Visualizing Dimension Coverage to Support Exploratory Analysis

Ali Sarvghad, Melanie Tory, and Narges Mahyar

Abstract—Data analysis involves constantly formulating and testing new hypotheses and questions about data. When dealing with a new dataset, especially one with many dimensions, it can be cumbersome for the analyst to clearly remember which aspects of the data have been investigated (i.e., visually examined for patterns, trends, outliers etc.) and which combinations have not. Yet this information is critical to help the analyst formulate new questions that they have not already answered. We observe that for tabular data, questions are typically comprised of varying combinations of data dimensions (e.g., what are the trends of *Sales* and *Profit* for different *Regions*?). We propose representing analysis history from the angle of *dimension coverage* (i.e., which data dimensions have been investigated and in which combinations). We use scented widgets [30] to incorporate dimension coverage of the analysts' past work into interaction widgets of a visualization tool. We demonstrate how this approach can assist analysts with the question formation process. Our approach extends the concept of scented widgets to reveal aspects of one's own analysis history, and offers a different perspective on one's past work than typical visualization history tools. Results of our empirical study showed that participants with access to embedded dimension coverage information relied on this information when formulating questions, asked more questions about the data, generated more top-level findings, and showed greater breadth of their analysis without sacrificing depth.

Index Terms—Dimension coverage, Tabular data, History, Empirical laboratory study, Exploratory data analysis, Scented widgets

1 Introduction

Dimension coverage information captures which dimensions in a tabular dataset have been explored so far, and in which combinations. We explore how revealing dimension coverage can facilitate data analysis. We illustrate how this information can be embedded within the interface widgets of a visual data analysis tool. We then present results of a user study demonstrating that dimension coverage information can help analysts to formulate questions.

During data analysis, an analyst constantly formulates and evaluates new questions or hypotheses about data. However, selecting a data subset to explore can be quite difficult [4, 18]. In fact, Lam identified *deciding-what-to-explore-next* as one of three key data analysis challenges [19] in exploratory data analysis (EDA). With current tools, analysts typically rely on memory to recall what questions they have asked and what they still need to do. However, factors such as limited short term memory and the recency effect (i.e. remembering recent items more clearly than those further in the past) [13] can impede recall. In other words, it can be difficult to maintain an awareness of dimension coverage. In addition, unfamiliarity with the shape and structure of the data [4, 32], vague analysis goals [32], and insufficient domain or visualization knowledge [32] can hinder question formation, potentially leading to under-exploration of the problem. Moreover, Wongsuphasawat et al. [32] argue that typical interfaces for constructing visualizations (i.e. manually mapping and filtering dimensions, making data transformations, and selecting visual encodings) may encourage premature fixation on specific questions, promoting depth-first exploration at the expense of breadth. How then, can we help analysts to formulate questions and encourage them to go both broad and deep?

In the case of tabular data, we observe that an analytic question is comprised of a combination of data dimensions and can be reasonably characterized by that dimension set. For example, a business analyst might start by asking "what is the relationship between *Profit* and *State*?" and next she may filter state to California and examine, "what is the relationship between *Profit*, *State:California* and *Products*?" Throughout her analysis, she constantly formulates and evaluates questions, each containing a combination of dimensions. This suggests that revealing dimension coverage information might help analysts recall what questions they have asked, and (perhaps more importantly), identify questions that have *not yet* been asked. Hence, in this research, we propose and evaluate the idea of visualizing dimension coverage to support data analysis.

To implement this idea, we extend scented widgets [30], a technique for embedding navigational cues into GUI widgets (see Figure 1). Specifically, we incorporate dimension coverage information directly into the interface elements of an exploratory analysis prototype (Figure 2). The theoretical support behind this method stems from the notion of information scent (i.e. attention pointers that assist a person in navigating an information space) introduced by Pirolli and Card [24]. This approach enables analysts to maintain an up-to-the-moment understanding of what data dimensions they have investigated and in what combinations.

Dimension coverage information is captured at the system level by visualization history modules that track and record visualization states (e.g., [15]). However, most visual representations of history provide very limited support for understanding dimension coverage because they focus instead on representing past states and/or actions. Previously, we introduced an alternative visualization of history to explicitly reveal dimension coverage information [27]. We demonstrated that this perspective could improve asynchronous collaboration, where one analyst does some work and then "hands-off" the work to a collaborator who continues the analysis. Providing analysts with information about which dimensions were previously investigated by their colleague reduced the duplication of work. This finding suggested that dimension coverage information might facilitate the flow of analysis in non-collaborative situations as well. Therefore, in the current paper, we investigate the effects of live dimension coverage information on individual data analysis. We also explore how this information can be integrated within a visual data analysis tool in a subtle way; our previous stand-alone representations were space inefficient and were not integrated with an analysis tool. We hypothesized that visualizing the coverage of dimension space would:

H1: increase the number of formulated questions,

H2: increase the number of findings, and

H3: increase the breadth of exploration without sacrificing depth.

- *Ali Sarvghad is with University of Victoria. E-mail: asarv@uvic.ca.*
- *Melanie Tory is with Tableau Research. E-mail: mtory@tableau.com.*
- *Narges Mahyar is with University of British Columbia. E-mail: nmahyar@ubc.ca.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx/

To evaluate our hypotheses, we built a prototype visual analysis tool (Figure 2) that used scented widgets to visualize dimension coverage. We then conducted a between-subjects user study to compare our tool to a baseline version without the dimension coverage information.

Contributions: In this research, we explore how dimension coverage information can support data analysis. We use an extension of scented widgets to incorporate this information into the interaction widgets of a prototype tool. Findings of our user study demonstrate that participants with access to real-time dimension coverage information formulate more questions, investigate a greater number of dimensions, and identify more top-level findings (see section 5.3 for definition) than participants without access to this information.

2 Related Work

Our work examines the value of representing dimension coverage information via scented widgets. Since dimension coverage can be thought of as one perspective on a visualization history, we first review prior research on history mechanisms for data analysis. Next, we present prior work on scented widgets and the types of past analysis information they have been used to reveal.

2.1 History Tools for Visual Data Analysis

History tools (e.g.,[8, 15, 17, 21]) track a person's past work as they analyze data with a visualization tool. They support visual data analysis workflow by enabling users to review, retrieve, and revisit past visualization states [15]. At the system level, visual history modules usually implement state-based or action-based architectures. In action-based models, individual or groups of user interactions are captured; these interactions typically result in a transformation of the system and/or visualization. In contrast, state-based history tools record information about the state of the system and/or visualization at specific times; these records can be used to replicate that system state at a later time. State-based history tools may also include analyst externalizations such as notes and annotations.

Visual representations of history usually correspond to the underlying architectural model. For action-based histories, the most common visual representation is a node-link graph, where nodes represent actions and connections show dependencies or precedence. This visual encoding can effectively depict dependencies and the flow of actions. GRASPARC [7], ExPlates [16], GraphTrails [11], VisTrails [6] and CzSaw [17] are a few examples of history tools that employ a node-link graph approach. One exception to this trend is SensePath [23], a provenance tool that represents its action-based history using a list of icons and textual descriptions. In contrast, state-based histories are typically visualized using a linear comic-strip-like list of captured states. Example analysis tools with a linear list history of visualization states are Heer et al.'s history for Tableau [15], Zhao et al.'s PivotSlice [33], and Mahyar et al.'s CoSpaces [21].

