Graphical Perception for Immersive Analytics

Matt Whitlock* Stephen Smart† Danielle Albers Szafir‡

University of Colorado - Boulder

Figure 1: We measure perceptual differences across data visualizations in virtual reality (left), augmented reality (middle) and desktop environments (right). We found that people's abilities to accurately interpret data using different visualization designs varies across modalities, offering empirically grounded insight into creating effective visualizations for these displays.

ABSTRACT

Immersive Analytics (IA) uses immersive virtual and augmented reality displays for data visualization and visual analytics. Designers rely on studies of how accurately people interpret data in different visualizations to make effective visualization choices. However, these studies focus on data analysis in traditional desktop environments. We lack empirical grounding for how to best visualize data in immersive environments. This study explores how people interpret data visualizations across different display types by measuring how quickly and accurately people conduct three analysis tasks over five visual channels: color, size, height, orientation, and depth. We identify key quantitative differences in performance and user behavior, indicating that stereo viewing resolves some of the challenges of visualizations in 3D space. We also find that while AR displays encourage increased navigation, they decrease performance with color-based visualizations. Our results provide guidelines on how to tailor visualizations to different displays in order to better leverage the affordances of IA modalities.

Index Terms: Human-centered computing—Visualization—Empirical studies in visualization; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality

1 INTRODUCTION

Immersive Analytics (IA) leverages immersive AR and VR technologies to support data analysis. IA relies heavily on data visualizations rendered in a 3D space. However, guidance for creating effective visualizations comes from studies on desktop displays. IA technologies offer new perceptual affordances like stereo viewing and embodied navigation that may change what it means to effectively visualize data. For example, while three-dimensional depth cues make visualizations harder to read on desktops [54], stereoscopic viewing may alleviate these challenges and make depth a viable channel to encode data. Marriott et al. posit that improved displays, the intuitive use of the third dimension, and the freedom to work beyond the monitor offer unique opportunities for data visualization in IA [32]. While prior efforts discuss benefits of AR and VR for data visualization, we have little empirical evidence for how to effectively design visualizations for immersive displays.

To address this limitation, we measure how different visualizations support effective data analysis across three display modalities: a desktop monitor, a virtual reality head-mounted display (VRHMD) and an augmented reality head-mounted display (ARHMD; Fig. 1). We anticipate that the perceptual and interactive affordances of these displays will affect how well people estimate values from data visualizations. Conventional desktops allow users to leverage a familiar interaction paradigm (mouse and keyboard), but their lack of stereoscopic rendering hinders depth perceptions. VR and AR provide users with stereo viewing and embodied navigation, but are relatively unfamiliar and require users to physically move around to navigate the visualization. AR additionally blends visualizations with real-world environments, adding additional referents that may alter data perception. While prior studies point to differences in user experience across these modalities [3, 5, 13], we instead seek to provide preliminary insight into how we might tailor our visualizations to best leverage different display modalities.

We provide insight into effective visualization design by evaluating how performance varies across modalities for different visual channels used to encode data. Inspired by classical graphical perception studies [8], we examine data visualized using color, size, depth, and orientation in scatterplots and height in barcharts. These channels represent a variety of designs used to encode data in multivariate visualizations and allow us to probe perceptual differences between channels. For example, blending physical and virtual objects in AR may shift color perceptions by changing the contrast between datapoints and their backgrounds, but stereoscopic viewing may make depth encodings more effective. We measure these effects using objective (accuracy and time-to-completion) and subjective (perceived confidence and ease of use) metrics.

We consider display type, encoding channel, analysis task, and dimensionality to provide initial guidelines for how effective visualization design varies across display modalities, offering preliminary steps toward generalizable models for immersive visualization design. We use these findings to rank channel effectiveness for differ-

*e-mail: matthew.whitlock@colorado.edu
†e-mail:stephen.smart@colorado.edu
‡e-mail:danielle.szafir@colorado.edu

2642-5254/20/$31.00 ©2020 IEEE
DOI 10.1109/VR46266.2020.00084

ent display types—comparable to Cleveland and McGill's canonical rankings for 2D visualizations [8]. We aim to provide visualization designers with empirically grounded guidelines for effective immersive visualizations that supplement designer intuitions in IA.

2 Related Work

Studies in data visualization measure how effectively different visualizations designs communicate patterns in data. IA tools can utilize these findings to inform system design. We draw on prior results in immersive analytics and graphical perception to inform our study.

2.1 Immersive Analytics

IA has received considerable attention with the increasing capabilities and availability of HMDs. Prior work has explored the use of IA in domains such as urban planning [1], computational fluid dynamics [30], economics [5], Internet of Things (IoT) applications [17], shopping [12], and maritime analytics [47]. These systems aim to empower users by enabling analysts to display and interact with data in novel ways beyond traditional 2D displays. For example, surgeons can use hands-free visualizations for teaching, enabling side-by-side comparison and remote presence [42, 51].

In creating IA systems, designers identify aspects of immersive displays that could benefit data visualization. For example, while problematic in 2D [54], the use of depth and increased display space may allow users to better understand complicated visualizations. Past systems provide anecdotal examples of these benefits in tasks requiring sensemaking across data tables [5, 9, 47], spatiotemporal data [14, 39], and high-dimensional data [11, 37]. However, empirical studies show that depth perception issues can be exacerbated in AR and VR, finding that people underestimate the depth of virtual objects [10, 25]. Visualizations could leverage different visual cues to account for these limitations in practice [31].

IA tools can also combine modalities to take advantage of trade-offs in different display types for visual analytics. For example, AR allows for the situated display of semantically relevant data by blending data with the physical contexts it describes. Recent efforts combine AR with 2D displays to leverage the strengths of each display modality [48, 52] but rely on intuition to determine when and how to use each. This intuition is used to transition between visualization types depending on context within a single display type [24, 56] or across display types at different levels of immersion [23, 40, 50, 52]. We look to extend these guidelines with a formal empirical analysis of design trade-offs across display types.

2.2 Empirical Studies of IA

The increased prevalence of IA has inspired a number of studies exploring the affordances of IA in different scenarios. For example, Laha et al. identify how aspects of VR fidelity, such as stereo vision and field of regard, impact performance for five different tasks in volumetric visualization [29]. Bach et al. investigated AR for exploratory analysis tasks in scatterplots such as distance estimation and cluster identification [3] but found limited benefits to AR for visualization viewing and interaction. Fafard et al. evaluated a similar set of tasks in Fishtank VR (FTVR), but found notable benefits of stereo vision and motion parallax for point clouds in spherical displays [13]. Kraus et al. found that participants more accurately identified clusters in VR scatterplots than on desktop scatterplot matricies and that the structure of VR visualizations influenced how people navigate data [27]. While these works provide insight into the types of tasks quantitatively suited to IA, they offer little insight into how to design for these platforms. We instead seek to evaluate visualization design for AR, VR and desktop interfaces, focusing on visualization perception across visual channels.

Prior studies of IA systems consider both perception and novel interaction techniques to complement new visualization systems. For example, Bach et al. evaluate not only use of AR for visualization tasks, but also the effectiveness of tangible interactions in aiding with these tasks [3]. Other works have explored how vibrotactile feedback [36] and spatial audio [22] can enhance performance with common visualization tasks. These works provide interesting insight into how we can effectively utilize the multimodal input and output in immersive display, but novel UIs make it challenging to isolate the perceptual benefits of immersive displays from the effects of interaction techniques. As interaction is a key component of IA systems, we consider effects of canonical interaction paradigms—embodied navigation in AR/VR or mouse and keyboard navigation on a 2D display—in measuring design differences across modalities.

2.3 Graphical Perception

The ways we visualize data determine how well people can extract different statistics about that data. For example, people can compare the position of two datapoints more accurately than their size [8]. Graphical perception experiments measure how well different visualizations allow people to extract statistical information from data. We can use the results of these studies to ground recommendations for visualizations that support specific tasks [2, 41] and even to automate visualization design [34]. However, the overwhelming majority of graphical perception studies focus on traditional 2D monitors. We have little empirical insight into how to best represent data in IA.

