultaneously.Visual analytics in such environments was pioneered by a collaborative tree analysis tool for digital tabletops from Isenberg and Carpendale [16], but similar work includes Lark [32], which externalizes data pipelines on a shared touch surface, and Cambiera [18], which captures documents read and queried within text collections.

2.3 Awareness and Coverage in Visual Analytics

Collaborative visual analytics requires particular attention to coordination mechanisms due to the complex nature of sensemaking. For example, the branch-explore-merge protocol [24] is a prime example of a sophisticated coordination mechanism that enables participants to branch from the shared state, explore the data independently, and merge back any new findings to the shared exploration.

Heer and Agrawala cite awareness as one of the main design considerations of collaborative visual analytics [14], naming notification and history mechanisms as key features. As a case in point, Baker et al. [4] proposed a notification technique for providing customized awareness to individuals assuming different roles in a collaborative setting. Similarly, Balakrishan et al. [5] provide awareness to users using shared visualizations. Finally, the Hugin [20] visual analytics tool provides awareness based on radar widgets [11, 12] and remote interactions [29].

Heer and Agrawala also propose *social navigation* [8], where the presence and activities of multiple users in a digital space

[2]http://public.tableau.com/

are recorded and visualized, as a way to aggregate the actions of multiple analysts in collaborative visual analytics [14]. One concrete approach based on social navigation is Scented Widgets [34], which embed visual representations of prior use in-situ on the interface elements—such as range sliders, lists, and hierarchies—themselves. In a similar vein, the collaborative brushing proposed by Isenberg and Fisher [18] for text documents was extended to tabular data by Hajizadeh and Tory [13]. Mahyar and Tory [22] take this even further by connecting collaborators' findings using an approach they call "Linked Common Ground."

One unique awareness aspect of collaborative visual analytics over traditional CSCW is the notion of *collective data coverage*. Sarvghad and Tory found that dimension coverage increases the breadth of exploration without sacrificing depth for a single user [28] and reduces work duplication in async. collaboration [27].

3 Supporting Team-First Visual Analytics

Being aware of the group's work in collaborative visual analytics (CVA) allows for subdividing tasks, avoiding conflicts, and improving communication. Here awareness can mean many things from noticing the presence of the collaborators to understanding the interactions and insights made by the group. However, providing complete awareness to the user can be a double-edged sword as the users can be significantly deviated from the actual sensemaking when overloaded with this information. One main goal for the team-first CVA design is to embed and blend the awareness information within the VA interface such that the users can perceive the group activity without requiring additional cognitive effort. In this section, we present the design space for group awareness in terms of awareness types and presentation, and then provide guidelines for effective awareness integration for the team-first design.

3.1 Capturing Group Activity

Traditional methods for awareness include presence and attention. However, sensemaking activities are complex with users going through multiple stages to gain insights [25]. Therefore capturing the data coverage and supporting communication is important.

3.1.1 Presence and Attention

The digital *presence* of collaborators implies interest solely based on their proximity and reduce conflicts during group activity. Further, knowing where the collaborators' attention is focused allows team members to understand their tasks and their interactions with the data. For example, Laufer et al. in Prezi Meeting [21] use avatars to represent the position and attention of collaborators within a presentation. In visual analytics, presence and attention have been explored using multiple techniques. Previous approaches allowed users to explicitly switch to see others' views, or show data items that are common to other collaborators' analyses [5, 22] to understand their focus and attention.

3.1.2 Analysis Coverage and History

The concept of *analysis coverage* [27] captures which parts of the data that a team is actively viewing or has viewed in the past (history). While it is not necessary that a team views the entire dataset, and further, it is not given that viewing data automatically yields all insights from it, it is still a useful metric on the completeness of a collaborative analysis. We distinguish three types of coverage:

- **Attribute Coverage:** The attributes that are currently being considered by the analyst. For example, if an analyst is viewing a bar chart capturing number of sports cars, sedans, coupes, and wagons in a cars dataset, he might select the bar containing sedans to filter the other views in the interface. The attribute coverage of the dataset would then be "car type".
- **Range Coverage:** The range of attribute values being examined. For example, suppose an analyst selects five cars of interest by filtering a specific range of values for gas and mileage attributes on the interface. These ranges would be considered as the range coverage.
- **Feature Coverage:** The connections between different dimensions of data being examined. For example, say an analyst is exploring sports cars with a high top speed. Feature coverage relates to providing information about the interesting connections between other attributes including cylinders, year of release, gas mileage, and horse power. One such connection can be that a lot of sports cars have poor gas mileage.

3.1.3 Communication and Deixis

Communication is a key part of effective collaborations, allowing team members to coordinate tasks and share insights. In an awareness visualization, communication can be facilitated directly using mechanisms such as textual, audio, or video chat. Another important aspect is supporting *deixis* [14]—essentially, the ability to point at elements of reference—to promote effective communication. For example, *collaborative brushing* [18] highlights selections made by a user on all of the remote displays for the entire team.

3.2 Presenting Group Activity

Presenting the group awareness in a VA interface—containing visualizations of a dataset of interest—based on the above categories quickly becomes a binary choice: should the awareness representation be separate (*ex-situ*) from the primary visualizations, or should it be integrated into (*in-situ*) said visualizations?

