Beyond the Physical: Examining Scale and Annotation in Virtual Reality Visualizations

Teoman Ulusoy
University of Calgary
ttulusoy@ucalgary.ca

Kurtis Danyluk
University of Calgary
ktdanylu@ucalgary.ca

Wesley Willett
University of Calgary
wj@wjwillett.net

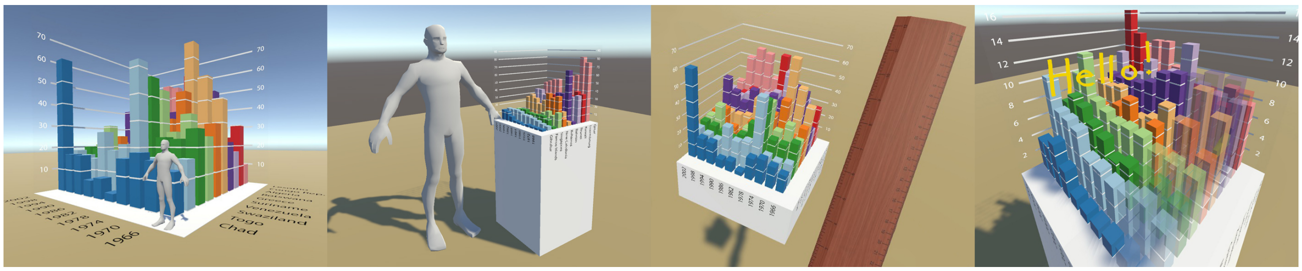

Figure 1: (From left to right) A room-scale visualization (~6.4m tall with 1.8m human for context), table-scale visualization (64cm), hand-scale visualization (12cm), and visualization with annotation and filtering tools.

ABSTRACT

We examine how viewers in virtual reality (VR) environments interact with simple data visualizations at scales ranging from hand-sized to room-sized. We also explore how the addition of virtual annotation and filtering tools affects how viewers solve basic data analysis tasks. We report on two studies, inspired by previous examinations of data physicalizations. The first study investigated how three visualization sizes, including hand-, table-, and room-scale versions, impact viewers' problem-solving behavior. A second study examined how interactive annotation and filtering tools might support new modes of use that transcend the limitations of physical representations. Our results highlight challenges associated with extreme scales, especially those that require navigation techniques other than physical locomotion, and hint at the potential of interactive annotation and filtering tools in VR visualization environments.

CCS CONCEPTS

• Human computer interaction • Interaction paradigms • Virtual Reality • User studies • Information Visualization

KEYWORDS

Virtual reality, Information visualization, Scale, Physicalization, Annotation, Immersive analytics.

Technical Report — Cite As:

Teoman Ulusoy, Kurtis Danyluk, Wesley Willett. (2018) Beyond the Physical: Examining Scale and Annotation in Virtual Reality Visualizations. URI: http://hdl.handle.net/1880/108786. Department of Computer Science, University of Calgary, Calgary, AB, Canada, T2N 1N4. October.

1 INTRODUCTION

New hardware and fabrication technologies are increasingly making it possible for data visualizations to transcend the limits of page and screen. Immersive visualization tools [6] promise to use virtual reality (VR), augmented reality (AR), and other technologies to embed representations of data in rich environments or in the context of everyday tasks. Meanwhile work on data physicalization has highlighted the potential of instantiating representations of data as real objects in physical spaces [20]. Yet the tradeoffs associated with representing data using these kinds of highly immersive virtual and physical representations remain poorly understood.

Early explorations of data physicalizations suggest that their tangible nature allows viewers to inspect, mark, and manipulate them more effectively than on-screen versions [18] and that viewers may find them more memorable than visualizations on paper [26]. VR and AR tools, meanwhile, lack the potential for tactile manipulation, but offer the potential for visualizations that transcend the limits of physical reality. Because they are not constrained by manufacturing complexity or even the limitations of real-world physics, visualizations on these platforms can easily be created in scales and configurations that would be impossible with physical objects. Moreover, they can support new kinds of physical interaction and manipulation, allowing viewers to reach *through* visualizations or dynamically change their form and behavior, while still appearing to retain many of the advantages of physicalizations.

We present an initial examination of the potential of simple immersive virtual visualizations by extending Jansen et al.'s studies of physicalizations [18] to VR environments. We first replicate Jansen et al.'s original experiment using virtual versions of their hand-sized physicalizations and compare these against versions at larger table- and room-sized scales. We then explore the addition of new tools enabled by the move to VR, including interactive annotation and filtering instruments. We find a clear preference for visualizations at hand- and table- scales as well as enthusiasm for both kinds of interactive tools. We also highlight the potential for VR systems to support new kinds of analysis via simple immersive interactions.

2 RELATED WORK

Our research builds directly on past work in virtual reality visualization, as well as recent work in data physicalization.

Virtual Reality for Infovis

Virtual reality (VR) is by no means a new field of research, with the first system created by Sutherland [27] in 1968. However, the debut of the Oculus Rift SDK in 2013, and the subsequent release of consumer headsets such as the HTC Vive and Windows Mixed Reality devices has renewed interest in the field. While the scientific visualization community has long embraced VR for showing 3D data with clear spatial embeddings, information visualization researchers are now increasingly looking for novel ways to display data using immersive VR [11,25].

Early investigations of abstract data visualizations in VR typically used either "fishtank" VR or CAVE systems which rely on head-tracking and static displays [2,10]. As early as 1993, Arthur et al. examined participants' ability to trace tree structures using a fish tank VR setup and found considerable speed and accuracy benefits [2]. Later work by Demiralp et al. further explored the impact of visualizations of different scales using both fishtank VR and the CAVE VR system [10]. Their findings highlighted the potential of VR visualization generally, while noting that fishtank VR was a better fit for most contexts, especially when visualizations were smaller than viewers' bodies.

In the last few years, however, the increasing availability of VR and AR head-mounted displays (HMDs) has resulted in a groundswell of new immersive information visualization tools. These include systems like Donalek et al.'s iViz [11] which adapt traditional abstract visualizations like scatterplots to shared 3D spaces, as well as more complex tools like Cordeil et al.'s ImAxes [8] which leverage the flexibility and openness of VR environments to create new abstract visualization types.

