Immersive Analytics Sensemaking on Different Platforms

Sebastian Blum
Technische Universität Berlin
Berlin, Germany
s.blum@campus.tu-berlin.de

Gokhan Cetin
SIAT, Simon Fraser University
Vancouver, Canada
gcetin@sfu.ca

Wolfgang Stuerzlinger
SIAT, Simon Fraser University
Vancouver, Canada
w.s@sfu.ca

Abstract

In this work we investigated sensemaking activities on different immersive platforms. We observed user s during a classification task on a very large wall-display system (experiment I) and in a modern Virtual Reality headset (experiment II). In experiment II, we also evaluated a condition with a VR headset with an extended field of view, through a sparse peripheral display. We evaluated the results across the two studies by analyzing quantitative and qualitative data, such as task completion time, number of classifications, followed strategies, and shape of clusters. The results showed differences in user behaviors between the different immersive platforms, i.e., the very large display wall and the VR headset. Even though quantitative data showed no significant differences, qualitatively, users used additional strategies on the wall-display, which hints at a deeper level of sensemaking compared to a VR Headset. The qualitative and quantitative results of the comparison between VR Headsets do not indicate that users perform differently with a VR Headset with an extended field of view.

Keywords

Visual analytics, large high-resolution displays, sparse peripheral displays, virtual reality, head-mounted displays, spatialization, clustering, visualization, sensemaking, immersive analytics

1. Introduction

1.1. Visual Analytics

Visual analytics (VA) is a science-based activity supporting sensemaking for large, complex datasets through interactive visual data exploration [1]. In VA, users reason and make sense of the data through interaction with visualizations of the data. In the past, this concept gave rise to products such as Tableau, Microsoft Power BI, QlikView, and others.

In most VA applications, users interact indirectly with data through widgets, such as sliders and menus, which control the visualization through modifying the underlying model parameters. In contrast, semantic interaction enables analysts to spatially interact with their visualizations directly within the visual metaphor, using interactions that derive from their analytic process [2]. Further, semantic interaction supports sensemaking better than indirect interaction [3]. In a spatial workspace, users can directly arrange documents spatially into clusters to convey similarity or relationships in the data [4]. Spatial workspaces also allow users to establish implicit relationships in large datasets [5].

The space limitations of normal desktop displays lead to a need to remove/reduce/hide (at least temporarily) existing visualizations to make room for new ones, requiring an additional cognitive effort from the users to remember the now invisible information. By radically increasing display size, this dynamic could change substantially and allow users to visually access more information at once. Comparisons can then be made visually and directly rather than relying on memory and imperfect mental models – which supports the usability principle of recognition over recall. With desktop monitors, we often face a tradeoff between the level of detail and the number of different objects that can be displayed. New forms of displays, such as large display surfaces or virtual reality environments, can thus potentially improve the efficiency of VA activities. On a large display, a flick of the eye or turn of the head is all that is required to consult a different data source [6]. Previous work has found that large, high-resolution displays (LHRDs) improve productivity over traditional desktop monitors [6–8]. We can expect this to hold for VA applications on large displays as well. The exploration of different user interface technologies for data visualization applications has never been a core topic in information visualization [9].

The combination of Virtual Reality (VR) technologies, 3D user interfaces and VA systems is a new approach to analyzing large or complex datasets. This research field is nowadays described by the term Immersive Analytics [9,10]. Working with VR systems in a professional environment and using these for data visualization or as an immersive analytical workspace provides "*an easy and natural path to collaborative data visualization and exploration*" [11]] [p. 609] and shows possibilities to "*maximize intrinsic human pattern recognition*". Previous research has already shown that VR can support insight discovery in (primarily) spatial application domains and in helping to more effectively investigate brain tumors [12], MRI results [13], shape perception [14], underground cave structure analysis [15], geo-scientific [16,17], or

paleontology questions [18]. Thus, we expect LHRDs and VR systems to improve VA activities. The work reported here aims to compare these two approaches through user studies.

1.2. Sensemaking in Immersive Environments

How analysts make sense of a given data set is a crucial part of their work. The way humans understand and process information in VA activities is well described by the *Sensemaking Loop* [19]. This sensemaking loop breaks the process down into several stages, as illustrated in figure 1. For Immersive Analytics, i.e., doing VA activities in immersive environments, previous research shows how each stage of the sensemaking loop might be improved or impaired by the capabilities or limitations of the system [20].
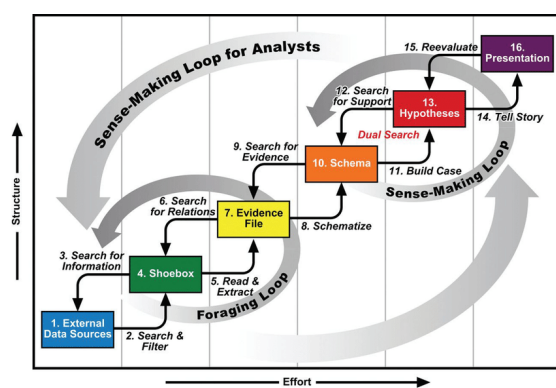

Figure 1: The sensemaking loop [19].

As seen in figure 1, the sensemaking loop involves a series of iterative steps for creating and evaluating a model for the data. Creating a model (bottom-up) involves finding information, extracting meaning, schematizing and building a case, and subsequently communicating that information. On the other hand, evaluating the model (top-down) involves re-evaluation, finding supporting evidence, finding relations in the information, or even finding basic information itself [19].

An integral part of the sensemaking process is foraging for information. To support such foraging, we could use represent data sources similarly to icons in an immersive environment, using either VR headsets or LHRDs. However in immersive environments, we can equip these representations with additional semantic meaning - such that the model of an engine might serve as a gateway for information about emissions, maintenance, power output, or other factors. One limitation of this approach is that there is much content that is not easily represented by icons, such as the cleaning budget for a department that sells cleaning products.

The second part of the sensemaking process is about synthesizing information, formulating hypotheses, and arranging supporting and contradictory evidence. Display space can play an important role in this process and assist in task completion, e.g., through the use of larger amounts of space for organization [6]. Further, the larger space offered by immersive systems provides a physical instantiation of the mnemonic device 'memory palace'.

Thus, different parts of a complex model could be compartmentalized to different (virtual) spatial locations. Analysts can then use the space not only for their collected information but also to organize and structure their analytical workflow and thought processes [20]. In this paper we focus on this aspect of sensemaking.

Some VA tasks require specific types of interaction methods, e.g., the method to select data during foraging might be different from that to express a hypothesis. Good user interfaces for immersive environments are subject to specific guidelines and concepts. Elmqvist's fluid interaction concept distinguishes different interaction types for selection, filter, sort, navigation, reconfiguration and labeling and annotating in VA tasks [21]. For example, techniques such as mouse interaction are not appropriate in immersive environments where the user is standing and not sitting in front of a display. Gestural interaction, e.g., pointing to and circling one or more elements, could be used instead. For such interaction, designing comfortable gestures is necessary to minimize the strain on the user.

As discussed above, there is great potential for integrating the VA process into immersive environments. Thus, we need to examine the specific components of the sensemaking process that can be enhanced by immersive technologies and use the advantages of such technologies to support the VA sensemaking process. Here we focus on the ability of the user to arrange the visualizations of data on a large display canvas.

