# Evaluating Strategies of Exploratory Visual Data Analysis in Multi Device Environments

Abeer Alsaiari[1,2] Jillian Aurisano[1,2] and Andrew Johnson[1,2]

[1]Department of Computer Science, University of Illinois at Chicago, Chicago, United States of America
[2]Electronic Visualization Laboratory

**Abstract**

*Supporting exploratory visual data analysis is essential when multiple analysts collaborate using multiple devices. Yet, we still have no full understanding of how the iterative process of analysis unfolds in complex settings. In this paper, we present the results from an exploratory study where six groups of three participants performed a collaborative visual data analysis task in a complex multi-user multi-device environment. We found that the course of the analysis happens at two levels. Within each level, we observed a set of exploration patterns. We present a categorization of the analysis structure in such a complex environment and discuss the implications of device affordances on this categorization. We also discuss this categorization in relation to the current structural assumptions of exploratory visual analysis.*

## 1. Introduction

Exploratory visual analysis (EVA) is an iterative process that involves cycles of visualization creation, interaction, and refinement. Existing models and frameworks of visual analysis process (e.g., [LTM17], [BM13]) touch on task analysis but not specifically for understanding how EVA is defined and structured [BH19]. Lam et al. [LTM17] presented a framework based on a review of 20 design study papers to describe the high-level analysis goals and how they can be achieved with low-level tasks identified from the review. Brehmer and Munzner [BM13] reviewed the literature on visualization tasks and derived a multi-level typology of visual analysis tasks. The typology comprises why and how a task is performed, and what are the input and output to complete it. These frameworks help to express high-level tasks as sequences of low-level tasks. However, the current literature of EVA provides different assumptions about how it is defined, and there is no consistent understanding about how it is structured. [BH19]. A recent work by Battle and Heer has made an effort to provide an understanding of exploratory visual analysis by evaluating how singular analysts behave when performing EVA tasks identified from the literature [BH19]. They found participants' analysis sessions have patterns and can overlap in some states [BH19]. When multiple analysts collaborate, it is even more essential to understand EVA process in order to support the analysis task and increase the collaboration efficiency. Building on their work, we conducted a study to examine EVA in a collaborative setting.

The focus of this paper is to evaluate strategies of EVA in a collaborative multi-user multi-device environment. We found that participants initiated different analysis paths exploiting the large space offered by the large display. These paths were evolved by visualizations created by users, either as a group, or individually using portable devices and then merging to the public work. Our choice of experiment design is justified by two reasons. First, settings of multiple devices allow different collaboration styles by enabling individual and group work, which is an essential factor in collaborative activity [IIH*13] [IFP*11]. Second, they enable the leveraging of different devices capabilities to support the analysis task [IIH*13] [BFE14] [HBED18]. The findings provide insights into understanding EVA in collaborative settings of multiple users and devices, and they can be evaluated in different contexts to find commonalities and differences. We observed certain impacts from the device affordances on how the analysis was performed and structured. The large display influenced the production of different analysis paths, and the performing of different exploration tasks such as compare, correlate, etc. In addition, The large space offered by the large display allowed for a multi-level exploration activity as shown in Figure 5. The presented categorization expand the definition of exploratory visual analysis by considering technological and social factors. Few research works addressed collaborative visualization around interactive surfaces. Mahyar [MST09] and Isenberg [ITC08] developed models of collaborative visual analysis processes around large displays. However, this work differs from previous work by addressing the analysis process from the dimension of analytical flow and structure.

## 2. Methodology

We conducted an exploratory study to assess user's analytical processes and strategies in a collaborative and multi-device setting. We used qualitative methods to analyze the collected data.

### 2.1. Apparatus

We used a cross-device framework , PolyVis, for collaborative visual data analysis [AJN19]. This framework integrates SAGE2 [MAN*14], a web-based tiled display wall middleware, with portable devices of different modalities: laptops, tablets, and AR headset (HoloLens). The framework enables the rapid construction and sharing of visualizations across devices. For device agnostic sharing and rendering of visualizations, the framework adopts the declarative visualization design using Vega [SRHH15] and Vega-lite [SMWH16].

