Evaluating the Benefits of Explicit and Semi-Automated Clusters for Immersive Sensemaking

Ibrahim A. Tahmid* Lee Lisle† Kylie Davidson‡ Chris North§ Doug A. Bowman¶

Center for Human-Computer Interaction and the Department of Computer Science
Virginia Tech

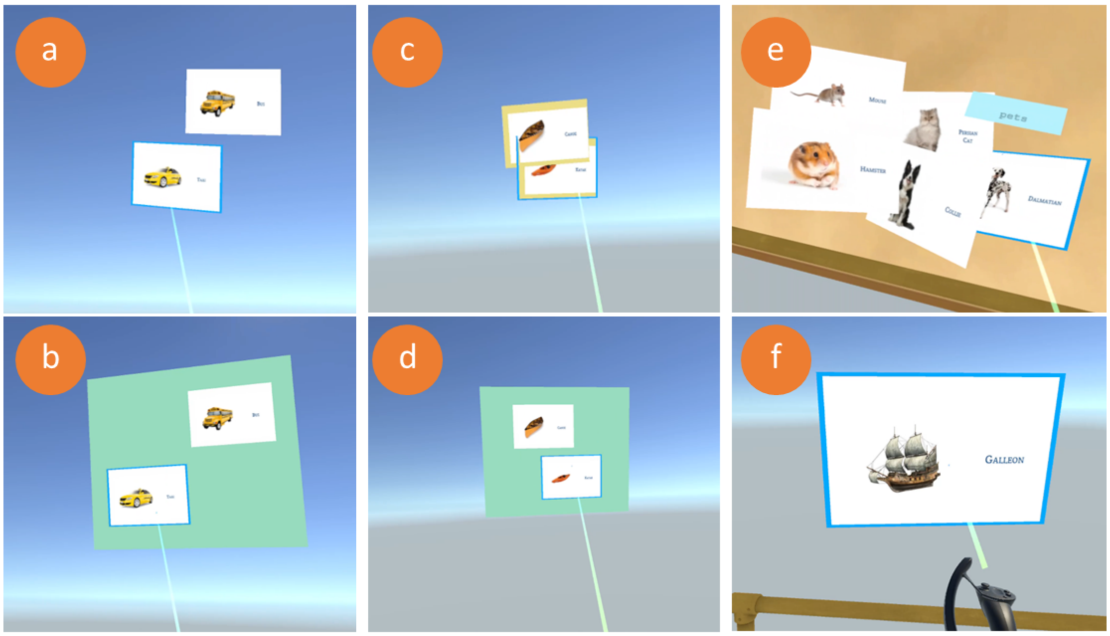

Figure 1: Clustering techniques in an immersive sensemaking tool. (a)&(b): Before and after forming an explicit cluster in the Proximity condition. Two or more documents in proximity to each other automatically create a new cluster; (c)&(d): Before and after forming an explicit cluster in the Overlap condition. The user intersects two documents manually to form a new cluster; (e): Informal cluster formed manually in the Freestyle condition; (f): Sample of a card in the Vehicle dataset;

Abstract

Immersive spaces have great potential to support analysts in complex sensemaking tasks, but the use of only manual interactions for organizing data elements can become tedious. We analyzed the user interactions to support cluster formation in an immersive sensemaking system, and we designed a semi-automated cluster creation technique that determines the user's intent to create a cluster based on object proximity. We present the results of a user study comparing this proximity-based technique with a manual clustering technique and a baseline immersive workspace with no explicit clustering support. We found that semi-automated clustering was faster and preferred, while manual clustering gave greater control to users. These results provide support for the approach of adding intelligent semantic interactions to aid the users of immersive analytics systems.

Index Terms: Virtual Reality; Human AI Collaboration; Semantic Interaction; Immersive Analytics; Clustering

*e-mail: iatahmid@vt.edu
†e-mail:llisle@vt.edu
‡e-mail:kyliedavidson@vt.edu
§e-mail:north@vt.edu
¶e-mail:dbowman@vt.edu

1 Introduction

Visual analytic tools have evolved over the years to support the continuous increase of large multimedia datasets [1, 6, 14, 19, 33]. These systems often aim to combine the computational power of the machines and the insightful perspective of human analysts. The analytic process relies on the users interacting with dataset elements, exploring plausible connections between them, organizing relevant information, and eventually solving a problem or making a decision which is often referred as *Sensemaking* [43]. Examples of sensemaking activities range from finding treatment for critical medical conditions [7, 8] to anticipating acquisition of bio-weapons by foreign nations [43] to understanding racial context among American soldiers during World War II [33] by collecting, organizing, and comprehending information from various sources. While each of these models address specific challenges, they all highlight the need to organize the data.

A recurring behavior in sensemaking is the act of grouping relevant documents in order to synthesize a common pattern, effectively forming a cluster to reduce the workspace clutter [37, 44] that has been observed in many visual analytic platforms [1, 21, 33, 55]. Prior work has shown that a machine could anticipate the user intent of creating such clusters by observing their interactions with the system [13, 16, 21]. Essentially the system could employ one of the many cluster identification algorithms [23] to automate the process of creating clusters, thus enabling the analysts to remain more focused on the high-level analysis of the data [18,19]. This approach of coupling the user's analytic interactions with the computational steps to identify cluster is often termed as *semantic interaction* [19,20]. For example, a user could move a document to a specific place with the intention of creating a spatial construct within the workspace. Therefore, a semantic-interaction-enabled model could anticipate this action, and create the spatial structure *for* the user.

ForceSPIRE [18] explored the benefits of semantic interactions on large two-dimensional displays, and showed how similarity could be represented simply by moving documents closer to each other. However, the parameters for the clusters were updated in the background, and there were not any explicit clusters for the users to interact with. In recent years, VR researchers have been investigating the benefits of analyzing 2D documents such as embodied notes [42], and maps [47] in the 3D space. Lisle et. al [33] developed Immersive Space to Think (IST) which allowed the users to create spatial layouts with 2D multi-media documents in an immersive space to perform sensemaking tasks.

In our research, we extended IST to investigate the mechanisms of leveraging semantic interactions to automatically create explicit clusters in a three-dimensional visual analytic tool. In our design work, we explored the following research question: RQ1: *What is an appropriate level of automation for a clustering feature in an immersive sensemaking system?* We also sought to address two additional questions through an experiment: RQ2: *How does an explicit clustering tool help the analysts organize an immersive workspace?* and RQ3: *What are the benefits and challenges of having semi-automated clusters in an immersive visual analytic system?*

With these questions in mind, we developed three different conditions: Freestyle (informal clusters, no semantic interaction), Overlap (explicit clusters, no semantic interaction), and Proximity (explicit clusters, semantic interaction). We consider documents that do not contain sufficient information to allow automatic clustering by the system without any user interaction. Both Overlap and Proximity allow the users to create and interact with explicit clusters. With Overlap, the users have more control over creating the clusters, while with Proximity, the system creates the clusters for the users by leveraging semantic interaction. Comparing Freestyle and Overlap would help us evaluate the potential benefits of explicit clusters, while comparing Overlap and Proximity would allow us to evaluate the effects of the semi-automated clustering approach, and the tradeoff between control and convenience. Our contributions include:

- The design of two interaction techniques providing explicit cluster support, and a 2.5D visualization technique for clusters.
- An assessment of the benefits of explicit clusters in completing sensemaking tasks in an immersive visual analytic system.
- An evaluation of the benefits and challenges of semi-automated clusters and user-controlled clusters.

