# The Impact of Immersion on Cluster Identification Tasks

M. Kraus, N. Weiler, D. Oelke, J. Kehrer, D. A. Keim, and J. Fuchs

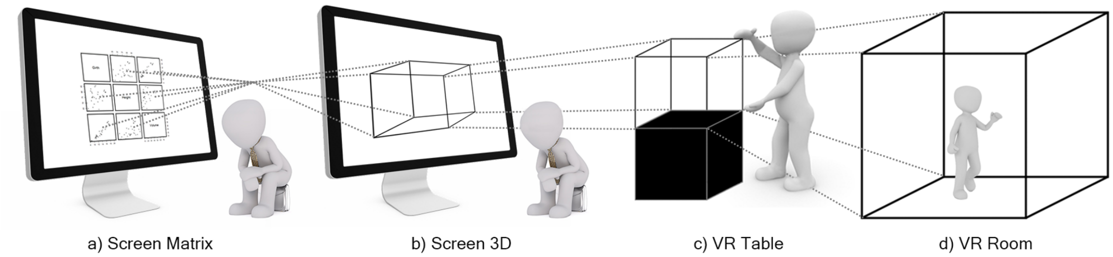



Fig. 1. A cluster identification task was performed and evaluated in four different visualization design spaces. Two screen-based methods, namely a scatterplot matrix (a) and a 3D scatterplot in a cube (b), and two visualizations in a VR environment: a 3D scatterplot on a virtual table (c) and a room-sized scatterplot (d). Gray lines emphasize transitions between visualization design spaces.

**Abstract**—Recent developments in technology encourage the use of head-mounted displays (HMDs) as a medium to explore visualizations in virtual realities (VRs). VR environments (VREs) enable new, more immersive visualization design spaces compared to traditional computer screens. Previous studies in different domains, such as medicine, psychology, and geology, report a positive effect of immersion, e.g., on learning performance or phobia treatment effectiveness. Our work presented in this paper assesses the applicability of those findings to a common task from the information visualization (InfoVis) domain. We conducted a quantitative user study to investigate the impact of immersion on cluster identification tasks in scatterplot visualizations. The main experiment was carried out with 18 participants in a within-subjects setting using four different visualizations, (1) a 2D scatterplot matrix on a screen, (2) a 3D scatterplot on a screen, (3) a 3D scatterplot miniature in a VRE and (4) a fully immersive 3D scatterplot in a VRE. The four visualization design spaces vary in their level of immersion, as shown in a supplementary study. The results of our main study indicate that task performance differs between the investigated visualization design spaces in terms of accuracy, efficiency, memorability, sense of orientation, and user preference. In particular, the 2D visualization on the screen performed worse compared to the 3D visualizations with regard to the measured variables. The study shows that an increased level of immersion can be a substantial benefit in the context of 3D data and cluster detection.

**Index Terms**—Virtual reality, evaluation, visual analytics, clustering.

## 1 Introduction

Different visualization design spaces, i.e., spaces in which a visualization is projected, exist. Visualizations often need to adapt to the given design space, which can change their level of immersion. An example of a common visualization design space is a two-dimensional space on a monitor screen. Any visualization that encodes a maximum of two attributes with one position can be displayed within this space (e.g., 2D scatterplot [9] or 2D parallel coordinates [24, 60]). Another visualization design space is created if an additional third attribute is encoded in the visual variable "position" (e.g., space time cubes [3, 29] or 3D scatterplots [37, 43, 45]). The level of immersion may already differ between the two exemplary design spaces (2D and 3D design space) as a higher degree of abstraction is necessary to display the same information in 2D as compared to a more natural display in 3D. For instance, a 3D scatterplot can be visualized in the 2D visualization space as a scatterplot matrix or after a PCA projection in a 2D scatterplot, both being more abstract than a 3D scatterplot visualized in a 3D visualization design space. The more familiar nature of the 3D data representation may lead to a more intense perception of immersion.

Over the last few years, augmented-, virtual-, and mixed-reality (AR, VR, MR) hardware and software have been on the rise, opening up new design spaces for visual analytics (VA) applications. Various examples of visualizations exist in VR, AR, and MR, either restricting the visualization's space to a small area [4] or allowing it to occupy the entire space around the observer [14, 30]. As the level of immersion with regard to the visualization differs largely between the two kinds, their visualization design spaces can be seen as two individual ones.

It is often not a trivial decision which design space is best suited for a specific task. There are several studies comparing visualizations in VREs to those in conventional design spaces, but they often focus on differences resulting from different visualization and interaction techniques [4, 58]. These studies do not capture how much of the differences in performance can be ascribed to those two factors and how much to the different levels of immersion. In this paper, we want to investigate how much influence the choice of the design space, and the associated level of immersion, has on the overall performance of visualizations. Since this is a rather broad question, we specifically focus on the task of cluster detection in scatterplot visualizations. Our study builds upon the work of Wagner Filho et al. [58] who investigated the effects of immersion provided by VREs. However, they compared the level of immersion provided by different interaction techniques and not the level of immersion provided by the design space. In particular, we investigated differences between four visualization design spaces, each having a certain level of immersion. In order to focus on differences due to the design spaces themselves, we minimized user interaction and used three-dimensional data in combination with a simple visualization.

- *M. Kraus, N. Weiler, D. A. Keim, and J. Fuchs are with University of Konstanz, Germany. E-mail: {matthias.kraus | niklas.weiler | daniel.keim | johannes.fuchs}@uni-konstanz.de.*
- *D. Oelke and J. Kehrer are with Siemens Corporate Technology, Munich, Germany. E-mail: {daniela.oelke | kehrer.johannes}@siemens.com.*


*Digital Object Identifier: 10.1109/TVCG.2019.2934395*

## 2 RELATED WORK

In this section, we provide a brief overview of the most related strains of research. First, we target research in which 2D visualizations were deployed and quantified for cluster identification tasks. Second, we present several examples of effects of immersion in various domains, which motivated our research to assess similar effects for InfoVis tasks. At the same time we outline how immersion was measured in previous work. We then focus specifically on 3D visualizations and 3D scatterplots since they are an integral part of the current work for the reason that they serve as a base visualization in the present study. Subsequently, we discuss advantages and disadvantages of data visualizations using stereoscopic displays as the current work investigates possible benefits and drawbacks of visualization design spaces in VR compared to screen-based ones by means of scatterplots.

### 2.1 Cluster Identification with static 2D Visualizations

For cluster identification tasks, a number of static 2D visualizations are commonly deployed, such as parallel coordinate plots [22, 62], dendograms [34, 66] and heatmaps [46, 56]. For the analysis of higher dimensional data, scatterplot matrices are a common technique for cluster identificaiton tasks [12, 23].

While various techniques for the visual exploration of previously extracted clusters in scatterplots exist [23, 25], the technique is also deployed for visual identification tasks of clusters. Cavallo et al. [7] propose a framework in which they make use of scatterplots to identify clusters. In their framework, they also deploy other techniques, such as silhouette plots and heatmaps. Etemadpour et al. [15] deployed an eye tracker to monitor user behavior when exploring 2D scatterplots for various tasks. They found that cluster density is more influential than cluster size in cluster identification tasks. They also discuss issues of cluster separation and cluster preservation for deployed dimensionality reduction techniques and their impact on user performance. Therefore, we included cluster density as an experimental side factor.

### 2.2 The Effects of Immersion

Several studies have shown that immersion can have a benefit in different fields, e.g., geology, architecture, and medicine. Examples from these areas show that increased immersion can foster spatial understanding and orientation [47] and increase focus capabilities of users by helping them to fade out distractions [5]. It was also shown that a higher level of immersion can increase task efficiency and effectiveness for spatial problem solving applications [19] and psychological treatment procedures (e.g., phobia treatment) [59]. Positive effects of immersion on learning performance in the context of medical education [20], on memorization [40], as well as for visualizing abstract visualizations [31, 41] have been reported.

In many studies it is just presumed, without elaboration, proof or reference, that VREs convey higher levels of immersion than screen-based mediums. However, to prove this assumption, some metric needs to be introduced measuring immersion. According to Slater et al. [51], immersion can be seen as a rather objective property of a system that introduces a subjective impression of presence to the user. Various researchers intended to measure immersion by quantifying system properties, such as resolution, field of view, degrees of freedom in movement and so on [21, 52]. This is, however, quite hard to quantify and measure. Witmer and Singer [61] propose to measure subjectively perceived presence and infer results back to immersion. In the past, researchers developed and applied several presence questionnaires under that premise [32, 42]. The most established one is the Presence Questionnaire (PQ) from Witmer and Singer [61].

### 2.3 3D Visualizations

Previous work has shown that 3D visualizations are often vulnerable to artifacts caused by the rendering of depth-related information, such as line-of-sight ambiguities, occlusion, and perspective distortion [18, 37]. Depending on the viewpoint, the visualization is distorted differently, hampering cognition and impeding the interpretation of distances and proportions between objects. Therefore, visual variables that perform well in 2D visualization spaces, such as length, size or position, may be less suitable in 3D visualizations due to depth distortion and the missing alignment with respect to a common baseline. Nevertheless, there are several advantages of 3D visualizations in general and, hence, various 3D visualization applications exist in different domains [35, 53]. Multiple studies show benefits of 3D in exemplary visualizations [18, 36], among others with regard to accuracy and efficiency. Moreover, studies indicate that 3D visualizations perform even better when inspected using stereoscopic displays due to a more natural, familiar and accurate perception of information [14].

