Supporting Awareness through Collaborative Brushing and Linking of Tabular Data

Amir Hossein Hajizadeh, Melanie Tory, and Rock Leung

Abstract—Maintaining an awareness of collaborators' actions is critical during collaborative work, including during collaborative visualization activities. Particularly when collaborators are located at a distance, it is important to know what everyone is working on in order to avoid duplication of effort, share relevant results in a timely manner and build upon each other's results. Can a person's brushing actions provide an indication of their queries and interests in a data set? Can these actions be revealed to a collaborator without substantially disrupting their own independent work? We designed a study to answer these questions in the context of distributed collaborative visualization of tabular data. Participants in our study worked independently to answer questions about a tabular data set, while simultaneously viewing brushing actions of a fictitious collaborator, shown directly within a shared workspace. We compared three methods of presenting the collaborator's actions: brushing & linking (i.e. highlighting exactly what the collaborator would see), selection (i.e. showing only a selected item), and persistent selection (i.e. showing only selected items but having them persist for some time). Our results demonstrated that persistent selection enabled some awareness of the collaborator's activities while causing minimal interference with independent work. Other techniques were less effective at providing awareness, and brushing & linking caused substantial interference. These findings suggest promise for the idea of exploiting natural brushing actions to provide awareness in collaborative work.

Index Terms—Collaboration, awareness, attentionally ambient visualization, brushing and linking, linked views, user study

1 Introduction

In visual analytics, we aim to support the entire process surrounding data analysis and sensemaking [35]. This process typically includes not just interacting with data visualizations, but also recording notes and findings, synthesizing hypotheses and evidence, documenting findings for dissemination, and frequently, collaborating with other analysts. Evidence has shown that some visualization tasks can be completed faster and with greater accuracy in collaborative settings [16, 20, 23, 24]. For example, Mark et al. [24] found that groups were more accurate than individuals at searching large datasets and free data discovery tasks. Analysts often work in teams, where team members with different experience and expertise work together to solve challenging problems. It is therefore critical to support the mechanics of collaboration along with other aspects of the data analysis process [18]. We argue that the approaches needed to support collaborative visual analytics are somehow different from other types of collaboration, because people need to understand their collaborators' actions in relation to the data being investigated.

One critical aspect of an effective collaboration is awareness (i.e. keeping track of what one's team mates are working on, what insights they have found, and what avenues of investigation they have pursued) [15]. Awareness helps team members to share information with each other at appropriate times, and avoid redundant work, thereby improving overall team performance [16]. We propose to exploit people's natural brushing actions (a common interaction technique when working with a visualization containing linked views [22]) as one mechanism of providing awareness to a collaborator. We present results of an experimental study in which one user's brushing actions were presented to a remote collaborator. We investigate the extent to which these awareness cues interfere with individual work, and whether they can provide awareness of another person's activities.

Awareness is important in all forms of collaborative work, but is especially challenging to support during synchronous distributed collaboration (i.e. same time, different place) [9]. In synchronous collocated (i.e. same place) collaboration, there are some inherent awareness mechanisms. A person can look over a collaborator's shoulder to see what the other person is doing, can observe gestures and physical actions, and can talk to inform their team-mates of how they are progressing in their work.

In distributed collaboration though, many of these awareness cues disappear, and it becomes more important to provide explicit awareness support through software [9]. We presume that collaborators may be connected through audio and video links, and may have a shared screen. These forms of support may be sufficient for closely coupled work, when collaborators are working closely together (e.g. discussing their findings or planning the next step). However, they may be insufficient during periods of loosely coupled collaboration (e.g. where participants have divided up the work and spend an extended period working independently). It is well known that same-time collaborative work naturally transitions between close and loose coupling [34, 14], and that tools need to support many such collaboration styles and the transitions between them. One way to support loosely coupled work styles, plus timely transitions to more closely coupled work styles, is to help people maintain an awareness of what their collaborators are working on [20].

We focus on enhancing awareness during synchronous but loosely coupled work by distributed teams. We view this as the most challenging situation in terms of awareness, and therefore the most important to support through software. Many mechanisms designed to increase awareness between users use extra screen real-estate or will result in a cluttered screen that can interfere with user's work and distract them from their primary task. Any visual representation of what the other person is doing will add more information to a potentially busy screen. Such representations could be distracting and may also require users to learn a new visual encoding. We aim to develop an awareness mechanism that (1) is effective at indicating to an information "consumer" what a "producer" is investigating, (2) exploits people's natural interactions with a visualization tool (i.e. requires no extra effort on the part of the producer), (3) re-uses existing visual encodings within the tool so that no new encodings need to be learned, and (4) causes minimal interference with the consumer's own independent work.

Brushing and linking is a mechanism commonly used in visualization tools to help people explore relationships between data in different but related views [22, 28]. Brushing lets the user select some data on

• *Amir Hossein Hajizadeh is with University of Victoria. E-mail: hajizade@uvic.ca.*
• *Melanie Tory is with University of Victoria. E-mail: mtory@cs.uvic.ca.*
• *Rock Leung is with SAP. E-mail: rock.leung@sap.com.*

Manuscript received 31 March 2013; accepted 1 August 2013; posted online 13 October 2013; mailed on 4 October 2013.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

one chart and see how it is reflected on other linked charts, as illustrated in Figure 1. For example, a user may select a cluster of people with high life expectancy, and see how this subset of data distributes over gender or country of origin. Brushing is an extremely common interaction technique in modern visualization tools for tabular data, and an effective strategy for understanding relationships between data subsets and dimensions. The data items that a person chooses to brush should be strongly related to the person's questions and interests. Therefore, we suspected that people should be able to gain some knowledge of each other's interests and investigative questions based on seeing brushing actions. Showing brushing actions to a distributed collaborator may therefore provide a useful awareness cue.

We aim to answer the following research questions: (1) Can an individual gain awareness of a remote collaborator's activities by viewing a representation of their brushing actions? and (2) What visual representation of brushing actions will maximize awareness while minimizing interference with individual work?

