

# IBM Research Report

## An Empirical Study of User Interaction Behavior during Visual Analysis

**David Gotz, Michelle X. Zhou**  
IBM Research Division  
Thomas J. Watson Research Center  
P.O. Box 704  
Yorktown Heights, NY 10598



# An Empirical Study of User Interaction Behavior During Visual Analysis

**David Gotz**

IBM T.J. Watson Research cCenter  
19 Skyline Drive, Hawthorne, NY 10578  
dgotz@us.ibm.com

**Michelle X. Zhou**

IBM T.J. Watson Research Center  
19 Skyline Drive, Hawthorne, NY 10578  
mzhou@us.ibm.com

## ABSTRACT

Visual analytics is an emerging discipline in which human analysts are tasked with using interactive visualization tools to make judgments and derive insight from large amounts of dynamically changing information. In a complex and long-running analysis process, we hypothesize that there exist meaningful structures of user interaction behavior. Moreover, we believe that these structures may be used to better understand a user’s analytic goals and reasoning, and to guide the design of a visual analytic system.

To validate our hypothesis, we have conducted an empirical study which examines structures of user behavior during visual analysis and their implications. Unlike previous findings, our study focuses on examining fine-grained user visual interaction behavior over the course of a realistic analysis task. We report both our observations and analysis which uncover two key structures of user visual analytic behavior. Both structures are found to influence user task performance and often reflect limitations or user-desired features of visual analytic systems. Based on our findings, we present a series of design recommendations outlining how these structures can be used to improve visual analytic systems.

## Author Keywords

Visual Analytics, Information Visualization, Smart Graphics, Sensemaking, User Study

## ACM Classification Keywords

H.5.2 [Information Interfaces]: User Interfaces I.3.6. [Methodology and Techniques]: Interaction Techniques

## 1. INTRODUCTION

Over the past several years, major advances in the field of information visualization have enabled people to explore large and complex data sets through interactive software that exploits the human visual system [1, 26]. These efforts have largely focused on *information presentation*, where the entire dataset to be explored is loaded into memory and visualized all at once. For example, a user may explore an entire hierarchical data set using a hyperbolic tree [12] or a tree map [25]. Furthermore, most existing visualization systems provide users with highly specialized tools to satisfy specific analytic goals that are known *a priori*. For example, if a user already knows that her goal is to examine the adjacency relations in a hierarchical data set, she may then

choose to use hierarchical edge bundles to view the desired information [8].

However, many real-world applications, such as business intelligence and market intelligence, require a more dynamic and flexible environment—one that is capable of handling massive amounts of dynamically changing information; and one that supports open-ended visual analysis where the desired analytic goals and insights cannot be known in advance [26]. More importantly, such environments must also be designed to exploit and support the complexities of human *analytical reasoning* during a continuously evolving analysis process [26]. As a result, the emerging discipline of *visual analytics* calls for an enlarged focus on developing effective visual interfaces to facilitate human analytical reasoning for complex analytic tasks.

Toward this goal, we are building *smart visual analytic systems* that can dynamically capture, interpret, and abstract users’ visual interaction activities to preserve insight provenance and to proactively assist users in their tasks. For example, our system may dynamically recommend to users suitable visualization tools or visual interaction paths based on their ongoing interaction behavior.

Our entire effort hinges on a critical hypothesis—that there exist structures of user interaction behavior during visual analysis, and that these structures can be observed and interpreted to help understand a user’s analytic goals and reasoning. In turn, such an understanding can be exploited to guide the design of a smart visual analytic system that can adapt its behavior to best meet a user’s changing analytic needs.

There are many studies that examine user behavior either in specific analytic tasks (e.g., intelligence analysis [14, 24]), in general sensemaking activities [16], or using specific visualization systems [2]. However, we are not aware of any empirical studies specifically designed to examine structures of user interaction behavior during visual analysis. We have therefore designed and conducted our own study to validate our hypothesis.

We asked 30 users to perform two open-ended visual analysis tasks. To avoid potential biases (e.g., variations in system robustness or visualization techniques), we have chosen to use two very different commercial-grade visualization systems as part of our study. Moreover, we studied the use of these systems on realistic visual analytic tasks that require users to visually analyze a large set of real-world data while deriving their answers.

In this paper, we report both our observations and analysis, including the identification of key behavioral structures detected during our study. Furthermore, we discuss the implications of these behavioral structures on user task performance and on the design of a visual analytic system. Based on our empirical findings, we also provide a series of recommendations for both ourselves and the wider research community to guide the design of better visual analytic systems. As a result, our work offers two unique contributions:

- **Empirical study and identification of user behavioral structures during visual analysis and their implications.** Unlike previous work, our study examines structures of user interaction activity during open-ended visual analysis tasks. Based on our observations and analysis, we have identified two prevailing structures of user interaction: patterns and trails. Here we use *pattern* to refer to a sequence of user visual interactions employed repeatedly to achieve a specific analytic goal (e.g., comparison). We define a *trail* as a chain of user activities that actually leads to a point of insight. We analyze the characteristics of both structures, and discuss their implications on user task performance and the functionality of visualization systems.
- **Design recommendations for visual analytic systems based on empirical findings.** We provide a set of concrete recommendations for designing visual analytic systems. Our recommendations are based upon our analysis of patterns and trails and their impact on users' visual analysis performance.

The remainder of this paper is organized as follows. We begin with a review of related work, followed by a description of our study methodology. We then present our key results, including a set of design recommendations based on our analysis.

## 2. RELATED WORK

Our work is closely related to three areas of research: information visualization, sensemaking, and the emerging discipline of visual analytics.

**Information Visualization.** The information visualization community has a long history of validating new technologies via empirical studies [2, 3]. Most often these evaluations focus on comparing a new visual technique to other related approaches [19]. To examine the effectiveness of different techniques, these studies often use a metric to assess end-user performance (e.g., time-to-completion or accuracy [27]). To facilitate a direct comparison of user performance, the tasks and metrics tested in these experiments are typically simple and carefully controlled (e.g., “Locate the file labeled 1990.htm” [11]).

