Assessing 2D and 3D Heatmaps for Comparative Analysis: An Empirical Study

Matthias Kraus
University of Konstanz
matthias.kraus@uni.kn

Katrin Angerbauer
University of Stuttgart
katrin.angerbauer@visus.uni-stuttgart.de

Juri Buchmüller
University of Konstanz
juri.buchmueller@uni.kn

Daniel Schweitzer
University of Konstanz
daniel.schweitzer@uni.kn

Daniel A. Keim
University of Konstanz
daniel.keim@uni.kn

Michael Sedlmair
University of Stuttgart
michael.sedlmair@visus.uni-stuttgart.de

Johannes Fuchs
University of Konstanz
johannes.fuchs@uni.kn

ABSTRACT

Heatmaps are a popular visualization technique that encode 2D density distributions using color or brightness. Experimental studies have shown though that both of these visual variables are inaccurate when reading and comparing numeric data values. A potential remedy might be to use 3D heatmaps by introducing height as a third dimension to encode the data. Encoding abstract data in 3D, however, poses many problems, too. To better understand this tradeoff, we conducted an empirical study ($N = 48$) to evaluate the user performance of 2D and 3D heatmaps for comparative analysis tasks. We test our conditions on a conventional 2D screen, but also in a virtual reality environment to allow for real stereoscopic vision. Our main results show that 3D heatmaps are superior in terms of error rate when reading and comparing single data items. However, for overview tasks, the well-established 2D heatmap performs better.

Author Keywords

virtual reality; visual analytics; heatmaps

CCS Concepts

•Human-centered computing → Heat maps; Virtual reality;

CHI'20, April 25–30, 2020, Honolulu, HI, USA

ISBN 978-1-4503-6708-0/20/04. . . $15.00

DOI: https://doi.org/10.1145/3313831.3376675

INTRODUCTION

Heatmaps are omnipresent in information visualization. They are frequently used not only as a basic module for novel applications and visualization designs [52, 51, 60], but also as a tool for presenting research results [42, 12, 59]. Heatmaps allow the analyst to quickly grasp a 2D distribution because of their capability to facilitate intuitive encoding of values by color in a 2D grid. The technique is also frequently deployed for comparison tasks, for example, to convey a temporal progression of 2D distributions. When heatmaps are used in a visual analytics pipeline to display intermediate results, visualizations have to be compared with visualizations from previous analysis steps to evaluate the improvement caused by parameter changes. Side-by-side comparisons of 2D distributions are also frequent tasks. For instance, Schreck et al. [52] deployed heatmap visualizations to compare distributions of different properties in Kohonen maps. In their application, the analyst can quickly identify attributes for which their neural network is optimized for. Besides, there are many examples in literature in which pre-post comparisons are presented. For instance, when two heatmaps are compared, one of which depicts values in their original form, and a second depicts the distribution after applying a change [18, 61, 43].

However, when comparing multiple heatmaps with each other, several issues emerge. In juxtapositioned comparisons, the analyst must locate a specific location in multiple heatmaps in order to compare individual values. This leads to a high cognitive load and a high potential error rate. Supportive methods like linking and brushing can alleviate this problem but might lead to additional issues, such as a perceptual distortion of values close to the highlight. Alternatively, multiple distributions could be joined into a single aggregation visualization, such as a difference map. Aggregations like this facilitate various tasks, such as offset extraction, but make others impossible (e.g., exact value extraction).

To overcome problems of juxtapositioned comparisons, one could extend the 2D heatmap by a third dimension, double encoding value by color and height. The resulting 3D heatmaps could then be superpositioned for comparison. This might facilitate local referencing in multiple 2D distributions and thus reduce the overall cognitive load without juxtapositioning while preserving all information of each distribution.

To date, however, a vast amount of work in the InfoVis domain has pointed against the use of 3D for abstract data. 3D is accompanied by fundamental issues such as occlusion [26], perceptual distortion [45], and the absence of a common baseline [47]. After due deliberation, however, we believe that there are good reasons why the applicability and usefulness of 3D heatmaps should be further investigated. Reason one are the shortcomings of traditional 2D heatmaps for comparative analysis, as described above. Reason two are new display technologies like Virtual and Augmented Reality (VR/AR) that are becoming more widespread and will necessitate us to think about proper data representations within them. It is not clear yet, in how far drawbacks of 3D data representations will persist in such environments, or if they might be reduced or even balanced out in comparison to conventional screen-based visualizations. Properties unique to VR approaches such as available degrees of freedom, cognitive immersion, or interaction possibilities have been identified as beneficial in many use cases. Among others, related research revealed benefits of VR in terms of improved spatial memory [38], learning performance [50], spatial understanding [14], the understanding of geometric models [62], and collaboration aspects [17]. Advantages of using VR to observe established visualizations could be demonstrated in several cases, for example for scatterplots [3], graph visualizations [20], and flow visualizations [23].

Towards better understanding of this new design space, we present a prototype for the interactive exploration of data distributions with 3D heatmaps. In addition to standard functionality for exploratory analysis, our prototype specifically supports comparative analysis tasks on multiple heatmaps. To do so, we employ a novel interaction metaphor, where users can shift stacked heatmaps into each other for spatial and numerical comparison. In a quantitative user study we focus on the prototype's capabilities for comparative analysis and compare our visualization approach of stacked 3D heatmaps with the conventional approach of juxtapositioned 2D heatmaps. Based on a literature review, we identified the most common types of tasks in comparative heatmap analysis, including *Lookup-Tasks*, *Locate-Tasks*, and *Overview-Tasks*. In addition to comparing the two different visualizations, we also tested the type of *Medium* in our experiment. Half of the participants conducted the study on a conventional monitor screen, the other half in a virtual reality environment, additionally allowing us to assess the impact of VR on the analytic performance of users.

In summary, we make the following two main contributions: *(i)* we present a 3D heatmap prototype in VR that supports the comparative analysis of multiple distributions, and *(ii)* based on this prototype, we present the results of an empirical study comparing the performance of 2D and 3D heatmaps for comparison tasks in virtual and conventional screen environments.

RELATED WORK

In the following, we will first provide a brief overview of where and how heatmaps were used in previous works. Next, we will summarize motives for deploying virtual reality as a medium for the observation of visualizations. Subsequently, we survey several existing approaches to comparative visualizations, with a focus on heatmaps.

Heatmap Visualizations

Heatmaps are a well-known technique to visualize continuous data. Their applicability and usefulness has been demonstrated in various domains, for example, in medicine for volume surface visualizations [55], in geography for temperature visualizations [15] or even for abstract trajectory analysis [52]. Often, heatmaps are used for the presentation of 2D distributions, which are the result of statistic evaluations [36, 16]. They are also frequently deployed for lining up and comparing two or more results, such as different experimental conditions or pre-post comparisons [52, 35, 49]. For the comparative analysis of 2D heatmaps, there is a large number of different techniques for merging two or more heatmaps into a blended view of them. For instance, Jo et al. [34] present various approaches to visualizing two density maps in one visualization using different blending techniques.

Three-dimensional heatmaps, also referred to as heightmaps, extend 2D heatmap visualizations by double-encoding the "heat" by a position as well (i.e., height). Of course, color can be replaced entirely or used to encode an additional attribute. Most commonly, heightmaps are associated with geographic visualizations such as OpenSpace [41] or Google Earth [25] in which landscape elevations are mapped to height. However, 3D heatmaps have also been deployed in a variety of visualizations for more abstract data, such as in sound analysis for frequency visualizations [51] or in medicine for the analysis of vascular movements [60]. Büschel et al. [10] used heightmaps to investigate spatial interaction in AR environments on 3D data visualizations. Tory et al. [56] present an empirical study for a search and value extraction task on scatterplots and 3D data landscapes. For the investigated task, the point-based spatialization was superior compared to the 3D heatmap-like representation. Our work adds to the line of work of empirically studying heatmaps, by focusing on the aspects of 2D vs. 3D heatmaps and the impact of stereoscopic perception and immersion on performance.