Grouping the information and generating clusters of related data is a part of sense-making in analytical tasks, which helps in the search for a way to encode data in a visual representation that helps to answer task-specific questions [22]. It takes place in the early, preparation stages of VA processes and its effectiveness can affect the efficiency of all following stages.

A core inspiration for this study was Endert et al.'s work [23], which presented the concept of semantic interaction that seeks to enable analysts to spatially interact within their analytical workspace. Following their work, we also believe it is important to observe how users organize their information, unaided by any algorithm, as this reveals insights about the human ability to understand large amounts of information through an interactive system. Thus, we do not use algorithmic clustering.

1.3. Large Display Systems

Large display systems are both qualitatively and quantitatively different from traditional displays. Reda et al. presented the results of a small-scale study to understand how display size and resolution affect insight. Although their results verify the generally accepted benefits of large displays, they also provide mixed results for extant work and propose explanations by considering the cognitive and interaction costs associated with visual exploration [24]. Other studies suggest that users working with large displays became less reliant on wayfinding aids in acquiring spatial knowledge. For example, Ni et al.'s experimental findings demonstrated the advantages of increased size and resolution [25]. As a general guideline, a LHRD was the preferred choice for IRVE applications,

since it facilitates both spatial navigation and information gathering.

Our research uses multiple immersive display systems. One of them is a LHRD system called V4-SPACE. We used it to investigate how LHRDs can support VA tasks for a single user (figure 2). During this experiment we observed how physical size, resolution, and content spatialization on the workspace affect user performance on LHRD in VA tasks. A detailed description of V4-SPACE is given in the next chapter.
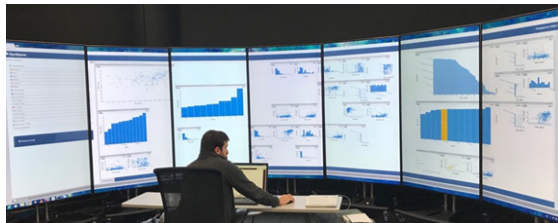

Figure 2: A user operating V4-SPACE

1.4. Head-Mounted Displays and Sparse Peripheral Displays

Extending VA system to be used within Virtual Reality (VR) systems poses challenges and opportunities. On the one hand, VR opens up new possibilities for data visualization, as virtual environments fully surround the user and the interface of the Human Machine Interface can be perceived to be more natural and intuitive. Thus, content can be placed and worked on all around the user. However, immersive VR technologies suffer from some restrictions, such as limited resolution, limited field-of-view, motion sickness, issues with interaction and gesture recognition, and with positioning objects in 3D space [9]. Even though current research in Immersive Analytics focuses primarily on low-level challenges [9], the above-mentioned obstacles still affect our research.

To address some of the above-mentioned issues of VR systems, we use a Sparse Peripheral Display HMD for our immersive VA study. The binocular field-of-view (FOV) of a standard VR HMD system is today about 84 degrees horizontal (110 degrees diagonal). By using a Sparse Peripheral Display (SPD), we can increase the field-of-view much by showing (low-resolution) content in the periphery of an HMD [26]. Increasing the FOV increases the spatial awareness of the user and might potentially even reduce simulator sickness.

A number of approaches and different systems to broaden the FOV through complex optical methods have been proposed. Fresnel lenses, such as the ones used in the StarVR and Wearality Sky, are challenging to manufacture and introduce optical distortions which are hard to mitigate [26]. Mirror based approaches [27] significantly increase the weight of their HMD devices.

In our study, we rely on previous research on SPDs [26] [28]. Xiao & Benko [26] presented a SPD that uses a lightweight, low-resolution and inexpensively produced array of LEDs, designed to surround the central high-resolution display of a Oculus Rift DK2 HMD. Their SPD expands the available FOV up to 190° horizontal. This SPD nearly fills the whole human FOV, which can span up to

210° giving the user a better perception of the virtual content that surrounds them (figure 3). Hashemian et al. [28] evaluated an improved SPD version in a HTC Vive setup, extending the Vives' FOV to 180° horizontally. They evaluated the performance of the SPD HMD in a spatial navigation study that used a previously presented navigational search paradigm [29,30]. Their results show that SPDs can provide a more natural experience of human spatial locomotion in virtual environments.

Figure 3: HTC Vive with a Sparse Peripheral Display extending the FOV to 180°.

2. A comparison of sense making behaviors in LHRD and HMDs

Based on knowledge about SPDs [26,28], large display systems [6,8] and immersive environments [9–11], combined with results from research about sensemaking [19,22] and productivity in VA [6,8,11,31,32], we performed two experiments in immersive analytics using different platforms. In Experiment I, we used a LHRD System and in Experiment II a SPD HMD with or without the SPD switched on.

To investigate sensemaking in each experiment and to compare how well both platforms used, we formulated the following hypotheses. We use the notation H1-II to signify the first hypothesis for the second experiment:

H1-I: Users spatialize their content through clustering information and exploit the advantages of a large display space.

H1-II: Users perform differently in a SPD HMD relative to an off-the-shelf VR HMD, resulting in different qualitative and quantitative outcomes.

H2-II: Users prefer a SPD over a standard VR HMD due to the larger FOV and thus a better sense of orientation in the space.

H3-II: SUS scores will show a higher sense of presence in the immersive environment with an SPD.

H1-C: Users perform better with a LHRD relative to a VR HMD, resulting in different qualitative and quantitative outcomes.

2.1. Global design decisions

To enable us to compare the results of both experiments, we aimed to maintain consistent design decisions across both experiments. Thus, we used (almost) identical experimental designs, using the same tasks for each experiment. We also tried to match the technological aspects of the HMD and the V4-Space as well as possible

by setting up the curvature and size of the virtual screen displayed in the HMD to be similar to the one of the physical V4-Space LHRD.

2.1.1. Task

In both experiments users were given a workspace populated with charts. Users could create new charts on the workspace, delete existing ones, resize as needed, and move them around freely. In both experiments, the participants' objective was to solve a VA task in an immersive environment. Participants were asked to complete two tasks subsequently.

In the first task, participants were asked to arrange the scatterplots into groups based on *similarity*. They were told that they could consider any similarity criterion and group according to their own perception of the data. We enforced a minimum of at least five (5) groups, but they could create more. Participants were required to group all the charts. After completion of the grouping, they were asked to explain their motivation behind the way they grouped the charts. The responses were recorded about all aspects they considered.

In the second task, participants highlighted each dimension in turn through clicking on the data dimension in the left panel. When a data dimension on the data panel is clicked, each chart using this dimension in the visualization panel is highlighted. This feature let users see if their grouping criteria had a "common dimension", i.e. charts having the same dimension in one of their axes are grouped. The task was to observe any patterns that might become apparent.

We observed in pilots that the classification behavior of novice users was sometimes purely based on the chart type, which has no relation with the visualized data. To ensure that classification decisions were not affected by chart type, only a single data type (numerical) and a single chart type (scatterplots) were used in the studies. The dataset we used for experiment I and experiment II has been the same US Census dataset. However, to randomize the data over the conditions in the VR tasks, we used an additional second dataset in experiment II.