2 Related Work

2.1 Sensemaking

Sensemaking is a cognitively difficult task that involves browsing large amount of data to extract meaningful content, and inferring new relations. Pirolli & Card proposed a model with two main loops that complete this process: foraging and sensemaking [43]. The foraging loop involves the analysts browsing the dataset to gather evidence, while the sensemaking loop relies heavily on the success of organizing the dataset by dividing it into clusters that help the analysts generate a better understanding of the underlying data [52].

Sensemaking tools such as Analyst's Notebook [2] and The Sandbox [56] enabled users to organize and make connections between data elements in a spatial workspace. Andrews et al. [1] demonstrated how increased space provided an external memory for the users while they made layouts to complete a sensemaking task. Immersive Space to Think (IST) took this a little further by providing the users an unconstrained three-dimensional space for their sensemaking process [34]. Some studies have explored possible designs for interacting with groups of documents in the immersive space. For example, Post-Post-it [32] was motivated by the physical metaphor of post-it notes, and they developed interactions for creating, removing, and merging multiple groups in the immersive space. Luo et al. [35] looked at the organization strategies followed in AR collaborative applications, relying on the furniture in the physical space to organize their documents.

We explore the effects of explicit clusters on users doing a simple sensemaking task with an immersive visual analytic tool.

2.2 Semantic Interaction

Semantic interaction is defined as the process where the system can *capture* the user action, *interpret* the user intent, *map* the user intent to the underlying change in the spatial structure, and *provide* visual feedback on the updated model within the visual metaphor [19]. Many of the sensemaking tools have found it useful to provide users with a workspace enabling them manually organize spatial representation of information [2,32,49,56]. Semantic interaction strives to enable a similar ability, without requiring the user to manually create the spatial structures [18]. Instead, the goal of the system is to learn from the interactions and co-create the layout.

ForceSPIRE [19] illustrated a set of semantic interactions for analytic process, and explained how they are associated with analytic reasoning. For example, highlighting some text would mark the importance of a phrase, or document movement would mean similarity/dissimilarity with nearby/far documents. However, in ForceSPIRE, the system largely takes the control of the layout away from the analyst.

In our research, we wanted to explore how we can give most of the spatial organization control to the analyst, with the system just helping to identify and make explicit cluster structures within the layout.

2.3 Clustering

Spatial analysis of data involves analysts rearranging documents and creating spatial constructs including clusters [1,21]. With explicit clusters, users are able to externalize their semantics of the information into the workspace [19]. Clusters serve the purpose of synthesizing timelines, classifications, or just organizing thoughts in an external knowledge space [34]. In one interesting approach in augmented reality, a system showed potential for organizing clusters around physical objects in the real world [35]. In fact, humans are so familiar with the notion of clusters in their regular 2D applications that their ideas for clusters have transferred to the 3D applications as well. Despite having three dimensional space with six degrees of freedom, the users tend to consider two documents as part of a cluster when one document is in proximity to or overlaps with another document in the same plane [34].

Considering the common strategy of using clusters, we suggest that it would be beneficial for the system to take care of the mundane steps of creating a cluster, leaving users to be more concerned about tasks that demand more human involvement. There are numerous algorithms to identify clusters [22,29,36,40,41,48]. However, most of these algorithms are designed to create clusters from an already existing group of objects scattered in a workspace. Furthermore, they are not equipped with parameters to adapt their outputs to accommodate users interacting with the individual documents in the cluster. Hence, they are not fit for identifying clusters in a visual analytic process where the workspace keeps changing to reflect the user's mental model. There are visual analytic systems with interactive clusters [15,46] that mostly focus on gaining insights from an aggregated visualization of the dataset, and deprioritize the individual data contents, thus making them unsuitable for sensemaking tasks. Hence, finding alternative approaches to create cluster identification algorithms for visual analytic tools remains an open research problem.

3 Goals and Research Questions

Our research was designed to address how automation in clustering can help participants organize their immersive analytics workspace more effectively. That led us to ask three broad research questions.

RQ1: What is an appropriate level of automation for a clustering feature in an immersive sensemaking system?

We wanted to know what amount automation of the clustering process could facilitate users' analysis, rather than designing a feature with too much automation that gets in the way. We reasoned that a fully automated clustering feature could make the user lose control over their workspace. What if, instead of thinking of automation as the removal of human involvement, we imagined it as the selective inclusion of human participation? The result would be a process that harnesses the efficiency of intelligent automation while remaining amenable to human feedback, all while retaining a greater sense of meaning. We addressed this question through an iterative design process (Section 4.1) before we moved on to our formal experiment.

RQ2: How does an explicit clustering feature help analysts organize an immersive workspace?

We developed a set of clustering interactions such as creating, expanding, removing, and merging clusters. Our hypotheses are that having explicit clusters would make analysts faster in organizing their workspace (H2a), would speed up the process of reorganizing an existing workspace (H2b), and would make the final layouts of the workspace more understandable and less ambiguous (H2c).

RQ3: What are the benefits and challenges of having semi-automated clusters in an immersive visual analytic system?

We aim to understand how automating some (or all) of the cluster interactions may affect user performance and satisfaction compared to fully user-controlled or no cluster interactions. We hypothesize that participants would prefer having more control over the clusters rather than depending on the semi-automated technique (H3).

4 Design Process

To address RQ1, we performed an iterative design process in which we explored the design space for semi-automated cluster interactions for immersive analytics. We evaluated various clustering algorithms and integrated the chosen algorithm into an interactive user experience. We implemented a semi-automated cluster creation technique based on this design, along with two comparison techniques.

4.1 Design Evolution

Algorithm First, we needed a fast, iterative algorithm that could detect clusters of arbitrary size, shape, and density. Most traditional approaches [36, 41] to clustering rely on prior knowledge of the number of the clusters K. Since the layout emerges throughout the sensemaking process, there is no way to know in advance how many clusters there will be. In addition, we need an algorithm to form clusters on the fly as the user is creating the spatial layout, rather than one that looks at a complete spatial layout to find the clusters.

Prior studies have proposed density-based [22, 29], and graph-based clustering [40,48], which do not require prior knowledge of K. Despite showing excellent performance on detecting clusters with uniform data distribution, and ideal shapes, they fall short on clusters varying in size, shape, density, and noise (e.g., DBSCAN can only identify spherical clusters). As clusters formed during sensemaking process can vary significantly in size, shape, and density [33], these clustering methods cannot be applied. However, the Dirichlet Process Mixture Model (DPMM) has shown promising results in detecting arbitrarily shaped clusters with no prior knowledge of K [3, 26, 27, 39]. In DPMM, every time a new data point is included, it either joins an existing cluster, or starts a new cluster [9, 24]. The flexibility makes DPMM particularly promising for users iteratively browsing documents to create clusters.