**PolyVis example scenario**: users start an exploration of earthquakes events and wells injection during 2010. First, they filter and mine data by year and location from the database. Then, they specify their visual representation (i.e. map) to overview the locations of these data points on the large display. Each map is attached with a bar code that can be scanned using tablet or phone's camera. The scanning will pull the map visualization to the portable device for further analysis. User on portable device creates charts such as scatter plot, line or bar charts and pushes them to the wall. To view a specific area from the map in 3D, user selects that area using the laptop to view the 3D representation of selected data points in HoloLens.

### 2.2. Study Design

We recruited 18 subjects, 6 groups of 3 participants from a pool of undergraduate and graduate students from the computer science department at the university. Participants were 13 male and 5 female, between ages 18 and 34. Participants spent between 45min to 1.5hrs in the study. Participants had some familiarity in visual data analysis ranging from medium to advanced. Each group performed visual analysis tasks using two geosciences datasets. The first dataset (earthquake dataset) contained information about earthquakes incidents in Oklahoma and California from the years 2000 to 2010. The second data set (wells dataset) contained information about the fracking activities in Oklahoma and California also from the years 2000 to 2010. The earthquake dataset is provided courtesy of http://service.iris.edu/ and the Wells injection dataset is provided courtesy of http://www.occeweb.com/. Participants were instructed to ascertain the relationship between injection volumes and pressure of fracking, and the frequency of earthquakes in the states of Oklahoma and California.

The study was conducted in a room of approximately 12.61 by 7.59 meters, equipped with a high-resolution large display. Overall display size is approximately 7.3 by 2.05 meters at a resolution of 11,520 by 3,240 pixels. Other portable devices were placed on a table in the middle for use during the study: one MacBook Pro (macOS Sierra, 2.4 GHz Intel Core i5), one 8" Samsung - Galaxy Tab A (32GB, Android 9 (Pie)), one 10" Samsung - Galaxy Tab A (64GB, Android 9 (Pie)), and one Microsoft HoloLens 1 (Windows Mixed Reality OS, Intel 32-bit (1GHz) CPU, 2 GB RAM). For further analysis that is beyond the scope of this paper, each of the portable devices was attached with Mocap markers for position and orientation tracking. In addition, three caps with attached Mocap markers were provided to the users. The tracking data were streamed to a Unity application portraying the physical space as a 3D model. The study was video and audio recorded using two cameras, one rear camera showing the full room from behind and one front camera showing the subjects interaction with the large display. Figure 1 illustrates the setup of the study.

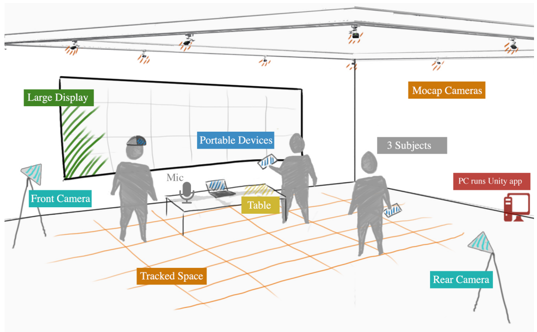



**Figure 1:** *Illustration of the study setup.*

### 2.3. Coding and Data Analysis

We collected data in the form of recorded videos, systems logs, tracking data, and questionnaires. About 420 minutes of videos were collected (an average of 70 minutes per session). The primary coder started a first pass of qualitative coding of the analysis structure and exploration patterns. Using an excel sheet, the coder wrote down for every created visualization how it is created (i.e. data attributes, visual encoding), its implicit or explicit relation to other visualizations if any, and the purpose of creating the visualization, interpreted from the context of participants discussion. Then, a flow diagram of the created visualizations during each session is plotted in chronological order with arrows indicating the first set of codes. In a second pass of coding with a secondary coder, they developed a theme of the analysis structure by grouping and refining the set of codes. Codes are available in our supplemental material online [†].

## 3. Findings

In this section, we report the observed exploration patterns and how they compose a structure of the analysis flow as: (a) a high-level analysis path initiation and (b) a low-level visualizations exploration that occur along paths.