3.2.1 Ex-Situ Representation

Ex-situ group awareness visualizations provide a separate view that captures presence, coverage, and communication aspects. For example, the group awareness representations introduced by Sarvghad and Tory [27]—circular dimension co-mapping and treemap designs—are ex-situ as they are presented in a separate view from the actual data visualizations. Ex-situ representations minimize clutter, because the view is separated from the primary visualization interface and the visual encoding can thus be designed freely. However, adding a new view requires splitting the user's attention and introduces a risk of change blindness.

3.2.2 In-Situ Representation

In-situ representations for group awareness are blended into the primary visualization interface that may contain multiple visual representations of a dataset. In this case, the group activity information can be either directly overlaid on the content of a data visualization within the VA interface resembling a shadow, or directly attached to a target visualization within the interface resembling a scented widget [34]. Either way, this is meant to capture the collaborators' selections and interactions. For both techniques, users can be distinguished through colors and labelling. These in-situ representations can make analysts be aware of what other team members are doing without having to divert their attention away from the main visualization window. However, information conveyed by these representations is limited compared to an ex-situ representation, which has its own dedicated space on the interface.

3.3 Designing Team-First Visual Analytics Tools

Given this design space, a team-first approach should provide awareness including presence, attention, and coverage information within the VA interface during group activity without deviating the user from the actual sensemaking activity. To develop our team-first VA tool, we used the following guidelines,

G1 Adapt the group awareness representation to the sensemaking scenario—target dataset and collaboration style (asynchronous/synchronous and distributed/co-located).

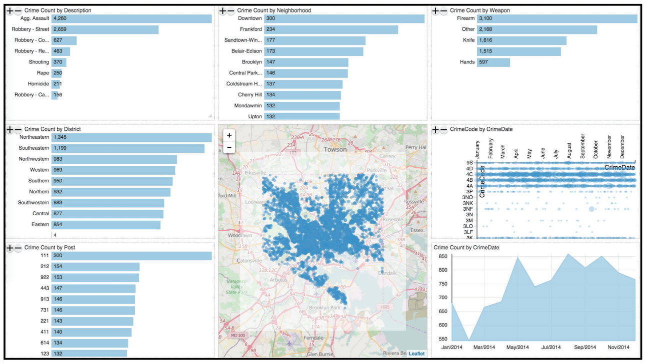

Figure 2: The InsightsDrive VA tool presenting a Baltimore crime dataset. By default, it shows bar charts (of crime count) for all the categorical variables, a line chart for visualizing the temporal component, and a map to capture the geographic data. Clicking the '+' button on top of each view allows for adding an extra dimension to the view to perform 2D analysis (e.g., bar chart turns into scatterplot).

G2 Target glanceable visual representations that convey group activity without heavyweight interactions and context switching, and allow further exploration if needed.

G3 Avoid visual clutter within in-situ and ex-situ awareness representations to aid in a quick understanding of group activity.

G4 Support customization of awareness representation, since users may be interested in different aspects of the group activity and have different perceptual capabilities (some can be faster at interpreting visualizations than others).

G5 Target extensible representations that can be applied to different visualization designs—line charts, bar charts, and graphs—to maintain consistency in awareness representation.

4 INSIGHTSDRIVE: A TEAM-FIRST CVA TOOL

Based on our design space and guidelines, we developed a prototype team-first visual analytics tool called INSIGHTSDRIVE (Figure 1). This tool was developed for multidimensional data with synchronous collaboration in mind; therefore, the group activity representations capture the current focus and selections of users (G1). It is currently most suited for distributed teams of analysts with a flat hierarchy since all the team members have access to the same type of features within the interface.

4.1 Interface

The actual visualization interface within our tool (Figure 2) contains multiple views, with each view showing a summary for a particular dimension within the dataset as a bar chart, line chart, or map visualization. Each view is interactive and allows selections, and uses brushing and linking to coordinate the other views. To provide a quick understanding of the group activity without cluttering the visual interface, our InsightsDrive combines ex-situ and in-situ representations to automatically capture presence, attention, and coverage of the collaborators (in a glanceable way while minimizing context switching from the actual activity) and support further exploration and customization (G2, G4).

4.2 Ex-Situ Representation

We use a separate interface widget to provide ex-situ awareness that unobtrusively docks to the main visualization window. This widget is collapsible (Figure 1) and uses limited interface space (G3). Since we target general multidimensional data, we have two visualization designs—parallel coordinates plot and scatterplot—to show the team's presence and analysis coverage within this ex-situ widget (further studied in Section 5). These particular representations

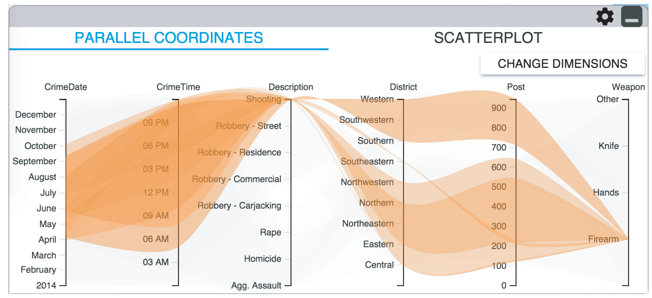

Figure 3: Parallel coordinates capturing coverage in a ex-situ widget in InsightDrive. To avoid clutter, the covered data points of each collaborator are clustered using hierarchical clustering into bands.