User Experience Design We needed to understand when and how this algorithm could be embedded in an analyst's workflow of organizing documents in the immersive space. We launched an informal exploratory study among the authors to observe the effects of different approaches. We envisioned a *human-AI collaboration* where the users are able to focus on the analysis, and the algorithm just *augments* the process by providing helpful assistance [50,51,53].

Creating spatial clusters has been associated with the semantic interaction of 'document movement' [19]. Hence, there are two possible answers to the question of 'when': the system automatically applies the algorithm after every document movement, or the user explicitly triggers the algorithm after multiple document movements. However, the latter approach prompted a challenge: the results of the algorithm after moving several documents may not correspond to the user's expectations. As we observed in our exploratory study, misplacement of even a single document prompted the user to take a step back, review all the documents since the last clustering, and make sure every document was properly put in their intended cluster. The whole process was time-consuming, annoying, and frustrating. So, we resorted to applying the algorithm after every document movement, which reduced the probability of a mistake, and reduced the realignment cost (even if the algorithm gave an undesired result, the user could rearrange the document immediately).

A second design consideration relates to which items should be considered for automated clustering after every document movement. We started with a fully automated system where all the documents were given as input to the algorithm. This resulted in a lot of changes to the workspace after each movement, which made users confused, and the outputs did not always match the user intentions. Hence, the users spent more time on rearranging the documents than completing the task. After a series of design iterations, we concluded that the user gets frustrated particularly when an already-formed cluster is restructured by the system through the addition or removal of unwanted document(s) without their involvement. So, we designed our system to only consider the subset of documents that were not yet in any cluster, and create cluster(s) with only those documents. That left the users in control of other cluster-related interactions such as expanding, removing, and merging (Section 4.3).

Finally, we needed to determine how the clusters should be visualized. Typically, users of immersive analytic tools tend to create clusters on two-dimensional planes, placing those planes in various depths [32, 34], thus creating a 2.5-dimensional visualization. Following a similar strategy, we created a rectangular plane that had the closest distance to all the documents in a cluster, and moved all the documents to that plane. We also found the smallest possible translations required for the documents to not overlap each other while reflecting the layout specified by the user. We determined the height and width of the rectangles based on spatial positions of the documents on the outer edges, and made sure the rectangle held every document in the cluster while leaving a small margin on the borders so that users could easily grab and move the clusters.

4.2 Techniques

In order to understand the effects of the semi-automated clustering technique, we implemented a semi-automated technique (Section 4.2.1), and a technique that was fully user-controlled (Section 4.2.2). We also had a control technique without any cluster interactions (Section 4.2.3). In all three techniques, the users could select and move individual documents with a ray-casting interaction.

4.2.1 Proximity

This technique allowed the users to create explicit clusters in addition to interacting with the individual documents. After every document movement, the DPMM algorithm was applied to all the documents that were not part of any cluster. The algorithm returned the documents with the clusters they belonged to. Based on the output, the system created rectangular clusters for the documents

(Figure 1(a)&(b)). Essentially, from the user's perspective, each cluster contained documents that were in proximity to each other.

4.2.2 Overlap

In addition to all the individual document interactions, this condition allowed the users to create explicit clusters by overlapping two documents with each other. Any documents that touched each other were highlighted with a yellow border (Figure 1(c)&(d)). This visual feedback let the user know about the consequence of their action in advance, so that the user could move the document away again if they did not want the cluster to be created. Releasing the document while it was still highlighted created a rectangular cluster that would hold all the documents that were touching each other.

4.2.3 Freestyle

The user did not have any explicit clusters in this condition (Figure 1(e)). They could move around individual documents, create labels, and move around individual labels.

4.3 Additional Cluster Interactions

In addition to creating clusters, the system allowed the users to interact with the clusters in other ways.

Cluster Movement Similar to document movement, the users could grab a cluster, move it around, and place the cluster anywhere in the immersive environment.

Cluster Expansion/Reduction/Rearrangement Users could expand an existing cluster by adding new documents to it. New documents could be added by overlapping them with the cluster. Clusters lighted up whenever they touched another document (not in the cluster) providing the user a visual feedback on their action. The users could also grab a document already in the cluster, and move it away causing the cluster to shrink in size. The documents could also be rearranged inside the cluster by grabbing a document, and moving it to another position of the same cluster plane.

Cluster Merger The users could also merge two or more clusters together to create one big cluster. A user could merge clusters by grabbing a cluster and bringing it closer to another cluster so that they overlap each other. All the clusters that were touching another cluster would light up to provide visual feedback.

Cluster Removal The users could remove a cluster by moving documents away from the cluster one by one until there was just one document left, and the cluster was deleted. In the Proximity condition, the lone document would then be considered as an input for the next run of the algorithm.

5 EXPERIMENTAL DESIGN

In this section, we present the various aspects of the experiment that we conducted to find out how explicit clusters help the analysts (RQ2), and investigate the benefits and challenges of the semi-automated clustering technique (RQ3).

We conducted a within-subjects experiment to evaluate the effect of our independent variable, clustering technique, on dependent variables, including interaction velocity, cluster movement time, cluster size, and the cluster validation score.

5.1 Experimental Task

The goal of our experiment was to understand how the ability to create/manipulate explicit clusters, and the ability of the system to automatically create new clusters, affected performance, usability, and strategy during a simple sensemaking task. With that goal, we chose a document classification task with common recognizable items from three datasets: *Food*, *Animal*, and *Vehicle*, which is similar to a card sorting task [54]. Each set consisted of 30 cards, each of which showcased an image of an item accompanied by its name on the side (Figure 1(f)).

The images of the dataset were chosen with the criteria that they should be familiar to most people and easily distinguishable. Additionally, we wanted to observe the reorganization strategy of the users when they encountered new data items while they already had an organized workspace. Hence, we split the datasets into two subsets of 15 cards. When the users completed the organization of the first subset, we presented the second half of the dataset, which the users had to incorporate in their already existing workspace.

We wanted to encourage the participants to create a tidy workspace with meaningful sets of cards that could be presented to an audience. Therefore, we likened the workspace to an exhibition room, and gave the participants the role of a curator. The participants were instructed to group the cards into clusters in a way that made sense to them. They were told that the final layout should be an exhibition space for kids who can learn to recognize the items from the curated groups. The participants also had to come up with a relevant label for each of the groups. For each condition, the participants had ten minutes to curate the exhibition space with 30 cards.

5.2 Apparatus

We used a Varjo XR-3 head-worn display [1] running on a desktop PC with an Intel i9-9900k processor and an NVIDIA GeForce RTX 2080 Ti graphics card. User movement was tracked by a SteamVR 2.0 Lighthouse tracking system covering a four-by-eight meter space that we kept clear of obstacles. The application was implemented using Unity v2020.3.9. The DPMM algorithm was implemented using the *sklearn* library in Python v3.8, communicated with the Unity application via a local server-client port.

