Modeling and Evaluating User Behavior in Exploratory Visual Analysis

Khairi Reda
Argonne National Laboratory
kreda@anl.gov

Andrew E. Johnson
University of Illinois at Chicago
ajohnson@uic.edu

Michael E. Papka
Argonne National Laboratory
& Northern Illinois University
papka@anl.gov

Jason Leigh
University of Hawai‘i at Mānoa
leighj@hawaii.edu

ABSTRACT

Empirical evaluation methods for visualizations have traditionally focused on assessing the outcome of the visual analytic process as opposed to characterizing how that process unfolds. There are only a handful of methods that can be used to systematically study how people use visualizations, making it difficult for researchers to capture and characterize the subtlety of cognitive and interaction behaviors users exhibit during visual analysis. To validate and improve visualization design, however, it is important for researchers to be able to assess and understand how users interact with visualization systems under realistic scenarios. This paper presents a methodology for modeling and evaluating the behavior of users in exploratory visual analysis. We model visual exploration using a Markov chain process comprising transitions between mental, interaction, and computational states. These states and the transitions between them can be deduced from a variety of sources, including verbal transcripts, videos and audio recordings, and log files. This model enables the evaluator to characterize the cognitive and computational processes that are essential to insight acquisition in exploratory visual analysis, and reconstruct the dynamics of interaction between the user and the visualization system. We illustrate this model with two exemplar user studies, and demonstrate the qualitative and quantitative analytical tools it affords.

Keywords
Exploratory visual analysis, visual exploration, evaluation, insight-based analysis

1. INTRODUCTION

Exploratory visual analysis represents one of the main use cases for interactive visualizations. Often, the primary reason for employing visualization is to enable users to explore their data from a flexible point of view so that they can acquire insights and develop new questions [12, 47, 48]. Visual exploration can be driven by undirected observation of patterns, outliers, and salient features in the visualization [44], or it can be guided by hypotheses, intuition, or prior exploratory goals [24]. These bottom-up and top-down processes interact in complex ways [18, 49], making visual exploration a highly fluid and emergent activity. Naturally, the manner in which this activity unfolds is likely to impact insight acquisition, and ultimately the success of the visualization. Therefore, when evaluating interactive visualizations, it is essential to examine how they are used in realistic scenarios to assess their effectiveness in scaffolding productive visual analysis.

Nevertheless, the visualization community has mostly focused on evaluating visualizations in terms of outcome, primarily by measuring low-level performance indicators such as task completion time and accuracy [10]. Similarly, techniques have been developed to characterize insights acquired by users in open-ended analysis scenarios [28, 40, 43]. These methods enable the evaluator to measure the visualization’s usefulness in supporting a particular analysis task, or in engendering new insights. However, it is difficult from these methods alone to get a sense of how and why a particular visualization was effective.

Recently, there has been interest in descriptive methods that focus on evaluating the problem-solving strategy of users as they engage in visual exploration [25]. Such methods often involve the analysis of interaction events from log files to reconstruct some of the dynamics between the user and the visualization tool [32]. Although this type of low-level analysis can sometimes reveal recurring interaction patterns that could shed a light on how users employ a particular visualization tool, it is often difficult to uncover the reasoning processes behind these interactions by looking exclusively at log files. This makes it difficult to understand whether and how the visualization was successful in engendering insights. To overcome this limitation, the evaluator could consider additional sources of information, such as verbal statements collected during interviews and think-aloud experiments, or by observing users’ viewing and interaction behavior from video recordings. This data can provide a richer picture of users’ problem-solving behavior and uncover potential usability problems in the visualization [17]. However, in exploratory analysis, there can be enormous variation in the high-level strategies of users, which are likely to be sensitive to the dataset as well as to differences in the problem-solving styles of individuals. Ul-
timately, it may not be clear what strategy constitutes an optimal exploratory behavior. What is therefore needed is an evaluation methodology that can bridge the gap between log-file-based analysis techniques, on the one hand, and qualitative, descriptive evaluation methods commonly employed in visualization design studies, on the other hand. Such a methodology could ultimately enable the evaluator to model the dynamic processes of visual exploration, characterize the analytic behavior of users, and quantitatively assess the effectiveness of the design in scaffolding analyses. In this article, we introduce such a method by extending traditional insight-based evaluation techniques [28, 40] to incorporate a wider range of mental processes and interactions events that are extracted from verbal protocols, video and audio recordings, and log files. We model the exploratory behavior of users using a Markov chain process and capture transitions between mental, interaction, and computational states. This model makes it possible to quantify a user’s probability of ‘moving’ between the processes that are essential to insight in visual analysis, while making it easier to characterize the high-level exploratory behavior of users with intuitive state transition diagrams.