3D scatterplots are used in various applications for visualizing multidimensional data [37, 43, 45, 65], e.g., to visualize network data [53] or a development over time in space time cubes [17, 35]. Sedlmair et al. [48] compared 2D scatterplots, 3D scatterplots and scatterplot matrices. They examined the effectiveness of these visualizations for separating clusters in datasets that have been transformed with the help of a dimension reduction technique. They found that 2D scatterplots could be used to perform the given task to a satisfactory extent, but that in most cases participants using scatterplot matrices outperformed others using 2D scatterplots. According to them, using 3D scatterplots for the examined task rarely helped, and sometimes even impaired the results. However, they only used data which previously was subject to a dimension reduction procedure and did not evaluate scatterplots in a VRE. Since they assumed that differences in performance between the designs mainly result from the data and not the users, the study was conducted with only two expert users. Each of them inspected and classified 816 scatterplots.

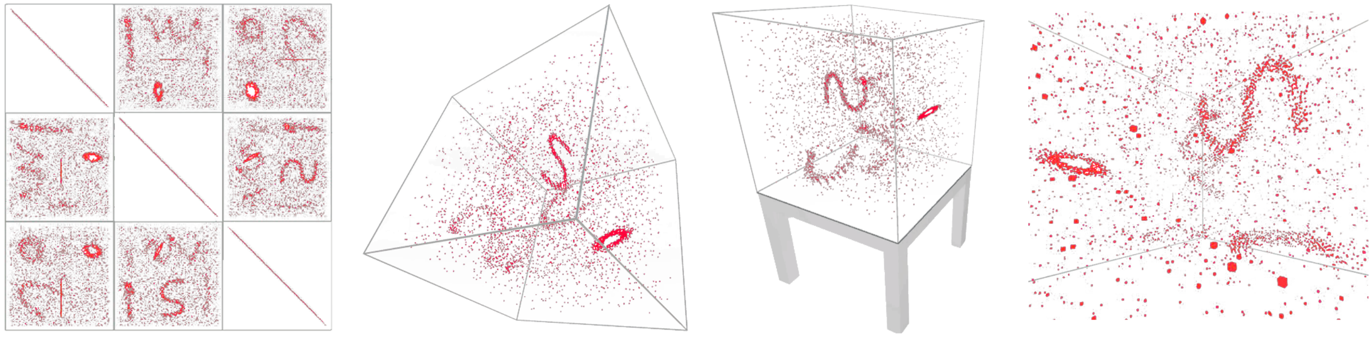
### 2.4 Evaluation of Stereoscopic Visualizations

Wagner Filho et al. compared 2D scatterplots with screen-based 3D scatterplots and VR-based 3D scatterplots [58]. Their tasks included finding nearest neighbors, finding the nearest class, identifying class outliers and comparing two classes to each other. Users in this study were faster using the 2D scatterplot and found it slightly more intuitive for the given tasks. On the other hand, participants were slightly more accurate and subjectively more engaged using the VR scatterplot. In a follow-up study, Wagner Filho et al. [57] present and evaluate an analysis environment in which the user is seated and interacting with scatterplot visualizations using gestures. The authors further investigated user capapabilities to evaluate dimension reduced 3D scatterplot visualizations in immersive and screen-based scenarios [16].

Prabhat et al. [38] conducted a study to evaluate environments differing in their level of immersion by means of different data analysis tasks. However, to the best of our knowledge, there is no study evaluating the impact of the degree of immersion in VREs on user performance during scatterplot analysis. By now, research has not extensively assessed the opportunities and disadvantages of design spaces in VR for abstract visualizations in VA tasks.

## 3 DESIGN SPACES

In this paper, we investigate user performance in a cluster identification task by means of scatterplots in four different visualization design spaces (see concept in Figure 1 and realization in Figure 2). The conducted study solely targets the visual detection of clusters in a dataset visualized as a scatterplot without encoding the cluster membership of data points and compares user performance in different visualization design spaces. In this section, each of the examined design spaces is briefly described. Subsequently, we reason why we chose the presented design spaces. In our basic scientific research approach, we consider three-dimensional data only. In many cases, multi-dimensional datasets can be effectively projected into 3D space using dimension reduction methods (e.g., PCA [27], t-SNE [54]). However, for truly high dimensional data, projections into 2D or 3D space might not be suitable for cluster identification tasks. In cases like that, three dimensions could be compared at a time in small multiple visualizations. We argue that we investigate basic visual perception and the users' capability to identify clusters in three-dimensional datasets. We only deploy the visual variable position and abstain from using additional visual variables (e.g., color, shape) to keep the experiment as simple as possible. Because



a) Screen 2D b) Screen 3D c) VR Table d) VR Room

Fig. 2. Representation of one exemplary dataset in all four investigated visualization design spaces. Except for the scatterplot matrix (a), all visualization design spaces had some kind of navigation available to inspect the visualization from different perspectives.

a maximum of three dimensions can be encoded in 3D visualizations exclusively by position, we focused on the reduction to three dimensions. Consequences of this constraint, in particular with regard to the 2D design space, are discussed in Section 8.

***Screen2D*: 2D on Screen –** The first design space is a 2D space on a monitor screen. To represent three-dimensional data in two-dimensional space, there are at least two intuitive options. One option is to use a dimension reduction technique, map the data to two dimensions, and visualize the resulting projection in a standard two-dimensional scatterplot. We decided against this approach as sometimes clusters vanish in the projection. Figure 3 depicts a two-dimensional projection of three-dimensional data displayed in Figure 2 after a PCA dimension reduction. The visualization demonstrates that, for some use cases, a PCA transformation can be unsuitable for cluster identification tasks. In the given example, only four out of six clusters are clearly distinguishable in the PCA projection (see Figure 3). More advanced dimension reduction techniques, such as t-SNE [54], often require a set of parameters that must be customized for each dataset to result in an optimal representation for the cluster identification task. In the case of our study, this would require additional user interaction and significantly increase interaction efforts for this visualization design space and consequently affect results. Moreover, due to individual adjustments of parameters, results of different participants would not be comparable anymore.

Another option is the display of 3D data in a scatterplot matrix representation, which is often used to visualize multi-dimensional data in various domains [6, 12, 44]. The scatterplot matrix is a projection of high-dimensional data into a 2D representation consisting of small multiples (2D scatterplots). For data with three dimensions (x, y, z), the resulting visualization is a compound of three different scatterplots (x&y, x&z, y&z) and rotated and mirrored versions of them as can be seen in Figure 2a. In our investigations, we chose to use a scatterplot matrix to visualize the data in this design space. *The observer is looking at a static, non-interactive scatterplot matrix on a screen.*

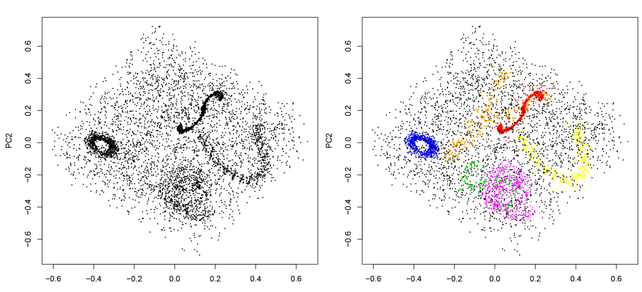



Fig. 3. PCA projection of data displayed in Figure 2. The dataset contains six clusters (highlighted on the right). Two clusters are hardly recognizable in the PCA projection (orange $Y$ and green $S$).

***Screen3D*: 3D on Screen –** The second design space is a 3D space on a monitor screen. The resulting visualization is a virtual 3D cube on a screen, containing the three-dimensional data as a 3D scatterplot. This design space is also frequently deployed in related works [12, 37, 45]. *The observer is looking at a projection of a 3D visualization on a screen, inspecting the data by rotating the scatterplot in arbitrary directions.*

***VRTable*: Miniature 3D in VRE –** The third design space is a restricted 3D space in a VRE. This design space is limited spatially so that the observer is able to walk around the visualization and observe it from outside. The resulting visualization is a 3D scatterplot on a virtual table in a VRE (table height: 75 cm; cube dimensions: 1 m × 1 m × 1 m; data point size: 2.5 mm). *The observer is standing in front of a virtual table with a 3D scatterplot on top of it, inspecting it by walking around the table.*

***VRRoom*: Room-Scaled 3D in VRE –** In the fourth design space, we adjusted the size of the 3D scatterplot to the size of the entire VRE (dimensions: 3 m × 3 m × 3 m; data point size: 7.5 mm). The entire space around the observer is used as visualization design space. *The observer is standing inside the visualization and inspects the scatterplot from within by walking and looking around.*

Design Decisions: In order to investigate the effects of immersion provided by design spaces, we aimed to create several different design spaces with varying degrees of immersion. First, we chose to introduce a 2D design space located on a 2D screen (*Screen2D*). This is a commonly used design space and can be seen as a baseline for the other design spaces. In line with the definition of immersion by Slater [50], we perceive a virtual object as more immersive if it reflects the characteristics of a real object. Therefore, to increase the degree of immersion provided by the design space, the resemblance with real-world objects has to be increased for virtual objects. This can be achieved by using a 3D design space located on a 2D screen (*Screen3D*). Thereby, data points are displayed more “naturally” as the real world is 3D itself. Moreover, with regard to scatterplots, we can easily perceive all three dimensions at the same time in a 3D environment, whereas heavy mental mapping is required to extract all dimensions of a data item from a scatterplot matrix.

