Immersive Visualization of Abstract Information: An Evaluation on Dimensionally-Reduced Data Scatterplots

Jorge A. Wagner Filho* Marina F. Rey† Carla M.D.S. Freitas‡ Luciana Nedel§

Institute of Informatics
Federal University of Rio Grande do Sul

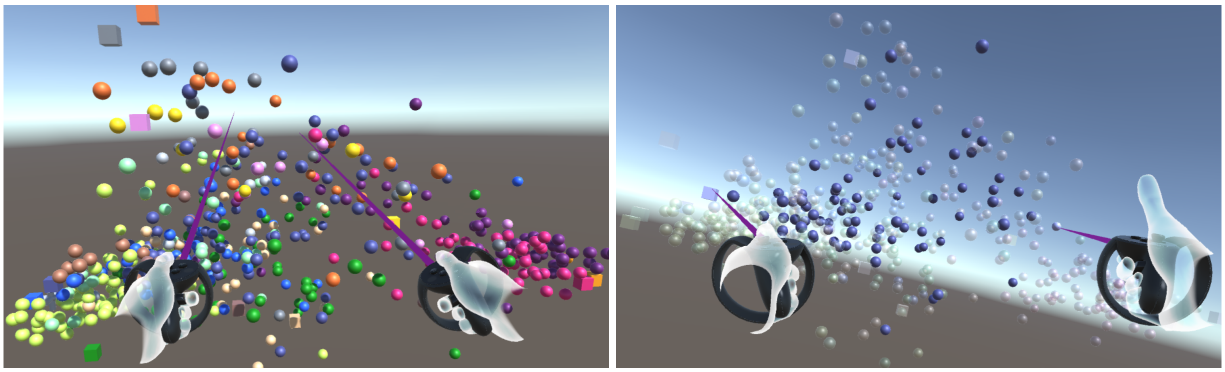


Figure 1: Proposed HMD-based immersive environment for the exploration of dimensionally-reduced data scatterplots. The user is equipped with two position-tracked hand controllers (left), being allowed to interact with the data through selection pointers (right).

Abstract

The use of novel displays and interaction resources to support immersive data visualization and improve analytical reasoning is a research trend in the information visualization community. In this work, we evaluate the use of an HMD-based environment for the exploration of multidimensional data, represented in 3D scatterplots as a result of dimensionality reduction (DR). We present a new modeling for this problem, accounting for the two factors whose interplay determine the impact on the overall task performance: the difference in errors introduced by performing dimensionality reduction to 2D or 3D, and the difference in human perception errors under different visualization conditions. This two-step framework offers a simple approach to estimate the benefits of using an immersive 3D setup for a particular dataset. Here, the DR errors for a series of roll call voting datasets when using two or three dimensions are evaluated through an empirical task-based approach. The perception error and overall task performance, on the other hand, are assessed through a comparative user study with 30 participants. Results indicated that perception errors were low and similar in all approaches, resulting in overall performance benefits in both desktop and HMD-based 3D techniques. The immersive condition, however, was found to require less effort to find information and less navigation, besides providing much larger subjective perception of accuracy and engagement.

Keywords: Immersive visualization, abstract information visualization, dimensionality reduction, 3D scatterplots.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities

*e-mail:jawfilho@inf.ufrgs.br
†e-mail:mfrey@inf.ufrgs.br
‡e-mail:carla@inf.ufrgs.br
§e-mail:nedel@inf.ufrgs.br

1 Introduction

Following consecutive breakthroughs in Virtual Reality (VR) research, the visualization community has progressively explored the use of immersive displays and new interaction devices to enhance analytical reasoning [5]. Nonetheless, even though immersive 3D visualizations present clear advantages and consolidated application for scientific spatial data [6, 11], it still remains largely unclear if and how these technologies can be properly applied to visualize abstract information [10, 13]. Some promising results have already been demonstrated, for example, in graph visualization [7, 17, 22].

In this work, we aim to expand this discussion, taking into consideration 3D scatterplots representing dimensionally-reduced data. Since this particular category of scatterplots, which is commonly applied for multidimensional data visualization, is always analyzed in terms of the distances between points, we hypothesize it could benefit from stereoscopic displays, egocentric points of view and more natural user interfaces, characteristics that are inherent to immersive analytical setups. We focus specifically on studying the application of an HMD-based environment to this problem (Fig. 1), in comparison to desktop-based alternatives, which correspond to the currently used solutions.

The use of 3D scatterplots has been controversial since long before the first uses of immersion, with related studies dating back to the 1970s [12]. In theory, a 3D representation allows clearer spatial separation, reduced overplotting and faster construction of a mental model [15]. Nevertheless, challenges such as difficulties in navigation, perspective distortion, foreshortening and occlusion have led multiple researchers to dismiss its utility. The use of 3D scatterplots for the representation of dimensionally-reduced data is often discussed. Adding an extra component could potentially reduce information loss in the process, but results from studies on quantifying visual analysis gains have been contradictory [15, 27]. Few authors, however, have investigated how immersion and stereopsis may impact on these issues. Moreover, most of them have only provided preliminary results, based on technologies which have advanced enormously over the past few years [1, 8, 26]. Therefore, we believe an updated and expanded investigation is still needed.

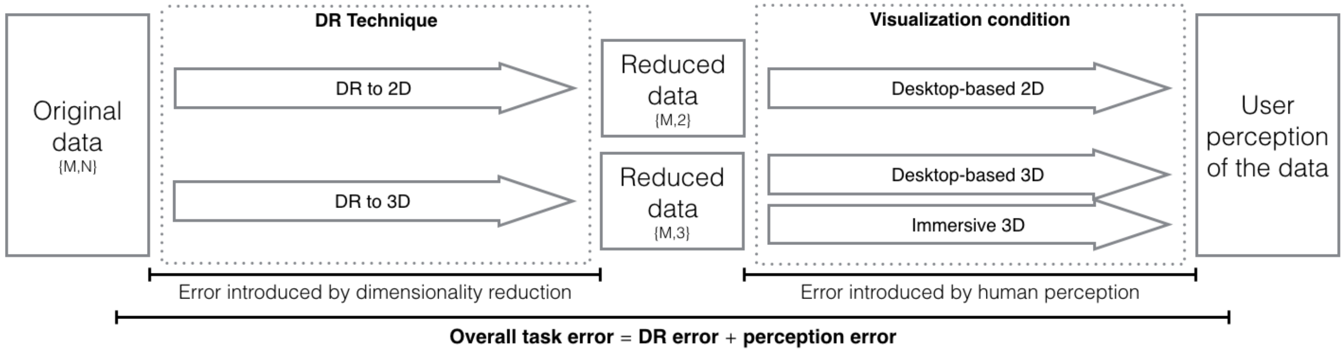


Figure 2: A visual model of the problem we target. The overall task performance for each scenario will be a result of the different errors introduced: by reducing the dataset to two or three dimensions and by using a desktop-based 2D, desktop-based 3D or immersive visualization approach.