The user held one Valve Index controller [2] to interact with the documents. As one of the most effective selection technique [4, 33], we chose a ray-casting method with depth control for selection and manipulation of documents within the workspace. A virtual ray emanated from the controller, and the first document the ray intersected could be selected by pressing down the trigger. The selected document stuck to the ray as the user moved the controller around. The user could use the joystick to change the distance of the selected document from its initial position along the ray. Text input for creating labels was achieved by using a pass-through AR "desk portal" that allowed the user to view and interact with a keyboard on a tracked wheeled desk. This afforded the user the ability to put down the Index controller and then type on a keyboard as they might do at a traditional computer workstation.

5.3 Participants

We recruited 27 participants (10 females) with a minimum age of 20 years and maximum age of 39 years old ($\mu = 26, \sigma = 4$). Six participants had no prior VR or AR experience, while ten had only used VR or AR once or twice, and the rest used VR or AR more than twice. Two participants wore contact lenses, while three used glasses, and the remaining 22 had uncorrected vision. The experiment was approved by the university's institutional review board.

5.4 Procedure

We split our study into five phases: pre-study, training, main study, subjective assessment, and post-experiment interview. The training, main study, and subjective assessment phases were repeated for each of the three conditions, while the pre-study and post-experiment interview phases were conducted once per participant.

Pre-Study During the pre-study phase, we sent out a questionnaire along with the recruitment email to collect demographic information, understand their experience with VR/AR, and to schedule a time for the experiment. On the scheduled day, we presented the participant an informed consent form to read and sign. We briefly explained the goal of our experiment, and allowed them to get familiar with the physical space. This phase lasted five minutes on average.

[1] https://varjo.com/products/xr-3/

[2] https://store.steampowered.com/app/1059550/Valve_Index_Controllers/

Training In this within-subjects study, the participant experienced all three conditions. We counterbalanced the order of the conditions according to a Latin square to avoid bias for any particular condition. For each condition, we demonstrated how the clustering technique worked with two physical cards followed by a training session in VR. We used a set of ten cards for this session that were not in the main datasets. If it was the first condition, we started the training by showing the participant the boundaries of the tracked area, and teaching them how to interact with the cards with the controller. The participant also learned how to create labels with the keyboard seen through the "Desk Portal". We proceeded to show the participant each feature of the clustering technique, and allowed them to explore the environment and practice the controls. To help in their learning, we gave them a set of tasks which were designed so that the participant got to try out all the features of the clustering technique. The participants spent 5-7 minutes on this phase.

Main Study When the participant completed the given tasks, and was satisfied with their preparation, we launched the main study that involved the participant performing the experimental task described in Section 5.1 with the clustering technique they just learned. An experimenter was always present in the room to a) ensure the participant did not hit any physical obstacle, and b) provide the second half of the dataset as soon as they were done with the first half. The participants took an average of 8.44 minutes per technique for this phase.

Subjective Assessment After indicating that they had completed the main phase, the participants then would start the in-VR presentation. This involved an experimenter posing as the audience and asking the participant for a tour of the exhibition space. This prompted the participant to describe the clusters they formed, and explain any spatial relationships they used during the classification. Following the presentation, we helped the participant take off the headset. We presented them a NASA Task Load Index (TLX) questionnaire [25] for measuring the mental workload after experiencing the condition, and a System Usability Scale (SUS) questionnaire [28] to collect the subjective assessment of the usability of the condition. This phase took five to ten minutes to complete per technique.

Post-Experiment Interview Upon completion of all three conditions, we wanted to understand the participant's preference. We isolated seven different relevant scales from the NASA TLX and the SUS questionnaires, and asked the participants to rank the three conditions based on: ease of use, comfort, performance, learnability, usefulness, mental workload, and physical workload.

Finally, we conducted a semi-structured interview with a series of open-ended questions. The interview was designed to gather feedback about the three conditions from different perspectives. We asked the participants about their user experience, what features they liked, and what features they would add to improve the organization experience. We also asked which of the three conditions they would choose for their daily organization tasks (and why), what features were useful or were frustrating, and if they had any general comments. This phase took between ten to fifteen minutes.

5.5 Data Collection and Measures

We collected a variety of data in order to measure the participants' actions, preference, and the cluster outputs during the experiment.

We screen-recorded the main study (seen from the participant's POV) using the Varjo Desktop Application's recording feature. This allowed us to review what the participant was doing, and revisit the evolution of their workspace organization after the experiment. We also kept a log of all user interactions with the cards and clusters (with associated time-stamps) in an external file. This gave us an opportunity to analyze their action intents and results with precision. We also recorded the final positions of all the cards, clusters, and the labels that allowed us to see the final layout of the workspace, and compare how neat the workspace was for the three conditions.

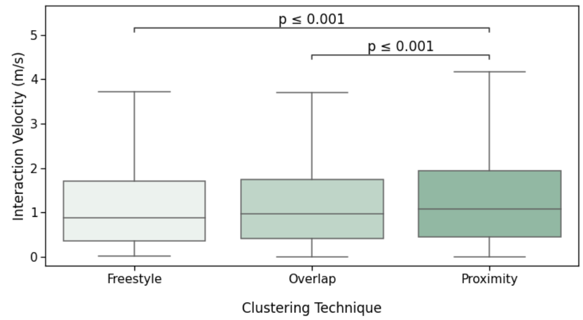

Figure 2: Interaction velocity in Proximity is higher than Freestyle and Overlap. Despite having clusters, interaction velocity in Overlap is not different from the interaction velocity in Freestyle.

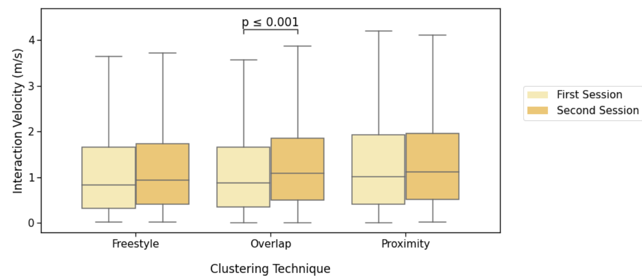

Figure 3: Interaction velocity changes in the second session. In Overlap, the users are faster in the second session.

We used the Qualtrics website to collect answers from the pre-study questionnaire, NASA TLX questionnaire, and the SUS questionnaire. The answers from the subjective assessment phase, and the post-experiment interview were recorded using Google Pixel's Recording app that also transcribed the audio. We reviewed the transcriptions for possible errors, and used them for further analysis.