Graphical perception studies typically focus on a particular encoding channel (e.g., color [43, 49]), a data type (e.g., time series [2, 19, 21]), or a visualization type (e.g., pie charts [26, 45]). These studies can generate a variety of insights into visualization design, ranging from A/B comparisons [41] to ranked lists of channels or designs [8] to quantitative perceptual models [38]. For example, Cleveland & McGill [8] measured how accurately people could compare two values using different visual channels. They created a ranked list of channels from most to least accurate which has long served as foundational guidance for effective visualization design. We adapt these ideas to understand how well channels support effective analysis across display modalities.

We ground our investigation in scatterplots. Scatterplots are among the most common visualization types used in graphical perception studies. They are both familiar to most users and allow designers to freely manipulate most visual channels, including each datapoint's size, color, and shape. Studies have used scatterplots to study how people estimate different statistics like means [18], correlation [20, 38], and similarity [35]. Other studies use scatterplots to measure how well people estimate values over specific channels like color [49] and shape [46]. Our study builds on this work to explore how people perceive different kinds of information across channels and modalities.

Results from graphical perception may not readily translate to other display factors. For example, AR visualizations are overlaid onto the physical environment where the colors and sizes of real-world objects may alter data perceptions [16]. McIntire & Liggett [33] discuss how stereoscopic displays could alleviate limitations of 3D visualizations by drawing on AR and VR studies in other contexts. Prior studies have evaluated visualization techniques like space-time cubes [15] and data highlighting [28] in immersive environments, but these studies emphasize techniques rather than specific design components. Closely related to our work, Kraus et al. [27] found scatterplot cluster detection to be more accurate in VR than on a desktop. We build on these results by measuring performance on an expanded set of tasks and modalities and exploring channels for encoding data beyond position. Our results extend traditional graphical perception approaches to provide empirically grounded guidance for effectively visualizing data on different platforms.

3 Methods

We conducted a 3 (display type) $\times$ 5 (encoding channel) $\times$ 3 (task) mixed factors graphical perception study to measure how

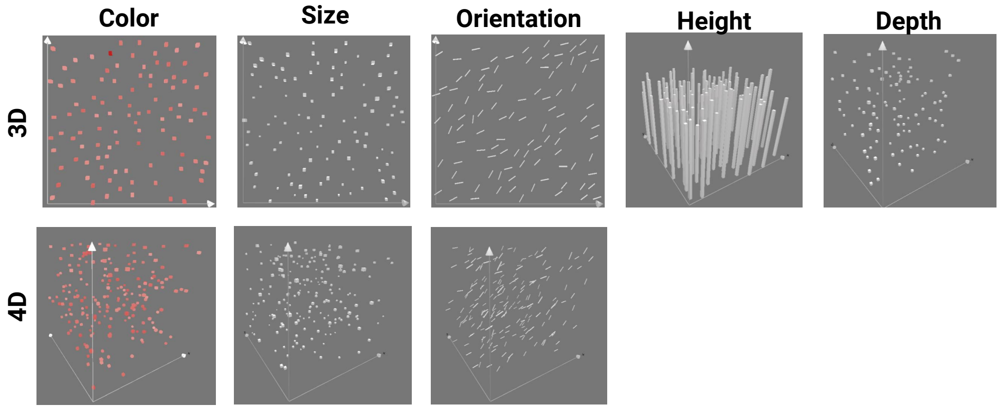

Figure 2: We tested analysis performance for eight visualization designs. Participants saw each visualization design twice per task: once to identify the greatest value, greatest area or direction of increasing trend and once to identify the lowest value, lowest area or decreasing trend.

well people perceive visualized data in IA. Our study tested visualizations on three different displays: desktop, VR using an HTC Vive, and passthrough AR using an HTC Vive and Zed mini (§3.2). We measured performance for visualizations with three and four data dimensions. Participants used these visualizations to estimate values over five encoding channels—color, size, depth, height, and orientation—building on canonical channels from graphical perception [8] (§3.1, Fig. 3). To better reflect the complexity found in real visualizations, we embedded the tested channels in 2D and 3D scatterplots and asked participants to conduct three statistical tasks—extrema detection, trend identification, and mean estimation—to measure how well each channel supports different analysis goals (§3.3). Encoding channel and task were within-subjects factors; display type was a between-subjects factor. Our study infrastructure is available at `https://github.com/CU-VisuaLab/ThreeDPerceptionStudy`.

We developed four hypotheses about objective and subjective performance. We crafted these hypotheses based on theoretical investigations [32, 33], prior studies [3, 27], and prior experience.

H1: Participants will have more difficulty disentangling depth and size on the desktop. When the camera vector is parallel to the axis, a small but near datapoint can appear the same as a large but distant one due to perspective projection. This confound can be rectified by changing viewpoint; however, we anticipate that stereo viewing can help participants more readily disentangle depth and size.

H2: AR and VR will perform comparably for all channels except color. Color in AR will be worse than in VR. AR and VR both provide immersive stereo experiences. However, AR renders data in the visual context of the real world, creating background color variations that may degrade color perceptions.

H3: Performance with height, depth, and visualizations with four data dimensions will correlate with scene navigation. Looking at a visualization from multiple angles will help reveal potentially hidden points in visualizations leveraging the X, Y and Z axes as points may otherwise be occluded by other points. We expect this correlation not to hold in cases where a single viewpoint will reveal all data values (e.g., color, size, or orientation on 2D scatterplots).

H4: Participants will prefer embodied navigation to the mouse and keyboard interface. Mouse-and-keyboard interaction is familiar, but can be cumbersome for navigating 3D spaces. Immersive interactions create more natural means for navigating data [3, 5].

3.1 Stimuli

Participants saw data with either three or four data dimensions. In these visualizations, one dimension mapped to the target visual channel and the remaining channels mapped to positions in a scatterplot. We use scatterplots as they are familiar to most users, allow us to readily control aggregate statistics (trends and means) over different regions of the data, and enable us to test different visual channels in an ecologically valid way without significant interference with other data dimensions. Data along each dimension ranged from 0 to 565, a range selected in piloting based on task and difficulty to show reasonable relative differences and variance within each visualization. Full details of data generation are provided in §3.3.

Visualizations with three data dimensions map 100 datapoints to positions on the X and Y axes (or the X and Z axes for height) and a third channel corresponding to our target channel. Visualizations with four data dimensions encode 200 datapoints on the X, Y, and Z axes and a fourth target channel (Fig. 3). Visualizations were created as prefabricated elements from DxR [44] in Unity, with sampling densities determined in piloting, and rendered at 232mm $\times$ 232mm screenspace to avoid performance biases due to size differences across modalities. Each visualization has a target question rendered above it and a legend to its right.

We tested five encoding channels: size, color, orientation depth, and height. We selected these channels based on their use in canonical studies [8] and as their performance may be impacted by stereoviewing (size, height, depth), projection (size, depth, orientation), or integration with real-world objects (color, height, size).

We encoded size using the edge length of cubes ranging from 4mm to 10mm in screen space when each visualization initially loads. For size visualizations, we attached fixed-size box colliders to each cube to retain consistent selection bounds. Color was mapped to the default sequential color scale in DxR (ranging from light grey (254,224,210) to red (222,45,38) interpolated in HCL). For orientation (comparable to direction in Cleveland & McGill [8]), we encoded data values as rotation about the Z axis where vertical bars represent 0 and horizontal bars represent the highest value. This mapping was determined in piloting. Depth encoded data as a position along the 232mm Z-axis. Height encoded data as the heights of bars originating at points along the XZ-plane and ranging in height by 96mm. We used height in the context of bar charts as, like scatterplots, they position values at the XY position of datapoints, but avoid confounds with vertical position and align values on a common plane. We did not test depth or height with four dimensional visualizations as these channels conflict with the Z position.