In this paper, we build upon the results of an initial pilot study [29]. Herein, we introduce a new model of the problem in hand, accounting for the two different factors that influence in the final task performance outcome (Fig. 2). We argue that the performance gains attained in a task are not just a function of the difference in perceptual accuracy presented by users under different visualization conditions, but rather of its interplay with the difference in errors introduced by reducing the dimensions of a particular dataset to two or three dimensions. This so called DR error component is dataset-dependent, depending on the particular complexity of the data structure. This means that, for a given dataset to benefit from a three-dimensional visualization condition, its content must be indeed well mapped to 3D. Moreover, the user must be able to perceive this added information appropriately, what can be challenging given the previously discussed issues associated with three-dimensional representations.

Based on this model, we propose an evaluation framework that aims to separately assess each of these variables. The *maximum potential performance* in 2D or 3D for our datasets is estimated through a task-based empirical approach. The *perception* and *overall task errors*, on the other hand, are assessed through a user study, comparing three alternative visualization conditions: desktop-based 2D (*2D*), desktop-based 3D (*3D*) and HMD-based immersive 3D (*IM*). Participants are subjected to a set of analytical tasks for two selected datasets, one with previously detected promised improvements in 3D (*D1*), and another one that, in theory, allows for similar performance in all representations (*D2*).

Our main contributions are: (1) an improved modeling of the problem in hand, (2) a task-based evaluation framework, (3) results from a comparative user study with 30 participants, and (4) reported user behavior and feedback for the proposed solution.

The remainder of this work is structured as follows. In Section 2 we briefly review related work. We introduce our evaluation framework in Section 3, while results are presented in Section 4 and discussed in Section 5. Section 6 presents our final remarks.

2 Related Work

2.1 Immersive Analytics

Immersive Analytics [5] is an ever-growing research area in the visualization community, concerned with applying novel display and interaction resources in combination to support immersive data visualization, and to improve the performance of typical analytical tasks. Early works in this area, starting in the 90's, explored the use of small spaces surrounded by retro-projected walls, denominated CAVEs [23]. Here, however, we are concerned about HMD-based environments, considering that the current technology provides adequate immersive capabilities with much more accessible requirements, both in terms of cost and space, and its exploration in the Infovis literature is still incipient. Donalek et al. [10] presented a very interesting early work in this direction. They implemented iViz, a platform for visualization of multidimensional data using an Oculus Rift HMD and a Leap Motion sensor for interaction. In their application, up to 8 data dimensions are mapped to different attributes of points in a 3D scatterplot.

Garcia et al. [13] discuss the great success already achieved by immersive applications in the scientific visualization context [6, 11], while abstract and multidimensional data visualization stayed behind. They point out, however, several cases of successful VR application to non-spatial data, for example in genomics, as encouragement for further explorations. Another field that has presented promising results for the application of immersive approaches is graph visualization. Ware and Mitchell [31] observed an order of magnitude increase over 2D displays in a path tracing task, using high resolution displays and a mirror stereoscope. Halpin et al. [17] also obtained significant performance improvements for fine-grained questions using a CAVE-like environment. Kwon et al. [22] explored different techniques in an HMD-based environment, proposing the use of a new spheric layout that offered performance increase especially for more difficult tasks. Cordeil et al. [7] presented a comparative study between CAVE-style and HMD-based environments for collaborative analysis of graphs, and were able to obtain high accuracy scores in both. Users in the HMD condition, nonetheless, were found to be substantially faster.

Zielasko et al. [33], who also explored a use case with graph analysis, presented an interesting discussion on the challenges and opportunities of an immersive analytical scenario named *deskVR*, where the user remains seated in his office chair during the immersive exploration of data. They believe that an immersive solution must be easily integrated to the analyst's workplace and workflow in order to be really adopted, and that the transition between real and virtual worlds must be seamless, so that the analyst may combine 2D and 3D environments according to the requirements of each specific task.

2.2 Dimensionality Reduction and 3D Scatterplots

In order to visualize very high dimensional data, we explore the use of dimensionality reduction (DR). DR methods aim to generate a more compact version of the information, yet maintaining the same characteristics of the original dataset. A popular example is the Principal Component Analysis (PCA) [19], a linear method which aims to position distant points in the original dataset far apart in the lower dimensional representation. Despite presenting several important applications, such as feature selection for algorithmic input, DR techniques are predominantly used for data visualization.

The use of three-dimensional representations, such as scatterplots, has been discussed for a long time in the literature. Ware [30], in his thorough discussion on spatial representations and depth cues, argued that the only two cues likely to be useful in a 3D scatterplot are stereoscopic depth and *structure-from-motion* (motion parallax and kinetic depth effect). The first should be more helpful to differentiate depths between near points, while the latter to differentiate more distant ones. Some authors have specifically investigated the use of monoscopic 3D scatterplots for visualization of dimensionally-reduced data, but results have been mixed. Gracia et al. [15] performed a user study evaluating point classification, distance perception and outlier detection tasks, and also applied several quality loss metrics from previous literature to affirm the advantage of using a third dimension. Sedlmair et al. [27], on the other hand, performed a data study where two annotators evaluated around 800 scatterplots in relation to cluster separability, and concluded that the interactive 3D versions never outperformed the 2D scatterplots (individually or in matrices), especially considering the added interaction cost. However, one should note that both works target different analytical tasks – this being compatible with our argument for a task-based evaluation.