While studies of this type are very valuable, their use in evaluating the effectiveness of visualization systems in support of long-running, complex tasks is limited [18, 21]. Moreover, there are few empirical studies that examine fine-grained user visual interaction behavior over the course of an extended analysis task. Since the goal of our work is to build interactive systems that support complex analysis

tasks, the study presented here is specifically designed to observe and understand user interaction behavior while a user performs realistic, open-ended visual analyses.

Given the greater complexity of the tasks in our study, the most closely related work to ours are the studies of Kobsa [10] and Saraiya et al. [23]. Both of these studies report observations of users performing complex analysis tasks. In each study, the reported observations were used to make comparisons across a set of tools and determine under which conditions each might be most effective. The focus of these studies is once again to validate or compare the effectiveness of individual visualization tools with respect to a set of performance metrics.

In contrast, our study has a starkly different focus. Instead of studying how specific visualization techniques influence task performance, our study is designed to examine characteristics of *user visual interaction behavior* and their impact on analysis results. In fact, in our study we normalize performance metrics across tools to remove any tool-specific differences. This allows us to identify common behavioral structures which lead to better visual analysis regardless of which visualization tool is being used.

**Sensemaking.** The progressive, non-linear nature of visual analysis is closely related to sensemaking—a process in which users iteratively update and refine their mental knowledge schemas as new information is discovered over the course of a task [16, 22]. Sensemaking theory has been used to motivate work in several areas ranging from web-based research [6, 20] to intelligence analysis [5, 15, 17].

Previous studies of sensemaking behavior (e.g., [14, 24]) have not generally focused on aspects unique to visualization environments. However, sensemaking platforms can certainly benefit from the capabilities of visualization tools. Recognizing this fact, some sensemaking systems have incorporated interactive visualization tools as part of a larger system [7, 30].

Unlike previous work, we neither propose a general model of sensemaking nor evaluate a specific sensemaking platform. Instead, our work is complementary to this related field of research. Our study is specifically designed to examine user interaction behavior when using visualization tools in support of sensemaking. Moreover, we propose a set of design recommendations based on our findings which are intended to encourage the development of smarter systems that support deeper and more effective sensemaking.

**Visual Analytics.** Despite the long history of progress in visualization and sensemaking research, there is growing recognition that a more focused effort is required to efficiently support complex analysis tasks. As a result, visual analytics, emerging as a new discipline, focuses on advancing the science of analytical reasoning as supported by interactive visualization interfaces [26]. Most of the existing work in visual analytics has focused on developing interactive visualization systems for specific analytic tasks. For example, there are systems to analyze a large body of text [4], to explore data relationships in a network [9], and to support interactive spatial and temporal analysis [29]. In contrast, our study is designed to directly support these sys-

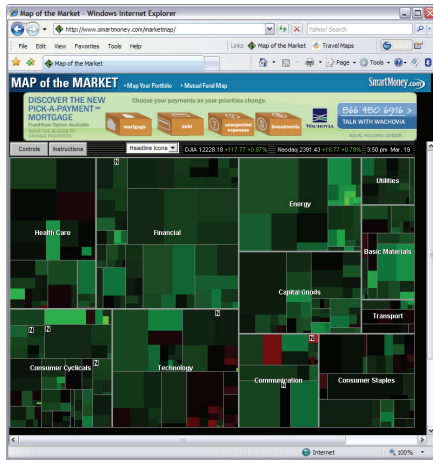


Figure 1. Participants assigned to the financial task were given access to the Map of the Market tool from SmartMoney.com.

tem development efforts, both for our own research as well as for the larger visual analytics community.

### 3. STUDY DESIGN AND METHODOLOGY

Our overall hypothesis postulates that prominent structures of user interaction behavior exist during visual analysis. Moreover, we believe that these structures can help us better understand a user's analytic goals and requirements. To validate our hypothesis, we designed and conducted a user study with two primary goals: (1) to determine if common structures of visual analysis behavior exist; and (2) if so, to characterize those structures in terms of what they represent and how they impact user performance. In this section, we describe several key aspects of our study, including the tasks, participants, methodology, and task performance metrics.

#### 3.1 Tasks

To observe how users behave when working on real-world visual analysis problems, we designed two typical tasks that can be accomplished using commercial-grade visualization tools with access to vast amounts of real data. In both tasks, participants were asked to provide answers to their assigned questions and collect evidence supporting their conclusions.

The first task asked users to assume the role of a stock market analyst and research specific market sectors for one of their clients. Participants were asked to characterize the market, looking for industry leaders and trends to develop investment recommendations. Each person was given access to the Map of the Market visualization tool from SmartMoney.com (Figure 1). They were also allowed to access any web resource directly linked to the tool. Such information includes interactive stock charts, analyst reports, and current events related to specific companies. To observe how users behave over similar tasks, we designed this assignment to contain two sub-tasks, each of which required users to analyze two distinct sectors: technology and finance. For the remainder of this paper, we refer to this as the *financial task*.

The second task had users research business travel facilities. Users were asked to produce a number of alternative sug-

gestions for lodging, dining, and entertaining in specific locations. Participants had to balance several competing priorities, such as budgetary constraints, proximity to corporate offices, and the quality of facilities. Participants were given access to a map-based tool (Figure 2) that is available within our company. This tool allows users to access a wide variety of information, including lodging, dining, and entertainment reviews from the public domain, as well as corporate databases and policies regarding hotels, rental cars, and office locations. As with the first task, participants were restricted to use only directly linked web resources. We also designed this assignment to contain two similar sub-tasks and asked users to research two distinct locations: San Francisco, CA and Burlington, VT. For the remainder of this paper, we refer to this as the *travel task*.