### 3.1. Two-level Structure of Analysis Sessions

We observed that the course of the analysis happens at two levels. Within each level, we observed a set of exploration patterns. At the higher level, participants created exploration paths that each consisted in a set of subtasks. To identify analysis paths, we coded times when participants individually or collaboratively specify a new subset of data points and start analyzing this subset through a set of visualizations. We observed three path emerging patterns. These patterns will be discussed below in detail. Along each analysis path, we also observed a set of view-to-view generation patterns.

[†] https://github.com/uic-evl/eurovis2020$_a$*lsaiari*

These are the low-level exploration patterns that occur within the larger cycles of the analysis. We below discus these observations and shed a light on how this structure corresponds to the current definition of exploratory visual analysis.

### 3.1.1. Initiation of Analysis Paths

We identified the higher level of the analysis structure by analysis paths that were taken by participants. Most of the time, participants start with one analysis path and subsequently start another analysis path to work along both paths in parallel or sequentially. In one case, we observed that participants initiated four analysis paths in parallel. All further created visualizations along analysis path compose analysis states. We observed three patterns of path initiation in our study as below.

**Parallel analysis paths**: This pattern of analysis was the most common among all groups since the open-ended task was exploring two datasets and make observations of possible correlation and comparison. Participants fetch subsets of data points from different datasets using same attributes value (*i.e. year = 2009, or location = "California"*). Then, they analyze these subsets of data using different visualizations and measures. The analysis of visualizations along one path was highly affected by the flow of the other path. For example, a participant requested to see the count of earthquakes over time after another participant merged the average volume over time of wells injection. Participants in general did organize visualizations from the same analysis path in a cluster. However, there were no specific patterns on how groups work on parallel paths. They alternated frequently between working on both at the same time and focusing on one for some time and then switching to the other one.

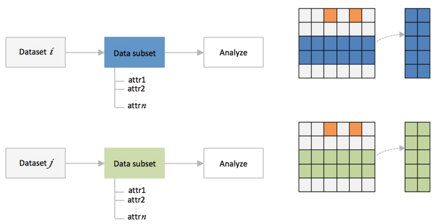



**Figure 2:** *Participants frequently subset data points by same attributes from different datasets to work on both sets in parallel.*

**Sequential analysis paths**: This type of analysis path takes place later in the analysis session. The Sequential term here refers to the temporal sequence and sequential course of analysis. Participants initiated this type of analysis path to re-examine formed hypotheses or observations by looking into different dimensions of the datasets. For example, participant P1 and P2 found no spatial correlation between earthquakes events and wells locations on the year 2009. P2 suggested: " *We can try another year...*" in order to test their hypothesis. That led to the initiation of a sequential analysis on a different dimension. Although taking this direction in the analysis session is less frequent, we believe that it is very important for validating hypotheses especially when working on large datasets. One way to support this analysis pattern in visualization systems would be to make it easy to regenerate analysis paths on different dimensions for rapid validation of hypotheses.

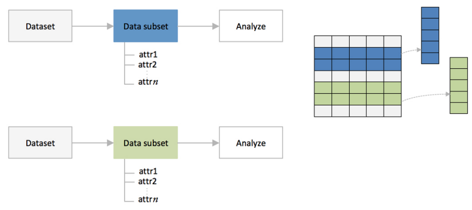



**Figure 3:** *Later in the analysis session, participants initiate a subsequent analysis on a different dimension of the data to validate hypotheses.*

**Focused analysis paths**: Taking this direction in the analysis happens when participants start with a larger scope of data subset and then drill down within dimensions. This is understandable to see if an observation within a larger subset of the data holds when filtered to smaller subsets. For example, participants started an analysis on a larger dimension space (location = "California" AND "Oklahoma") and after forming some hypotheses, they drilled down to verify that the hypothesis is still correct per state. Focus analysis can be considered as special validation task similar to sequential analysis. We observed that participants were also focused on their analysis along the last two patterns of paths. They start the analysis with a larger set of attributes and then they focus on attributes of interest from which they form their hypotheses. We believe that this is because of the gained knowledge from the exploratory analysis on previous dimensions space.