Concerning immersive environments, Arms et al. [1] performed a comparative evaluation of the visualization of multidimensional data projected to two and three dimensions, achieving better cluster identification results in the virtual environment and serving as inspiration for our study. However, they explored a CAVE environment, and suffered from heavy technological limitations at the time, especially regarding interaction. Raja et al. [26] also explored the application of immersive VR to 3D scatterplots in a CAVE environment, observing favorable results when including large field-of-regard, head-tracking and stereopsis. Their user study, however, was very initial, with only four subjects. A later study with 32 users was performed with similar indications, but failed to present statistical significance [25]. Babaee et al. [2] proposed a new metric to compare DR techniques in terms of structure preservation, based on a communication channel model. They visualized datasets of images reduced to three dimensions in immersive CAVE-like environments.

In a preliminary pilot study [29], we obtained results and feedback that indicated clear issues with the implementation, the protocol and the unstructured evaluation methodology, leading to our much improved problem modeling and a more controlled experiment. In that study we also used Razer Hydra controllers, which were now replaced by a more accurate alternative.

3 Evaluation Framework

In this section we describe the evaluation framework designed to assess the two errors that may affect the overall task performance.

3.1 Hypotheses

We defined five hypotheses for our evaluation purposes. As mentioned before, D1 is a dataset that presents potential information gain in 3D, and D2 one that does not.

- H1 The perception error will be smaller in *IM* than in *3D*, specially due to the stereopsis.
- H2 The overall task error in *IM* will be smaller than in *3D* or *2D* for *D1*.
- H3 The overall task error in *IM* will be at least as good as in *3D* or *2D* for *D2*.
- H4 *2D* is expected to be the quickest, given its inherent smaller cost for navigation and interaction.
- H5 The benefits provided by immersion, such as a more natural interaction and an egocentric view of the data [5], will be reflected on the subjective user evaluations.

3.2 Targeted Data

In this work, we visualize roll call voting data from the Brazilian Chamber of Deputies. We consider this domain very appropriate for our goal in this work due to the very high dimensionality of its datasets (each roll call is a dimension), its consistent application of DR techniques in the literature (a survey was published by Spirling and McLean [28]) , and the easy definition of semantically meaningful analytical tasks.

We extracted information about the votes of each deputy and the official vote instruction given by each party represented in the Chamber for every roll call in the last four four-year legislatures from the Brazilian Congress: 52nd (451 roll calls), 53rd (619 roll calls), 54th (428 roll calls) and 55th (493 roll calls). For each legislature, we constructed a voting matrix where all deputies and parties are represented by M lines, and roll calls are represented by N columns. Each (i, j) cell is then attributed a value depending on the ith deputy or party vote on the jth roll call: -1 for "no", 1 for "yes" or 0 for abstention or absence. Following previous works [9, 20], Principal Component Analysis (PCA) by Singular Value Decomposition [14] is then applied to this matrix, resulting in $min(N, M)$ principal components. For visualization purposes, only the first two or three are considered, and seen as a political spectrum. Euclidean distances in these representations indicate how similarly or differently deputies have voted in the given period.

3.3 Analytical Tasks

The axes in a scatterplot obtained from a DR method correspond to artificial, uncorrelated dimensions synthesized by an algorithm and, in general, have no semantic meaning. Instead, all information presented is encoded through the distance between points in the scatterplot. Our set of defined analytical tasks is thus composed of tasks relative to different competencies in distance judgments: perception of near, medium and far distances. All tasks were designed to be simple and atomic (i.e., combinable for more complex analyses), but we believe they constitute a representative subset of the typical tasks of a data analysis in this specific domain.

- T1 *Selection of a deputy's closest deputy*. In this near-distance perception task, the user is requested to select the closest deputy to a given one.
- T2 *Selection of a deputy's closest party*. A more difficult variation of the previous task (since deputies are usually positioned between multiple parties), where the user is requested to select the closest party to a given deputy. It can also be seen as a point classification task, where the user is reclassifying deputies in parties according to vote coherence.
- T3 *Identification of party outlier*. In this task, the user must identify the member of a specific party which is furthest located from the official party position.
- T4 *Selection of a party's closest party*. Also a variation of T1, but exploring different competencies since parties are more distributed on the spectrum.

3.4 Task-Based DR Error Assessment

The efficiency of a DR method when representing a dataset in a lower dimension is highly dependent on the data geometry. This implies that, while some datasets will benefit from an extra dimension, others will already be well represented in 2D. In fact, several metrics try to quantify the information gain of adding a third dimension to a DR data scatterplot. Gracia et al. applied 11 metrics in their study with 12 DR algorithms and 12 real-world datasets to affirm that the loss of quality when reducing from 3 to 2 dimensions accounts, in average, for 30.4% of the total DR loss [15]. A simple and commonly used metric in the case of PCA is the proportion of variance contributed by each principal component, given by its eigenvalue. This information is usually plotted in a *scree plot* (see Fig. 3) and used to estimate the dataset *intrinsic dimensionality*.

However, from a practical point of view, it is generally difficult to estimate how the information loss from 3D to 2D, even if it exists,

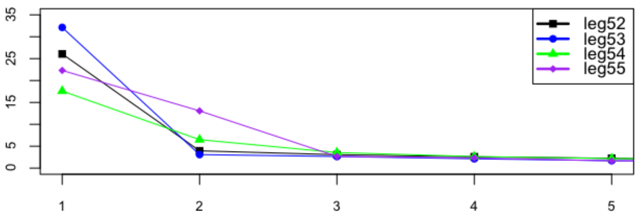


Figure 3: Scree plot of the proportion of variance contributed by each of the 5 first principal components in our 4 datasets.

will impact on the user's analytical performance. Moreover, it is hard to conjecture whether the trade-off between information loss and the clearer and simpler visualization provided by 2D is worth it.

We approach this issue in an empirical, task-based way, by computing a user's *maximum potential performance* in 2D and 3D. This is done by simulating the minimum average error a user would achieve in each scenario if he/she were able to perceive the presented information with absolute accuracy, for all possible instances of a task. For example, if the task is selecting the closest deputy to a given one (T1), the correct answers to all 513 deputies according to the information presented in 2, 3 or all dimensions are calculated and compared. Euclidean distances between points in the corresponding set of dimensions are used, and average errors are always calculated in the original vote matrix.

