International Journal of Data Analytics
Volume 1 • Issue 2 • July-December 2020

The Impact of Utilizing a Large High-Resolution Display on the Analytical Process for Visual Histories

Haeyong Chung, University of Alabama in Huntsville, USA
https://orcid.org/0000-0001-5168-1190

Andrey Esakia, Virginia Tech, USA

Eric Ragan, University of Florida, USA

ABSTRACT

Visual history can be helpful in building awareness of various investigative documents exemplified by active visual artifacts that encourage the growth of and access to different threads of investigation. On the downside, however, with the growth of an analysis history, it becomes more difficult to keep track of the workflow and the decision-making processes throughout the analysis. This article explores the concept of supporting visual history through branching functionality for analyzing a cybersecurity dataset in a spreadsheet format using a large high-resolution display (LHRD). To support the findings, the authors conducted a qualitative study to investigate the effect of screen size on the analytical process and the support of visual history. A comparison of participants' analytic processes found differences between the two different display setups and revealed that the LHRD participants tended to take advantage of the visual history spatially during their analytical investigations.

KEYWORDS

Cybersecurity, Large High-Resolution Displays, Physical Navigation, Provenance, Visual Analytics, Visual History

1. INTRODUCTION

Analysts across different domains are often tasked with complex exploratory investigations involving large amounts of data. For example, digital forensics analysts often examine data to identify threats among interconnected pieces of evidence from various sources (Pirolli & Card, 2005). Biologists analyze data through iterative cycles of computational analysis, visualization generation, and hypothesis testing (Li et al., 2011). In the domain of cybersecurity, digital forensics often requires analysis of large amounts of digital storage and records of network activity to investigate suspicious behaviors and accurately identify criminal activity (Goodall & Tesone, 2009). For such analytic tasks in real-world situations, analysts commonly rely on a multitude of conventional tools such as Excel, web browsers, text editors, and command-line tools as their main tools due to their complex task requirements and lack of support for rapid foraging across different applications (Endert, Andrews,

DOI: 10.4018/IJDA.2020070106

Fink, & North, 2009). A shared challenge in analysis work is keeping track of process history (Madanagopal, Ragan, and Benjamin, 2019).

Although many general tools have well-determined purposes for analysis, keeping track of the history of the analysis process is difficult due to the complex nature of hypothesis exploration. Specifically, analysts often undertake difficult investigations leading to frequent acts of hypothesis development and testing. As such, it may be difficult to keep track of the many logic paths along with the corresponding data views created via different tools. Overload from various analysis artifacts and data views can make it difficult for the analysts to recall and explain how they arrived at a certain state in the investigation. Such challenges with reporting or quality control can affect the entire investigation (Madanagopal, Ragan, and Benjamin, 2019). Furthermore, the reduced awareness of the analysis artifacts can lead to redundant or unproductive work due to the inability to notice and reuse earlier findings. A survey of workplace practices shows analysts often resort to substandard solutions such as saving numbered versions of files as a means of preserving the history of the process (Fink, North, Endert, & Rose, 2009). Such an approach for maintaining history items can lead to even more confusion because the large number of files eventually becomes meaningless and makes the history of the investigation process virtually irretrievable.

To alleviate these problems in data analytics and sensemaking with visual histories, researchers and developers have designed workflow and provenance tools to help manage, capture and revisit the analytic processes, e.g., (Wenwen et al., 2009; Silva, Freire, & Callahan, 2007; Bavoil et al., 2005). These tools help keep track of the steps taken to create different data views throughout the progression of exploratory analysis to make it easier for analysts to review their work. Many provenance tools represent the steps of the analytic process visually, as is often referred to as a **visual history**, e.g., (Singh et al., 2011; Gotz & Zhou, 2009; Heer & Shneiderman, 2012). Visual history applications take advantage of the large number of **history items** indicating snapshot views denoting process milestones and separate lines of investigation, e.g., (Bavoil et al., 2005; Li et al., 2011). By spatially organizing visual history items that pertain to different analysis tasks or datasets on the screen, an analyst is afforded with a visual history that provides an easy to access record of the workspace and analytic process in real time, revealing patterns of branching histories of data items. The core objective for the visual history in the context of the analyst workspace is to minimize any hinderance to task performance caused by the burden of revisiting previous data states. The applications supporting visual histories leave trails of **process landmarks** as representations of key decisions, divergences in the process, or changes to the data. Prior studies have found that even the final state of the analysis workspace can serve as visual history to enable significant improvements to memory and recall of the analysis process (Ragan, Goodall, & Tung, 2015). Previously, Singh et al. (2011) explored the potential benefits of large high-resolution displays for maintaining visual histories. The authors explored the utility of additional display space for supporting persistent visual history; however, they did not evaluate the effects of workstation space on the process of generating the visual history.

In this paper, we investigate whether display space affects the ability to represent visual histories of analytic provenance in the context of digital forensics, extending Singh et al.'s prior work (2011). We define visual history as the state of analyst's workspace where a visible record of data states is active, fully visible, and immediately accessible throughout the duration of the analysis. LHRDs enable users to see more information and detail by offering wide fields of view, large physical screen space, and high DPI (dot per inch) capacity, and supporting natural, physical navigation of the information space (Ball, North, & Bowman, 2007; Yost, Haciahmetoglu, & North, 2007). We examine whether having the ability to create process history with live and active data views on LHRDs can change the analytic process. To explore this concept, we created an Excel add-in that simplifies the creation of visual histories supporting branching history represented by multiple spreadsheets (i.e., data files). As a way to explore the effectiveness of the tool, we conducted a study with eight participants that performed a digital-evidence analytics task. The study compares how analysis processes are affected by differences in large and small-screen display setups. Based on the findings from the study, this paper also proposes a set of features aimed at enhancing visual history tools for LHRD.

2. RELATED WORK

Our work shares inspiration with prior research in analytic provenance, visual analytics, and cyber analytics. Ample display space makes it possible to capture the analytic reasoning process and represent it as a visual history of the analyst's actions. Many visualizations have been presented to help cyber analysts keep track of the analytical process history via solutions such as chronological ordered thumbnails each capturing the state of the analytical process, e.g., (Dunsmuir et al., 2010; Heer, Mackinlay, Stolte, & Agrawala, 2008; Heer & Shneiderman, 2012; Jankun-Kelly, Kwan-Liu, & Gertz, 2007; Kreuseler, Nocke, & Schumann, 2004; Kwan-Liu, 1999; Lee & Grinstein, 1995; Li et al., 2011). However, similar to the numbered file-naming strategy, chronological iconic history requires extra effort to decode the meaning of the visualization in the context of an ongoing investigation.

Providing a more streamlined approach, Bavoil et al. (2005) presented *VisTrails*, a provenance management system for exploratory computational tasks. The system provides detailed provenance information that allows users to understand how scientific computation and visualizations progress over time. *Analytic provenance* facilitates the reasoning process and streamlines workflow in order to reproduce and share the analysis results (Wenwen et al., 2009; Silva et al., 2007). Furthermore, analytic provenance emphasizes the analytic process through visualization and interaction in order to promote insight (North et al., 2011; Ragan et al., 2015). The sensemaking process originates from the analytic provenance via repeated generation and testing of hypotheses that in turn arise from the analytic investigation. Kadivar et al. (2009) introduced a visual analytics tool (*CzSaw*) that monitors user interactions and translates them into a set of scripts. The scripts are also interpreted to produce visual history of the process. The history is visualized by a chronological set of screenshots that represent various stages of user driven investigative processes. Dunne et al. proposed *GraphTrail* (Dunne et al., 2012), an analytics tool that supports visual history. In particular, this tool would attract analysts that deal with large network datasets. With GraphTrail, they performed a long-term field study with archeologists and determined that their visual history approach does benefit the users.

Our work is also motivated by the visual analytics systems that allow users to spatially organize hypotheses and evidence on large screen space. The screen space can be used as both a synthesis space, where users specially situate various artifacts that form a semantic layer that ultimately constitutes the sensemaking process, and also as a *space for process*, in which the screen space is used to accommodate history views that represent the analytical process itself (Singh et al., 2011). For instance, Sandbox is designed to support an open workspace where users can move and organize digital objects on the display space for external representations (Wright et al., 2006). Andrews et al. (2010; 2012) presented a visual analytics tool based on large displays called Analyst's Workspace. Several studies have provided evidence that larger displays can lead to user performance benefits and improvements in various cognitive and visualization tasks, e.g., (Andrews, Endert, & North, 2010; Bi & Balakrishnan, 2009; Shupp, Andrews, Dickey-Kurdziolek, Yost, & North, 2009). Their study results indicate that LHRD plays a crucial role as it offers space for sensemaking in its natural form. Similar to these tools, our approach emphasizes spatial organization of multiple documents, enabling the analyst to better leverage the larger screen space. As such, instead of creating references to prior states in the form of icons or shelved files, users can keep the past states of data views on the screen as fully visible active windows.