For all conditions except color, we used an ambient gradient light from light gray (RGB (242, 242, 242)) to mid-gray (RGB (127, 127, 127)) and back to light gray in the Unity tri-light ambient lighting model. We supplemented this light with a camera-sourced light gray point light (RGB (187, 187, 187)), halfway between pure white and the mid-gray desktop and VR background (§3.2) with an intensity of 75%. In piloting, we found these lighting settings uniformly lit the data to properly show shading on 3D datapoints. The low intensity point light aligned with the participant's gaze removed shadows from the front of the datapoints without emphasizing points directly in front of the camera. For color visualizations, we used only a flat white light since gradient ambient lights non-uniformly lit points in ways that interfered with their perceived color.

3.2 Display Types

To explore the effects of display type on visualization performance, we built a common study infrastructure in Unity using DxR [44]. Where possible, we designed comparable interactions across the desktop, AR and VR to avoid confounds from different UIs. Each modality used industry standard navigation and selection techniques (mouse and keyboard for desktop and embodied navigation and gaze and controller-based selection for AR and VR). All visualizations rendered using a computer with an Intel i7 processor, 8 GB RAM, and an NVidia GeForce GTX 1070. The study space consisted of a 4m $\times$ 2.7m room with one overhead light. To mimic real-world conditions, the room contained objects typical to an office environment: a desk with the desktop computer, a whiteboard, a TV, and a frosted glass wall (Fig. 1). To maximize the navigation space, all furnishings were wall-mounted or against the glass wall.

3.2.1 Desktop

The desktop condition rendered full-screen visualizations on a mid-gray ($L^* = 50$) background displayed on a 28-inch Samsung U28E590D 4K monitor. We constrained the camera to the same 110° field of view as the HTC Vive. Participants navigated the scene with a mouse and keyboard using default Unity key mappings. Left/Right/Up/Down arrow keys allowed participants to pan along the XY plane, and the scroll wheel moved along the Z axis.

Participants used a Logitech Anywhere MX mouse with continuous scroll to more closely replicate continuous embodied HMD navigation. Participants rotated the camera by moving the mouse while pressing the right mouse button. Dragging the mouse in the desired direction adjusted the camera's rotation about the X and Y axes. We used the Windows default cursor and left click for selection.

3.2.2 Virtual Reality

VR visualizations on the HTC Vive rendered using the same mid-gray background ($L^* = 50$). As the Vive is a fully immersive VRHMD, we did not use the same mouse-and-keyboard paradigm as on the desktop. We instead implemented data navigation and selection based on findings from prior literature and industry standards.

Modern HMDs afford embodied data navigation: participants navigate scatterplots by physically moving around the scene. However, selection with HMDs is less standardized. Chatterjee et al. found a gaze-tracked cursor with gestural confirmation comparable to a mouse [7]. We built on these results to enable data selection using a gaze-based cursor and the HTC Vive remote. Similar to the Microsoft HoloLens, a small black cylindrical cursor was rendered directly in the middle of the participant's field of view. Participants "clicked" to select an object by pulling the back trigger of the Vive remote, mirroring the standard selection mapping for the Vive.

3.2.3 Augmented Reality

The AR interface mirrors the VR condition except it utilized a Zed mini camera attached to the front of the Vive for passthrough AR. Visualizations rendered over streamed images of the real world rather than a fixed grey background. As opposed to the see-through displays like the Microsoft HoloLens, the Zed mini augments entire field of view provided by the Zed with virtual elements, leading to a comparable field of view between the AR and VR conditions. Using a pass-through camera also allowed visualizations render with the same color fidelity as in VR. Due to bandwidth limitations, the Zed mini could only feed image data at 720p resolution in order to retain a reasonable framerate (60 fps) and field of view (100.7° diagonally). Given the size and simplicity of marks in our scatterplots, we do not anticipate this resolution had a significant impact on our results.

3.3 Tasks & Datasets

Participants completed three tasks–extrema identification, trend detection, and mean estimation—to evaluate how well different designs support a variety of statistical judgments (Fig. 2). These tasks require both identifying relevant information from individual values and combining magnitudes or changes across multiple values [2]. For symmetry, half of the stimuli required participants to select the largest value or mean or increasing trend, and the other half elicit the smallest value or mean or decreasing trend.

Our visualizations label the target channel as temperature data to contextualize these tasks within familiar measures. We generated synthetic data from Gaussian distributions to regulate the intended effects and task difficulty. Scatterplot point positions were sampled from uniform random distributions along X, Y and Z (where needed) with no overlapping points. Difficulties were calibrated in piloting, targeting an overall accuracy of 70% to avoid ceiling and floor effects. We generated 24 datasets per task $\times$ dimension and randomly mapped them to visualizations for each participant.

Extrema Tasks: Extrema tasks asked participants to identify the maximum or minimum value within a dataset. For this task, temperature values were drawn from within the lowest 82nd percentile of a Gaussian distribution ($\mu = 10, \sigma = 1$). The extrema point reflected values between the 83rd and 91st percentile of the data. We also generated datasets with mirrored conditions for tasks identifying the lowest value: all but one of the temperatures lie in the upper 82% of the distribution, and extrema data points were between the 9th and 17th percentile. We confirmed post-hoc that the target values corresponded to the largest/smallest value after data sampling.

Participants were instructed to select a datapoint using the mouse- or gaze-based cursor, which removed all data but the selected datapoint and loaded a confirmation menu asking them to confirm their selection. Selecting "No" (a mistaken selection) returned participants to the task while selecting "Yes" initiated the next stimuli.

Quadrant Means: Our second task asked participants to identify the quadrant of the visualization with the highest or lowest mean temperature. The X, Y, and (for visualizations with four data dimensions) Z axes were halved, yielding 4 quadrants for three dimension datasets and 8 octants for four dimensions. Datasets were generated using four Gaussian distributions where each quadrant has a different mean ($\sigma = 1$). The differences between the highest and next highest quadrant means ranged from 6% to 12% of the value range.

To avoid having to navigate the visualization to select a quadrant, participants selected any point within the target quadrant to provide an answer. Selection then removed all data except the selected quadrant while participants confirmed their selection. In training, we instructed participants not to use the interface to isolate individual quadrants, but instead only to verify their selection.

Trend Detection: Trend tasks asked participants which direction the data is either increasing or decreasing. For visualizations with three data dimensions, they chose one of eight directions (two parallel to each axis and four towards the corners). For four data dimensions, they chose one of six (two parallel to each axis).

X, Y and Z positions were first randomly generated. Temperature values were then drawn from a Gaussian distribution ($\mu = 50$, $\sigma = 15$) and multiplied by an offset function:

$$f(x,y,\alpha) = \alpha \times (d_x x + d_x y) \tag{1}$$

where x and y correspond to the point's x and y, $d_x, d_y, \varepsilon[-1,0,1]$ control the trend direction, and α is a scaling constant controlling the trend strength. As an example for a trend in the +y direction, each temperature datapoint is a drawn from the Gaussian distribution and increased by α times the y position ($d_x = 0, d_y = 1$) to create a trend toward +y. We varied α between 0.5 and 0.1 based on piloting. For

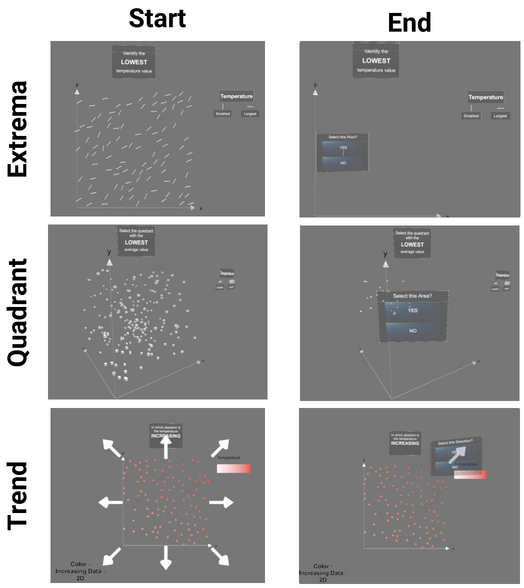

Figure 3: Participants completed extrema, mean, and trend tasks. For mean and extrema, users selected a point to isolated either it (top) or its corresponding quadrant (middle). For trend tasks, selecting a trend arrow isolated that arrow (bottom). Participants then confirmed that the selected point, quadrant, or arrow was the intended target.

all trend fits, $f(x, y, \alpha)$ was constrained such that the line of fit ran through the origin. Trend direction was verified by linear regression.