Fig. 6 presents results for our four datasets over the different tasks introduced in Sect. 3.3. As expected, different legislatures result in different potential contributions for the third dimension. We identify two particular scenarios: for the 54th legislature, all tasks appear to benefit from its inclusion – for T1, T2 and T3, it is the dataset with the largest performance improvement (observing Fig. 3, this was indeed the dataset with the smallest variance explained by the two first components combined). For the 52nd legislature, on the other hand, all tasks appear to be equally well performed in both scenarios – for T1, T2 and T4, it is the dataset with the least gain. From now on, we will refer to these datasets as D1 (54th legislature) and D2 (52nd legislature). In the remaining of this paper, we will assess how the task performance is affected by the user perception of the third dimension (under desktop and HMD-based conditions) in both of these cases.

3.5 User Study: Perception & Overall Error Assessment

3.5.1 Visualization Conditions

The implementations for our three studied visualization conditions are based on those used in our pilot study, and were updated to include feedback provided by the participants. Both two and three dimensional virtual environments (VEs) were implemented using the Unity game engine. The 3D version can be explored either through desktop-based (monoscopic, non head-tracked) or HMD-based (stereoscopic, head-tracked) setups.

In both desktop-based VEs, explored through a 22” 1080p display, controls were implemented using only mouse and keyboard, as in a traditional data analysis setup. In *IM*, our implementation choice, looking for providing a more natural and immersive interaction, was to use two selection rays, which are controlled by position-tracked Oculus Touch hand controllers (see Fig. 1). Accurate virtual representations of the users' hands and of the controllers are also shown, increasing the feeling of embodiment and serving as anchors to the real world [33]. This environment is explored through an Oculus Rift CV1 HMD (formed by two 1080×1200 displays), with the user seated in a swivel chair. Several guidelines were employed to minimize possible discomfort: the speed of movement is slow and constant; user control of the camera is maximized; no near ground was included to avoid uncomfortable rapid ground plane

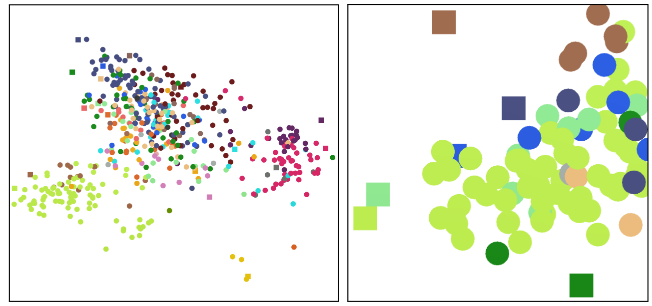


Figure 4: In the *2D* condition, data points are distributed along screen space (left), and the user is allowed to zoom and pan (right).

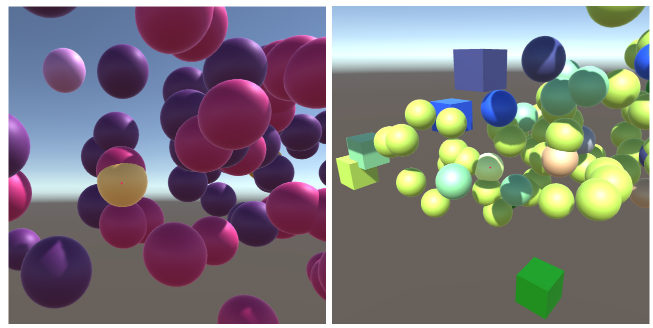


Figure 5: In the *3D* conditions, the user is allowed to freely navigate through the data, which is distributed along a 3D virtual environment.

changes; and adequate hardware was employed to minimize latency and lag. In the event of teleportation to a new position, such as in the beginning of a task, a camera fade is also applied [32].

All VEs explore the same visual encodings: colors for political parties and shape for different categories of points – circles or spheres for deputies and squares or cubes for official parties positions. They also all offer the same set of possible interactions: a user may click on a point to show/hide its name (using double click or a specific button in the controller) and may highlight the whole party of any point to inspect its relative position (using right click or the inner trigger in the controller, to emulate a grabbing action). Labels are shown upon selection during the familiarization phase, to aid in the comprehension of the representation semantics. During the tasks, they remain hidden to avoid potential use of previous knowledge.

The setups differ, however, in the forms of navigation and interaction. In *2D*, the user can zoom in/out and pan the screen (see Fig. 4). In both 3D versions (see Fig. 1 and Fig. 5), the user can navigate freely in all directions, through gaze-directed flying [24]. He/she is allowed to move forwards, backwards, vertically or laterally, using keyboard keys or the left controller joystick, and also rotate the camera, moving the mouse or using the right controller joystick. This metaphor is meant to be simple to learn [3] and enable an egocentric view, placing the user inside the data representation.

Moreover, while, in *2D*, selection is done by the mouse cursor and, in *IM*, by the pointer rays, in *3D*, it is also gaze-directed, implemented by a reticle cursor in the center of the screen, so that the mouse movement can be used to rotate the camera. The 3D environments also include a ground and sky background and illumination from above for orientation purposes [16].

3.5.2 Experiment Design

Our user study was implemented through a within-subjects protocol, combining *3 visualization conditions x 4 tasks x 2 datasets*. The target population, recruited on campus, was composed of 30 subjects

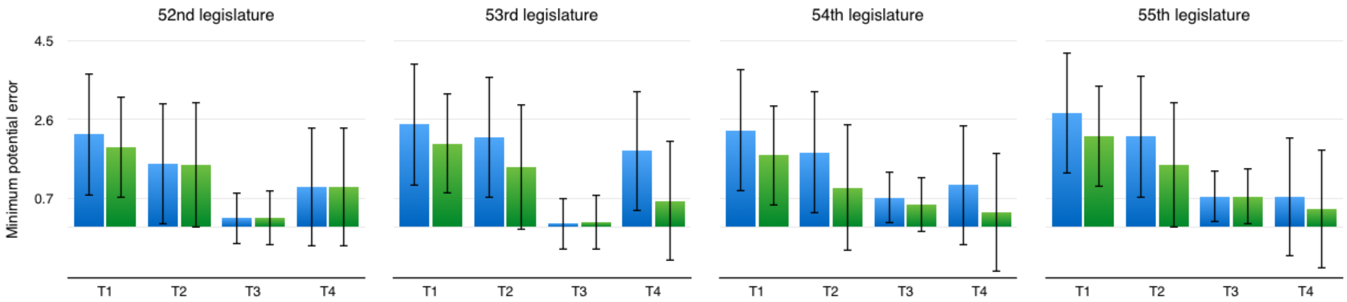


Figure 6: Results of our task-based analysis of the minimum potential average error a user could achieve both in 2D (blue) and 3D (green), were he/she always capable of perceiving accurately the distances represented. Error bars present the corresponding standard deviations.