However, our study presented here is fundamentally different because of the analytical task and how users spawn process landmarks (or history items) that in turn build and support visual histories. The research of Andrews et al. (2010) studied the use of space for a textual analytics task that involved reading, searching, and making connections among entities in an intelligence analysis scenario involving a large number of documents. In contrast, the task in our study involves the manipulation of multiple spreadsheet windows originating from a small set of initial files. The goal of the analytical task is to identify a small number of specific events related to illegal actions, while the study itself focuses on understanding the use of branching functionality as a means of supporting the visual analysis history.

Lastly, Fink et al. (2009) recognized that screen space and high resolution are important for supporting the context of cyber analytic workspaces. Singh et al. (2011) presented a large-display tool for a visual history of the cyber analytic process expanding on Fink's design principles. The key design goal for the prototype tool was to support persistent data views via branches of new data views from the existing views. The persistence of views facilitates the process of reading and referencing history as the full analysis content remains visible on the display. Previous work has also shown the importance of persistent visibility for the effectiveness of memory support with large displays (Ragan et al., 2012). Our work extends prior research by (1) further examining whether effective use of visual history is influenced by increased display space, and (2) studying how screen space affects the analytical process through a user study.

3. BRANCHING HISTORY ITEMS

Fink et al. (2009) previously observed that two types of branching operations: **duplication** and **subsetting** are commonly used when conducting investigations with visual history. The subset operation creates a new history item containing a copy of the selected subset of a history item (Figure 1a,b,c), while the duplicating operation indicates the creation of the same copy of a history item (Figure 1a,d). It is crucial to keep different versions or parts of the same data file in order to make them readily available for different analytic tasks on large displays.

To facilitate such branching support, we created a custom Excel add-in that supports both window duplication and subsetting operations using a single mouse click (Figure 2), based on prior research (Fink et al., 2009; Singh et al., 2011). The subset operation creates a new Excel window containing a copy of the selected subset of data in the current window. To create a duplicate, users first select an empty cell and then press any of the arrow buttons (Figure 2b) to create a subset in a newly created window in the direction indicated by the arrow. The data in the new window is automatically saved with a name that includes the timestamp for the moment the file was created. Newly created windows are automatically resized to fit its data contents. We also provide a separate auto-resize feature to automatically resize a window to match the displayed data. To perform the auto-resize, the users simply select a window and press the AutoResize button (Figure 2c). The add-in also contains shortcuts for common analysis tasks of spreadsheet heatmapping (Figure 2d), sorting (Figure 2e), and highlighting (Figure 2f). Heatmap coloring can be applied to a range of selected cells of numbers by applying an intuitive color scheme where green stands for lower values, yellow for medium values, and red for high values (Figure 3). Sorting of selected columns can be done in ascending or descending orders. Lastly, highlighting selected cells can be done by pressing the red, yellow, or green button.

4. QUALITATIVE STUDY

To explore the impact of display size on the use of visual histories, we conducted a qualitative study using the context of an analysis task with a cybersecurity themed digital forensics dataset (Grinstein, Scholtz, Whiting, & Plaisant, 2009).

4.1 Research Questions and Hypothesis

The main goal of the study was to learn about how screen size affects the analysis process and strategies involving the creation and use of visual histories. The core research questions include:

Figure 1. Branching a source window in two ways: subsetting and duplication. The windows with blue boarders represent final branched windows from the IPLOG window.

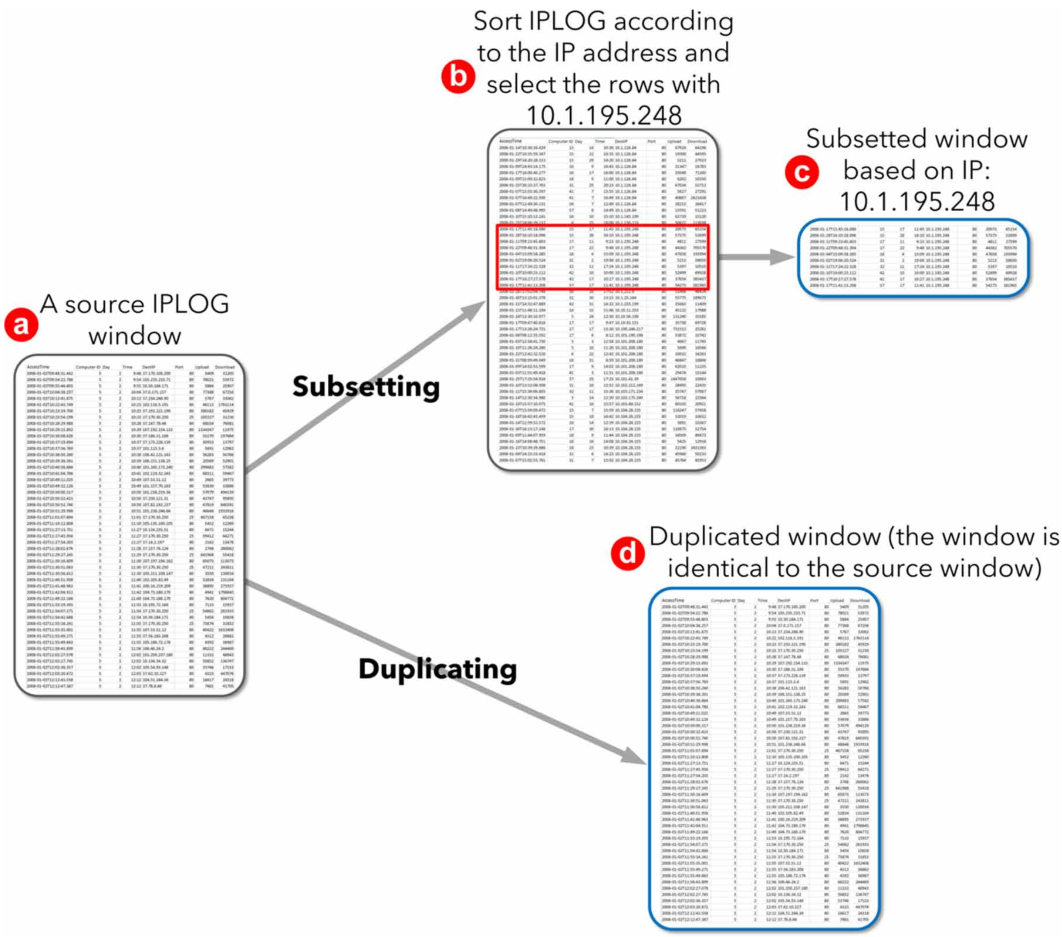
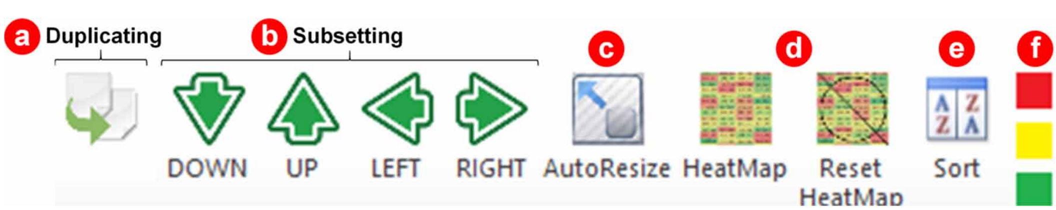
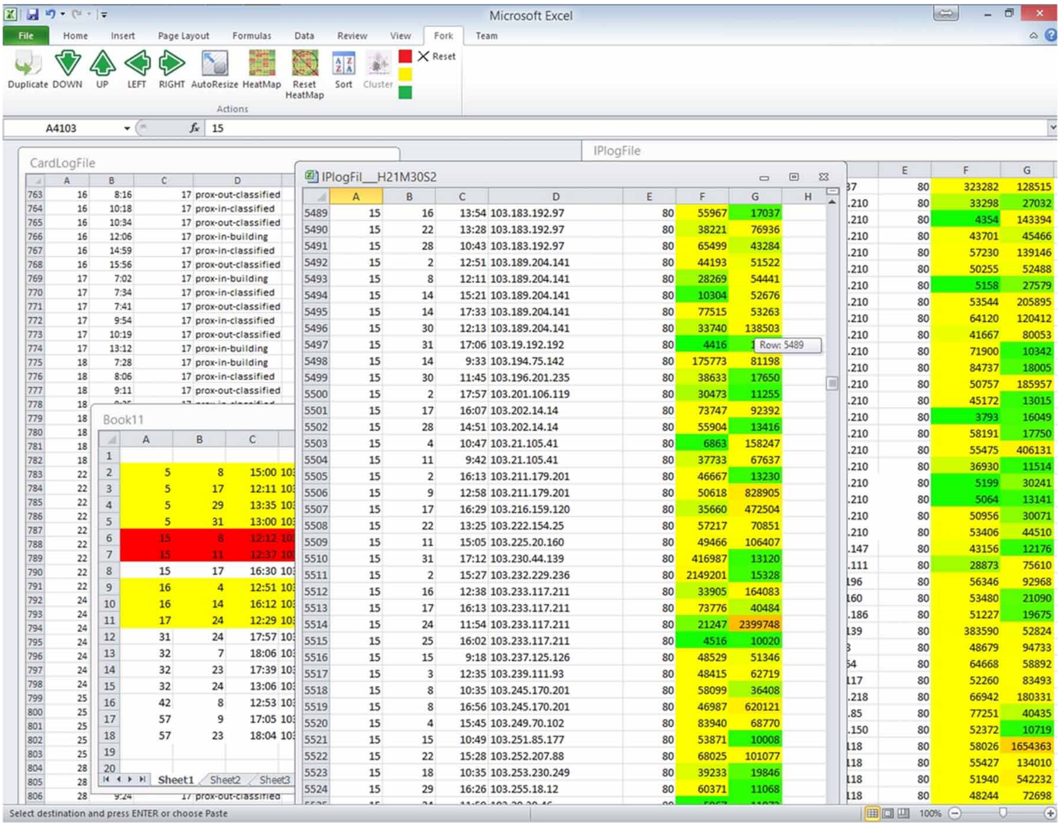


Figure 2. Buttons for the extra features of the Excel plug-in



- **RQ1.** How do participants create and employ visual histories for their analytical investigations?
- **RQ2.** How do participants create and organize process landmarks or history items utilizing the large screen space throughout their analysis?
- **RQ3.** How do analytic processes compare between the LHRD group and the small-screen group when using involving visual histories?