We also observed in pilots that VR controllers/wands did not provide sufficient performance. Thus, we decided to use the same input device, a high-resolution gaming mouse, in both experiments.

2.1.2. Experimental Design

We used a convergent parallel mixed methods design for both experiments to collect both quantitative and qualitative data. We acquired quantitative data by measuring the number of groups created in the clustering. Additionally, to gain qualitative data and to get insights into the participants' thoughts, we asked the participants to think aloud while they were performing the tasks.

In experiment I, each participant performed primarily the first task. We collected qualitative data to gain an insight into the participants behavior and of the potential advantages of a large display space.

In the second experiment, each participant was exposed to two conditions with the VR HMD: one with the SPD display on and another with the SPD switched off. This resulted in a within-subject design. The task described above was the same for each condition. Accounting for the repeated exposure to the task in the second condition and to consider its implicit learning, we randomized the order of the SPD/Non-SPD condition for each participant to minimize the bias and ensure counterbalancing. Further, we used two different datasets which we randomized over the SPD/Non-SPD conditions. Additional quantitative data for Experiment II were acquired by measuring the completion time of the first task.

Qualitative and quantitative data of both experiments were later used to investigate differences between the platforms. To avoid learning effects across both experiments, we used a between-subjects design to compare between the LHRD and SPD/Non-SPD HMD.

2.1.3. Data Collected

As part of the instruction phase for each experiment, participants filled in a questionnaire regarding their background and their knowledge of VA, VR and immersive environments like LHRD. Participants were asked to use the think-aloud protocol while performing the task and to share their thoughts about their clustering choices. Participants used the same protocol to share their observations about the highlighted dimensions in relation to their own grouping.

During the tasks, users' answers were recorded via free-form text boxes and were stored in a survey system. Users' clustering activity was tracked in two ways: First, we screen-recorded their activities. Second, we recorded the positions and information of each chart for each final participant's grouping.

To be able to better understand the raw screen recordings, we watched the user carefully at each step of the procedure and recorded observations, e.g., through jotting down signs of frustration at specific points in time, which further informs the analysis of the results.

At the end of the study, we performed an audio-recorded post-study interview with each participant, taking typically less than 5 minutes. This interview was semi-structured. We asked participants a list of questions regarding the tasks they completed, the software tool, the system, what they liked, what was challenging and/or confusing, and whether they had any further feedback.

2.2. Apparatus

The most notable components of the apparatus for this study are a LHRD for the first experiment and a HMD with/without SPD for the second experiment. We further used a VA tool called DynSpace to carry out analytical tasks. DynSpace runs as a web application on a Node.js server with MongoDB running in the backend to support data collection during experiments [33]. We chose a web application because of its ubiquitous accessibility and flexibility in terms of both client and server side technologies and to ease for future expandability on large display screens.

2.2.1. V4-Space

V4-SPACE consists of a 1x7 array of large, tiled displays, each a vertically oriented 85" 4K Samsung Smart TV. The user sits in front of a desk at the center of the semi-circular arrangement of the TVs, see figure 2. We put an additional 21" monitor for auxiliary tasks before the user, below the line of sight to the large display. The main display has 15120 x 3840 pixels, which makes V4-SPACE a 58-megapixel system with 52 PPI pixel density. While the aspect ratio of a single display is 9:16, the system ratio is 63:16, about 4:1. The main display is 7.41 m by 1.88 m. In the 1x7 grid, there are no horizontal bezels and only 6 vertical ones.

V4-SPACE is controlled by a single computer with an Intel i7-6700K 4GHz processor with four PCI Express Gen 3 slots. The displays are driven by two nVidia Quadro M5000 cards, which provide four 4K outputs each, hardware synchronized through an nVidia Quadro Sync card. The auxiliary desktop monitor (which was not used for VA activities, and solely served to display a questionnaire form) is connected via an nVidia Quadro K620 graphic card. V4-SPACE relies on the nVidia Mosaic driver functionality, which presents all seven large displays as a single display to the user.

The system is designed for a single user who has a fixed position in front of the display system, about 3.3 m from each monitor. At this distance V4-SPACE is a "super-retina" display[34]. The display system is arranged in a circular arc (approximately 131° horizontal field-of-view, FOV) such that each monitor is equidistant to the user. This avoids information legibility issues due to non-uniform distances. As a limited form of physical navigation, the user can simply rotate their head or rotate a swivel chair to look at different parts of V4-SPACE.

Interaction is through keyboard and mouse. To support the high resolution of V4-SPACE, we use a Razer DeathAdder Chroma 10000 PPI optical gaming mouse, which permits the user to perform pixel-accurate pointing on the large display surface.

2.2.2. SPD HMD

For the HMD we used an off-the-shelf HTC Vive VR headset which was extended with a custom built Sparse Peripheral Display (SPD) to broaden the field-of-view of the virtual environment (figure 3). The SPD consists of 256 RGB LEDs placed on a 2-layer flexible printed circuit board. The LEDs are set in a radial array and the flexible PCS is folded to tightly fit around the lenses of the headset. The LEDs themselves are controlled by two LED drivers (Texas Instrument TLC5951), which provide 12-bit PWM color control to each color of the LED RGB triplet (36 bits of color per LED). Its LED drivers are controlled by an Cortex-M3 microcontroller (Cypress CY8C5288 LTI-LP090), which handles all communication with the computer through USB. The SPD update rate is set to 100Hz and the horizontal display FOV consists of approximately 180° (see [28] for a similar setup). We switched the SPD on and off for the two conditions investigated in this experiment.

2.2.3. Common software platform

The VA tool DynSpace is a browser-based VA tool written in JavaScript. Its user interface consists of two two panels: a Data Panel on the left side showing data dimensions and the main visualization panel on the right. Users can select data dimensions in the left one and drag-and-drop them into charts in the main visualization panel for analysis.

The visualization panel contains data charts that show relations between selected data dimensions. Each chart is contained in a rectangular sub-panel that a user can move, resize, add, or delete, and shows a 2D data plot. All plots are coordinated through brushing and linking.

DynSpace was designed to aid analysis of complex datasets [34]. Initially it displays a number of charts, each generated automatically by picking random pairs of data dimensions. This set of initial charts uses about half of the workspace in either of the experiments. The initial display of charts is a simple array with no clustering. Enough free space is left for the user to arrange the spatial layout of the content as they wish, e.g., by moving charts and creating clusters.

DynSpace uses a grid-based layout manager that enforces complete visibility of charts at all times by not allowing charts to partially or completely overlay one another. The available space is divided into invisible rows and columns so that the clusters always appear as an array.

3. Experiment I

The purpose of experiment I was to observe user behavior during a classification task in a LHRD.

3.1. Methods

The spatialization task used a simplified version of DynSpace to display 2D plots visualizing relations among a subset of the 2016 US Census Dataset. This study took place on the V4-SPACE display (figure 4).
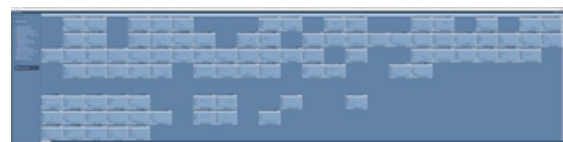

Figure 4: The instance of DynSpace that spans the display space of our LHRD. Each square is a chart of a part of the data.