3D Visualizations and Virtual Reality

In general, 3D visualizations are not the preferred solution for abstract data as they are accompanied by flaws like occlusion and perceptual distortion [39]. Sedlmair et al. [53] compared the performance of 2D and 3D scatterplots for cluster verification tasks on dimensionality reduced data and, based on their results, strongly advise against using 3D visualizations for this task. However, in the recent past, virtual reality devices, such as Oculus Rift or HTC Vive, have gained attention in the field of information visualization. Dwyer et al. coined the term 'Immersive Analytics' by defining it as "the use of engaging,

embodied analysis tools to support data understanding and decision making" [19]. In their book, Marriott et al. [37] provide a collection of papers that characterize this research area. Among others, they point out that immersive analytics can be an opportunity for decision making and knowledge generation even for abstract data. The deployment of virtual reality environments for 3D visualizations has proven advantageous in some cases [27, 62, 48]. According to Donalek et al. [17], improved depth perception in VR leads to a better overall perception of the datascape geometry and a better understanding of the data in graph visualizations. Similarly, Erra et al. [20] found a beneficial effect of VR on graph exploration tasks. Etemadpour et al. [21] conducted several studies comparing stereoscopic visualizations to projection-based ones. They found that surface-based visual encoding benefits more from a VR setting than point-based renderings.
We add to this line of research by assessing design factors that have not been studied so far. For comparative tasks, superpositioned 3D heatmaps could pose a benefit compared to conventional juxtapositioned heatmaps when inspected in VR.

Comparative Visualizations

The comparison of two or more data sets is a frequent task in visual analysis. Hence, a vast amount of different visualization techniques and approaches exist for comparing several visualizations. Most commonly applied are side-by-side visualizations [52, 33]. With this approach, the observer has to find the same position in each visualization, which can be a tedious and inaccurate task. Linking and brushing can be deployed to ease this process. For time-series comparisons, Gleicher et al. [24] presented several strategies. Besides juxtaposition and superposition, the signals can be merged by calculating a difference signal and displaying it instead. However, each of the named strategies has its benefits and drawbacks. Alabi et al. [1] surveyed various techniques to compare surface visualizations. Among others, they listed the usage of transparency in combination with overlapping surfaces, the partitioning of the surfaces into slices aligned in alternating order, and the utilization of semi-opaque textures. Multiple coordinated views can be used to look at one data set from different perspectives. This design consists of different windows in which different projections or visualizations of the same data entity are displayed [32, 40, 8]. For comparative analysis, any comparison technique can be used separately in each coordinated view.

In our work, we add to this strain of research by assessing the performance of superpositioned 3D heatmaps for typical comparative tasks. Moreover, we assess the impact of immersion on user performance by deploying the compared types of visualization on a conventional screen and in a virtual reality environment.

PROTOTYPE DESCRIPTION

In this section, we will first discuss some general design considerations for visualizing 3D heatmaps in VR, and how we implemented them in our prototype. Subsequently, we will focus on comparing heatmap visualizations and explore the possible advantages of 3D and VR. Finally, we will investigate how the interactive embedding of 2D visualizations in the 3D design space can help the user to overcome the disadvantages of 3D representations. In the following section we will then focus on evaluating one aspect of the prototype in more detail: its capabilities for comparative analysis.

Design Considerations

In 3D heatmaps, the third dimension can be used to double encode the value by color and height. This strengthens the encoding since the value is additionally encoded by the more powerful visual variable "size" [4]. However, this comes at a price: drawbacks caused by the nature of 3D visualizations such as occlusion and perceptual distortion appear on the scene. Previous research has demonstrated a potential benefit of VR in various immersive analytics use cases such as improved spatial memory, more natural interaction capabilities, and better depth perception [17, 27, 62]. In order to compensate for disadvantages caused by the three-dimensionality of the visualization, we deploy the visualization in a virtual reality environment. Concerning comparative tasks, the three-dimensional visualization has the advantage that superpositioning is possible. This could ease spatial referencing and reduce the mental workload compared to juxtapositioned 2D visualizations. To follow up on this presumption, we strive to evaluate the performance of 3D heatmap visualizations for comparative analysis tasks in virtual reality environments. We assume that the three-dimensional visualization, in combination with stereoscopic vision, could be advantageous for such tasks. Therefore, we developed a prototype that provides a platform for 3D heatmap visualizations and the associated functionality for explorative analysis. In the following, we present the three most important design considerations for our prototype:

1. To facilitate an interactive visual exploration workflow, we follow Shneiderman's well-established "information seeking mantra" [54]: overview first, zoom and filter, and details on demand. A 3D heatmap visualization is used as a base visualization that provides the user with an overview of the distribution. The visualization environment must supply the functionality to adapt the visual encoding (color coding) and the representation of data (sampling rate, normalization). Furthermore, it must provide the functionality to enlarge areas of interest (zoom) and to filter data. It must also provide the functionality to extract exact value information for points of interest.

2. The prototype must provide the functionality to display multiple visualizations simultaneously, thus enabling comparative analysis tasks. We decided to align the 3D heatmaps horizontally in order to obtain a common plane of reference and, at the same time, provide the metaphor of natural landscapes. Individual layers of heatmaps should be movable only in the vertical direction, preserving spatial referencing. Shifting could help users to identify connected surfaces and partially overcome problems associated with occlusion. For instance, one heatmap can be shifted through the other until its surface pierces through the surface of the other at a specific location of interest. The extent of the required shift then indicates the offset of values at the given location.

3. Although 3D visualizations have several disadvantages, they also offer new design possibilities and metaphors. In order to

take advantage of 2D visualizations, the visualization design space should not be limited to 3D. Users should be able to seamlessly transition data visualized in the 3D representation into 2D projections of the same data. The user must be able to create 2D aggregations from the 3D representation and display them in the visualization environment in order to overcome drawbacks associated with 3D visualizations.

Prototype: Base Visualization

The visualization environment was developed using the gaming engine Unity3D [57] and consists of a 5m×5m room with a table in the center, surrounded by white walls. The 3D heatmap visualization is placed on top of the table and framed by axes at each corner. The visualization can be inspected through an HTC Vive Pro [31] head-mounted display (HMD) by walking around the table.

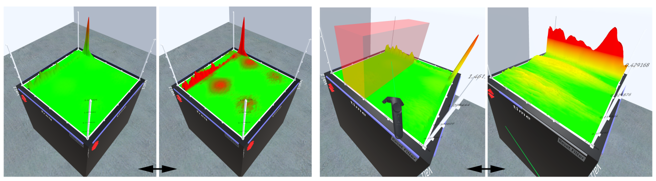

Figure 1. The user can interactively change the color encoding (left). Data ranges can be selected and applied as a filter to zoom the visualization to the selected data range (right).

As depicted in Figure 1 (top), the visual encoding can be adapted interactively. The user can change the coloring of the heatmap, emphasizing certain data ranges. Further, it is possible to select data ranges in the heatmap by hovering over an arbitrary axis and zoom into the selected data range (see Figure 1, bottom). Details on certain locations can be obtained by clicking on it, which opens an information pop-up.

Comparison of Multiple Heatmaps

Comparing two visualizations to find correlations or other coherences is a frequent task [24]. While conceivable, directly comparing two heatmaps with one using color and the other using opacity to encode the values is perceptually extremely challenging and not feasible in practice. Thus, in traditional visualization environments, side-by-side or selective, confined comparisons are state of the art. Another option is to merge the data sets of both visualizations, which are to be compared into a new visualization, e.g., by creating difference maps or aggregated views. Juxtapositioned small multiples increase the cognitive load in comparative tasks since the user has to coordinate his or her attention between two or more different visualizations. I.e., a particular location in one heatmap has to be located in a second heatmap in order to compare the values. Aggregated visualizations that simultaneously encode information of two or more distributions often lack crucial information. For instance, when creating a difference map of two heatmaps, only the value offset is displayed, whereas the absolute level of values disappears. Of course, these views can be displayed additionally to juxtapositioned small multiples, but this further increases the overall mental workload. Interaction methods, such as linking and brushing, can reduce the mental effort, but the fundamental challenges remain.

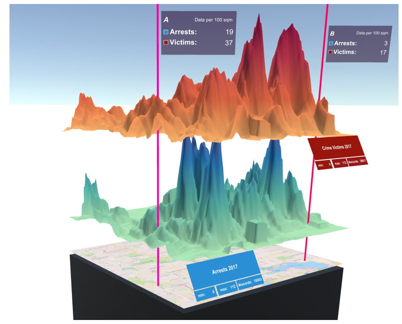

Figure 2. Stacked 3D heatmaps for comparative tasks. In this example, one heatmap (blue) shows the frequency of arrests at any location in a district of Baltimore. The second heatmap (red) depicts locations where crimes were reported by victims. The analyst can select any location on the map to get detailed information about that location (pillars).