Recent surveys of visualization history tools [25, 26] identify recalling past actions and visualization states as one of the most common uses of visualization histories. Action and state-based history models work well for this use case [15], but are much less effective at supporting an understanding of dimension coverage and a high level picture of the analysis done so far [27]. State-based history models intrinsically contain information about dimension coverage but their typical linear representations do not make this information very accessible. To understand what has been done, what has been left out, and what combinations of dimensions have been investigated, users need to sequentially review the list and rely on memory or notes to keep track of dimension coverage. This process is time consuming and prone to error. PivotSlice [33] made this information slightly more apparent by providing a list of recently examined attributes. However, it only showed the most recently used ones, and the main purpose was to support reuse of filters rather than to reveal dimension coverage.

In a recent study comparing linear history with a view specifically designed to reveal dimension coverage [27], we found that participants with access to dimension coverage information were faster and more accurate at answering questions about dimensions used in the analysis. Our experimental results also demonstrated that analysts with access to dimension coverage information were more likely to investigate aspects of the problem that were left uninvestigated in previous work by a collaborator. However, the dimension coverage designs in this earlier work were built as standalone history tools independent of the data analysis tool and could therefore only be used post-analysis. In addition, the focus of our prior investigation was the effect on asynchronous collaboration. In the current paper, we expand on our prior work by investigating the value of dimension coverage information for ongoing work by a single analyst.

2.2 Scented Widgets for Analysis History

Our work integrates dimension coverage information directly into a visual data analysis tool by using scented widgets. There is a rich body of research that investigates embedding different kinds of information within GUI elements. Phosphor [5] superimposed a halo effect on recently used interface widgets to assist users in noticing changes that had taken place in the interface. Derthick [10] and Eick [12] introduced modified versions of slider controls that visually embedded information in the widget. Depending on the design, this information could be related to the data values in the dimension that the slider is bound to or values of a different dimension. For example, a slider that is bound to City (i.e. that allows users to pick a city name) could contain an embedded visualization showing the average number of frost-free days for each city [12].

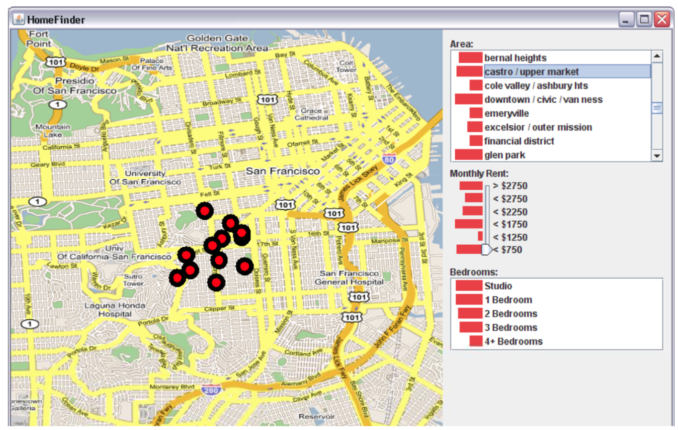


Fig. 1: Re-implemented HomeFinder with scented widgets that provide social navigation cues [30]. Bar charts embedded in the controls reveal information about the frequency of other people's investigation of data values. For example, few people looked at 4+ Bedrooms.

In [30], Willett et al. introduced guidelines for designing scented widgets. As one example, they used scented widgets to provide social navigation cues. They re-implemented HomeFinder [31], a map-based housing search tool. Information about prior house searches were embedded in the dynamic query widgets to help people better understand which data values had been investigated by other users (Figure 1). This work has three main differences from our research: type of exploratory task, visualized information and context. In their case, the exploration only required filtering of data values. Each question about data included a fixed set of dimensions (Area, Rent, Number of Bedrooms etc.) and users only manipulated filtering. On the other hand, we consider exploratory analysis tasks that require investigating varying combinations of dimensions. For example, a business analyst might explore many performance indicators (e.g., Profit, Sales, Return on Investment etc.) in relation to other attributes. To support this type of analysis, we reveal *dimension* coverage information (embedded in dimension name widgets) rather than *data value* coverage (embedded in value name widgets), and extend scented widgets to capture *co-investigation* information (i.e. which dimensions were considered in combination). Finally, in our case, the context of analysis is single professional users as compared to online collaborative social data analysis. Most often, the main goal of the latter is to enjoy while the former

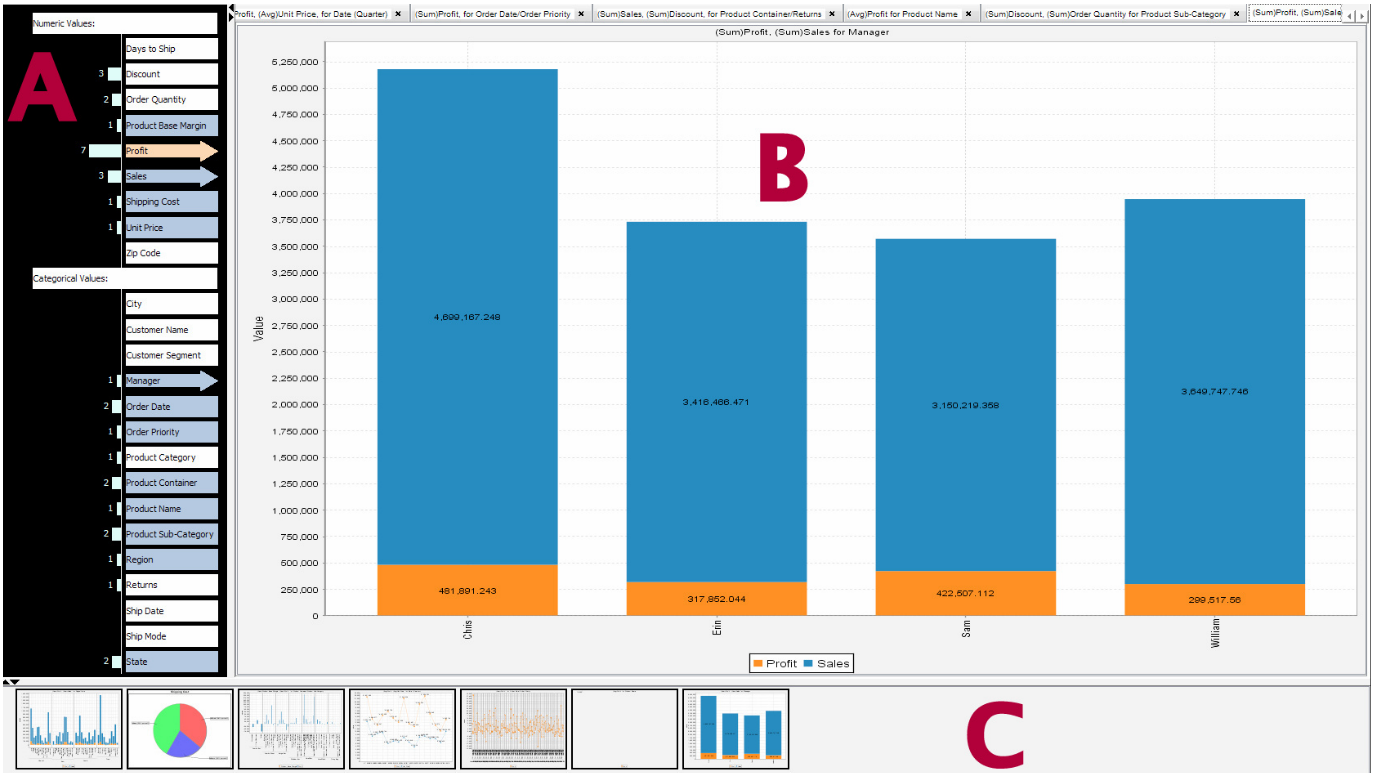


Fig. 2: Visual analysis prototype that reveals dimension coverage information. (A) Scented View reveals the dimension coverage information using scented widgets. Bar charts to the left of each dimension name indicate how frequently each has been investigated. Co-investigated dimensions are revealed through colouring (blue) when one or more target dimensions are selected (orange). Arrows indicate dimensions included in the currently displayed chart. (B) Visualization panel. (C) Sequence View shows a chronologically-ordered list of created charts.