This article grew out of a paper presented at BELIV’14 [36], a workshop that was dedicated to discussing novel evaluation methods for visualizations. In this expanded article we contribute a theoretical foundation for the proposed methodology and describe one additional case study to further demonstrate its applicability. We first discuss related work in visualization evaluation and place our work in context. We then present the conceptual foundation for our methodology and discuss procedures to implement it in empirical studies of interactive visualizations. We then present two exemplary user studies that demonstrate how one could utilize the methodology to characterize the exploratory behavior of participants as they engage in visual analysis. Findings from the two studies were previously reported in [37, 38]. The discussion here, however, is focused on how the proposed methodology was employed in either study to support the stated evaluation and analysis goals. We discuss commonalities between the two studies, reflect on our proposed methodology, and consider its limitations.

2. RELATED WORK

We review some of the prominent evaluation validation models for visualizations and discuss where our methodology lies with respect to these models. We then discuss related evaluation methods that are designed to assess the data analysis and problem-solving processes of users.

2.1 Visualization validation models

The topic of evaluation has been gaining increasing attention from visualization researchers and practitioners alike, reflecting the need for robust methods to validate visualization research [10]. Nevertheless, evaluation remains a challenging aspect of visualization analysis and design [13]. One of the main sources of complexity is the diversity of contexts in which visualizations are used, making it difficult for the evaluator to choose appropriate methods for their evaluation needs. Consequently, some research has been done in an attempt to organize the plethora of evaluation methodologies into unified models. One example is Lam et al.’s taxonomy of evaluation scenarios, which is intended to guide researchers in deciding on what evaluation goals to consider, and which methods are appropriate to address various questions that arise in user studies [22]. Their framework suggests two broad evaluation foci: the visualization (i.e., visual encoding), or the data analysis process it scaffolds. The methodology we propose here aims at characterizing the exploratory behavior of users and how that behavior is influenced by visual encodings and interactions in the visualization. In Lam et al.’s taxonomy, our method is thus concerned with evaluating visual data analysis and reasoning, and assessing the visualization’s effectiveness in supporting these processes. With respect to Munzner’s visualization analysis model which characterizes visualization design and validation as a nested process [26], our methodology can be used to validate interaction design and visual encoding. Specifically, we provide a method to examine how a particular visualization design shapes the analytic behavior of its users, and whether it supports (or hinders) insight acquisition.

2.2 Evaluating the data analysis process

One important consideration in designing empirical visualization studies is the choice of a representative set of tasks. Visual analysis tasks range from low-level operators, such as finding clusters in a graph and acquiring detailed information about a specific data point [1, 42], to highly-complex, ill-defined tasks, such as characterizing cause-effect relationships and gathering intelligence from document collections [2, 30, 47]. Nevertheless, evaluation efforts have predominantly focused on validating visualizations under well-defined tasks by measuring low-level performance indicators (e.g., task completion time and accuracy) in controlled experiments [10]. However, users often employ visualizations in an open-ended manner without knowing a priori what tasks they want to accomplish, or what is it that they are looking for in the data. Consequently, controlled experiments often fail to capture the complexity of user behaviors and interactions that are likely to occur in the real world [31]. In the last few years, there has been a steady increase in empirical studies that have attempted to assess users’ visual analysis processes in realistic and open-ended scenarios [20]. Generally, these studies can be categorized under four themes:

- Informal case studies that report on feedback from users about their experience, and how the visualization supports their analysis (e.g. [39]).
- Qualitative observation of user interaction with the visualization tool (e.g. [4]), and their problem-solving strategy [25].
- Quantitative analysis of the insights acquired by users in open-ended, exploratory experiments [11, 28, 40]
- Quantitative analysis and modeling of users’ low-level interaction logs in open-ended, exploratory scenarios [9, 32].

Informal observational studies provide a rich, high-level description of user workflows, but the evaluation risks being sensitive to individual differences and high-level variations in strategy of users. Insight-based evaluation provides concrete measures of the visualization’s usefulness in spurring
discoveries, but does not address questions on how users acquire such insights, making it difficult to use these results alone in improving the design. Lastly, analysis of event logs could uncover interesting patterns in user interaction, but reconstructing users’ analytical processes from such interaction logs remains doubtful.