(20 male/10 female; average age of 25.2, ranging from 17 to 50), who had not taken part in the pilot study. Regarding previous contact with involved technologies, 76% reported at least average familiarity with first person games and gamepads, and 60% with motion controllers. However, 60% had low or no familiarity at all with HMDs.

Each participant experienced all conditions in alternating order, to minimize learning biases. The subject was always initially allowed to get familiar with the corresponding controls while exploring the 55th legislature dataset. Then, he/she was asked to perform, as accurately as possible and without specific training, each of the tasks described in Sect. 3.3 six times in a row, being three in dataset D1 and three in D2. The order of presentation of the datasets in each task is alternated between users, but task order is preserved. Between different conditions, the scatterplots are mirrored with relation to the vertical and/or horizontal axes, so as to minimize the possibility of using previously viewed information. The specific task questions presented were selected as follows: for each task and dataset, 10 different sets of 9 points were randomly selected (3 for each condition). Each of these sets was used by three users, alternating the conditions, so that, in the end, every point selected once in one condition was also selected once in the others. The purpose of selecting multiple sets of random points instead of just one is to maximize the representation of different possible situations in the data, and to cross validate the results [15]. Also to maximize representation, in tasks involving deputies, repetition was not allowed even between sets (this way, these tasks explore 90 out of the 513 possible deputy points) – for party tasks, this is not possible due to their smaller number, and so repetition is not allowed just inside the sets.

In all tasks, one point is shown blinking, and the user must point and click to choose the corresponding answer. Following previous experiences from our pilot study, we opted to block semantically impossible answers (e.g. a party outlier that is not from the given party), so as to reduce noise resulting from accidental clicks or misunderstandings. When this is the case, the user hears a negative audio feedback. Otherwise, a positive sound is played, and the camera is teleported back to the initial overview position.

After each technique, subjective opinion questionnaires were applied, including usability-related (SUS) questions [4]. SSQ [21] was applied pre and post VR exposure to evaluate well-being effects. In the end, users were also allowed to compare all the techniques according to different criteria. The complete experiment took about 45 minutes.

4 Results

4.1 Quantitative Results

4.1.1 Perception Error

Perception errors were calculated as the differences in Euclidean distances, in two or three dimensions, between the one from the given point to the user's answer and the one to the correct answer in the representation. They refer, therefore, to the errors with relation to the information shown, and not to the original data. The better the user was able to perceive the distances in the representation, the closer to zero this error will be. Fig. 8 presents results for all tasks (in this analysis, we do not differentiate between datasets). Since we were not able to verify normality under Shapiro-Wilk tests, non-parametric Friedman tests were executed. Post-hoc tests are implemented using the Wilcoxon-Nemenyi-McDonald-Thompson test [18]. Surprisingly, no significant differences were observed in any task (p-values .8, .5, .8 and .4, respectively), neither between *3D* and *IM* nor between both and *2D*. H1 was, therefore, not confirmed.

4.1.2 Overall Task Error

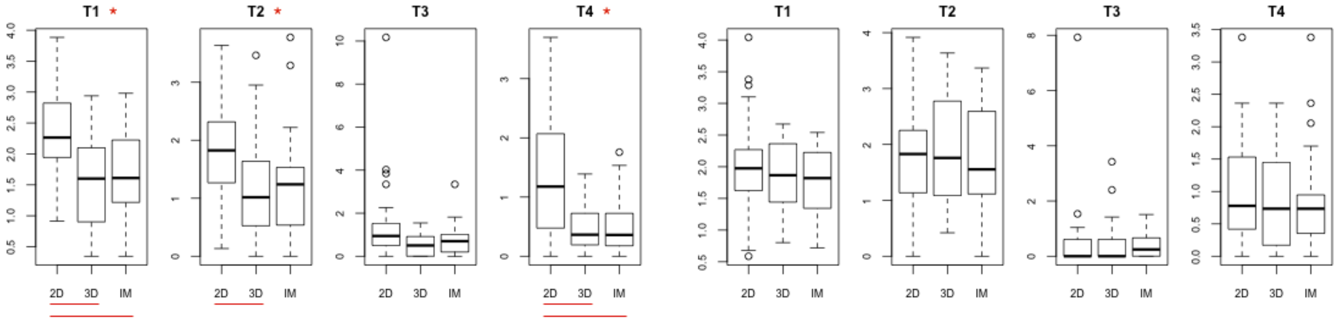
Overall task errors were calculated as the differences in Euclidean distances, in the original vote matrix, between the one from the given point to the user's answer and the one to the correct answer in the real multidimensional data set. They are expected, therefore, to be the combination between the expected DR errors seen in Fig. 6 and the perception errors seen in Fig. 8. The results for all tasks in D1 and D2 are shown in Fig. 7. Friedman and the Wilcoxon-Nemenyi-McDonald-Thompson post-hoc tests were again used, and the significant pairwise differences, when found, are indicated with red lines.

For D2, no task presented significant differences between conditions (p-values .6, .6, .93 and .85, respectively). This confirms our hypothesis H3, i.e., *IM* is at least as good as *2D* for the dataset that has the least expected information gain with the use of the third dimension. D1, on the other hand, presented significant differences for all tasks except T3, which can be considered almost significant (p-values were .002, .007, .06 and .01). All indicated pairwise differences presented $p < .01$. In T2, *2D* and *IM* presented a trend of significance with $p = .08$. Notably, however, H2 could not be confirmed, since *3D* and *IM* were never significantly different.

4.1.3 Task Completion Time

As expected in H4, *2D* was significantly faster in time than the two other conditions in all tasks (always with $p < .001$). *3D* and *IM* did not present significant differences with each other in any case.

Distance perception tasks T1 and T4 were the quickest to be solved, with average times of 8.4s, 23.1s and 26.2s for *2D*, *3D* and *IM* in the former, and 8s, 16.3s and 15.8s for the latter. The outlier identification and classification tasks took longer, especially in the 3D conditions. This was already expected due to their higher difficulty, since frequently there are multiple possible answers (observe Fig. 4, Fig. 5 and Fig. 1 (right)). Average times were 10.5s, 29.8s and 33.1s for T2, and 10.5s, 23.6s and 30.2s for T3.