is to discover and present [22]. More importantly, our research investigates the value and effect of seeing traces of one's own past analysis, rather than trails of investigation left by other unknown people.

3 Incorporating Dimension Coverage Information into Visual History

In order to investigate how dimension coverage information would influence exploratory analysis, we designed and implemented a history module that provided three distinctive representations of analysis history. The most important of these was *Scented View*, which used scented widgets to reveal dimension coverage information. We also included two complementary history views: a data values coverage view (*Data View*) and a traditional linear list of past states (*Sequence View*). These views were embedded within a prototype tool for visual data analysis (Figure 2).

3.1 Scented View

Scented View reveals dimension coverage information using scented widgets (Figure 2A). Embedding information that is logically relevant to a GUI element (such as a textbox showing a dimension's name) makes information easy to discover and readily available. This was our main rationale behind using a scented widget approach.

We designed Scented View to support two primary tasks: 1) understanding which dimensions have been investigated versus which have been left out, and 2) understanding which combinations of dimensions have been examined (i.e. co-investigation relationships). We also aimed to support the secondary task of understanding frequencies of use. For instance, Figure 2A shows a greater focus on Profit (investigated in 7 charts) than Sales (investigated in 3 charts). From a visualization design perspective, the first primary task required distinguishing between investigated and uninvestigated dimensions (a categorical concept). Since it is not advisable to alter shape and/or spatial position of interface elements (a rectangular textbox cannot suddenly become triangular or move to some other part of the GUI), these two channels could not be used to encode the information. Therefore, similar to [30], for every investigated dimension, we placed a bar to the left of the textbox that contained the dimension's name. The length of the bar encoded the magnitude of investigation; this used position encoding, the most powerful visual encoding channel [9, 22], to encode the magnitude of investigation.

The second task required discovery of relationships between dimensions (i.e. dimensions investigated together within the same chart). Similar to the first task, and because of the limitations imposed by working with GUI elements, containment, grouping and proximity could not be used for encoding relationships. We considered drawing lines between the labels to visualize connections (e.g. lines traveling from sales to Profit, State and City to show co-investigation) but this design would add clutter and visual obstruction to the dimension panel. Therefore we opted to use colour hue (blue and orange) to encode dependencies and rely on interaction to reveal this information on demand. Figure 2A shows how this information was conveyed. When the user selected a dimension (by clicking on the name label), the background of the dimension textbox changed to orange and the background of any other co-investigated dimension(s) became blue. Furthermore, a user could select more than one dimension from the list to investigate higher-order relationships. While this visual encoding choice needed to be learned, it had several advantages: 1) the selected item(s) were clearly indicated by orange colour, making it easy to add to or change the selection, 2) selected items (orange) and their co-investigated items (blue) clearly stood out from the list, and 3) the encoding only minimally changed the standard appearance of interface widgets. This interaction approach and colour scheme were also easily learned and understood by participants in our earlier user study (albeit within a very different visual encoding) [27].

Dimension coverage information for individual dimensions was constantly present in the user interface. Co-investigation information,

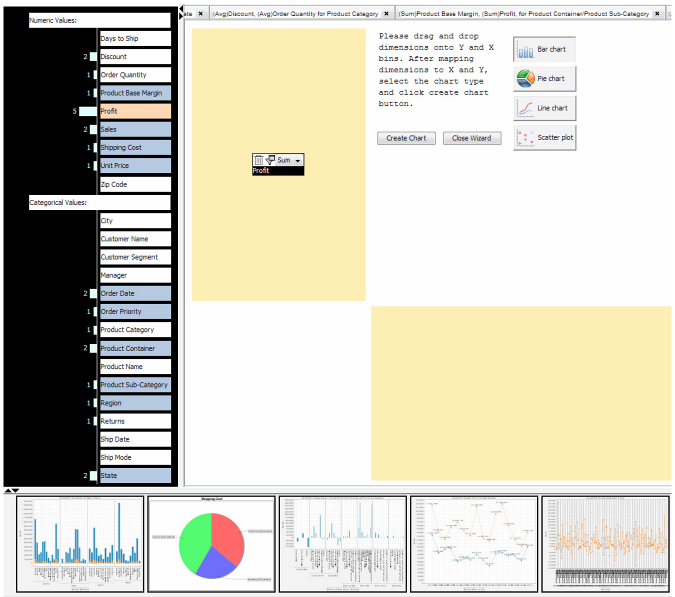


Fig. 3: Automatic presentation of co-investigation information at the time of creating a new chart. As the user drags and drops dimensions into X or Y bins, the view changes to show prior co-investigation of the dimensions. In this example, as soon as the user dropped Profit in the Y bin, the colours changed to highlight dimensions that were previously co-investigated with Profit.

however, was not continuously present; it became available in two different ways. One way to attain this information was by selecting one or more dimensions (clicking names in the list). This approach required the user's explicit interaction with the view. For example, in Figure 2A, the user has selected Profit; all dimensions ever considered with Profit have changed to blue. A user could also select a combination of dimensions (e.g. Profit+Sales+City) to discover their co-investigation with the remaining dimensions. In the second approach, the system automatically represented co-investigation information while the user actively created a new chart. For example, in Figure 3, the user has just mapped Profit to the Y axis; at this point, ten other dimension widgets turn blue to remind the user about which other dimensions have been previously co-investigated with Profit. The view keeps updating as the user maps more dimensions on X or Y.

The rationale behind this design was to support two critical use cases involving co-investigation information. In the first use case, the user intentionally paused analysis to review and recall co-investigation information. In the second use case, automatic presentation of this information could help the user avoid duplicating earlier work and assist them to formulate novel questions on the fly. Selecting dimensions by either approach (direct selection via mouse click or by the system during chart creation) also filtered the content in Sequence View (discussed below). Filters were removed after the user deselected dimension(s) or after a new chart was created.

3.2 Sequence View

Sequence View showed past visualization states in a linear list format (Figure 2C). This view mirrored the typical linear list approach to visualization history (e.g., [15, 21, 33]). In this approach, thumbnail images of past charts are ordered chronologically, labeled with information such as chart name. Our aim with Sequence View was to help users to quickly review and reuse past states. We included this view as a representative of typical history designs, because we felt it was complementary to our new views, and to enable a comparison to an appropriate baseline in our study. Thumbnail image size was proportional to the height of the Sequence View panel. If desired, a user could maximize the Sequence View panel to the entire width and height of the window to browse and compare large images side by side. Hovering the mouse over a thumbnail image showed a tooltip with detail information about the chart (e.g., (SUM)Sales, (SUM)Profit for Region, City, Product Category).