Our research questions arose from a design challenge we encountered while developing a visualization tool for synchronous distributed (same time, different place) collaboration (Figure 2). Our goal with this tool was to explore design options to support collaborative business intelligence. In particular, the tool supported collaborative analysis of business data (e.g. sales data tables) by enabling analysts to create and examine multiple linked charts. While it was fairly straightforward to design shared and individual workspaces to support both closely coupled and loosely coupled activities, it was less clear how brushing should be handled. Should collaborators be able to see each other's brushing actions? If so, when? Should all details be shown or just limited detail? During closely coupled collaboration, it seemed clear that having an identical shared view (including brushing) could help collaborators to discuss and analyze data together. This is supported by previous work indicating that shared visualizations enable people to ground their discussions [16]. But the best design for loosely coupled collaboration was not obvious. Seeing another person's brushing actions could provide some awareness of what the other person is working on, but might also be distracting or otherwise interfere with individual work. What was the best approach to take?

We note that the general idea of collaborative brushing and linking has been previously introduced by Isenberg et al. [19]. This earlier work was inspirational to us and encouraged us to consider the possible value of collaborative brushing and linking. However, substantial differences in the task and data type made it non-obvious how to implement collaborative brushing in our situation. In particular, Isenberg et al. designed a tool for collaborative analysis of document collections rather than tabular data. In their work, "collaborative brushing and linking" was a feature explicitly added to the design for the purpose of supporting awareness. A visual indicator was shown to a collaborator when there was a common search term or a common document had been viewed or returned as part of a search. Individual user activities consisted of searching for and reading documents. In contrast, in our work, brushing and linking is the primary form of interaction for investigating data. We propose to further exploit those interactions for the secondary purpose of providing awareness.

We conducted an experimental study to compare three different methods of showing another person's brushing actions, as shown in Figure 3. Brushing & linking (top right) shows exactly what the collaborator sees while they brush the data (the selection plus corresponding items in all other views). Selection (Bottom left) shows only the collaborator's selected item. Persistent selection (Bottom right) is the same as selection, except that selected items remain highlighted for some time (40 seconds in our study) and gradually fade; thus multiple successive selections may be visible simultaneously. Our results showed that persistent selection was the most effective awareness mechanism, and that it can be implemented in a way that causes minimal interference with individual work.

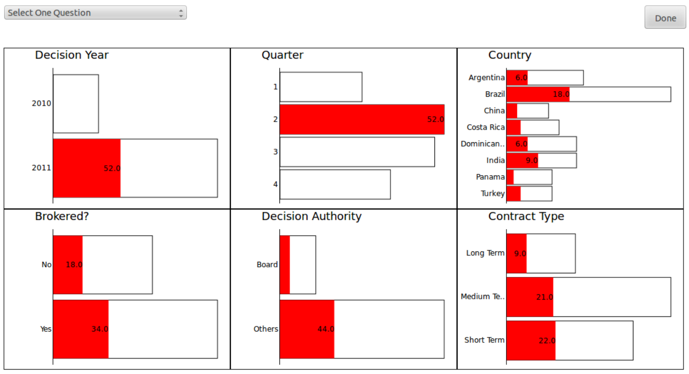

Fig. 1. Example of brushing and linking in the charts used for our study. Quarter 2 has been selected by the analyst in the top middle chart. Data corresponding to Quarter 2 are highlighted in all other charts. Partial bar highlighting indicates the proportion of the data that match Quarter 2.

Fig. 2. Screen shot of a distributed collaboration prototype we developed. Each user has their own work area where they can create charts. Shared parts of collaborators' workspaces may be viewed by selecting the appropriate tab.

2 Related Work

2.1 Awareness in Collaborative Work

Importance of establishing common ground between members of a collaborating team has been acknowledged in several areas, such as visual analytics [6, 7] and linguistics [8]. It is well known that common ground is a requirement for successful communication [6] and that even an imperfect shared knowledge will reduce cost of the communication between collaborators [16]. Weaver [36] mentioned that maintaining a visual common ground is the most difficult problem in collaborative visualization.

Heer and Agrawala [16] emphasized that the awareness of others' activities is an important source for common ground in collaborative work. Such awareness includes knowledge of what has been done and what are the possible next steps. Isenberg [20] found that collaborative brushing and linking as an awareness mechanism can avoid redundant work and empower analysts to validate and build upon each other's results.

In the context of collaborative visualization, awareness has been provided using a wide range of possible designs. At least four different categories of information can be shared among collaborators; what has been done, who has done it, when was it done, and why [33]. Based on the context and scenario a subset of these four categories will be presented to the user. There are also at least four different approaches to presenting awareness information. In some cases the indications are directly encoded in the shared workspace, e.g. using color to indicate changes [12, 19, 33]. Tam et al [33] also used text labels to represent type of action, while different colors were used to show "who" had

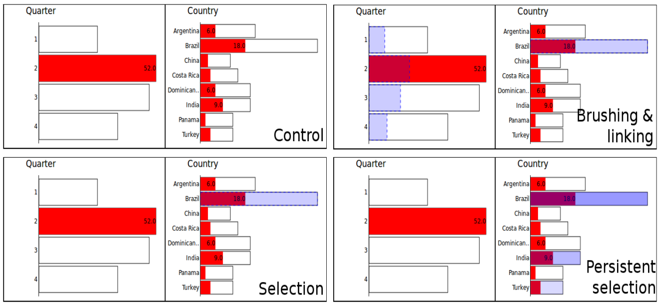

Fig. 3. Four collaborative brushing and linking conditions used in our study. In the study, participants saw six linked charts, but here only two are shown for simplicity. Red = own brushing, Blue = collaborator. (Top left) Control condition with no awareness mechanism, (Top right) Brushing & linking, (Bottom left) Selection, (Bottom right) Persistent selection.

changed it. Gross [12] used color darkness as the indication for the change, which faded away over time. This approach is similar to what we used in our persistent selection condition.

As a different approach, user activities can be aggregated and abstracted [16]. Brush et al. [4] have studied use of notification as an example of this approach. In some other designs a separate space is allocated to the awareness mechanism, such as timeline [10], or user activities list [17]. Finally, some awareness approaches rely on active user participation, for example through chat or comment threads [17] or annotations [21].

Awareness mechanisms have clear benefits for collaboration, but they do not come without costs. Grudin [13] states that every additional effort put into a collaborative tool should have direct values for its users. Also some studies (e.g. [11, 14]) have pointed to the cost introduced to collaborative effort by the communication media, including the means provided to keep participants awareness of each others' progress. McCrickard et al. [26] proposed a comprehensive model to understand benefits and costs in designing notifications and ambient systems. Their model tries to incorporate possible design goals (interruption, reaction and comprehension) in a successful system design. In our approach, there is no extra effort expended to producing the information since the brushing actions are part of the normal analysis process. However, there is potential for the visualization of those actions to be disruptive to collaborators in addition to providing awareness benefits.