Presenting a 3D object on a 2D screen usually introduces perspective distortions [18]. These distortions change how a person perceives the object and, therefore, may reduce immersion as the object reflects the characteristics of a real object to a lesser extent. This effect can be avoided by using VREs. Therefore, we deployed VR in the third and fourth design space. In the third design space (*VRTable*), we introduced the restriction that the user can only observe the visualization from outside and is not able to enter the visualization itself. We argue that this restriction is insofar reasonable as the same restriction applies to the previous design spaces. Removing the restriction (*VRRoom*) may lead to an even more increased level of immersion as the user enters the visualization itself and is fully enclosed by it.

In order to validate our hypothesis that the level of immersion increases in each design space (see Figure 2, a to d), we conducted a supplemental sub-study, in which we investigated solely this specific issue (Section 4).

## 4 Pre-Study: Levels of Immersion

Among others, the property *level of immersion* discriminates visualization design spaces. According to Slater [50], immersion describes how much a system preserves the fidelity of sensory modalities. To confirm differences between the proposed design spaces (presented in Section 3) with regard to their level of immersion, we conducted a pre-study. As it is hard to directly measure the properties of the system, we rely on the approach of Witmer and Singer [61], i.e., measuring presence and referring it back to immersion. In this pre-study, we evaluated participants' level of self-reported immersion for each design space. Further subjective observations, opinions and perceptions of participants concerning the design spaces (e.g., abstractness, preference) were gathered.

### 4.1 Study Description and Hypothesis

The only experimental factor of this study was *visualization design space*. All four design spaces introduced in Section 3 were examined by means of a within-subjects design. We hypothesize that the design spaces can be sorted by their level of immersion as follows: *Screen2D* is the least immersive design space, followed by *Screen3D*, *VRTable* and *VRRoom*. As there was no reason to assume an impact on participants' physical or mental health, no institutional review board (IRB) was consulted for the study. Also, the participants could abort the study at any time.

After a training session in all design spaces, 12 participants (six female, six male) conducted one cluster identification task in each design space. The order of designs and used datasets were counterbalanced. All datasets had similar properties, contained between five and seven clusters and were enriched with the same amount of noise. Subjects were asked to identify and count all clusters in the data and to report their result to the examiner. After each of the four trials, participants completed a questionnaire. Both a multiple measures questionnaire for immersion (consisting of 18 questions) and a single measure of immersion (consisting of one question) were used. The first question served as a single measure of the subjectively perceived immersion in the respective design space: "How immersed did you feel in the virtual environment?" (see Appendix A.4.1). The following set of 18 questions were adopted from questionnaires by Regenbrecht et al. [42] (IPQ), Witmer and Singer [61] (PQ), Lessiter et al. [32] (ITC) and Jennett et al. [26] (IEQ). We carefully selected questions that fit all design spaces as well as the current task. Therefore, we excluded, for instance, questions that are explicitly aimed at gaming experiences in VREs. The resulting questionnaire we used is attached in the Appendix (A.4.2). After the completion of all four trials, we conducted a semi-structured interview (see Appendix A.4.3 for the structure of the interview). At the end of the experiment, participants received 10 € as compensation. The apparatus of this study was similar to the one in the main study described in Section 5.4.

### 4.2 Results of Pre-Study

#### 4.2.1 Questionnaires

Statistical tests were performed using IBM SPSS Statistics (version 24). In this section, we only report significant results. A Bonferroni correction was applied (to control for multiple testing) and, hence, all effects are reported at a .008 level of significance. A detailed overview of all results can be found in the Appendix (A.4).

To evaluate differences in the level of immersion between design spaces with regard to the single measure of immersion, a non-parametric Friedman test was deployed ($\chi^2(3) = 22.49$, $p < .001$). We used a non-parametric test because of skewed distributions. Wilcoxon signed-rank tests were computed as post hoc tests to follow up this

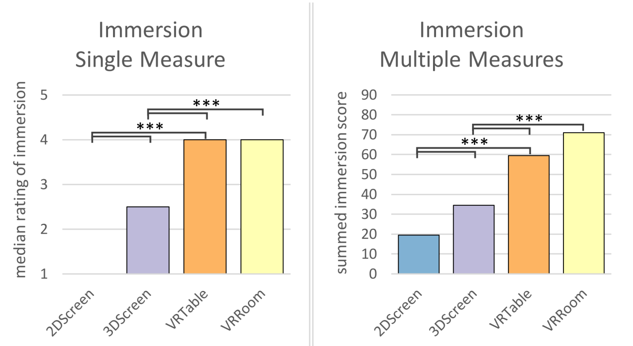



Fig. 4. *Measures of Immersion* – Left: single measure question of subjectively perceived immersion. Right: multiple measure questionnaire on immersion.

finding (see Figure 4 left). The post hoc tests revealed that the subjective experience of immersion was significantly lower in the *Screen2D* design space ($Mdn$ = 1.00) as well as in the *Screen3D* design space ($Mdn$ = 2.50) compared to both VR spaces, namely *VRTable* ($Mdn$ = 4.00) and *VRRoom* ($Mdn$ = 4.00): *Screen2D-VRTable*: $z = -2.84$, $p$ = .001; *Screen2D-VRRoom*: $z = -2.75$, $p$ = .002; *Screen3D-VRTable*: $z = -2.85$, $p$ = .001; *Screen3D-VRRoom*: $z = -2.61$, $p$ = .003.

For the multiple measure of immersion (i. e. the immersion questionnaire), immersion scores were computed by summing up participants' responses to all 18 questions. The same statistical approach was used as for the analysis of the single measure of immersion ($\chi^2(3) = 24.23$, $p < .001$). As depicted in Figure 4 (right), Wilcoxon signed-rank tests showed that the level of immersion was significantly lower in both the *Screen2D* design space ($Mdn$ = 19.50) and the *Screen2D* design space ($Mdn$ = 34.50) than in the two VR spaces, namely *VRTable* ($Mdn$ = 59.50) and *VRRoom* ($Mdn$ = 71.00): *Screen2D-VRTable*: $z = -2.90$, $p$ = .001; *Screen2D-VRRoom*: $z = -2.90$, $p$ = .001; *Screen3D-VRTable*: $z = -2.94$, $p < .001$; *Screen3D-VRRoom*: $z = -2.87$, $p$ = .001.

#### 4.2.2 Interview

The evaluation of the interview questions on abstraction and presence, which can be regarded as substitute variables for immersion [61,63], revealed the predicted order of design spaces (see median user ratings depicted in Figure 5, left and center). Participants perceived the VR design spaces as less abstract and therefore more natural compared to the two screen-based ones (*Screen2D*: $Mdn$ = 5; *Screen3D*: $Mdn$ = 2.5; *VRTable*: $Mdn$ = 2; *VRRoom*: $Mdn$ = 2). Particularly for the subjective user rating of how present they felt in the respective design space, the assumed pattern emerged (*Screen2D*: $Mdn$ = 1; *Screen3D*: $Mdn$ = 2; *VRTable*: $Mdn$ = 4; *VRRoom*: $Mdn$ = 5).

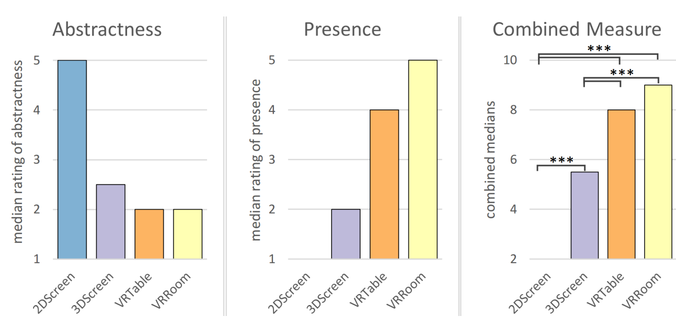



Fig. 5. *Interview* – Median user ratings for the design spaces with regard to abstractness (left) and presence (center). Participants were asked to rate the abstractness and presence of each design space on a five-point Likert scale from 1 = *not abstract/not present* to 5 = *very abstract/very present*. Right: Combined median of the abstractness and presence scores used for statistical evaluation and as a measure of immersion.

We conducted a Friedman test ($\chi^2(3) = 30.28$, $p < .001$). Bonferroni-corrected Wilcoxon signed-rank tests revealed significant differences between all design spaces: *Screen2D-Screen3D* ($z = -2.89$, $p = .001$), *Screen2D-VRTable* ($z = -3.27$, $p < .001$), *Screen2D-VRRoom* ($z = -3.28$, $p < .001$), *Screen3D-VRTable* ($z = -2.75$, $p = .002$), *Screen3D-VRRoom* ($z = -2.97$, $p = .001$), *VRTable-VRRoom* ($z = -2.67$, $p = .005$).

These results are supported by the interview question, in which the participants were asked to sort the design spaces by the amount of perceived presence. All subjects put *VRRoom* in first place ($n = 12$) and *Screen2D* last. Only one participant put *Screen3D* in second place and *VRTable* in third place – all others put the design spaces in the expected order.

### 4.3 Conclusion

Overall, our pre-study supports the previously stated hypothesis and verifies the assumed order of design spaces with regard to the level of immersion:

$$Screen2D < Screen3D < VRTable < VRRoom.$$

## 5 Main Experiment

As shown in previous research, the degree of immersion can have an effect on spatial cognition and memorability in various contexts [11, 33]. Some studies even indicate correlations between the degree of immersion and efficiency in cluster identification, distance estimation, and outlier detection tasks in scatterplot visualizations [4,41]. However, many existing studies use a variety of different interaction techniques individually for each design space, disguising possible effects caused solely by characteristics of the different design spaces. In order to avoid possible confounding factors resulting from different interaction techniques, we limited our study to an absolute minimum of interaction techniques. No institutional review board (IRB) was consulted for the study as there was no reason to assume any impact on participants' physical or mental health. Participants could end the study at any point.