3.1.1. Participants

There were nine participants, P0 to P8, four of whom were male. Ages ranged from 18 to over 40. Some were undergraduate students participating for course credit, others were volunteers with at least a bachelor's degree. Through pilot studies, we ensured that all users could read the text on the displays from the default chair position without problems.

Participants were asked about their familiarity with seven VA terms and concepts in the pre-study survey. On average, they were familiar with 4.4 of those.

Since the study did not require specific domain knowledge, we were able to use mostly novice users for this study. 5 participants out of 9 did not have any VA experience. One reported less than a year of experience and three reported 1 to 3 years of experience. Seven of them had never used any

visualization or VA tools, whereas one had used D3.JS and another used R in a statistics course for a term.

In a pre-study survey we asked users whether they could interpret data from scatter plots. They ranked their ability of interpreting data correctly from a scatter plot on a scale from 1 to 5. The results were 3x "Sometimes", 4x "Mostly" and 2x "Yes, always" answers; which respectively stood for 3, 4 and 5 on the scale used.

3.1.2. Procedure

We first instructed participants that they could rotate/swivel the chair to see the full display, but that the chair had to remain in the same location until the end of the experiment, to retain the same distance to all displays in the system. Leaning back and forward was permitted, as desired. Next, we trained them in basic, yet frequent operations on V4-Space: keeping track of the cursor, how to find it when they lost track of it, and switching between the LHRD and the auxiliary monitor. Then, users were introduced to DynSpace along with some practical tips regarding the usage. We assured users that they could ask questions during the experiment and that they should communicate their thoughts around the tasks or whenever they experienced issues. We recorded any direct or indirect feedback from the users during the tasks through note-taking.

3.2. Results

Reporting the results for experiment I, we use the notation P3-I to signify participant three from the first experiment.

3.2.1. Clustering

Three participants created only the minimum allowed five groups and four clustered their plots into 6 groups. The other two participants created 8 and 16 groups respectively. The participants used one of the following three equally prevalent strategies when clustering: Similar visual appearance (P1-I, P2-I, P7-I), commonality in labels (P0, P4-I, P5-I), or common topics (P3-I, P6-I, P8-I). When building clusters, charts were either arranged horizontally or vertically (P0-I, P1-I, P4-I, P5-I), roughly circular (P6-I, P8-I), or in a mixed arrangement (P2-I, P3-I, P7-I).

For some users, there was no perceptible relation between different clusters in the workspace (P0-I, P1-I, P4-I, P5-I). For P3-I, cluster shapes were determined by cluster types. For the rest (P2-I, P6-I, P7-I, P8-I), the distance between clusters decreased as the similarity between clusters increased. For these participants, relations between clusters were reflected by cluster separation. For some participants, there was either no (P0-I, P1-I) or only a minimal (P3-I, P5-I, P8-I) distance between clusters. For some participants (P4-I, P6-I) the bezels strongly influenced the cluster arrangement. For a few users (P2-I, P7-I), the distance between clusters varied depending on the inter-cluster relations.

3.2.2. Space Usage

Out of nine participants, seven used the entire width display provided by the system. If a user runs out of free space, DynSpace permits vertical scrolling via the mouse wheel (but not horizontal scrolling). Surprisingly, P0-I and P1-I used only about 2/7 and 3/7 of the space, respectively, and relied heavily on vertical scrolling.

3.2.3. Navigation Techniques

Six of the participants used physical navigation frequently during the tasks, i.e., they rotated their head and/or body back and forth to "access" all parts of the LHRD visually. The others kept their gaze mostly focused on a subset of the displays.

4. Experiment II

The purpose of Experiment II was to observe sensemaking in a VR environment and investigate whether an increased field of view in VR via a SPD has an effect on VA tasks.

4.1. Methods

We used a HTC Vive with a SPD to show the virtual environment through Unity 3D. The VR environment consisted of a curved display surface, as seen in figure 5.
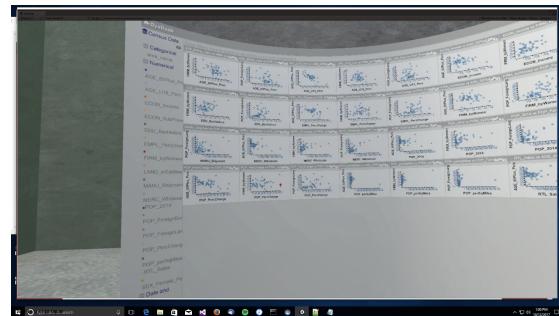

Figure 5: Curved screen surface in the virtual environment showing the scatterplots.

We duplicated several desktop windows into the virtual environment and displayed different web applications, including DynSpace, using the Awesomeium plugin for Unity 5. We also redirected mouse and keyboard events to enable interaction with the content of these windows. We displayed more than a single window to make sure that participants could not get lost in the virtual environment. The largest window was placed in front of the user and showed DynSpace with 64 scatter plots on the VR surface to perform the VA task. Participants used a computer mouse to interact with the DynSpace user interface which was placed on a table in front of the participants. We used a screen-capture software to record the participants' interactions with the system and to collect the task data.

4.1.1. Procedure

Prior to the study, participants were instructed in the specifics of VR, SPD and VA systems and were asked to give informed consent. They were further asked to agree to screen capturing and audio recording while taking part in the study. Participants were verbally instructed on the procedure for the experiment and exposed to a short training session prior to the actual task. A researcher was always present to collect observations, provide instruction, and for technical support, if needed.

Participants sat on a chair located in the middle of a lab environment, with a little table in front of them to use the computer mouse. The surrounding area was emptied so they could not touch or bump into anything while being immersed to the virtual world.

After each condition, each participant completed a Slater-Usoh-Steed questionnaire to account for their perception of presence. This has been additional collected data compared to experiment I.

4.1.2. Participants

In total, 7 (new) subjects (5 female; mean age was 21.28 ± 2.05 years) participated. Subjects were students of the local university and recruited from a local subject pool. The participants were unaware of the purpose of the study. The experiment was approved by the ethics committee of the university.

A pre-study survey indicated that the participants had little experience in VA tasks (3x "None", 3x "less than a year", 1x "1-3 years") and have been uniformly familiar with VR Headsets (3x "Yes", 4x "No").

4.2. Results

For the results of experiment II, we use the notation P4-II to signify participant four from the second experiment.
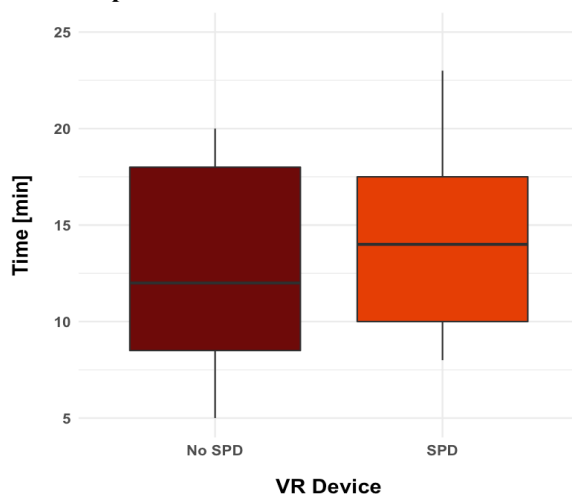
4.2.1. Completion Time

Figure 6: Boxplots showing the time needed to complete the subtasks in either the SPD or the no-SPD condition.