Stacked 3D heatmaps in VR naturally lean themselves towards interaction techniques that allow intuitive comparisons of two or more heatmaps. Instead of coordinating attention between several visualizations, as it is the case with small multiple visualizations, we propose to vertically shift heatmaps into each other. In doing so, values can be compared along the vertical axis through a specific point of interest. Figure 2 depicts two horizontally aligned heatmaps stacked on top of each other. For optimal visual differentiation, each displayed 3D heatmap should have a distinct color scale. Each value in the heatmaps is double encoded by value and height, where color and height can be used to compare values within the same heatmap, and height can be used for efficiently compare values of two heatmaps. The user can shift the heatmaps interactively along the vertical axis. Shifting one heatmap relative to another eases the detection of coherence between them when peaks of one heatmap appear and rise through the other (see Figure 3). For exact value comparison, heatmaps can be snapped in with aligning axes to establish a common baseline for the visualizations.

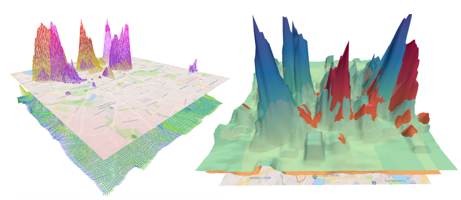

Figure 3. Superimposed heatmaps, one encoding arrests and the other reported crimes. Meshing (left) or transparency (right) can be used to overcome occlusion. The geo-map layer can be used as a cutting plane to serve as a common plane of reference for shifted peaks or to cover distracting parts of the visualization (left).

Besides shifting the heatmaps and the base map arbitrarily, the prototype provides further parameterization options: As shown in Figure 3, a base geo-map can provide further spatial reference (left), and the appearance of the heatmap itself can be customized to display a meshed surface (left) or a semi-transparent surface (right). As these representations allow users to look through surfaces, the user gets the possibility to see how strongly the values of the different heatmaps correlate in certain areas. The color maps can be adjusted interactively.

For better readability and comparability, a user can interactively place labels anywhere on the map to compare values of one or more different points between all displayed heatmaps (see Figure 2, labels A and B). The pink lines perpendicular to the map support the comparison as visual cues, pointing out the selected position on all layers.

Transformation & Projection from 3D to 2D

Three-dimensional visualizations can have some disadvantages in visual analytics tasks. For example, occlusion and perspective distortion can occur. To overcome problems arising from 3D visualizations, we integrated the possibility to create 2D projections from selections in the 3D visualization seamlessly. The user can select the data to be projected using the selection box tool (see Figure 4). By pressing and holding the trigger button on one of the sides of the selection box, an aggregation is generated and attached to the controller. It can then be placed on a wall by releasing the button at the desired position. The side of the selection box that was clicked determines by which axis the data is aggregated. As if the selection box had been compressed, only the selected side remains. Projections can be arbitrarily organized, deleted, or supplemented by annotations (drawing function). For the aggregation, currently, the third dimension that needs to be reduced is averaged to demonstrate the concept. Of course, any other aggregation function can also be used.

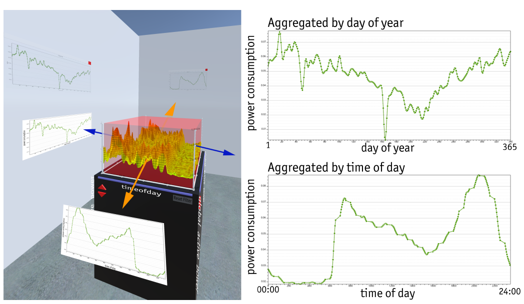

Figure 4. 2D projections of 3D heatmap. In this example, all data is selected (red selection box, left). Two 2D line charts (right) were created interactively by selecting two different sides of the selection box, resulting in different forms of aggregation. Example: single household power consumption over one year.

STUDY: HEATMAPS FOR COMPARATIVE ANALYSIS

In order to evaluate the previously presented approach of stacked 3D heatmaps, we conducted a user study with 48 participants. We compared the novel approach with the conventional analog approach of juxtapositioned 2D heatmaps. To assess the role of VR, we also considered the 2D version in a virtual reality environment.

Study Design

The conducted study comprises two experimental factors: *Medium* (Screen, VR) and *Dimensionality* of the visualization (2D, 3D – see Figure 5). We used a between-subjects design to avoid learning effects. The sample was divided into four groups, each of which is a combination of the two factors (*Screen2D*, *Screen3D*, *VR2D*, *VR3D*).

The 2D condition (Figure 5, left) consists of four heatmaps. The upper two heatmaps are the distributions that have to be compared. Each of the bottom two heatmaps is an aggregation of the upper two distributions. The first one (bottom left) is a difference map (values of heatmap A minus values of heatmap B), and the second one (bottom right) displays the sum of the values of both heatmaps (normalized to range between 0 and 10). The aggregation views were added to make a fair comparison to state-of-the-art methods for comparative analysis [53]. We used a color scale (blue-white-red) that is frequently deployed for comparative tasks of heatmaps with negative values [46, 28, 7, 12]. In the *VR* condition the visualization was attached to a wall standing in the virtual environment and participants were able to move in the virtual space, whereas in the *Screen* condition the 2D visualization was centered on the screen and no motion interactions were provided.

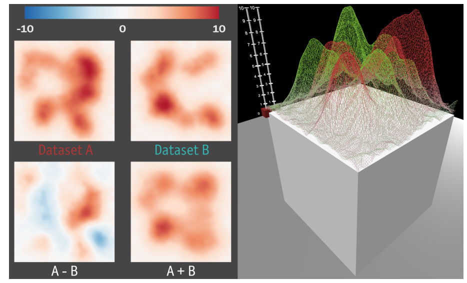

Figure 5. Heatmap visualizations. In the 2D condition (left), two 2D heatmaps represent the distributions to be compared (left, top). Additionally, two aggregated heatmaps show combinations of both distributions (left, bottom). In the 3D condition (right), two 3D heatmaps are superpositioned on top of a cube. Each of them has a uniform color for visual distinction.

The visualization of the 3D condition is depicted in Figure 5 on the right. Each distribution is visualized as a meshed 3D heatmap with a uniform color for visual distinction. We abstained from using unique color gradients for each heatmap for improved visual distinctness of the two distributions. To increase the controllability and fairness of the study, we removed advanced interaction capabilities, such as filtering, and creating 2D projections from 3D heatmaps. Participants were able to spatially navigate in the visualization environment using either keyboard and mouse (*Screen* condition) or body movements (*VR* condition). As a further interaction, it was possible to shift each heatmap up and down by dragging its red anchor cube vertically. Apart from these, there were no further interaction capabilities in the study (e.g., no zooming, no adaption of color maps).

While we sought to make the comparison between 2D and 3D conditions as fair as possible, there are some limitations stemming from the interactive nature of the design space that we intend to study. Limitations caused by these differences in interaction capabilities are discussed in the limitation section.

Comparative Tasks

For an appropriate selection of tasks, we surveyed all IEEE Vis papers of the last five years. We identified 54 papers that used heatmaps for comparative tasks and classified them into four types of tasks provided in the visualization tasks taxonomy by Brehmer and Munzner [6]: *Lookup*, *Locate*, *Browse* and *Explore*.

We did not distinguish whether the comparative task was the main focus of the paper or just implicitly mentioned in the presentation of the results. Based on the comparative tasks found, we created an abstract version of the tasks with the aim of reflecting the common purpose of each category. Due to limited controllability in the study for default *exploration* (location and target unknown) and *browsing* (location is known and target unknown) tasks, we refrained from adopting them directly. Instead, we identified an essential element from both types of tasks and merged them into a single *Overview* task. For *exploration* and *browsing* tasks, the analyst must understand the overall distribution of the heatmaps to be compared.

Lookup

Target and location are known. For comparative analysis, this means that the value at a specific location has to be extracted from two different heatmaps in order to be compared with each other. For instance, Wang et al. [59] developed a visualization technique for networks in which they deployed heatmaps in the background to visualize density. They compared multiple of these heatmaps with each other by picking out a location of interest in a heatmap and reference the same location in a second heatmap. Various other examples for comparative tasks on heatmaps exist that compare values at a specific location in multiple heatmaps [42, 12, 58]. To avoid the unfair comparison between the visual variables color and height, we abstained from asking participants to extract exact values from heatmaps and created a task in which participants should estimate the distances between pairs of locations and compare the relative difference of distances. Therefore, we placed two markers in each heatmap. For each heatmap, participants should estimate the value offset and compare it relatively to the other one. Instead of asking for the total difference, participants should only indicate - with “Yes” or “No” - whether the value offset in the first heatmap is higher than the value offset in the second heatmap (see Figure 6). To indicate the positions to be compared, we inserted colored markers into the respective visualization. In the 2D condition, the 3D pins could be perceived as colored dots when inspected from above.