3.3 Data View

Data View represented the prior coverage of data values for each dimension (Figure 4). This information was only available on demand by hovering the mouse over a dimension's label on the list for 2 seconds. This design decision was based on the findings of our previous study [27], in which data values coverage information was constantly present in the user interface but was not used much by participants (for a similar analysis task and data set). This result suggested that data values coverage could be available on-demand.

Hovering the mouse popped open a tooltip-like pane that contained a visualization of the data value coverage for that dimension. The visual encoding depended on the dimension type (quantitative versus ordinal/nominal). We used a tag cloud for ordinal and nominal values where font size encoded the frequency of prior investigation. Figure 4 (bottom) shows that for Product Categories, the analyst had focused more on "Binders and Binder Accessories" than on "Appliances". For quantitative dimensions (e.g. Sales), we used a bar to encode values within the dimension and colour saturation to convey information about the magnitude of investigation. Darker shades indicated data values that were included in more charts (i.e. other values were filtered out). For example, in Figure 4 (top), all values of the Profit dimension are represented in the bar, ordered smallest to largest from left to right. In this specific example, the analyst focused most on values ranging from approximately 43 to 441.

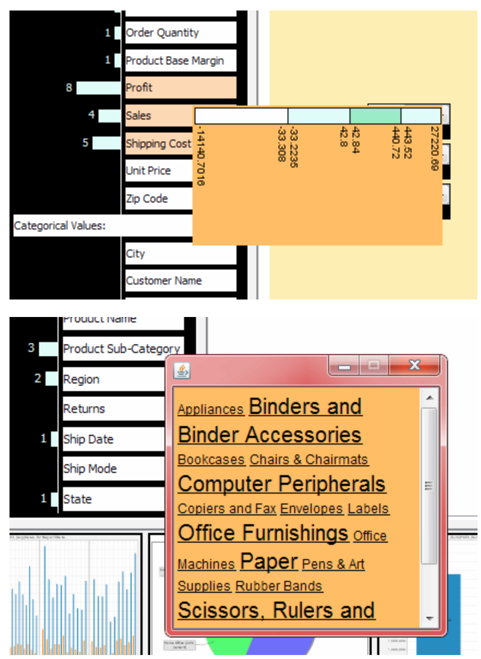


Fig. 4: Two examples of Data View, showing pop-ups after hovering the mouse on Profit (top) and Product Sub-Category (bottom). In the top, the range of [42.84, 440.72] has been investigated more than the other values. In the bottom, some Product Sub-Categories have been investigated more than others.

3.4 Prototype Implementation

We implemented our three history views within a visual data analysis prototype that enabled a user to create a variety of basic statistical charts from tabular data. The key feature of our prototype was the use of scented widgets for live presentation of dimension coverage within interface elements. The prototype was built in Java and used the JFreechart [1] API for chart creation. While the available chart types were not elaborate, and the tool's functionality was less comprehensive than commercial tools, we performed two pilot user studies (with identical procedure to the evaluation study described later) to ensure that the prototype provided a sufficient analysis environment in which to investigate the idea of embedded dimension coverage information.

Figure 2 shows a screen shot of the prototype. After connecting to the data source, a list of data dimensions was created and placed in Scented View (Figure 2A). Dimensions were divided into categorical and numeric. Dimension names were ordered alphabetically. Our chart creation design was loosely based on the shelf approach in Polaris [28] and Tableau [2]: the user could select dimensions from the list and drag and drop them into vertical and horizontal shelves in the chart pane (Figure 3) that respectively represented X and Y axes. After a dimension was mapped to an axis, if required, the user could filter values and apply simple statistical operations such as sum, average and standard deviation. Next, the user would select a chart type amongst the available options: bar, stacked bar, line, pie or scatter. To enable analysts to create complicated charts, they could map multiple dimensions on each axis. Each new chart was placed in a new tab in the charts pane (Figure 2B). A user could switch between the charts by clicking on the respective tab at the top of the panel. Each tab contained the title of the chart (e.g. [Avg.] Profit, [Avg.] Sales, [Avg.] Shipping Cost, City) to help identify it. Selecting a tab caused the shape of the dimensions that were involved in the corresponding chart to change to arrows (Figure 2A). The rationale behind this design was to visually assist a user to quickly identify the dimensions plotted in the currently selected chart.

4 Evaluation - Method

We conducted a controlled between subjects laboratory experiment to evaluate how access to dimension coverage information would influence the analysis process and its outcomes. Specifically, we tested the three hypotheses described in the introduction, namely that dimension coverage information would cause participants to ask more questions (H1), produce more findings (H2), and increase the breadth of their analysis without sacrificing depth (H3).

4.1 Experiment Design

We compared the full prototype described in section 3 to a baseline version that was identical in design and functionality except that dimension and data value coverage information was removed . Specifically, in the baseline version 1) there were no bars next to investigated dimensions' labels to show the frequency of investigation, 2) interaction with the list of dimensions provided no insight into the co-investigation of dimensions through colour-coding, and 3) Data View was removed. Similar to the full version, baseline version users could filter Sequence View by selecting dimension(s) from the dimension list. The background of the selected dimension(s) became orange but the co-investigated dimensions (if any) remained unchanged. Baseline version users could review history sequentially (by looking at visualizations one by one) or selectively (by filtering to show only visualizations with certain dimensions). We chose this experimental design, comparing the identical tool with features enabled versus removed, so that we could conclude that any difference in number of questions, number of findings, and breadth of analysis (dependent variables) were caused by the additional dimension coverage information (independent variable).

4.2 Participants

We recruited 20 business students (12 graduate, 8 senior undergraduate, 4 male, 16 female, average age of 25) through online advertisement and posters across campus. We selected business students to ensure that participants had the necessary domain knowledge to investigate a finance-related problem. We only recruited participants who reported having an understanding of business data analysis and experience with creating different types of statistical charts such as bar, line and scatter plots. To minimize the effects of gender on the outcomes (considering the much larger population of female participants), we randomly assigned eight female participants and two male participants to each condition (full or baseline). All participants signed a consent form approved by the University of Victoria's human research ethics office.

4.3 Procedure

We began with an introduction (approximately 15 minutes) to the task, data and tool (either full or baseline version). After the introduction, participants practiced (approximately 30 minutes) using the tools by doing short warm up tasks with an example sales dataset different from the dataset used for the actual task. The practice task required working with all the main features of the system. In particular, participants practiced how to create charts by dragging and dropping dimensions onto X & Y shelves, how to filter data and perform statistical operations, how to use Scented and Data views to obtain coverage information (full version only), and how to review and reuse work history using Sequence View. An experimenter was present during the practice session and participants could ask questions about the software, task and history file. A list of the supported user interactions and their outcomes was left with the participant to be referred to (if needed) during the actual analysis task. After the introductory part, participants were left alone and given one hour to perform the main task. Pen and paper were provided in case participants wanted to take notes. Participants were told that there were no constraints on what they could record in their notes. The analysis session was followed by a short interview and a questionnaire.