(a) D1

(b) D2

Figure 7: Overall task errors (w.r.t original data) observed for each dataset. Asterisks and red lines indicate occurrence of statistical significance.

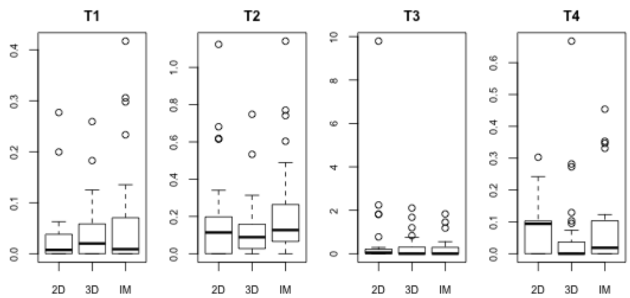


Figure 8: Results for perception errors under the different conditions and tasks. They are given by the average differences between the Euclidean distances from the task point to the user answer and to the correct one, in two or three dimensions.

4.2 User Behavior

4.2.1 Navigation Patterns

The monitoring of user navigation patterns in both three dimensional conditions showed that navigated distances were consistently longer in *3D* in comparison to *IM*. More specifically, they were 18% longer in T1 ($p = .01$, under a paired Wilcoxon signed-rank test), 20% longer in T2 and T3 ($p = .03$ and $p = .07$, respectively), and 30% longer in T4 ($p = .004$). This was not reflected in faster completion times, as seen previously, probably due to the slower navigation speed adopted in the immersive scenario, or because we did not ask users to care about the time. Many users complained about the slow speed and not being able to increase it, but we believe this contributed to minimize the occurrence of simulator sickness.

Similar behaviors were also observed in terms of accumulated camera rotation, which was 48% larger for *3D* in T1 ($p = .0004$), 23% larger in T2 ($p = .1$) and 38% larger in T4 ($p = .003$). The only exception was T3 (rotation 8% smaller, $p = .62$), what is explained by the different nature of this task (perception of long distances).

Considering that our protocol ensures that a task performed in one condition will always be performed by another user in the other conditions, these differences are not related to task difficulty, but to the interaction and visualization techniques themselves. The enabled navigation forms were also similar in both conditions. The rotation difference may be partly due to the different fields-of-view (FOV) in both scenarios (60 degrees in *3D* and 96 in *IM*). Another plausible explanation for navigation and rotation variations may reside, however, in the different depth cues provided. As discussed by Ware [30], a very important cue for the inspection of clouds of points, besides stereopsis, is structure-from-motion.

Finally, navigated distance was also found to be, as expected, consistently inversely correlated with perception error, particularly for *3D*. Pearson correlations between the two metrics were -0.66 and -0.49 for *3D* and *IM*, respectively. While at least three users were observed to develop the strategy of assuming points positions to obtain egocentric perceptions of distance, most adopted allocentric points of view.

4.2.2 Hand Usage

Considering specifically the immersive condition, we were particularly curious about how users would adapt to the two-handed embodied interaction metaphor. All hand movements and interactions (point selections and party highlights) were thus recorded. Observed right hand use was much more pronounced, as was already expected given that only one participant had reported being exclusively left-handed. Nonetheless, an interesting result was that hand usage varied according to the task requirements. Average numbers of interactions with the left and right hands were, respectively, 1.1 and 7.7 for T1, 1.0 and 8.7 for T2, and 1.0 and 6.3 for T4 (ratios of 6.6, 8.4 and 6.1). For T3, where a common approach was to highlight the party with one hand and select the party outlier with the other (see Fig. 1 (right)), this changed to 4.2 and 13.8 (a ratio of 3.2, less than half of the other tasks). Moreover, while in T1, T2 and T4, less than 30% used both hands to interact, in T3 this was done by 63% of the users.

Another interesting observation was that the differences in average hand movement were much smaller than the ratios in effective interactions, suggesting the users consistently moved both hands together despite using one of them much more frequently. Average hand translation per task was about 1.0 meter for the left hand, and 1.2 meters for the right one.

4.3 User Feedback

4.3.1 Preferences and Usability

All visualization conditions were well rated with relation to usability, without significant differences ($p = .11$). SUS questionnaire scores were 81.5 for *2D*, 77 for *3D*, and 76.6 for *IM* (standard deviations 12.6, 15.3 and 20.2). We believe this successfully reflected our efforts to optimize our implementations (in the pilot study, the ratings for the previous versions had been scored 83.1, 61.3 and 68.3, respectively).

In post-technique interviews for all conditions, at least 75% of the participants also agreed it was easy to navigate and interact, achieving 90% in some cases (*2D* navigation, *3D* and *IM* interactions) (see Fig. 9). Users appeared to be able to complete the tasks with less effort in *IM*, with 24 agreeing that it was easy to find information in this representation, compared to 21 in *3D*, and 16, in *2D*. However,

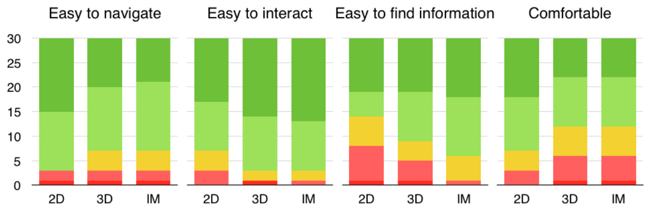


Figure 9: User Likert-scale agreements to the different assertions, ranging from completely disagree (dark red) to completely agree (dark green). All techniques were well rated in terms of ease of navigation and interaction. Noticeably, however, *IM* was better rated in terms of ease to find information, and performed as well as *3D* for comfort.

no significant differences were found in the Likert-scale questions – Friedman tests indicated p = .14 for navigation, .26 for interaction, .15 for information finding and .14 for comfort.

Users were also asked to rank the different experienced conditions according to different criteria. Interestingly, despite the similar quantitative results achieved for both *3D* and *IM*, 18 users perceived *IM* as the most accurate of all, against only 3 for *3D*. 19 users indicated *2D* as the least accurate condition, probably because, with one less dimension, points were clearly less well distributed in space. A Friedman test on the mean rankings for accuracy (p = .005) indicated significant differences between *IM* and *2D* (p = .005) and near significance between *IM* and *3D* (p = .052), but no significance between *2D* and *3D* (p = .71). *2D* and *IM* tied in the dispute for the title of most intuitive (p = .16), with 13 votes each, what is rather surprising considering the ubiquitousness of 2D interfaces (actually, *2D* was also voted 10 times the least intuitive, versus 7 of *IM*).