#### 3.2 Participants

We recruited 30 people with varied gender and age to participate in our study. Participants were divided into two groups: fifteen participants were assigned to the financial task and fifteen were assigned to the travel task. Based on our user profiling questionnaire, 73% of the participants reported using visualization tools of some kind at least weekly. However, most of their experience was with common, web-based visualizations such as map-based tools. Two-thirds reported to engage in information analysis activities at least weekly.

In contrast to the widespread user experience with visualization and information analysis in general, very few participants had used the visual analytic systems available to them for this study. In fact, 90% of participants had never used their assigned tool, and of the remainder none were regular users. Not surprisingly, users assigned to the travel task were significantly more likely to identify themselves as experienced in the domain of their task ( $p < 0.01$  for Pearson's correlation test). However, participants' domain knowledge varied widely from novice to experienced for both tasks.

#### 3.3 Methodology

We asked each participant to perform one of the two analysis tasks. At the beginning of each study, participants were

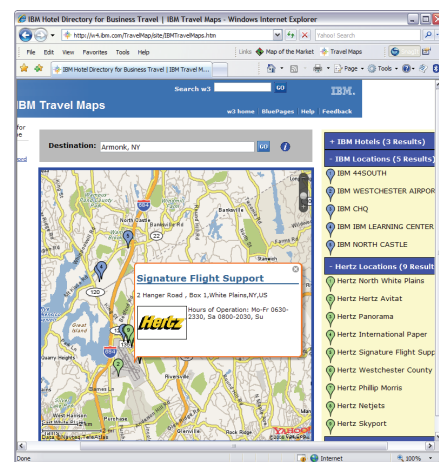


Figure 2. Participants assigned to the travel task were provided access to the Travel Maps tool from IBM, a web-based environment built using Yahoo! Maps.

given a pre-task questionnaire to gather general information about her/his experience and familiarity with the domain and visual analytic system for the task. We then gave a short tutorial on their assigned visual analytic system. Participants were then allotted 30 minutes to perform the task. Finally, each participant was asked to complete a post-task questionnaire which included a discussion with a study moderator.

In addition to the visual analysis environment, participants were provided tools for documenting their analysis process. Specifically, they were provided with pencil and paper for handwritten notes. In addition, they could use a screen capture tool to record visual evidence by taking snapshots of the visualization environment.

We audio-taped and video-taped each participant’s session for later analysis. Throughout each session, the moderator encouraged each person to “think out loud” as s/he performed the assigned task to gather additional insight into their analytical reasoning.

Furthermore, the moderator manually recorded a detailed log of every analytic step taken by the participant during the assigned task. The log, which we call an *action history*, provides an itemized list of all user analytic actions arranged linearly based on the order in which the actions were performed.

### 3.4 Metrics for Measuring User Performance

As part of our analysis of the study’s results, we quantitatively evaluated a participant’s performance in completing the assigned task. However, it is difficult to directly measure performance because visual analytic tasks are inherently non-procedural, open-ended activities in which human judgment is an essential component [26]. For this reason, there is typically no single “right” answer to be found, but rather a series of human judgments leading to a collection of relevant insights and conclusions.

We therefore evaluated user performance in terms of two sets of task requirements: (1) how many of the key questions of a given task were addressed by the participant, and (2) which of those insights were substantiated by collected evidence. Based on these quantitative features, we then measured a user’s performance using the PARADISE evaluation method [28]. Specifically, we calculated the Kappa coefficient ( $\kappa$ ) for each user, assigning a performance score in the range  $\kappa \in [0,1]$  where  $\kappa = 1$  represents the best possible performance. We calculated  $\kappa$  independently for each task group to normalize for differences among visualization environments and task requirements.

## 4. RESULTS AND ANALYSIS

Our study provides a number of valuable insights into how humans perform visual analysis. Most importantly, it helps us to identify the logical structures of a user’s visual discovery behavior. In this section, we present our analysis of the study and discuss the implications of our findings on user performance and system design. We present our analysis in two sections: *Patterns of Activity* and *Trails to Insight*. Each section corresponds to an important structure of user inter-

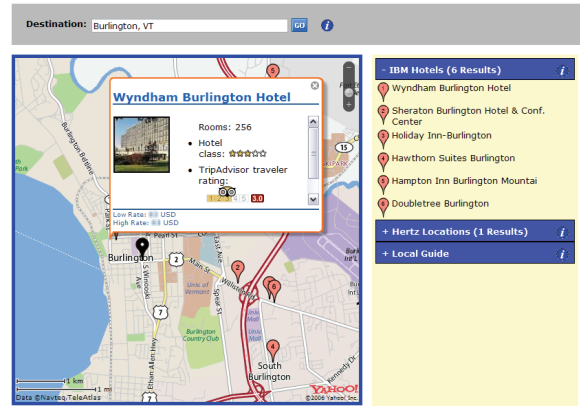


Figure 3. During the travel task, several users performed a sequence of Inspect actions, or a *scan pattern*, to view call-outs for individual hotels as shown above. The pattern was performed to enable visual comparison of non-geographic hotel attributes (e.g., room rate), a behavior not directly supported by the system.

action behavior identified in our study. Within each section, we present both our quantitative and qualitative analyses. Motivated by our findings, we conclude each section with a series of recommendations for the design of effective visual analytic systems.

### 4.1 Patterns of Activity

In our experiments, perhaps the most prominent feature we observed was that participants typically invent ad-hoc action sequences, or *patterns*. We define a pattern as a short sequence of three or more visual actions performed iteratively by a user to accomplish a specific low-level analytic goal, such as visual comparison. Patterns are performed repeatedly by users throughout their task to iteratively achieve similar analytic requirements.

For example, one user in the travel task performed the following pattern to iteratively bring up a visual callout (an *Inspect* action) for each of three hotels:

$$P_1 = \{Inspect(Hotel_1), Inspect(Hotel_2), Inspect(Hotel_k)\}$$

The above pattern was performed to accomplish the analytic goal of visual comparison. This same pattern, with different parameters (e.g., different hotels), was performed multiple times throughout the same user’s analysis. Figure 3 shows a screenshot of what the user saw as a result of one of the Inspect actions. We refer to patterns like  $P_1$  as *simple patterns* as they are formed by repeating a single action.