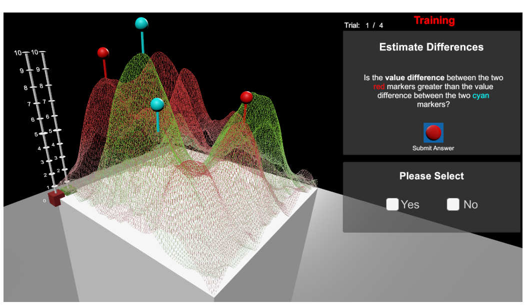

Figure 6. Study interface. Interaction board is attached to the right border of the screen in the *Screen* conditions and attached to the left VR controller in the *VR* conditions.

Locate

Target is known, but the location is unknown. For comparative analysis, this describes a class of tasks in which the analyst visually searches for common characteristics in both heatmaps. For instance, Papadopoulos et al. [44] visualized experimental results of user movement as heatmaps. They then visually compared the heatmaps of different tasks and focused on finding common hot-spots in several heatmaps. Various analogue examples can be found in literature [2, 18, 61]. We used a task where participants had to find two locations where both heatmaps had an equally intense hot spot. Participants were asked to point out the identified shared peaks. Using the respective input device of the condition, participants could click on a heatmap to create a marker at the selected position. Markers could be re-positioned arbitrarily. There were precisely two shared peaks in each heatmap pair in all trials of this task.

Overview

For many tasks of the classes *Explore* and *Browse*, it plays a vital role in keeping track of the entire distribution. Borkin et al. [5] compared pairs of eye-tracking fixation heatmaps in an exploratory manner. They did not only search for new, interesting properties in both heatmaps (*explore*), but also picked out locations of interest and investigated correlations between the two heatmaps at that position (*browse*). Many other examples of papers exist, in which such exploring and browsing tasks were applied [22, 43, 11]. In most of the tasks, an overview is a crucial factor for solving the task efficiently. Therefore, we deployed a task with which we could assess how well the overall distributions in two heatmaps can be compared. Half of the presented pairs of heatmaps were flipped and rotated versions of each other with different noise levels. So the overall structure was similar, but the overall appearance was slightly different. Participants had to judge if the second heatmap was a transformed version of the first one. Like this, we could assess if participants were able to keep track of the overall distribution in each heatmap.

Data

In order to conduct a controlled user study, we created a set of distinct distribution pairs with certain, measurable characteristics. For each task, 14 distribution pairs were generated. For each distribution, we placed 10 to 20 Gaussian kernels randomly on a 100×100 pixel grid. Each position in the grid can contain a value between 0 and 1. The kernel size (30 - 60 pixel) and the peak value at the center of the kernel (0.5 - 1.0 value points) vary randomly within the specified ranges. In a pilot study, we experimented with different parameter settings and identified the one used as the one with the best results. Participants were able to solve roughly 50% of the tasks correctly. In the end, we added random noise to each distribution (0.0 - 0.4 value points).

Lookup

As this task does not need any further constraints, distribution pairs were generated as described above. This results in pairs of distinct distributions as depicted in Figure 7 (left).

Locate

For this task, precisely two locations in both distributions of a pair need to have the same value. Therefore, we added only 8-18 random peaks in each distribution and added two more common peaks in both of them. This guarantees that for each pair, exactly two peaks exist that are of the same height and at the same location (see Figure 7, center).

Overview
For this task, half of the distribution pairs were not altered (see Figure 7, left). For each pair of the other half, one distribution was randomly generated. Its counterpart was then generated by rotating it one to three times by 90° and flipping it on the horizontal axis between each rotation with a probability of 50%. The noise was applied to each distribution separately. This results in distribution pairs with one being a mirrored and rotated version of the other as depicted in Figure 7 (right).

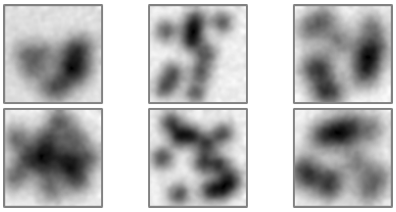

Figure 7. Sample distribution pairs for each task. Left: *Lookup*, center: *Locate*, right: *Overview*.

In order to be able to qualitatively assess differences in the four conditions for real-world data as well, we included an open discussion trial in which we showed real-world data to the participants and discussed it. For this, we used the Baltimore crime dataset depicted in Figure 2.

Procedure
Participants were welcomed and gave written informed consent. They were then introduced to the topic by reading an information sheet. The study supervisor immediately clarified questions that arose during the reading. After participants were familiar with how to interpret the base visualization (2D/3D heatmap), they were prepared to start the study trials by sitting down in front of the monitor or fitting the HMD on their heads.
The experiment was structured into three main parts. First, participants executed a block of 42 trials in total (four training trials + ten trials in each task). For each of the three tasks, they first completed four training trials in which they were introduced to the current task and confronted with the correct answer. Once they fully understood the task, participants completed ten trials without any support from the study supervisor. The order of tasks was counterbalanced (Latin Square design). The order of the deployed data was randomized (each task had a pool of datasets).
Second, participants completed ten memorization tasks. In this task, participants had to indicate for one heatmap whether the given distribution was part of the previous five trials. Half of the distributions shown were selected from the ten available candidates, and the other half were new distributions that were not shown in any of the previous trials.
Third, a real-world crime dataset was displayed and discussed with the participants. While viewing the visualization, the study supervisor explained the dataset. We showed two distributions, one depicting arrests in the city of Baltimore and the other showing reports of crimes. Subsequently, the study investigator asked several questions to determine if the visualization was well understood. Next, an open discussion on the situation in Baltimore was initiated.
After these three main parts, participants were asked to fill in three questionnaires: NASA Task Load Index (TLX), System Usability Scale (SUS), and a custom questionnaire. Finally, they were compensated for participating (10 €).

Apparatus
The experiment took place in laboratories at the University of Konstanz and the University of Stuttgart. In addition to the participant, a study supervisor was present in each session. Participants in the *Screen* condition sat in front of a 24" monitor with a resolution of 1920 × 1200 pixel. In this condition, participants interacted with the study software using a mouse and a keyboard. In the *VR* condition, participants were equipped with an HTC Vive Pro [31] and two Vive controllers.

Sample
A sample of $N = 48$ participants (28 female, 20 male) was recruited via invitations on social media channels, mailing lists, and flyers distributed around the universities. Most of the participants did not have much experience with virtual realities ($Mdn = 2$), heatmaps ($Mdn = 1.5$), and information visualization ($Mdn = 2$). Medians represent experience ratings of users on a scale from *very few* = 1 to *very much* = 5.

Dependent Variables
For each task (*Lookup*, *Locate*, *Overview*), we assessed the error rate and the task completion time. The error rate for the tasks *Lookup* and *Overview* was calculated as the percentage of incorrectly answered trials. For the *Locate* task, we calculated for each participant whether both markers were set within a small radius (10 cm) around the ground truth position of the shared peaks. The error rate was then calculated similarly to the other two tasks. The task completion time was measured as the time between two button clicks (display of visualization and login of the answer).
We assessed participants' capability to recall distributions from the last five trials. For the evaluation, we calculated a memorization rate as the percentage of correctly selected answers.
The NASA Raw Task Load Index (TLX [29]) was used to assess users' overall task load in the respective constellation of *Medium* and *Dimensionality*. Besides, participants completed the System Usability Scale (SUS [9]) to provide feedback on each condition.
Additionally, participants filled in a custom questionnaire assessing their subjective opinion. For instance, the perceived difficulty of each task, the certainty of participants' answers and the level of perceived immersion.
For the real-world data discussion, the study supervisor took notes during the conversation. Additionally, we recorded the entire conversation to encode it after the study in a video analysis procedure. We filtered out where participants had difficulties interpreting the visualizations correctly or comparing the two distributions. Moreover, we summarized which aspects participants mentioned as drawbacks and benefits of the particular condition.

Hypotheses

Based on observations from a pilot study, initial user feedback on our base visualization environment, and related literature, we derived the following hypotheses. All hypotheses were tailored to the conducted study, but could partially be extended to a broader scope.

H1 *Lookup-Task*: We expect participants to perform better with regard to reading values in the 3D condition, due to a more meaningful encoding of values by the variable height instead of color [4]. We further expect that VR poses a benefit compared to the screen representation because of stereoscopic vision due to previous research findings. Hibbard et al. [30], for instance, attributed better depth estimation and improved appreciation of 3D shapes and positions of objects to stereoscopic perception. We, therefore, expect the *VR3D* group to perform best in this task.

H2 *Locate-Task*: We expect lower error rates and task completion times in the 3D condition because of users' improved capability to detect highly granular value changes due to the deployment of the visual variable height instead of color [4]. Moreover, due to superposition, no mental mapping from one heatmap to the other is required in the 3D condition, which should also be reflected in higher performance.