Another potential cost of awareness mechanisms is screen real estate devoted to the awareness tool. Some evidence indicates that people are willing to sacrifice screen real estate for beneficial information, providing the information is useful and there is sufficient real estate to begin with. For example, a field study by Cadiz et al. [5] on a peripheral awareness interface indicated that people were willing to give up a portion of their screen to stay aware of important information related to daily tasks.

2.2 Attentionally Ambient Visualizations

Our visualizations of a collaborator's brushing actions are attentionally ambient; in other words, they are intentionally designed to be able to move in and out of a viewer's attentional focus [27] (e.g., by forming a secondary visual layer so that they can be ignored if desired). Pousman and Stasko define ambient information systems by several criteria: important but not critical information; can move from the periphery to the focus of attention and back again; subtle changes to reflect updates in information (should not be distracting); and are aesthetically pleasing and "environmentally appropriate." ([29, 30], e.g., [31]). Attentionally ambient visualizations extend the idea of ambient visualizations to an attention-based space rather than a location-based one. Closely related to ambient visualizations are peripheral visualizations (i.e. visualizations designed to be in one's peripheral vision).

As an example of early work in peripheral visualization, Bartram et al. [2] studied motion as visual encoding technique. They focused on how different motion parameters influence detectability and attentional distraction from a separate primary task. Results indicated that a slow linear motion can be a good compromise, with low distraction and high detection rate. These results suggested to us that showing a full representation of one's brushing and linking actions to a collaborator could be quite distracting since it may involve many rapid motion events; this might cause interference with primary task performance.

Later, Somervell et al. [32], conducted a similar study. They showed that density in peripheral visualizations and an attentionally demanding secondary task both have negative impacts on performance at a primary task. They also concluded that longer-lasting peripheral visualization are easier to interpret, which is in line with [2] and the hypotheses for our study. Persistence also improves comprehension and memorability of the awareness visualization [25].

3 Study Design

In our study, we compared three different ways a collaborator's actions can be visualized: brushing & linking, selection, and persistent selection (see Figure 3) in order to assess which provides the greatest awareness while causing minimal interference with individual work. To this aim, we asked participants to perform a primary analysis task, while simultaneously viewing an integrated visualization of what a "remote participant" was doing. Unbeknown to the participants, the remote participant was actually fictitious (implemented as a script, as described in the Apparatus section). At the end of the study session, we measured participants' awareness of the remote collaborator's activities by asking them to identify questions that were in common with the collaborator. We also assessed the visualization interference by measuring completion time of the task, accuracy of their answers, and perceived distraction.

We focused on quantitative measures because we saw this as a useful starting place for investigating feasibility of awareness mechanisms. Another advantage of our approach is that it is not subject to reporting biases; the awareness effects that we hoped to see are not in the focus of the user's attention during the tasks, and therefore might not be easily self-reported. Our hope was that we could identify a mechanism that provided a measurable awareness benefit while not causing substantial interference with individual work. After identifying such a mechanism, then future (perhaps more qualitative) studies could explore nuances of the design and influences on collaborative analysis in a more realistic setting.

The experiment used a between-subjects design with four conditions: a control condition (no visualization of the other person's actions) plus three visualization conditions (brushing & linking, selection, and persistent selection).

3.1 Participants

We recruited 40 participants among students of the University of Victoria. Participants had various backgrounds such as Engineering, Chemistry and Humanities. They were from different levels of their education: 26 undergraduate students and 14 graduate students. The average age for the participants was 21 years, and there were 20 females and 20 males among them. They were paid $10 each for their participation. These participants were randomly assigned to one of four different conditions, with 10 participants for each condition. None of the participants appeared to be familiar with the concept of brushing and linking.

3.2 Apparatus

Participants were told that a remote collaborator was sharing their screen and working simultaneously, but there was no audio or video link. The remote collaborator was actually simulated and implemented as a script. Prior to the study, we recorded the actions of an experimenter playing the role of a participant. During each study session (except for control conditions), these recorded actions were played back as an animation. Although technically the script was pre-recorded, this setup was designed to emulate synchronous, distributed collaboration. Participants believed that they were seeing the actions of a live collaborator.

This approach reduced the number of participants needed (no need to recruit pairs for each study session). More importantly, the behaviour of the collaborator was identical (i.e., controlled) in each study session and it could not influence the results. The length of the animation was 213 seconds, which covered the whole task time for 22 participants; of the remaining participants, 12 finished their tasks within one minute after the animation stopped.

We used a prototype we developed to enable brushing and linking for visual analysis tasks. Each user worked with a PC running the Ubuntu Linux operating system. The prototype was developed using Javascript language, employing jQuery [1] and D3 [3] libraries. Participants used Firefox 17 as the web browser to see and interact with the prototype. Screen resolution was 1920 x 1280. Participants interacted with the prototype using a mouse. Time spent on each question was recorded by the prototype.

3.3 Tasks

3.3.1 Primary task (data analysis)

Using our prototype, users could see six charts derived from a tabular data set. The six charts showed one single measure over six different dimensions (refer to Figure 1). We used US export statistics from US Census data. In the main phase of the study, the number of contracts was plotted against the following six dimensions using bar charts:

1. Year of export;
2. Quarter of export;
3. Destination of the export (the country);
4. Whether or not the contract was brokered;
5. Decision authority that was responsible for approving or declining the contract; and
6. Type of the contract (short-, medium-, or long-term).

The participants' primary task was to answer six analytical questions individually using the prototype. Two examples of the questions are "Which decision authority worked on more contracts in decision year 2011?" and "Which countries have more short term contracts than medium term ones?". Users could interact with the system to select one dimension value at a time, and see brushing of their selection on the other five charts (Figure 1). All of the questions could be answered using a single selection so multiple selection was not necessary.