Figure 3. A sample view showing highlighted data items on a small display. This view also shows cluttered screen space with overlapping windows.



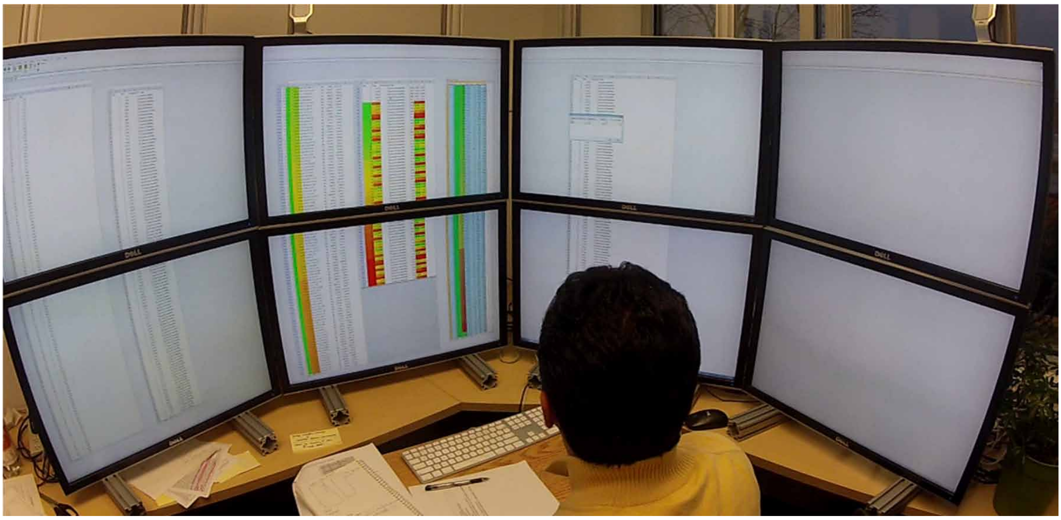
Prior to conducting the study, we hypothesized that the visual history with a larger display would help users better understand and more efficiently explore spreadsheet data.

4.2 Study Method

The study used two display configurations of different sizes (small and large). The small display setup used a 21-inch HD (1920x1024 pixels) resolution monitor. The large display setup used a personal workspace composed of a 4x2 grid of 30-inch monitors with total resolution of 10240x3200 pixels (see Figure 4). Both displays were connected to a Windows PC, and participants used Microsoft Excel with our Excel Add-in (described in Section 3). Each participant completed the task once using one of the two displays.

Dataset and Task: The analytic task and dataset for this study are based on a cybersecurity investigation from the VAST 2009 Challenge digital evidence dataset (Grinstein et al., 2009). The challenge is about detecting suspicious computer activity at a fictitious embassy. In this fictitious scenario, an embassy employee is suspected of transmitting sensitive data to an external criminal organization. The data consists of network traffic logs and proximity card logs that capture employee access to the certain areas in the embassy and associated computers. The challenge has several sub-challenges—we chose a challenge about identifying computers used to send information to an outside organization. This sub-challenge is representative of complex digital investigations and encourages both broad analyses of trends and deep analyses of specific entities (e.g., events related to specific employees and IP addresses).

Figure 4. The large display setup used in the study



The data consist of two CSV (comma-separated values) files: proximity card logs (PROXLOG) and network traffic logs (IPLOG).

- **The PROXLOG file** represents the embassy employee's proximity card usage logs. The proximity cards are used to access the embassy building and the restricted areas within. The dataset contains the following columns: *time of use*, *employee ID*, and *type* of card event (*prox-in-building, prox-in-classified, prox-out-classified*).
- **The IPLOG file** represents the computer activity and network traffic log file. Each employee of the embassy uses a different desktop computer with a static IP address. The IPLOG file contains data related to computer activity and network traffic associated with each computer's IP address. Specifically, the file includes *accesstime, source IP, destination IP, employee ID, network port information, upload rates, and download rates*.

By analyzing both files, the study participants were asked to identify which computer(s) was used to send information to certain external computers and describe how much information was sent. The task has a known ground-truth solution (Grinstein et al., 2009). Participants were given 90 minutes to investigate the data set and come up with a solution. At the beginning of the study, all participants were instructed that computer activity should be considered suspicious if an employee associated with a computer is not at the computer at the time of the activity. Additionally, all participants were given a document with additional information about the dataset 50 minutes into the study. The document contained information about the ports and IPs intended to help narrow down the digital evidence to a more manageable size. By providing this extra hint, we simulated a real-world situation where analysts receive new information in the midst of investigation. While the hint helped the participants narrow down the dataset, the analytical aspect of the task still remained challenging.

Participants: We recruited eight (three female, five male) graduate students with computer science backgrounds and basic knowledge of Microsoft Excel. All participants reported being confident and comfortable with the task of data analysis via Excel. Participants were compensated at the rate of $20 per hour. The study followed a between-subjects design; four participants were randomly assigned to the large LHRD setup, and the other four to the small screen.

Study Procedure: For each study session, the experimenter first explained and demonstrated the functionality of the Excel Add-in and answered any questions about the application. Upon explaining the task and the data set, participants were given 90 minutes to conduct the investigation. Upon completion of the task or reaching the time limit, the experimenter interviewed each participant. The entire study session took approximately two hours.

Data Collection: We recorded 1920x1080 resolution video of both display configurations and automatically captured screenshots every five seconds. We also recorded audio and video of the participants and the displays. Additionally, throughout each study session, we performed detailed observations and took notes of participant actions that seemed informative with regards to the analytical process. The post-study interviews were semi-structured and asked the participants to explain (a) their solutions to the analysis task, (b) what they did to arrive at their solutions, and (c) how they used to the application to conduct the investigation. We asked participants about their use of the provided add-in features. While all participants had the same set of common questions, the interview was flexible in that the interviewer could ask follow-up questions based on the initial responses. The interviews concluded with a series of questions based on the individual observations during the investigation process. The interviewer asked participants if they could explain specific actions or events performed during the investigation, thus clarifying the semantics associated with the actions. Explanations about specific actions and uses of the add-in were helpful in establishing a broader understanding of the usage of the branching histories.

Data Analysis: After the completion of all study sessions, we analyzed the video, screenshots and interview data. From our observations, we determined visible history items (spreadsheet windows) that could be interpreted as processes landmarks. We were particularly interested in how users created the visual landmarks during their analysis and what features they used for this purpose. We sought to analyze how process landmarks differed between users and how they recalled the investigation process upon completion of the task. We started by analyzing the video recordings and the screenshots. The video portion of the observational data was helpful for capturing temporal actions (e.g., interactions with the Excel add-in) and mapping them to physical participant actions. We performed selective coding on the video and screenshots to identify instances of duplication and subsetting from the existing data windows. For each window, we recorded when each window was created and closed as a way of objectively characterizing the user behavior. Additionally, we recorded time durations when windows were occluded by other windows (a window was considered to be occluded if part of it was hidden by another window). We also used the video and screenshots to record the instances of data highlighting, heatmapping, and sorting.

While it was not possible to infer rationale or motivation for every user action, we were able to map the interview notes and experimenter observations to many actions from the video. We sought to understand participant rationale and motivation for their actions (e.g., branching windows, duplicating windows, closing windows, or highlighting data).