H3 *Overview-Task*: We expect higher performance in the 2D condition due to juxtapositioned visualizations. Etempadpour et al. [21] identified the loss of overview in virtual environments as a critical issue. In side-by-side views, each heatmap can be observed separately while overplotting in the 3D condition hampers the perception of individual structures.

H4 *Memorability*: We expect participants to perform better in the *VR* condition due to increased spatial memory [13]. The spatial component supports the memorization of outstanding features. In VR, all heatmaps and their components are related to a physical location (immersion), whereas on the screen, no direct mapping is established.

Results

We report significant results of our quantitative analysis as well as qualitative feedback from the real-world data discussion. All statistical tests were performed using IBM SPSS Statistics (version 25) and are based on a significance level of $\alpha = .05$. For each dependent variable, we first tested whether the data was normally distributed (Kolmogorov-Smirnov). Depending on the outcome, we used either a one-way independent ANOVA for normally distributed data or its non-parametric counterpart, the Kruskal-Wallis test. Task load (TLX) and usability (SUS) were the only two dependent variables with normally distributed data. As post hoc tests, we deployed the Tukey-HSD test or the Mann-Whitney test (non-parametric). All info graphics depict mean values with error bars indicating the standard error of the mean. Asterisks indicate significant differences (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$).

Error Rate

Error rates differed significantly between the four investigated conditions for the tasks *Lookup* ($H(3) = 13.62, p = .003$) and *Overview* ($H(3) = 23.88, p < .001$). Figure 8 depicts pairwise comparisons between the four conditions. In the *Lookup* task,

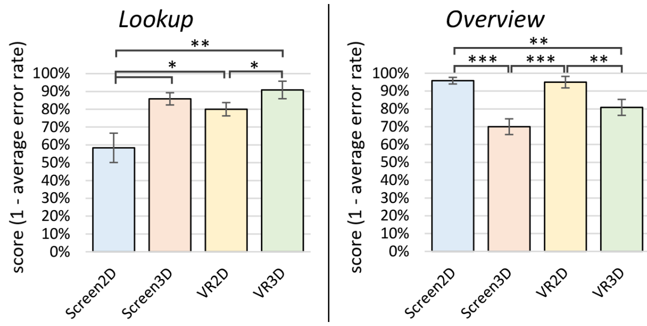

Figure 8. Average user performance scores in the four conditions. The score is calculated as 1 - error rate, where the error rate is the percentage of incorrect answers. The score reflects user performance. Only significant results are reported.

users scored significantly lower in the *Screen2D* condition ($Mdn = 0.65$) compared to all other conditions: *Screen3D* ($Mdn = 0.90, U = 29.50, z = -2.48, p = .013, r = -.36$), *VR2D* ($Mdn = 0.85, U = 38.00, z = -1.99, p = .047, r = -.29$), and *VR3D* ($Mdn = 1.00, U = 102, z = -2.89, p = .004, r = -.42$). Additionally, for the VR conditions, participants in the *VR2D* condition performed worse than in the *VR3D* condition ($U = 108, z = -2.51, p = .012, r = -.36$). Thus, except for the pairwise comparison between *Screen3D* and *VR2D*, participants performed worse in 2D conditions compared to 3D conditions.

For the task *Overview*, participants performed significantly better in the 2D conditions compared to both 3D conditions: *Screen2D* ($Mdn = 1.00$) lead to better results than *Screen3D* ($Mdn = 0.75, U = 7.00, z = -3.86, p < .001, r = -.56$) and *VR3D* ($Mdn = 0.80, U = 103.50, z = -2.82, p = .005, r = -.41$). Similarly, the performance was better in the *VR2D* ($Mdn = 1.00$) condition than in the *Screen3D* ($U = 88.5, z = -3.67, p < .001, r = -.53$) and *VR3D* condition ($U = 105, z = -2.75, p = .006, r = -.40$).

When comparing conditions solely based on the independent variable *Dimensionality*, for all three tasks differences emerge (see Figure 9): *Lookup-Task* ($H(1) = 9.96, p = .002$), *Locate-Task* ($H(1) = 7.12, p = .008$), and *Overview-Task* ($H(1) = 21.92, p < .001$).

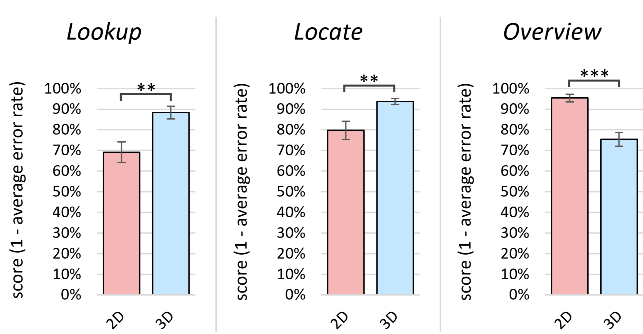

Figure 9. Average user performance scores by *Dimensionality*. The score is calculated as 1 - error rate, where the error rate is the percentage of incorrect answers. The score reflects user performance.

Task Completion Time

As depicted in Figure 10, the tasks *Lookup* ($H(3) = 8.52, p = .036$) and *Overview* ($H(3) = 24.90, p < .001$) revealed significant differences between groups. In the *Lookup* task, participants required more time in the *Screen3D* condition ($Mdn = 21.51$s) compared to both VR conditions: *VR2D* ($Mdn = 9.56$s, $U = 23.00, z = -2.83, p = .005, r = -.41$), and *VR3D* ($Mdn = 11.69$s, $U = 38.00, z = -1.96, p = .050, r = -.28$).

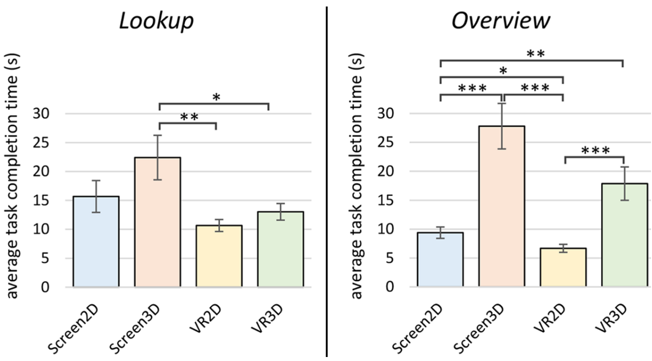

Figure 10. Average task completion times in seconds per task and condition. Only significant results are reported.

In the *Overview* task, participants required significantly more time in both 3D conditions compared to 2D conditions. They were faster in the *Screen2D* condition ($Mdn = 9.04$s) compared to the *Screen3D* condition ($Mdn = 30.42$s, $U = 17.00, z = -3.18, p = .001, r = -.46$) and in the *VR2D* condition ($Mdn = 6.24$s) compared to the *VR3D* condition ($Mdn = 15.79$s, $U = 8.00, z = -3.70, p < .001, r = -.53$). When comparing the two 2D conditions, participants performed faster in VR ($U = 38.00, z = -1.96, p = .050, r = -.28$).

Memorization

With regard to the calculated memorization score, no significant differences emerged between groups or single variables (*Medium*, *Dimensionality*).

Task Load & Usability

Figure 11 depicts the results of the NASA TLX questionnaire and the SUS questionnaire. For the task load an overall difference between groups could be detected ($F(3,44) = 3.67, p = .019, \omega = 0.38$). Post hoc tests revealed only one significant difference between single groups: the condition *Screen2D* ($M = 29.51, SD = 16.35$) was perceived as less demanding than the condition *VR3D* ($M = 49.31, SD = 13.25$).

In addition, the groups differed in terms of assigned usability scores ($F(3,44) = 3.99, p = .013, \omega = 2.91$). Post hoc tests revealed a significant difference between the two VR conditions. Participants evaluated the *VR2D* condition ($M = 32.33, SD = 5.71$) with higher usability scores than the *VR3D* condition ($M = 24.17, SD = 7.59$).