However, the potential benefits and trade-offs associated with various VR design choices for abstract data visualization remain poorly understood. Initial studies have highlighted the effectiveness of immersive VR environments with stereoscopic and motion-based depth cues for particular kinds of visualization tasks including graph analysis [23]. Experiments have also shown that HMDs compare favorably against much more costly CAVE systems [9]. Work on immersive unit visualization [17] has also showcased the potential for VR to support transitions between multiple scales, supporting both high-level analysis and detailed examination of individual data points within the same continuous environment. So far, however, this research gives little guidance as to which scales are the most effective for various tasks and datasets.

Physical Visualizations

Meanwhile, work on data physicalization has identified a variety of benefits for highly immersive, physical instantiations of data [20]. Interestingly, this growing body of research attributes many of the positive characteristics of these physical representations to their ease of manipulation and exploration, as well as their strong physical presence – traits which VR and AR tools are increasingly able to approximate. As such, our paper is heavily inspired by Jansen et al.'s fundamental work in which they compared the performance of physicalizations against on-screen equivalents and investigated multiple factors (including stereoscopic depth cues and tangible manipulation) that contribute to the performance differences between them [18].

Studies by Berard et al. have also begun to explore the interstitial space between physical and virtual systems, examining novel "handheld perspective-corrected displays" which can project complex interactive puzzles and other objects onto simple volumetric props [3]. Interestingly, participants were able to solve complex 3D puzzles faster and more accurately when using projected virtual objects than with physically printed ones – likely because the virtual objects did not suffer from the poor contrast, occlusions, and other shortcomings of the physical materials.

3 GOING BEYOND THE PHYSICAL

Data physicalizations allow viewers to leverage their real-world perceptual and physical abilities to inspect and interpret data, using interactions that build on familiar metaphors and expectations from the physical world. Initial work in this space highlights how physicalizations can provide a variety of benefits, including support for physical manipulation and locomotion [20] and may encourage greater memorability [26] and engagement [16]. However, physicalizations can be complicated, difficult, and impractical to construct – especially as their scale and degree of interactivity increases. Tabletop systems like EMERGE [28] and inFORM [12], for example, required long-term, concentrated engineering efforts to develop and maintain. Meanwhile the few examples at even larger room- and building-scales are mostly art installations, whose goals are aesthetic or communication-oriented, rather than analysis-focused.

VR systems, meanwhile, offer many of the advantages of physicalizations, providing increasingly vivid immersion and presence facilitated by binocular and motion-based depth cues, realistic interactions, and increasing levels of visual realism – without the prohibitive costs. As a result, VR tools offer the opportunity to create VR visualizations that would be difficult or impossible in the physical world. For example, virtual environments can accommodate visualizations at extreme scales and levels of detail without material costs or space constraints. Similarly, virtual visualizations can support interactive manipulations that would be limited by the physics of real-world objects, including dynamically changing visualizations' materials, sizes, or transparency. Virtual environments may also make it easier to design and implement tools and interactions to support common tasks like filtering, selection, and annotation.

As an initial exploration, we explore the potential for VR interfaces that build on the kinds of simple chart designs and interactions that show clear benefits in the physical world. Specifically, we use virtual reality prototypes to replicate and extend foundational studies of simple data physicalizations. This allows us to examine the impact of larger visualization scales and test new kinds of interactive tools, while still preserving many of the norms associated with simple, physical charts.

4 EXPERIMENT SETUP

We based both our visualization and experiment designs off of those developed by Jansen et al. [19]. In their studies, participants used small 3D physical bar charts as well as 2D and 3D on-screen versions (Figure 2) to complete a series of simple data analysis tasks. The studies also compared the same physicalizations against on-screen versions that used stereo depth cues and supported rotation using physical props. Based on these explorations, Jansen et al. concluded that the advantages of the physical visualizations likely related to participants' ability to simultaneously manipulate and inspect the object while using their fingers to mark and compare items of interest. This direct interaction, combined with the high visual fidelity of the physical object, helped participants compensate for problems like occlusion that routinely plague 3D visualizations on screens.

VR visualizations, unlike their 3D on-screen counterparts, have the potential to offer many of these same kinds of interactions, allowing viewers to manipulate and inspect virtual objects much as they would physical ones. Recent VR systems also offer levels of immersion and visual fidelity that are increasingly able to approximate the appearance and behavior of real-world settings. To examine the viability of VR visualizations for these same kinds of tasks, we replicated and extended Jansen's original design. This allowed us to benchmark against their results while also exploring some of the new affordances of VR by examining the impact of visualizations at three dramatically different scales.

Visualization Design

We used a 3D bar chart design which emulates the physical charts created by Jansen et al. with only a few differences. Like the originals, our charts (Figure 1) feature a 10 × 10 array of bars with a white base and black labels. We also retained the same bar widths, spacing, aspect ratio, and color palette. The back sides of our versions are entirely transparent, with floating axis lines and values. To increase legibility, we added higher-contrast tick marks on the bars themselves. We also increased the size of the category labels and aligned them more closely with their respective bars. As in Jansen et al.'s study we used this chart template to generate a variety of different charts each using 10 years worth of development statistics from Gapminder[1] organized by country. We opted to use percentages, rather than raw counts or intervals, for all axes to reduce the potential for participant confusion related to units and magnitudes.

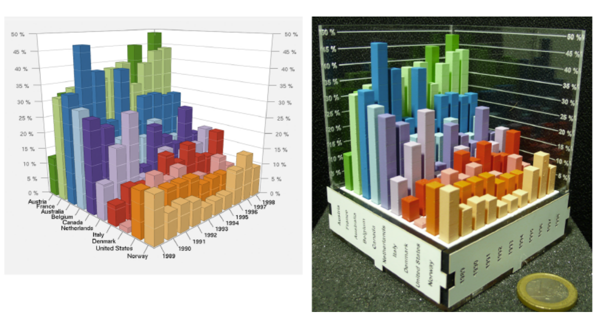

Figure 2: 3D on-screen visualization and physicalization from Jansen et al. [16]. ***(Image permissions pending.)***

Virtual Environment

We conducted our experiments using a test environment that we implemented using Unity and which supports a range of different VR headsets including the HTC Vive and Windows Mixed Reality devices. For our studies, we used an HTC Vive installed in a $2.5m \times 2.5m$ tracked area in an open-plan research space. Related studies have explored the use of alternative control schemes for VR interaction, including gestural hand-tracking [30]. However, we chose to use a pair of Vive controllers, based on recent studies that suggest they have a lower learning curve and more stable tracking than gesture-recognition systems like the Leap Motion controller [14]. Participants held a pair of controllers at all times during the studies. In the virtual environment, these controllers appeared either as a small visualization or as $30cm$ virtual ruler, depending on the study condition. We added the virtual ruler based on feedback from pilot studies in which participants repeatedly attempted to use the virtual controller as a level or measuring implement.