Users also performed *compound patterns*—iterative behaviors during which users repeat entire chains of actions. For example, a participant in our study performed the compound pattern  $P_2$  to gather scattered information about individual hotels. This pattern includes an *Inspect* action followed by two *Link* navigations to access relevant hotel and amenity information. The participant repeated  $P_2$  multiple times to perform a multi-dimensional comparison of several hotels in the same area.

$$P_2 = \{Inspect(Hotel_i), Link(Review_i), Link(Amenities_i)\}$$

As illustrated by both  $P_1$  and  $P_2$ , analysts invent and perform patterns to achieve specific analytic goals, such as information gathering or comparison. Moreover, we believe that such patterns are executed primarily to compensate for real or perceived limitations in the visualization environment. In particular, users may not know how to use the available tool to perform the desired analytic goal directly, or the tool is incapable of the desired functionality.

For example, consider pattern  $P_1$ , via which a user compared a set of hotel prices to determine which represented the best value. The user developed this pattern precisely because direct price comparison was not available as part of the provided map-based system. If this function were present, the user would not need to invent a new procedure with repeated actions to achieve the required comparison.

This interpretation of patterns corresponds well to the subjective comments provided by participants following the completion of their tasks. For example, one participant who performed several  $P_1$  patterns stated that “sorting on data” and “comparisons” were functions that they found missing in the map-based visual analytic system used for the travel task.

The action history logs recorded during our study show that over 96% of participants employed some form of repetitive action patterns within their analysis. According to our calculations, the observed rate is far greater than the 9.6% frequency we would expect to see had users performed their actions in a random order. This provides overwhelming evidence ( $p < 0.01$  for one-tailed Binomial test) that the repetitive patterns found in the participants’ histories are indeed a primary structure of visual analysis interaction behavior.

While the patterns performed by participants in our study varied widely, there were two patterns that were extremely prevalent (performed over 40% of our participants):

- **Scan Pattern:** a sequence of three or more *Inspect* actions examining similar objects. For example,  $P_1$  shows a scan pattern executed to compare hotels. Participants performed different variations of the scan pattern, each to compare multiple data objects of the

same type (e.g, restaurants or hotels). Scans were performed by 50% of the participants in our study.

- **Flip Pattern:** a sequence of three or more *Constrain* actions that alters the subset of information being visually analyzed. For example,  $P_3$  shows a flip pattern executed by several participants in the financial task to examine temporal trends within the market. Flips were performed by 43.3% of the participants in our study.

$$P_3 = \{Constrain(t = 52weeks), Constrain(t = 26weeks), Constrain(t = 52weeks), Constrain(t = YTD)\}$$

The pattern above ( $P_3$ ) contains an action sequence in which a user “flips” the time constraint back and forth from 52 weeks to other settings (e.g., 26 weeks and Year-To-Date). This pattern allowed for the comparison of a single stock’s performance at different time intervals, something not directly supported by the visual analytic system. As stated by one participant, “I’m just going back and forth to compare.” Figure 4 shows three screenshots captured during a flip pattern performed using the Map-of-the-Market tool.

In general, both the scan and flip patterns directly reflect significant limitations in the visual analytic systems used in our study. In both cases, multi-dimensional visual comparison is not supported along the dimensions required to complete the task. The scan pattern was significantly more prevalent in the travel task ( $p < 0.01$ ) because the map tool did not support side-by-side comparisons of hotels or restaurants. As a result, participants performed scan patterns over sets of objects to compare non-geographic attributes (e.g., prices or ratings).

Meanwhile, the flip pattern was significantly more prevalent in the financial task ( $p < 0.01$ ) because the tree-map does not directly support visual comparison along the time dimension. As a result, participants performed flip patterns by changing the time constraints back and forth across different values to compare stock performance over time.

**Pattern Evolution.** While the definition of patterns presented above stipulates a strict repetition of a static action



Figure 4. Participants regularly performed *flip patterns* while completing the financial task to visually compare information over several time intervals including (a) close of market, (b) 26 weeks, and (c) 52 weeks. During a flip pattern, users iteratively flip back and forth between constraint values to seek out information (e.g., temporal trends) not directly conveyed by the visualization. (Screenshots show the Map of the Market tool from SmartMoney.com.)



sequence, our analysis shows that a more fluid process is often at work. The ad hoc patterns developed to achieve a specific analytic goal frequently evolve over time as users improve their technique. We observed two types of pattern evolution: (1) expansion and (2) compression.

*Pattern expansion* occurs when, over time, a given pattern evolves into a similar but longer pattern. For example, one user who performed pattern  $P_1$  repeatedly during the early phase of his analysis discovered that he could gather additional information and make a more informed conclusion by taking a closer look at each hotel. This user expanded the simple pattern  $P_1$  into the compound pattern  $P_4$ . A similar expansion was performed by multiple participants.

$$P_4 = \{ \text{Inspect}(\text{Hotel}_i), \text{Link}(\text{Review}_i), \text{Link}(\text{Amenities}_i), \\ \text{Inspect}(\text{Hotel}_j), \text{Link}(\text{Review}_j), \text{Link}(\text{Amenities}_j), \\ \text{Inspect}(\text{Hotel}_k), \text{Link}(\text{Review}_k), \text{Link}(\text{Amenities}_k) \}$$

In contrast, *pattern compression* occurs when a given pattern evolves over time into a shorter pattern to achieve the same goal more efficiently. Pattern compression typically represents an improvement in technique gained through experience. For example, consider pattern  $P_5$ :

$$P_5 = \{ \text{Inspect}(\text{Company}_i), \text{CreateNotes}(\text{Company}_i), \\ \text{Link}(\text{News}_i), \text{CreateNotes}(\text{Company}_i) \}$$

A participant in our study performed the compound pattern  $P_5$  several times to examine a number of different companies. As the pattern shows, the participant initially reviewed a list of recent news articles for each company he examined.