4.4 Task

The open-ended exploratory analysis task required participants to evaluate the business performance of an online retailer using a sales dataset and identify any positive or negative performance indicators. The task was based on typical SWOT (Strength, Weakness, Opportunity and Threats) analysis tasks that help large organizations to evaluate their business venture. All of the participants reported that they were familiar with this type of analysis, as it is taught in business and commerce programs. Following are the instructions given to participants: "You are a business data analyst in a large online retailer. You should explore the performance data for the past 4 years and identify trends/outliers in the data that are indicative of strong and/or poor performance." We used the Superstore sales dataset provided by Tableau Public. This dataset contains sales information for four years and has 24 data dimensions and 8400 records. No specific directions or restrictions were imposed to influence or direct the participant's focus. Participants were given 60 minutes for the task (exclusive of introduction and practice time) and all participants fully used this time (differing by only a few minutes).

4.5 Data Capture

Participants were asked to think aloud. We recorded video and audio of all the sessions and interviews. The video camera was pointed at the screen to capture the screen contents as well as user actions that could not be logged by the system (e.g., tracing the dimension list with one's finger). Each participant's analysis work (i.e., charts created) were recorded by the system. Both base and full versions of the prototype automatically logged user interactions with the tool (e.g., selecting a dimension or reloading a chart). In addition, full version users gave their assessment of the scented widgets by answering a Likert style questionnaire.

5 Evaluation - Data Analysis and Findings

The following subsections describe the data analysis and findings related to each hypothesis, followed by interview and questionnaire results. We used a combination of both qualitative and quantitative techniques in order to best assess each hypothesis.

5.1 H1: Effect on the Number of Questions Asked

H1 speculated that providing dimension coverage information would increase the number of questions asked. To evaluate this hypothesis, we first identified and categorized instances of questions through multi-pass open video coding. We qualitatively analyzed transcribed alouds to identify and count the number of questions asked by each participant.

Following Liu and Heer [20], we considered a question as “an indication of desire to examine certain aspects of the data.” The rationale behind adopting this definition was the similarity between tasks (exploratory), data types (tabular), and the type of data analysis in the studies. In addition, this definition could be objectively assessed by a human coder. A question did not need to end with a question mark. Following is an example question from the study: “I want to look at Sales and Profit for Product Categories”. We only considered utterances that we could confidently identify as a question and did not take into consideration vague and incomplete ones such as “...now I'm gonna look at [didn't continue]”. Although very rare in both conditions, we also excluded questions that were logically invalid. For instance, one participant investigated the relationship between Sales and Container Type (types of containers used for shipping items to customers, e.g. large box): logically, the shipping container should have no impact on Sales. On the other hand, we did consider valid questions that were inconclusive and did not result in a particular finding. This decision was made based on the fact that a relevant question about data can and should be asked even if it does not yield any results. Each identified question was timestamped for future fast retrieval.

Table 1 shows the number of valid questions for each condition and Figure 5 shows the distribution of questions per condition. The result of a two-tail independent t-test (p <0.0001, t=31.623, df=9) showed that the full prototype group asked more valid questions on average.

Table 1: Total number of valid questions for each condition.

Condition	Count of valid questions	Average	StdDv
Baseline	94	9.7	3.3
Full	187	18.7	6.6

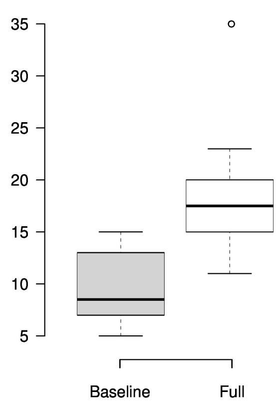


Fig. 5: Boxplots showing count of valid questions asked by participants in each condition.

Next, we investigated why and how providing live coverage information (the difference between conditions) resulted in an increase in the number of questions by full version users. Using each question's time stamp, we reanalyzed videos and identified utterances right before the formation of the question (if any). After extensive multi-pass analyses, we identified two types of utterances related to the question formation process (Generative & Recollective). Generative utterances represented question formation activities that did not necessarily require the analyst to remember work so far (e.g., “Let's start with Profit and Sale for Regions”). On the other hand, Recollective utterances were indicative of a need to remember prior work (e.g., “let me see... what can we analyze more”). We identified a total of 101 Recollective (full=51, baseline=50) and 64 Generative (full=33, baseline=31) utterances. Interestingly, these numbers were very similar across conditions even though the total number of questions differed, likely because not all questions were preceded by intelligible utterances. Since Generative utterances marked questions that were not reliant on remembering history, we focused our further analysis on Recollective utterances.

Next, for each Recollective utterance, we further analyzed the videos to understand if and how participants interacted with the history views. We identified the interplay between participant and tool (if any) and what part of the GUI was targeted. To detect the GUI target, we relied on three sources: 1) user interaction with the tool (e.g., clicking a dimension's name in Scented View or the Dimension List, browsing charts in Sequence View, opening Data View), 2) physical gestures (e.g., tracing the list of the dimensions with a finger or a pen and reading the dimension names aloud), and 3) participants' alouds (e.g., “oh it [Scented View] says I missed Returns”). Table 2 shows some typical examples of Recollective utterances, the targeted interface widget, and the user interaction.

Table 2: Examples of Recollective utterances for each condition. This table also shows the interface target that participants were interacting with at the time of producing utterances and the interaction itself.

Condition	Utterance	GUI Target	Interaction
Baseline	“did I do Order Quantity?”	Sequence View	Browsing
Baseline	“what else [dimension] we have here?”	List of dimensions	Tracing list with finger
Baseline	“let me see, what can I analyze more”	Sequence View	Browsing
Full	“what next? oh, didn't check [Product] containers”	Scented View	Tracing list with mouse pointer
Full	“lets go back and see. with Profit, I have looked at Sales and City”	Scented View	Clicking
Full	“did I consider Days to Ship with Product Category?”	Scented View	Clicking

Figure 6 shows the breakdown of Recollective utterances based on condition and GUI target. The total number of identified Recollective attempts were almost equal, 50 and 51 for baseline and full conditions respectively. As shown in the figure, full version users relied heavily on Scented View to recall prior work while formulating new questions (82% of cases). To achieve the same goal, baseline version users relied mostly on the Sequence View (62% of cases). Interestingly, we found that in the other 38% of cases, baseline version users referred to the list of dimensions (even though it did not contain any dimension coverage information) and tried to recall what had been done by looking or tracing through the list. This suggests that users intrinsically expect this view to help them to remember their prior analysis. The extensive use of Scented View for recalling past work by full version users corroborates the speculation that the availability of dimension coverage information helped people to formulate questions. In addition, we analyzed question formation activity for each group over the analysis session time, but the patterns were sporadic and inconclusive.

5.2 H2: Effect on the Number of Findings

H2 posited that revealing dimension coverage information would result in more findings. Following Liu and Heer [20], we define a finding as one of the following:

Fig. 6: GUI targets that participants interacted with while making Recollective utterances. Full version users frequently referred to Scented View whereas baseline version users relied on Sequence View and the list of dimensions.