Participants indicated trend direction using arrows in the periphery of each scatterplot (Fig. 2). Participants were instructed to select the arrow that best represents the direction in which datapoints are either increasing or decreasing. Selecting an arrow loads a confirmation menu that allows the participant to confirm the selection.

3.4 Study Design & Procedure

We explored effects of display and channel on data analysis using a mixed factors design with display type between participants and all other factors within participants. Each participant completed 48 trials blocked by task with block order counterbalanced between participants to avoid learning effects and to maximize the number of trials each participant completed in the session. 24 trials asked participants to identify a maximum or increasing value and 24 a minimum or decreasing value.

We recruited 42 participants from the local campus community (25 Male, 15 Female, 1 Non-binary, 1 DNR). Participant ages ranged from 18-35 (μ=22.3, σ=4.64). Participants reported relatively low familiarity with AR (38% of participants reporting prior familiarity with the technology), moderate familiarity with VR (62%), and low familiarity with 3D modeling tools (33%) and game engines (33%). Data more than 70% of the way to the opposite answer was excluded for task misunderstanding (84 of 2,016 total trials). As was the intention in piloting, we achieved an overall binary accuracy (right/wrong, agnostic of magnitude) of 71%.

The study consisted of three phases: (1) consent, (2) task trials, and (3) demographics and compensation. Each block of the formal study contained three phases: (1) training, (2) testing, and (3) subjective questionnaire. After explaining the study premise and acquiring informed consent, participants were outfitted with a headset or seated at the desktop. They were instructed in the use of the modality and then began the first study block.

In each block, participants performed eight training tasks—one for each combination of channel and dimensionality. Participants acclimated to the eight visualizations they would see, navigation in the environment, and the interactions required to make selections, asking for clarifications when needed. Participants had to correctly answer each training trial before moving to the next training trial.

Participants completed 16 test trials in each task block. These trials consisted of two stimuli per channel $\times$ data dimensionality, with combinations randomly ordered. For each stimuli, participants indicated their answer through the mechanisms described in §3.3 and confirmed before moving to the next trial. To avoid potential confounds from switching between designs and highest/lowest value tasks, trials were separated by a five second screen displaying the encoding channel, dimensionality, and task for the next trial.

After completing all testing trials in a block, participants responded to 10 Likert-scale questions on a 2D monitor. In these questions, participants rated their agreement (1-7) for the following statements: "I found using *Channel* made it easy to identify differences in data values." and "I felt confident in my selections when using *Channel*." This questionnaire provided subjective feedback as well as a break between task blocks to mitigate fatigue effects.

After the study, we administered a questionnaire with Likert questions about overall experiences, open-ended feedback, and a demographics survey (materials: `https://osf.io/sj8v2/`). We then compensated participants with a $10 Amazon gift card.

3.5 Measures

We measured performance through both objective (error and time-to-completion) and subjective (ease of use and confidence) measures. We computed the percent error, normalized as a percentage of the possible value range for each task. We use relative error rather than absolute correctness to account for not only the ability to identify the correct answer for a given task but the precision with which people can estimate that answer. For extrema tasks, error reflected the difference between selected value and the greatest/smallest value. For mean tasks, error reflects the difference between the mean of the selected region and the greatest/smallest mean. For trend, we computed error as the percentage difference in slope between the best fit line and the selected direction. We measured time to completion as the time elapsed between when the visualization loaded and when the participant confirmed a selected datapoint, region or direction.

To understand the influence of different display modalities and visualization designs on data navigation, we additionally measured two elements of participant interaction: positional and rotational displacement. We collected subjective preference of different visual channels through Likert scale questions after completing all of trials for each task. We then asked overall subjective questions about the interface and subjective perceptions of the visualizations at the end of the study. A final questionnaire also included a demographics survey and space for open-ended feedback.

4 Results

We evaluated objective results using two Multivariate Analysis of Variance (MANOVA) tests with error rate and time-to-completion as dependent variables: a 3 (display type) $\times$ 3 (task) $\times$ 5 (channel) full factorial MANOVA for the three variable visualizations and a 3 (display type) $\times$ 3 (task) $\times$ 3 (channel) MANOVA for the four variable visualizations. Considering task as a factor in our analysis allows us to account for variance due to task while also measuring overall performance differences. We divide our statistics into two tests as we could not test all encoding channels in each dimension. However, we provide descriptive statistics for relevant differences (means and 95% confidence intervals) to allow the reader to compare across three and four data dimensions. We used Tukey's Honest Significant Difference Test (HSD) for general post-hoc comparisons and Least Square Means Contrast tests ($\alpha = .05$) for pairwise comparisons of specific channel-display combinations. We used MANOVAs to

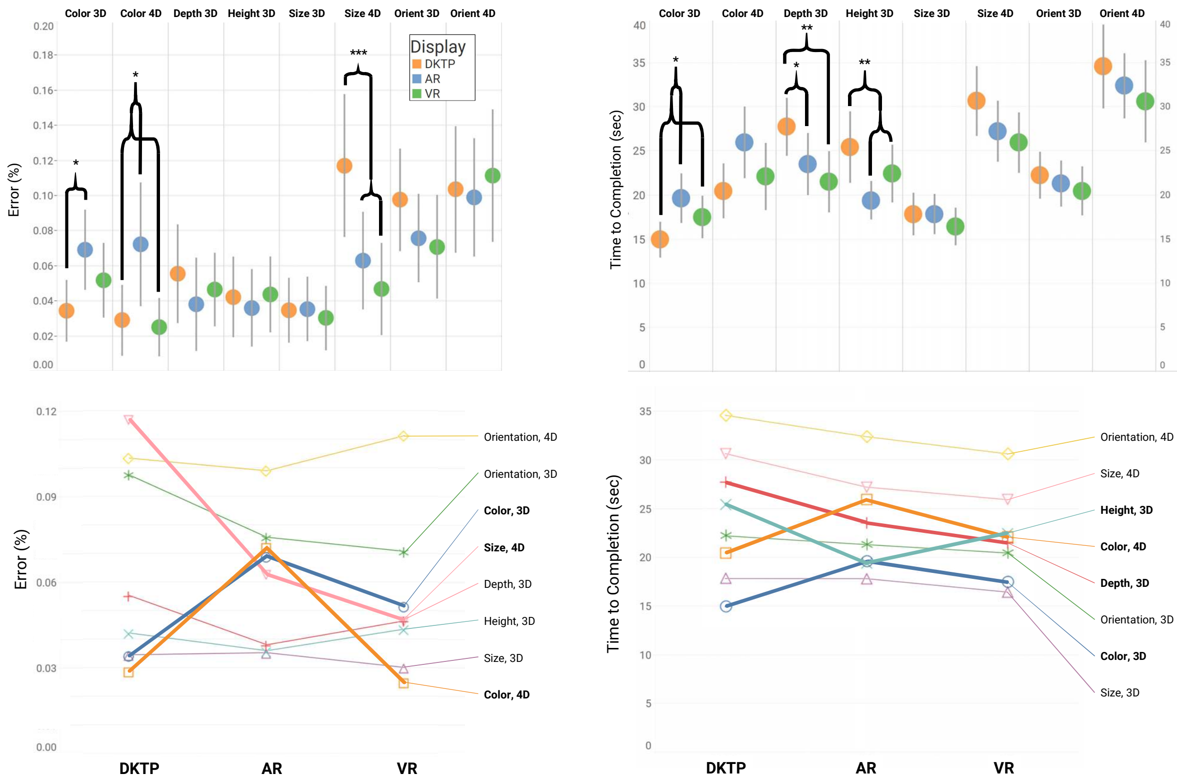

Figure 4: We compared objective performance through error and time to completion measures across displays for channel-dimension combinations. * indicates significance with $p < .05$ and *** indicates $p < .001$ (top). Error bars denote 95% confidence intervals. The line graphs (bottom) show how rankings of channel-dimension pairs differ for Desktop, AR and VR displays. Bold lines between displays indicate significant differences across displays ($p < .05$).

explore navigation strategies across display types and scale construction to analyze subjective feedback from Likert-scale questions.