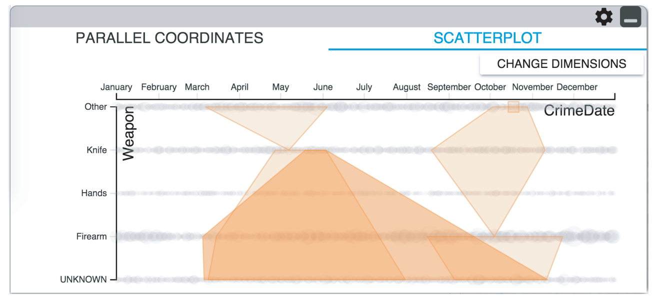

Figure 4: Scatterplot awareness on ex-situ widget in InsightsDrive. Two dimensions are chosen to create a scatterplot. The points viewed by a collaborator are clustered and shown as the regions.

can apply to any multidimensional dataset and are also extensible (G5). This widget can also provide methods for communication.

4.2.1 Ex-Situ: Parallel Coordinates

A parallel coordinates view (Figure 3) can represent all of the data points in the dataset and the respective coverage of each team member. We use agglomerative clustering to create bands [9] to quickly understand the covered data points (G2), while avoiding clutter (G3). Transparency of each band encodes the fraction of the total number of data points it contains. Hovering over a band highlights the encoded points on collaborators' interface [18]. Axes can be added, removed, and reordered for customization (G4). This parallel coordinates view makes it easy to see sequential selections based on the band transitions (e.g., user first selects a range on dimension X and then dimension Y). However, showing coverage by aggregation comes at a cost as individual point-level information is lost.

4.2.2 Ex-Situ: Scatterplot

While a scatterplot matrix can provide an overview of the data on all dimensions, SPLOMs yield high clutter and require significant display space. As an alternative to displaying all dimension combinations, we use a single scatterplot with editable axes to make the awareness widget compact (Figure 4). Again we use agglomerative (hierarchical) clustering to visualize clusters of covered points in the scatterplot as two-dimensional regions.

4.3 In-Situ Representation: Selection Shadows

We visualize the selections made by all other team members as "shadows" (Figure 5) in the background of each individual visualization (G2, G3). These selection shadows are coded with a unique color and label assigned to each collaborator. Shadows are adapted to the underlying visualization—appearing as borders to bars in bar charts and as colored regions in line charts and maps (G5).

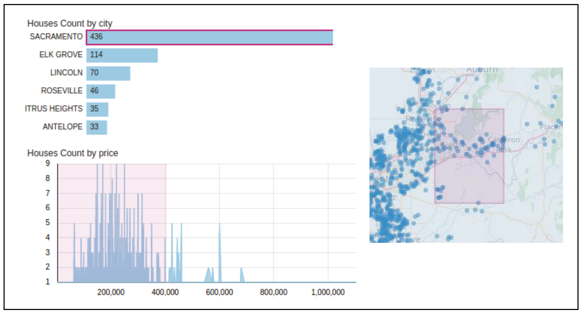

Figure 5: Selection shadows for in-situ awareness. Based on collaborative brushing [18], shadows show selections that collaborators have made as color-coded shapes in the background of each view.

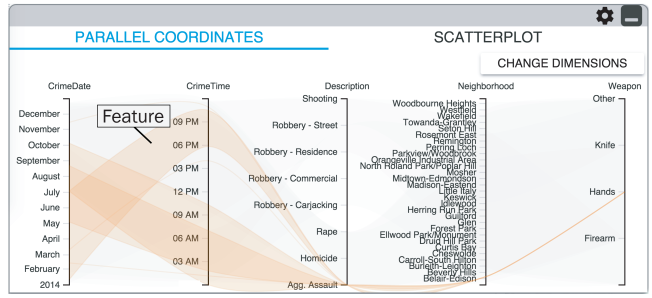

Figure 6: Example feature from the dataset covering Assaults with Hands in the Bel-air Edison neighborhood. It appears that crimes happening after 6pm mostly occurred in the first half of the year.

5 User Study 1: Utility of Ex-Situ Awareness

We conducted an exploratory user study to compare the utility of our ex-situ representations—parallel coordinates and scatterplots—to understand their affordances in conveying awareness.

5.1 Study Design

Participants. We recruited 6 participants (1 female, 5 male) between the ages of 18 and 45 from the student population within our university campus. They were paid $10 for participation. All participants self-reported as proficient computer users and as experienced with using visualizations for data analysis.

Dataset. We used a Baltimore crime dataset[3] that contains 11 attributes including date, time, location, description, and weapons used. We picked this dataset to enable investigative sensemaking by using questions related to trends and anomalies. Sessions were held in a lab setting using the InsightsDrive tool on a Google Chrome browser of a Macbook Pro (15-inch display; 1440×900 resolution).