3.3.2 Secondary task (awareness)

Participants in the three visualization conditions were also asked to perform a secondary task related to the remote collaborator's actions. These participants were told that they would be seeing some information about the brushing actions of another remote participant who was working on answering similar questions about the same data set using an identical screen. Thus participants would see their own brush actions as well as a visualization of the remote participant's brushing actions. We started by explaining about awareness and the reason it is important in a collaborative setting. We believed it was important for participants to know why they would be presented with a representation of another person's actions, since in real collaborations people would know this. However, we were careful not to draw too much attention to the secondary awareness task, because we were afraid that participants would focus too much of their attention on the collaborator's actions, whereas our design intention was for it to serve as background information, and one of our goals was to determine whether it disrupted primary data analysis work. Instructions explaining the situation were carefully designed, informed by pilot studies, to minimize influence on the performance of the primary task. Thus, apart from this information, participants were not told about the secondary task until the end of the study. At the end, they were informed that three out of six questions they had answered were in common between their task and their collaborator's. Then they were asked to guess which three questions were in common. The remote collaborator's actions were not shown to the control group so these participants were not asked to perform the secondary task.

3.4 Visualization Conditions

We used three different visualization techniques to show the remote collaborator's actions. Each is separately explained below.

3.4.1 Brushing & linking

One obvious way to keep one aware of what a remote person is doing on a shared display is to show an exact copy of their actions on the shared display. Our brushing & linking condition did this by visually representing both the remote participant's selected item plus any linked items (as the remote person would see them). To make the visualization distinguishable from the participant's own selection, we used different colors (red for one's own actions, blue for the collaborator's). To make the awareness visualization less disturbing we used 20% opacity. The color and opacity level were selected after trying various options. We aimed to find a configuration that was clearly visible and distinguishable from the main selection color (red), while also forming a secondary visual layer.

Figure 3 (Top right) shows an example of this visualization. The user has selected Quarter 2 from left chart, but simultaneously sees what the other user has selected (i.e. Brazil from right chart) as well as the linked items in other views. This visualization does not explicitly show which bar was selected. This information can be inferred because that bar will be fully colored, but we did not specifically instruct participants about how to do this.

3.4.2 Selection

In contrast to brushing & linking, which blindly mirrors what the remote collaborator sees, we can selectively choose what part is more informative to share among users. In order to understand what a person is interested in, we conjectured that the main selection is the most important part, and the linked items are secondary. Showing the selection only should result in a less cluttered view with more focused information. Therefore we chose our second visualization to visually represent only the selected item, without any linked items. Figure 3 (Bottom left) shows a snapshot of this visualization. In this condition one can precisely see what the collaborator has selected at the present time. The visualization updates when the selection changes or the remote collaborator deselects the currently selected item. The selection is visualized with 20% opacity.

3.4.3 Persistent selection

We assumed that limiting the visualization to the user's selection would make it less cluttered and more focused on important information. However, both the selection and brushing & linking conditions have a transient nature, so the user may miss some of the previous selections. Inspired by [20] and the results from [32], we also designed a persistent selection condition (Figure 3, bottom right). In this condition, the user saw current and previous items selected by the collaborator, ordered by their temporal occurrence. Opacity was used to show time since the selection was made. All selections except the current one faded out over time and eventually disappeared, and the current one started disappearing when it was deselected.

Using this visualization, a participant will not only see the most recent selection, but will also have information about earlier selections made by the collaborator. In addition to making it less likely that a selection will be missed, this approach puts the current selection in context. As an example, if the most recent selection is "Brazil", and we see that the previous selections are also countries, this is an indication of that the person is comparing countries regarding some factor.

In this condition the most recent selection had an opacity of 40%, and each decaying selection had its opacity reduced 1% every second. As a result, each former selection lasted for 40 seconds. During our study, this design meant that the last 4 or 5 selections were usually visible. Also, the questions that were used for the script of the fictitious collaborator were designed such that each required working on a different chart from the previous question; this minimized the overlap of consecutive selections.

3.5 Procedure

Before the start of the study, participants were asked to read and sign the consent form, and fill out a demographic questionnaire. In that questionnaire they were asked about their age, major and level of study, and also if they had experience with visualization applications or had participated in any project or experiment that included data or visual analysis tasks.

After filling in the questionnaire, the experimenter then presented the prototype to the participant and explained brushing and linking concepts using a sample data set (E-fashion). The experimenter worked through two sample questions with the participant to make sure they were comfortable with the concept of brushing and linking. Sample questions were chosen so that they covered different kinds of questions used in the primary task.

When the participants felt confident about brushing and linking, they were briefed about the actual dataset. Different dimensions of the data and their meaning were verbally explained to the participant. At this point, participants in visualization conditions (brushing & linking, selection and persistent selection) were told about their collaborator. The control group participants were informed about the presence of the collaborator, but because they did not see any representation of that person's actions no further information was provided to them.

Participants were told that the other person would be working on a "different" set of questions, which "may have" some questions in common with them, but would be randomly ordered. However, we also emphasized that they would be working independently of their collaborator. Here participants in visualization conditions (not control group), were instructed about the visualization assigned to them, as the mechanism of awareness.

At this point they were prepared to start the primary task. Participants had to select each question individually, work with the charts to find the answer and then write down their response on a blank paper provided to them. A drop down list containing six options, labelled as "Question 1" through "Question 6", was located on the top left corner of the screen (Figure 1). Selecting each of the options showed text of the question in the spacing between the drop down and the charts. At the start of the experiment none of the questions were visible to the participant and after making a selection, only one question was present on the screen at a time.

The questions had one to four word answers, where each word was a value in a specific dimension (e.g. "Quarter 1" or "Brazil, China, and India"). Participants clicked on a button labelled "done" (situated on the screen) after they had found the answer but before writing it down on the paper. Timing of the questions was done by the software, and included the time spent to read the question, but not time to write the answer.

After finishing the primary task, participants in visualization conditions were provided with a printed copy of the questions and were told that out of the six questions they had answered, three were in common with their collaborator. They were asked to state which questions they thought were in common by marking beside those questions. We suggested that they could refer to what they remembered from the other person's actions and see if they could relate the actions to one of the questions.

Finally, participants filled in the post-study questionnaire. The questionnaire was printed on a letter sized paper and handed to the participant. All groups answered a question about how comfortable they were about the primary task. There was also an optional field provided for additional comments. The three visualization groups also rated their perception of the distraction caused by the awareness mechanism on a five-level Likert scale.