To focus our discussion on relevant results, we report significant first order effects and interaction effects relevant to our hypotheses. The full statistical analysis of all interaction effects and anonymized data are available as supplemental material at `https://https://osf.io/sj8v2/` and are summarized in Fig. 4.

4.1 Objective Results

Task: We found a significant effect of task on error in visualizations with three data dimensions ($F(2,1119) = 32.12$, $p < .0001$): participants were significantly less accurate at identifying trends ($6.73\% \pm 1.17$) than extrema ($3.36\% \pm .71$) or quadrants ($2.15\% \pm .53$). Task had a different effect on time-to-completion (TTC) for visualizations with three data dimensions ($F(2,1119) = 3.39$, $p = .0041$). Unlike with error, mean tasks were significantly faster ($18.07s \pm 1.07$) than extrema tasks ($\mu = 24.44s \pm 1.34$). We additionally found effects of task for four data dimensions ($F(2,659) = 13.42$, $p < .0001$). Participants again took longer to complete extrema tasks ($\mu = 32.06s \pm 2.62$) than mean tasks ($\mu = 27.05s \pm 2.13$), but also took longer finding extrema than estimating trends ($\mu = 24.09s \pm 2.02$).

Channel: Channel impacted overall error in visualizations with both three and four data dimensions. For three dimensions, error rate was significantly higher for orientation ($6.65\% \pm 1.40$) than with any other channel ($F(4,1119) = 8.67$, $p < .0001$). In visualizations with four data dimensions, participants were significantly more accurate with color ($3.33\% \pm 1.29$) than with size ($6.00\% \pm 1.5$) and orientation ($8.8\% \pm .020$; $F(2,659) = 11.90$, $p < .0001$).

We found similar effects for completion time ($F_{3D}(4,1160) = 14.54$, $p < .0001$; $F_{4D}(2,700) = 18.16$, $p < .0001$). For both three and four data dimensions, color ($\mu_{3D} = 17.35s \pm 1.42$, $\mu_{4D} = 22.81s \pm 2.14$) was significantly faster than orientation ($\mu_{3D} = 21.34s \pm 1.55$; $\mu_{4D} = 32.65s \pm 2.52$). With three dimensions, color ($\mu = 17.36 \pm 1.42$) and size ($\mu = 17.39s \pm 1.31$) were also significantly faster than depth ($\mu = 24.59 \pm 2.08$) and height ($\mu = 22.40 \pm 1.9$), whereas with four dimensions, color was significantly faster than size ($\mu = 27.92 \pm 2.12$).

Channel $\times$ Display: We found no significant effects of display on error or completion time but found interaction effects between display and channel for both measures. In visualizations using color ($F_{3D}(1,40) = 3.98$, $p = .046$; $F_{4D}(1,40) = 3.86$, $p = .050$), AR displays ($\mu_{3D} = 5.53\% \pm 1.86$; $\mu_{4D} = 5.53\% \pm 3.09$) resulted in significantly higher error than desktop ($\mu_{3D} = 2.89\% \pm 1.52$; $\mu_{4D} = 2.51\% \pm 1.95$) and than VR for visualizations with four data dimensions ($2.05\% \pm 1.47$), likely from the color variation introduced by real-world objects in participants' field of view. We found a significant effect of display on error for size in visualizations with four data dimensions ($F(1,40) = 11.16$, $p = .0009$): Desktop displays ($9.67\% \pm 3.41$) induced more error than VR ($3.27\% \pm 1.52$) or AR ($\mu = 5.13\% \pm 2.26$) displays. We anticipate that it was more difficult to disentangle size changes in data from size changes due to perspective projection without stereo viewing. Viewing in an immersive space may reduce interference between size and depth, resulting in performance closer to that of size alone (Fig. 4, top left). We found no significant effects of display on error for size with three data dimensions or for orientation, height or depth.

Display had a significant effect on TTC for depth encodings ($F(2,39) = 8.90$, $p = .003$), where desktop displays ($27.73s \pm 3.35$) resulted in significantly greater time to completion compared to VR ($22.53s \pm 3.87$) and AR ($23.51s \pm 2.60$). Display also had a significant effect on TTC for height encodings ($F(1,40) = 6.77$, $p = .009$), where desktop displays ($25.43s \pm 4.13$) required significantly more time than AR ($19.40s \pm 2.20$) or VR ($\mu = 22.43 \pm 3.30$). These findings echo observations from error: depth, height, and size likely benefit from the added stereo cues of immersive HMDs. As participants often noted that they reoriented the visualization to estimate height and depth on the desktop, these results may also stem from participants being less efficient at desktop navigation.

Display had a significant effect on TTC for color encodings with three data dimension visualizations ($F(1,40) = 3.85$, $p = .05$), where AR ($19.62s \pm 2.84$) took significantly longer than VR ($\mu = 17.50 \pm 2.42$) and desktop ($\mu = 14.95 \pm 2.06$) displays. With four data dimensions, color differed marginally across displays ($F(1,40) = 3.39$, $p = .066$), being marginally slower in AR ($25.94s \pm 4.13$) than desktop ($20.46s \pm 3.15$) and VR ($22.12s \pm 3.85$) displays, again pointing to potential effects of background contrast. We also observed a marginally significant effect of display on time to completion for size encodings in four data dimension visualizations ($F(1,40) = 2.95$, $p = .086$): desktop displays ($30.66s \pm 4.05$) required marginally more time than AR ($27.22s \pm 3.54$) and VR ($25.93s \pm 3.51$) displays. As with error, this difference suggests that stereo viewing may help disentangle size from depth. We did not see significant effects of display on TTC for orientation or size visualizations with three data dimensions.

4.2 Navigation

We hypothesized that data navigation would also vary across displays and visualization designs. To quantitatively evaluate these differences, we conducted a MANOVA to test the effect of display on position and rotation. We again used Tukey's HSD for post-hoc comparisons. More difficult stimuli generally required more time and more interaction to answer. To decouple effects of prolonged interaction due to visualization design, we normalized total distance traveled by dividing by TTC. We also include tests run on the raw positional and rotational distance traveled in our web supplement, though the results are similar to those from the normalized measures.

Both positional ($F(2,39) = 9.35$, $p < .0001$) and rotational ($F(2,39) = 321.00$, $p < .0001$) distance traveled varied significantly across display types (Fig. 5). For position, AR displays ($.131m/s \pm .0055$) resulted in significantly more positional motion than VR ($.119m/s \pm .0051$) or desktop ($.114m/s \pm .0067$) displays. For rotational distance traveled, AR displays ($23.87deg/s \pm .67$) resulted in significantly more rotation than VR displays ($21.21deg/s \pm .68$). Desktop had the least overall rotation ($11.82deg/s \pm .74$). We hypothesize that the significantly higher motion in AR stems from participants feeling more comfortable moving around in an environment where they have nearly full situational awareness. The low motion in desktop displays may also suggest difficulty in using mouse-and-keyboard interactions to navigate a 3D space.