Tasks and Protocol. Each task consisted of the participant following the awareness visualization (either parallel coordinates plot or scatterplot) while a VA expert (the study investigator) answered a question about the dataset. The participants were asked the speak out the observations (think-aloud protocol) they make from the awareness representation as the expert interacts with the interface to figure out the answer. The participants worked on eight tasks in the experiment: four with parallel coordinates and four with scatterplot (order counterbalanced across participants in the study). The motivation behind this methodology was to verify to what extent the participants can follow the activity of their collaborator in terms of presence and analysis coverage (attribute, range, and feature) just by viewing the ex-situ awareness representations. For this reason, the participants did not interact with the interface or the investigator directly during the experiment session. The candidate question list used for the tasks was generated by two VA experts using InsightsDrive. It consisted of questions related to four high-level visual analytics tasks: specific value identification, trend identification, extrema detection, and comparison of two data items. The list includes questions such as,

- What is the most common weapon used in April?
- In what neighborhoods do most shootings occur?
- During what time of the day did assaults with a firearm most happen in the central district?
- What do crimes happening in Downtown in the Spring and Fall seasons have in common?

Procedure. Participants first went through a training procedure where the assigned awareness representation (parallel coordinates or scatterplot) was demonstrated. They then proceeded with the tasks, and later repeated the process with the other awareness representation. Finally, they named their preferred awareness visualization. Each session lasted for less than 40 minutes.

The participants were asked to think-out-aloud during the tasks. Screen and audio recordings were captured for participants' answers as well as comments during the session.

5.2 Results and Observations

Here, we report the observations made based on how participants used the visualizations on our ex-situ widget in InsightsDrive.

5.2.1 Parallel Coordinates

All participants were able to easily identify which dimensions have been selected by just looking at the intersections of the bands with the axis of the parallel coordinates. The **attribute coverage** was the primary visual feature the participants observed after every expert interaction due to the change in shape of the bands (P3 described it as, "*when a dimension is selected, it appears like the free-flowing bands [that cover the entire space] are tied to a specific range [on a dimension]*"). All participants followed their observation of the attribute coverage with an observation of a **range coverage** aspect almost immediately. Typically, this was about the coverage over the date, time, district, crime description, and weapon dimensions.

Participants P3, P4, and P5 made complex observations related to **feature coverage** (Figure 6). For example, when the expert was viewing the street robberies, P4 remarked that there are a lot of crimes in the Southeastern and Central districts that happen after 9am in the morning. This specific feature is apparent due to our clustering approach. Beyond this, the participants could also sense the presence and attention of the collaborator based on the changes. However, a potential drawback (P2 and P6) was that the dimensional ordering in the parallel coordinates affected the perception of coverage. Overall, participants made more observations from parallel coordinates (2-4 per task) than scatterplot (1-2 per task).

5.2.2 Scatterplot

Participants typically took longer to interpret the scatterplot visualization due to the inherent need to switch dimensions to get a complete perspective of the coverage. This was expected from the use of a scatterplot as it can only capture coverage on two dimensions at once. Participants in this scenario focused on the range coverage (all) and feature coverage (P3, P5, P6). For example, when the expert was viewing crimes happening in the Fall months, P5 remarked that "*crimes are [evenly] distributed on the weapons dimension, but knife is more commonly used during September to December, while firearms for August to October*". We observed that the process of understanding the awareness on scatterplots can be viewed as the opposite of parallel coordinates. In parallel coordinates, the participants interpret the coverage top-down (e.g., by first examining

[3]https://data.baltimorecity.gov/

attribute coverage, then examining more specific details about the data space if possible). In contrast, they try to comprehend scatterplots bottom-up (e.g., by looking at individual data points first). This is because the participants had to look at 2D distributions and also explicitly switch between scatterplot dimensions, and therefore they made (range and feature) observations about the data on the current two dimensions first.

5.2.3 Subjective Feedback

All participants preferred parallel coordinates for the ex-situ group activity widget because, (1) it was harder to interpret clusters in scatterplots than bands on parallel coordinates, and (2) scatterplots require switching between dimensions.

6 User Study 2: Ex-Situ vs. Combination

InsightsDrive provides both in-situ shadows and ex-situ coverage widget as a balanced way to provide awareness. We were interested in observing the tradeoffs of the combination of in-situ and ex-situ over just ex-situ awareness on time and accuracy measures, when a team of analysts (participants) try to solve a practical visual analytics task involving decision making. Note that just having in-situ awareness by itself is not ideal for capturing presence and providing complete coverage on the dimensions since this can inundate each view with shadows and highlights based on the group activity, making it hard to follow. Hence this condition is not considered in the user study. Also, based on the previous study, we decided to use only the parallel coordinates plot for the ex-situ widget as it was the preferred visualization and led to more observations about the attribute, range, and feature coverage.

6.1 Study Design

Participants. We recruited 20 participants (6 female, 14 male) between the ages of 18 and 45 from the student population within our university campus. They were paid $10 for participation. All participants self-reported as proficient computer users and 18 of them had previously used visualization for data analysis. Participants were grouped into 10 teams based on their availability for the study. Participants in 9 teams knew each other, but only participants in one team worked with each other in a professional situation before.