### 5.1 Study Design

Our main experimental factor was the *visualization design space*. Besides, we investigated the impact of noise level, cluster shape, and cluster density. All three study side factors (noise, shape, density) were introduced to examine if the designs are differently robust to dataset characteristics and to ensure that our main results are generalizable to different kinds of datasets. A prototype, developed specifically for the purpose of this study, was used for the execution of the study.

Visualization Design Space: As main experimental factor the design spaces introduced in Section 3 were examined. In each design space, an adaption of a scatterplot was displayed (scatterplot matrix, 3D scatterplot). Figure 2 shows one exemplary dataset in all four designs.

Noise Level: The first additional experimental factor was the level of introduced noise. With regard to the noise level, two kinds of datasets were generated. One contained 1,000 additional randomly positioned points (low noise level), the other one 3,000 additional noise points (high noise level).

Cluster Shape: As a second additional experimental factor, the shape of clusters was manipulated. Half of the datasets contained convex clusters (spheres, capsules, discs), the other half contained non-convex clusters (spirals, donuts, y-shapes, s-shapes, sinus-curved pipes). We statistically counterbalanced noise level and cluster shape, i.e., all four possible combinations (low-noise & convex, low-noise & non-convex, high-noise & convex, and high-noise & non-convex) occurred equally often.

Cluster Density: The third additional experimental factor was cluster density. Two different types of clusters were created with regard to cluster density. We used the DBSCAN algorithm, introduced by Ester et al. [13], as a measure to distinguish between dense and sparse clusters. For sparse clusters the parameters $t_1 = \{ MinPts = 10, \varepsilon = 0.15 \text{ m} \}$ were used as thresholds, and for dense clusters $t_2 = \{ MinPts = 30, \varepsilon = 0.10 \text{ m} \}$. The two parameter sets were systematically refined during several trial dataset generation procedures to generate two visually distinguishable types of clusters. The clustering was performed in a cube with the dimensions 2 m × 2 m × 2 m. With the lower threshold $t_1$, all clusters, but nothing else, should be found by the DBSCAN algorithm, and with the higher threshold $t_2$, solely all dense clusters should be found. In contrast to the other three side factors, each dataset contained both sparse and dense clusters at the same time. However, the error rate was measured separately for both types of clusters.

### 5.2 Procedure

The experiment was structured in four blocks (see Appendix B.1). Each block was dedicated to one visualization design space (*Screen2D*, *Screen3D*, *VRTable*, *VRRoom*). In each block, the participant completed four trials by pointing to all clusters found and reporting the overall count to the study supervisor. Each trial had a different dataset. The order of blocks was structurally alternated with the only constraint that always the two screen-based and the two VR design spaces were directly after each other. We chose to introduce this constraint because pretests showed that some participants experienced varying levels of discomfort after switching in or out of the VRE. Half of the participants started with VR design spaces, half of them with screen-based ones. Participants were systematically assigned to one order.

At the beginning of the experiment, written informed consent was obtained from the participants and they were asked to fill in a questionnaire assessing demographic variables. After that, participants completed four blocks, each beginning with a training session for the respective visualization. A total of three practice trials had to be completed before the first real trial of the block could start. In each trial block, participants completed eight tasks. At the end of the second block, participants were again asked to fill in a brief questionnaire examining participants' memory of the last completed trial (see Appendix B.4.1). A third questionnaire was administered after the last block, collecting information about personal preferences and subjective opinions about the four visualizations (see Appendix B.4.2). Finally, participants were thanked and received a monetary compensation for participating (10 €). During the experiment, sound, video, and position data were recorded.

### 5.3 Data and Task

Sedlmair et al. [49] proposed a taxonomy of visual cluster separability factors in scatterplots. They describe various factors of clusters, such as shape, size, or number of items, that affect the observer's capability to identify the centroid of each cluster in dimensional reduced datasets presented as 2D scatterplots. For the generation of our study datasets, we varied the identified variables shape, size and density. We first created a set of 16 different clusters. In this context, we refer to clusters as areas with a higher density of data points compared to the surrounding areas. In order to guarantee the cluster property and also a consistency over all clusters, we applied the DBSCAN algorithm after creating the clusters interactively in Unity. Two different types of clusters were prepared with regard to density. Dense and sparse clusters had to be found as only clusters by the DBSCAN algorithm with a certain parameter set ($MinPts = 10$, $\varepsilon = 0.15$). Sparse clusters must not be found using another parameter set ($MinPts = 30$, $\varepsilon = 0.1$), which should only detect all dense clusters (see Section 5.1).

Finally, we generated 32 study datasets as compositions of rotated and flipped versions of the previously created clusters. Additionally, we created a set of 20 extra datasets for training trials. Subsequently, we added a certain amount of noise to each dataset (50% of the datasets with high noise level). Half of the datasets contained only convex clusters, and the other half only non-convex shaped ones. Each dataset was constructed carefully so that all clusters were potentially identifiable

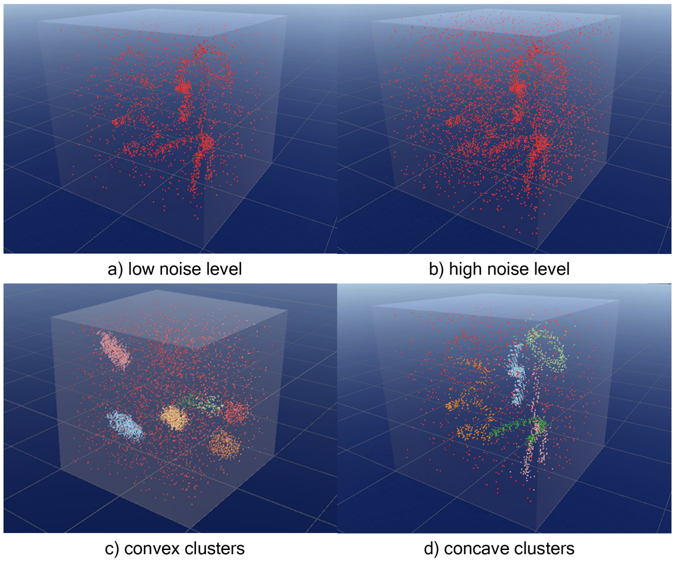



Fig. 6. Four sample datasets illustrating different properties. Top: low noise condition (left) and high noise condition (right). Bottom: convex clusters condition (left) and non-convex clusters condition (right).

in the scatterplot matrix (i.e., no cluster was occluded in all views). Exemplary datasets from both conditions are depicted in Figure 6.

Although all our datasets were created with three attributes (one coordinate each for the *x-axis*, *y-axis*, and *z-axis*), we do not see this as a limitation of our study. Higher dimensional data can be transformed into 3D data by projection techniques like a PCA. However, the type of projection and its settings has a major impact on how well clusters can be identified in the resulting visualization. For this study. Therefore, we created datasets natively in three dimensions and abstained from deploying dimension reduction techniques as it is common practice in real-world applications. We only aim to investigate the effects of immersion provided by the design spaces, which should not be affected by a preceding data transformation step.

For the entire experiment, the task performed by participants remained the same, even though interaction and visualization techniques differed. The task was to identify clusters in a scatterplot visualization, to point at them, and to count up all clusters. Participants were asked to point at found clusters (with the mouse or VR controller) and report their detection to the study supervisor. At the end of each trial, they indicated the overall count of found clusters.

### 5.4 Apparatus

The experiment took place in a quiet, closed room at the University of Konstanz. Participants were individually invited to the laboratory. Besides the participant, the examiner was the only person present. During two blocks (screen-based visualization design spaces), the participants sat in front of a 24" monitor with a resolution of $1920 \times 1200$ pixel. In those blocks, participants interacted with the study software solely with the mouse as input device. During the remaining two blocks, participants were equipped with a Vive HMD and one Vive controller as a pointer. In those blocks, participants were initially positioned at a specific starting point. During the task, they were allowed to walk freely through the room within the bounds of the virtual environment (which were visually highlighted in the VRE as blue walls). In the *VRTable* visualization design space, participants were additionally instructed not to walk into the virtual table.

### 5.5 Sample

A sample of $N = 18$ participants (5 female, 13 male) was recruited using short notices distributed around the university. Most of the participants had none or only little experience with scatterplot matrices (66.6%), but had experienced a VRE at least once before (72.2%). We introduced a training phase at the beginning of each of the four blocks in order to minimize any effects resulting from different levels of experience. Participants were aged 19 to 41 years ($M = 26, SD = 4.87$). Three participants were still in high school, nine held a Bachelor's degree, and six a Master's degree. The background of the participants was quite diverse with eight having a computer science background (44%) and the rest from various domains without advanced computer science knowledge.

### 5.6 Dependent Variables

To compare differences caused by changes in the independent variables (visualization design space, noise level, cluster shape, and cluster density), we analyzed multiple dependent variables. For each trial, the error rate was calculated as the percentage of clusters not found. All trials were recorded for later video analysis to count errors and find frequent patterns in participant behavior. Participants were instructed to point at identified clusters throughout the trials using the mouse (screen) or the laser pointer attached to the Vive controller (VR). After the study, we analyzed the recordings by coding which clusters were found in each trial. All videos were encoded by at least two people to avoid counting errors. In addition, the task completion time was logged. For all VR trials, the VR headset was tracked (head position and orientation). Besides, two questionnaires were issued gathering information about personal preferences and the memorability of data (count, shapes, and positions of clusters) in a previously completed task. An overview of all gathered data is provided in the Appendix (B.2).

### 5.7 Hypotheses

Based on subjective indications from two exploratory pilot studies and in part based on results of studies presented in the related work section, we derived the following hypotheses. All hypotheses refer to the deployed tasks and variants of scatterplot visualizations.