Qualitative Feedback

Throughout the real-world data discussion, various statements were frequently made by participants. Most of them referred to the *Dimensionality* of the visualization. In the *2D* condition, several participants mentioned that the comparison was hampered due to the requirement of switching between

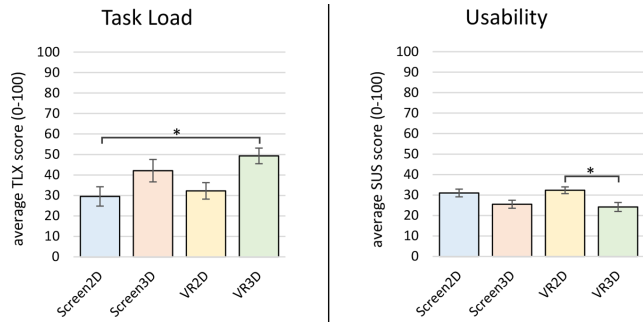

Figure 11. Left: average task load scores of the TLX Questionnaire. High values indicate a high subjectively perceived task load. Right: average usability scores of the SUS Questionnaire. High values indicate high perceived system usability.

two different visualizations and finding one position in the two heatmaps ($n = 5$). Moreover, it was difficult for them to make out small value changes from the linear color gradient ($n = 4$). Opinions differed regarding the usefulness of aggregation views: Two participants said that they were particularly helpful and two said they saw no benefit in them.

For the *3D* condition, participants indicated the low resolution of the surface grid ($n = 4$) as a hindering factor. Two mentioned that the capability to shift one layer into the other eases the distinction between the two layers and the detection of commonalities and differences. Participants also mentioned occlusion and overlap as factors that limited their overall performance. In the *Screen3D* condition, participants found the interaction with keyboard and mouse unfamiliar ($n = 4$). In the case of the *VR3D* condition, two participants emphasized that stereoscopic vision was advantageous to observe the 3D structure of the heatmaps, to distinguish them, and find differences such as common peaks.

DISCUSSION

The *Lookup* task required the user to compare pairs of values extracted from heatmaps. Results show that for each medium (*Screen* and *VR*) participants had lower error rates in the *3D* condition. With regard to the task completion time, no significant differences emerged. However, a reversed trend appeared, which is reflected in higher median task completion times in the *3D* condition. This could be due to increased interaction effort with more degrees of freedom to find the optimal perspective on the visualization. Participants' statements also underline their difficulty in perceiving the exact value from colors in 2D heatmaps. Based on these results, we can, therefore, accept hypothesis $H1$.

In the *Locate* task, participants should find commonalities in the comparative analysis. They were asked to scan two heatmaps for positions where both have equally high hot spots. The statistical analysis showed no significant differences between the four conditions in terms of error rate and task completion time. Thus, $H2$ has to be rejected. However, when comparing only *2D* and *3D* conditions, participants performed better with regard to the error rate when using 3D heatmaps. Statements of participants also reflect an advantage of the 3D representation of the heatmaps. The ability to shift heatmaps into each other allows the user to identify the offset between two heatmaps at any position quickly. By shifting one heatmap up and through the other, its peaks rise through the surface of the first one. This could also be helpful for similar tasks where smaller correlations between two heatmaps are of interest.

In the *Overview* task, participants required significantly more time in the *3D* conditions. Moreover, they performed worse in the *3D* conditions in terms of the error rate. Therefore, $H3$ can be accepted. This could be due to an improved overview caused by juxtapositioned heatmaps. In the side-by-side visualizations, the overall distribution can be observed more quickly, whereas perspective distortion, overlapping, and occlusion makes it hard to observe the entire shape of a distribution.

Concerning memorization scores, no significant differences emerged between conditions. $H4$ can, therefore, not be accepted. This could possibly be due to the fact that the visualization was limited to a small space in the virtual environment. Participants were not required to move around a lot. If participants were standing on the 3D heatmaps and surrounded by the visualization, they possibly could have made more use of their spatial memory for the memorization task.

Differences between the two 2D conditions (*Screen* vs. *VR*) were partially significant (e.g., task performance in the *Lookup* task). This unexpected result could be due to different actual sizes in which visualizations were perceived with the respective mediums. While the visualization in the *Screen* condition was limited to a 24" screen, participants could approach the virtual wall in the *VR* condition and thus perceptually enlarge the 2D visualization. Future work could pursue this finding by controlling the size of the visualization in VR.

Moreover, future work could follow up on our findings and extend the study design by an independent variable for the appearance of 3D heatmap surfaces. As Tory et al. [56] discovered for non-immersive 3D landscapes, the effectiveness of the visualization increased when double encoding values with height and color. Additionally, it would be interesting to qualitatively assess a potential benefit introduced by a hybrid visualization design space that facilitates the seamless transition between the 3D visualization and 2D projections.

LIMITATIONS

As with all empirical work, our study comes with limitations. Most importantly we note that the compared dimensionality conditions (2D and 3D) differ substantially in their visual representation and the interaction capabilities offered. We strove to minimize the degrees of freedom for both representations by providing only the tools necessary to complete the given tasks. In the 3D version, the possibility to shift heatmaps up and down is a crucial component of the technique itself, while, in previous research, the 2D version is commonly used without any interaction. This choice might have influenced our results. There is a variety of possible interaction techniques for 2D heatmaps that could potentially increase its performance. Future work might, for example, introduce a tool to interactively filter or select value ranges in the 2D condition. Similarly, we decided not to use double encoded 3D heatmaps. Instead, we tried to evaluate the impact of using 3D by comparing uniformly colored 3D meshes with flat 2D heatmaps. Again, a large amount of different visualization design options exists that could have had an impact on our results, such as colored and semi-transparent surfaces.

With regard to the lookup task, the way the task was set might have affected user performance. Colored markers were placed on the surface in the 3D condition and participants were asked to compare relative value offsets by solely considering the provided annotations (markers). Since we displayed the 3D heatmaps as meshes, participants were able to look at the visualization from the side, reducing the task to a vertical offset comparison task. This favors the 3D condition for this task. Therefore, choosing a different visualization design, such as double encoded surfaces, might lead to different results.

In the 2D condition, we used a blue-white-red color map that is often used for comparative tasks in heatmaps. The choice of the color map can substantially impact user performance though. Therefore, the use of other color maps, which, for example, highlight zero values more clearly, might increase the performance of users in the 2D condition.

We only assessed the comparative analysis of two heatmaps at a time. If more than two heatmaps are compared simultaneously, a matrix of heatmaps might be more scaleable than superpositioned heatmaps. Also, most of the participants did not have much experience with heatmaps. In particular, for aggregated 2D heatmaps, we expect a steep learning curve, which could increase the performance of expert users. Hardware constraints caused by the current state of technology for HMDs may also have affected the overall performance of participants in the VR setting.

CONCLUSION

We presented an approach for the comparative analysis of heatmaps. In a quantitative user study with 48 participants, we compared our approach to a common alternative of juxtapositioned 2D heatmaps. In addition to comparing the two different types of visualizations, we assessed the impact of immersion on the overall performance of users. Results of the user study indicate that for value extraction tasks and property detection tasks, the 3D approach outperforms the conventional visualization in terms of lower error rates, but requires more time. Juxtapositioned 2D heatmap visualizations, on the other hand, were providing a better overview of both distributions, allowing a better comparison on higher levels. We can conclude that 3D heatmap visualizations can indeed be a suitable representation in specific comparative analysis tasks. However, analysts should always consider the cost-benefit ratio when introducing 3D visualizations for abstract data. A possible solution, yet to investigate, is to make use of hybrid design spaces, cherry-picking benefits of both (2D and 3D) visualization design spaces.

ACKNOWLEDGEMENTS

This work was funded by the European Union's Horizon 2020 VICTORIA project under grant agreement No. 740754 and from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 251654672 – TRR 161.

REFERENCES

[1] Oluwafemi S Alabi, Xunlei Wu, Jonathan M Harter, Madhura Phadke, Lifford Pinto, Hannah Petersen, Steffen Bass, Michael Keifer, Sharon Zhong, Chris Healey, and others. 2012. Comparative visualization of ensembles using ensemble surface slicing. In *Visualization and Data Analysis 2012*, Vol. 8294. International Society for Optics and Photonics, SPIE, Burlingame, California, United States, 82940U. DOI: http://dx.doi.org/10.1117/12.908288

[2] Eric Alexander and Michael Gleicher. 2016. Task-driven comparison of topic models. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 320–329. DOI: http://dx.doi.org/10.1109/TVCG.2015.2467618

[3] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister. 2018. 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 (Jan 2018), 457–467. DOI: http://dx.doi.org/10.1109/TVCG.2017.2745941

[4] Jacques Bertin. 2010. *Semiology of Graphics: Diagrams, Networks, Maps*. Esri Press, Redlands.