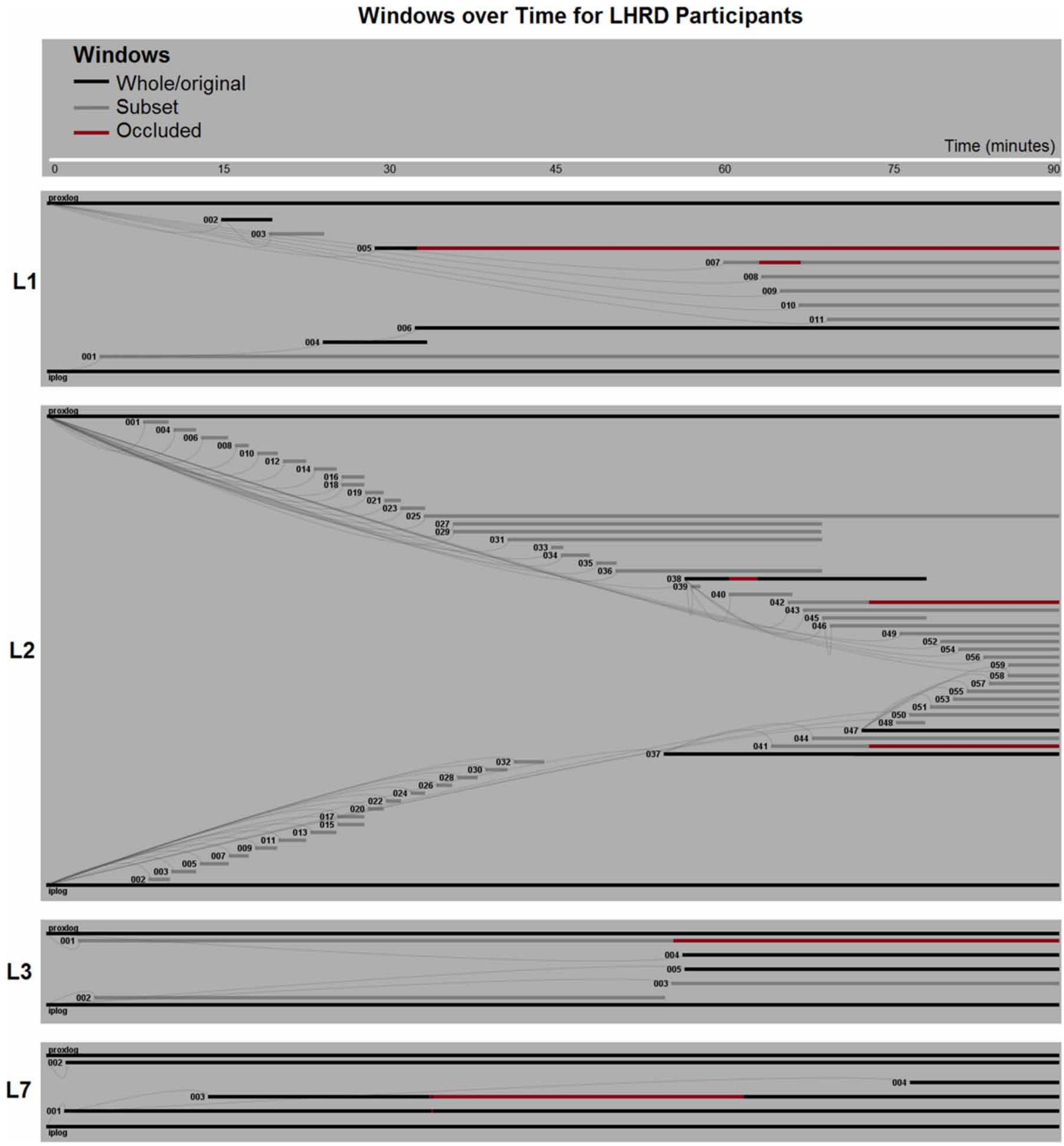
5. RESULTS AND ANALYSIS

In this section, we focus on examining the process landmarks (composed of highlighted or branched spreadsheet windows), use of visual histories, and spatial organization. We describe results for specific participants using numbers prefixed with “L” for LHRD users (L1, L2, L3, and L7) and “S” for small display users (S4, S5, S6, and S8). Results of our study show differences in analysis behaviors between the LHRD and the small screen groups.

5.1. Process Map

To better understand overall analysis processes, we examined instances of (a) window duplication, (b) subsetting, and (c) occlusion to create detailed process maps for each of the eight participants (see Figure 5 and Figure 6) to visualize the investigation process based on the collected study data.

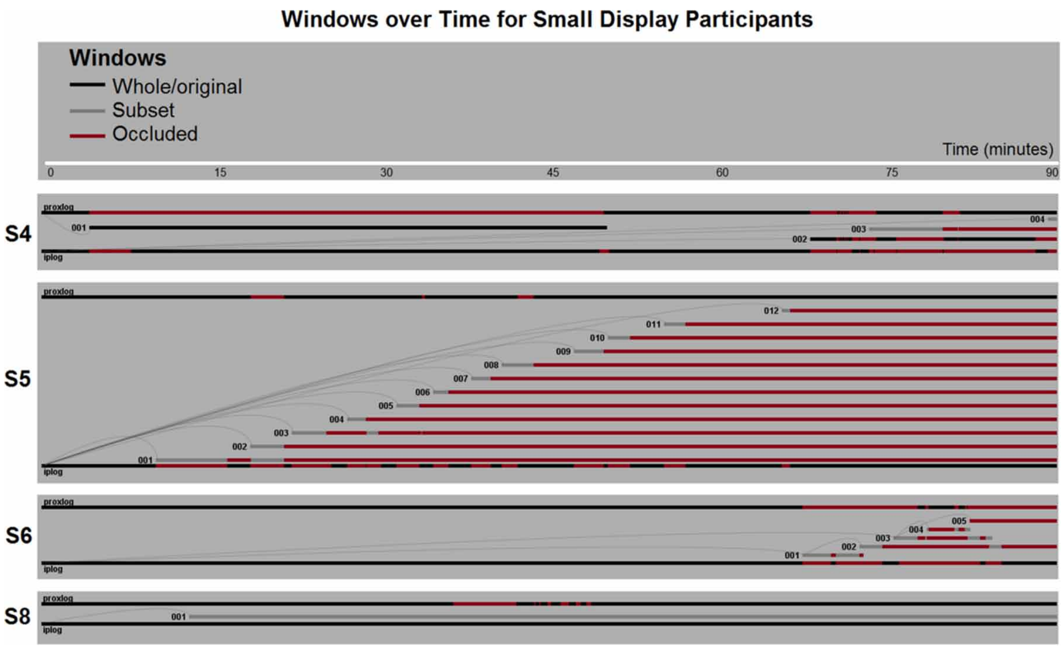
Figure 5. Window management map for LHRD participants. Each horizontal bar represents an open window, with black bars representing whole data view, gray bars representing subsetted views, and red segments indicating window occlusion. Thin gray lines indicate branching behaviors (subsetting and duplication) from source windows.



The figures capture the state of the windows as well as the basic user-initiated actions—namely duplication and subsetting. In terms of the state, the figures illustrate whether the windows branched from the original data or occluded.

Figures 5 and 6 show the results of all participants' analytics processes with visual histories during their study sessions in terms of window duplication, subsetting, and occlusion to create detailed process maps. These figures reveal the process taken by each user based on window management, with Figure 5 showing LHRD participants and Figure 6 showing the small-display participants. In both figures, the horizontal axis represents the passage of time. The colored black, gray, and red bars are window identifiers that show the status of all windows open at any given times. The black bars

Figure 6. Window management map for small display participants. Each horizontal bar represents an open window, with black bars representing whole data view, gray bars representing subsetted views, and red segments indicating window occlusion. Thin gray lines indicate subsetting and duplication from source windows.



represent a complete data window (either the original view or a duplicate), and gray bars represent subsetted windows. The red segments on the bar indicate window occlusion. The bars are connected with thin curved lines that indicate the origins of where the windows were branched. That is, following one of the curved gray hairlines from a bar on the left to a different bar on the right means that the window identifier on the right was created from the window identifier on the left. The small numbers above next to each bar are window identifiers from the coding process for each user's process map.

As shown in the process maps of both groups, we can observe that the LHRD participants left a large number of windows open, and they did a significant amount of branching. The branched windows were rarely occluded by other ones on the large displays. On the other hand, the small-display participants did not create many branches (the behavior is represented thin gray lines in Figure 6) except S5. However, S5's branched windows were occluded (the red segments indicate window occlusion.)