Over time, however, the participant decided that it would be more efficient to bypass the news gathering step during their initial survey of individual companies. As a result, the participant compressed their initial pattern  $P_5$  into the more concise  $P_6$ , a shorter pattern performed multiple times later in their session.

$$P_6 = \{ \text{Inspect}(\text{Company}_i), \text{CreateNotes}(\text{Company}_i) \}$$

The evolution of patterns over the course of an analysis implies that users are progressively refining a set of ad hoc techniques, or templates, which they can re-use at a future time to satisfy similar analytic goals. This finding is consistent with observations in related studies which have found similar re-use at other scales [5, 13].

**Patterns and Insight.** Patterns can be useful structures for understanding how analysts go about their visual analysis tasks. We observed earlier in this section that patterns typically correspond to specific analytic goals, and that they map to actual or perceived limitations in a visual analysis environment. Moreover, our results indicate that analysts who perform patterns are more inclined to develop deeper and less obvious insights during their analysis.

Our study found that participants who performed patterns were significantly more likely to discover connections between pieces of information found at different points of

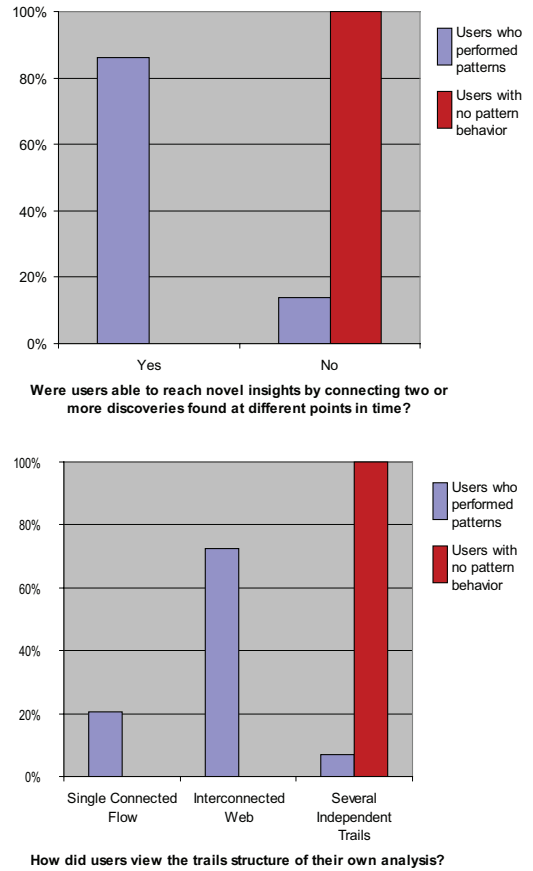


Figure 5. Users who performed patterns are significantly more likely to form joint conclusions and view their analysis as an interconnected web of trails. Both traits are signs of information synthesis activity.

time to form a joint conclusion ( $p < 0.05$ ). Similarly, when participants were asked to describe their own visual analytic workflow, those who performed patterns were significantly more likely view their work as “several interconnected lines of inquiry” rather than “several independent lines of inquiry” ( $p < 0.05$ ). Both results are shown in Figure 5.

These characteristics are extremely desirable in visual analytics because they are the hallmarks of information synthesis [26]. By combining individual facts together, an analyst performing information synthesis can move beyond information recall and actually create new information as a joint conclusion.

For example, one participant in the financial task identified (at different points in time) two companies with strong stock performance. The same participant then provided a much deeper insight by concluding that the performance of the first company was linked to the second company after reading news about a rumored merger. We observed that in this example, the participant used patterns which led him to perform a more systematic analysis. As a result, he was able to discover the important link between the stock performance of two of the companies identified in his analysis.

**Implications on Visual Analysis Design.** Our study shows that pattern behavior is ubiquitous throughout a typical

user’s visual analysis behavior, regardless of domain, expertise, or tool. Because patterns are developed ad hoc by analysts to meet important analytic goals, our characterization of patterns has several implications on visual analysis tool design. Below we provide two recommendations based upon our findings.

**RECOMMENDATION 1.** Patterns are a widespread behavioral structure (performed by over 96% of users in our study) that often reflect real or perceived limitations within a given visualization tool. The two most prevalent patterns found in our study (the scan pattern and the flip pattern) indicate a user requirement for flexible multi-dimensional visual comparison. Visual analytic systems should therefore be designed to maximize a user’s capabilities to perform visual comparison along all desired dimensions.

**RECOMMENDATION 2.** Users who exhibit pattern behavior are significantly more likely to form conclusions by combining two or more insights ( $p < 0.05$ ), and significantly less likely to treat their task as a series of independent inquiries ( $p < 0.05$ ). Both are desirable traits of information synthesis activity. However, because patterns reflect real or perceived limitations in functionality, patterns also indicate points at which users are in need of assistance. Therefore, to encourage deeper and more efficient analysis, visual analysis environments should be designed to: (1) recognize when users are initiating a known pattern; and (2) proactively assist the user in completing their desired analytic goal more efficiently by recommending the right tools at the right time. For example, a system capable of recognizing pattern  $P_1$  (performed to compare hotel ratings) could proactively suggest a chart-based visualization tool capable of directly supporting the intended visual comparison.

#### 4.2 Trails to Insight

*Trails* were the second prominent behavioral structure observed during our study. Trails are chains of user-initiated actions which lead to points of insight, and therefore embody the logical provenance of individual discoveries.