The virtual environment for the studies was larger (up to 100m × 100m) than the tracked area and included only minimal ornamentation. We used a simple floor plane in a neutral color and a default skybox. Because the virtual space was much bigger than the physical lab environment, we allowed participants to augment their physical locomotion with virtual flight. Participants could control flight by pointing either controller in the desired direction, then clicking and holding forward on the controller's touchpad. Participants could also fly backwards by clicking and holding back on the touchpad. This allowed participants to fly in any direction without having to change the orientation of their body. The movement did not use any form of acceleration and moved the user at a constant rate of $1.8m/s$.

Tasks

We used the same three types of basic chart-reading tasks introduced by Jansen et al. in their original study:

Range Task. Indicate the range of values for a country.
Order Task. Sort the values for a year in ascending order.
Compare Task. Locate three specific country-year pairs and determine which one has the lowest value.

[1] https://www.gapminder.org/

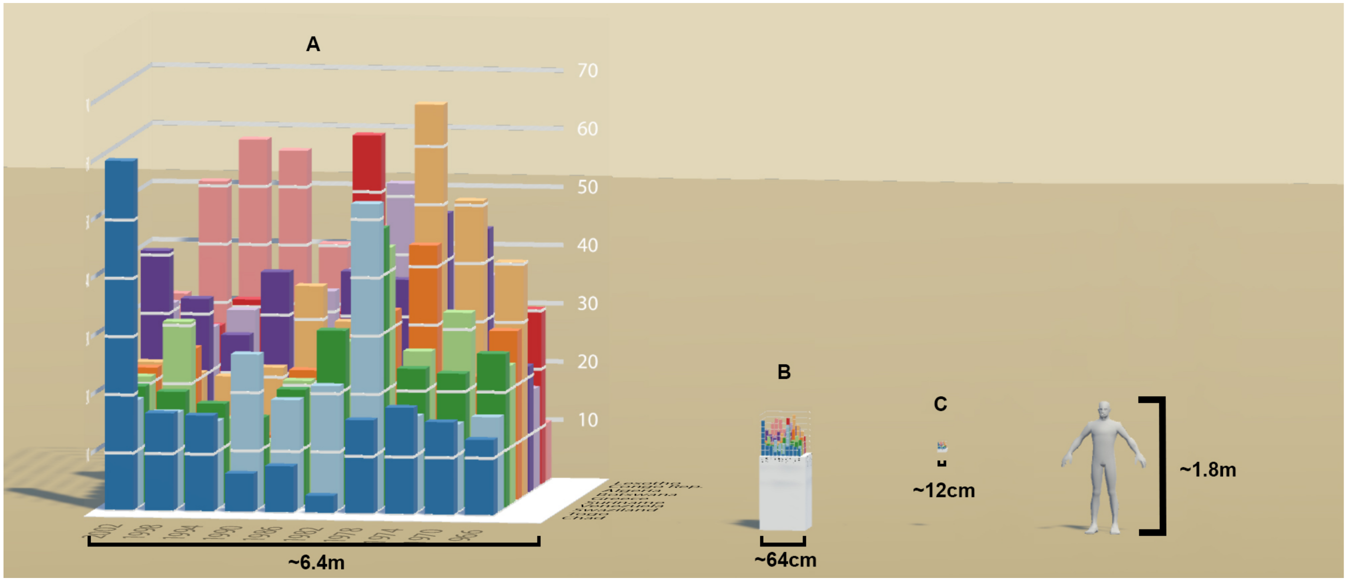

Figure 3: (A) Room-scale, (B) table-scale, and (C) hand-scale visualizations with 1.8m tall human figure for reference.

Jansen et al. used a tablet on which participants could see the study prompts and record their responses. Because we were concerned about participants' ability to enter responses on a virtual version of this same interface, we instead displayed task prompts on a *question board* (Figure 3) attached to one of the controllers. Participants could summon or dismiss the question board as needed using the trigger on the controller. When using larger visualizations, the question board appeared to the right of one of the rulers, while for the smaller hand-scale visualizations it appeared behind the chart itself. Participants were free to choose either hand to hold the question board.

Upon completing each task, participants reported their answers verbally to an experimenter who was seated 1-2m away. This experimenter manually recorded participant timing data and advanced participants from one task to the next. Throughout the studies we also captured audio and video streams from the headset for follow-up analyses.

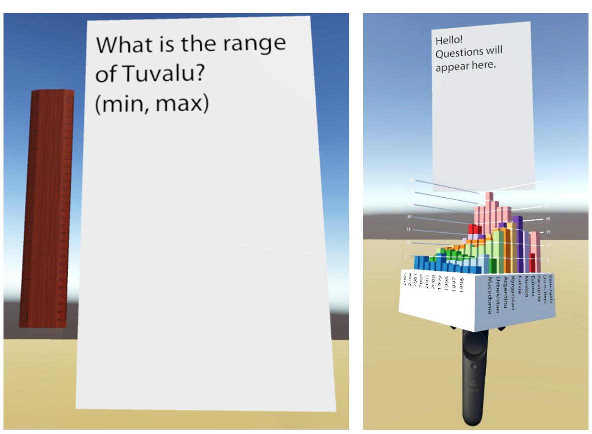

Figure 4: Question board design and location relative to the ruler tool (left) and hand-scale visualization (right).

5 STUDY ONE – SCALE

In our first experiment, we examined how the scale of a virtual visualization affected participants' ability to perform basic analysis tasks, and explored how the scale of the visualization changed their overall experience.