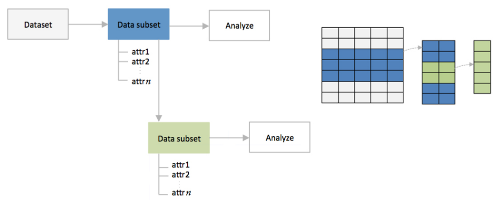



**Figure 4:** *Participants later in the analysis session initiate a focused analysis path by drilling down into a dimension space.*

### 3.1.2. View-to-view Generation along Analysis Paths

The second level that we coded focused on generated visualizations that compose the evolution of analysis paths. Participants specify visual encodings for all produced visualization resulting as a new analysis state. The production of visualization can have an implicit or explicit relationship to another visualization in the analysis session. In other words, participants produced some visualization by explicitly referring to another visualization either to directly compare, correlate, etc.

Some other visualizations are produced to be explored solely without explicit relationship to another visualization (implicit to the

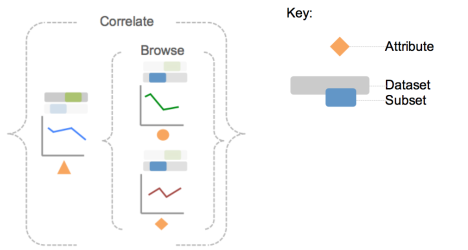
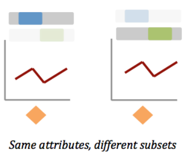
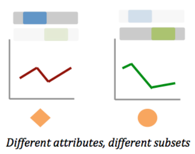



**Figure 5:** *Exploration patterns can occur at different levels. Components of Browse task used to find if they correlate to another attribute.*

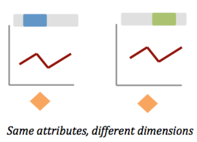
context). We categorized those exploration patterns based on the goal of exploring every visualization. Those categories we present below are not a one-to-one relationship. That is, they can work in hierarchy as shown in Figure 5 where items in a low-level exploration task compose a counterpart for a higher exploration task. For simplicity, we decomposed those exploration patterns and explained below.



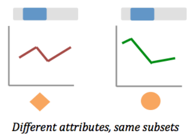

**Compare**: It takes place when participants create visualizations that share an attribute and visual representation. It's a common exploration task to find similarities and differences between data attribute. Compare can take place between attributes of data subsets from same or different population. For example, comparing the average depth of earthquakes in different years output an understanding of attribute trends over subsets from same data population. On the other hand, comparing the average depth of earthquakes and wells in a specific location outputs similarities and differences between an attribute trends over subsets from different data populations.




**Correlate**: As compare, correlate was also a common exploration task to infer the dependency of one attribute on another. Participants created many visualizations, mostly, of the same representation but different attributes to infer possible relationships. We observed that correlating different attributes form different subsets was the most common. In very few cases, it involved the same attributes (i.e. depth) from different data subsets. As mentioned earlier, the presented exploration patterns are not one-to-one. In many cases, participants correlated more than one visualization to infer attributes relationship.




**Validate**: It takes place within the larger cycle of the validation process. It involves the creation of visualization of similar attributes and representation but different dimension space of data points. Participants aim from this task to validate observations made on an earlier visualization from a different dimension space or data subset. It was less frequent than other patterns as the validation process was limited due to the hard tracking of the analysis flow.




**Browse**: We observed that browsing attributes usually takes place mostly at the beginning of the analysis session. At this point participants are aiming to navigate through the data for interesting attributes that may involve interesting trends. After browsing, participants proceed to form different analysis tasks. When participants browse set of attributes, there is no obvious relationship between them, but they form the take away points for further analysis.

### 3.2. Temporal Relationships of Observed Patterns

At the higher level, some paths are initiated naturally after another. For example, a focused analysis path is a drill-down of an earlier analysis path, and it was started later in the analysis session. Sequential paths also naturally come after an earlier analysis path. However, we observed that one team initiated a sequential path, for validation, along the current analysis path. This is because the team planned its analysis strategy at the beginning of the analysis session. Therefore, we believe it is important to provide flexibility in designing visualization tools to support different analysis styles and strategies. Working on both, focused and sequential path, is independent of their ancestor paths. So by default, participants can work on them independently later in the session or along with their ancestors. On the other hand, the goal of initiating a parallel path is to be working on them in parallel with another analysis path. Therefore, they come sometime after the initiation of an analysis path. When participants were working on more than one analysis path, there were no specific patterns on how they alternated between them. At the lower level of exploring different visualizations, the temporal relationships of creating and working on visualization were more dynamic. Participants switched frequently between creating visualizations for browsing, comparing, correlating and validating. They worked forward and backward along the analysis paths.