3.6 Hypotheses

Our hypotheses for the study were as follows:

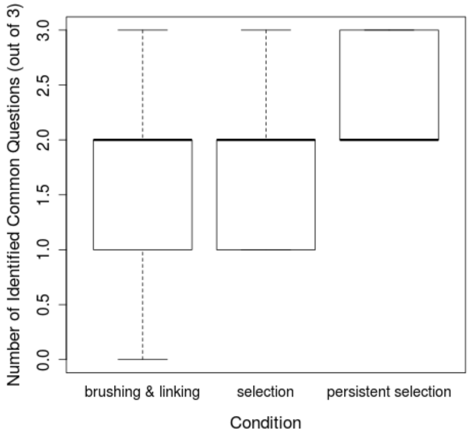

Fig. 4. Box plots of the number of correctly identified common questions in the secondary task (out of 3).

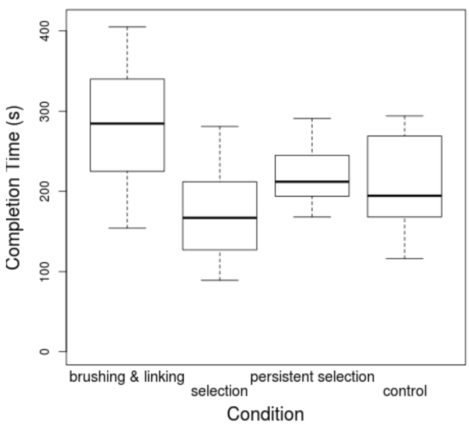

Fig. 5. Box plots showing primary task completion time for the four conditions.

1. All visualization techniques will be better than chance with respect to guessing the common questions (secondary task), but persistent selection will outperform brushing & linking and selection (because it is less transient and therefore less likely to be missed).

2. For completion time on the primary task, the order will be: control $<$ selection and persistent selection $<$ brushing & linking. We anticipate that the three visualization conditions may cause some interference with the time to complete tasks and brushing & linking will be the most interfering due to its visual complexity.

3. There will be no differences in the number of errors made in the primary task in visualization groups compared to the control group. We anticipate that people will still complete the tasks, but will just take longer if there is interference.

4. Brushing & linking will be rated as more distracting on the primary task than selection and persistent selection.

4 Results

After running the study, we analyzed the data using a number of statistical methods. We first checked whether our measures fit a normal distribution using Q-Q plots. Among the five measures, only completion time matched a normal distribution. Therefore we used ANOVA for analysing time, and the non-parametric Kruskal-Wallis test for other measures.

If a significant effect was identified in any of the measures, pairwise t-tests (for time) or pairwise Wilcoxon rank-sum tests (for all other measures) were used to investigate which pairs were significantly different. In all of these cases, we used Bonferroni correction to correct for multiple comparisons.

4.1 Awareness of Collaborator's Actions

We studied awareness of the remote collaborator's activities by counting the number of common questions correctly guessed by the participant (the secondary task). Participants had three questions in common with the remote collaborator, out of the six questions they were working on as their primary task. This measure has a range of 0 to 3, i.e. zero if none of the common questions is guessed correctly, to three if all common questions were correctly identified.

Figure 4 illustrates that the number of correctly identified common questions was higher with persistent selection than with brushing & linking or selection conditions. We analyzed the results by comparing the participants' scores to the score predicted by random chance (i.e. 50%). Although participants were not forced to guess all three common questions and in some cases they refused to guess any, correctly guessing the common questions may have happened by chance. Therefore each of the conditions was compared to a theoretical value of 50% (1.5 out of 3), using a Wilcoxon test. This analysis showed that persistent selection was significantly better than chance (V=55, $p < 0.01$), while brushing & linking (V=29, p=0.916) and selection (V=41.5, p=0.145) were not significantly better than chance.

We also compared the awareness measures between visualization conditions using the Kruskal-Wallis rank sum test with condition as a between-subject factor. We found a marginally significant effect of condition on awareness ($X^2(2)$=5.5, p = 0.061). Running a pair-wise Wilcoxon test with Bonferroni correction showed that persistent selection was more accurate than brushing & linking with marginal significance.

4.2 Primary Task Completion Time

One goal of our study was to determine if adding any of the selected visualizations would result in interference with the user's primary task. We chose time performance as the primary indication of interference. We assumed that if any of the visualizations interfered participants' work, it would result in a longer time to perform the tasks compared to the control condition.

As shown in Figure 5, completion time was highest with the brushing & linking condition, and lowest with the selection condition. The primary task included six questions, and each took a different amount of time to answer, ranging from 16 to 54 seconds. To account for the effect of question differences on completion time, we ran a two-way ANOVA on time with condition as a between-subjects factor and question as a within-subjects factor. We note that questions were not in random order; however, we do not consider this a problem because comparing between the questions is not the purpose of our analysis.

Results from the ANOVA showed significant main effects of condition ($F(3)$ = 5.2, $p < 0.01$) and question ($F(5)$ = 35.2, $p < 0.001$), but no significant interaction between question and condition.

After this step, we ran pair-wise t-tests on completion time with condition as the factor, employing Bonferroni correction. The results showed that selection was significantly faster than brushing & linking ($p < 0.01$) and the control group was faster than brushing & linking with marginal significance (p = 0.051). Other pairwise comparisons were not significantly different.

4.3 Primary Task Accuracy

Another measure related to interference is the number of correct answers, or the accuracy in answering the questions. Participants had to answer six questions, each having exactly one correct answer. Figure 6 summarizes the number of correctly answered questions, and illustrates that selection had somewhat higher average accuracy than the other conditions.

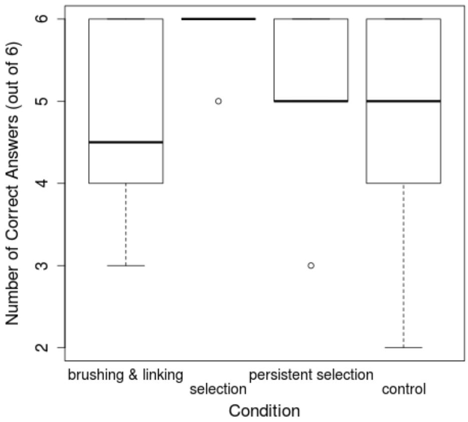

Fig. 6. Box plots showing primary task accuracy (number of correct answers, out of 6).