Visualization Scales

We explored three different visualization scales: hand scale (Figure 3c), table-scale (Figure 3b), and room-scale (Figure 3a).

We modelled the **hand-scale** visualization directly after the visualizations Jansen et al. used in their original study. This version of the visualization was roughly 12cm across. We increased the size of labels specifically to maintain their legibility on the 2160×1200 pixel Vive display, but otherwise attempted to replicate the size and level of detail of Jansen's physical prototypes. We attached the hand-scale visualization directly to the top of one of the two Vive controllers, allowing participants to reorient and inspect it by moving that hand.

For the **table-scale**, we increased the visualization's size to 64cm across, similar to the size of tabletop bar-chart displays like EMERGE [28] (Figure 5) and shape-changing displays like Relief [24] , Tangible Cityscape [29] and inFORM [12]. We placed the visualization atop a virtual plinth with a default height of 1m. At the start of each study we interactively adjusted the height of the plinth based on feedback from the participant to ensure that the visualization was easy for them to see and reach.

The **room-scale** visualizations were considerably larger, measuring 6.4m to a side. This scale was inspired by large-scale installations like Richard Burdett's population-density models of major cities [5], the walkable age pyramid created by Atelier Brükner [4] , and the eCLOUD [21] and airFIELD [13] sculptures

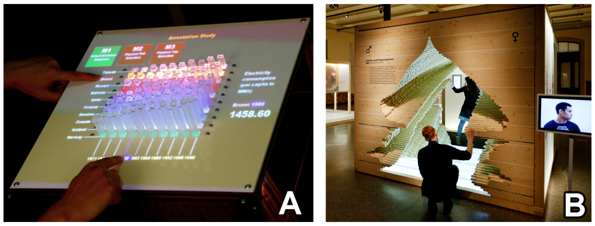

Figure 5: (A) EMERGE tabletop display [24]. (B) Walkable age pyramid by Atelier Brükner [3]. *(Image permissions pending.)*

by Dan Goods, Nik Hafermaas, and Aaron Koblin – all of which allow viewers to explore data by physically walking through, under, and around it. The overall design was similar to the other scales. However, in order to increase the legibility of the country and year labels on the axes, we reoriented them to lie horizontally on the floor next to the visualization rather than underneath it.

Procedure

The experiment consisted of 3 blocks with 3 tasks per block. Each block used a visualization at a different scale. All participants saw the visualizations in order, starting with the hand scale then advancing to the table and room scales. We gave participants a few minutes before each block to become accustomed to the current scale.

At hand scale, we allowed participants to experiment before deciding which hand they wished to use to hold the visualization. To reduce the risk of simulator sickness, we gave participants the option to sit while using the hand- and room-scale visualizations since neither required full body locomotion. We also allowed participants to take off the HMD between each block to rest. We used 3 different datasets and rotated the order for each participant. The experiment lasted 45 minutes on average, with the time spent in VR being roughly 20 minutes.

We recruited a total of 9 participants. Seven were familiar with VR and 7 participants were familiar with data visualizations. All participants were also familiar with video games. All participants had normal or corrected-to-normal vision and were able to clearly read all the chart and axis labels within the VR environment.

Measures

We recorded two measures of performance for each task: error rate and time-on-task. Both of these measures match Jansen et al. and our error calculations also mirror theirs. *Time-on-task* measured the interval between the time when the question appeared on the question board and the time at which the participant verbally stated their final answer (as recorded by the experimenter). We computed the error for *range* tasks using the absolute difference between the participant's answer and the correct minimum or maximum, divided by the total axis range. For *order* tasks we computed the normalized Kendall Tau distance between the participant's answer and the correct order. Finally, for *compare* tasks we tallied the total number of incorrect answers. We normalized each of the final error scores to give a value between 1-0. For all quantitative measures we report averages and 95% confidence intervals (CIs).

For more qualitative feedback, we administered two surveys after participants finished their trials. These included a digital questionnaire to assess participant likes and dislikes as well as a paper questionnaire containing a copy of Plutchik's Wheel of Emotion to gauge participants' emotional response. Plutchik's wheel consists of radial layout containing 24 adjectives describing a range of different emotions. We asked participants to circle any emotions that described how they felt when interacting with each scale and provide qualitative explanations for their choices.

Results

As in Jansen et al.'s original study, error rates were similarly low across all scale conditions and participants. As a result, our quantitative results focus predominantly on timing. We follow with notable observations from the survey responses.

Error Rate. The mean error rate for the *range* tasks (0.03, CI = [0.01, 0.05]) was very low. However, we saw higher rates for the *compare* tasks (0.16, CI = 0.04, 0.30) and *order* tasks (0.24, CI = [0.046, 0.436]). The rate for order was skewed higher by multiple instances in which participants forgot to report one of the ten values. However, we saw no clear relationship between this mistake and any particular scale.

Time on Task. We computed average time-on-task by participant for each task and condition (Figure 4). We observed no clear differences in performance between hand and table scale for any of the tasks. For the range and order tasks, however, the room scale condition was considerably slower than the other two. Participants took an average of 13.9s longer to perform *range* tasks with the room-scale visualization (47.2s, CI = [36.7s, 57.7s]) than with the table-scale version (33.3s, CI = [25.3s, 41.4s]). This difference was even more dramatic for *order* tasks, where participants took an average of 46.9s longer for room-scale (116.7s, CI = [86.5s, 146.8s]) than table-scale (69.8s, CI = [50.0s, 89.5s]).

Comparison to Physicalizations. For all scales and task types, participants in our study performed considerably more slowly than participants in Jansen et al.'s original experiment [18]. As Figure 4 shows, our conditions were routinely 10-30s slower on average than Jansen et al.'s *physical* condition, and closer to the

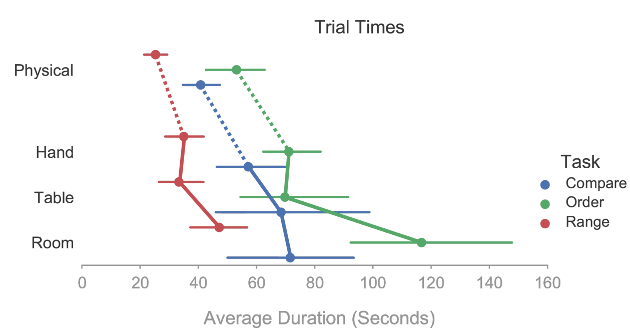

Figure 6: Study 1. Time-on-task for each scale (*hand, table, room*) × task (*compare, order, range*) combination. Results from Jansen et al.'s *physical* condition [18] are shown for context (top). Error bars show 95% CIs.

times observed for their on-screen 3D bar charts. We suspect this slowdown may be due, at least in part, to our timing and reporting procedure, in which participants gave their answers verbally rather than entering them on a tablet. These exchanges with the experimenter likely introduced some delay in each trial, and the presence of the experimenter may have encouraged participants to more carefully consider each response before reporting it aloud.