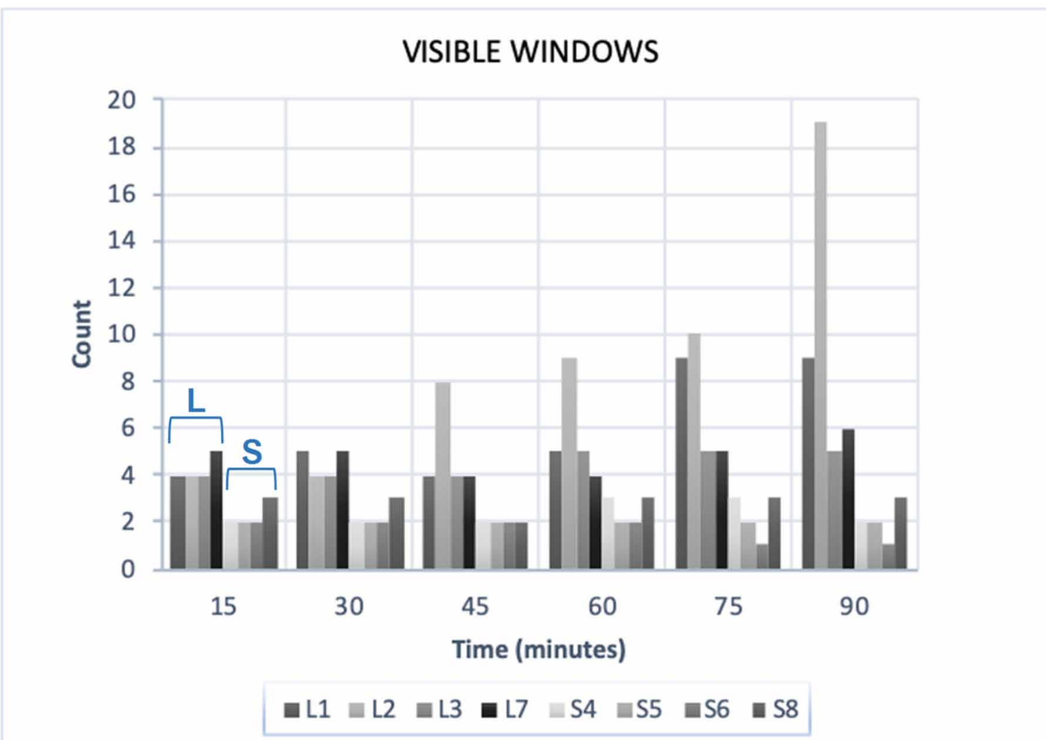
5.2. Process Landmarks

We were interested in identifying process landmarks that serve as an explicit record of the origin of an analytic subtask or divergence in the broader analytic process. Such instances of process landmarks act as the building blocks for the formation of the visual history. Our observations and video analysis demonstrated that all eight users created process landmarks in the following two ways: (1) **highlighting** data entries within windows (Figure 3) and (2) **branching** (subsetting and duplicating) data entries into new windows (Figure 1). All participants used highlighting as a way to reduce the information by making the distinction between the processed and unprocessed data more explicit while maintaining the context of the original data file. Subsetting and duplication, on the other hand, served as a way to focus on a subset of entries while simultaneously separating the set from the source window. By subsetting and duplicating entire sets of entries, users effectively created "sandboxes" where they could interact without interfering with the state of the source windows. This behavior is true for both groups, yet we observed several notable differences in how participants from the two groups created process landmarks.

Keeping separate windows for new and different information. One difference between the two display groups was in situations where users had to change the status of a landmark. For instance, S4 highlighted certain blocks of data as red early in the investigation. Then, after receiving the hint at the 50-minute mark, participant user manually removed certain red highlights. When asked why she did so, the participant explained that the meaning associated with the highlights had changed upon receiving the additional hint. L2, L3, and L7 from the LHRD group also performed highlighting of data files during the first 50 minutes, however, after receiving the hints, they created duplicates of the initial data files and continued working with the duplicates. When asked to explain their actions, participants unanimously claimed that they performed the duplications of the data windows in order to pursue new ideas sparked by the hints. By performing the duplication, the three LHRD participants preserved the history of their prior investigation threads by keeping the highlighted entries unchanged.

Exploiting persistent visibility of history items. Another difference comes from the number of branches and duplicates that participants created. Our video coding revealed that all four participants from the LHRD group had more visible data windows than the participants from the small-screen group over time (see Figure 7). On the small screen, the maximum number of simultaneously visible windows was three while on the large display it was 20. In order to fit additional windows on the small display, the participants periodically adjusted zoom levels, resized window borders, and adjusted the gaps between window boundaries. The participants from the LHRD group, on the other hand, never adjusted the zoom levels or adjusted window boundaries.

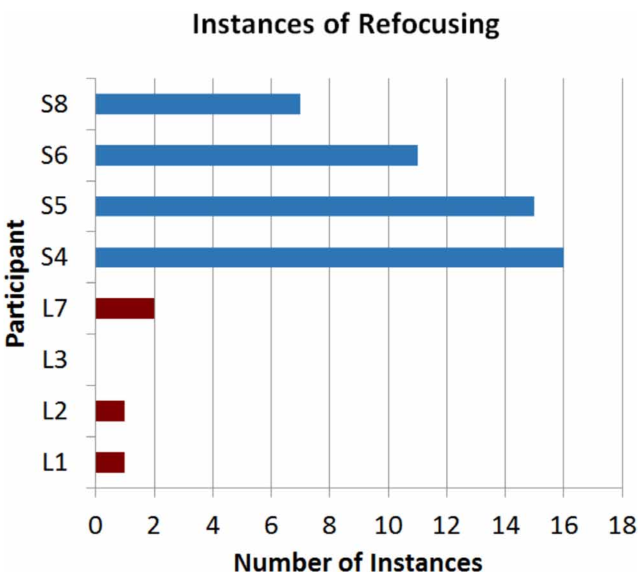
Figure 7. Visible windows for participants over time. LHRD users (L) had more windows visible than small-screen users (S). Over time, the number of open and visible windows increased for the large display participants.



We also observed a difference in how participants from the LHRD and small-display groups created process landmarks. Three of the four large-screen participants (L1, L2, and L7) used highlighting and branching features (i.e., duplicating and subsetting) throughout the investigation process. Particularly, as shown in Figure 5, the significant amount of branching was done by participants L1 and L2. In addition to branching and highlighting, L7 also explicitly copied certain findings into an empty column of the original window and highlighted the suspicious records with red. The participants from the small display group also used branching and highlighting throughout the investigation, albeit within the limited space imposed by the display.

From the post-study interviews, we discovered that the process of branching windows was perceived to be less helpful for the small display users because it was more difficult to keep track of the branched windows and it required them to manually make the occluded windows visible. Figure 8 shows how frequently the small display participants had to switch their active window focus (by

Figure 8. Total instances of refocusing for all participants. Small-screen users switched the focus of the active window more frequently than the LHRD users. The blue bars represent the small display group, and the brown bars illustrate the large display group. To refocus windows, the users actively clicked on the visible parts of the windows thus making the content fully visible.



clicking on the occluded windows with their mouse) in order to show specific occluded windows on top of other windows. Our observations and the post-study interview confirm that the small display space constraints and windows occlusions were the core factors explaining the preference to use highlighting as the primary tool for capturing the investigation process.

The post-investigation interviews revealed a difference in how the users described their analysis processes between the two groups. When asked to describe the process, the small-screen group participants relied on virtual navigation approaches (scrolling and searching widgets) and tended to get lost in searching for the history items. They often failed when asked to identify the windows while relevant to their investigation process. It was not uncommon to see the small-screen participants scroll through open spreadsheets in search of the highlighted entries. This highlights the importance of persistent visibility of items, as also indicated by (Andrews et al., 2010; Chung et al., 2014; Ragan et al. 2012). For example, while S5 used branching extensively (see Figure 6), window occlusions caused by the limited space made it difficult for the participant to explain the purpose of all the branched windows. Participants who used branching in the LHRD group, on the other hand, were able to point to the important data views and highlights without noticeable effort by simply pointing to relevant content on the large display space.

5.3. The Rationale of Branching and Maintaining History Items

From the video review, we observed that all eight participants used branching of windows in two ways: subsetting and duplication (Figure 1). The post-study interviews revealed three types of reasons for branching and subsetting.

Duplicate windows to keep the original one intact. At the end of each study session, we asked the participants to explain why they duplicated or "subsetted" the data. L7 had two uses for duplication. First, he duplicated windows in order to keep the original dataset intact. Then, upon receiving the hints, he duplicated once more to create more windows for the same data because the hints allowed him to pursue a new angle of investigation. L1, L2, L3 and S4 used duplication as a way to restart their investigation from the unaltered source files while keeping the manipulated files as backup. Thus, duplication was used to save and preserve the original source, for "fresh starts," and for creating multiple views of the same data. As for the subsetting functionality, all eight participants noted they used it to focus on particular subsets of data in safe isolation from the source files.

Subset windows to break the analysis into sub-problems. While subsetting behavior was seen in both groups, each group had one "highly active subsetter" (L2 and S5), who branched out more than nine times within 30 minutes. When asked about their strategies and the overall use of branching in the interview, L2 explained that he sought to exhaustively eliminate possible records by simply subsetting out certain employee's IPLOG and PROXLOG daily records and explicitly checking the daily data for any evidence of suspicious activity. It should be noted that L2 closed all those branches that had no signs of suspicious activity. L2 used a divide-and-conquer style approach by first dividing the dataset into smaller and more manageable subsets and then eliminating irrelevant subsets.

Similar to L2, the small-display participant S5 made frequent successive subsets (see S5 in Figure 6) and explicitly checked for signs of suspicious activity. Unlike L2, S5 simply made new branches right after finishing the branch verification tasks, leaving the branches open. However, despite leaving these branches open, S5 never revisited them. In the interview, S5 explained that leaving the branches open had no meaning to her because she intentionally treated those branches as processed and thus did not care about other windows completely occluding those branches. Similar feedback was received from three out of the four participants from the small display group, with S8 being the exception. Interestingly, S8 had no occluded branches until the end of the study. The participant explained that he only used the two original data windows and then tried to solve the problem computationally (i.e., by creating formulas in Excel) using a separate window. After failing to solve computationally, S8 decided to keep the single branch on the screen and adjust the sizes of the source windows in a way that would make all three windows visible.