No participant exceeded the 60 minute limit of the study. Completion time of each task varied from 5 to 20 minutes in the No-SPD condition ($M = 12.85$, $SD = 6.12$) and from 8 to 23 minutes ($M = 14.28$, $SD = 5.40$) in the SPD-condition (see figure 6).

A Shapiro-Wilk test showed a normal distribution for both, the No-SPD and the SPD condition. Variance homogeneity between conditions exists. An ANOVA showed no significant results of the condition on the working time ($F(1,6) = 3.614$, $p = 0.105$, $\eta^2 = 0.018$).

Participant P7-II created by far the biggest number of groups (SPD: 15, No SPD: 17). The participant noted afterwards, that they wanted to solve the task as soon as possible. Since the participant only relied on the commonality of labels, they performed merely a pattern recognition task on the naming of the plots, rather than truly making sense of the data.

In contrast, participant P3-II solved the task very carefully and even changed their grouping strategy in the second condition. They took about 20 minutes for grouping through "commonality in labels" (No SPD) and 23 min. for grouping by similar visual appearance (SPD).

4.2.2. Number of Clusters

The number of groups (figure 7) created in each condition varied from 6 to 17 in the No-SPD condition ($M = 9.42$, $SD = 3.69$) and 6 to 15 in the SPD condition ($M = 8.42$, $SD = 3.20$). A Shapiro-Wilk test showed a normal distribution for the No-SPD condition and not for the SPD condition. Variance homogeneity between conditions exists. We performed a parametric test since it is adequate robust to possible violations of the normality assumption [35]. The ANOVA identifies a significant difference between conditions for the number of groups created ($F(1,6) = 7.000$, $p = 0.0382$, $\eta^2 = 0.024$).
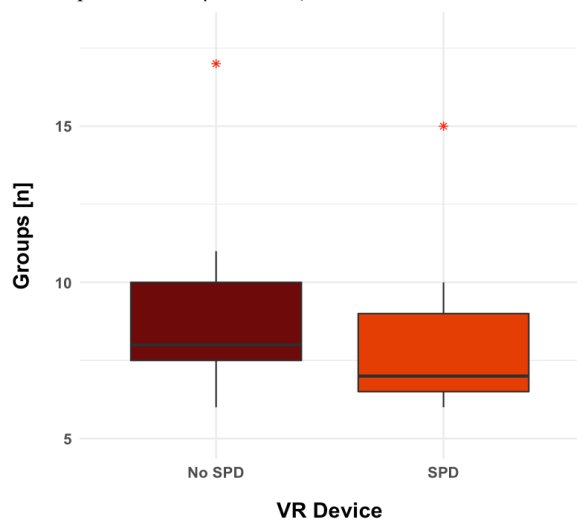

Figure 7: Boxplots showing the number of clusters in each subtask in either the SPD or the no-SPD condition.

4.2.3. Clustering Strategy

Participants used two different strategies for grouping plots in experiment II. The first strategy was to group by a commonality in labels. The second one was to group by a similar visual appearance of the scatterplots shown, e.g., (broadly) increasing plots vs. plots with no visible trends. While we expected that participants would use the latter strategy more frequently, the first one was used more widely. We could not identify a considerable difference of the clustering strategy between the SPD and No-SPD conditions.

However, we observed a shift in strategy, from an initial grouping by commonality in labels to a grouping by similar visual appearance (illustrated in figure 8). This shift seems to be because of the succeeding subtasks. After the first subtask, one participant (P3-II) mentioned that they expected the visual appearance of the scatterplots to be more important for making sense of the data, rather than just ordering them by a commonality in axis labels. Thus,

this participant grouped the graphs in the second condition according to the visual appearance of the scatterplots.Figure
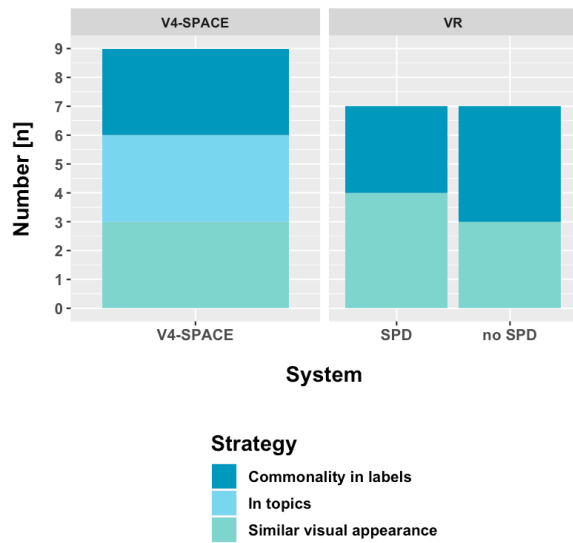
8: Barcharts showing the strategy used for clustering in V4-SPACE and VR setups (for both experiments).

4.2.4. Cluster Shapes

We observed three different kinds of arrangements participants used to classify the data: horizontally and vertically arranged, roughly circular, or a mixed arrangement (see figure 9). This classification is similar to arrangements observed in experiment I
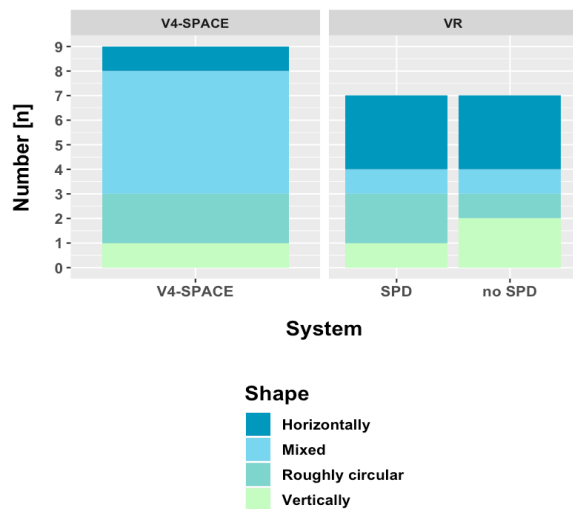

Figure 9: Barcharts showing the shape of the clusters used for clustering in V4-SPACE and VR setup

Interestingly, the arrangement of clusters seems to directly correspond to the used clustering strategy. When grouping by commonality in labels, the arrangement of the clusters was horizontal or vertical. The strategy to use similar visual appearances changed the arrangement of the cluster shapes to a more "prototypical" arrangement of clusters with roughly circular shapes. Participant (P3-II) even commented that this seemed to be „more intuitive". This indicates that the arrangement of clusters may be directly related to the clustering strategy

4.2.5. Advantages and Disadvantages of Sparse Peripheral Displays

Asking the Participants what HMD condition they preferred, the outcomes reveal the same equal level of preference for "No SPD preferred", "No difference experienced" or "SPD preferred" (see figure 10). One participant (P1-II) noted that they did not notice the SPD change between the tasks.

Those participants who did not like the SPD condition mentioned that they perceived the SPD display to be too bright and thus as distracting while solving the VA task. One participant (P7-II) felt that the SPD impaired perception and that it actually "took away" from the experience. The same participant mentioned that the SPD was "some sort of separately perceived reality" to them.