We compared overall motion interactions between AR, VR, and desktop as a function of block order. AR and VR displays saw no significant differences in positional or rotational navigation between blocks, whereas block order did significantly influence positional navigation on desktop displays ($F(1,40) = 12.29$, $p = .0005$). Participants panned the camera significantly less during the first task block ($.103m/s \pm .010$) than in second ($.125m/s \pm .011$) or third ($.127m/s \pm .0124$) task blocks. This finding indicates that at least some of the effects of display type may be due to difficulties in interaction; however, the effect size and differences between AR and VR indicate that AR encourages more data navigation.

4.3 Subjective Feedback

We analyzed subjective feedback by first performing scale construction on the 10 Likert scale questions completed per task block. A factor rotation over the six questions for each channel (two per channel per task block) yielded significantly correlated scales

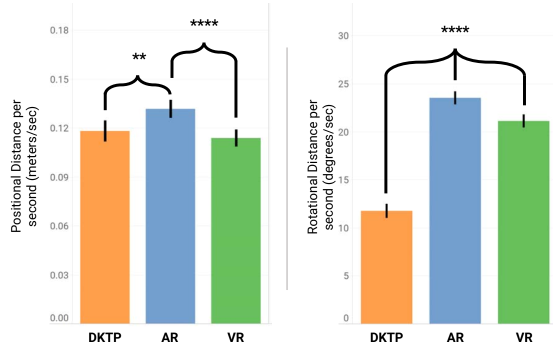

Figure 5: Analyzing position and rotation across displays, we found that participants navigated the data more in AR than in VR or on the desktop. "**" indicates $p < .01$ and "****" indicates $p < .0001$. Error bars denote 95% confidence intervals.

for each channel, with Crobach's $\alpha \geq 0.8$: color ($\alpha = .80$), size ($\alpha = .85$), height ($\alpha = .85$), depth ($\alpha = .81$), and orientation ($\alpha = .83$). We constructed a scale for each channel combining responses across questions. Using a MANOVA to evaluate the effect of display type on the constructed scales with Tukey's HSD again used for post-hoc comparisons, we found that participants preferred to use color ($F(2,39) = 6.632$, $p = .004$) on desktop displays ($5.48 \pm .57$) over AR ($4.12 \pm .63$). We found the opposite effect for height ($F(2,39) = 3.59$, $p = .037$): people preferred height in AR ($6.19 \pm .47$) to height on desktop displays ($5.21 \pm .69$). These results likely indicate unique affordances of AR combining data with the real world: while the color of real world objects complicates color estimation, people can use size contrast to better compare heights.

We conducted an additional factor rotation with the five Likert scale questions from the final questionnaire. These questions measured perceived location of the data relative to the viewer and to other data points and the perceived intuitiveness, utility, and task relation of the display navigation. The rotation yielded significantly correlated scales across questions of perceived location ($\alpha = .81$) and across navigation utility and task relation ($\alpha = .80$). We analyzed display performance for each scale independently using ANOVAs.

Participants reported significantly lower perceived spatial comprehension of the data on desktop displays ($4.21 \pm .74$) than in AR ($5.57 \pm .66$) or VR displays ($5.42 \pm .84$; $F(2,39) =$, $p = .015$). We found only a marginal effect on navigation ($F(1,39) = 3.58$, $p = .066$): participants found navigation marginally more useful in AR ($6.00 \pm .42$) than on desktop ($5.32 \pm .63$) displays. These results collectively suggest that people feel better able to spatialize and navigate data in immersive environments.

5 Discussion

We conducted a mixed factors study measuring graphical perception across desktop, AR, and VR displays. We summarize significant objective and subjective metrics that indicate key design decisions for IA. These design decisions begin to enumerate the space of effective IA design and illustrate the need for deeper empirical understanding into the trade-offs of visualizations across displays.

Size in 4D: Estimating size with four data dimensions is significantly more challenging on desktop displays than in VR or AR, partially supporting *H1*. Visualizations using depth to encode values typically use perspective projection to communicate depth. This projection means that size and depth are intertwined, as was reflected in the poor performance of size on desktops. However, as hypothesized in prior work [32, 33], depth became more effective and size easier to distinguish from depth in AR and VR. While traditional visualization guidelines strongly advocate against 3D visualizations using depth, stereoscopic viewing may alleviate many concerns with these designs, echoing Kraus et al.'s findings for scatterplot cluster detection [27]. Future work should more deeply explore how AR and VR might shift recommendations about 3D visualization design.

Color in AR: Color was significantly harder to use in AR than in other modalities (*H2*). Participants noted that "color in AR was particularly difficult due to interference from the background color of the room." *(P30)*. The red hues in our color scale did not conflict with visible objects in the experimental space. However, future work should consider how to intelligently select visualization colors for IA. One strategy is to draw on techniques for colorblind accessibility [55] to ensure more robust color design. This result also suggests that VR applications should consider their choice of backgrounds. For example, skyboxes and terrains may cause contrast effects when navigating data that could degrade visualization effectiveness.

AR vs. VR: Aside from color visualizations, objective performance was comparable between AR and VR, supporting *H2*. We observed some participants use their bodies or objects around them as referents to help with size, height, and depth estimation in AR. For example, participants pointed "with [their] fingers at an individual point to keep track of it while [they] looked elsewhere" *(P15)*. Some attempted similar strategies in VR despite the lack of embodiment, holding a hand out to "act out touching the data points [they] wanted to remember and compare for later" *(P1)*. These strategies suggest that people may leverage embodiment to interpret IA visualizations. Systems could use these observations to enhance data interaction.

Differences in Engagement & Interaction: Participants were significantly slower at analyzing depth and height on the desktop than in immersive HMDs. This effect correlated with the amount of interaction (*H3*). Height required moving across viewpoints to mitigate occlusion; however, many participants navigated to the top or sides of a visualization to maximize depth comparisons despite this motion introducing occlusion. Desktop participants consciously reflected on their interaction strategies in open-ended responses, citing the importance of "spinning all the way around the 3-D graphs" *(P14)* and "turning to the different sides of the 3D tasks" *(P29)*. These strategies mitigated potential sources of error by shifting depth estimates to position estimates, but required additional time and energy compared to AR and VR. In assessing the design of the mouse and keyboard interface against the AR/VR interface, we note the lack of other conditions where the desktop condition was significantly slower. This provides some indication that differences in interaction did not strongly affect time to completion results.

Subjective measures also reflect these interaction differences: participants were more confident using depth and height in AR and VR, reflecting engagement and performance dependencies found in Bach et al. [3]. Designers may wish to prioritize depth and height differently across modalities: on the desktop, people must navigate the visualization to isolate these channels but can more readily make these comparisons in AR and VR with less navigation.

Differences in interaction strategies extended beyond depth and height: participants engaged with data differently in all three modalities, partially supporting *H4*. For example, participants became more efficient with the mouse and keyboard over time, moving more on the first task than subsequent tasks. However, we did not see these differences in AR or VR, indicating that embodied navigation was more intuitive. Our subjective metrics support this hypothesis: people found navigating in AR qualitatively easier than the desktop.

Analyzing interaction data revealed additional differences in data engagement in AR and VR. People moved significantly more in AR than VR. We anticipate these differences reflect increased situational awareness and embodiment offered by AR: people can see the space they are moving in. This comfort may suggest different uses of physical space for configuring visualization systems in AR. Put simply, people may make better use of space outside of their field of view in AR to create constellations of visualizations that they can navigate more fluidly and comfortably than in VR, essentially turning the world into an analytical canvas. Future work in exploratory design and empirical studies could better inform these possibilities.