Dataset and Apparatus. We used a simulated real estate dataset with 10 attributes including address, bedrooms, bathrooms, size, and price, as well as distances from closest school, shopping mall, university, and golf course. This dataset helped us develop simple relatable tasks that can be controlled for the study purposes. Participants worked in a lab setting similar to the previous study. During the user study session, participants sat opposite to each other without being able to see each other's displays. Beyond following the awareness representations, communication through speech was the only means for them to consolidate their work during the tasks. This choice replicates a distributed collaboration scenario in this study.

Tasks and Protocol. We used decision making tasks (four types) about real estate (house) search for the participants in our study. Each of these tasks involved giving a specific set of constraints (e.g., within 2 miles from a school) to each participant in a group and asking them to interact based on the constraints and coordinate with their collaborators to find the best choice.

- **Task 1 (T1):** Here, only one house in the dataset satisfies the constraints given to the participants. The participants would have to make appropriate selections based on their constraints and use the awareness visualizations to understand their collaborators' constraints. They then find the candidate houses on their interface based on their awareness of the group activity, discuss them with their collaborator, and pick a house.
- **Task 2 (T2):** There are multiple houses satisfying the constraints in this task. The participants follow a similar procedure as Task 1, but now they need to consolidate and pick one final house among the satisfying ones. We were interested in seeing how they would come to consensus and if it would change the performance.
- **Task 3 (T3):** There is no house satisfying the constraints in this task. Therefore, the participants need to negotiate to reach a compromise on some constraints to make a decision.
- **Task 4 (T4):** This task is similar to Task 3, but the participants are now aware of all the constraints (even the ones given to other participants in their group).

Example constraints include,

- Find a house within $200,000 price.
- Find a house within 5 miles from the closest school.
- Find a house with more than 2 bedrooms.

Participants worked on a total of eight tasks during the study: four (one per task type) with in-situ and ex-situ combination, and four with just the ex-situ awareness. For each task, groups of two participants worked as a team, along with a VA expert (the study administrator). The expert user added one more constraint to the task while encouraging the other participants to talk to each other. The expert user did not participate in the discussion between the two participants. This is a variant of the pair analytics protocol [1], modified for collaborative studies, giving the study administrator unfettered insight into the collaborative work. The time taken during each task from introducing the constraints to reaching a final consensus was measured. This represents the speed at which the participants become aware of the group activity and consolidate with the other, and thus captures the collaboration dynamics to an extent within this controlled setting for teams of two participants. The answers were also analyzed to evaluate their accuracy (as discussed in Section 6.3.2).

Experimental Factors. The awareness technique (T) and the task type (Q) are the factors influencing the group performance. For the awareness technique, we tested two conditions:

- EX+IN: This involved using the InsightsDrive multi-dimensional dashboard for the real estate dataset with both awareness techniques: in-situ shadows and ex-situ widget.
- EX: Only the ex-situ widget with parallel coordinates was used to gain a complete awareness of the group activity.

The order of tasks and conditions was counterbalanced.

Procedure. Participants first trained with the assigned visual analytics interface by demonstrating the visualizations, interactions, and awareness representations. They were given a set of training questions to answer and could return to the training again if needed. Following this, they worked on the four tasks with their group. They then moved on to the second awareness condition and repeated the same procedure. At the end of the session, they individually filled a questionnaire providing feedback about the perceived usability of the awareness representations for solving the tasks. The participants' comments and answers were audio recorded. Each session lasted for less than one hour.

6.2 Hypotheses

H1: Participants will be faster when both in-situ and ex-situ awareness is provided, since it can balance the participant attention between the actual interface (in-situ) and ex-situ components.

H2: Participants will be more accurate when both forms of awareness are provided as this may give a high-fidelity awareness. The in-situ representation in EX+IN captures the user interaction on the VA dashboard itself and can ensure that the collaborator does not miss any group activity due to split attention.

6.3 Results

Here we report the results from the statistical analysis of the time and accuracy measures collected during the sessions.

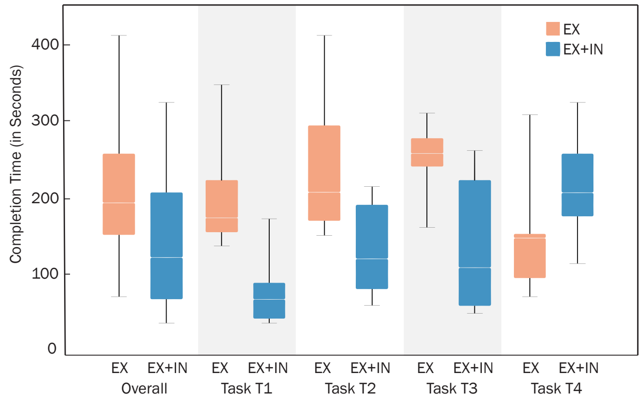

Figure 7: Differences between the task completion times. Statistical analyses revealed that having both forms of awareness (EX+IN) was faster than just ex-situ (EX) for tasks T1, T2, and T3.

6.3.1 Time

We first analyzed the time taken by the participants to solve the tasks for the two techniques and the four tasks using repeated-measures analysis of variance (Table 1). The technique had a significant effect, but an interaction between task and technique was also found to be significant. The combination of in-situ and ex-situ awareness (EX+IN) (M = 139 sec, SD = 79 sec) was faster than the ex-situ only condition (EX) (M = 207 sec, SD = 75 sec).