## 4. Conclusion

We presented a synthesized characterization of the analysis structure from observed analysis behaviors. Our characterization comprises tow levels of exploration patterns. Briefly, participants performed a set of exploration patterns within larger cycles of analysis paths. These findings help us to build a comprehensive view and understanding of exploratory visual analysis. In collaborative contexts, this understanding would help to provide tool support enhancing task performance and collaboration efficiency. Since these findings are grounded to the study setup, they can be evaluated in different collaborative settings to find commonalities and differences. In the future, we aim to understand how variances in the setup, tool design, or analysts' experience may affect exploration strategies. Here, we highlight some limitations of the study that should be noted. The presented categorization of the analysis structure and strategies is what we observed using two datasets of view sets of attributes. Therefore, we cannot generalize our observations to other types of datasets, which may contain dozens of attributes. Exploration strategies may differ with large datasets of many attributes. In addition, the used visualization framework supports the basic types of visualizations, which might not be sufficient for complex and large datasets that require advanced visualizations.

## References


[AJN19] ALSAIARI A., JOHNSON A., NISHIMOTO A.: Polyvis: Cross-device framework for collaborative visual data analysis. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (2019), IEEE, pp. 2870–2876. 2

[BFE14] BADAM S. K., FISHER E., ELMQVIST N.: Munin: A peer-to-peer middleware for ubiquitous analytics and visualization spaces. *IEEE Transactions on Visualization and Computer Graphics 21*, 2 (2014), 215–228. 1

[BH19] BATTLE L., HEER J.: Characterizing exploratory visual analysis: A literature review and evaluation of analytic provenance in tableau. In *Computer Graphics Forum* (2019), vol. 38, Wiley Online Library, pp. 145–159. 1

[BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics 19*, 12 (2013), 2376–2385. 1

[HBED18] HORAK T., BADAM S. K., ELMQVIST N., DACHSELT R.: When david meets goliath: Combining smartwatches with a large vertical display for visual data exploration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), ACM, p. 19. 1

[IFP*11] ISENBERG P., FISHER D., PAUL S. A., MORRIS M. R., INKPEN K., CZERWINSKI M.: Co-located collaborative visual analytics around a tabletop display. *IEEE Transactions on visualization and Computer Graphics 18*, 5 (2011), 689–702. 1

[IIH*13] ISENBERG P., ISENBERG T., HESSELMANN T., LEE B., VON ZADOW U., TANG A.: Data visualization on interactive surfaces: A research agenda. *IEEE Computer Graphics and Applications 33*, 2 (2013), 16–24. 1

[ITC08] ISENBERG P., TANG A., CARPENDALE S.: An exploratory study of visual information analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2008), pp. 1217–1226. 1

[LTM17] LAM H., TORY M., MUNZNER T.: Bridging from goals to tasks with design study analysis reports. *IEEE transactions on visualization and computer graphics 24*, 1 (2017), 435–445. 1

[MAN*14] MARRINAN T., AURISANO J., NISHIMOTO A., BHARADWAJ K., MATEEVITSI V., RENAMBOT L., LONG L., JOHNSON A., LEIGH J.: Sage2: A new approach for data intensive collaboration using scalable resolution shared displays. In *10th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing* (2014), IEEE, pp. 177–186. 2

[MST09] MAHYAR N., SARVGHAD A., TORY M.: Roles of notes in co-located collaborative visualization. *Collaborative Visualization on Interactive Surfaces-CoVIS'09* (2009), 13. 1

[SMWH16] SATYANARAYAN A., MORITZ D., WONGSUPHASAWAT K., HEER J.: Vega-lite: A grammar of interactive graphics. *IEEE transactions on visualization and computer graphics 23*, 1 (2016), 341–350. 2

[SRHH15] SATYANARAYAN A., RUSSELL R., HOFFSWELL J., HEER J.: Reactive vega: A streaming dataflow architecture for declarative interactive visualization. *IEEE transactions on visualization and computer graphics 22*, 1 (2015), 659–668. 2