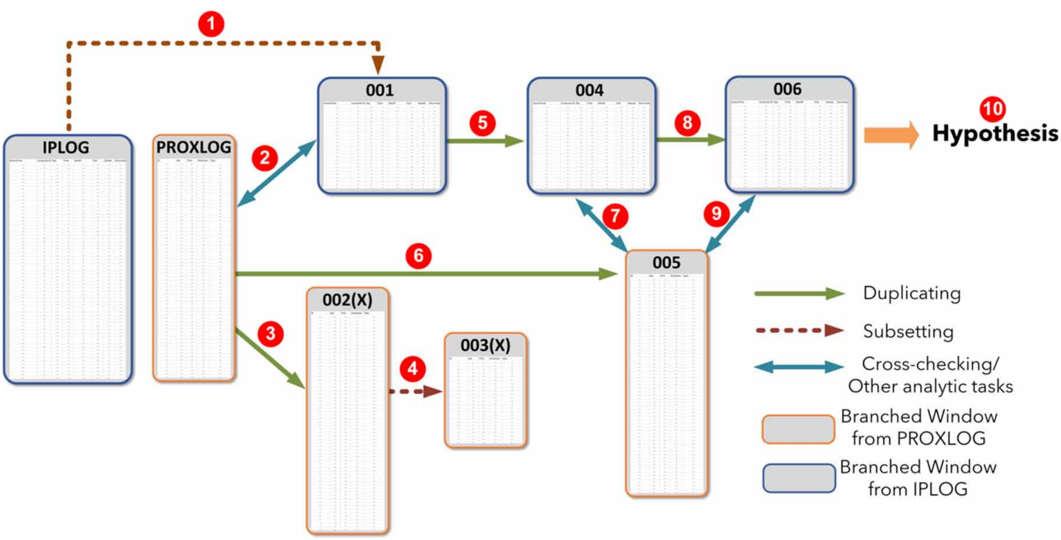
Assign roles to different windows in analysis. While there were some cases of users arbitrarily keeping windows visible, there was a noticeable pattern behind the rationale for keeping or closing windows. While the reasons for closing windows varied, each participant followed specific criteria. Participants only closed windows that had been processed or those that were byproducts of excessive/unnecessary duplication. L1 and L2, for instance, closed some of the branch duplicates due to consumed space. The participants kept the windows open when they perceived the windows to serve as either (a) a source of new subsets, (b) a container for findings, (c) a subset to be analyzed, or (d) a duplicate that created a multi-perspective view.

5.4. Use of Visual Histories

Throughout the study, we examined the use of visual history with and its influence on the participant analytical processes. During each study session, we observed how participants accessed windows on the displays. All four LHRD group participants accessed the windows on the display by physically turning towards the windows, thus employing physical navigation (Ball et al., 2007) and leveraging the immediate accessibility and visibility of the windows.

It is notable that L1 and L2 demonstrated prominent usage of visual history. We attribute this to the fact that L1 and L2 produced a number of branched windows, and as a result, their investigative findings were scattered across several windows, unlike the other two large-screen participants. For instance, L1 started out by sorting a particular column in one spreadsheet window (IPLOG) and then branched out subset "001" to investigate a particular hypothesis about a network port (Figure 9-1). Next, he applied multiple sort operations inside "001" followed by heatmapping. L1 started a verification process for selected entries in "001" by crosschecking selected IP activity records against the proximity card activity records (Figure 9-2). To facilitate this process, he first duplicated PROXLOG into "002" (Figure 9-3) and then created subset "003" (Figure 9-4). However, several minutes later, he closed both "002" and "003" and then duplicated "001" into "004" (Figure 9-5). After several minutes of scrolling, highlighting, and sorting inside "004", he went back to the PROXLOG data and created another duplicate, "005" (Figure 9-6), which had all the original columns intact. After several minutes of crosschecking IP records from "004" against card event records from "005" (Figure 9-7), L1 once again refreshed his sources of IP and event information by duplicating "004" into "006" (Figure 9-8). From this point onwards, he kept sorting, scrolling and highlighting inside "006" and crosschecking select values from "006" against "005" (Figure 9-9). Interestingly, L1 highlighted different cells of the same data files ("004" and "006") and performed cross-referencing between each file and "005" separately. After the 50-minute mark, he found instances of suspicious activities inside "006" (Figure 9-10). He branched out multiple subsets from PROXLOG as evidence

Figure 9. An example of a visual history created by L1. The numbered labels (from 1 to 10) indicate a sequence of the analytic tasks. Branched windows with blue boarders were originated from IPLOG, and ones with orange boarders were branched from PROXLOG. For window 002 and 003, 'X' indicates windows that were closed.



of the suspicious nature of highlighted IP records in "006". L1 demonstrated use of a visual history by going back to a previous window (from "005" to PROXLOG) and using it for the current thread of investigation. The participant explained that he decided to go back to PROXLOG and branch out suspicious records from it because he thought that "005" was modified and could compromise the new thread of investigation he had started after receiving the hints. These observations show that L1

Figure 10. Large display users encoded different semantic meanings or work areas according to spatial positions of windows. L2 maintained the original data files in the middle of the screen as the primary work area (the red box), both of which were regarded as the most important. The red arrows indicate the direction in which L2 moved windows of secondary importance, branching out from the primary work area.

benefited from having access to the original version of PROXLOG because he was able to revisit the file and use it once again.

Small-screen participants tried to revisit windows, including those that had not been displayed for prolonged periods. However, the instances of revisiting of older windows came at the price of dealing with occlusions. The occlusions were the direct result of the lack of space on the small display since the screen space on the small display generally could not accommodate more than three windows in a usable way. Thus, the lack of space acted as a constraint preventing visual histories from being established and leveraged by the small-screen participants.

5.5. Spatial Organization

Spatially, all LHRD participants derived branches and duplicates in the same direction—from left to right. This comes as no surprise since the LHRD participants started with original source files in the leftmost (the first) column of the screen. Thus, they built their visual histories as extensions of the original files as they remained in their original positions. All four LHRD participants used the available space near the center of the display as a synthesis space for data exploration and manipulation. The participants kept the windows open if they contained data that had not been decisively ruled out as irrelevant. During the post-study interview, participants explained that the area around the center of the screen was used as their preferred synthesis space because of the sitting position and the distance to the screen. This space represents the “sweet spot” of the large display as the sitting position allows users to achieve the shortest distance from their eyes. We also observed that participants would often use this area as the place for branched windows from these two original data files (Figure 10). Interestingly, all LHRD group participants performed similar actions when it came to managing windows that did not need additional processing. Participants simply moved such windows to either left or right from the center of the screen. For instance, L2 moved all the small branches with findings to the right and a duplicate view to the left, closer to the original files.

6. DISCUSSION

In this section, we discuss the results of the qualitative study, the use of large screen space for the analytic tasks with visual histories, limitations in our study, and design implications for future visual analytics tools that supports the visual history on LHRD.

6.1. Main Differences in Analysis Process Between the LHRD and Small Screen Group

The results of the study show participants from the large screen group used the branching functionality to form visual histories, thus leveraging the quick access to the branched history items (i.e., the content inside the branched windows) to rapidly switch their threads of investigation. We observed the following differences in the process of visual history formation between the LHRD and small-display groups:

- **Exploiting Physical Navigation:** Persistent visibility of multiple windows significantly affected the use of virtual and physical navigation to create and interact with visual histories between the two groups. The LHRD enables users to take advantage of large physical space (Andrews et al., 2010; Chung et al., 2014; Ragan et al., 2012) by continued visual availability of windows. Specifically, the LHRD users had more windows visible to participants as compared to the small-screen users who employed virtual navigation such as adjusting zoom levels, resizing window borders, and minimizing gaps between window boundaries to fit more data into view. Conversely, we observed no instances of zooming out or resizing windows among large-display participants. For the refocusing issues, the small-screen users had to switch the active window more frequently than the LHRD users, because of the added burden of dealing with window occlusions. In contrast, the LHRD users did not need to perform such actions, because the larger space was able to freely accommodate multiple windows.
- **Branching windows as process landmarks:** Although both large- and small-screen participants used branching and highlighting throughout the investigation process, there was a difference between the two groups in the types of process landmarks they utilized. The LHRD participants often created separate duplications and subsets of windows and used them to solve various smaller sub-tasks. By performing duplications, the LHRD participants actively preserved the process history by maintaining the windows with different process landmarks. On the other hand, the small-screen participants only employed highlighting as a way to track the investigation process. Based on the post-study interview findings, the space constraints and the occlusions of windows on the small display were the major reasons for using highlighting instead of branching.
- **Multiple roles of branched windows:** While all participants in both groups employed branching, the main goals for duplication or subsetting differed. The large-display group participants kept windows open when they thought the contents of the window could be useful (see Section 4.2.). Since large displays support keeping multiple windows open and visible, the LHRD participants divided the IPLOG and PROXLOG windows into smaller successive branches (e.g., daily logs for a specific IP address), quickly analyzed each of branched windows, and kept the branched windows open. The small-display participants also created multiple branches and investigated the data in such windows for signs of suspicious activity. However, despite leaving the branched windows open, they rarely revisited them and mostly ignored the inevitable window occlusion caused by the limited screen space.