The above limitations have motivated researchers to attempt to combine multiple procedures when evaluating interactive visualizations, including analysis of think-aloud protocols, interactions logs, and eye tracking data [7, 23]. Our methodology builds on the idea of integrating multiple streams of data to give the evaluator a more holistic view of how users interact with visualization systems. By simultaneously looking at users’ interaction patterns and mental operations, we get a richer picture of their visual analysis process. This enables the evaluator to study how users acquire insights from their interaction with the visualization. Behind this methodology is a Markov-chain-based behavioral model, which supports both qualitative and quantitative analyses of user behavior in exploratory visual analysis.

3. METHODOLOGY

Our goal in this methodology is to characterize the exploratory behavior of users so that we can begin to understand how this behavior is potentially influenced by variations in the design of the visualization interface. Ultimately, this would enable us to evaluate the effectiveness of visualizations for exploratory analysis. However, rather than simply assessing the outcome of the visual analytic activity, we want to detect potential patterns in how users interact with a particular visualization, and understand how such interaction can either foster or hinder insight. We believe these two questions can be addressed at a micro-analytic level. In this sense, a micro characterization of the visual analytic process relates to patterns in the transitions between ‘mental’, ‘interaction’, and ‘computational’ states. That is, moments when the user is performing mental computation in his/her head, and moments when he/she is interacting with the visualization, and hence offloading some of the information processing onto the visualization tool.

This perspective reflects a higher-level characterization compared to Amar et al’s analytic operators [1], but also constitutes a finer-grained analysis compared to descriptive methods typically employed in observational user studies [25]. We discuss the rationale behind our choice of this unit of analysis and propose a framework for modeling user behavior in exploratory visual analytics. Following that, we outline an analysis procedure for inferring mental and interaction states from verbal transcripts, log files, and video recording.

3.1 Behavioral model

Visual analysis is a cognitively distributed process [19], in which the user interacts closely with a visualization tool to explore an information space and acquire new insights. In this role, the visualization tool harnesses the computational power of modern computers to process and transform large quantities of information into visual representations. The user, on the other side, contributes a “flexible pattern fnder coupled with an adaptive decision-making mechanism” [49]. The cognitive function of the system is thus a product of mental and computational processes that are coordinated through interaction with the visualization tool. Figure 1 presents a simplified conceptual framework of the visual analytic system based on this perspective [29, 41]. In this model, information processing is distributed among three main spaces:

- Mental space: comprises cognitive computation carried out mentally be the user (e.g., deduction, generalization, and hypothesis formulation).
- Interaction space: comprises actions performed by the user to modify the state of the visualization (e.g., filtering and navigation).
- Computational space: comprises data-processing and analysis algorithms executed by the computer (e.g., data normalization and cluster analysis).

To effectively model the behavior of this system, we need to be able to recognize and capture analytic phenomena that occur across the three spaces. From the evaluator’s perspective, the inner working of these processes maybe difficult to discern. For instance, it may be difficult to characterize the full decision-making process a user employs to make a domain observations about the data, or to infer the user’s mental model. However, the evaluator can delineate abstract, domain-independent processes that they believe to be relevant to the exploratory task. For instance, the evaluator can detect and tag insight-generating mental operations such as deductions, generalizations, and hypotheses by analyzing users’ verbal statements. The evaluator can also easily recognize different types of interactions, such as filtering and navigation, by recording these as events in a log file or coding them from videos or screenshots. Similarly, invocation of computational processes, such as cluster analysis and machine learning, can be recorded into a log file. Once the relevant processes are recognized and tagged, a user’s analysis activity can then be characterized by the sequence of these processes.

For example, in a graph-based social network visualization, a user may first set a goal to explore highly-connected individuals, by consecutively selecting nodes that appear to have a higher number of incident edges. The user may observe a certain attribute that is shared by these nodes, leading him/her to hypothesize that the attribute is correlated with leadership roles. This may in turn prompt the user to set a secondary goal to explore subgroups centered around
high-degree nodes, by invoking cluster analysis and creating a cluster view of the network. This sequence of operations together with the various input parameters (e.g., number of expected clusters) defines a unique exploratory trajectory [21]. Although one may be interested in the above trajectory, it is likely that such trajectories are unique to individual users and/or independent exploratory sessions. Therefore, rather than being concerned with a particular sequence, we are interested in potential patterns sequences emerging from multiple exploratory sessions, for different individual users.

As the fundamental activity in visual analytics is centered on the successive transformation of visual and mental representations (by means of interaction) in support of insight, we focus on characterizing transitions between mental, interaction, and computational processes. This point of view enables us to address a wide range of evaluation questions, by considering the probability that certain interactions will eventually lead to insights. For instance, in the example above, the evaluator may be interested in knowing whether the cluster view is useful for eliciting new questions about the network, by quantifying the probability of a transition from create cluster view to formulate hypothesis. More generally, the evaluator could be interested in knowing what mental processes are likely to occur after two or more transformations following exposure to the cluster view. This type of analysis can be carried out by representing visual exploration as a Markov chain process [27]. In this model, the states correspond to mental, interaction, or computational processes within the visual analytic system, with edges reflecting the probability of moving between states. Figure 2 illustrates this model.