When asked about the SPD experience without a focus on solving a VA task, the SPD was preferred by a majority of participants, potentially due to the perception of less or no motion sickness. This result is contrary to the results of [28], where they observed an increase in motion sickness in using a SPD. Further, one participant (P3-II) mentioned that the SPD gave him a better sense of orientation. This helped him to navigate better between the content in the second subtask (the highlighting of the dimensions).
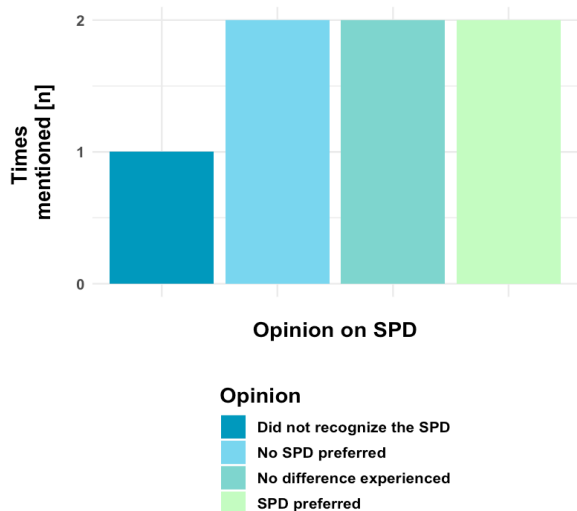

Figure 10: Preferences of the participants regarding the SPD headset.

4.2.6. Presence Questionnaire

The overall Slater-Usoh-Steed (SUS) score for each subject was quite high, resulting in a high prevalence of participants being present in the virtual environment in general (figure 11). An ANOVA identified no significant results of the condition on the SUS score ($F_{(1,12)} = 0.025$, $p = 0.878$, $\eta^2 = 0.0021$). Thus, our data disconfirm hypothesis H3-II.

Only a single participant (P7-II) gave the SPD condition a conspicuously higher SUS score than the no-SPD condition. The participant is the same who identified the SPD to be some sort of separately perceived reality (see section 4.2.5). Interestingly, the same participant also mentioned the SPD to be less preferable than the No-SPD condition, which contradicts their SUS rating.

P3-II identified the SPD condition as more preferable and that it gave them a broader sense of orientation during the task. Looking at the participants' SUS scores, there was no significant difference between the conditions though. Overall, the SUS scores of P3-II were notably lower than any other participants' scores.
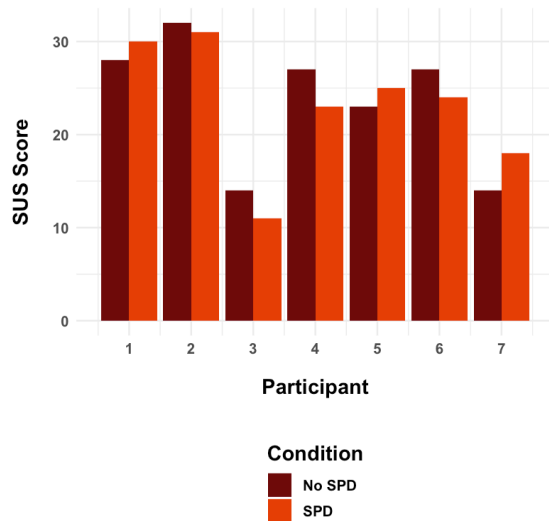

Figure 11: Results of the Slater-Usoh-Steed questionnaire for each participant and each HMD condition.

4.2.7. Highlighting of Dimensions

The highlighting of the dimensions was a solely qualitative task. For those participants who grouped the plots by a commonality in labels, any highlighting of the dimensions directly matched the arranged groups. It is interesting to see that even though the participants did not know the requirements of the second subtask at the beginning, they arranged the groups by the dimensions in the first subtask. This was helpful for the (future) second subtask. Yet, no participant discovered and mentioned this. Most of the participants did not identify any reasons as to why some dimensions matched their groupings and why some did not. Not surprisingly, the participants who grouped according to similar visual appearance tried to make sense of their arrangement far more often than the participants who grouped by the commonality in labels. Their groups were arranged more likely by different dimensions and matched the highlighting less well. Participant (P5-II) mentioned it might have been better for the second subtask to arrange the graphs by the commonality in labels as opposed to arranging the clusters in a roughly circular manner. Using the think-aloud protocol, most participants who grouped by similar visual appearance mentioned the potential utility of cross references between the arranged groups according to the dimensions. Participant (P6-II) arranged the graphs by commonality in labels. However, faced with the second subtask, the participant tried to explain the dimensional matching only based on the scatterplot appearance and did not rely on the labels. However, the participant was still not able to come up with a reason as to why some plots where in different groups in their arrangement.

Overall, most of the participants found the second subtask of making sense of the data and highlighting the dimensions challenging. One potential explanation is the relative lack of experience with VA in our participants.

5. Discussion

We can confirm Hypothesis H1-I, that users spatialize their content through clustering information and exploit the advantages of a large display space, through the observations of the users' clustering behaviour in experiment I. Instead of treating the plots as a single group of data, they classified the plots through various strategies and used the full display space during sensemaking to create different clusters of data plots.

Our second experiment, does not support any of the hypotheses H1-II, H2-II or H3-II, as we did not find any significant results that users perform differently in a SPD HMD compared to a common VR Headset.

Combining the quantitative results for completion time and number of clusters with the qualitative data of clustering strategy and cluster shapes, we are unable to discern a clear picture if users perform differently in a SPD HMD relative to an off-the-shelf VR HMD. Thus, our hypothesis H1-II is not supported.

We see no support for hypothesis H2-II, since we are unable to reach a clear conclusion on whether the SPD condition was preferred by the users while solving a VA task. The (low-resolution) larger field-of-view did not seem to have had a strong effect on the outcome. Finally, we observed no difference on the Slater-Usoh-Steed questionnaire, resulting in an unmet hypothesis H3-II. Using a between-subjects design we compared the results of experiment I and experiment II to investigate differences in sensemaking between the LHRD system and the VR HMD with and without the SPD. As seen in figure 8, 12 and 9, there are several differences between the results of experiment I and experiment II.

As we did not record the completion time in the first Experiment, we were not able to compare times between the experiments. We thus analyzed quantitative data in terms of the number of created clusters to compare both experiments. A Shapiro-Wilk test showed a normal distribution for the No-SPD condition, but not for the SPD and the LHRD. Variance homogeneity between the three exists. Since parametric tests are robust to the violation of the normality assumption [35], we performed an ANOVA over V4-SPACE and HMD results (V4-Space: $M = 7.0$, $SD = 3.5$; No-SPD condition: $M = 9.42$, $SD = 3.69$; SPD condition: $M = 8.42$, $SD = 3.20$). The results showed no significant difference between the number of created cluster in V4-Space and the no-SPD HMD ($F_{(1,12)} = 1.257$, $p = 0.284$, $\eta^2 = 0.095$) and the V4-SPACE and the SPD HMD ($F_{(1,12)} = 0.449$, $p = 0.515$ $\eta^2 = 0.036$).