Table 1: Effects of technique (T) and task (Q) on time (repeated-measures ANOVA—all assumptions satisfied).

Factors	df, den	F	p
Awareness technique (T)	1, 80	23.83	$<.001$
Task type (Q)	3, 80	3.09	.033
T * Q	3, 80	9.71	$<.001$

We then analyzed the individual differences between the techniques for each task using paired T-tests. We found that the technique factor led to a significant difference in time for tasks T1 (t(9) = 3.79, $p = .004$), T2 (t(9) = 3.00, $p = .015$), and T3 (t(9) = 4.49, $p = .002$). For these tasks, the in-situ and ex-situ combination led to better performance (Figure 7). This **confirms** hypothesis **H1** for tasks T1, T2, and T3.

6.3.2 Accuracy (Distance)

Accuracy meant different things across the four tasks. For tasks T1 and T2, accuracy was the correctness of the decision made (whether the final house selected satisfied the constraints). All groups responded to these tasks correctly by picking the house that satisfies the constraints. Therefore, there was no difference across conditions for these tasks.

For T3 and T4, which do not have a correct answer, accuracy is based on the concession distance that defines how closely the selected house matched the constraints (similar to the one used by McGrath et al. [24]). This concession distance is defined as the normalized euclidean distance between the selected house and the boundaries of the collective constraints given to the group. For instance, for price range $\leq$ \$200,000 constraint, the boundary on the price attribute is \$200,000. During computation of this normalized distance, the attribute distances between the selected house and the constraints are scaled down by the overall range of the particular attribute in the entire dataset. For this reason, attributes in the dataset with higher values in general (e.g., price compared to distance) still have the same influence over the distance measure as others.

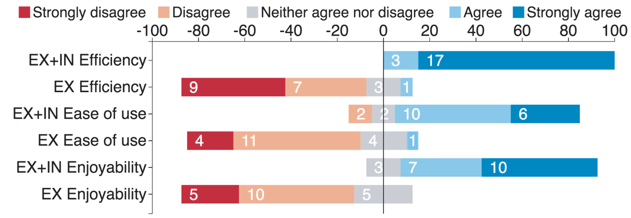

Figure 8: Differences between the condition with both forms of awareness (EX+IN) and ex-situ awareness only (EX) in terms of the Likert-scale ratings. Each bar in this chart captures the number of participants who gave the corresponding rating.

Repeated-measures analysis of variance applied to this measure revealed significant differences across two techniques based on the effects shown in Table 2.

Table 2: Effects of technique (T) and task (Q) on distance (accuracy) (repeated-measures ANOVA—all assumptions satisfied).

Factors	df, den	F	p
Awareness technique (T)	1, 40	6.14	.020
Task type (Q)	1, 40	5.87	.022
T * Q	1, 40	16.07	$< .001$

Paired T-tests applied to the distance for the individual tasks revealed significant differences only for T4 (t(9) = -7.73, $p < .001$). In T4, the normalized distance was higher for the ex-situ + in-situ condition (M = .40, SD = .07) than just ex-situ (M = .22, SD = .06). The differences were not significant for T3. Hypothesis **H2** is therefore **not confirmed**.

6.3.3 Subjective Ratings

The participants rated the awareness techniques on separate 5-point Likert scales for efficiency, ease of use, and enjoyability. This was analyzed using non-parametric Friedman tests and significant differences were found for all three scales (significance level: $p <$ .001). As evidenced in Figure 8, having both forms of awareness (EX+IN) was perceived to be more efficient, easy to use, and enjoyable than just ex-situ (EX). Almost all participants agreed to these questions for the condition with both forms of awareness, while disagreeing in case of ex-situ (EX) condition (Figure 8). Note that this questionnaire was given after the tasks on both awareness conditions were completed, so the responses are comparing the ex-situ technique to the combination of the ex-situ and in-situ techniques.

7 Discussion

Below, we reflect more broadly on the results of our studies, and present implications for the design of team-first CVA tools.

7.1 Explaining the Results

The user studies provided an understanding of the effectiveness of our awareness techniques. The first user study revealed the affordances of the parallel coordinates plot and scatterplot representations within the ex-situ awareness widget. The parallel coordinates plot conveyed all three forms of coverage with the attribute and range coverages more apparent to the participants and features tracked by a few. Overall, parallel coordinates plot led to more observations and was also preferred for tracking collaborator's coverage than scatterplots. This was because it was easier for participants to interpret 1D bands in parallel coordinates plot than 2D regions in the scatterplot (which also required switching dimensions).

In the second user study, the combination of in-situ shadows and ex-situ awareness with parallel coordinates proved to be faster and more easily usable than just the ex-situ awareness. This was in the context of visual analytics tasks that require the group to make decisions based on their visual exploration. These decision-making tasks (type 1-4) needed different levels of coordination between the participants as they needed to locate candidate houses, understand the constraints of others, and coordinate with them to find a common answer. For tasks T1-T3, they needed to understand the coverage and the specific selections of their collaborators to propose candidate choices. In the condition with both forms of awareness (EX+IN), participants found the in-situ shadows to be very useful in seamlessly revealing the collaborator's selection, and they then used the parallel coordinate plot to quickly understand the matches between their collaborator's and their own coverage. In contrast, participants spent more time interpreting the specific interactions along with the total dataset coverage on the just ex-situ condition (EX) for two reasons, (1) they need to follow multiple bands to interpret the selections as they occurred, and (2) the plots are inundated with bands as three users interact and they cannot focus on the actual interface while following the ex-situ widget.