**H1** VR vs. Screen – Error Rate: We expect error rates to be lower when participants work with VR visualization design spaces. This hypothesis is based on Filho et al. [16] and Arns et al. [2] experiments on the analysis of multi-dimensional data in 3D scatterplots. They report on beneficial effects of immersion with regard to distance and structure perception. Both properties are crucial for cluster identification.

**H2** VR vs. Screen – Task Completion Time: Participants will be more active and need more time to complete the task when working in VR visualization design spaces compared to them working in screen-based ones. Bach et al. [4] came to the conclusion that participants need more time in AR environments because they move more, take extra time to explore the visualization, and are new to the device. We expect similar findings in our VR settings.

**H3** VR vs. Screen – Memorability: Participants will show better memory performance when working in VR design spaces compared to them working in screen-based design spaces. Previous studies have shown that in certain VR scenarios the spatial memory is crucially better compared to applications on the screen due to a more natural navigation [10].

**H4** VR vs. Screen – Subjective Preference: Visualizations in VR visualization design spaces will come more naturally to the participants than the ones in screen-based design spaces. This hypothesis is based on the assumption that the level of abstraction of VR visualizations should be relatively small as, for instance, distances can be measured in "real" measures such as inches or centimeters. Additionally, the possibility to navigate the data space like in the real-world (e.g., walking around or rotating the head) is expected to increase the engagement of participants [4].

**H5** Full Environment vs. Restricted Area – Error Rate: Comparing the VR visualization design spaces, participants will perform worse in the totally immersive design space (*VRRoom*) compared to the *VRTable* design space with regard to the error rate. This hypothesis is based on the assumption that participants will miss clusters due to blind spots (clusters behind, underneath or above the observer) or a possible loss of orientation due to the missing overview as reported by Etempadpour et. al. [14].

## 6 Results of Main Study

We report significant results of our quantitative analysis, as well as qualitative feedback.

### 6.1 Statistical Analysis

All statistical tests were performed using IBM SPSS Statistics (version 24) and are based on a significance level of $\alpha = .05$. To evaluate differences between the visualization design spaces related to the error rate, i.e., the percentage of clusters not identified, a Friedman test was used. Due to serious violations of assumptions, in this case we have decided against an ANOVA and opted for its non-parametric counterpart. Wilcoxon signed-rank tests were computed as post hoc tests. Moreover, a one-way repeated measures ANOVA was applied to compare the time participants required for performing the task (completion time). Mauchly's sphericity test was used to confirm the sphericity assumption needed for a one-way repeated measures ANOVA.

In case of a significant omnibus $F$-test, we report the results of Bonferroni-corrected pairwise comparisons. Finally, head rotation data were analyzed using a paired samples $t$-test. Note that time data and head rotation data were log-transformed because of skewed distributions. Shapiro-Wilk tests were used to check the assumption of normality after the log transformations and before the $t$-tests.

### 6.2 Error Rate

Error rates differed significantly between the visualization design spaces ($\chi^2(3) = 40.67$, $p < .001$). As depicted in Figure 7 (left), Wilcoxon signed-rank tests revealed that with regard to the error rate participants performed significantly worse in the design space *Screen2D* ($Mdn = 16.67\%$) compared to all other design spaces: *VRTable* ($Mdn = 0\%$, $z = -4.87$, $p < .001$), *VRRoom* ($Mdn = 0\%$, $z = -4.57$, $p < .001$) and *Screen3D* ($Mdn = 14.29\%$, $z = -3.95$, $p < .001$).

When also taking noise into account, there was a significant difference in error rates between the low noise ($Mdn = 8.45\%$) and the high noise condition ($Mdn = 12.47\%$; $z = -2.24$, $p < .025$). For each visualization design space, error rates increased with an increasing noise level. However, the resulting change differed between the design spaces: The difference in error rates between the low noise and the high noise condition was 6.54% for *Screen2D*, 1.3% for *Screen3D*, 5.74% for *VRTable* and 5.54% for *VRRoom*. Statistical tests showed significant differences between noise conditions in both VR design spaces (*VRTable*: $t(17) = -2.27$, $p < .05$, $r^2 = .23$; *VRRoom*: $t(17) = -2.19$, $p < .05$, $r^2 = .22$).

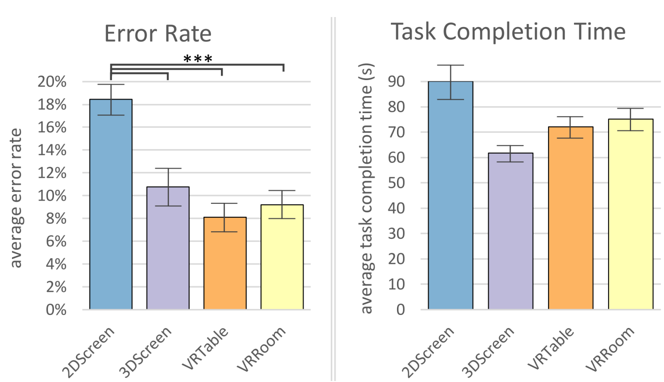



Fig. 7. Average error rate and completion time as a function of visualization design space. Bars indicate the 95% CI of the mean, asterisks significant differences between design spaces (*** $p \leq .001$). Note that for statistical analysis task completion times were log transformed because of skewed distributions, while in this figure original data is displayed.

With regard to the side experimental factors cluster shape and cluster density, no significant differences emerged with respect to the error rate.

### 6.3 Task Completion Time

As depicted in Figure 7 (right), the average completion time in the *Screen3D* design space ($M = 61.79$ s) was the lowest, followed by *VRTable* ($M = 72.14$ s), *VRRoom* ($M = 75.19$ s) and *Screen2D* ($M = 90.12$ s). Task completion times differed significantly between the four design spaces, $F(3,51) = 4.4$, $p < .01$, $\eta_p^2 = .206$. Bonferroni-corrected post hoc tests were applied. After correcting for alpha error accumulation, none of the pairwise comparisons reached significance. The experimental side factor noise level had no significant influence on task completion time.

### 6.4 Memorability

In the questionnaire which was administered after the second block participants were asked to recall the count, shapes, and positions of all clusters in the last completed trial. Results show that, with regard to the error rate, participants performed better in the *VRTable* design space ($M = 0\%$) compared to all other design spaces (*Screen2D*: $M = 43.33\%$; *Screen3D*: $M = 32.67\%$; *VRRoom*: $M = 20.42\%$). The percentages reflect how many clusters of the previously found clusters could not be remembered with the correct shape. To prevent training effects, each participant performed the memory task only once. Therefore, the sample size per design space is rather small ($n \approx 4$).

### 6.5 Subjective Preference

As part of the final questionnaire, participants were asked to rank the visualizations by difficulty (1 = *easy* to 4 = *hard*). As Figure 8 depicts, ranks assigned to visualizations in the *Screen3D* design space show a positive skewness (ranks 1 & 2: 61.1%; ranks 3 & 4: 39.9%). To visualizations in the *Screen2D* design space, participants only assigned the lowest ranks 3 (22.2%) and 4 (77.8%). To visualizations in the *VRRoom* design space, mainly middle ranks were assigned (rank 1: 11.1%; ranks 2+3: 88.9%). The distribution of visualizations in the *VRTable* design space is positively skewed with the mass center on the upper ranks (ranks 1 & 2: 83.3%; ranks 3 & 4: 16.7%).

In accordance with these results, 50% of the participants mentioned the *VRTable* design space as their preferred design space, 33.3% the *Screen3D* design space and 16.7% the *VRRoom* design space. In contrast, none of them indicated the *Screen2D* design space as their preferred visualization design space.

Regarding disadvantages and opportunities perceived by participants, several findings emerged. As benefits of VR visualizations, participants rated VR design spaces to be more comprehensive ($n = 8$), intuitive ($n = 5$) and to provide a better overview when the visualization is inspected from outside (*VRTable*, $n = 2$). Moreover, participants mentioned that naturally changing the perspective (moving the head) helps to grasp the visualization ($n = 4$). As drawbacks, participants mentioned poor overview in the *VRRoom* environment ($n = 9$), increased expenditure of time ($n = 4$) and expensive hardware ($n = 2$).

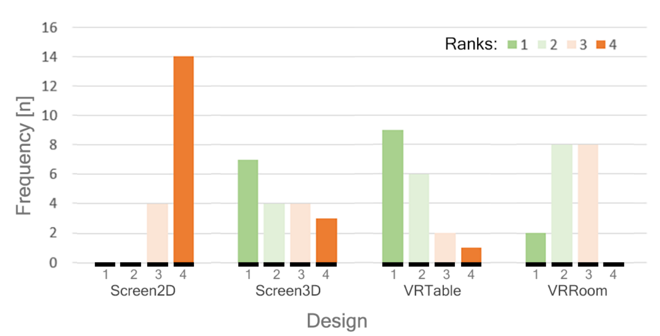



Fig. 8. Subjective preference: ranks (1 = *easy* to 4 = *hard*) assigned to the four visualization design spaces by the participants.

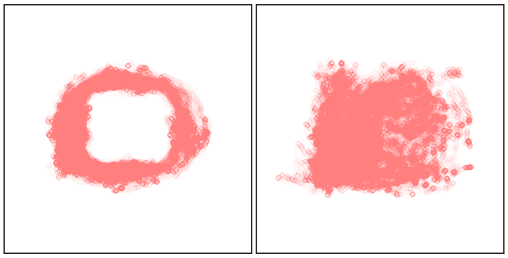


Fig. 9. Top-down view on the VRE. Participants' movements while solving study trials in the design space *VRTable* (left) and *VRRoom* (right). In the *VRTable* environment, participants were explicitly asked not to walk into or through the virtual table. Except for this, the area covered is approximately the same. However, in the *VRRoom* environment, participants covered roughly twice as much distance compared to the *VRTable* environment.