6 Results

6.1 Quantitative Analysis

6.1.1 Interaction Velocity

We define *Interaction velocity* as the distance the participant moves a card per second, and we use this as a measure of interaction efficiency. A one-way ANOVA revealed a statistically significant effect of condition on interaction velocity ($F = 18.3$, $p \leq 0.001$). In post-hoc analysis with Bonferroni correction, we found very strong evidence (Figure 2) that the mean interaction velocity in Proximity is larger than both Freestyle ($p \leq 0.001$, cohen's d$= 0.3$), and Overlap ($p \leq 0.001$, cohen's d$= 0.2$). There was no significant difference in mean interaction velocity between Freestyle and Overlap .

We ran further analysis to understand how the participant behaviors changed over time during a condition. As mentioned in Section 5.1, each condition was split into two sessions. For each condition, we performed Student's t-test to understand how the interaction velocity changed in the second session compared to the first session (Figure 3). We found that interaction velocity significantly increased in the second session of Overlap compared to the first session ($t = 11.09$, $p \leq 0.001$, cohen's d$= 0.2$). There was no significant difference of interaction velocity between the two sessions for either Freestyle or Proximity.

6.1.2 Reorganizing the Workspace

We wanted to analyze the efficiency of the participants in the different conditions when they had to rearrange an existing workspace. We looked at the time taken to move entire clusters, and how it varied depending on the number of documents in the cluster.

Both Overlap and Proximity conditions had explicit clusters, thus

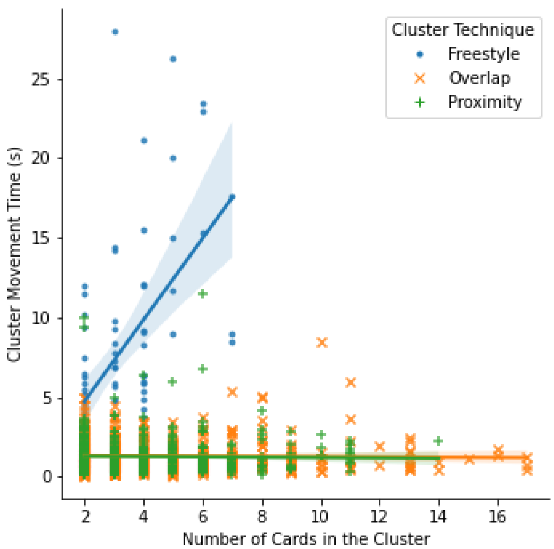

Figure 4: Cluster movement times in each condition, compared to number of cards in the cluster. The blue line is the regression line for Freestyle, while the blue shaded area is the confidence interval for the regression estimates. In Overlap and Proximity, the cluster movement time stays constant (orange and green lines respectively).

allowing us to analyze the cluster movement time from the log files. However, the Freestyle condition did not have any explicit clusters, making the analysis of the cluster movement time non-trivial. We analyzed this by reviewing the screen capture of the participant actions in the Freestyle condition. We identified a group of cards as a cluster if a) the participant created a label close to the group of cards, or b) the intra-card distance was smaller than the distance to all the adjacent groups of cards. The cluster identification was verified by the in-VR presentation of the workspace by the participant at the end of their session. We considered an action as a cluster movement if the participant moved two or more documents of the same cluster from one place to another through two or more consecutive actions.

We plotted the cluster movement times against the number of cards in each cluster (Figure 4). Pearson's correlation test revealed that the cluster movement time in Freestyle is positively linearly correlated with the number of cards in the clusters ($coeff_{freestyle} = 0.56$), while in Overlap and Proximity, the cluster movement time has no significant correlation with the number of cards in the cluster ($coeff_{overlap} = -0.023$, $coeff_{proximity} = -0.017$).

6.1.3 Cluster Size

As we saw in Figure 4, the participants tended to create larger clusters in Proximity and Overlap than in Freestyle. This prompted us to analyze the effect of condition on average cluster size which was determined by the number of cards in the cluster.

A one-way ANOVA revealed that there was a statistically significant effect of condition on the cluster size ($F = 10.34$, $p \leq 0.001$). In post-hoc analysis with Bonferroni correction, we found evidence (Figure 5) that the mean cluster size of both Overlap and Proximity are significantly higher than the mean cluster size of the Freestyle condition (between Freestyle and Overlap: $p \leq 0.001$, cohen's d $= 0.5$, between Freestyle and Proximity: $p \leq 0.05$, cohen's d$= 0.4$).

6.1.4 Workspace Neatness

To calculate the neatness of the workspace quantitatively, we measured the Silhouette Scores (SS) [45] of each of the participants for the three conditions. SS is a measure of how similar an object is to its cluster compared to other clusters. SS is computed using the

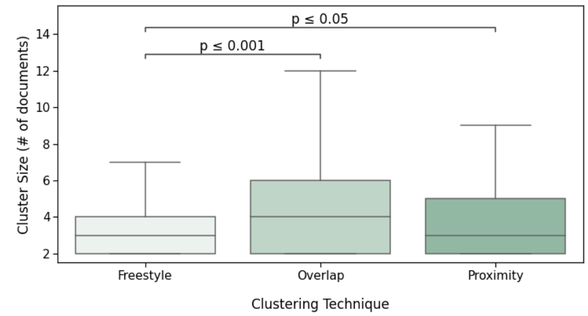

Figure 5: Average cluster size in Freestyle is lower than Overlap and Proximity.

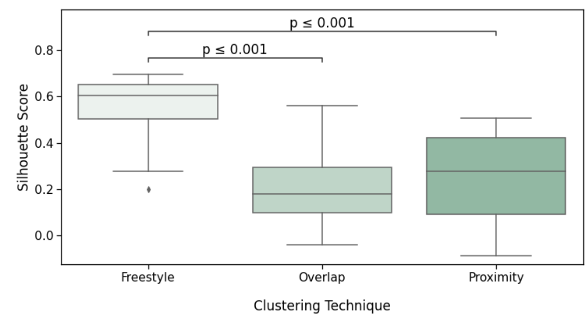

Figure 6: Silhouette Score in Freestyle is higher than the Silhouette Scores in Overlap and Proximity.

mean intra-cluster distance, and the mean nearest-cluster distance for each sample. A high value (max=1) of the SS would mean that the clusters are well-defined while a low value of zero would indicate that the clusters are hardly distinguishable in the workspace.

We found that condition had a statistically significant effect on SS. Post-hoc analysis with Bonferroni correction revealed (Figure 6) that the mean SS of both Overlap and Proximity are significantly lower than the mean SS of the Freestyle condition (between Freestyle and Overlap: $p \leq 0.001$, cohen's d$= 2.2$, between Freestyle and Proximity: $p \leq 0.001$, cohen's d$= 1.8$).

6.1.5 Workload and Usability

We found no significant difference in the NASA TLX Scores among the three conditions, nor did we find any significant difference in the System Usability Scales [11] among the three conditions. However, the average SUS scores for all three conditions were higher than 85, which is considered "Excellent" [5].

6.2 Qualitative Analysis

For qualitative analysis, we went through the answers in the post-experiment interview. We started by transcribing the interviews for each participant, and putting them all in the same document. First, we generated initial codes for each response. We found common themes among the codes, and proceeded with defining and naming them. We went through the initial set, and looked for similar themes that could be combined, leading us to a smaller set of themes. After a series of iteration, we ended up with three major themes describing factors participants felt most influenced their experience with clusters in the experiment: Convenience, Control, and Creativity.