Feedback and Observations

We found that all participants chose to hold the hand scale visualization in their dominant hand. Similarly, for the table and room scales all participants held the question board in their dominant hand. Across all scales, participants seldom used rulers for measuring, and instead used them as a pointing device to help track and recall specific bars. All participants repeatedly clipped through the visualization using the rulers, but none tried to clip through the visualization using their face or body in the hand or table scale. In general, participants treated the table-scale visualization as if it were occupying real space and did not attempt to move into it. However, at room scale, participants had no problem clipping through the bars with their body for short instances while flying – though most were quick to physically move out of bars if they accidentally stopped inside of one.

Much of the feedback we received from participants reflected a desire for better tools. Four participants specifically asked for the ability to mark or select bars in order to keep track of them. One participant also suggested filtering tools to hide rows or columns that obscured their view.

Table-scale. Overall, participants voiced a strong preference for the table-scale visualizations (Figure 7), with several noting that they felt the most natural to interact with. In contrast, only one participant thought the table-scale was the hardest to use, with their critique emphasizing the additional physical movement required to navigate around the chart. Participants' emotional responses to this scale on the Plutchik wheel (Figure 8) were largely positive. However, 3 participants expressed *annoyance*, with 2 of them unhappy about the lack of additional tools, and one who disliked how taller bars obscured smaller ones.

Hand-scale. Only a few participants reported a preference for the hand scale visualizations. A single participant out of the 9 indicated that hand-scale was the easiest to use, explaining that they found it simpler to manipulate and examine than the other scales. The same participant also believed that they had performed the fastest with the hand scale, but added that they might have been more accurate (albeit slower) with the larger table-scale. Two participants identified hand-scale as the most difficult to use, noting that they found it hard to accurately see obscured bars on such a small chart and that it was difficult to level the chart to compare values. Participants' responses on the Plutchik wheel showed a mix of *interest* (based on the visual appeal of the small model) and *annoyance* (mostly related to controller shake).

Room-scale. Participants responded the least positively to room scale, with none of the participants indicating that it was the easiest to use and six calling it the most difficult. Most cited the flight mechanism as the main shortcoming, with 3 reporting that

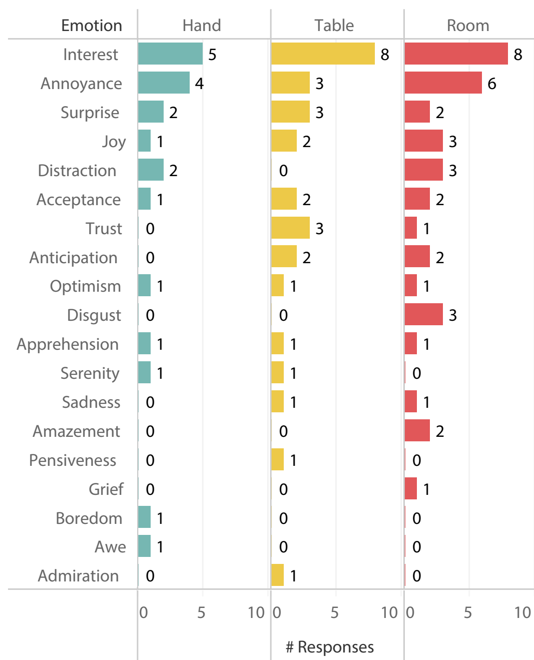

Figure 8: Study 1. Participant responses on Plutchik's wheel of emotions for each visualization scale.

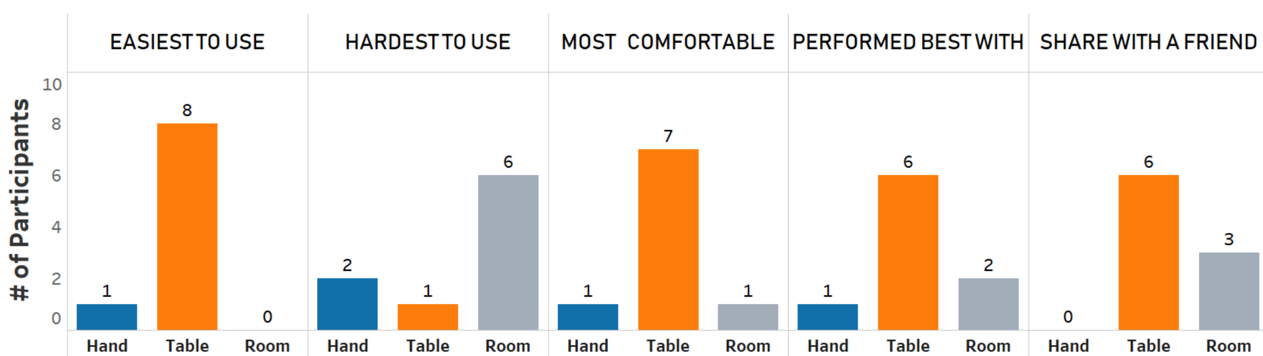

Figure 7: Study 1. Post-study survey results by scale.