We defer our formal definition of trails until later in this section. We begin instead with an example. Consider the action sequence  $T_1$ , performed by a participant assigned to the travel task.

$$T_1 = \{Query_1, Query_2, Pan_3, Pan_4, Inspect_5, Inspect_6, Query_7, Inspect_8, Pan_9, Query_{10}, Filter_{11}, CreateNotes_{12}\}$$

Sequence  $T_1$  was performed while a user was beginning their analysis of travel facilities in Burlington, VT. The user started with a pair of queries (a search for Burlington; then a request to add hotel information to the map). She then panned the map twice and inspected a pair of individual hotels to quickly survey the available options. The user then requested that information about tourist attractions be added to the visualization via  $Query_7$ . After another *Inspect* to

examine a third hotel, the user re-centered the map with a *Pan* action. The user then queried for rental car locations ( $Query_{10}$ ) before quickly removing them in the subsequent step after deciding it wasn’t helpful. The sequence terminates with a *CreateNotes*<sub>12</sub> action representing the user’s creation of notes to record their insight that downtown Burlington is centrally located for both hotels and tourist attractions.

The sequence of actions in  $T_1$  represents only a small fraction of the user’s overall action history. Nonetheless,  $T_1$  is a small snippet of activity which captures the provenance of a single important insight: that downtown Burlington is the best location.  $T_1$  is therefore the logical path followed by the user to arrive at their conclusion.

As exemplified by  $T_1$ , a *trail* is a sequence of visual analysis activity consisting of two parts: (1) an initial *exploration phase* and (2) a terminal *insight action*. During the exploration phase, a user performs a potentially long sequence of actions to explore through the available data. The exploration phase includes actions such as queries, filters, and visual brushing of data subsets. A user executes these actions as they logically navigate the available information, as in steps one through eleven in  $T_1$ .

The exploration phase continues until the user develops a new insight, at which point they perform a final insight action (e.g., taking notes or bookmarking a visualization) to record the discovery. Every insight recorded by a participant in our study was observed having a similar trail structure: a series of exploration actions followed by an insight action. In our study, trails were in fact present in the activity history of 100% of our participants.

The exploration phase in a trail captures the logical reasoning that led to the associated insight. It may therefore be an important behavioral structure to help address one of the important challenges facing the visual analytics research community: the need to model, capture, and preserve the analytical reasoning behind an analyst’s conclusions [26]. Many previous attempts at capturing this reasoning (e.g. argument trees [15]) require direct manual construction, which may distract users from their main task at hand. In contrast, the trails identified in our study can potentially be automatically captured “behind the scenes” without any added burden to users.

**Interconnected Trails.** The example trail  $T_1$  can be understood independent of other trails. However, our observations show that many trails are interconnected. This observation reflects a progressive visual analysis workflow, in which insights from one stage of an analysis motivate future lines of inquiry.

For example, the participant who performed  $T_1$  later used the associated insight (i.e., the downtown region recorded via *Create*<sub>12</sub>) as the starting point when analyzing the available dining options as illustrated in trail  $T_2$ . We use the ‘ $\cap$ ’ operator to represent trail dependency.

$$T_2 = T_1 \cap \{Inspect_{13}, Inspect_{14}, Inspect_{15}, CreateNotes_{16}\}$$



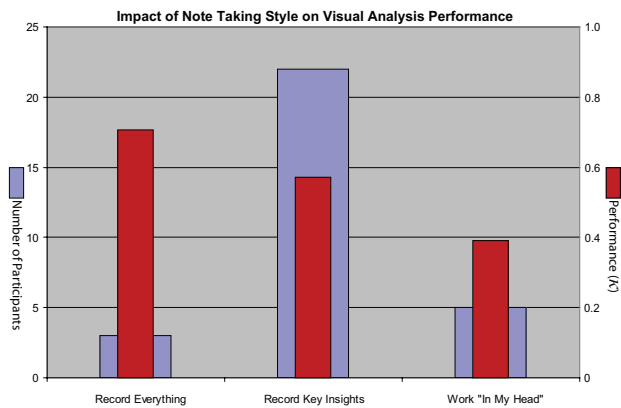


Figure 6. Users had various approaches to recording insights during their analysis. Those who documented more of their discoveries performed significantly better (based on the  $\kappa$  metric) than those who tried to mentally track their insights.

As shown above, trail  $T_2$  began by building upon the discovery of  $T_1$ : “downtown Burlington is the best location.” The user performed a scan pattern (actions 13-15) to visually compare a number of restaurants in the region defined by  $T_1$ . Then,  $T_2$  ends with the creation of an additional note to record the best available restaurant.

As this example illustrates, interconnections are built between trails as users progressively discover new insights. These dependencies were observable in the manually recorded action histories captured during the study. Moreover, 70% of our participants described their own visual analysis process as “several interconnected flows of analysis.” Furthermore, over 80% reported that they combined two or more separate discoveries to form a joint conclusion.

Perhaps most importantly, the presence of interconnected trails was highly correlated with the ability to form joint conclusions based on two or more discoveries found at different points in time ( $p < 0.05$ ). As mentioned earlier, this process of creating new information based on multiple discoveries is a hallmark of information synthesis, which is a critical visual analysis activity.

**The Value of Recording Insights.** Every participant in our study performed insight actions (e.g., taking notes and/or capturing visualization screenshots) to record their discoveries. These actions are an essential component of the structure of trails, serving as terminal markers for individual trails. However, the value of recording insights extends well beyond its use in defining trails.

As part of an end-of-study questionnaire, we asked each participant to identify which of the following three styles of analysis most closely described their approach to their assigned task:

- “I collected everything I found, and then selected the most important insights at the end of my task.”
- “I collected only the important insights as I found them.”

- “I worked mostly in my head and recorded the key insights in the end.”

The distribution of responses to this question and the associated performance of each group is shown in Figure 6. Those that said they worked mostly “in their head” without recording their insights performed significantly worse (as measured by the  $\kappa$  metric) on the task compared to the other groups ( $p < 0.05$ ).