### 6.6 Space Utilization and Motion

Except for the area in which the table was located (i.e., the area participants were instructed not to cross), participants used approximately the same amount of space in the two VR visualization design spaces (see Figure 9). However, total walking distances varied significantly between the design spaces, $t(17) = -8.80$, $p < .001$, $r^2 = .82$. In the fully immersive environment (*VRRoom*), participants covered considerably more distance ($M = 32.20$ m, $SD = 0.14$) compared to the less immersive design space (*VRTable*: $M = 16.77$ m, $SD = 0.18$).

The two VR environments varied significantly concerning the head rotations of participants, $t(17) = -8.80$, $p < .001$, $r^2 = .82$. In the fully immersive environment (*VRRoom*), participants tended to look around much more ($M = 32196°$, $SD = 0.03$) compared to the less immersive design space (*VRTable*: $M = 16771°$, $SD = 0.04$).

### 6.7 Video Analysis

In order to evaluate the experimental trials, we manually examined the videos of each trial. While watching the videos, not only participants' final answer, but also observations throughout the entire task were noted down in a database. We identified several mistakes that were made repeatedly by participants. In particular, four frequent scenarios could be observed: (1) the participant "finds" a cluster twice (double count), (2) the participant finds all clusters, but skips one in the final counting, (3) the participant counts a sparse and a dense cluster as one and (4) the participant detects a sparse cluster, but neglects it as noise. For each of these scenarios, we manually counted the number of occurrences.

The final comparison revealed that all double count-errors were made in screen-based design spaces (*Screen2D*: $n = 7$, *Screen3D*: $n = 2$). Moreover, most adding up-errors (missing to count a cluster in the end) appeared in screen-based design spaces (*Screen2D*: $n = 7$, *Screen3D*: $n = 4$, *VRTable*: $n = 2$, *VRRoom*: $n = 3$). Only in the *Screen2D* design space, it occurred that participants counted a sparse and a dense cluster as one ($n = 2$). Mainly in VR-based design spaces, participants tended to neglect detected sparse clusters as noise (*Screen2D*: $n = 1$, *Screen3D*: $n = 2$, *VRTable*: $n = 1$, *VRRoom*: $n = 5$).

## 7 Discussion

In this section, the results, as well as their implications, are discussed with the focus being on accuracy, efficiency, memory, and orientation. In the course of the discussion, we will address all hypotheses.

### 7.1 Accuracy

H1 implies that in cluster identification tasks error rates are directly influenced by the degree of immersion present in the respective design space when comparing screen-based design spaces with VR-based ones. Specifically, we assumed that participants perform better in design spaces characterized by higher immersion levels (VR-based design spaces). The results partially correspond to that assumption. In case of the *Screen2D* design space, significant differences emerged. As a basis for this hypothesis it was suggested, among other things, that VR visualizations come more naturally to participants in comparison with abstract visualizations or non-stereoscopic 3D visualizations on the screen (cf. H4). Video analysis revealed that situations containing a loss of orientation or navigational problems mostly occurred in the screen-based design spaces. This indicates improved navigation and orientation capabilities in VREs, which again could be due to a better spatial memory (see Section 7.3). Hypothesis H4 is also strongly supported by qualitative feedback. Multiple participants stated to prefer VR visualizations due to a more comprehensive and intuitive representation of the data. Moreover, participants tended to classify VR visualizations as rather easy to work with compared to screen-based ones (in particular, the scatterplot matrix was frequently rated as the most difficult visualization). For the two VR visualization design spaces differed, no significant difference was found in terms of accuracy. Hence, we cannot confirm hypothesis H5.

### 7.2 Efficiency

Contradicting hypothesis H2, no significant differences emerged between pairwise-compared visualization design spaces in terms of task completion time. Nevertheless, the hypothesis can partially be accepted as the statistical analysis revealed a main effect of visualization design space on task completion time and an almost significant difference between the design spaces *Screen3D* and *VRRoom*. Moreover, the average completion time in the design space *Screen3D* was lower compared to the average completion time in both VR design spaces. One reason for that could be the requirement for the user to be more active in VR design spaces. Instead of sitting in front of a computer screen and operating a mouse, the participant had to move and look around. Tasks in the *Screen2D* scenario required on average much more time than in all other design spaces. This could partially be due to a high learning curve for scatterplot matrices due to small multiples. Participants had to mentally match data points in different visualizations in order to avoid counting a cluster twice or missing one. However, the evaluation of participants familiar with scatterplot matrices did, as well, not reveal a difference. Corroborating the second part of the hypothesis (activity of participants), the total walking distance and the total head rotation differed significantly. The means of both attributes are on average approximately twice as large in the *VRRoom* design space. Possible reasons can be derived from video analysis and user feedback. Participants had to change their position more often in the *VRRoom* design space in order to prevent occlusion or blind spots and they had to turn their heads 360 degrees in order to observe the entire visualization space. One trade-off of the "natural" navigation in VR design spaces is the necessary activity compared to conventional mediums. Especially for long sessions, the increased physical effort could lead to fatigue, which in turn could affect accuracy and efficiency. Therefore, if using VR design spaces, present findings suggest favoring the *VRTable* design space as it minimizes the required physical activity.

### 7.3 Memory and Orientation

Participants performed better with regard to memorizing previously identified clusters in VR visualization design spaces compared to screen-based ones. In the *VRTable* scenario, participants had the least difficulties remembering all clusters and their shapes correctly. Moreover, video analysis revealed that more memory-related errors, such as double counts or missing counts, occurred within screen-based design spaces. Therefore, H3, which states an advantage of VR visualization design spaces in terms of memorability, can be considered confirmed.

After working with the abstract visualization (scatterplot matrix), participants had most difficulties to recall all found clusters. We assume that the higher level of abstraction compromises users' orientation capabilities, as building a mental model of the small multiples is necessary to notice connections between clusters in different windows of the scatterplot matrix (e.g., to find one cluster in all views). An increased level of difficulty, accompanied by the requirement for a mental model, is also evident from user feedback. Participants voted the *Screen2D* design space to be the most difficult and least preferred design space.

## 8 Limitations and Generalizability

Some limitations need to be taken into account. It is discussable whether and to what extent our findings are generalizable and transferable to other visualizations in the given visualization design spaces. We argue that most of the findings rather refer to properties of the design spaces than to characteristics of the individual visualizations (e.g., immersion, spatial memory, orientation, or navigation). Nevertheless, it has to be investigated if found distinctions between design spaces also emerge if alternative visualizations are employed. Changing the type of visualization or allowing more advanced interaction techniques might redistribute assigned characteristics to the visualization design spaces and influence final outcomes.

The *Screen2D* visualization design space is fundamentally different from the other design spaces hampering pairwise comparisons. The scatterplot visualization in the *Screen2D* design space is fixed to a certain viewpoint and does not provide any interactions aside from pointing on clusters. During the generation of datasets, we made sure that every single cluster is potentially detectable in the scatterplot matrix visualization as well and avoided pairs of clusters that overlap in all small multiples of the matrix. Additionally, the data used for the experiment was three-dimensional. In the all 3D design spaces (*Screen3D*,*VRTable* and *VRRoom*) the data was visualized in its natural space whereas in the *Screen2D* scenario, multiple 2D scatterplots had to be displayed to compensate for the third dimension.

Another limitation of the present study is the exclusive deployment of a cluster identification task. Compared to the *VRTable* design space, the *VRRoom* design space helps to reduce occlusion since the visualization occupies the entire virtual environment of the observer. However, this comes at the price of tremendous overview loss. These properties likely have a different impact on cluster identification tasks compared to other visual analytics tasks. Future studies should investigate whether a combination of the *VRRoom* and *VRTable* design spaces are preferable for specific tasks. One has to keep in mind that excessive interaction and switching between the two design spaces could impair some of the benefits, such as improved spatial memory capabilities. For 3D scatterplots, Yu et al. [64] presented a toolset of effective selection techniques in 3D pointclouds. In future research, such advanced techniques for the accurate selection of clusters could be implemented to assess if participants found the entire cluster. Also, advanced techniques that support the detection of clusters could be deployed, such as highlight-planes presented by Prouzeau et al. [39]. Besides the impact of interaction, it would be interesting to assess properties of the screen deployed. For instance, a larger screen with higher resolution might lead to higher levels of perceived immersion and increase task performance.

A larger sample size would have been beneficial to assess every experimental side factor accurately. However, we argue that the experimental side factors (noise, shape, density) were mainly deployed to guarantee the stability of results in the analysis of the visualization design spaces. We analyzed them as additional factors, but set the focus on the comparison of results for different visualization design spaces. Even though there was an exhaustive training session, and statistically no difference between experts and non-experts emerged, different outcomes could have emerged if we had conducted the study only with experts. We deployed two different kinds of datasets with regard to their noise level, much higher levels of noise could have changed the performance of users differently in each visualization design space.

One major limitation of our study is the restriction to three-dimensional data, favoring 3D design spaces, and thereby introducing a bias. However, we argue that our foundational research is targeting cases where dimension reduction to two dimensions is impossible or not advisable (e.g., see Figure 3). For truly high dimensional data a projection to 3D space might not make sense for cluster identification. We chose to focus on three dimensions as this is the maximum number of dimensions that can be encoded by the visual variable ‘position’ at a time in all deployed design spaces. However, this favors the 3D scatterplot visualizations as, for instance, if more than three dimensions had been represented in the visualization, a scatterplot matrix would have outperformed the three-dimensional scatterplots due to its dynamic scalability with regard to the number of dimensions. In addition to that, the 2D design space was disadvantaged as a higher learning curve can be expected for scatterplot matrices. More than half of the participants had none or few experience with scatterplot matrices.