Our findings confirmed that the LHRD participants created and employed visual histories with process landmarks generated throughout the investigation process. The LHRD participants' investigative sub-tasks and findings were distributed among a higher number of smaller windows, and the use of visual histories was more convenience and accessible due to the windows on the screen. Indication of this finding are the number of visible windows during the investigation and the increase in number of windows throughout the investigation (Figure 7). Furthermore, the amount of space and the scale of the visual history on the LHRD formed an environment that allowed the participants to leverage physical navigation ultimately contributing to a better awareness of the investigation artifacts (e.g., the windows and the content inside the windows) when compared to the small screen participants.

6.2. Additional User Studies

Our qualitative study replicated the existing visual analytics studies that also factored in several essential variables, including the number of participants, analysis tasks, data analysis/collection methods, etc. Specifically, our efforts extend prior studies that have emphasized the value of space for visual analytics processes, featuring large high-resolution displays (Endert, Fiaux, & North, 2012; Andrews & North, 2012; Ragan, Goodall, & Tung, 2015), notecards on a tabletop (Robinson, 2008), and multiple displays (Chung et al., 2014). As Kang et al. (2009) emphasized, it is important to assess the qualitative value of a visual analytics system and related process instead of relying solely on statistical results.

This exploratory study revealed non-trivial differences in analytic processes and usage of visual histories between users of large- and small-screen displays. For example, the results suggest that greater availability of display space encouraged user to (1) create visual histories and (2) use visual histories to improve their task performance and overall investigation awareness. It must be noted, however, that due to the small subject population, we could not conclusively pinpoint how the use of LHRDs helped improve the analysis performance or demonstrate statistical significance.

Hence, follow-up studies are needed to evaluate the effects of display sizes and types on users' analysis performance involving the use of visual histories and the ability to recall the steps of the analytic process in a quantitative manner. For example, future studies can adopt quantitative methods for evaluating recall of analysis processes outlined by Ragan and Goodall (2014) and demonstrated in (Ragan, Goodall, & Tung, 2015). Future research can also consider quantitative task performance based on existing ground-truth solutions to the cybersecurity analysis challenges (Grinstein et al., 2009). Subsequent studies can evaluate participant scores for answers correctly identified to assess quantitative difference between different visual analytics environments.

We also note that our results are constrained due to the experimental setting spanning only a single session per participant, and we hypothesize the implications of display space would be even more pronounced over longer durations. A longitudinal study with more experienced investigators could provide additional valuable insights—as would the investigation of the effectiveness of a longitudinal analysis with visual history over longer periods of time (i.e., days, weeks, or a month).

6.3. Design Implication and Future Work

Our study identified several challenges related to visual analytics with visual histories on LHRD. These challenges suggest important avenues for advancing visual analytics tools for both visual history and LHRD. In this section, we discuss a new visual representation and user interface techniques for addressing the challenges we identified in the study.

Visualizing branch relationships among history items. As we observed, participants needed to create a number of the duplicated and subsetted windows, even though they did not use them immediately. Participants kept these windows open, forming a visual history for their analyses. However, since the information and the tasks ended up scattered and disconnected across separate windows, one challenge associated with the visual history on the large display is to maintain awareness of relevant windows. For instance, after branching out or duplicating several windows, a participant might lose track of where a specific window had originated. Thus, to address this problem, we can use visual links to visually connect branching information among windows (Chung & North, 2018; Waldner, Puff, Lex, Streit, & Schmalstieg, 2010) with the goal of focusing/guiding the analyst's attention to the relationships between the original file and the branched windows originating from the original file.

Minimizing steps for creating new branched history items. The study results demonstrate that the Excel plug-in supported a straight-forward process for creating a branching history directly through the workspace, though ability to reference the history depended on display size. Ideally, the process of branching should be natural, and the steps required for creating new process artifacts should also be minimized. For instance, our tool facilitates the creation of a new subset window by simply selecting a range of a specific data item and clicking a button (Figure 2b). From our observations and participant responses, this functionality was simple to use. However, selecting spreadsheet cells in a large range may not be easy for a large-sized branch since it requires users to click and drag the mouse over a large group of cells (or rows of the spreadsheet). Thus, we will investigate new interaction techniques that can help users create branches of the investigation with the least possible effort.

Promoting Physical Navigation. With LHRDs, users are granted the freedom of creating information space of a desired size to explicitly act as visual history. However, we observed that use of large displays with traditional input devices (mouse and keyboard) create certain challenges in facilitating physical navigation. For instance, in our study, users complained of difficulty navigating the mouse cursor large distances due. These interface issues for better supporting physical navigation are essential for large-display research (Esakia et al., 2014), and they need be addressed in the context of specific applications and user interfaces, e.g., (Endert et al., 2011).

Supporting Automatic Branching. The presented study focused on manual branching by users, but another possibility is to consider automation in branching. Research with experts in professional analysis settings suggest the clear need for such automation to support provenance management (Madanagopal, Ragan, and Benjamin, 2019). Our findings of typical branching behaviors could serve as a basis for developing such tools. While this study focused exclusively on a spreadsheet dataset, in the future we will consider creating a visual history environment that is generalizable to other tools used in various digital investigations, e.g., WireShark (Chappell & Combs, 2010), X-Ways (Shavers & Zimmerman, 2013), etc. With a software agnostic visual history, the benefits of enhancing investigative process would not be limited to a particular type of task, but instead could be applicable to a wide variety of tasks involving large and complex datasets.

7. CONCLUSION

Many digital forensics analysts cope with large amounts of data and perform exploratory investigations that involve pursuing multiple hypotheses. It is important for data exploration and analysis tools to support the flow and history of an investigation for complex analysis tasks. When data manipulation is involved, it is crucial to maintain different versions of the data and to make them readily available. Visual history of the investigative process can maintain the elements associated with such processes in their active forms as part of the natural analysis workspace. Our study shows that, with large screen space, users can naturally generate, access, and leverage multiple history items with the help of persistently visible and easily accessible artifacts directly available through the analysis workspace.

By comparing the use of visual history between users of large and small screen displays, our study shows that the amount of screen space can influence the analytic process involving visual histories shown through a spatial representation. LHRD participants consistently created multiple versions of data windows and used these views to track progress and insight, to gain multiple perspectives on the data, and to preserve and reuse previous steps of their process by directly leveraging their process history artifacts. Overall, the LHRD participants showed less difficulty with recalling and presenting process landmarks. The small-screen participants, in contrast, did not show consistency with the creation of multiple branches of data windows. The small-screen size had a distracting effect on the analysis process because users had to manage multiple overlapping windows and excessive vertical scrolling within the data files.

While large displays have been previously shown to demonstrate advantages for visual pattern finding tasks in large-scale visualizations (Fink et al., 2009; Yost et al., 2007), the results of our study suggest that large displays may also provide significant benefits for the task of managing the analytical process and associated analytics artifacts. Our study shows that visual histories can be used to bolster visual analytics processes by supporting better awareness of different threads of investigation, mainly via interactions in the form of memory externalization (based on spatial organizations of windows) and physical navigation. With such interactions and improved awareness, users can experience reduced demand on the mental workload thus leaving more cognitive resources for the analytical task at hand.

REFERENCES

Andrews, C., Endert, A., & North, C. (2010). Space to think: large high-resolution displays for sensemaking. *Proceedings of ACM CHI'10*, 55-64. doi:10.1145/1753326.1753336

Andrews, C., & North, C. (2012). Analyst's Workspace: An embodied sensemaking environment for large, high-resolution displays. *Proceedings of IEEE VAST'12*, 123-131. doi:10.1109/VAST.2012.6400559

Ball, R., North, C., & Bowman, D. A. (2007). Move to improve: promoting physical navigation to increase user performance with large displays. *Proceedings of ACM CHI'07*, 191-200. doi:10.1145/1240624.1240656

Bavoil, L., Callahan, S. P., Crossno, P. J., Freire, J., Scheidegger, C. E., Silva, C. T., & Vo, H. T. (2005). Vistrails: Enabling interactive multiple-view visualizations. *Proceedings of IEEE VIS 05*, 135-142.

Bi, X., & Balakrishnan, R. (2009). Comparing usage of a large high-resolution display to single or dual desktop displays for daily work. *Proceedings of ACM CHI'09*, 1005-1014. doi:10.1145/1518701.1518855

Chappell, L., & Combs, G. (2010). *Wireshark network analysis: the official Wireshark certified network analyst study guide*. Protocol Analysis Institute, Chappell University.