We also found that visual analysis performance (the  $\kappa$  metric) was strongly correlated to the number of visualization screenshots recorded during a session ( $p < 0.01$ ). In addition, our results show a significant correlation between the number visualization screenshots and a participant’s ability to develop interconnected lines of inquiry during their analysis ( $p < 0.05$ ).

**Capturing Technique in Trails.** A key aspect of visual analysis observed during our study is the widespread re-use of analytic technique. The simple patterns described earlier in this paper model this re-use at the micro level, while compound patterns capture re-use at slightly greater scales. Similar repetition of techniques also occurs at even higher levels.

As part of our study design, both the travel and financial tasks contained two distinct but similar sub-tasks. For example, the travel task asked participants to perform the same analysis for two cities. Similarly, the financial task asked users to perform the same analysis for two sectors of the market.

In both tasks, there was a high degree of re-use of technique between the two sub-tasks. Over 90% of users reported using identical (43.3%) or similar techniques (50.0%) to address the two sub-tasks. Only 6.7% used entirely different techniques, and our results show that those participants were significantly less likely to perform information synthesis by connecting two or more insights into a joint conclusion ( $p < 0.05$ ).

Just as users invented ad hoc patterns to achieve transient analytic goals such as visual comparison, successful analysts also created longer and less rigid re-usable procedures for gathering specific types of insights. For example, one participant in the travel task performed the following procedure in both sub-tasks when choosing the best restaurants:

- $T_i$ : Look for corporate office locations in desired city.
- $T_{i+1}$ : Find quality hotels close by to office locations.
- $T_{i+2}$ : Find highly rated restaurants located near selected hotels.

The above procedure is represented within a user’s action history as a set of interconnected trails, one for each of the intermediate insights uncovered during the process. Each trail is composed of a sequence of actions. A similar set of trails was performed by this user for both sub-tasks.

**Implications on Visual Analysis Design.** Trails were created by every participant in our study as a by-product of their natural analytical behavior. These trails, which capture an analyst’s reasoning behind specific insights, hold poten-

tial as valuable artifacts within visual analysis environments. Motivated by our findings, below we provide a series of recommendations.

**RECOMMENDATION 3.** Trails, which correspond to the logical paths via which insight are discovered, were evident in the action histories for 100% of the participants in our study. Visual analytic systems should be designed to automatically record and preserve trails to serve as representations of insight provenance.

**RECOMMENDATION 4.** Re-use of analytic technique was widely present in our study (over 90% of participants) and correlated strongly ( $p < 0.05$ ) with improved information synthesis activity. Therefore, visual analytic systems should encourage re-use by capturing and exposing previous trails and their associated insights. Ideally, these trails can be parameterized and serve as starting points for future analysis. For example,  $T_i$  has a parameter “city” which a user would need to change to apply a trail developed for San Francisco, CA to the city of Burlington, VT.

**RECOMMENDATION 5.** Insight actions, such as visualization bookmarking, led to significantly improved task performance ( $p < 0.01$ ) and information synthesis ( $p < 0.05$ ). Visual analysis environments should be designed to encourage increased insight action activity

by providing easy-to-use note taking and visualization bookmarking capabilities.

## 5. CONCLUSION

A critical challenge in the emerging field of visual analytics is to develop visualization systems that can effectively support the human analytical reasoning process. In support of our own ongoing research to attain this goal, we conducted an empirical study specifically designed to examine structures of user behavior during visual analysis. Our findings and recommendations contribute to both our own future work as well as to the greater visual analytics community.

First, we reported our observations and analyses from an empirical study of users’ visual analysis behavior. Our study recorded behavior logs and interviews with 30 participants performing one of two open-ended visual analysis tasks. The results of our study suggest two prominent behavioral structures of visual analysis activity: patterns and trails.

Patterns are short ad hoc sequences of visual actions performed iteratively by users throughout their task. Patterns were performed by over 96% of the participants in our study, providing overwhelming evidence ( $p < 0.01$ ) that such behavior is a primary structural artifact in visual analysis activity. Based on our characterization of this behavior, we provided two concrete recommendations for exploiting pat-

Feature	Recommendation
Pattern	<b>RECOMMENDATION 1:</b> Patterns are a widespread behavioral structure (performed by over 96% of users in our study) that often reflect real or perceived limitations within a given visualization tool. The two most prevalent patterns found in our study (the scan pattern and the flip pattern) indicate a user requirement for flexible multi-dimensional visual comparison. Visual analytic systems should therefore be designed to maximize a user’s capabilities to perform visual comparison along all desired dimensions.
	<b>RECOMMENDATION 2:</b> Users who exhibit pattern behavior are significantly more likely to form conclusions by combining two or more insights ( $p < 0.05$ ), and significantly less likely to treat their task as a series of independent inquiries ( $p < 0.05$ ). Both are desirable traits of information synthesis activity. However, because patterns reflect real or perceived limitations in functionality, patterns also indicate points at which users are in need of assistance. Therefore, to encourage deeper and more efficient analysis, visual analysis environments should be designed to: (1) recognize when users are initiating a known pattern; and (2) proactively assist the user in completing their desired analytic goal more efficiently by recommending the right tools at the right time. For example, a system capable of recognizing pattern $P_1$ (performed to compare hotel ratings) could proactively suggest a chart-based visualization tool capable of directly supporting the intended visual comparison.
Trail	<b>RECOMMENDATION 3:</b> Trails, which correspond to the logical paths via which insight are discovered, were evident in the action histories for 100% of the participants in our study. Visual analytic systems should be designed to automatically record and preserve trails to serve as representations of insight provenance.
	<b>RECOMMENDATION 4:</b> Re-use of analytic technique was widely present in our study (over 90% of participants) and correlated strongly ( $p < 0.05$ ) with improved information synthesis activity. Therefore, visual analytic systems should encourage re-use by capturing and exposing previous trails and their associated insights. Ideally, these trails can be parameterized and serve as starting points for future analysis. For example, $T_i$ has a parameter “city” which a user would need to change to apply a trail developed for San Francisco, CA to the city of Burlington, VT.
	<b>RECOMMENDATION 5:</b> Insight actions, such as visualization bookmarking, led to significantly improved task performance ( $p < 0.01$ ) and information synthesis ( $p < 0.05$ ). Visual analysis environments should be designed to encourage increased insight action activity by providing easy-to-use note taking and visualization bookmarking capabilities.