Especially in the domain of molecular biology, 3-dimensional representations of molecular surfaces are often used, e.g., to investigate the size of genes, to compare proteins, or to identify substructures in electron tomography [28]. These spatial structures are comparable to point clouds visualized in 3D scatterplots. Therefore, we expect our results to be also true for similar tasks in such settings. Similarly, our findings could be applicable for applications with flow visualizations [55], spatio-temporal visualizations [1] and graph visualizations [8] in which entities have to be identified in a large 3D environment. Although not significant, participants performed better in the *VRTable* condition compared to *VRRoom*. When analyzing the subjective feedback, participants reported that they were missing an overview of the data when being entirely immersed in the *VRRoom* design space. We expect this circumstance to be independent of the visualization technique used. As a consequence, researchers should think about techniques to provide an overview of the data in VR environments.

To generally assess the possibilities of abstract visualizations for VA purposes in VR, future research should compare specific scenarios (task + visualization + data) in various visualization design spaces. The ultimate goal would be to establish some rules of thumb, advising one to avoid certain VA tasks and visualizations in VR design spaces and to favor the usage of others.

## 9 Conclusions

We presented a user study with 18 participants examining differences between four visualization design spaces with regard to cluster identification in scatterplots. The four employed visualization design spaces differed in their degree of immersion as confirmed by an additional study. Two of the design spaces were observed using a standard computer monitor (2D and 3D spaces on screen) and two using VR HMDs (restricted area in VRE and entire VRE). While the results show that more immersive visualization design spaces generally fit better to the given task, a fully embracing analysis environment may not be the best choice for scatterplot analysis due to a lack of overview and blind spots. Hence, for cluster identification tasks in scatterplots, results suggest favoring a restricted area in a VRE as visualization design space. It is difficult to give a general recommendation when to use screen-based design spaces and when to deploy HMDs. We found that for scatterplot visualizations it can be beneficial to convey information by using three-dimensional VR design spaces if the task is to identify clusters in three-dimensional data. Results imply that thereby memory and orientation capabilities are increased. In comparison to abstract representations, 3D visualizations tend to be more comprehensive (maximally by using stereoscopic perception) and therefore ease the identification of clusters. However, abstract visualizations deliver more detail on single points or groups of points as extracting exact information from 3D visualizations can be difficult for humans due to distortion and a missing common baseline for comparing values that refer to multiple axes. Overall, we can state that VREs can indeed provide suitable design spaces for abstract visualizations such as scatterplots. Moreover, it became apparent that getting an overview of three-dimensional data can be enhanced by means of VR due to a more natural navigation, and better orientation and memorability capabilities.

### Acknowledgments

The work was supported and partially funded by Siemens Corporate Technology. This work has also received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 740754. We thank Philipp Meschenmoser and Tanja Allgaier for their hands-on support.

## REFERENCES


[1] N. Andrienko, G. Andrienko, and P. Gatalsky. Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages & Computing*, 14(6):503–541, 2003. doi: 10.1016/S1045-926X(03)00046-6

[2] L. Arms, D. Cook, and C. Cruz-Neira. The benefits of statistical visualization in an immersive environment. In *Virtual Reality, 1999. Proceedings IEEE*, pp. 88–95. IEEE, 1999. doi: 10.1109/VR.1999.756938

[3] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale. A Descriptive Framework for Temporal Data Visualizations Based on Generalized Space-Time Cubes. *Computer Graphics Forum*, 36(6):36–61, 2017. doi: 10.1111/cgf.12804

[4] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister. The Hologram in My Hand: How Effective is Interactive Exploration of 3D Visualizations in Immersive Tangible Augmented Reality? *IEEE Transactions on Visualization and Computer Graphics*, 24(1):457–467, 2018. doi: 10.1109/TVCG.2017.2745941

[5] R. Bidarra, D. Gambon, R. Kooij, D. Nagel, M. Schutjes, and I. Tziouvara. Gaming at the dentist's serious game design for pain and discomfort distraction. In *Games for Health*, pp. 207–215. Springer, 2013. doi: 10.1007/978-3-658-02897-8_16

[6] D. B. Carr, R. J. Littlefield, W. L. Nicholson, and J. S. Littlefield. Scatterplot matrix techniques for large N. *Journal of the American Statistical Association*, 82(398):424–436, 1987. doi: 10.1080/01621459.1987.10478445

[7] M. Cavallo and Ç. Demiralp. Clustrophile 2: Guided Visual Clustering Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):267–276, 2019. doi: 10.1109/TVCG.2018.2864477

[8] Chang-Sung Jeong and A. Pang. Reconfigurable disc trees for visualizing large hierarchical information space. In G. Wills and J. Dill, eds., *InfoVis'98: Proceedings of the IEEE Symposium on Information Visualization*, pp. 19–25,. IEEE Computer Society, 2002. doi: 10.1109/infvis.1998.729555

[9] W. S. Cleveland and R. McGill. The many faces of a scatterplot. *Journal of the American Statistical Association*, 79(388):807–822, 1984. doi: 10.1080/01621459.1984.10477098

[10] A. Cockburn and B. McKenzie. Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI '02*, p. 203, 2002. doi: 10.1145/503376.503413

[11] M. P. Czerwinski, M. van Dantzich, G. Robertson, and H. Hoffman. The Contribution of Thumbnail Image, Mouse-over Text and Spatial Location Memory to Web Page Retrieval in 3D. In *Press*, pp. 163–170, 1999. doi: 10.1.1.33.774

[12] N. Elmqvist, P. Dragicevic, and J. D. Fekete. Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1141–1148, 2008. doi: 10.1109/TVCG.2008.153

[13] X. Ester, M., Kriegel, H. P., Sander, J., & Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Kdd*, vol. 96, pp. 226–231, 1996. doi: 10.1016/B978-044452701-1.00067-3

[14] R. Etemadpour, E. Monson, and L. Linsen. The effect of stereoscopic immersive environments on projection-based multi-dimensional data visualization. In *Proceedings of the International Conference on Information Visualisation*, pp. 389–397, 2013. doi: 10.1109/IV.2013.51

[15] R. Etemadpour, B. Olk, and L. Linsen. Eye-tracking Investigation During Visual Analysis of Projected Multidimensional Data with 2D Scatterplots. In *2014 International Conference on Information Visualization Theory and Applications (IVAPP)*, pp. 233–246. IEEE, 2014. doi: 10.5220/0004675802330246

[16] J. A. W. Filho, C. Dal, S. Freitas, M. F. Rey, and C. M. D. S. Freitas. Immersive Visualization of Abstract Information: An Evaluation on Dimensionally-Reduced Data Scatterplots. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 483–490, 2018. doi: 10.1109/VR.2018.8447558

[17] P. Gatalsky, N. Andrienko, and G. Andrienko. Interactive analysis of event data using space-time cube. In *Proceedings of 8th International Conference on Information Visualisation*, pp. 145–152, 2004. doi: 10.1109/iv.2004.1320137

[18] A. Gracia, S. González, V. Robles, E. Menasalvas, and T. Von Landesberger. New insights into the suitability of the third dimension for visualizing multivariate/multidimensional data: A study based on loss of quality quantification. *Information Visualization*, 15(1):3–30, 2016. doi: 10.1177/1473871614556393

[19] K. Gruchalla. Immersive well-path editing: Investigating the added value of immersion. In *Proceedings - Virtual Reality Annual International Symposium*, pp. 157–164, 2004. doi: 10.1109/VR.2004.1310069

[20] F. Gutiérrez, J. Pierce, V. Vergara, R. Coulter, L. Saland, T. Caudell, T. Goldsmith, and D. Alverson. the Impact of the Degree of Immersion Upon Learning Performance in Virtual Reality Simulations for Medical Education. In *Journal of Investigative Medicine*, vol. 55, p. S91, 2007. doi: 10.1097/00042871-200701010-00099

[21] C. Hendrix and W. Barfield. Presence within virtual environments as a function of visual display parameters. *Presence: Teleoperators and Virtual Environments*, 5(3):274–289, 1996. doi: 10.1162/pres.1996.5.3.274

[22] D. Holten and J. J. Van Wijk. Evaluation of cluster identification performance for different PCP variants. *Computer Graphics Forum*, 29(3):793–802, 2010. doi: 10.1111/j.1467-8659.2009.01666.x

[23] C. Hurley. Clustering visualizations of multivariate data. *Journal of Computational and Graphical Statistics*, 13(4):788–806, 2004. doi: 10.1198/106186004X12425

[24] A. Inselberg and B. Dimsdale. Parallel Coordinates for Visualizing Multi-Dimensional Geometry. In *Computer Graphics 1987*, pp. 25–44. Springer, 2011. doi: 10.1007/978-4-431-68057-4_3

[25] H. Janetzko, M. C. Hao, S. Mittelstädt, U. Dayal, and D. Keim. Enhancing scatter plots using ellipsoid pixel placement and shading. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, pp. 1522–1531. IEEE, 2013. doi: 10.1109/HICSS.2013.197

[26] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton. Measuring and defining the experience of immersion in games. *International Journal of Human Computer Studies*, 66(9):641–661, 2008. doi: 10.1016/j.ijhcs.2008.04.004