Chung, H., & North, C. (2018). SAViL: Cross-display visual links for sensemaking in display ecologies. *Personal and Ubiquitous Computing*, *22*(2), 409–431. doi:10.1007/s00779-017-1091-4

Chung, H., North, C., Self, J. Z., Chu, S., & Quek, F. (2014). VisPorter: Facilitating information sharing for collaborative sensemaking on multiple displays. *Personal and Ubiquitous Computing*, *18*(5), 1169–1186. doi:10.1007/s00779-013-0727-2

Dunne, C., Henry Riche, N., Lee, B., Metoyer, R., & Robertson, G. (2012). GraphTrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. *Proceedings of ACM CHI'12*, 1663-1672. doi:10.1145/2207676.2208293

Dunsmuir, D., Baraghoush, M. Z., Chen, V., Joorabchi, M. E., Alimadadi, S., Lee, E., . . . Woodbury, R. (2010). *CZSaw, IMAS & Tableau: Collaboration among teams: VAST 2010 Grand Challenge award: Excellent student team analysis.* Paper presented at IEEE VAST'10. doi:10.1109/VAST.2010.5653035

Endert, A., Andrews, C., Fink, G. A., & North, C. (2009). Professional analysts using a large, high-resolution display. *Proceedings of IEEE VAST'09*, 273-274. doi:10.1109/VAST.2009.5332485

Endert, A., Fiaux, P., Chung, H., Stewart, M., Andrews, C., & North, C. (2011). ChairMouse: leveraging natural chair rotation for cursor navigation on large, high-resolution displays. *Proceedings of ACM CHI'11*, 571-580. doi:10.1145/1979742.1979628

Endert, A., Fiaux, P., & North, C. (2012). Semantic interaction for visual text analytics. *Proceedings of ACM CHI'12*, 473-482.

Esakia, A., Endert, A., & North, C. (2014). Large display interaction via multiple acceleration curves and multifinger pointer control. *Advances in Human-Computer Interaction*, *2014*, 12. doi:10.1155/2014/691507

Fink, G. A., North, C. L., Endert, A., & Rose, S. (2009). Visualizing cyber security: Usable workspaces. *Proceedings of IEEE VizSec 2009*, 45-56.

Goodall, J. R., & Tesone, D. R. (2009). Visual analytics for network flow analysis. *Proceedings of IEEE Cybersecurity Applications & Technology*, 199-204. doi:10.1109/CATCH.2009.47

Gotz, D., & Zhou, M. X. (2009). Characterizing users' visual analytic activity for insight provenance. *Information Visualization*, *8*(1), 42–55. doi:10.1057/ivs.2008.31

Grinstein, G., Scholtz, J., Whiting, M., & Plaisant, C. (2009). VAST 2009 challenge: an insider threat. *Proceedings of IEEE VAST 2009*, 243-244.

Heer, J., Mackinlay, J., Stolte, C., & Agrawala, M. (2008). Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, *14*(6), 1189–1196. doi:10.1109/TVCG.2008.137 PMID:18988963

Heer, J., & Shneiderman, B. (2012). Interactive dynamics for visual analysis. *Queue*, *10*(2), 30. doi:10.1145/2133416.2146416

Jankun-Kelly, T. J., Kwan-Liu, M., & Gertz, M. (2007). A Model and Framework for Visualization Exploration. *IEEE Transactions on Visualization and Computer Graphics*, *13*(2), 357–369. doi:10.1109/TVCG.2007.28 PMID:17218751

Kadivar, N., Chen, V., Dunsmuir, D., Lee, E., Qian, C., Dill, J., & Woodbury, R. (2009). Capturing and supporting the analysis process. *Proceedings of IEEE VAST'09*, 131-138.

Kang, Y.-a., Gorg, C., & Stasko, J. (2009). Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study. *Proceedings of IEEE VAST'09*, 139-146. doi:10.1109/VAST.2009.5333878

Kreuseler, M., Nocke, T., & Schumann, H. (2004). A History Mechanism for Visual Data Mining. *Proceedings of IEEE INFOVIS'*04, pp. 49-56. doi:10.1109/INFVIS.2004.2

Kwan-Liu, M. (1999). Image graphs-a novel approach to visual data exploration. *Proceedings of IEEE Visualization'*99, 81-88.

Lee, J. P., & Grinstein, G. (1995). An Architecture for Retaining and Analyzing Visual Explorations of Databases. *Proceedings of IEEE Visualization'*95, 101. doi:10.1109/VISUAL.1995.480801

Li, G., Bragdon, A. C., Pan, Z., Zhang, M., Swartz, S. M., Laidlaw, D. H., . . . Chen, J. (2011). *VisBubbles: a workflow-driven framework for scientific data analysis of time-varying biological datasets.* Paper presented at the SIGGRAPH Asia 2011 Posters, Hong Kong, China. doi:10.1145/2073304.2073333

Madanagopal, K., Ragan, E. D., & Benjamin, P. (2019). Analytic Provenance in Practice: The Role of Provenance in Real-World Visualization and Data Analysis Environments. *IEEE Computer Graphics and Applications*, *39*(6), 30–45. doi:10.1109/MCG.2019.2933419 PMID:31395538

North, C., Chang, R., Endert, A., Dou, W., May, R., Pike, B., & Fink, G. (2011). Analytic provenance: process+ interaction+ insight. *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, 33-36.

Pirolli, P., & Card, S. (2005). Sensemaking Processes of Intelligence Analysts and Possible Leverage Points as Identiðed Through Cognitive Task Analysis. *Proceedings of International Conference on Intelligence Analysis*, 5, 6.

Ragan, E. D., Endert, A., Bowman, D. A., & Quek, F. (2012). How spatial layout, interactivity, and persistent visibility affect learning with large displays. *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 91-98. doi:10.1145/2254556.2254576

Ragan, E. D., Endert, A., Sanyal, J., & Chen, J. (2015). Characterizing provenance in visualization and data analysis: An organizational framework of provenance types and purposes. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 31–40. doi:10.1109/TVCG.2015.2467551 PMID:26340779

Ragan, E. D., & Goodall, J. R. (2014). Evaluation methodology for comparing memory and communication of analytic processes in visual analytics. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization* (pp. 27-34). ACM. doi:10.1145/2669557.2669563

Ragan, E. D., Goodall, J. R., & Tung, A. (2015). Evaluating how level of detail of visual history affects process memory. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2711-2720). ACM. doi:10.1145/2702123.2702376

Robinson, A. C. (2008). Collaborative synthesis of visual analytic results. *Proceedings of IEEE VAST'08*, 67-74.

Shavers, B., & Zimmerman, E. (2013). *X-Ways Forensics Practitioner's Guide*. Newnes.

Shupp, L., Andrews, C., Dickey-Kurdziolek, M., Yost, B., & North, C. (2009). Shaping the Display of the Future: The Effects of Display Size and Curvature on User Performance and Insights. *Human-Computer Interaction*, *24*(1-2), 230–272. doi:10.1080/07370020902739429

Silva, C. T., Freire, J., & Callahan, S. P. (2007). Provenance for Visualizations: Reproducibility and Beyond. *Computing in Science & Engineering*, *9*(5), 82–89. doi:10.1109/MCSE.2007.106

Singh, A., Bradel, L., Endert, A., Kincaid, R., Andrews, C., & North, C. (2011). Supporting the cyber analytic process using visual history on large displays. *Proceedings of the 8th International Symposium on Visualization for Cyber Security*, 1-8. doi:10.1145/2016904.2016907

Waldner, M., Puff, W., Lex, A., Streit, M., & Schmalstieg, D. (2010). Visual links across applications. *Proceedings of Graphics Interface 2010*, 129-136.

Wenwen, D., Dong Hyun, J., Stukes, F., Ribarsky, W., Lipford, H. R., & Chang, R. (2009). Recovering Reasoning Processes from User Interactions. *IEEE Computer Graphics and Applications*, *29*(3), 52–61. doi:10.1109/MCG.2009.49 PMID:19642615

Wright, W., Schroh, D., Proulx, P., Skaburskis, A., & Cort, B. (2006). The Sandbox for analysis: concepts and methods. *Proceedings of ACM CHI'06*, 801-810. doi:10.1145/1124772.1124890

Yost, B., Haciahmetoglu, Y., & North, C. (2007). Beyond visual acuity: the perceptual scalability of information visualizations for large displays. *Proceedings of ACM CHI'07*, 101-110. doi:10.1145/1240624.1240639

Haeyong Chung is an assistant professor of computer science at University of Alabama in Huntsville. His research interests include visual analytics, information visualization, and supporting sensemaking with display ecologies and large high-resolution displays. He received his Ph.D in Computer Science from Virginia Tech.

Andrey Esakia is a Collegiate Assistant Professor at Virginia Tech in the Department of Computer Science.

Eric D. Ragan is an Assistant Professor in the Department of Computer & Information Science & Engineering at the University of Florida, United States. He directs the Interactive Data and Immersive Environments (INDIE) lab, which conducts research of human-computer interaction, visual analytics, virtual reality, and explainable intelligent systems. He received his Ph.D. in Computer Science from Virginia Tech.