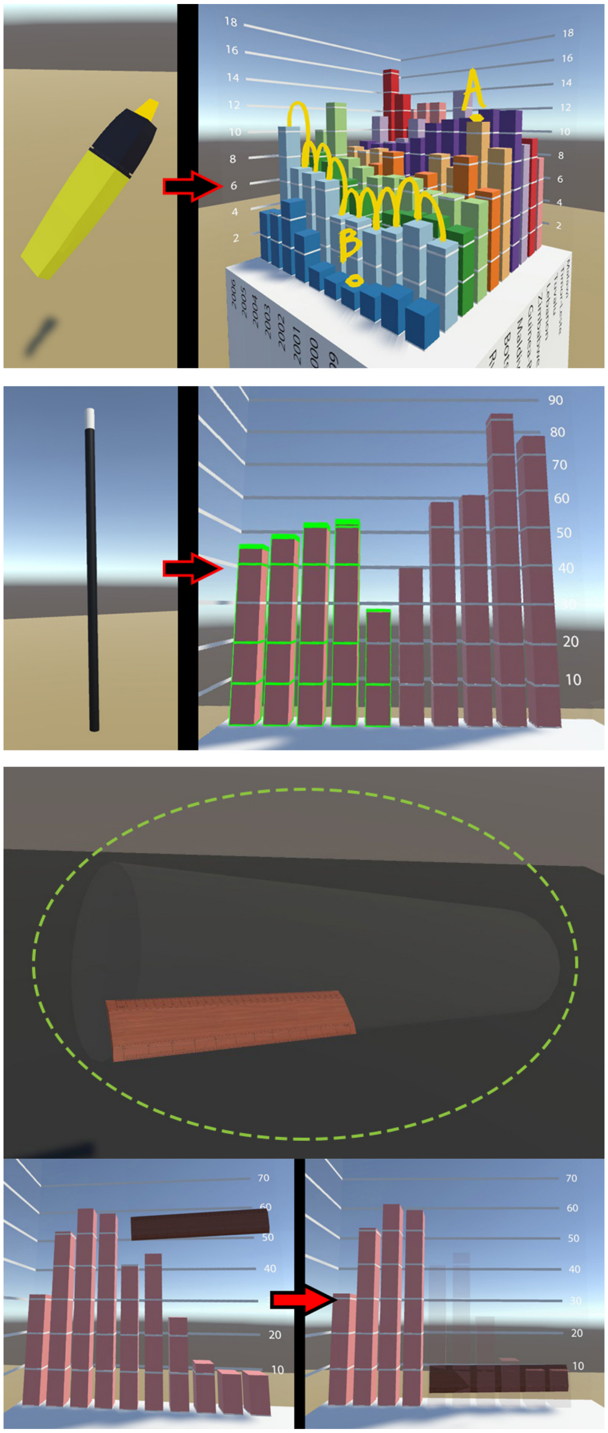

Figure 9: The drawing stylus (top), highlighting wand (middle), and filtering volume (bottom).

flying made them mildly motion sick. However, 2 participants thought they performed best with the room scale since the large bars allowed them to spot differences in height more easily. Overall, participants were more vocal about their emotional reaction to the room-scale than to other sizes, with participants highlighting an average of 4 reactions on the Plutchik wheel (compared to an average of 3.2 reactions for table-scale and 2.2 for hand-scale). The valence of the responses was divided, with 8 participants showing *interest*, but 6 participants also expressing *annoyance*. Three participants indicated *disgust*, specifically in response to the minor motion sickness they experienced.

6 STUDY TWO — ANNOTATION & FILTERING

Following our initial examination, we conducted a second experiment to explore how virtual annotation and measuring tools might alter the experience of using VR visualizations. Current VR tools lack the haptic feedback mechanism necessary to enable the kinds of manual exploration, comparison, and marking with the fingers that are possible with physical models [18]. In contrast, however, virtual environments make it much easier to implement simple interactions which might support many of the same strategies. In response to feedback from participants in our first experiment, we chose to examine two simple mechanisms for annotating charts that might serve as alternatives to touch-based comparison and marking. We also explored the potential for simple implicit filtering tools to combat the issues with occlusion that impede the legibility with 3D charts in both physical and virtual settings.

VR Tools

Tools for Annotation. We designed two annotation tools — a drawing stylus and highlighting wand — which differ primarily in terms of their expressiveness and complexity.

The **drawing stylus** (Figure 9-top) is a simple 3D paintbrush, similar to those in VR drawing applications like Google's Tilt Brush [31]. The stylus allows viewers to draw strokes in midair, creating flexible free-form annotations. These strokes are not affected by gravity or collisions and remain anchored in space relative to the chart. Viewers can create new strokes by holding the trigger and then drawing in space and around objects. The stylus produces yellow strokes about 1.5*cm* across for the table scale and smaller 1*cm* strokes for the hand scale. The strokes have no shading, ensuring high contrast against the more muted colors in the visualizations. However, we allow strokes to cast shadows on the chart itself, further reinforcing the spatial relationship between them. Viewers can also erase strokes using a second tip, summoned by pressing the touchpad on the controller.

The **highlighting wand** (Figure 9-middle) is a much more minimalist implement which supports highlighting but not more flexible annotation. Viewers can highlight bars one at a time by touching them with the wand and pressing the trigger. Highlighted bars receive a bright green outline visible from all directions but retain their original base color. Viewers can toggle highlights off by touching the wand to a bar and pressing the trigger a second time.

Tools for Filtering. We implemented support for filtering via **filtering volumes** (Figure 9-bottom) – transparent cylindrical regions 20*cm* in diameter and 75*cm* long attached to each virtual controller so that they envelop the area around the viewers' hand and arm. When a viewer reaches into the visualization, any bars that collide with the cylinder become semi-transparent, making it possible to examine objects behind them. For virtual rulers, we align the volume with the tool's left edge. This allows viewers to use the ruler to prune the visualization, selectively hiding small sets of bars or deploying the volume as a cutting plane to slice through the entire chart. We also include a filtering volume around each of the annotation tools, allowing viewers to annotate and highlight near the center of the visualization without occlusion from chart elements in the foreground. Based on feedback from pilot studies, we chose to make the volume slightly opaque rather than completely transparent. This slight opacity helps viewers to more precisely understand the extent of the volume and predict how it will behave.

Procedure

Our second study used the same overall tasks and procedure as the first. While the general design of the visualizations did not change from the first experiment, we generated a fresh set of charts – again using data from Gapminder. Based on feedback from the first study, we also doubled the size of the hand-scale chart and rotated it along the y-axis by 45 degrees. These changes provided more space to use the tools and also helped reduce wrist strain. The table-scale visualization design remained the same. We omitted the room-scale, based on the negative feedback from the first study and the practical challenges associated with using the annotation tools on such a large visualization.