[5] Michelle A Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. 2016. Beyond memorability: Visualization recognition and recall. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 519–528. DOI: http://dx.doi.org/10.1109/TVCG.2015.2467732

[6] Matthew Brehmer and Tamara Munzner. 2013. A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics* 19, 12 (2013), 2376–2385. DOI: http://dx.doi.org/10.1109/TVCG.2013.124

[7] Matthew Brehmer, Jocelyn Ng, Kevin Tate, and Tamara Munzner. 2016. Matches, mismatches, and methods: multiple-view workflows for energy portfolio analysis. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 449–458. DOI: http://dx.doi.org/10.1109/TVCG.2015.2466971

[8] Dominique Brodbeck and Luc Girardin. 2003. Design study: Using multiple coordinated views to analyze geo-referenced high-dimensional datasets. In *Proceedings International Conference on Coordinated and Multiple Views in Exploratory Visualization*. IEEE, CMV, London UK, 104–111. DOI: http://dx.doi.org/10.1109/CMV.2003.1215008

[9] John Brooke and others. 1996. SUS-A quick and dirty usability scale. *Usability Evaluation In Industry* 189, 194 (1996), 189–194.

[10] Wolfgang Büschel, Patrick Reipschläger, Ricardo Langner, and Raimund Dachselt. 2017. Investigating the Use of Spatial Interaction for 3D Data Visualization on Mobile Devices. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*. ACM, New York, NY, USA, 62–71. DOI: http://dx.doi.org/10.1145/3132272.3134125

[11] Jan Byška, Mathieu Le Muzic, M Eduard Gröller, Ivan Viola, and Barbora Kozlikova. 2016. AnimoAminoMiner: Exploration of protein tunnels and their properties in molecular dynamics. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 747–756. DOI: http://dx.doi.org/10.1109/TVCG.2015.2467434

[12] Nan Cao, Conglei Shi, Sabrina Lin, Jie Lu, Yu-Ru Lin, and Ching-Yung Lin. 2016. Targetvue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 280–289. DOI: http://dx.doi.org/10.1109/TVCG.2015.2467196

[13] Andy Cockburn and Bruce McKenzie. 2002. 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*. ACM, New York, NY, USA, 203. DOI: http://dx.doi.org/10.1145/503376.503413

[14] Robert Coulter, Linda Saland, T Caudell, Timothy E Goldsmith, and D Alverson. 2007. The effect of degree of immersion upon learning performance in virtual reality simulations for medical education, In Medicine Meets Virtual Reality. *Studies in health technology and informatics* 15 (2007), 155. DOI: http://dx.doi.org/10.1097/00042871-200701010-00099

[15] J Huw Davies. 2013. Global map of solid Earth surface heat flow. *Geochemistry, Geophysics, Geosystems* 14, 10 (2013), 4608–4622. DOI: http://dx.doi.org/10.1002/ggge.20271

[16] Madhusmita Dehingia, Narayan C Talukdar, Rupjyoti Talukdar, Nageshwar Reddy, Sharmila S Mande, Manab Deka, Mojibur R Khan, and others. 2015. Gut bacterial diversity of the tribes of India and comparison with the worldwide data. *Scientific Reports* 5, 18563 (2015), 2045–2322. DOI:http://dx.doi.org/10.1038/srep18563

[17] Ciro Donalek, S. George Djorgovski, Scott Davidoff, Alex Cioc, Anwell Wang, Giuseppe Longo, Jeffrey S. Norris, Jerry Zhang, Elizabeth Lawler, Stacy Yeh, Ashish Mahabal, Matthew J. Graham, and Andrew J. Drake. 2014. Immersive and Collaborative Data Visualization Using Virtual Reality Platforms. *CoRR* abs/1410.7670 (2014), 609–614. DOI: http://dx.doi.org/10.1109/BigData.2014.7004282

[18] Soumya Dutta, Chun-Ming Chen, Gregory Heinlein, Han-Wei Shen, and Jen-Ping Chen. 2017. In situ distribution guided analysis and visualization of transonic jet engine simulations. *IEEE transactions on visualization and computer graphics* 23, 1 (2017), 811–820. DOI: http://dx.doi.org/10.1109/TVCG.2016.2598604

[19] Tim Dwyer, Kim Marriott, Tobias Isenberg, Karsten Klein, Nathalie Riche, Falk Schreiber, Wolfgang Stuerzlinger, and Bruce H Thomas. 2018. Immersive analytics: An introduction. In *Immersive Analytics*. Springer, 6330 Cham, Switzerland, 1–23. DOI: http://dx.doi.org/10.1007/978-3-030-01388-2_1

[20] Ugo Erra, Delfina Malandrino, and Luca Pepe. 2019. Virtual reality interfaces for interacting with three-dimensional graphs. *International Journal of Human–Computer Interaction* 35, 1 (2019), 75–88. DOI: http://dx.doi.org/10.1080/10447318.2018.1429061

[21] Ronak Etemadpour, Eric Monson, and Lars Linsen. 2013. The effect of stereoscopic immersive environments on projection-based multi-dimensional data visualization. In *Proceedings of the International Conference on Information Visualisation*. IEEE, London, UK, 389–397. DOI: http://dx.doi.org/10.1109/IV.2013.51

[22] Rebecca Faust, David Glickenstein, and Carlos Scheidegger. 2019. DimReader: Axis lines that explain non-linear projections. *IEEE transactions on visualization and computer graphics* 25, 1 (2019), 481–490. DOI: http://dx.doi.org/10.1109/TVCG.2018.2865194

[23] Andrew S Forsberg, David H Laidlaw, Andries Van Dam, Robert M Kirby, GE Kafniadakis, and Jonathan L Elion. 2000. Immersive virtual reality for visualizing flow through an artery. In *Proceedings Visualization 2000. VIS 2000 (Cat. No. 00CH37145)*. IEEE, Los Alamitos, CA, USA, 457–460. DOI: http://dx.doi.org/10.1109/VISUAL.2000.885731

[24] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D Hansen, and Jonathan C Roberts. 2011. Visual comparison for information visualization. *Information Visualization* 10, 4 (2011), 289–309. DOI: http://dx.doi.org/10.1177/1473871611416549

[25] Google Earth 2019. Google Earth Computer Program. https://www.google.com/earth/. (2019). Accessed: 2019-09-17.

[26] Antonio Gracia, Santiago González, Víctor Robles, Ernestina Menasalvas, and Tatiana Von Landesberger. 2016. 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 (2016), 3–30. DOI: http://dx.doi.org/10.1177/1473871614556393

[27] K. Gruchalla. 2004. Immersive well-path editing: investigating the added value of immersion. In *IEEE Virtual Reality 2004*. IEEE, Chicago, IL, USA, 157–164. DOI:http://dx.doi.org/10.1109/VR.2004.1310069

[28] Tobias Günther and Holger Theisel. 2017. Backward finite-time Lyapunov exponents in inertial flows. *IEEE transactions on visualization and computer graphics* 23, 1 (2017), 970–979. DOI: http://dx.doi.org/10.1109/TVCG.2016.2599016

[29] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, Amsterdam, Netherlands, 139–183. DOI: http://dx.doi.org/10.1016/S0166-4115(08)62386-9

[30] Paul B Hibbard, Alice E Haines, and Rebecca L Hornsey. 2017. Magnitude, precision, and realism of depth perception in stereoscopic vision. *Cognitive Research: Principles and Implications* 2, 1 (2017), 25. DOI:http://dx.doi.org/10.1186/s41235-017-0062-7

[31] HTC Vive 2019. HTC Vive Developer Website. https://www.vive.com/eu/. (2019). Accessed: 2019-08-27.

[32] Alexander Hubmann-Haidvogel, Arno Scharl, and Albert Weichselbraun. 2009. Multiple coordinated views for searching and navigating web content repositories. *Information Sciences* 179, 12 (2009), 1813–1821. DOI: http://dx.doi.org/10.1016/j.ins.2009.01.030

[33] Victoria Interrante, Henry Fuchs, and Stephen M Pizer. 1997. Conveying the 3D shape of smoothly curving transparent surfaces via texture. *IEEE Transactions on Visualization & Computer Graphics* 2 (1997), 98–117. DOI:http://dx.doi.org/10.1109/2945.597794

[34] Jaemin Jo, Frédéric Vernier, Pierre Dragicevic, and Jean-Daniel Fekete. 2018. A Declarative Rendering Model for Multiclass Density Maps. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 470–480. DOI: http://dx.doi.org/10.1109/TVCG.2018.2865141

[35] Paul Klemm, Kai Lawonn, Sylvia Glaßer, Uli Niemann, Katrin Hegenscheid, Henry Völzke, and Bernhard Preim. 2016. 3D regression heat map analysis of population study data. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 81–90. DOI: http://dx.doi.org/10.1109/TVCG.2015.2468291

[36] David Krakov and Dror G Feitelson. 2013. Comparing performance heatmaps. In *Workshop on Job Scheduling Strategies for Parallel Processing*. Springer, Berlin, Heidelberg, 42–61. DOI: http://dx.doi.org/10.1007/978-3-662-43779-7_3