There were also differences in the time spent between the four tasks used in the second study. For tasks T2 and T3, the added ambiguity in the final answer increased the time spent on the task compared to task T1. This explains the additional time needed for resolving ambiguity within the group, but the overall trend remains similar across the tasks T1-T3—the condition with both forms of awareness condition was faster (Figure 7). However, for task T4 this trend was reversed for accuracy measures. This is due to the fact that the participants knew the constraints of their collaborators up front, which eliminated the need to follow the in-situ representations in EX+IN condition and made them irrelevant.

7.2 Implications for Team-First Visual Analytics

The motivation for team-first design of CVA tools is to help the analysts focus on the actual sensemaking process while quickly following the group activity integrated into the visual interface. Our intention in this paper is to develop a CVA tool that follows an effective team-first design—thus helping the individuals and the team. Our claim is not that team-first tools are always better, but that there are team-first designs that can be beneficial in collaborative VA. As such, our specific instantiation through in-situ and ex-situ components proved to be advantageous for coordination within the group during decision making. In-situ selection shadows added an additional descriptive layer to existing data visualizations to show the focus of collaborators within the context of the visualization. This seems to create a good starting point to understand the collaborator's presence and attention. Coupled with this, the ex-situ widget provided expressive combinations of data coverage—attribute and range coverage were quickly interpreted from the parallel coordinates plot (Section 5)—to quickly come to a consensus. In the absence of in-situ, our ex-situ representation made understanding selections a heavyweight operation that also requires switching from the main interface; thus, slowing the team's workflow.

While InsightsDrive is created to be a team-first tool, by necessity it also contains analyst-first components (e.g., brushing-and-linking). This is because to support collaborative sensemaking, there first needs to be enough support for the sensemaking process of a single analyst. Within InsightsDrive, the brushing-and-linking interactions were used by the participants to get started with the constraints given to them. However, the strength of our team-first approach comes from its support for better team performance.

Overall, the complementary combination of in-situ and ex-situ awareness allows for quickly understanding the interactions and the coverage components, and improved the performance significantly for our decision-making task; thus, providing an exemplar design of team-first CVA. We are interested in exploring alternative team-first designs and theorizing the team-first VA paradigm in the future.

7.3 Limitations

We identify several limitations with our experiments and designs.

- The studies were conducted with specific awareness designs for multidimensional datasets in synchronous collaboration. As such, it is hard to generalize the findings to other settings.
- Our awareness designs are not built for supporting simultaneous activity from more than three analysts. To scale to larger groups, (1) aggregation techniques [9] need to be taken into consideration to avoid showing a multitude of bands in the ex-situ parallel coordinates, and (2) the number and intensity of the visual shadows in in-situ representations should be minimized based on the user's focus. In that case, more details about the group activity can be shown on demand (either when the user hovers over or selects specific views in the interface).
- The tasks chosen for our studies are not representative of all possible visual analytic tasks. However, we hope our results will initiate new research into studying other VA tasks.

8 Conclusion and Future Work

We have presented a team-first perspective into designing collaborative visual analytics tools, wherein the design of the tool focuses on supporting the needs of the team as a whole using representations of group activity. We have further demonstrated a concrete implementation of such an approach that provides an ex-situ awareness widget as well as an in-situ collaborative brushing technique. Our implementation records interactions across multiple collaborators and visualizes them using multiple awareness visualizations. Results from our user studies unraveled the affordances of our ex-situ representations in conveying awareness and also revealed that our particular balance between in-situ and ex-situ components for team-first design was effective for collaborative visual analytics. Future work in this space should focus on improved awareness visualizations, direct notifications, connections between insights, suggestions for organizing the group work, better coordination mechanisms, and team-first CVA tools for other collaboration settings.

Acknowledgments

This work was supported by the U.S. National Science Foundation award #1539534. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agency.

References

[1] R. Arias-Hernandez, L. T. Kaastra, T. M. Green, and B. Fisher. Pair analytics: Capturing reasoning processes in collaborative visual analytics. In *Proceedings of the Hawaii International Conference on System Sciences*, pages 1–10, 2011.

[2] S. K. Badam and N. Elmqvist. PolyChrome: A cross-device framework for collaborative web visualization. In *Proc. of the ACM Conference on Interactive Tabletops and Surfaces*, pages 109–118, 2014.

[3] R. M. Baecker. *Readings in Groupware and Computer-Supported Cooperative Work*. Morgan Kaufmann Publishers, San Francisco, 1993.

[4] D. Baker, D. Georgakopoulos, H. Schuster, A. Cassandra, and A. Cichocki. Providing customized process and situation awareness in the collaboration management infrastructure. In *Proceedings of the IFCIS International Conference on Cooperative Information Systems*, pages 79–91, 1999.