Thus, we may disconfirm hypothesis H1-C and can only state that users do not seem to perform differently in an HMD relative to a LHRD. Users did not perform better in a LHRD relative to a VR HMD according to our investigations.
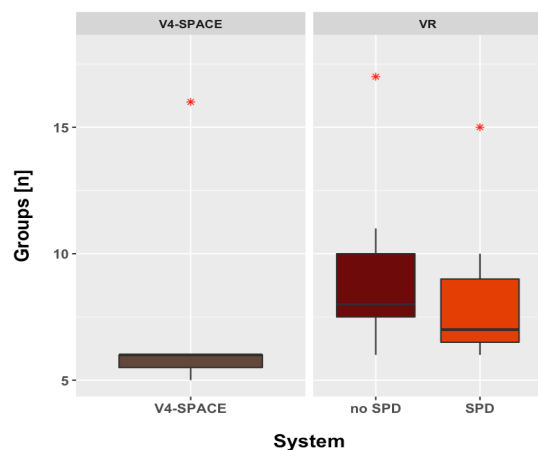
Figure 12: Number of clusters generated in V4-SPACE and the VR setups.

Clustering Strategy

In terms of the clustering strategy, in the second experiment a subset of users relied on grouping based on the commonality of labels. Yet, as the participants progressed through the experiment, we observed a transition towards a strategy based on similar visual appearance of scatterplots in some participants. This could be considered to be a more reflective way of grouping the plots. We hypothesize that grouping by similar visual appearance corresponds to a higher-level sensemaking activity, since it involves more active reflection about the data. In the first experiment, we observed an additional, third strategy. While about a third of the participants based their strategy on the visual appearance of the scatterplots and another third based it on the commonality of axis labels, the remaining third adopted a clustering approach based on common topics of plots. This strategy is more complex and reveals that some were actively trying to make sense of the data.

Using this strategy, participants neither clustered according to a pattern recognition/visual matching strategy, regardless of what the data presents; instead, they aimed to bring charts together based on what the data might be about. Even if the axis labels were different, users thought about the concepts and what those concepts are about and clustered the plots so that each went into a corresponding category of topics.

Cluster Shape

We also compared the shape of the clusters the participants generated. In the first experiment the participants' rearrangement efforts resulted mostly in simple shapes, either highly rectangular (aspect ratio substantially different from 1) or roughly circular (aspect ratio around 1). Classifying cluster shapes by aspect ratio (large ratio: horizontal; small: vertical) yields the following grouping (with examples shown in figure 13):

- Horizontal: 1 participant.
- Vertical: 3 participants.
- Roughly circular: 5 participants.

For several participants, no clear pattern of cluster arrangement could be identified. P0-I and P1-I did not have any space between clusters arranged in a random order. For P4-I and P5-I, clusters were separated by some distance. However, the spacing seemed uniform and clusters appeared similar. In contrast, other participants used cluster position to convey meaning. As an example, P3-I used the cluster shape to indicate what the clusters represented. While P3-I categorized groups looked somewhat similar, their group of uncategorized plots looked different from other clusters in terms of shape and the distance from other clusters.

We observed similarities in cluster shapes in experiment II as well. Participants' cluster shapes were defined similarly to the first experiment (horizontal or vertical, roughly circular, mixed arrangement). As in experiment I, horizontal or vertical arrangement was used by participants that did not seem to think deeply about their clustering strategy.

When users appeared to make more sense of the data, they started to create more traditional "roughly circular" clusters and used shape and space as a tool for sensemaking within the available workspace.
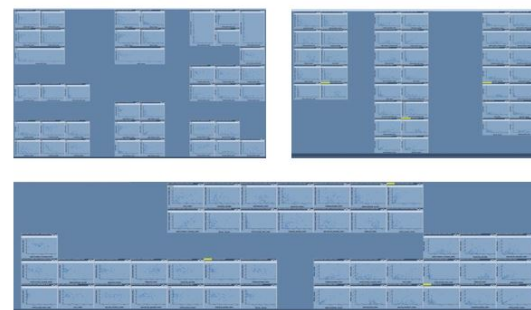

Figure 13: Variation in aspect ratios used by participants. P6-I used a roughly circular arrangement (aspect ratio ~1, top left) while P4-I used a vertically-oriented layout (aspect ratio <<1, top right) and P5-I a more horizontal organization (aspect ratio >> 1, bottom).

5.1. Summary

We conducted two sensemaking studies in immersive environments, one on a large high-resolution display (LHRD) called V4-SPACE and one using a HMD-based VR system with a sparse peripheral display option that we turned on and off. Using equivalent tasks, we observed users' behaviors in those environments and compared the outcomes within the same environment and the two separate display environments to each other.

The first exploratory study aimed to identify challenges and yielded qualitative observations for VA tasks in a LHRD environment. We obtained new empirical information about users' approaches when asked to prepare a large volume of data for analytical tasks on large displays. Even though all users started with the same state, the results of experiment I suggest that different users follow varying classification and spatial organization strategies. We observed clear distinctions in terms of clustering strategies, space usage, and preferred navigation techniques.

In the second experiment, we investigated the utility of a sparse peripheral display condition on a VR HMD. The presence of the SPD yielded mixed responses. While some users found it 'too bright', 'distractive' or thought that it reduces the immersion; others thought it causes less or no

motion sickness or that it supported a better sense of orientation. Quantitative analysis yielded a significant larger number of clusters created in the SPD than in a common VR HMD.

The comparison of both experiments showed no significant difference of the performed quantitative data in V4-SPACE and the VR SPD HMD. A comparison of the qualitative data yielded an additional strategy used in V4-SPACE and a slightly varying distribution of cluster shapes. The results show initial tendency that users perform different in both immersive environments, leaving the possibility for future research.

5.2. Limitations

A limitation we have to consider is the small sample size of n = 9 in experiment I and n = 7 of experiment II. While this limits the strengths of the insights we can derive, we point out that the effort to run the studies was high (and the SPD broke during the second experiment for the eight participant).Further, our participants were university students with limited VA experience. We observed that some participants in experiment II could not tell what was happening when highlighting dimension. We believe the lack of VA experience might have impacted how much sense the participants were able to extract from the data.

We tried to carefully match experimental conditions for both environments through using the same input device, apparent size of the workspace, amount of information displayed and the tasks that participants had to perform. Yet, there were also technological limits to this, such as the maximum resolution of the available VR headset system being far inferior to what is available on a LHRD.

Since the participants could terminate the task whenever they wanted, the duration to solve the task mostly depended on their motivation and their clustering and sensemaking strategies for the data. Further, since there was no single correct solution for the task, each participant used a different classification strategy with their own criteria. This limitation introduces noise into the main goal of comparing sensemaking in HMD and LHRD technologies and makes it more difficult to draw strong conclusions. In future work, it would be interesting to repeat the experiment on a conventional computer environment as a reference baseline.

A limitation for the second experiment in specific is using a computer mouse as an input device for the VR HMD. Yet, as mentioned above, we found that a VR controller was not sufficiently accurate to enable users to interact with the details of the charts. Since a mouse is not a typical HMD input device, it might make its interaction with the system unnatural. While participants could not see the physical mouse they were using while being immersed in the HMD, none of the participants reported any issues with this.

References

1. Shrinivasan YB, van Wijk JJ. Supporting the analytical reasoning process in information visualization. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM; 2008. pp. 1237–1246.