In terms of time, subjective perceptions confirmed the quantitative observations, with *2D* being placed behind both *3D* ($p < .001$) and *IM* ($p < .001$), and no significant differences between *3D* and *IM* (p = .55). Finally, 25 participants classified *IM* as the most engaging condition, compared to *2D* ($p < .001$) and *3D* (p = .002). This is probably related in large part to the novelty of the display and interaction technologies being used, but may also refer in part to its sense of immersion and egocentric point of view. Differences between *2D* and *3D* were not significant (p = .07).

4.3.2 Simulator Sickness

Simulator sickness is still a major issue in immersive environments, especially when non-physical navigation is employed. Despite following multiple guidelines (Sect. 3.5.1), we still observed significant well being effects on part of the subjects. Fig. 10 displays, ordered from least to most severe, the observed VR exposure impacts on the SSQ scores [21] for all participants.

Noticeably, while around 60% reported only minor symptoms (to the left of the red line), the others presented quite significant discomfort levels. Many users reported that this was minimized (though not avoided) when employing physical movements, for example, to rotate the camera, instead of using the alternative joystick control. User results did not appear, however, to be impacted by the occurrence of discomfort, with a Pearson correlation of only -0.1 between SSQ scores and average perception errors in *IM*.

5 Discussion

The results obtained from the user study offered many insights. The most surprising was certainly the absence of significant distance perception differences between *2D*, *3D* and *IM*, contradicting previous beliefs about the suitability of monoscopic 3D scatterplots and also our hypothesis H1. We believe that this is related to the fact that our desktop-based 3D environment, implemented in a powerful game engine, does not resemble typical 3D scatterplots. Designed in an effort to enable a fair comparison with its HMD-based counterpart, it provided game-like first person navigation and a multitude of depth cues (including perspective, occlusion, shading and structure-from-motion). As a consequence, both *3D* and *IM* were able to present the promised information gain for dataset D1, with significant or almost significant differences to *2D* with relation to the original voting data in all tasks. Both techniques were also able to present similar performance to *2D* in dataset D2. These facts confirmed part of hypotheses H2 and H3.

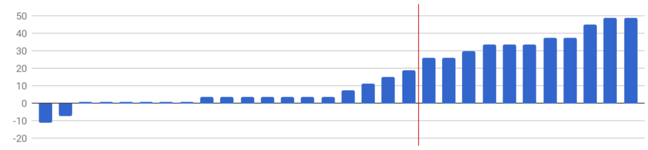


Figure 10: SSQ score impacts post-VR exposure for all 30 participants, ordered from least to most severe. Around 60% of them presented only minor symptoms (to the left of the red line), but others presented quite significant discomfort levels.

Analyzing behavioral and subjective results, however, a series of differences between *3D* and *IM* appears. An equivalent performance appears to have taken considerable less effort in the immersive scenario, given that, under this condition, users were required to navigate up to 24% less, and agreed more often that information was easy to find. This could benefit higher-levels tasks, such as cluster detection, which requires estimating multiple pairwise distances at the same time – nonetheless, this should be verified by future studies. Subjectively perceived accuracy was also much larger for *IM* than for *3D*, despite their similar results. This was also observed during our post-test interviews, when many participants described being convinced of a better performance within the immersive scenario. However, it could be argued that this might potentially generate over-confidence in incorrect observations. *IM* was also labeled the most engaging, what we believe may be, at least in part, linked to its natural interaction and egocentric point of view, as stated in our hypothesis H5. Despite around 40% of the users presenting significant levels of discomfort due to simulator sickness, *IM* was also well rated in terms of usability through the SUS questionnaire, with a similar score to *3D*. Task completion times were, as expected (H4), around 3 times slower in *IM* than in *2D*, due to the navigation and interaction costs incurred by the third dimension. However, no significant differences were observed between *IM* and *3D*, despite the slower navigation provided. It is important to mention that the time results refer to the particular 3D navigation metaphor employed (free-flying), and could vary under alternative implementations. We believe approaches that allow embodied manipulation of the dataset from a fixed position would sacrifice the egocentric view but minimize time requirements and possibly also the simulator sickness.

We acknowledge that these results are still narrow in the aspect that we only explored one data domain (roll-call data analysis) and one DR technique (PCA). Our objective here was not, however, to assert the universal applicability of immersive scatterplots for the exploration of multidimensional data. Our point is, instead, to demonstrate that current off-the-shelf VR technologies may effectively aid in analytical tasks for abstract information visualization in some cases, even challenging previous beliefs about three-dimensional representations. This area is still incipient, and much more work is still needed to construct more specific guidelines about when *IM* is or is not a recommended approach. Here, we proposed a task-based analysis approach to this end, and comparative user studies for the assessment of perception errors.

6 Conclusion and Future Work

In an effort to extend discussions about Immersive Analytics to new contexts, we presented an evaluation on a particular representation for multidimensional data: 3D scatterplots obtained using PCA. We modeled the overall visualization task error as the result of the combination between the error introduced by dimensionality reduction and the one introduced by human perception. Through a task-based empirical approach, we selected two different datasets in the domain of roll call analysis: one with promising information gain in 3D and one without such gain. In a user study, we observed that perception errors were similarly low both in desktop-based and HMD-based conditions. Task performance was therefore improved with the addition of the third dimension regardless of immersion, when the data enabled so. Nonetheless, the HMD-based condition required smaller effort to find information and less navigation, besides offering a much larger subjective perception of accuracy and engagement.

As future work on this topic, we plan to implement new comparative evaluations, focusing on different tasks or data domains, and to improve our immersive environment to mitigate user discomfort and minimize completion times (for example, exploring different navigation metaphors). Moreover, we intend to expand our study into other forms of immersive information visualization, aiming to broaden the discussion about this important application field.

Acknowledgments

We acknowledge the financial support from CNPq and CAPES, Brazilian research and scholarships funding agencies. We also thank the participants in the user study for their availability and interesting comments, and the reviewers for their contributions.