The notion of transition probability, as opposed to process sequence, enables us to generate a behavioral model that captures typical behavior by inferring average transition probability from multiple individual users over several exploratory sessions. During this process, it is important to maintain actual ordering of the mental, interaction, computational processes (based on their timestamps) to correctly calculate transition probabilities. In addition to maintaining the ordering of processes, the evaluator may also be interested in measuring the time spent at each state. While it is possible to accurately infer the temporal ordering of processes, from timestamped verbal statements and event logs, delineating the exact temporal boundaries (and hence the time spent at each state) may be difficult, particularly for mental states that involve non-verbally evidenced reasoning chains. Furthermore, due to differences in analytical abilities and visualization literacy, users may exhibit wide variation in the time spent at each state, particularly for interaction states that require a series of pointer-based actions. For this reason, we focus on analyzing transitions between states in the Markov chain model, as opposed to measuring times spent at each state. However, when generating an average Markov chain from multiple users in open-ended experiments, we normalize transition frequency for each user by the total length of time it took him/her to complete the experiment (excluding training time). This has the effect of converting transition frequencies to rate units, enabling us to quantitatively compare the behavior of participants who spend different times on the experiment.

The main limitation of our Markov chain model is its memoryless nature. We hence lose the ability to observe high-level strategies that manifest in long exploratory trajectories. However, we shall see that even with this memoryless model, we are able to observe how users employ a particular visualization tool, and reveal differences in behavior given variations in the visualization.

3.2 Analysis procedure
Our methodology is primarily focused on characterizing behavioral patterns in open-ended, think-aloud user studies, in which participants are given a visualization tool and asked to freely explore a dataset. We therefore assume that the evaluator has access to the following pieces of data. We also assume that it is possible to synchronize the different data sources to a global time reference (i.e., it is possible to reconstruct a global ordering for events extracted from these difference sources):

- Video/audio recordings and/or screenshots. Ideally, this data should clearly depict the contents of the visualization display as well as the participants’ interactions, viewing behavior, and verbal utterances.
- Text transcripts of participants’ verbalizations produced from the audio record [15].
- Log files containing a record of events in the visualization tool, including user interactions and computational analysis processes performed on the data.

The first analysis step in our proposed methodology is to identify mental, interaction, and computational processes that comprise the backbone of the exploratory task, and develop an appropriate coding scheme to capture them. These identified processes will comprise the states in the Markov chain model. Once a coding scheme is developed, the evaluator proceeds to analyze the data and tag evidenced instances
of these processes. This involves the coding of qualitative data (i.e., transcripts and videos) as well as extracting events from structured data (i.e., log files). The second step is to analyze the flow between these processes by creating Markov chains that depict behavior of users. We describe both of these steps in detail.

### 3.2.1 Establishing core computational, interaction, and mental states

The first task for the evaluator is to devise a coding scheme to identify and tag processes that are relevant to the exploratory task and the visualization tool. Mental processes can be inferred from verbal statements uttered by subjects, whereas computational events (e.g., invocation of cluster analysis), can be identified from log files. In most situations, it is relatively straightforward to extract interaction events from log files. Alternatively, the evaluator can infer and tag user interactions directly from the videos or screenshots.

The evaluator may at first adopt an open coding approach, by looking for recurring interactions and behaviors evidenced from the video artifacts, verbal transcripts, and log files. The evaluator can also be informed by knowledge acquired during the visualization design process to establish which interactions are crucial in providing task-support, and what mental processes are relevant to the analysis task. Generally, in exploratory visual analysis, the evaluator may be interested in capturing mental processes that reflect the articulation of exploratory goals or insight acquisition (e.g., observation making and hypothesis generation). Additionally, the evaluator is likely to be interested in interactions that imply intention to search the information space, such as navigation and querying. It is also desirable to merge related codes into more abstract categories to reduce the complexity of the resulting model. Such abstraction and grouping of related codes can also be informed by established interaction taxonomies in visualizations [50] or task typologies [8]. For instance, panning and zooming, scrolling, and following hyperlinks can be coded as ‘navigation’, while clustering and multidimensional scaling can be regarded as ‘structure analysis’ processes. The final set of categories in the coding scheme will correspond to the states in the Markov chain model.