6.2.1 Convenience

The participants preferred their system to be easy, fast, and convenient to make sure not much thought was required for organizing the clusters. The participants found that although the Freestyle condition was the easiest to learn among the three, having explicit clusters definitely made the organization part of the task more convenient.

They said it was faster, and it required less awareness of the cards in the clusters individually, thus reducing *cognitive effort*:

> "Definite clusters reduce effort to manually organize the images in a way that defines a group." (P019)

All but three of the participants preferred one of the conditions with explicit clusters (Proximity or Overlap), rather than the no-cluster condition (Freestyle). 22 out of 27 participants (81.5%) specifically called out the explicit clusters to be a useful feature, particularly when they had to reorganize their workspace:

> "I can move all of them as a group to wherever you like. Also, I can remove one card without breaking the entire group. That was the most convenient feature." (P016)

However, the comparison of the two conditions with explicit clusters was not as straightforward. While some participants liked the Overlap condition for having full control over the clusters, more participants (20 out of 27) preferred the Proximity condition specifically because of its similarity with the Freestyle condition in terms of simplicity. Participant P014 put it this way:

> "It [Proximity] was as easy as Freestyle, with the added benefits of having clusters."

Also, the participants did not like overlapping cards as "you have to consciously make an effort to bring the images very close to each other, and see the color changes to make a cluster" (P003).

6.2.2 Control

The desire to have control over the workspace divided the participants into two factions. One faction (18 out of 27 participants) thought having control over every aspect of the clusters was distracting, and preferred to have the system take care of the clusters while they focused more on completing the task. They chose the Proximity condition over the Overlap condition because of its assistive feature:

> "As clustering is kind of auto, I don't have to think about whether it's cluster or not. That saves my time." (P001)

Participants also shared how the Proximity condition offered better performance by making them faster in their organization task:

> "As long as the cards were close enough, it grouped itself. I was able to organize them much quicker." (P010)

The rest of the participants were in favor of the Overlap condition because it offered full control over their workspace. Since there was a definite action to create clusters (overlap), the participants were always aware of the newly developed clusters prior to their existence. This gave them a sense of control over their workspace. There were three participants who were frustrated with the Proximity condition as they were losing control over their workspace because of unwanted clusters. According to Participant P002,

> "Sometimes [Proximity] would create cluster by itself even though I did not want it to ... I would have to put more effort to put them far enough so that they do not create a cluster."

Furthermore, the visual feedback on their action (highlighting the overlapping documents) helped them to see into the future before making their final decision. That enabled them to make split second decision changes without causing major updates to the workspace:

> "It's only easier to Overlap than bringing stuff closer and just *hoping* them to make a cluster." (P015)

6.2.3 Creativity

While most of the participants liked having clusters, there were three participants who preferred the Freestyle condition over the other two. All three of them reiterated the necessity of having creativity in their organization that they could not achieve if there were explicit clusters. They wished to have control over the three-dimensional positions of the documents in the same cluster to develop clusters with various size and shape. Participant P004 said,

> "Freestyle was the quickest to adapt and also the way to be the most creative."

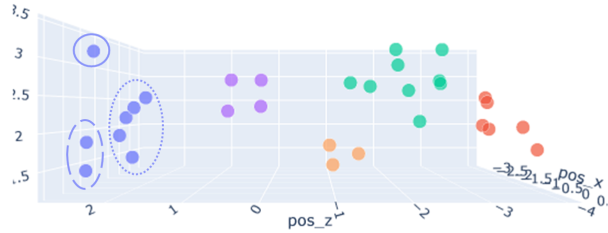

Figure 7: Final layout of participant P012 who preferred the Freestyle condition. They created sub-clusters (marked by dotted and solid circles) inside a bigger cluster (left-most).

For example, Figure 7 shows the final layout of a participant who preferred Freestyle for the creativity it allowed. They created five clusters. In the left-most cluster, which was labelled "*run on ground*" by the participant, they made subdivisions inside the cluster. They kept regular ground transports (bus, taxi, rickshaw, bicycle, go-cart) in one sub-cluster, while keeping the two ground transports used in construction (crane, tractor) in another sub-cluster. The sole document on the top represents a military ground transport (Humvee) that the participant kept separate from the others.

6.3 Suggested Features

Finally, participants presented several additional features that they would want in a clustering tool. First, in addition to creating clusters, participants wanted the system to automatically create or suggest labels for the cluster based on its contents. Second, even in this simple classification task, participants wanted to create subgroups inside each group. Third, participants wanted to have auditory feedback for when clusters were created, expanded, or deleted.

7 Validation Study

Although our findings from analyzing the interaction velocity supported our hypothesis regarding the clusters making the participants faster (H2a), they went against our hypothesis regarding the participants preferring having more control over the clusters (H3). This prompted us to revisit our techniques from a critical perspective. We noticed that even though the aggregated results showed the Overlap technique to make the participants slower, they were a lot faster by the end of their task compared to when they started. Further investigation into the video screenings of the participants' sessions revealed that many participants were actually having issues to meet the criterion of Overlap to create clusters. They ended up performing consecutive interactions with the same document over a longer period of time, thus, making them slower in their overall classification task. In their interviews, they mentioned that they were struggling at first because they were having trouble with the ray-casting technique to bring the cards to an overlapping position with another card. As evidenced by prior studies, ray-casting does not afford rotation of a card in place since rotating the controller also causes translation of the card at the end of the ray [10, 31]. The farther the card is, the worse the effect becomes. In addition, the joystick to change distance between the card and the participant was too sensitive, thus, making it harder to cause a card to align with another. With every push or pull of the joystick the card would either go behind or come to the front of another card instead of overlapping it.

We hypothesized that these issues could be the deciding factors for the participants to prefer Proximity over Overlap. Therefore, we replaced the selection method with HOMER [10], a technique where after selecting a card with a ray, instead of the card becoming attached to the end of the ray, the virtual hand moves to the card position and is attached to the card. Thus, HOMER uses the metaphor where the user is grabbing the card with their own hand regardless of the card's distance from the user, enabling them to rotate the card in place by rotating the physical hand. Upon releasing the card, the virtual hand would return to the position of the physical hand.

With the HOMER modification, we ran a small within-subjects validation study with four participants and two conditions: Overlap and Proximity (since we wanted to see if HOMER improved the results for Overlap). The results from this validation study turned out to be comparable with the original experiment. We found that participants still struggled in the first session of Overlap, and they got faster by the end of the second session. There was still no effect of session on interaction velocity for Proximity. Two participants preferred Overlap because of having more control, and two participants chose Proximity because of its ease of use. Since there was no overwhelming evidence that the change to HOMER changed our results, we concluded that the ray-casting method was not the primary factor for the participants to prefer the semi-automated technique in the original experiment. Rather, they liked the Proximity technique for its intuitive automated assistance in creating clusters.