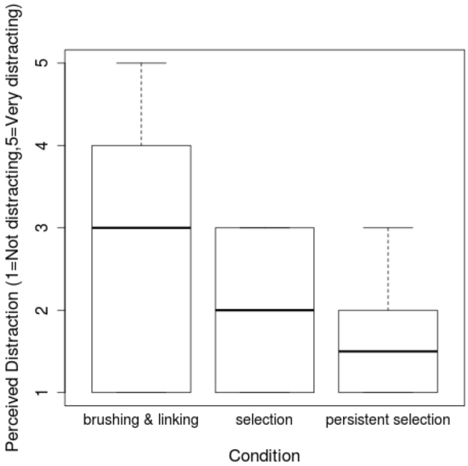

Fig. 7. Box plots of perceived distraction for the three visualization types. Distraction was measured on a 5-point Likert scale (1=Not distracting, 5=Very distracting).

The Kruskal-Wallis test showed a significant effect of condition ($X^2(3)$ = 8.4, p = 0.037). However, post hoc tests were only marginally significant ($p < 0.1$). Corrected pair-wise Wilcoxon tests showed that the selection group was marginally more accurate than the brushing & linking group (p = 0.098), and than persistent selection group (p = 0.051).

4.4 Perceived Distraction

Results of participants' distraction ratings are shown in Figure 7. Brushing & linking tended to be rated as more distracting than the other conditions. The Kruskal-Wallis test showed a marginally significant effect of condition on perceived distraction ($X^2(2)$ = 5.3, p = 0.071). However corrected pair-wise Wilcoxon tests did not show any significant differences between the three visualization conditions.

4.5 Qualitative Comments

During the experiments, we obtained interesting feedback from participants. These comments were usually either mentioned to the experimenter after the study, or written in the optional comment field in the post-study questionnaire.

Two participants mentioned that when they saw the other person working, they were encouraged to work faster. They somehow felt an urge to work faster to catch up with their collaborator. This effect was not an intended outcome of our design and was surprising to us. We note that the recorded animation had a duration (213 seconds) that was very close to, but slightly faster than, the average of participants' total time (218 seconds). It was also interesting that those two participants had a total time much lower than the average time for all participants and also for the participants in their condition. One of the participants was from the persistent selection condition and the other one belonged to the selection group.

Also, one participant reported using the awareness mechanism as a way to validate their own approach to the task. The participant stated that they observed the collaborator's method to make sure they were doing the task in the same way. As our participants were unfamiliar with the concept of brushing before participating in this study, this point demonstrates one possible advantage of awareness mechanism, i.e. assuring the users of their method if it is aligned with others, or correcting them in earlier stages if they are performing it wrong. Of course, this only has the desired effect if the collaborator uses an effective strategy; an inexperienced collaborator could have the opposite effect.

Participants' comments also indicated that our visual encoding was effective at making the awareness information attentionally ambient. A few participants in our brushing & linking condition commented that even though brushing was occupying a big portion of their screen and it was constantly changing, "it really didn't distract" them, "because of its subtle [blue] color and transparency."

5 Discussion

In summary, our findings demonstrate that persistent selection enabled the highest awareness of a collaborator's activities, while causing minimal interference with individual work. In this section, we first revisit each of our hypotheses to discuss the results in detail. Following this, we discuss more general implications of our results and design considerations.

5.1 Discussion of Hypotheses

H1: All visualization techniques will be better than chance with respect to guessing the common questions (secondary task), but persistent selection will outperform brushing & linking and selection.

This hypothesis was partially supported. Although average scores for both selection and persistent selection groups were higher than chance, our analysis showed only persistent selection to be significantly better than chance. This confirms our hypothesis that persistent selection would provide the best level of awareness. However, it is uncertain whether the other two techniques provided any awareness benefit.

H2: For primary task completion time, the order will be: control $<$ selection & persistent selection $<$ brushing & linking.

This hypothesis was partially supported. We expected all three visualization conditions to be disruptive and therefore the participants in those groups to be slower than the control group. However, our results showed that only brushing & linking was slower than the control group (with marginal significance). This supports our conjecture that the visual clutter and motion caused by the collaborator's actions in the brushing & linking condition would interfere with individual work. Although some participants in the brushing & linking condition reported that they did not find the collaborator's activities disruptive (due to good color choices), the cluttered nature of this condition nonetheless appeared to cause interference. The fact that our other visualizations took no longer than the control is very promising; it suggests that some of the collaborator's actions can be shared without negatively influencing individual performance. It is worth noting that the average completion time for the selection condition was surprisingly faster than the control group, but this difference was not statistically significant.

Selection and persistent selection were faster than brushing & linking (significant for selection). This demonstrates that simply showing everything that a collaborator sees is not the best solution. A selected representation of a collaborator's actions can reduce interference while providing the same or better levels of awareness.

H3: There will be no differences in the number of errors in visualization groups compared to the control group (primary task).

Looking at accuracy of answers, we see that selection was better than brushing & linking and persistent selection (with marginal significance). However none of the visualization conditions were significantly different than the control group. Although we cannot statistically conclude that the conditions had identical effects on accuracy (and it is unlikely that they did), our findings are generally in line with our hypothesis that visualization conditions would not be worse than the control group in terms of answer accuracy. Thus, none of the visualizations appeared to significantly interfere with performing the primary task in terms of accuracy.

H4: Brushing & linking will be rated as more distracting than selection and persistent selection.

This hypothesis was not supported. Although brushing & linking had a higher average distraction rating, this effect was not significant. This might be explained by participants' comments that the color was subtle enough that they could focus on their work and ignore the awareness representations completely.

5.2 Implications for Design

Our results demonstrate that visualizing brushing actions in an attentionally ambient way can enable loosely coupled collaborators to gain awareness of each other's activities within a shared display space. The most effective design for the awareness mechanism is likely to be selective in what is shown (i.e. show only the collaborator's selections, not everything they see) and include some temporal persistence of the awareness visualization (i.e. so that the collaborator's selections do not immediately disappear when the collaborator selects something new).

We found that a cluttered awareness visualization, like the brushing & linking condition, causes interference with a user's primary task. Therefore, selective filtering of the visualization is required. In our study, brushing & linking seemed to be showing more than what was needed for the participant to answer the secondary task. However, what makes an effective summary may be task-specific, and we should take this into consideration when designing attentionally ambient awareness mechanisms for other tasks. We may also need to visualize awareness information with different levels of detail depending on the phase of collaboration and role of the awareness information. Most importantly, we expect that it is important to show more detailed information about brushing & linking during closely coupled work (to support conversations about the data), but to show less detailed information (only selections) when participants work in a loosely coupled fashion.