It is important to emphasize that, given the different evaluation scenarios and the diversity of behaviors that are likely to emerge in exploratory visual analysis, the evaluator should have a wide latitude in deciding on what processes and phenomena to capture in the model. Ultimately, this decision will be driven by the evaluation goals, but can also be informed by what the evaluator sees in the data. In both of the two exemplar studies we present, we started with an open coding process, refining the codes over multiple passes to arrive at the final coding scheme. However, we were also informed by earlier studies on insight-based evaluation, and hence we incorporated some of the categories developed by Saraiya et al. to tag insights reported by participants [40].

A complicating factor arises when delineating the state boundaries in systems with processes that span two of the three spaces in Figure 1. In particular, visual analytic systems often tie interactions in the visual space with background computational processes that attempt to infer the user’s mental model and steer the computational analysis accordingly [3, 6, 14]. When establishing states for these closely coupled processes, the evaluator has the choice of merging interaction and computation into a single state, or treating them as two separate states. A good guideline here is to consider whether the interaction-coupled computational process can be invoked by way of another state, in which case it warrants a distinct state in the model. Otherwise, the computational process can be merged with the interaction state resulting in a combined computational-interaction state.

### 3.2.2 Markov chain modeling

After establishing and tagging the various processes in the system, we create Markov chain models that reflect users’ analytic behaviors by highlighting transitions between these processes. First, the evaluator establishes a process-sequence for the exploratory activity of individual users; for every evidenced instance of a mental, computational, or interaction process we add an event to the sequence while maintaining the temporal-ordering of events. The resulting sequences can be thought of as an extended log that captures processes within the combined human-computer cognitive system. Once a process-sequence is established for each individual user, the next step is to create a Markov chain model from this sequence. The states in the Markov chain correspond to categories of mental, interaction, or computational processes established in the first analysis stage (see previous section). For every pair of consecutive events in the sequence, we record one transition between the two corresponding states. For instance, if the sequence indicates that a user had reported a hypothesis after performing a pan and zoom interaction, we record a transition from ‘navigation’ to ‘hypothesis formulation’. We also record reflexive transitions, which start and end at the same state. For example, two distinct observations reported successively by the user can be recorded as a reflexive transition from and to ‘observation making’. The resulting Markov chains can be visualized as node-link diagrams or adjacency matrices to depict transition frequency (or probability) between the states. This analysis produces a separate Markov chain for every individual participant in a user study. However, the evaluator can also create an ‘average’ model that reflects typical user behavior by averaging transition frequencies from multiple participants. In this case, transition frequencies can be normalized by the length of the exploratory activity to account for time variations in open-ended experiments.

Visualization of the Markov chain model provides a convenient way for the evaluator to get a qualitative sense of users’ exploratory behaviors, to and recognize interdependencies between different analytic processes. Furthermore, the Markov chain model is amenable to quantitative analysis, which enables the evaluator to conduct comparative analysis across multiple experimental conditions to reveal potential effects for design variations or individual differences on user behavior. In the following sections, we illustrate how both of these analyses can be performed with two exemplar studies. The first is a case study aimed at evaluating the usefulness of a particular visualization tool for the analysis of ensemble data. The goal of the second study was to empirically investigate the general effects of increasing the physical size and resolution of the visualization interface on exploratory behavior and insight acquisition.
We used a 19 Megapixels tiled display with physiological dimensions of 7x3 meters. The visualization environment divided the screen vertically into a number of semi-independent workspaces. Workspaces can be set to group collections of ant trajectories based on the underlying experimental condition, thus enabling the user to juxtapose and compare different trajectory groups. Trajectories were arranged in a small-multiples layout [46] within each workspace. The different workspaces were given distinct background tints to make them more distinguishable. Figure 3 illustrates the visualization environment.

4.2 Interactive features

Navigating the dataset is achieved by changing the experimental condition associated with each workspace via a set of filters. For instance, one workspace can be set to show trajectories of ants captured east of the colony’s main foraging trail, while a second workspace can be set to show trajectories of ants captured on the trail while carrying a seed.

To facilitate comparison and correlation, we provided a visual query-by-example feature that lets the user interactively select trajectory segments and highlight experiments in which the insects had exhibited similar spatio-temporal movement patterns. This is achieved by enabling the user to brush the background of a single trajectory, causing the visualization to highlight trajectory segments that spatially intersects with the painted selection (see Figure 3, inset).

Additionally, a temporal filter enables the user to limit the visualization to trajectory segments that corresponds to motion within a specified time window (such as the first 30 seconds of the experiment).