[27] Y. Koren and L. Carmel. Robust linear dimensionality reduction. *IEEE Transactions on Visualization and Computer Graphics*, 10(4):459–470, 2004. doi: 10.1109/TVCG.2004.17

[28] B. Kozlíková, M. Krone, M. Falk, N. Lindow, M. Baaden, D. Baum, I. Viola, J. Parulek, and H. C. Hege. Visualization of Biomolecular Structures: State of the Art Revisited. *Computer Graphics Forum*, 36(8):178–204, 2017. doi: 10.1111/cgf.13072

[29] M.-J. Kraak. The Space-Time Cube revisited from a Geovisualization Perspective. In *Proceedings of the 21st International Cartographic Conference (ICC)*, pp. 1988–1996, 2003. doi: 10.1.1.102.9231

[30] O. H. Kwon, C. Muelder, K. Lee, and K. L. Ma. A study of layout, rendering, and interaction methods for immersive graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 22(7):1802–1815, 2016. doi: 10.1109/TVCG.2016.2520921

[31] B. Laha, K. Sensharma, J. D. Schiffbauer, and D. A. Bowman. Effects of immersion on visual analysis of volume data. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):597–606, 2012. doi: 10.1109/TVCG.2012.42

[32] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff. A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence: Teleoperators and Virtual Environments*, 10(3):282–297, 2001. doi: 10.1162/105474601300343612

[33] J. M. Loomis, J. J. Blascovich, and A. C. Beall. Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, and Computers*, 31(4):557–564, 1999. doi: 10.3758/BF03200735

[34] B. J. T. Morgan and A. P. G. Ray. Non-uniqueness and Inversions in Cluster Analysis. *Applied Statistics*, 44(1):117, 2006. doi: 10.2307/2986199

[35] T. Nakaya and K. Yano. Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. *Transactions in GIS*, 14(3):223–239, 2010. doi: 10.1111/j.1467-9671.2010.01194.x

[36] L. Nelson, D. Cook, and C. Cruz-Neira. Xgobi vs the c2: Results of an experiment comparing data visualization in a 3-d immersive virtual reality environment with a 2-d workstation display. *Computational Statistics*, 1(1):39–52, 1999. doi: 10.1007/PL00022704

[37] H. Piringer, R. Kosara, and H. Hauser. Interactive focus+context visualization with linked 2D/3D scatterplots. In *Proceedings - Second International Conference on Coordinated and Multiple Views in Exploratory Visualization, CMV 2004*, pp. 49–60, 2004. doi: 10.1109/CMV.2004.1319526

[38] Prabhat, A. Forsberg, M. Katzourin, K. Wharton, and M. Slater. A comparative study of Desktop, fishtank, and cave systems for the exploration of volume rendered confocal data sets. *IEEE Transactions on Visualization*

*and Computer Graphics*, 14(3):551–563, 2008. doi: 10.1109/TVCG.2007.70433

[39] A. Prouzeau, B. Ens, B. H. Thomas, C. Robin, and T. Dwyer. Scaptics and Highlight-Planes : Immersive Interaction Techniques for Finding Occluded Features in 3D Scatterplots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019. doi: 10.1145/3290605.3300555

[40] E. D. Ragan, A. Sowndararajan, R. Kopper, and D. Bowman. The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments. *Presence: Teleoperators and Virtual Environments*, 19(6):527–543, 2010. doi: 10.1162/pres_a_00016

[41] D. Raja, D. A. Bowman, J. F. Lucas, and C. North. Exploring the Benefits of Immersion in Abstract Information Visualization. In *Proceedings of the 8th Immersive Projection Technology Workshop*, pp. 61–69, 2004. doi: 10.1.1.205.235

[42] H. Regenbrecht and T. Schubert. Real and illusory interactions enhance presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 11(4):425–434, 2002. doi: 10.1162/105474602760204318

[43] G. Reina and T. Ertl. Volume Visualization and Visual Queries for Large High-Dimensional Datasets. In O. Deussen, C. Hansen, D. Keim, and D. Saupe, eds., *Eurographics / IEEE VGTC Symposium on Visualization*. The Eurographics Association, 2004. doi: 10.2312/VisSym/VisSym04/255-260

[44] T. M. Rhyne, A. M. MacEachren, M. Gahegan, W. Pike, I. Brewer, G. Cai, E. Lengerich, and F. Hardisty. Geovisualization for Knowledge Construction and Decision Support. *IEEE Computer Graphics and Applications*, 24(1):13–17, 2004. doi: 10.1109/MCG.2004.1255801

[45] H. H. Robert Kosara, Gerald N. Sahling. Linking Scientific and Information Visualization with Interactive 3D Scatterplots. *In Proceedings of the 12th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)*, pp. 133—-140, 2004. doi: 10.1.1.10.7799

[46] T. Schreck, J. Bernard, T. Tekušová, and J. Kohlhammer. Visual cluster analysis of trajectory data with interactive Kohonen maps. *VAST'08 - IEEE Symposium on Visual Analytics Science and Technology, Proceedings*, 8(1):3–10, 2008. doi: 10.1109/VAST.2008.4677350

[47] P. Schuchardt and D. A. Bowman. The benefits of immersion for spatial understanding of complex underground cave systems. In *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology*, p. 121. ACM, 2007. doi: 10.1145/1315184.1315205

[48] M. Sedlmair, T. Munzner, and M. Tory. Empirical guidance on scatterplot and dimension reduction technique choices. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2634–2643, 2013. doi: 10.1109/TVCG.2013.153

[49] M. Sedlmair, A. Tatu, T. Munzner, and M. Tory. A Taxonomy of Visual Cluster Separation Factors. *Computer Graphics Forum*, 31(3pt4):1335–1344, 2012. doi: 10.1111/j.1467-8659.2012.03125.x

[50] M. Slater. A note on presence terminology. *Presence Connect*, 3(January):1–5, 2003.

[51] M. Slater, B. Lotto, M. M. Arnold, and M. V. Sanchez-Vives. How we experience immersive virtual environments: The concept of presence and its measurement. *Anuario de Psicologia*, 40(2):193–210, 2009. doi: 1988-5253

[52] M. Slater and S. Wilbur. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616, 1997. doi: 10.1162/pres.1997.6.6.603

[53] Soon Tee Teoh, Kwan-Liu Ma, S. Wu, and Xiaoliang Zhao. Case study: Interactive visualization for Internet security. In *vis*, pp. 505–508, 2003. doi: 10.1109/visual.2002.1183816

[54] L. van der Maaten and G. Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008. doi: 10.1007/s10479-011-0841-3

[55] T. Van Walsum, F. H. Post, D. Silver, and F. J. Post. Feature extraction and iconic visualization. *IEEE Transactions on Visualization and Computer Graphics*, 2(2):111–119, 1996. doi: 10.1109/2945.506223

[56] J. Van Wijk and E. Van Selow. Cluster and calendar based visualization of time series data. In *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis' 99)*, pp. 4–9,. IEEE, 2003. doi: 10.1109/infvis.1999.801851

[57] J. A. Wagner Filho, C. M. Freitas, and L. Nedel. VirtualDesk: A Comfortable and Efficient Immersive Information Visualization Approach. *Computer Graphics Forum*, 37(3):415–426, 2018. doi: 10.1111/cgf.13430

[58] J. A. Wagner Filho, M. F. Rey, C. M. D. S. Freitas, and L. Nedel. Immersive Analytics of Dimensionally-Reduced Data Scatterplots. *2nd Workshop on Immersive Analytics*, 2017.

[59] D. G. Walshe, E. J. Lewis, S. I. Kim, K. O'Sullivan, and B. K. Wiederhold. Exploring the Use of Computer Games and Virtual Reality in Exposure Therapy for Fear of Driving Following a Motor Vehicle Accident. *CyberPsychology & Behavior*, 6(3):329–334, 2003. doi: 10.1089/109493103322011641

[60] E. J. Wegman. Hyperdimensional data analysis using parallel coordinates. *Journal of the American Statistical Association*, 85(411):664–675, 1990. doi: 10.1080/01621459.1990.10474926

[61] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240, 1998. doi: 10.1162/105474698565686

[62] D. Yang, E. A. Rundensteiner, and M. O. Ward. Analysis guided visual exploration of multivariate data. In *VAST IEEE Symposium on Visual Analytics Science and Technology 2007, Proceedings*, pp. 83–90. IEEE, 2007. doi: 10.1109/VAST.2007.4389000

[63] S. N. Yang, T. Schlieski, B. Selmins, S. C. Cooper, R. A. Doherty, P. J. Corriveau, and J. E. Sheedy. Stereoscopic viewing and reported perceived immersion and symptoms. *Optometry and Vision Science*, 89(7):1068–1080, 2012. doi: 10.1097/OPX.0b013e31825da430

[64] L. Yu, K. Efstathiou, P. Isenberg, and T. Isenberg. CAST: Effective and Efficient User Interaction for Context-Aware Selection in 3D Particle Clouds. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):886–895, 2016. doi: 10.1109/TVCG.2015.2467202

[65] D. Zeckzer, D. Gerighausen, and L. Müller. Analyzing Histone Modifications in iPS Cells Using Tiled Binned 3D Scatter Plots. In *2016 Big Data Visual Analytics, BDVA 2016*, pp. 1–8. IEEE, 2016. doi: 10.1109/BDVA.2016.7787042

[66] A. Zelenyuk, D. Imre, Y. Cai, K. Mueller, Y. Han, and P. Imrich. SpectraMiner, an interactive data mining and visualization software for single particle mass spectroscopy: A laboratory test case. *International Journal of Mass Spectrometry*, 258(1-3):58–73, 2006. doi: 10.1016/j.ijms.2006.06.015