Table 1. We provide five recommendations for visual analysis tool design based on our observations and analysis.

terns in the design of future visual analysis environments. Our recommendations suggest that patterns can be used to determine user-required visualization tool functionality. Furthermore, patterns can be recognized by smart visual analytic systems to proactively assist users in completing their desired analytic goals.

Trails are chains of user visual interaction activity which lead to points of insight. Trails were ubiquitous throughout our study and performed by 100% of the participants. Moreover, the trail structure for most users was interconnected, representing the progressive nature of insight discovery. We provided three recommendations for exploiting trails in designing future visual analytic systems. These recommendations encourage the use of trails for insight provenance and the development of different tools for supporting insight-based activity (e.g., visualization bookmarking).

Motivated by the results of our empirical study, our primary direction of future research consists of incorporating the recommendations made here into our ongoing work. This includes the automatic detection of patterns and trails as representations of user context for both visualization recommendation and insight provenance. Once realized, we plan to conduct additional user studies to empirically measure the value to end-users of automated detection of patterns and trails.

## REFERENCE

- [1] S. Card, J. Mackinlay, and B. Shneiderman, editors. *Readings in Information Visualization*. Morgan Kaufmann, 1999.
- [2] C. Chen and M. Czerwinski, editors. *International Journal of Human Computer Studies, Special Issue on Empirical Evaluation of Information Visualizations*, volume 53. 2000.
- [3] C. Chen and Y. Yu. Empirical studies of information visualization: a meta-analysis. *Int J. on Human Computer Studies*, 53(5):851–866, 2000.
- [4] E. H. Chi, L. Hong, J. Heiser, S. K. Card, and M. Gumbrecht. Scintindex and scenthlights: Productive reading techniques for conceptually reorganizing subject indexes and highlighting passages. *Information Visualization*, 6, 2007.
- [5] P. Cowley, L. Nowell, and J. Scholtz. Glass box: An instrumented infrastructure for supporting human interaction with information. In *Proc. of Hawaii International Conference on System Sciences*, 2005.
- [6] D. Gotz. The scratchpad: Sensemaking support for the web. In *Proc. of the Inter. WWW Conf. Posters*, 2007.
- [7] D. Gotz, M. X. Zhou, and V. Aggarwal. Interactive visual synthesis of analytic knowledge. In *IEEE VAST*, Nov 2006.
- [8] D. Holten. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. In *Proc. of InfoVis*, 2006.
- [9] H. Kang, C. Plaisant, B. Lee, and B. B. Bederson. Netlens: Iterative exploration of content-actor network data. *Information Visualization*, 6, 2007.
- [10] A. Kobsa. An empirical comparison of three commercial information visualization systems. In *Proc. of InfoVis*, 2001.
- [11] A. Kobsa. User experiments with tree visualization systems. In *Proc. of InfoVis*, 2004.
- [12] J. Lamping and R. Rao. Visualizing large trees using the hyperbolic browser. In *Proc. of ACM CHI*, pages 388–389, New York, NY, USA, 1996. ACM Press.
- [13] I. Mani and G. L. Klein. Evaluating intelligence analysis arguments in open-ended situations. In *Proc. of Inter. Conf. on Intel. Analysis*, 2005.
- [14] E. Morse, M. P. Steves, and J. Scholtz. Metrics and methodologies for evaluating technologies for intelligence analysts. In *Proceedings of the International Conference on Intelligence Analysis*, 2005.
- [15] N. J. Pioch and J. O. Everett. Polestar - collaborative knowledge management and sensemaking tools for intelligence analysts. In *Proceedings of CIKM*, 2006.
- [16] P. Pirolli and S. Card. Information foraging. *Psychological Review*, 106:643–675, 1999.
- [17] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proc. of the Inter. Conf. on Intelligence Analysis*, 2005.
- [18] C. Plaisant. The challenge of information visualization evaluation. In *Proc. of Advanced Visual Interfaces*, 2004.
- [19] C. Plaisant, J. Grosjean, and B. B. Bederson. Spacetree: Supporting exploration in large node link tree, design evolution and empirical evaluation. In *Proc. of InfoVis*, 2002.
- [20] Y. Qu and G. W. Furnas. Sources of structure in sensemaking. In *Proc. of ACM CHI Short Papers*, 2005.
- [21] J. Redish. Expanding usability testing to evaluate complex systems. *J. of Usability Studies*, 2:102–111, 2007.
- [22] D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card. The cost structure of sensemaking. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1993.
- [23] P. Saraiya, C. North, and K. Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Trans. on Visualization and Computer Graphics*, 11, 2005.
- [24] J. Scholtz, E. Morse, and T. Hewett. An analysis of qualitative and quantitative data from professional intelligence analysts. In *Proceedings of the International Conference on Intelligence Analysis*, 2005.
- [25] B. Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics*, 11:92–99, 1992.
- [26] J. J. Thomas and K. A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Press, 2005.
- [27] F. van Ham and J. J. van Wijk. Beamtrees: Compact visualization of large hierarchies. In *Proc. of InfoVis*, 2002.
- [28] M. A. Walker, D. J. Litman, C. A. Kamm, and A. Abella. Paradise: A framework for evaluating spoken dialogue agents. In *Proc. of 35th Annual Meeting of the ACL*, 1997.
- [29] C. Weaver, D. Fyfe, A. Robinson, D. Holdsworth, D. Peuquet, and A. M. MacEachren. Visual analysis of historic hotel visits. *Information Visualization*, 6, 2007.
- [30] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort. The sandbox for analysis - concepts and methods. In *Proc. of CHI 2006*, 2006.