Using the query-by-example brush and the temporal filter in concert, the user can specify a movement pattern he/she is curious about. Brushing actions propagate across all workspaces, instantly revealing trajectories that exhibit spatio-temporal similarities with a perceptually salient high-light. This action can performed in-place without disrupting the small-multiples layout, or modifying the contents of the display, which presumably reduces the cognitive overhead associated with this operation. Additional details about the design decisions are described in [35].

4.3 Analysis

The participant was given an hour to explore and analyze her data and instructed to think aloud. The session was video and audio-recorded, and a verbal transcript was produced from the audio record. We employed a two-pass open coding process to analyze the verbal transcript and the video data. In the first pass, we focused on capturing the user’s low-level interaction with the visualization tool, by tagging interactions such as changing the contents of the workspaces, utilizing the query-by-example brush, and filtering by time.

We elected to code interactions from the video record as opposed to extracting them directly from log files, in attempt to get a deeper understanding of what actions the user was trying to perform. Additionally, we also coded observations, questions, and hypotheses from the verbal transcript, based on Saraiya et al’s insight categories [40]. In the second pass, we grouped all interactions under two high-level categories, which we presume to incur varying cognitive costs. We also dropped infrequent mental operations (e.g., question formulation), while adding a separate category for observations that involved an apparent decision-making process. Ulti-
Figure 4: A state transition diagram illustrating the Markov chain model. The states correspond to key processes identified from the verbal transcripts and the video data. Mental states are marked in blue whereas interaction states are marked in green. The weights of arcs reflect transition frequency between states.

The coded mental and interaction events were sorted by timestamp and merged into one sequence, which was used to create the Markov chain model (see Figure 4). As this was a case study involving a single user, we chose to represent them as opposed to probability.

4.4 Findings

Once workspaces were configured, the user spent significant time investigating relationships between trajectories that were visible on the display using the query-by-example feature, before switching to different groups. Interestingly, this action coincided with the articulation of new hypotheses and questions in most of the time. For instance, upon seeing that ants captured east of the colony’s main foraging trail exhibit a direct movement towards the west side, the user brushed the west side and noticed a majority of them had a red highlight (see Figure 3, inset). This led the user to hypothesize that ants would attempt to head in the direction of the colony’s trail when released in an effort to locate pheromone cues that would lead them back to the colony’s nest. Upon seeing how different workspaces reacted to brushing, the user would quickly articulate a different pattern to be tested, which would in turn trigger a new hypothesis. For example, upon seeing that ants captured on the trail did not exhibit a similar directed motion, the researcher proceeded to brush the starting point in the trajectory with a green color, hypothesizing that “off-trail ants should start green and turn black faster [than their on-trail counterparts], because they know where they’re going”.

This behavior is neatly illustrated in the Markov chain model; the major feature in Figure 4 is the strong interrelation between Query by example and Hypothesis formulation, which manifests in frequent transitions between these two states. Conversely, transitions to Workspace management are less frequent indicating a reduction in the rate of virtual navigation, as one would expect with a large display interface [5]. Interestingly, there are relatively few transitions to Observing outliers, which suggests that the exploratory activity was predominantly driven by top-down goals that reflect the current working hypothesis, as opposed to non-directed search in the dataset. This behavior is also consistent with our design goal of enabling the user to explore a large hypothesis space regarding the effects of environmental context on insect behavior.

The state transition diagram provides evidence to support the efficacy of small-multiples when coupled with a visual query-by-example feature in promoting the exploration of large ensemble datasets. The analysis also raises some interesting questions on the effects of using larger displays with more pixels on exploratory visual analysis. Could these displays, by reducing the need for cognitively disruptive interactions (e.g., virtual navigation and window management operations), encourage users to invest more effort in exploring their data, and ultimately acquire more insights? In the following section, We employ our proposed evaluation methodology to investigate this question in a comparative user study.

5. STUDY II

The second study was aimed at characterizing the general effects of increasing the size of the display on user behavior and insight acquisition in exploratory visual analysis. We empirically investigate this effect within the context of an open-ended user study involving the visual analysis of crime patterns in a major metropolitan area. The visualization interface was also built around a small-multiples design, enabling participants to explore crime categories in multiple years, and in different parts of the city. To study the effect of display size, we varied the physical size (and resolution) of the visualization display, which implicitly modulated the amount of information a participants can see at a time. We give an overview of the study design, focusing the discussion on how we applied our proposed methodology to characterize and quantitatively compare the exploratory behavior of participants under two experimental conditions. We refer the reader to [37] for complete details of the study.
5.1 Study design