[5] A. D. Balakrishnan, S. R. Fussell, and S. Kiesler. Do visualizations improve synchronous remote collaboration? In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 1227–1236, 2008.

[6] H. Chung, C. North, J. Z. Self, S. L. Chu, and F. K. H. Quek. VisPorter: facilitating information sharing for collaborative sensemaking on multiple displays. *Personal and Ubiquitous Computing*, 18(5):1169–1186, 2014.

[7] H. H. Clark and S. E. Brennan. Grounding in communication. In L. B. Resnick, J. M. Levine, and S. D. Teasley, editors, *Perspectives*

on Socially Shared Cognition, pages 127–149. Amer. Psych. Assoc., 1991.

[8] P. Dourish and M. Chalmers. Running out of space: Models of information navigation. In *Proceedings of the Conference on Human Computer Interaction*, 1994.

[9] Y.-H. Fua, M. O. Ward, and E. A. Rundensteiner. Hierarchical parallel coordinates for exploration of large datasets. In *Proceedings of the IEEE Conference on Visualization*, pages 43–50, 1999.

[10] C. Gutwin and S. Greenberg. Design for individuals , design for groups: Tradeoffs between power and workspace awareness. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, pages 207–216, 1998.

[11] C. Gutwin and S. Greenberg. Effects of awareness support on groupware usability. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 511–518, 1998.

[12] C. Gutwin, M. Roseman, and S. Greenberg. A usability study of awareness widgets in a shared workspace groupware system. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, pages 258–267, 1996.

[13] A. H. Hajizadeh, M. Tory, and R. Leung. Supporting awareness through collaborative brushing and linking of tabular data. *IEEE Trans. Vis. Comput. Graphics*, 19(12):2189–2197, 2013.

[14] J. Heer and M. Agrawala. Design considerations for collaborative visual analytics. *Information Visualization*, 7(1):49–62, 2008.

[15] J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 1029–1038, 2007.

[16] P. Isenberg and S. Carpendale. Interactive tree comparison for co-located collaborative information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1232–1239, 2007.

[17] P. Isenberg, N. Elmqvist, J. Scholtz, D. Cernea, K.-L. Ma, and H. Hagen. Collaborative visualization: definition, challenges, and research agenda. *Information Visualization*, 10(4):310–326, 2011.

[18] P. Isenberg and D. Fisher. Collaborative brushing and linking for co-located visual analytics of document collections. *Computer Graphics Forum*, 28(3):1031–1038, 2009.

[19] P. Isenberg, A. Tang, and M. S. T. Carpendale. An exploratory study of visual information analysis. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 1217–1226, 2008.

[20] K. Kim, W. Javed, C. Williams, N. Elmqvist, and P. Irani. Hugin: A framework for awareness and coordination in mixed-presence collaborative information visualization. In *Proceedings of the ACM Conference on Interactive Tabletops and Surfaces*, pages 231–240, 2010.

[21] L. Laufer, P. Halacsy, and A. Somlai-Fischer. Prezi meeting: collaboration in a zoomable canvas based environment. In *Proceedings of the ACM Conference on Human factors in Computing Systems*, pages 749–752, 2011.

[22] N. Mahyar and M. Tory. Supporting communication and coordination in collaborative sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1633–1642, 2014.

[23] G. Mark and A. Kobsa. The effects of collaboration and system transparency on CIVE usage: an empirical study and model. *Presence: Teleoperators & Virtual Environments*, 14(1):60–80, 2005.

[24] W. McGrath, B. Bowman, D. C. McCallum, J. D. Hincapié-Ramos, N. Elmqvist, and P. Irani. Branch-explore-merge: facilitating real-time revision control in collaborative visual exploration. In *Proceedings of the ACM Conference on Interactive Tabletops and Surfaces*, pages 235–244, 2012.

[25] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of the International Conference on Intelligence Analysis*, volume 5, pages 2–4, 2005.

[26] A. C. Robinson. Collaborative synthesis of visual analytic results. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*, pages 67–74, 2008.

[27] A. Sarvghad and M. Tory. Exploiting analysis history to support collaborative data analysis. In *Proceedings of the Graphics Interface Conference*, pages 123–130, 2015.

[28] A. Sarvghad, M. Tory, and N. Mahyar. Visualizing dimension coverage to support exploratory analysis. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2016.

[29] A. Tang, C. Neustaedter, and S. Greenberg. VideoArms: embodiments for mixed presence groupware. In *Proceedings of the British Human-Computer Interaction Conference*, volume 20, pages 85–102, 2006.

[30] J. C. Tang. Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies*, 34(2):143–160, 1991.

[31] J. J. Thomas and K. A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press, 2005.

[32] M. Tobiasz, P. Isenberg, and S. Carpendale. Lark: Coordinating co-located collaboration with information visualization. *IEEE Trans. Vis. Comput. Graphics*, 15(6):1065–1072, 2009.

[33] F. B. Viégas, M. Wattenberg, F. van Ham, J. Kriss, and M. M. McKeon. ManyEyes: a site for visualization at internet scale. *IEEE Trans. Vis. Comput. Graphics*, 13(6):1121–1128, 2007.

[34] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, 2007.