5.1 Limitations & Future Work

Our study represents a preliminary exploration of graphical perception for IA. With continued exploration of graphical perception in IA, the research community can build models for generating suitable visualizations depending on display, channel and dimensionality, as has been done for 2D visualizations [38]. Toward that end, future work could extend our approach to more display types or mixed display configurations. Popular see-through ARHMDs offer less color fidelity (e.g., inability to project black and issues of opacity in strong physical lighting) and smaller fields of view, but users can directly see the immediate environment, which our results suggest may influence data interpretation and engagement. Studies should also extend to other tasks, such as cluster identification to directly connect to results from prior studies [22,36,37], to other data types (e.g., streaming or time series data), to an increased number of data points, and to other encoding channels (e.g., motion) and visualization types (e.g., pie charts or line graphs).

Orientation, though noted as a useful cue in prior studies [8], led to low performance across all tasks. While we anticipate that this reflects the use of the encoding channel in the context of a data visualization rather than as isolated value pairs, better understanding the limitations of orientation as a channel is key future work. Many immersive visualizations utilize rotation or orientation [14,29,57]. Future studies should evaluate differences in orientation across tasks targeted by these systems and about multiple rotation axes.

We only tested one color encoding, the default DXR red ramp [44], thus not considering the breadth of colors that could be used in visualizations. Testing different color ramps and strategies to help users identify subtle color differences (e.g., edge enhancement or recoloring) could increase the viability of AR color encodings [55].

We focused only on channels with minimal interference with position. Understanding how our results generalize across other combinations of channels (e.g., their separability) across displays could extend prior models for multivariate visualizations in 2D [46] and inform new visualization designs. To draw principled conclusions about visualization design, we mirrored stimuli across displays as closely as the technologies allowed. In doing so, we omitted several features that may make tasks easier in immersive environments. Participants mentioned features such as "a middle vertex [for orientation] where you're looking with faint lines going out in all directions" *(P12)* or being able "to position the data plot rather than moving around" *(P0)*. In IA, gestural and raycast-based interactions [6,53] and drop shadows [10] can help infer depth and distance. Immersive displays can also provide multimodal feedback using spatial audio [22], voice commands [4], or vibration [36] for data interaction. Future work should take a broader view on how interaction may support efficient and accurate IA.

6 Conclusion

While interest in IA is growing, we still have limited guidance for creating effective IA visualizations. We conducted a graphical perception study to measure data estimation performance across visualization designs and display modalities. These findings offer new insight into the utility of depth, color, and other cues for immersive displays. We also provide preliminary insight into differences in navigation strategies across displays that could inform new interaction designs. We see this work as preliminary steps towards empirically grounded guidelines for effective immersive analytics design.

Acknowledgments

The authors thank David Burlinson and Keke Wu for figure and video support. This work was supported by NSF Award #1764092.

REFERENCES

[1] L. Al Bondakji, A.-M. Chatzi, M. Heidari Tabar, L.-M. Wesseler, and L. C. Werner. Vr-visualization of high-dimensional urban data. pp. 773–780. eCAADe; Lodz University of Technology, 2018. Available Open Access publishedVersion at https://depositonce.tu-berlin.de//handle/11303/8544.

[2] D. Albers, M. Correll, and M. Gleicher. Task-driven evaluation of aggregation in time series visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, p. 551–560. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2556288.2557200

[3] B. Bach, R. Sicat, J. Beyer, M. Cordeil, and H. Pfister. The hologram in my hand: How effective is interactive exploration of 3d visualizations in immersive tangible augmented reality? *IEEE Transactions on Visualization and Computer Graphics*, 24(1):457–467, Jan 2018. doi: 10.1109/TVCG.2017.2745941

[4] S. K. Badam, A. Srinivasan, N. Elmqvist, and J. Stasko. Affordances of input modalities for visual data exploration in immersive environments. In *Workshop on Immersive Analytics, IEEE Vis*, 2017.

[5] A. Batch, A. Cunningham, M. Cordeil, N. Elmqvist, T. Dwyer, B. H. Thomas, and K. Marriott. There is no spoon: Evaluating performance, space use, and presence with expert domain users in immersive analytics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):536–546, Jan 2020. doi: 10.1109/TVCG.2019.2934803

[6] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. *SI3D*, 97:35–38, 1997.

[7] I. Chatterjee, R. Xiao, and C. Harrison. Gaze+gesture: Expressive, precise and targeted free-space interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ICMI '15, p. 131–138. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2818346.2820752

[8] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554, 1984.

[9] M. Cordeil, A. Cunningham, T. Dwyer, B. H. Thomas, and K. Marriott. Imaxes: Immersive axes as embodied affordances for interactive multivariate data visualisation. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, pp. 71–83. ACM, New York, NY, USA, 2017. doi: 10.1145/3126594.3126613

[10] C. Diaz, M. Walker, D. A. Szafir, and D. Szafir. Designing for depth perceptions in augmented reality. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 111–122, Oct 2017. doi: 10.1109/ISMAR.2017.28

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

[12] N. A. ElSayed, B. H. Thomas, K. Marriott, J. Piantadosi, and R. T. Smith. Situated analytics: Demonstrating immersive analytical tools with augmented reality. *Journal of Visual Languages & Computing*, 36:13 – 23, 2016. doi: 10.1016/j.jvlc.2016.07.006

[13] D. Fafard, I. Stavness, M. Dechant, R. Mandryk, Q. Zhou, and S. Fels. Ftvr in vr: Evaluation of 3d perception with a simulated volumetric fish-tank virtual reality display. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 533:1–533:12. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300763

[14] J. A. W. Filho, W. Stuerzlinger, and L. Nedel. Evaluating an immersive space-time cube geovisualization for intuitive trajectory data exploration. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):514–524, Jan 2020. doi: 10.1109/TVCG.2019.2934415

[15] J. A. W. Filho, W. Stuerzlinger, and L. Nedel. Evaluating an immersive space-time cube geovisualization for intuitive trajectory data exploration. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):514–524, Jan 2020. doi: 10.1109/TVCG.2019.2934415

[16] J. L. Gabbard, J. E. Swan, and A. Zarger. Color blending in outdoor optical see-through ar: The effect of real-world backgrounds on user interface color. In *2013 IEEE Virtual Reality (VR)*, pp. 157–158, March 2013. doi: 10.1109/VR.2013.6549410

[17] J. A. Garcia Macias, J. Alvarez-Lozano, P. Estrada, and E. Aviles Lopez. Browsing the internet of things with sentient visors. *Computer*, 44(5):46–52, May 2011. doi: 10.1109/MC.2011.128

[18] M. Gleicher, M. Correll, C. Nothelfer, and S. Franconeri. Perception of average value in multiclass scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2316–2325, Dec 2013. doi: 10.1109/TVCG.2013.183

[19] A. Gogolou, T. Tsandilas, T. Palpanas, and A. Bezerianos. Comparing similarity perception in time series visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):523–533, Jan 2019. doi: 10.1109/TVCG.2018.2865077

[20] L. Harrison, F. Yang, S. Franconeri, and R. Chang. Ranking visualizations of correlation using weber's law. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1943–1952, Dec 2014. doi: 10.1109/TVCG.2014.2346979

[21] J. Heer, N. Kong, and M. Agrawala. Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, p. 1303–1312. Association for Computing Machinery, New York, NY, USA, 2009. doi: 10.1145/1518701.1518897

[22] B. Hoppenstedt, T. Probst, M. Reichert, W. Schlee, K. Kammerer, M. Spiliopoulou, J. Schobel, M. Winter, A. Felnhofer, O. D. Kothgassner, and R. Pryss. Applicability of immersive analytics in mixed reality: Usability study. *IEEE Access*, 7:71921–71932, 2019. doi: 10.1109/ACCESS.2019.2919162

[23] T. Horak, S. K. Badam, N. Elmqvist, and R. Dachselt. When david meets goliath: Combining smartwatches with a large vertical display for visual data exploration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173593

[24] J. Irizarry, M. Gheisari, G. Williams, and B. N. Walker. Infospot: A mobile augmented reality method for accessing building information through a situation awareness approach. *Automation in Construction*, 33:11 – 23, 2013. Augmented Reality in Architecture, Engineering, and Construction. doi: 10.1016/j.autcon.2012.09.002

[25] J. A. Jones, J. E. Swan, II, G. Singh, E. Kolstad, and S. R. Ellis. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, APGV '08, pp. 9–14. ACM, New York, NY, USA, 2008. doi: 10.1145/1394281.1394283

[26] R. Kosara and D. Skau. Judgment error in pie chart variations. In *Proceedings of the Eurographics / IEEE VGTC Conference on Visualization: Short Papers*, EuroVis '16, p. 91–95. Eurographics Association, Goslar, DEU, 2016.