- **Observation: "a piece of information about the data that can be obtained from a single state of the visualization system".** For example, "I see that lots of round-tables are sold in Texas".
- **Generalization: "a piece of information acquired from multiple visualization states".** For example, "In the south, sales of furniture are higher than any other product category".

This definition excludes any common sense or intuitive conclusions that participants made about the data, in order to isolate findings to only those that were supported by the analysis tool. To collect findings, we used a multi-pass open-coding approach to qualitatively analyze participants' alouds and their notes. We manually transcribed all participants' alouds from the video recordings, using Transana [3] for video analysis. Later, using the transcribed data, we identified and counted findings for each participant. Each finding was time-stamped to enable cross-referencing with video. We also reviewed participants' notes and extracted all the recorded findings. We only considered relevant and correct findings and excluded all the vague and incomplete instances, such as "...it seems that Sales are [mumbling something]". We also ignored findings that were based on an invalid statistical function or an incorrect interpretation of data. For example, if using sum of values instead of average resulted in an incorrect interpretation of data, that finding was ignored. Table 3 shows examples of extracted findings for each condition.

Two independent researchers coded the transcribed utterances. First, both coders analyzed 4 randomly selected experimental sessions (2 full and 2 baseline sessions). Next, each coder independently analyzed 8 of the remaining sessions (4 full plus 4 baseline, assigned randomly). For the sessions analyzed by both coders, we included only those findings where both coders were in agreement. Inter-coder reliability was 0.89, calculated using Krippendorf's alpha.

Table 3: Examples of participants' findings for each condition

Condition	Finding
Baseline	"...in product categories, technology shows a strong performance..."
Baseline	"...Kansas and New Mexico have biggest negative profit..."
Baseline	"...in customer segments, consumer [segment] had the steadiest performance in all 4 years..."
Baseline	"...home-office and small business profit and sales are increasing..."
Full	"...office supplies in west and furniture in east have poor sales..."
Full	"...spikes in order processing times in Q3 and Q4 in 2010..."
Full	"...in west, California has the strongest performance..."
Full	"...profit has gone down from 2009 to 2010..."

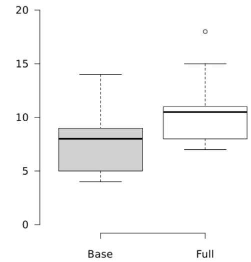


Fig. 7: Total count of findings by participants in each condition.

Figure 7 shows the total count of findings by participants in each condition. Because the data could not be transformed to fit a normal distribution, we analyzed the results using the non-parametric Mann-Whitney test. Although full version users asked more questions on average, the Mann-Whitney test showed that this difference was only marginally significant (w = 24.5, $p < 0.055$, Cohen's d=0.419).

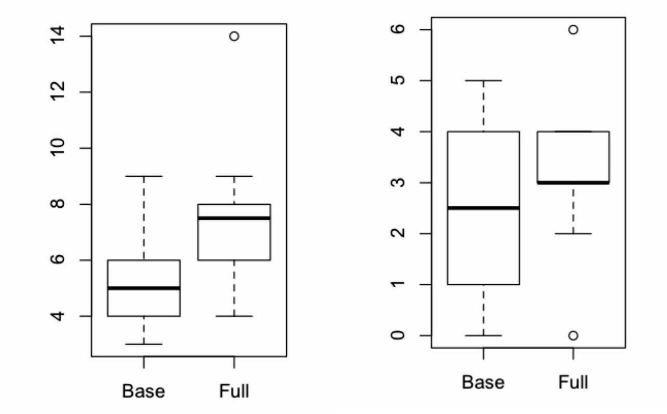
5.3 H3: Effect on the Breadth of Analysis

H3 posited that dimension coverage information would increase the breadth of analysis without sacrificing depth. To evaluate H3, we conducted another multi-pass qualitative analysis to investigate the *process* leading up to participants' findings (findings were those identified in the analysis of H2). Using the timestamps associated with findings and questions, we found the question corresponding to each finding (if any). Looking more closely at questions and findings enabled us to categorize findings into two categories of *top-level* and *drill-down*. Top-level findings are the result of starting a new line of inquiry. Drill-down findings are the result of drilling in on top-level findings. For example, a participant examined "what is the relationship between Profit, Returns (i.e. returned merchandise) and Regions". She observed, "West loses lots of profit because of returns". Triggered by this finding, she formulated the next question as "Which Product Category in Region [filter: West] has biggest Returns". Consequently, she discovered that "lots of furniture was returned in west". In this example, the former and latter findings are top-level and drill-down findings respectively. Each finding was only considered either top-level or drill-down. In the previous example, if the drill-down finding had in turn triggered a further investigation to discover what furniture items were returned the most, we would still have considered it as a drill-down finding.

Conceptually, top-level and drill-down findings can be considered to represent breadth and depth of exploratory analysis. Top-level findings involved investigating a new aspect of the problem (i.e. a new line of inquiry). For a top-level finding, an analyst created a new question with a focus different from the previous question (e.g. shifting focus from Profit to Sales). Therefore, a greater number of top-level findings suggests a larger breadth of analysis. On the other hand, drill-down findings involved continued investigation of the same problem. Though there may have been some changes in dimensions or filtering, a question resulting in a drill-down finding essentially followed the same analysis path as the preceding question. As a result, more drill-down findings suggests a greater depth of analysis.

Figure 8 depicts the count of top-level and drill-down findings in the two conditions. We performed separate Mann-Whitney tests to compare full and baseline groups in terms of their top-level and drill-down findings. We found that participants who used the full prototype produced more top-level findings than those who used the baseline (w=21, $p < 0.030$, Cohen's d=1.05). Full version users also produced slightly more drill-down findings on average, but this difference was only marginally significant (w=42.5, $p < 0.059$, Cohen's d=0.36).

As a second metric of breadth, we also examined the number of dimensions that each participant considered in their analysis (see Figure 9). Using the questions identified for evaluating H1, we counted the number of dimensions considered by each participant. Over the same period of time (60 minutes per participant), full version users



(a) Top-level

(b) Drill-down

Fig. 8: Count of top-level and drill-down findings by participants in each condition.

considered an average of 16.6 dimensions (SD= 2.7), versus 13.6 (SD=3.1) for baseline version users. A two-tail Welch Two-Sample t-test showed a statistically significant difference between the averages ($t(16.3) = 2.43$, $p < 0.027$). In line with our analysis of top-level findings, this result shows that full version users exhibited a greater breadth of exploration.

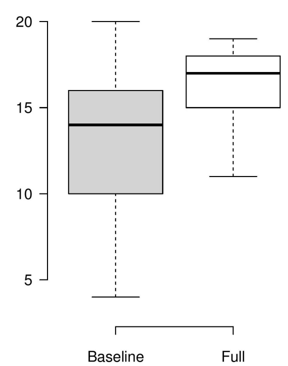


Fig. 9: Dimensions considered by Full and Baseline tool users.

5.4 Questionnaire & interview Results

At the end of the study session, full version users filled out a questionnaire that elicited information on their experience. The questionnaire consisted of two parts. The first part asked participants what activities they found Scented, Data and Sequence Views most useful for. Figure 10 summarizes participants' responses.

The second part of the questionnaire asked participants to rate the overall usefulness and understandability of each view using a five-point Likert scale. All participants strongly agreed or agreed that the views were understandable. As depicted in Figure 11, 7 out of 10 participants were uncertain about the overall usefulness of Data View. On the other hand, participants all agreed or strongly agreed that dimension coverage information (i.e. Scented View) was useful.