References

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

[2] M. Babaee, M. Datcu, and G. Rigoll. Assessment of dimensionality reduction based on communication channel model; application to immersive information visualization. In *2013 IEEE International Conference on Big Data*, pp. 1–6. IEEE, Oct 2013. doi: 10.1109/BigData.2013.6691726

[3] D. Bowman, E. Kruijff, J. J. LaViola Jr, and I. P. Poupyrev. *3D User Interfaces: Theory and Practice, CourseSmart eTextbook*. Addison-Wesley, 2004.

[4] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194):4–7, nov 1996.

[5] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, et al. Immersive analytics. In *2015 Big Data Visual Analytics (BDVA)*, pp. 1–8. IEEE, Sept 2015. doi: 10.1109/BDVA.2015.7314296

[6] D. Coffey, N. Malbraaten, T. Le, I. Borazjani, F. Sotiropoulos, and D. F. Keefe. Slice wim: A multi-surface, multi-touch interface for overview+detail exploration of volume datasets in virtual reality. In *Symposium on Interactive 3D Graphics and Games*, pp. 191–198. ACM, New York, NY, USA, 2011.

[7] M. Cordeil, T. Dwyer, K. Klein, B. Laha, K. Marriott, and B. H. Thomas. Immersive collaborative analysis of network connectivity: Cave-style or head-mounted display? *IEEE transactions on visualization and computer graphics*, 23(1):441–450, Jan 2017. doi: 10.1109/TVCG.2016.2599107

[8] A. V. Datey. *Experiments in the use of immersion for information visualization*. PhD thesis, Virginia Tech, 2002.

[9] F. G. de Borja and C. M. Freitas. Civisanalysis: Interactive visualization for exploring roll call data and representatives' voting behaviour. In *Proc. SIBGRAPI*, pp. 257–264. IEEE, aug 2015. doi: 10.1109/SIBGRAPI.2015.34

[10] C. Donalek, S. G. Djorgovski, A. Cioc, A. Wang, J. Zhang, E. Lawler, S. Yeh, A. Mahabal, M. Graham, A. Drake, et al. Immersive and collaborative data visualization using virtual reality platforms. In *2014 IEEE International Conference on Big Data (Big Data)*, pp. 609–614. IEEE, Oct 2014. doi: 10.1109/BigData.2014.7004282

[11] M. Drouhard, C. A. Steed, S. Hahn, T. Proffen, J. Daniel, and M. Matheson. Immersive visualization for materials science data analysis using the oculus rift. In *2015 IEEE International Conference on Big Data (Big Data)*, pp. 2453–2461. IEEE, Oct 2015. doi: 10.1109/BigData.2015.7364040

[12] M. A. Fisherkeller, J. H. Friedman, and J. W. Tukey. An interactive multidimensional data display and analysis system. 1974.

[13] R. J. García-Hernández, C. Anthes, M. Wiedemann, and D. Kranzlmüller. Perspectives for using virtual reality to extend visual data mining in information visualization. In *2016 IEEE Aerospace Conference*, pp. 1–11. IEEE, March 2016. doi: 10.1109/AERO.2016.7500608

[14] G. H. Golub and C. Reinsch. Singular value decomposition and least squares solutions. *Numerische Mathematik*, 14(5):403–420, apr 1970. doi: 10.1007/BF02163027

[15] 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

[16] G. Gray. *Navigating 3d Scatter Plots in Immersive Virtual Reality*. PhD thesis, University of Washington, 2016.

[17] H. Halpin, D. J. Zielinski, R. Brady, and G. Kelly. Exploring semantic social networks using virtual reality. In *International Semantic Web Conference*, pp. 599–614. Springer, Berlin, Heidelberg, 2008. doi: 10.1007/978-3-540-88564-1_38

[18] M. Hollander, D. A. Wolfe, and E. Chicken. *Nonparametric statistical methods*. John Wiley & Sons, 2013.

[19] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

[20] A. Jakulin and W. Buntine. Analyzing the us senate in 2003: Similarities, networks, clusters and blocs. *Preprint. Available at http://kt. ijs. si/aleks/Politics/us_senate. pdf*, dec 2004.

[21] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, jul 1993.

[22] 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, Jul 2016. doi: 10.1109/TVCG.2016.2520921

[23] S. Manjrekar, S. Sandilya, D. Bhosale, S. Kanchi, A. Pitkar, and M. Gondhalekar. Cave: An emerging immersive technology–a review. In *Computer Modelling and Simulation (UKSim), 2014 UKSim-AMSS 16th International Conference on*, pp. 131–136. IEEE, 2014.

[24] M. R. Mine. Virtual environment interaction techniques. *UNC Chapel Hill CS Dept*, 1995.

[25] D. Raja. *The Effects of Immersion on 3D Information Visualization*. PhD thesis, Virginia Polytechnic Institute and State University, 2006.

[26] D. Raja, D. Bowman, J. Lucas, and C. North. Exploring the benefits of immersion in abstract information visualization. In *Proc. Immersive Projection Technology Workshop*, pp. 61–69, 2004.

[27] 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, Dec 2013. doi: 10.1109/TVCG.2013.153

[28] A. Spirling and I. McLean. The rights and wrongs of roll calls. *Government and Opposition*, 41(4):581–588, 2006. doi: 10.1111/j.1477-7053.2006.00204.x

[29] J. A. Wagner Filho, M. F. Rey, C. M. Freitas, and L. Nedel. Immersive analytics of dimensionally-reduced data scatterplots. In *2nd Workshop on Immersive Analytics*. IEEE, 2017.

[30] C. Ware. *Information Visualization: Perception for Design*. Elsevier, 2012.

[31] C. Ware and P. Mitchell. Visualizing graphs in three dimensions. *ACM Transactions on Applied Perception (TAP)*, 5(1):2, 2008.

[32] R. Yao, T. Heath, A. Davies, T. Forsyth, N. Mitchell, and P. Hoberman. Oculus vr best practices guide. *Oculus VR*, 2014.

[33] D. Zielasko, B. Weyers, M. Bellgardt, S. Pick, A. Meibner, T. Vierjahn, and T. W. Kuhlen. Remain seated: towards fully-immersive desktop vr. In *2017 IEEE 3rd Workshop on Everyday Virtual Reality (WEVR)*, pp. 1–6, Mar 2017. doi: 10.1109/WEVR.2017.7957707