2. Endert, Alex, Patrick Fiaux, and Chris North. Unifying the sensemaking loop with semantic interaction. IEEE Workshop on Interactive Visual Text Analytics for Decision Making at VisWeek. 2011;

3. Endert A, Fiaux P, North C. Semantic Interaction for Sensemaking: Inferring Analytical Reasoning for Model Steering. IEEE Trans Vis Comput Graph. 2012;18: 2879–2888.

4. Endert A, Fox S, Maiti D, Leman S, North C. The Semantics of Clustering: Analysis of User-generated Spatializations of Text Documents. Proceedings of the International Working Conference on Advanced Visual Interfaces. New York, NY, USA: ACM; 2012. pp. 555–562.

5. Marshall CC, Rogers RA. Two Years Before the Mist: Experiences with Aquanet. Proceedings of the ACM Conference on Hypertext. New York, NY, USA: ACM; 1992. pp. 53–62.

6. Andrews C, Endert A, North C. Space to think: Large, High-Resolution Displays for Sensemaking. Proceedings of the 28th international conference on Human factors in computing systems - CHI '10. 2010; 55–64.

7. Baudisch P, Good N, Bellotti V, Schraedley P. Keeping Things in Context: A Comparative Evaluation of Focus Plus Context Screens, Overviews, and Zooming. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 2002. pp. 259–266.

8. Czerwinski M, Smith G, Regan T, Meyers B, Robertson G, Starkweather G. Toward Characterizing the Productivity Benefits of Very Large Displays. Interaction. 2003;3: 9–16.

9. Chandler T, Cordeil M, Czauderna T, Dwyer T, Glowacki J, Goncu C, et al. Immersive Analytics. 2015 Big Data Visual Analytics, BDVA 2015. 2015. doi:10.1109/BDVA.2015.7314296

10. Marriott K, Schreiber F, Dwyer T, Klein K, Riche NH, Itoh T, et al. Immersive Analytics. Springer; 2018.

11. Donalek C, Djorgovski SG, Cioc A, Wang A, Zhang J, Lawler E, et al. Immersive and Collaborative Data Visualization Using Virtual Reality Platforms. 2014 IEEE International Conference on Big Data Immersive. 2014; 609–614.

12. Zhang S, Keefe DF, Dasilva M, Laidlaw DH, Greenberg BD. An Immersive Virtual Environment for DT-MRI Volume Visualization Applications : a Case Study. IEEE. 2001;Vi: 1–5.

13. Chen J, Cai H, Auchus AP, Laidlaw DH. Effects of Stereo and Screen Size on the Legibility of Three-Dimensional Streamtube Visualization. IEEE Trans Vis Comput Graph. 2012;18: 2130–2139.

14. Demiralp C, Jackson CD, Karelitz DB, Zhang S, Laidlaw DH. CAVE and Fishtank virtual-reality displays: a qualitative and quantitative comparison. IEEE Trans Vis Comput Graph. 2006;12: 323–330.

15. Ragan ED, Kopper R, Schuchardt P, Bowman DA. Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task. IEEE Trans Vis Comput Graph. 2013;19: 886–896.

16. Helbig C, Bauer HS, Rink K, Wulfmeyer V, Frank M, Kolditz O. Concept and workflow for 3D visualization of atmospheric data in a virtual reality environment for analytical approaches. Environ Earth Sci. 2014;72: 3767–3780.

17. Hsieh T-J, Chang Y-L, Huang B. Visual analytics of terrestrial lidar data for cliff erosion assessment on large displays. Satellite Data Compression, Communications, and Processing VII. International Society for Optics and Photonics; 2011. p. 81570D.

18. Laha B, Bowman DA, Socha JJ. Effects of VR System Fidelity on Analyzing Isosurface Visualization of Volume Datasets. 2014;20: 513–522.

19. Pirolli P, Card S. The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis. Proceedings of international conference on intelligence analysis. 5: 2–4.

20. Stuerzlinger W, Dwyer T, Drucker S, Görg C, North C, Scheuermann G. Immersive Human-Centered Computational Analytics. In: Marriott K, Schreiber F, Dwyer T, Klein K, Riche NH, Itoh T, et al., editors. Immersive Analytics. Cham: Springer International Publishing; 2018. pp. 139–163.

21. Elmqvist N, Moere AV, Jetter H-C, Cernea D, Reiterer H, Jankun-Kelly TJ. Fluid interaction for information visualization. Inf Vis. SAGE Publications; 2011;10: 327–340.

22. Russell DM, Stefik MJ, Pirolli P, Card SK. The Cost Structure of Sensemaking. Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 1993. pp. 269–276.

23. Endert A, Fiaux P, North C. Semantic Interaction for Visual Text Analytics. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 2012. pp. 473–482.

24. Reda K, Johnson AE, Papka ME, Leigh J. Effects of Display Size and Resolution on User Behavior and Insight Acquisition in Visual Exploration. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 2015. pp. 2759–2768.

25. Ni T, Bowman DA, Chen J. Increased Display Size and Resolution Improve Task Performance in Information-Rich Virtual Environments. Proceedings of Graphics Interface 2006. Toronto, Ont., Canada, Canada: Canadian Information Processing Society; 2006. pp. 139–146.

26. Xiao R, Benko H. Augmenting the Field-of-View of Head-Mounted Displays with Sparse Peripheral Displays. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16. 2016. pp. 1221–1232.

27. Nagahara H, Yagi Y, Yachida M. Super wide viewer using catadioptrical optics. Proceedings of the ACM symposium on Virtual reality software and technology. ACM; 2003. pp. 169–175.

28. Hashemian AM, Kitson A, Ngyuyen-Vo T, Benko H, Stuerzlinger W, Member, et al. Evaluating a New Sparse Peripheral Display in a Head-Mounted Display for Vection and a Navigational Search Task.

29. Nguyen-Vo T, Riecke BE, Stuerzlinger W. Moving in a box: Improving spatial orientation in virtual reality using simulated reference frames. 2017 IEEE Symposium on 3D User Interfaces (3DUI). 2017. pp. 207–208.

30. Riecke BE, Bodenheimer B, McNamara TP, Williams B, Peng P, Feuereissen D. Do we need to walk for effective virtual reality navigation? physical rotations alone may suffice. Proceedings of the 7th international conference on Spatial cognition. Springer-Verlag; 2010. pp. 234–247.

31. Thomas JJ, Cook KA, Process AD. Illuminating the path: The research and development agenda for visual analytics. IEEE Computer Society. 2005; 184.

32. Bi X, Balakrishnan R. Comparing Usage of a Large High-Resolution Display to Single or Dual Desktop Displays for Daily Work. CHI '09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009. pp. 1005–1014.

33. El Meseery M, Wu Y, Stuerzlinger W. Multiple Workspaces in Visual Analytics. 2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA). 2018. pp. 1–12.

34. Cetin G, Stuerzlinger W, Dill J. Visual Analytics on Large Displays: Exploring User Spatialization and How Size and Resolution Affect Task Performance. 2018 International Symposium on Big Data Visual and Immersive Analytics (BDVA). 2018. pp. 1–10.

35. Rasch D, Guiard V. The robustness of parametric statistical methods. Psychol Sci Q. 2004;46: 175–208.