Persistent selection was the only condition that performed significantly better than chance in informing the participants of their collaborator's activities. This also supports previous findings on peripheral visualization [32] and persistent brushing [20], which suggest that a peripheral visualization must persist to enable it to be noticed. Therefore, in designing peripheral visualization for secondary tasks we could consider slowing down the decay rate of the indicators in order to give more time for the user to notice and switch between their primary and secondary tasks. This may be especially critical when the primary task is attention-consuming or time-constrained.

Finally, our persistent selection visualization, which only shows the selected items, could conceivably be applied to a wider range of actions beyond brushing, as long as the actions involve selecting dimensions or values. It might be beneficial to visually represent the type of action performed, or to represent positive indications of interest in an item (e.g. brushing) differently from indications of non-interest (e.g. filtering).

Brushing, filtering, and other data-centric actions have another benefit: they are tied to the underlying data and therefore can be understood by the machine [16]. Therefore, it may be possible to exploit this attribute to propose intelligent methods to integrate and aggregate users' actions that likely will reduce the amount of data shown to a collaborator.

5.3 Design Details for Further Study

Some design details should be further investigated before directly deploying our visualization approaches to similar collaborative situations. As an example, for our persistent condition, we selected a fall-off speed based on pilot studies with our fictitious collaborator data. The optimal fall-off speed for a broader range of people and situations would likely be different and would need to be determined. Most importantly, the fall-off speed needs to be investigated in the context of a longer, more in-depth data analysis session, since our study sessions were very short. A simple speed-based solution may actually not be optimal either. Developing an intelligent algorithm to identify groups of related actions and set the decay rate based on them may turn out to be a more valuable approach.

Alternative color palettes and transparency levels should be considered too. It should be possible to increase the transparency level, which would reduce the visual salience of the background visualization, while still keeping it detectable. Additionally, our visualization uses the colour channel to distinguish one's own brushing and linking actions from those of the collaborator. In visualizations that use colour coding for another purpose, this is potentially problematic. For single users, visualizations typically either use saturation to indicate selections, or reserve one colour as the "selection colour". Our awareness visualizations may require reserving two colours instead of one (reducing the space of colours that can be used for colour coding) or using three levels of saturation (probably most saturated for one's own selections, second most saturated for the collaborator's selections, and least saturated for non-selected items). Future studies should investigate how these encoding schemes interact with color coded data.

5.4 Limitations and Future Work

Our study should be viewed as a starting point that shows promise for the idea of providing awareness through visualization of brushing actions. However, substantial additional research is needed to explore how to apply this idea in a more diverse range of situations.

To begin with, our study examined single selection (i.e. selecting exactly one bar). This was a logical place to start, but future work should consider the effects of multiple selections. Multiple selections will increase the amount of data shown to a collaborator for selection and especially persistent selection views. The amount of clutter may not be a problem in real practice, or might be mitigated by intelligently aggregating or summarizing the selections, but this issue needs further examination. In this case, we also need to devise ways to visually represent repeated selections in the persistent visualization (e.g. if a person selects Canada & Brazil, followed by Canada & Germany).

We also focused on bar charts, whereas in a more realistic situation analysts use a variety of chart types, such as time series plots and scatter plots. In addition, although our study did involve multiple linked views, analysts are likely to use a larger and more diverse collection of views. We focused on bar charts for simplicity and consistency in the study design. We believe it should be possible to extend the idea to other types of charts, but finding an effective representation of selections will require some design exploration. Since persistent selection was most effective in bar charts, we would like to extend this technique to a variety of chart types and ensure that it can be understood. Multiple linked views (beyond the six in our study) will add an additional level of complexity, and we should investigate how this influences the effectiveness of persistent selection. In practice, we would also expect the two collaborators to be simultaneously examining different views of the same data set (e.g., different charts). This is unlikely to be a big problem for persistent selection as long as the collaborator's selected items can be shown in some currently visible plot. However, when the collaborator's selected items are not included in any of the visible charts, a different approach will be needed. For that case we may want to highlight the relevant dimension in a visible list of dimensions or add other informative indicators. It will also be useful to explore how to present other actions that may indicate interest in the data, such as filtering and aggregation.

Evaluating the persistent selection idea in a more realistic context will also require recruiting professional analysts as participants. Since our study focused on basic issues of perceptibility and interference of the awareness visualizations, we were able to use fairly simple tasks. Our student participants seemed comfortable with these tasks and were fairly successful at answering the questions. However, more complex tasks would likely be approached differently by professional analysts as compared to students.

Our measure of awareness, i.e. number of guessed common questions, showed promising results that people can gain awareness through visualization of each other's brushing actions. We emphasize that even though participants were not informed of the secondary task before the study, those in the persistent selection group were able to use what they recalled, to guess the common questions effectively. Devising a quantitative measure of awareness was actually very challenging. Awareness is a somewhat "fuzzy" concept, and it is not clear what attributes would best indicate that a person is "more" or "less" aware. Our common question strategy provides one quantitative metric that we consider useful. However, additional measures of awareness could improve the precision of awareness measurements or add diversity by including other aspects of awareness. For example, it would be useful to have a more holistic way to measure a participant's understanding of their collaborator's activities, or to assess their understanding of a collaborator's higher-level strategy or work process.

As our own future work we would like to consider situations where there are more users involved, they work in both loosely and closely coupled styles, and do not share an identical workspace. We would also consider how to apply persistent selection to other types of charts and how to visually represent multiple selections.

6 Conclusion

We compared three methods of visualizing a collaborator's brushing actions on a shared multi-view visualization (brushing & linking, selection, and persistent selection), as a way of providing awareness of the collaborator's activities. We measured which design provides the best level of awareness while maintaining minimal interference with a participant's primary task.

Our experiment demonstrates that people can gain awareness of a collaborator's activities by viewing the collaborator's selections as they investigate data through brushing and linking. In particular, the persistent selection visualization technique enabled people to perform better than chance on a secondary awareness task, while causing minimal interference with their primary tasks. Other techniques (selection, brushing & linking) were less effective at providing awareness, and brushing & linking caused significant interference. Our results reveal that there is value to revealing data selections to collaborators via an attentionally ambient visualization, even when collaborators are working in a loosely coupled fashion. We also demonstrated that only the selection needs to be shown to support awareness, and that persistence of the selection is important to increase the likelihood that it will be noticed.