[37] Kim Marriott, Falk Schreiber, Tim Dwyer, Karsten Klein, Nathalie Henry Riche, Takayuki Itoh, Wolfgang Stuerzlinger, and Bruce H Thomas. 2018. *Immersive Analytics*. Vol. 11190. Springer, 6330 Cham, Switzerland. DOI: http://dx.doi.org/10.1007/978-3-030-01388-2

[38] Leonel Merino, Johannes Fuchs, Michael Blumenschein, Craig Anslow, Mohammad Ghafari, Oscar Nierstrasz, Michael Behrisch, and Daniel A Keim. 2017. On the Impact of the Medium in the Effectiveness of 3D Software Visualizations. In *2017 IEEE Working Conference on Software Visualization*. IEEE, Shanghai, China, 11–21. DOI: http://dx.doi.org/10.1109/VISSOFT.2017.17

[39] Tamara Munzner. 2008. Process and Pitfalls in Writing Information Visualization Research Papers. In *Information Visualization*, Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North (Eds.). Springer-Verlag, Berlin, Heidelberg, 134–153. DOI: http://dx.doi.org/10.1007/978-3-540-70956-5_6

[40] Steven Noel, Michael Jacobs, Pramod Kalapa, and Sushil Jajodia. 2005. Multiple coordinated views for network attack graphs. In *Visualization for Computer Security, 2005.(VizSEC 05). IEEE Workshop on*. IEEE, Minneapolis, MN, USA, 99–106. DOI: http://dx.doi.org/10.1109/VIZSEC.2005.1532071

[41] Open Space 2019. OpenSpace Computer Program. https://www.openspaceproject.com/. (2019). Accessed: 2019-09-17.

[42] Cicero AL Pahins, Sean A Stephens, Carlos Scheidegger, and Joao LD Comba. 2017. Hashedcubes: Simple, low memory, real-time visual exploration of big data. *IEEE transactions on visualization and computer graphics* 23, 1 (2017), 671–680. DOI: http://dx.doi.org/10.1109/TVCG.2016.2598624

[43] Cesar Palomo, Zhan Guo, Cláudio T Silva, and Juliana Freire. 2016. Visually exploring transportation schedules. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 170–179. DOI: http://dx.doi.org/10.1109/TVCG.2015.2467592

[44] Charilaos Papadopoulos, Ievgeniia Gutenko, and Arie E Kaufman. 2016. VEEVVIE: visual explorer for empirical visualization, VR and interaction experiments. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 111–120. DOI: http://dx.doi.org/10.1109/TVCG.2015.2467954

[45] Harald Piringer, Robert Kosara, and Helwig Hauser. 2004. Interactive focus+context visualization with linked 2D/3D scatterplots. In *Proceedings - Second International Conference on Coordinated and Multiple Views in Exploratory Visualization, CMV 2004*. IEEE, London, UK, 49–60. DOI: http://dx.doi.org/10.1109/CMV.2004.1319526

[46] Jorge Poco, Angela Mayhua, and Jeffrey Heer. 2018. Extracting and retargeting color mappings from bitmap images of visualizations. *IEEE transactions on visualization and computer graphics* 24, 1 (2018), 637–646. DOI: http://dx.doi.org/10.1109/TVCG.2017.2744320

[47] Nicholas F Polys and Doug A Bowman. 2004. Design and display of enhancing information in desktop information-rich virtual environments: challenges and techniques. *Virtual Reality* 8, 1 (2004), 41–54. DOI: http://dx.doi.org/10.1007/s10055-004-0134-0

[48] Daniel Probst and Jean-Louis Reymond. 2018. Exploring Drugbank in Virtual Reality Chemical Space. *Journal of Chemical Information and Modeling* 58 (6 2018), 1731–1735. DOI: http://dx.doi.org/10.26434/chemrxiv.6629150.v1

[49] Andy Pryke, Sanaz Mostaghim, and Alireza Nazemi. 2007. Heatmap visualization of population based multi objective algorithms. In *International Conference on Evolutionary Multi-Criterion Optimization*. Springer, Berlin, Heidelberg, 361–375. DOI: http://dx.doi.org/10.1007/978-3-540-70928-2_29

[50] Eric D Ragan, Ajith Sowndararajan, Regis Kopper, and Doug A Bowman. 2010. The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments. *Presence: Teleoperators and Virtual Environments* 19, 6 (2010), 527–543. DOI: http://dx.doi.org/10.1162/pres_a_00016

[51] Harold Rodriguez, Diane Beck, David Lind, and Benjamin Lok. 2008. Audio Analysis of Human/Virtual-Human Interaction. In *Intelligent Virtual Agents*, Helmut Prendinger, James Lester, and Mitsuru Ishizuka (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 154–161. DOI: http://dx.doi.org/10.1007/978-3-540-85483-8_16

[52] Tobias Schreck, Jürgen Bernard, Tatiana Von Landesberger, and Jörn Kohlhammer. 2008. Visual cluster analysis of trajectory data with interactive Kohonen Maps. In *2008 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, Columbus, OH, USA, 3–10. DOI: http://dx.doi.org/10.1109/VAST.2008.4677350

[53] Michael Sedlmair, Tamara Munzner, and Melanie Tory. 2013. Empirical Guidance on Scatterplot and Dimension Reduction Technique Choices. *IEEE Transactions on Visualization and Computer Graphics* 19 (2013), 2634–2643. DOI: http://dx.doi.org/10.1109/tvcg.2013.153

[54] B. Shneiderman. 1996. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*. IEEE, Boulder, CO, USA, 336–343. DOI: http://dx.doi.org/10.1109/VL.1996.545307

[55] Paul M Thompson, Kiralee M Hayashi, Greig De Zubicaray, Andrew L Janke, Stephen E Rose, James Semple, David Herman, Michael S Hong, Stephanie S Dittmer, David M Doddrell, and others. 2003. Dynamics of gray matter loss in Alzheimer's disease. *Journal of neuroscience* 23, 3 (2003), 994–1005. DOI: http://dx.doi.org/10.1523/jneurosci.23-03-00994.2003

[56] M. Tory, D. Sprague, F. Wu, W. So, and T. Munzner. 2007. Spatialization Design: Comparing Points and Landscapes. *IEEE Transactions on Visualization and Computer Graphics* 13, 06 (nov 2007), 1262–1269. DOI: http://dx.doi.org/10.1109/TVCG.2007.70596

[57] Unity 3D 2019. Unity 3D Developer Website. https://www.unity3d.com/. (2019). Accessed: 2019-09-17.

[58] Hong Wang, Yafeng Lu, Shade T Shutters, Michael Steptoe, Feng Wang, Steven Landis, and Ross Maciejewski. 2019b. A Visual Analytics Framework for Spatiotemporal Trade Network Analysis. *IEEE transactions on visualization and computer graphics* 25, 1 (2019), 331–341. DOI: http://dx.doi.org/10.1109/TVCG.2018.2864844

[59] Xumeng Wang, Wei Chen, Jia-Kai Chou, Chris Bryan, Huihua Guan, Wenlong Chen, Rusheng Pan, and Kwan-Liu Ma. 2019a. GraphProtector: A Visual Interface for Employing and Assessing Multiple Privacy Preserving Graph Algorithms. *IEEE transactions on visualization and computer graphics* 25, 1 (2019), 193–203. DOI: http://dx.doi.org/10.1109/TVCG.2018.2865021

[60] Niels Willems, Huub Van De Wetering, and Jarke J Van Wijk. 2009. Visualization of vessel movements. In *Computer Graphics Forum*, Vol. 28. Wiley Online Library, IEEE, Chichester, UK, 959–966. DOI: http://dx.doi.org/10.1111/j.1467-8659.2009.01440.x

[61] Wenchao Wu, Jiayi Xu, Haipeng Zeng, Yixian Zheng, Huamin Qu, Bing Ni, Mingxuan Yuan, and Lionel M Ni. 2016. Telcovis: Visual exploration of co-occurrence in urban human mobility based on telco data. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 935–944. DOI: http://dx.doi.org/10.1109/TVCG.2015.2467194

[62] Song Zhang, Cagatay Demiralp, Daniel F Keefe, Marco DaSilva, David H Laidlaw, Benjamin D Greenberg, Peter J Basser, Carlo Pierpaoli, Ennio Antonio Chiocca, and Thomas S Deisboeck. 2001. An immersive virtual environment for DT-MRI volume visualization applications: a case study. In *Proceedings of the conference on Visualization'01*. IEEE, San Diego, CA, USA, 437–440. DOI: http://dx.doi.org/10.1109/VISUAL.2001.964545