We included six conditions (3 tools × 2 scales) each with three tasks. We used 6 different datasets and permuted both condition and dataset order using a Latin square design. We used the same task types and measurements as the first study. The virtual environment and question board remained the same.

We recruited 12 participants for the second experiment. Seven participants had experience with creating or viewing data visualizations. Six of the participants had no experience with VR before this study, but 11 participants had experience with video games in various genres. We refer to these participants below using the codes P1-P12. As in the first study, participants had the opportunity to take a break in between task blocks, each of which took less than 5 minutes. The experiment lasted 50 minutes on average, with the time spent in VR being roughly 30 minutes.

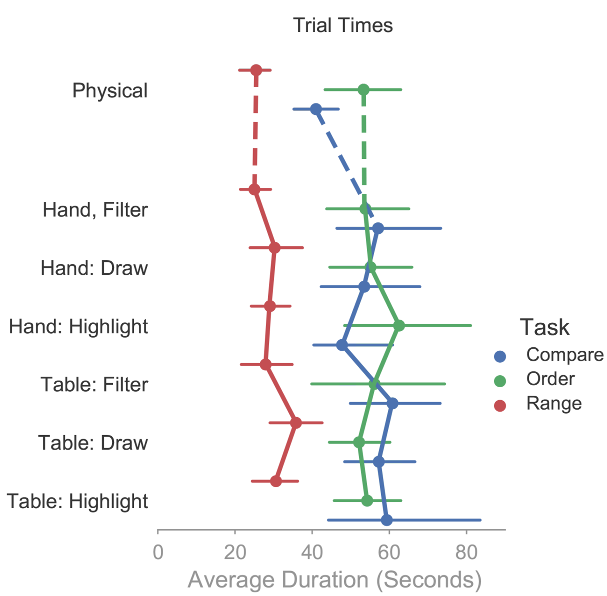

Figure 10: Study 2. Time-on-task for each scale (*hand, table*) × tool (*filter, draw, highlight*) × task (*compare, order, range*) combination. Results from Jansen et al.'s *physical* condition [18] are shown for context (top). Error bars show 95% CIs.

Results

Error Rate. As in the first study, participants' error rates were low (M_{Range}= 0.0184, M_{Order}=0.1037 $M_{Compare}$= 0.1111) and we saw no clear relationship between error rate and either scale or tool.

Time on Task. Participants' task times (Figure 10) were more consistent than in the previous study. Range tasks were markedly faster than the order or compare tasks and results for all three task types were somewhat faster than in the first study. We saw very little discernable difference between the six conditions, and average times for all six aligned much more closely with those reported by Jansen et al. for their physical charts.

Survey Results. Responses from the post-study survey showed that, as in study 1, a majority of participants (9 of 12) preferred the table-scale visualization over the hand-scale (Figure 11). Advocates for the table scale argued that it was more stable and more comfortable to work around, with P11 calling it "easier to comprehend" and P8 noting that "because you can move around it's more comfortable to view the charts". Others stressed that the table reminded them more strongly of a physical object, with P1 writing that the table-scale made it "easier to spatially keep track of things in my mind. [It] felt more hands-on, like I was interacting with something physical/tangible."

When annotating, the majority of participants (7) preferred the highlighting wand over the drawing stylus, while 2 participants indicated a mixed preference. P12 responded that both tools were equally helpful while P9 responded that they used the tools infrequently and only on the compare tasks. Overall, participants most strongly preferred the combination of table-scale visualization and highlighting wand, and most disliked the hand-scale with filtering only. Several participants (P4, P10, and P12) specifically noted that they had a hard time remembering bars of interest when they did not have access to either annotation tool.

Feedback and Observations

In their feedback, all 12 participants expressed a preference for some combination of filtering and annotation tools, rather than a static visualization. Moreover, we observed that all participants

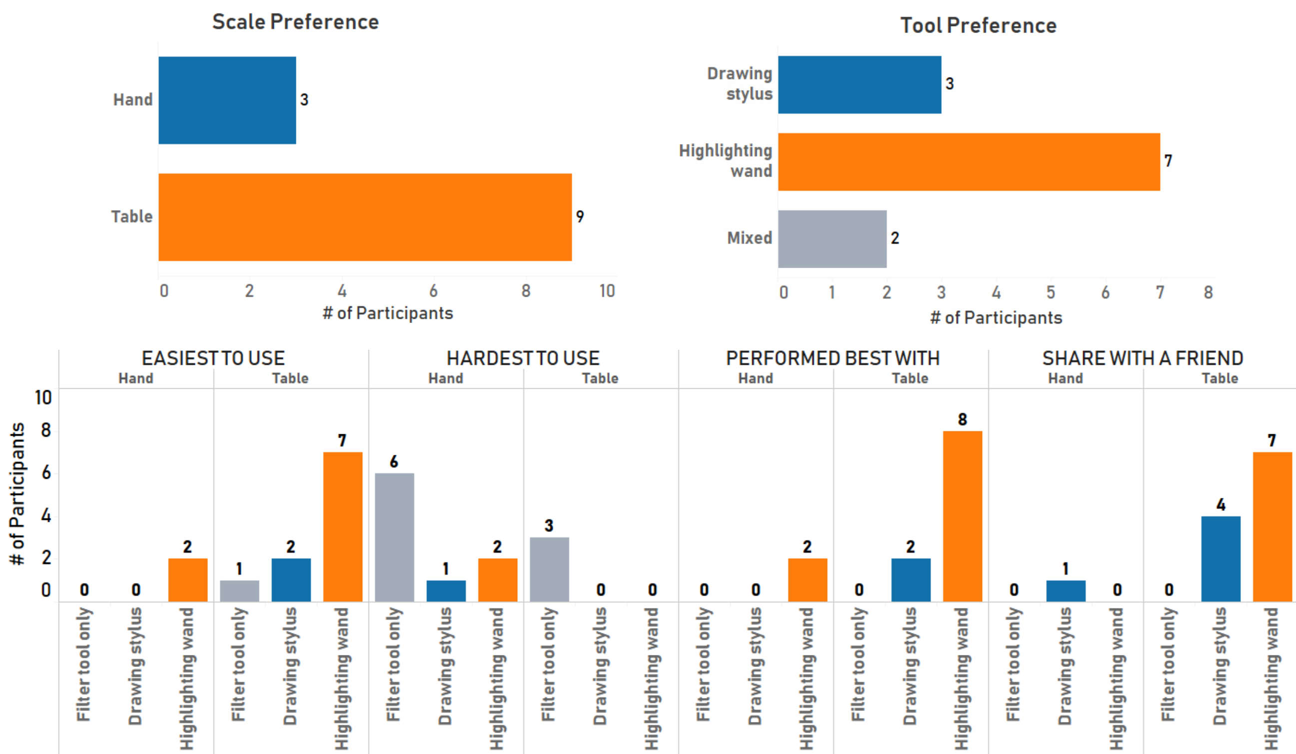

Figure 11: Study 2. Post-study survey results by scale and tool.