Our results show promise for exploiting analysts' natural brushing actions to build a collaborative awareness mechanism. Future research can build upon these results to design effective awareness mechanisms for collaborative visualization. Other design possibilities can be explored to elevate the level of awareness or to extend it to other situations and different types of analysis.

Acknowledgments

We would like to thank our colleagues in University of Victoria VisID lab and SAP for their support and invaluable feedback. This work was supported by grants from SAP and the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] http://jquery.com.

[2] L. Bartram, C. Ware, and T. Calvert. Moving icons: detection and distraction. In *Proc. IFIP TC*, volume 13, pages 157–165, 2001.

[3] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2301–2309, 2011.

[4] A. Brush, D. Bargeron, J. Grudin, and A. Gupta. Notification for shared annotation of digital documents. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, pages 89–96. ACM, 2002.

[5] J. Cadiz, G. D. Venolia, G. Jancke, and A. Gupta. Sideshow: Providing peripheral awareness of important information. *Microsoft Research, Collaboration, and Multimedia Group, Tech. Rep*, 2001.

[6] Y. Chen, J. Alsakran, S. Barlowe, J. Yang, and Y. Zhao. Supporting effective common ground construction in asynchronous collaborative visual analytics. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 101–110. IEEE, 2011.

[7] M. C. Chuah and S. F. Roth. Visualizing common ground. In *Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on*, pages 365–372. IEEE, 2003.

[8] H. H. Clark and S. E. Brennan. Grounding in communication. *Perspectives on socially shared cognition*, 13(1991):127–149, 1991.

[9] P. Dourish and V. Bellotti. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pages 107–114. ACM, 1992.

[10] C. H. Ganoe, J. P. Somervell, D. C. Neale, P. L. Isenhour, J. M. Carroll, M. B. Rosson, and D. S. McCrickard. Classroom BRIDGE: using collaborative public and desktop timelines to support activity awareness. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 21–30. ACM, 2003.

[11] D. Gergle, R. E. Kraut, and S. R. Fussell. Language efficiency and visual technology minimizing collaborative effort with visual information. *Journal of Language and Social Psychology*, 23(4):491–517, 2004.

[12] T. Gross, W. Wirsam, and W. Graether. Awarenessmaps: visualizing awareness in shared workspaces. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems*, pages 784–785. ACM, 2003.

[13] J. Grudin. Why cscw applications fail: problems in the design and evaluationof organizational interfaces. In *Proceedings of the 1988 ACM conference on Computer-supported cooperative work*, pages 85–93. ACM, 1988.

[14] C. Gutwin and S. Greenberg. Design for individuals, design for groups: tradeoffs between power and workspace awareness. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 207–216. ACM, 1998.

[15] C. Gutwin and S. Greenberg. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)*, 11(3-4):411–446, 2002.

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

[17] J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and voyeurs: supporting asynchronous collaborative visualization. *Communications of the ACM*, 52(1):87–97, 2009.

[18] 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.

[19] P. Isenberg and D. Fisher. Collaborative brushing and linking for co-located visual analytics of document collections. In *Computer Graphics Forum*, volume 28, pages 1031–1038. Wiley Online Library, 2009.

[20] P. Isenberg, D. Fisher, M. R. Morris, K. Inkpen, and M. Czerwinski. An exploratory study of co-located collaborative visual analytics around a tabletop display. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 179–186. IEEE, 2010.

[21] R. Kang and S. Kiesler. Do collaborators' annotations help or hurt asynchronous analysis. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion*, pages 123–126. ACM, 2012.

[22] Q. Li and C. North. Empirical comparison of dynamic query sliders and brushing histograms. In *Information Visualization, 2003. INFOVIS 2003. IEEE Symposium on*, pages 147–153. IEEE, 2003.

[23] N. Mahyar, A. Sarvghad, M. Tory, and T. Weeres. Observations of record-keeping in co-located collaborative analysis. *HICSS 2013*, Jan 2013.

[24] 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.

[25] D. S. McCrickard, R. Catrambone, and J. T. Stasko. Evaluating animation in the periphery as a mechanism for maintaining awareness. In *Proceedings of the IFIP TC. 13 Conference on Human Computer Interaction*, pages 148–156, 2001.

[26] D. S. McCrickard and C. Chewar. Attuning notification design to user goals and attention costs. *Communications of the ACM*, 46(3):67–72, 2003.

[27] A. V. Moere. Towards designing persuasive ambient visualization. In *Issues in the Design & Evaluation of Ambient Information Systems Workshop*, pages 48–52. Citeseer, 2007.

[28] C. North and B. Shneiderman. Snap-together visualization: can users construct and operate coordinated visualizations? *International Journal of Human-Computer Studies*, 53(5):715–739, 2000.

[29] Z. Pousman and J. Stasko. A taxonomy of ambient information systems: four patterns of design. In *Proceedings of the working conference on Advanced visual interfaces*, pages 67–74. ACM, 2006.

[30] Z. Pousman, J. T. Stasko, and M. Mateas. Casual information visualization: depictions of data in everyday life. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1145–1152, 2007.

[31] J. Rodgers and L. Bartram. Exploring ambient and artistic visualization for residential energy use feedback. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2489–2497, 2011.

[32] J. Somervell, D. S. McCrickard, C. North, and M. Shukla. An evaluation of information visualization in attention-limited environments. In *Proceedings of the symposium on Data Visualisation*, volume 2002. Citeseer, 2002.

[33] J. Tam and S. Greenberg. A framework for asynchronous change awareness in collaborative documents and workspaces. *International Journal of Human-Computer Studies*, 64(7):583–598, 2006.

[34] A. Tang, M. Tory, B. Po, P. Neumann, and S. Carpendale. Collaborative coupling over tabletop displays. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1181–1190. ACM, 2006.

[35] S. van den Elzen and J. J. van Wijk. Small multiples, large singles: A new approach for visual data exploration. In *Computer Graphics Forum*, volume 32, pages 191–200. Wiley Online Library, 2013.

[36] C. Weaver. Is coordination a means to collaboration? In *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on*, pages 80–84. IEEE, 2007.