We used a between-subjects design with a single independent variable: The display size (Small vs. Large). Half of the participants undertook the study using a small display while the other half experienced the visualization environment on a large display. The study took place in the CAVE2 environment which consists of a cylindrically shaped 18x4 tiled display [34]. The small display condition utilized 3 of the 18 available columns available in CAVE2, giving participants a resolution of 4,098x3,072 (12.5 Megapixels) and approximately a 40-degree field-of-view. The large display condition utilized 13 columns, giving participants a resolution of 17,758x3,072 (54.5 Megapixels) and approximately a 190-degree field-of-view. Figure 5 illustrates these two conditions. We employed the same visualization interface in both experimental conditions. However, the size of the display served to implicitly modulate the number of small-multiple views that can be seen simultaneously; while the small display required participants to frequently switch between views, the large display afforded the ability to see more view side-by-side, thus reducing the need for view switching.

5.2 Participants and procedures

We recruited ten volunteers to participate in the study (four female). The experiment began with a 15 minutes training session with the experimenter explaining the task and demonstrating the visualization environment and its various interactive features. Participants were then given 150 minutes (2.5 hours) to explore the Chicago crime dataset, and instructed to think aloud during the activity and report interesting observations, salient patterns and outliers, correlations, trends, as well as hypotheses that explicate their observations. Participants, however, were free to end the experiment earlier if they felt that they had exhaustively explored the dataset. The session was video and audio recorded.

5.3 Analysis

Our main focus in this analysis was on the video and audio data, which contained a record of participants’ verbal utterances as well as the state of the visualization, at the time. Since our main goal in this experiment was to study the effects of increasing the size of the display on participants’ exploratory behavior, we focused on capturing insight-generation mental processes as a key component of exploratory visual analysis. Building on the coding scheme we developed from the first study, we utilized Saraiya et al.’s insight categories [40]. Additionally, we were interested in how and when participants formulated their exploratory goals during the study.

Two participants were selected for an initial coding pass. We then refined the codes in a second pass, merging and dropping infrequent categories. The final coding scheme for the verbal data consisted of the following codes:

- **Observation**: A unit of knowledge acquired from looking at and interacting with the visualization.
- **Hypothesis**: A conjuncture made by the participant, often as a result of making a series of observations.
- **Goal**: A statement reflecting an exploratory objective.

The three codes above comprised three core mental states: *Make observation, Formulate hypothesis*, and *Form goal*, respectively. As for interaction states, we were interested in two general classes of interactions:

- **Layout-preserving interaction**: Comprises actions that do not result in major changes to the state and layout of the visualization environment. Interaction events coded here include brushing-and-linking and panning the map.
- **Layout-changing interaction**: Comprises action that result in major changes to the state of the visualization, potentially requiring the participant to rebuild his/her ‘mental map’ [33]. This category comprised the following interactions: Creating, closing, positioning views, changing the year or the crime category in one or more views.

We coalesced the above codes into two core interaction states (corresponding to the two classes above): *Brush, link, pan map* and *Modify layout*. We hypothesize that these two classes of interactions would incur varying cognitive costs, and expect them to occur with varying frequency in the two experimental conditions. Furthermore, we were interested in whether and how differences in the incidence of these two interactions would affect transitions to mental states.

5.4 Results

Combining coded verbal statements and interactions into a single, timestamp-ordered sequence, we created an extended activity log for each participant, which was in turn used to create a Markov chain for each participant using the procedure outlined in earlier (see Methodology: Analysis procedure). To account for differences in exploration time, we normalize transition frequency for each participant by the time the participant had spent on the study. Figure 6 illustrates the Markov chains for participants S5 (who undertook the experiment with the small display) and L5 (who utilized the large display). From the transition diagram, we observe that S5 had to perform layout-disruptive operations more frequently compared to L5. Moreover, we observe higher transition probability to goal setting and hypothesis formulation states in L5’s diagram, which suggests that the participant was able to devise top-down exploratory goals more often during the study.
Figure 6: Two state transition diagrams illustrating differences in the exploratory behavior of participant S5 who used the small display (left), and participant L5 who utilized the large display (right). Mental states are depicted in blue whereas interaction states are shown in green. The weight of edges represent transition probability between states (with log transformation applied); darker arrows denote higher transition probability.

To understand systematic variations in participant behavior between the two experimental conditions, we averaged transition frequencies for participants under the same condition, which yielded two ‘average’ Markov chains corresponding to the small and the large displays. To guarantee equal contributions from each participant to the average diagrams, we normalized transition frequencies for each participant by the time it took him/her to complete the activity. Figure 7 shows the two average transition diagrams side-by-side, illustrating important differences in the overall exploratory behavior of participants. Figure 8 highlights these variations with a ‘difference’ diagram and transition matrix, showing the relative changes in transition probabilities as a result of increasing the display size.