In the interviews, all full version users reported that they relied on Scented View to recall prior work. In particular, they stated that they found the dimension coverage and co-investigation information very useful. Following are a few examples of participants' comments about Scented View: "I definitely used this [Scented View] a lot, it was quite nice to see what variables I did", "It's nice to see what I have done", "I found the changing colours [i.e. co-investigation information] very useful", "colour coding helped me to [see] what I wanted". Although

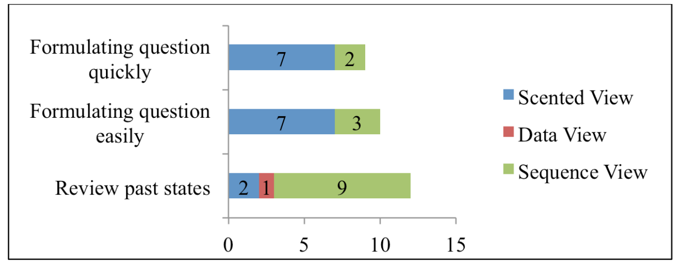


Fig. 10: Tasks for which different history components were considered useful. Sequence View was most useful for reviewing past states. Scented View was considered helpful for formulating questions quickly and easily. Data view was not considered very useful.

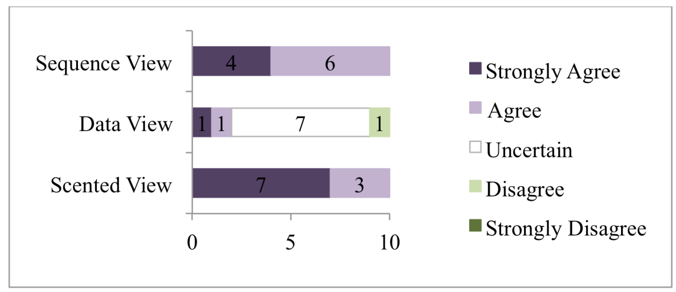


Fig. 11: Rated overall usefulness of different history views. Participants were asked to agree or disagree with a statement that each view was useful.

all the full version users valued dimension coverage information, six out of ten participants reported they did not use the quantitative magnitude information (encoded as bar length). For example, one participant said "I think bars helped me to see what variables I did/did not [do], but I didn't really read the numbers". None of the participants reported that they used Data View for carrying out their analysis task. One participant said "it did pop open a few times, but I didn't really check it" and another participant said "this [pointing at Scented View] was enough for me".

6 Discussion

H1 was strongly supported: we found a statistically significant difference between conditions for the number of questions asked, with full version participants asking almost twice as many questions on average than baseline participants. Our analysis of Recollective utterances showed that when users needed to understand what they had already done in order to formulate a new question, full version users relied on Scented View and baseline version users relied on Sequence View.

Sequence view is inherently limited in providing first-hand insight into the coverage of dimension space. We observed that acquiring coverage information from Sequence View consisted of many steps, starting with filtering or browsing the list to find target visualizations, investigating individual visualizations to extract coverage information, and remembering which dimensions were included in these visualizations (note: no one recorded this coverage information on paper). Fewer steps were required for Full version users to acquire the same information because dimension coverage information was constantly present in the interface. Recent research [19] indicates that executing physical sequences is one of the main contributors to the overall cost of interacting with a visualization tool. The recall task was undoubtedly costlier for baseline version users, which is likely one of the key reasons why they formulated fewer questions overall.

H2 was not supported: while there was a trend towards more overall findings for full version participants, this difference was only marginally significant. We speculate that there was insufficient statistical power in our experiment and that more participants might have revealed a significant difference here. However, full version participants did produce significantly more top-level findings (full avg.=7.6, baseline avg.=5.2).

Our results showed strong support for H3. Full version users demonstrated greater breadth in their analysis by identifying significantly more top-level findings than baseline users. They also investigated significantly more dimensions. These results indicate that full version users followed more lines of inquiry overall. We also note that this analysis breadth was not at the expense of depth. Drill-down findings represented additional details within the same line of inquiry (i.e. depth). We found no significant difference between full and baseline users in terms of drill-down findings (if anything, there was a trend towards full version users finding more of these as well).

We attribute the observed greater breadth of analysis by full version users to the availability of Scented View. This view facilitated discovery of "what has not been done", and consequently enabled the analyst to identify new relevant avenues of analysis. Note that we are not arguing that the breadth of analysis is more important than the depth (nor vise-versa). Yet, prior research has suggested that under some circumstances such as exploring new data, it is beneficial to encourage increased breadth, as it can reduce the likelihood of empty results [14] and prevent premature fixation on a single aspect of data [32]. In line with Wattenberg and Kriss [29], we speculate that encouraging breadth may be most useful when exploring a new dataset; the benefits may diminish as analysts become more familiar with the shape and structure of their data and clarify their analysis goals.

Questionnaire results indicated that full version users found Sequence View most suitable for reviewing past states. On the other hand, they found Scented View to be more helpful when forming questions. Interestingly, baseline users referred to the list of dimensions in 19 out of 50 (38%) attempts to recall analysis (Figure 6), even though this view did not contain any dimension coverage information in the baseline condition. This suggests that people may intuitively expect this view to help them in recalling dimension coverage information, and that our choice to embed this information within the list using scented widgets was appropriate.

Interestingly, results of the interviews with full version users showed they all found information about "what has been investigated and in what combinations" more useful than the frequency information (i.e. bar length and the number). This suggests that the Scented View could be further compacted by reducing the bar to a smaller mark that is either present or absent.

Another interesting finding relates to the relative value of dimension coverage information as compared to data values coverage. While Scented View was highly used and considered very useful, the opposite was true for Data View. None of the participants reported using Data View to help them formulate questions (Figure 10). In addition, only two out of ten participants agreed or strongly agreed that this view was useful (Figure 11). These results suggest that dimension coverage information may be much more important than data values coverage. Although we could not experimentally isolate the value of Data View because it was in the same condition as Scented View, ratings and qualitative observations suggested that participants found Scented View to be much more useful. It is of course possible that the value of Scented View was related to the specific task in our study or to the view's visual prominence in the interface. In [30], the exploratory analysis task involved investigating a constant set of dimensions by manipulating the filtering of values. In such a case, being able to visually understand which data values were explored versus left out might be more important to formulating new questions. Nonetheless, evidence here suggests that dimension coverage information is more useful than data values coverage in the more general case.

Our experimental results are subject to some additional caveats including: use of student participants, task design, metrics, and design choices. While we paid careful attention to the study setting (e.g., recruiting participants with some business knowledge and some experience in visual data analysis), it is not clear to what extent the student participants behaved like real analysts. We also consulted a business PhD student to carefully model the analysis task, but real exploratory business data analysis tasks may require extra steps such data wrangling and use of multiple data sources and tools. Although we designed the history views as well as the VA tool through an iterative design process, some of our findings may be related to our specific design choices. One specific issue is the scalability of our design. Although the tested data set contained a moderately large number of dimensions (24), it is not clear how well this design would work for a larger number of dimensions when scrolling is required. One possible approach could be to add an overview of the Scented View with all the dimensions constantly visible. Finally, measuring concepts like analysis breadth is very challenging. We introduced some new metrics to capture this concept but they are undoubtedly incomplete and could be refined in future research. These caveats all limit the degree to which our results can be generalized; future studies are needed to confirm our results for a wider population.