8 Discussion

Even though we started our study with a motivation to automate the clustering process, our design exploration revealed how a fully automated system could do more harm than good which aligns with findings from prior works [17]. Users were frustrated, annoyed, and spending more time on fixing the automated outputs rather than progressing with their analysis. Through an iterative co-design process with users, we were able to reduce the automation such that the system had control over only the creation of clusters, while users took over the other aspects of cluster interaction. Essentially, we ended up employing human-in-the-loop design in cluster interaction, making it a semi-automated clustering technique. Although we found an appropriate automation level, there were still some issues with automation having undesired effects.

In our experiment, we found that Proximity made interaction faster than the other techniques (partially supporting H2a), and the participants preferred it over Overlap because of it being easier to use, even though both had explicit clusters. However, upon further examination, the analysis shows that the participants were slower in Overlap only for the first session. By the end of the second session, they became as fast in Overlap as in Proximity. This suggests that Overlap required some learning, but was not inherently slower.

We also found evidence to support our hypothesis (H2b) that explicit clusters made it easier for participants to complete their task, particularly when they had to reorganize an already existing workspace. We showed, and the participants reiterated, that the moving time for clusters in the Overlap and Proximity techniques was faster, and independent of the number of documents, while cluster moving time in the Freestyle technique increased linearly with the number of cards in the cluster. However, the participants could reach a faster moving time with a multi-object selection feature in the Freestyle technique. However, finding a standard multi-object selection technique is still an open research problem [12,30,38], thus, making it out of our study's scope. In addition, as the participants found it easier to make changes with explicit clusters, they ended up creating larger clusters when they had explicit clusters, which allowed them to move many documents at the same time.

However, the clusters did not make the workspace more understandable and less ambiguous from a quantitative point of view. Silhouette Scores (SS) for Overlap and Proximity tended to be close to zero, indicating that there were actually overlaps among the clusters, while SS for Freestyle was closer to one, implying that the clusters formed in that condition were more distinct, contrary to our hypothesis (H2c). Participants in the IST have showed similar trends in prior studies [34] where they created dense spatial layouts to organize their mental models. However, our video analysis reveals that despite having overlaps in the clusters, the users considered the workspace more presentable, as the explicit borders around each of the clusters helped them distinguish one cluster from the other. On the other hand, in Freestyle, the participants had to depend solely on the inter-cluster distance to keep them distinct. We believe this finding has a greater implication in working with larger datasets as it shows that having explicit clusters allows the users to require less space, and yet have distinguishable layout.

Between the two conditions with explicit clusters, the participants preferred the semi-automated Proximity technique. Even though they acknowledged that the Overlap technique allowed them to have more control over their layout, they liked Proximity because of its simplicity and ease of use, which contradicts our hypothesis (H3). Participants found it convenient to create clusters in collaboration with an AI which made their spatial organization process *"as easy as freestyle with the added benefits of the explicit clusters"*.

Finally, our experiment extracted the desired features that users want in an immersive clustering tool. The Proximity technique showed promise with its ease of use, and the ability to make the users faster. The participants liked Overlap for its control over the workspace, and instinctive visual feedback for cluster formation. Even the Freestyle technique was preferred by some of the participants because of the freedom it offers. An ideal clustering technique should have the convenience of automatic clustering, but also give the users a sense of control with meaningful visual feedback. Once the clusters are formed, the size, shape, and the structure of the clusters should be open to customization.

9 Limitations

We kept our task simple so that the users could complete the classification task easily with the clustering technique they were provided. This allowed us to disregard the complexities around navigating complex datasets, and narrow our focus on comparing the users' perception of clustering techniques. Future studies need to design a task involving more than simple classification, and using a more complex dataset involving large textual documents, if we want to have a more ecologically valid evaluation of clustering techniques for sensemaking tasks in immersive visual analytic tools.

10 Conclusions and Future Work

In this paper, we presented the results of our investigation into how automating clustering techniques can help analysts working with an immersive visual analytic tool to organize documents during a sensemaking task. We found that users are not comfortable with fully automated systems, as they can tend to deviate from the user intentions. We proposed a semi-automated clustering technique that proved to make users faster, and the users found it more convenient to create spatial structures in collaboration with an AI. This suggests there is significant potential for other intelligent assistance features in complex immersive analytic workflows. We showed how the semi-automated approach can be improved by adding visual feedback, and affording more creativity to the users. In addition, we found that explicit clusters made the final layouts formed by the users more presentable, requiring less space, and clusters allowed the users to reorganize their workspace in a faster fashion independent of the document count in the cluster.

One of the future directions of this research will be investigating the effects of the clustering techniques on a more complex task involving textual datasets. We will also explore additional intelligent features for immersive sensemaking, such as auto-labelling clusters, searching for new similar documents from a larger data space, and updating the level of details of the clusters based on the user's needs.

Acknowledgments

This work was supported in part by NSF I/UCRC CNS-1822080 via the NSF Center for Space, High-performance, and Resilient Computing (SHREC).

References

[1] C. Andrews, A. Endert, and C. North. Space to think: large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 55–64, 2010.

[2] C. Andrews and C. North. Analyst's workspace: An embodied sensemaking environment for large, high-resolution displays. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 123–131. IEEE, 2012.

[3] C. E. Antoniak. Mixtures of dirichlet processes with applications to bayesian nonparametric problems. *The annals of statistics*, pp. 1152–1174, 1974.

[4] F. Argelaguet and C. Andujar. A survey of 3d object selection techniques for virtual environments. *Computers & Graphics*, 37(3):121–136, 2013.

[5] A. Bangor, P. Kortum, and J. Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.

[6] L. Besançon, A. Ynnerman, D. F. Keefe, L. Yu, and T. Isenberg. The state of the art of spatial interfaces for 3d visualization. In *Computer Graphics Forum*, vol. 40, pp. 293–326. Wiley Online Library, 2021.

[7] S. K. Bhavnani. Domain-specific search strategies for the effective retrieval of healthcare and shopping information. In *CHI'02 extended abstracts on human factors in computing systems*, pp. 610–611, 2002.

[8] S. K. Bhavnani, R. T. Jacob, J. Nardine, and F. A. Peck. Exploring the distribution of online healthcare information. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems*, pp. 816–817, 2003.

[9] D. Blackwell and J. B. MacQueen. Ferguson distributions via pólya urn schemes. *The annals of statistics*, 1(2):353–355, 1973.

[10] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pp. 35–ff, 1997.

[11] J. Brooke. Sus: a "quick and dirty'usability. *Usability evaluation in industry*, 189(3), 1996.

[12] J. Cashion, C. Wingrave, and J. J. LaViola Jr. Dense and dynamic 3d selection for game-based virtual environments. *IEEE transactions on visualization and computer graphics*, 18(4):634–642, 2012.