Figure 7: Two transition diagrams illustrating ‘average’ behavior of participants under the small (left) and large (right) display conditions.

The large display diagram is marked by a decrease in the transition to the Modify layout state (column 1 of the transition matrix in Figure 8), indicating that participants were less likely to initiate layout-disruptive interactions on the large display. We also see decreased transition probability to Brush, link, pan map, indicating that participants were also less likely to initiate brushing-and-linking and map panning operations (column 2 of the transition matrix). However, generally, we see an increased tendency for participants to transition from interaction states to insight-generating mental states with the large display (columns 4 and 5). Furthermore, we see an increased likelihood for participants to remain in these states with the large display (cells [4,4] and [5,5] of the adjacency matrix). In particular, post-hoc analysis indicates a significant increase in the probability of remaining in the Make observation state (t(8) = 4.995, p < .001), indicating that participants are more likely to make a series of consecutive observations.

Overall, these variations suggest that the large display was more effective at triggering insight-generating mental processes and keeping participants in the ‘cognitive zone’ [16], where they are likely to continue to acquire insights. Furthermore, we see a slight increase in the tendency of participants to transition to Form goal, which suggests that they were able to set more ambitious exploratory goals for themselves. One possible explanation for this is that the cognitive costs involved in the pursuit of such goals was perceived to be less on the large display, given the ability to simultaneously consult a larger number of views with little need for disruptive navigation interactions.

6. DISCUSSION

The two studies provide evidence to suggest positive effects for adopting larger displays in exploratory visual analysis. Specifically, we see a change in user behavior marked by a reduction in the transition to interaction states, particularly for layout-disruptive interactions that are likely to incur higher cognitive costs, such as Workspace management in Figure 4 and Modify layout in Figure 7. In both studies, the Markov chains suggest that users are adapting their strategy by reducing the rate of interaction and focusing instead on mental processing of the information conveyed by the visualization. Interestingly, this reduction in interaction rate was accompanied by increased likelihood in transitioning to and remaining in insight-generating mental states. Our Markov chain models enabled us to qualitatively describe these effects, and to quantify differences in user behavior given variation in display size.
The methodology we proposed affords us a micro-view onto the analytic behavior of users as they engage in open-ended data exploration with the aid of interactive visualizations. From this perspective, we are able to shed a light onto behavioral patterns that occur at a relatively fine temporal grain. Such analysis compliments existing evaluation methodologies for visualizations, which typically concern themselves with the high-level user strategy, or, alternatively, focus exclusively on the analysis of interaction patterns derived from event logs. Our methodology can be regarded as a middle-ground approach between these two endpoints of the spectrum. By incorporating mental, interaction, and computational processes in our analysis, we are able to provide a richer view onto the processes that are essential to insight in exploratory visual analysis. While the two studies described in this article were aimed at investigating the affordances of large high-resolution displays, the methodology and the analyses presented should be applicable to a wide variety of evaluation scenarios.

However, our proposed methodology has a number of limitations that may restrict its applicability and feasibility in some evaluation scenarios. First, the Markov chain model is entirely memoryless, which makes it incapable of expressing the high-level strategy of users. Analyzing high-level analytical processes requires a stateful model that can capture sequences spanning multiple mental, interaction, and computational operations, which is not currently possible with our model. A second limitation lies in the time that is necessary to gather the data for this type of evaluation, which could limit its applicability. In particular, the analysis and coding of verbal transcripts is an inherently labor-intensive process [45]. Compared to traditional qualitative evaluation, our methodology requires significantly more time and effort, which can make it impractical for a large number of participants.

7. CONCLUSIONS

We proposed a methodology for modeling and characterizing user behavior in exploratory visual analysis. We model the micro-analytic behavior of users using a Markov chain process, capturing transitions between mental, interaction, and computational processes that are essential to insight acquisition. We demonstrated this model and the analytical tools it affords with two example studies that investigated the effects of increasing the size of the visualization interface on users’ exploratory behavior. We believe that the methodology is particularly useful in revealing quantitative and qualitative differences in user behavior, given variations in the characteristics or the design of the visualization interface.

8. ACKNOWLEDGMENTS

This article originally appeared as part of the 5th BELIV workshop on novel evaluation methods for visualizations. We thank the workshop organizers and attendees for the insightful discussions, which ultimately led to this expanded article. We would also like to thank the anonymous reviewers for their comments and feedback on an earlier manuscript. This work was supported in part by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357.

References