7 Conclusion and Future Work

We examined the value of providing dimension coverage information to support exploratory analysis of unseen data. We illustrated how this information can be incorporated into the interface of a visual analysis tool by using scented widgets. Our results demonstrated that this approach could increase the number of questions asked about data and expand the breadth of analysis without sacrificing depth.

In the future, we would like to explore dimension coverage designs that could emphasize the most important dimensions. It is possible that some dimensions intrinsically contain more information related to an analysis task. If we could identify these, we could place them in a more prominent position on the list (e.g. top). One possible approach is to rank the list of dimensions automatically based on the frequency of investigation or through data mining or machine learning techniques. We would also like to investigate alternative means of representing data values coverage. One possible alternative is to temporarily replace the frequency bar next to a dimension's name with data coverage information when a dimension is selected. In addition, it would be interesting to explore mechanisms to reveal co-investigation information involving only subsets of data values.

We tested the benefits of visualizing dimension coverage information in a specific EDA setting where analysts create visualizations to answer questions and evaluate hypotheses. Further research is required to evaluate if and how this approach would benefit other types of analysis and how it would interact with other visual analytics approaches such automated recommendation of visualizations or dimensions. In addition, it is important to further evaluate our approach with expert business analysts since our study involved student participants.

Our scented widget design is specific to tabular data. Extension of the data-centric history idea to other types of data that do not have this discrete tabular nature is not obvious. However, our work does demonstrate that showing people a summary of their past work can help them to decide what to do next. This finding should have wider generalizability and implications beyond one specific data type. Figuring out how to apply this idea to other data, however, would require completely rethinking the design and the content to show. We would like to explore such extensions in future work.

We would also like to examine our approach of providing dimension and data space coverage in a collaborative setting. In particular, we would like to extend our scented widgets with capabilities to distinguish among the activities of multiple users in a collaborative team. For example, if four people are all exploring the same dataset, it could be helpful to see what aspects each person has worked on. When working simultaneously, a dimension centric view might serve as a helpful awareness mechanism, indicating which parts of the data are being neglected and might be worthy of exploration. Such extensions would need to consider how to present the contributions of different users without cluttering the interface.

Acknowledgments

This research was supported by grants from NSERC and GRAND. We would like to thank Anirban Kar for his help in designing the study task.

REFERENCES

[1] JFreeChart. http://www.jfree.org/jfreechart. Accessed: 2016-03-28.

[2] Tableau. http://www.tableau.com. Accessed: 2016-03-28.

[3] Transana. http://www.transana.org. Accessed: 2016-03-28.

[4] J. Abello, F. Van Ham, and N. Krishnan. Ask-graphview: A large scale graph visualization system. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):669–676, 2006.

[5] P. Baudisch, D. Tan, M. Collomb, D. Robbins, K. Hinckley, M. Agrawala, S. Zhao, and G. Ramos. Phosphor: explaining transitions in the user interface using afterglow effects. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pages 169–178. ACM, 2006.

[6] L. Bavoil, S. P. Callahan, P. J. Crossno, J. Freire, C. E. Scheidegger, C. T. Silva, and H. T. Vo. Vistrails: Enabling interactive multiple-view visualizations. In *Proceedings IEEE Conference on Visualization*, pages 135–142. IEEE, 2005.

[7] K. Brodlie, A. Poon, H. Wright, L. Brankin, G. Banecki, and A. Gay. Grasparc-a problem solving environment integrating computation and visualization. In *Proceedings IEEE Conference on Visualization*, pages 102–109. IEEE, 1993.

[8] S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, C. T. Silva, and H. T. Vo. Managing the evolution of dataflows with vistrails. In *Proceedings. 22nd International Conference on Data Engineering Workshops*, pages 71–71. IEEE, 2006.

[9] W. S. Cleveland and R. McGill. An experiment in graphical perception. *International Journal of Man-Machine Studies*, 25(5):491–500, 1986.

[10] M. Derthick, J. Harrison, A. Moore, and S. F. Roth. Efficient multi-object dynamic query histograms. In *IEEE Symposium on Information Visualization*, pages 84–91. IEEE, 1999.

[11] C. Dunne, N. Henry Riche, B. Lee, R. Metoyer, and G. Robertson. Graphtrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1663–1672. ACM, 2012.

[12] S. G. Eick. Data visualization sliders. In *Proceedings of the ACM symposium on User Interface Software and Technology*, pages 119–120. ACM, 1994.

[13] M. Glanzer and A. R. Cunitz. Two storage mechanisms in free recall. *Journal of verbal learning and verbal behavior*, 5(4):351–360, 1966.

[14] M. Hearst. *Search user interfaces*. Cambridge University Press, 2009.

[15] J. Heer, J. D. Mackinlay, C. Stolte, and M. Agrawala. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1189–1196, 2008.

[16] W. Javed and N. Elmqvist. Explates: spatializing interactive analysis to scaffold visual exploration. In *Computer Graphics Forum*, volume 32, pages 441–450, 2013.

[17] N. Kadivar, V. Chen, D. Dunsmuir, E. Lee, C. Qian, J. Dill, C. Shaw, and R. Woodbury. Capturing and supporting the analysis process. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 131–138. IEEE, 2009.

[18] A. Kobsa. An empirical comparison of three commercial information visualization systems. In *IEEE Symposium on Information Visualization*, pages 123–130. IEEE, 2001.

[19] H. Lam. A framework of interaction costs in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1149–1156, 2008.

[20] Z. Liu and J. Heer. The effects of interactive latency on exploratory visual analysis. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2122–2131, 2014.

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

[22] T. Munzner. *Visualization Analysis and Design*. CRC Press, 2014.

[23] P. H. Nguyen, K. Xu, A. Wheat, B. Wong, S. Attfield, and B. Fields. Sensepath: Understanding the sensemaking process through analytic provenance. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):41–50, 2016.

[24] P. Pirolli and S. Card. Information foraging. *Psychological review*, 106(4):643, 1999.

[25] E. D. Ragan, J. R. Goodall, and A. Tung. Evaluating how level of detail of visual history affects process memory. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 2711–2720. ACM, 2015.

[26] A. Sarvghad, N. Mahyar, and M. Tory. History tools for collaborative visualization. *Collaborative Visualization on Interactive Surfaces*, 2009.

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

[28] C. Stolte, D. Tang, and P. Hanrahan. Polaris: A system for query, analysis, and visualization of multidimensional relational databases. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):52–65, 2002.

[29] M. Wattenberg and J. Kriss. Designing for social data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):549–557, 2006.

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

[31] C. Williamson and B. Shneiderman. The dynamic homefinder: Evaluating dynamic queries in a real-estate information exploration system. In *Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 338–346. ACM, 1992.

[32] K. Wongsuphasawat, D. Moritz, A. Anand, J. Mackinlay, B. Howe, and J. Heer. Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):649–658, 2016.

[33] J. Zhao, C. M. Collins, F. Chevalier, and R. Balakrishnan. Interactive exploration of implicit and explicit relations in faceted datasets. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2080–2089, 2013.