[13] R. Chang, M. Ghoniem, R. Kosara, W. Ribarsky, J. Yang, E. Suma, C. Ziemkiewicz, D. Kern, and A. Sudjianto. Wirevis: Visualization of categorical, time-varying data from financial transactions. In *2007 IEEE symposium on visual analytics science and technology*, pp. 155–162. IEEE, 2007.

[14] K. A. Cook and J. J. Thomas. Illuminating the path: The research and development agenda for visual analytics. Technical report, Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2005.

[15] S. Das, B. Saket, B. C. Kwon, and A. Endert. Geono-cluster: Interactive visual cluster analysis for biologists. *IEEE Transactions on Visualization and Computer Graphics*, 27(12):4401–4412, 2020.

[16] W. Dou, D. H. Jeong, F. Stukes, W. Ribarsky, H. R. Lipford, and R. Chang. Recovering reasoning processes from user interactions. *IEEE computer graphics and applications*, 29(3):52–61, 2009.

[17] S. M. Drucker, D. Fisher, and S. Basu. Helping users sort faster with adaptive machine learning recommendations. In *IFIP Conference on Human-Computer Interaction*, pp. 187–203. Springer, 2011.

[18] A. Endert, P. Fiaux, and C. North. Semantic interaction for sensemaking: inferring analytical reasoning for model steering. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2879–2888, 2012.

[19] A. Endert, P. Fiaux, and C. North. Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 473–482, 2012.

[20] A. Endert, P. Fiaux, C. North, et al. Unifying the sensemaking loop with semantic interaction. In *IEEE Workshop on Interactive Visual Text Analytics for Decision Making at VisWeek 2011*, 2011.

[21] A. Endert, S. Fox, D. Maiti, S. Leman, and C. North. The semantics of clustering: analysis of user-generated spatializations of text documents. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 555–562, 2012.

[22] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, vol. 96, pp. 226–231, 1996.

[23] V. Estivill-Castro. Why so many clustering algorithms: a position paper. *ACM SIGKDD explorations newsletter*, 4(1):65–75, 2002.

[24] T. S. Ferguson. A bayesian analysis of some nonparametric problems. *The annals of statistics*, pp. 209–230, 1973.

[25] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988.

[26] R. Huang, G. Yu, Z. Wang, J. Zhang, and L. Shi. Dirichlet process mixture model for document clustering with feature partition. *IEEE Transactions on knowledge and data engineering*, 25(8):1748–1759, 2012.

[27] H. Ishwaran and L. F. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173, 2001.

[28] P. W. Jordan, B. Thomas, I. L. McClelland, and B. Weerdmeester. *Usability evaluation in industry*. CRC Press, 1996.

[29] D. Kellner, J. Klappstein, and K. Dietmayer. Grid-based dbscan for clustering extended objects in radar data. In *2012 IEEE Intelligent Vehicles Symposium*, pp. 365–370. IEEE, 2012.

[30] R. Kopper, F. Bacim, and D. A. Bowman. Rapid and accurate 3d selection by progressive refinement. In *2011 IEEE symposium on 3D user interfaces (3DUI)*, pp. 67–74. IEEE, 2011.

[31] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017.

[32] J. H. Lee, D. Ma, H. Cho, and S.-H. Bae. Post-post-it: A spatial ideation system in vr for overcoming limitations of physical post-it notes. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–7, 2021.

[33] L. Lisle, X. Chen, J. E. Gitre, C. North, and D. A. Bowman. Evaluating the benefits of the immersive space to think. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 331–337. IEEE, 2020.

[34] L. Lisle, K. Davidson, E. J. Gitre, C. North, and D. A. Bowman. Sensemaking strategies with immersive space to think. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 529–537. IEEE, 2021.

[35] W. Luo, A. Lehmann, H. Widengren, and R. Dachselt. Where should we put it? layout and placement strategies of documents in augmented reality for collaborative sensemaking. In *CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2022.

[36] J. MacQueen. Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability*, pp. 281–297, 1967.

[37] R. Mander, G. Salomon, and Y. Y. Wong. A "pile" metaphor for supporting casual organization of information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 627–634, 1992.

[38] R. A. Montano-Murillo, C. Nguyen, R. H. Kazi, S. Subramanian, S. DiVerdi, and D. Martinez-Plasencia. Slicing-volume: Hybrid 3d/2d multi-target selection technique for dense virtual environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 53–62. IEEE, 2020.

[39] R. M. Neal. Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2):249–265, 2000.

[40] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14, 2001.

[41] K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell. Text classification from labeled and unlabeled documents using em. *Machine learning*, 39(2):103–134, 2000.

[42] M. Olaosebikan, C. Aranda Barrios, L. Cowen, and O. Shaer. Embodied notes: A cognitive support tool for remote scientific collaboration in vr. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 1–6, 2022.

[43] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, vol. 5, pp. 2–4. McLean, VA, USA, 2005.

[44] D. E. Rose, R. Mander, T. Oren, D. B. Ponceleon, G. Salomon, and Y. Y. Wong. Content awareness in a file system interface: implementing the "pile" metaphor for organizing information. In *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 260–269, 1993.

[45] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[46] D. Sacha, M. Kraus, J. Bernard, M. Behrisch, T. Schreck, Y. Asano, and D. A. Keim. Somflow: Guided exploratory cluster analysis with self-organizing maps and analytic provenance. *IEEE transactions on visualization and computer graphics*, 24(1):120–130, 2017.

[47] K. A. Satriadi, B. Ens, M. Cordeil, T. Czauderna, and B. Jenny. Maps around me: 3d multiview layouts in immersive spaces. *Proceedings of the ACM on Human-Computer Interaction*, 4(ISS):1–20, 2020.

[48] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.

[49] J. Stasko, C. Görg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization*, 7(2):118–132, 2008.

[50] H. Strobelt, J. Kinley, R. Krueger, J. Beyer, H. Pfister, and A. M. Rush. Genni: Human-ai collaboration for data-backed text generation. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1106–1116, 2021.

[51] D. Wang, J. D. Weisz, M. Muller, P. Ram, W. Geyer, C. Dugan, Y. Tausczik, H. Samulowitz, and A. Gray. Human-ai collaboration in data science: Exploring data scientists' perceptions of automated ai. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–24, 2019.

[52] K. E. Weick, K. M. Sutcliffe, and D. Obstfeld. Organizing and the process of sensemaking. *Organization science*, 16(4):409–421, 2005.

[53] J. D. Weisz, M. Muller, S. Houde, J. Richards, S. I. Ross, F. Martinez, M. Agarwal, and K. Talamadupula. Perfection not required? human-ai partnerships in code translation. In *26th International Conference on Intelligent User Interfaces*, pp. 402–412, 2021.

[54] J. Wenskovitch and C. North. An examination of grouping and spatial organization tasks for high-dimensional data exploration. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1742–1752, 2020.

[55] S. Whittaker and J. Hirschberg. The character, value, and management of personal paper archives. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 8(2):150–170, 2001.

[56] W. Wright, D. Schroh, P. Proulx, A. Skaburskis, and B. Cort. The sandbox for analysis: concepts and methods. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 801–810, 2006.