[27] M. Kraus, N. Weiler, D. Oelke, J. Kehrer, D. A. Keim, and J. Fuchs. The impact of immersion on cluster identification tasks. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):525–535, Jan 2020. doi: 10.1109/TVCG.2019.2934395

[28] A. Krekhov, S. Cmentowski, A. Waschk, and J. Krüger. Deadeye visualization revisited: Investigation of preattentiveness and applicability in virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):547–557, Jan 2020. doi: 10.1109/TVCG.2019.2934370

[29] B. Laha, D. A. Bowman, and J. J. Socha. Effects of vr system fidelity on analyzing isosurface visualization of volume datasets. *IEEE Transactions on Visualization and Computer Graphics*, 20(04):513–522, apr 2014. doi: 10.1109/TVCG.2014.20

[30] J.-R. Lin, J. Cao, J.-P. Zhang, C. van Treeck, and J. Frisch. Visualization of indoor thermal environment on mobile devices based on augmented reality and computational fluid dynamics. *Automation in Construction*, 103:26 – 40, 2019. doi: 10.1016/j.autcon.2019.02.007

[31] M. Luboschik, P. Berger, and O. Staadt. On spatial perception issues in augmented reality based immersive analytics. In *Proceedings of the 2016 ACM Companion on Interactive Surfaces and Spaces*, ISS Companion '16, pp. 47–53. ACM, New York, NY, USA, 2016. doi: 10.

1145/3009939.3009947
[32] K. Marriott, J. Chen, M. Hlawatsch, T. Itoh, M. A. Nacenta, G. Reina, and W. Stuerzlinger. *Immersive Analytics: Time to Reconsider the Value of 3D for Information Visualisation*, pp. 25–55. Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-030-01388-2_2
[33] J. P. McIntire and K. K. Liggett. The (possible) utility of stereoscopic 3d displays for information visualization: The good, the bad, and the ugly. In *2014 IEEE VIS International Workshop on 3DVis (3DVis)*, pp. 1–9, Nov 2014. doi: 10.1109/3DVis.2014.7160093
[34] D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer. Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):438–448, Jan 2019. doi: 10.1109/TVCG.2018.2865240
[35] A. V. Pandey, J. Krause, C. Felix, J. Boy, and E. Bertini. Towards understanding human similarity perception in the analysis of large sets of scatter plots. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 3659–3669. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2858036.2858155
[36] A. Prouzeau, M. Cordeil, C. Robin, B. Ens, B. H. Thomas, and T. Dwyer. Scaptics and highlight-planes: Immersive interaction techniques for finding occluded features in 3d scatterplots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pp. 325:1–325:12. ACM, New York, NY, USA, 2019. doi: 10.1145/3290605.3300555
[37] 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.
[38] R. A. Rensink and G. Baldridge. The perception of correlation in scatterplots. In *Computer Graphics Forum*, vol. 29, pp. 1203–1210. Wiley Online Library, 2010.
[39] B. Resch, A. Wichmann, and N. Göll. Usability in 4d ar: Visualising multi-temporal real-time geo-data in augmented reality environments. *International Journal of Interactive Mobile Technologies (iJIM)*, 9(4):23–33, 2015.
[40] J. S. Roo and M. Hachet. One reality: Augmenting how the physical world is experienced by combining multiple mixed reality modalities. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, pp. 787–795. ACM, New York, NY, USA, 2017. doi: 10.1145/3126594.3126638
[41] B. Saket, A. Endert, and . Demiralp. Task-based effectiveness of basic visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(7):2505–2512, July 2019. doi: 10.1109/TVCG.2018.2829750
[42] I. M. Sauer, M. Queisner, P. Tang, S. Moosburner, O. Hoepfner, R. Horner, R. Lohmann, and J. Pratschke. Mixed reality in visceral surgery: development of a suitable workflow and evaluation of intraoperative use-cases. *Annals of surgery*, 266(5):706–712, 2017.
[43] K. B. Schloss, C. C. Gramazio, A. T. Silverman, M. L. Parker, and A. S. Wang. Mapping color to meaning in colormap data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):810–819, Jan 2019. doi: 10.1109/TVCG.2018.2865147
[44] R. Sicat, J. Li, J. Choi, M. Cordeil, W. Jeong, B. Bach, and H. Pfister. Dxr: A toolkit for building immersive data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):715–725, Jan 2019. doi: 10.1109/TVCG.2018.2865152
[45] D. Skau and R. Kosara. Arcs, angles, or areas: Individual data encodings in pie and donut charts. In *Computer Graphics Forum*, vol. 35, pp. 121–130. Wiley Online Library, 2016. doi: 10.1111/cgf.12888
[46] S. Smart and D. A. Szafir. Measuring the separability of shape, size, and color in scatterplots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300899
[47] S. Y. Ssin, J. A. Walsh, R. T. Smith, A. Cunningham, and B. H. Thomas. Geogate: Correlating geo-temporal datasets using an augmented reality space-time cube and tangible interactions. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 210–219, March 2019. doi: 10.1109/VR.2019.8797812
[48] S. Su, V. Perry, and V. Dasari. Comparative study for multiple coordinated views across immersive and non-immersive visualization systems. In J. Y. Chen and G. Fragomeni, eds., *Virtual, Augmented and Mixed Reality. Multimodal Interaction*, pp. 321–332. Springer International Publishing, Cham, 2019.
[49] D. A. Szafir. Modeling color difference for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):392–401, Jan 2018. doi: 10.1109/TVCG.2017.2744359
[50] B. Thoravi Kumaravel, F. Anderson, G. Fitzmaurice, B. Hartmann, and T. Grossman. Loki: Facilitating remote instruction of physical tasks using bi-directional mixed-reality telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, pp. 161–174. ACM, New York, NY, USA, 2019. doi: 10.1145/3332165.3347872
[51] N. Weibel. New frontiers for pervasive telemedicine: From data science in the wild to holopresence. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '17, pp. 276–281. ACM, New York, NY, USA, 2017. doi: 10.1145/3154862.3154912
[52] M. Whitlock, K. Wu, and D. A. Szafir. Designing for mobile and immersive visual analytics in the field. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):503–513, Jan 2020. doi: 10.1109/TVCG.2019.2934282
[53] C. A. Wingrave and D. A. Bowman. Baseline factors for raycasting selection. In *Proceedings of Virtual Reality International*, 2005.
[54] J. Zacks, E. Levy, B. Tversky, and D. J. Schiano. Reading bar graphs: Effects of extraneous depth cues and graphical context. *Journal of experimental psychology: Applied*, 4(2):119–138, 1998. doi: 10.1037/1076-898X.4.2.119
[55] Y. Zhao, E. Cutrell, C. Holz, M. R. Morris, E. Ofek, and A. D. Wilson. Seeingvr: A set of tools to make virtual reality more accessible to people with low vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300341
[56] S. Zollmann, D. Kalkofen, C. Hoppe, S. Kluckner, H. Bischof, and G. Reitmayr. Interactive 4d overview and detail visualization in augmented reality. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 167–176, Nov 2012. doi: 10.1109/ISMAR.2012.6402554
[57] S. Zollmann, C. Poglitsch, and J. Ventura. Visgis: Dynamic situated visualization for geographic information systems. In *2016 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pp. 1–6, Nov 2016. doi: 10.1109/IVCNZ.2016.7804440