actively used some or all of the tools to inspect the visualization and to help externalize their thinking processes.

During the *compare* task all 12 participants used the annotation tools as a way to remember important bars, either by highlighting them or indicating them with a simple visual mark like a line or dot. One participant initially attempted to write numerical values on bars but gave up after finding that writing with the Vive controller was difficult. Generally, participants used marking and highlighting tools in much the same way that participants in Jansen et al.'s original study used their fingers – marking important elements as a form of external memory that allowed them to identify and then revisit those data values.

A smaller subset of participants annotated during the *range* tasks, with 8 participants using the highlighting wand and only 5 participants using the drawing stylus. Those who used the stylus adopted a similar set of strategies to the compare task, with 4 drawing dots and lines to mark important values and 1 writing numerical values on the chart.

In the *ordering* task, we saw the reverse, with 9 participants annotated with the drawing stylus, while only 4 highlighted with the wand. Here, participants used the stylus in several different ways: drawing a lines through or on top of the row of interest, drawing a mark on the label for that row, and marking bars as they answered aloud to ensure they did not miss any. Those who used the highlighting wand generally marked bars in the relevant row, then used the filtering volume to single that row out. Only one participant used this highlighting method for the hand scale.

7 DISCUSSION

Across both studies, hand- and table-scale visualizations exhibited very similar performance, but the table-scale was much more favorably received by participants. Both sizes were small enough that viewers could examine the entirety of the visualization using relatively small physical movements. However, the larger table-scale visualizations allowed viewers to assess differences in bar heights more easily. Participants also seemed to prefer physical locomotion around the static table-scale visualization to the combination of physical movement and manipulation necessary with the hand-scale chart. Moreover, the relatively low-resolution displays and imperfect position tracking of current-generation VR headsets created a number of imperfections in the hand-scale visualizations that may have made them less convincing to viewers than the larger table-sized views.

The poor performance and mixed responses for the bigger room-scale visualizations reflect the underlying challenges associated with exploring and manipulating visualizations at large scales and over physical distances [1]. We expect that other locomotion methods such as teleportation may help address some of the specific concerns about disorientation and motion sickness raised by our participants. However, the larger depth and height and diminished reachability of visualizations at this scale also limit the annotation, filtering, and data manipulation tools that can be used with them. Despite these shortcomings, nearly all of our participants still expressed considerable interest in the room-scale visualization, which highlights the potential for these kinds of larger virtual visualizations as communication and teaching tools. Moreover, VR and AR visualizations that support transitions between multiple scales [17] may have the potential to balance the trade-offs between both large- and small-scale approaches.

VR vs. Physicalization

In comparison to their physical counterparts in Jansen et al.'s original study, participants' generally performed more slowly with our VR charts. While part of this difference may be due to differences in our procedures for recording times and responses, virtual reality versions of these virtual charts at any scale appear unlikely to substantially outperform basic physical versions for these simple tasks. Given the current state of VR and tools, visual realism and the lack of tactility represent the main divides still separating physicalizations from VR visualizations.

The degree to which visual realism plays a role in the perception or interpretation of abstract visualizations remains open for debate. Based on their evaluations, Jansen et al. speculated that a lack of realism might hinder performance for on-screen representations. However, Berard et al.'s work on handheld projection-mapped displays highlights how the lack of occlusion and higher contrast of a virtual object can actually *improve* performance over using a physical one [3]. Still, specific limitations of modern VR hardware like the vergence-accommodation conflict — wherein the apparent focal depth of virtual objects diverges from the actual distance of the VR display from the eye — may indeed hinder viewers' ability to comfortably use certain virtual visualizations [15]. Moreover, these effects are the most pronounced for nearby objects like our hand-scale visualizations, where incorrect focus cues are more likely to lead to divergence [22].

Finally, Jansen et al.'s initial results suggest that support for direct touch and physical manipulation were likely the biggest advantage of their physical prototypes [18]. Haptic displays or shape-changing interfaces capable of replicating this degree of tactility for VR and AR visualizations remain a very distant prospect. However, our participants' active use of annotation tools and virtual props (like the rulers in our studies) to perform many of the same kinds of marking and manipulation operations is promising. These findings raise the possibility that tools and interactions which enable viewers to inspect, manipulate, and externalize their though process visually on top of VR and AR visualizations could provide many of the same advantages as physicalizations. Hybrid techniques, which combine virtual and physical approaches by fusing tactile input and output devices with more elaborate VR and AR visuals [7], are also promising.

8 CONCLUSION & FUTURE WORK

We considered the use of simple virtual bar charts at three scales and using several annotation and filtering mechanisms. Our results find clear advantages for tabletop-scale VR visualizations, which strike a balance between readability and reachability that allows viewers to both examine and manipulate them easily. Our findings also showcase the value of virtual annotation tools, which can potentially provide many of the benefits typically associated with data physicalizations to their VR equivalents.

However, research on both physicalization and immersive VR/AR visualization are still emerging fields. As a result, considerable additional work is needed to understand the advantages and disadvantages of these techniques for more complex and realistic visualization types, interaction techniques, tasks, and scales. In that space, VR visualization tools represent an opportunity to leverage many of the benefits of the physical world, while also transcending its limitations.

REFERENCES

1. Christopher Andrews, Alex Endert, Beth Yost, and Chris North. 2011. Information visualization on large, high-resolution displays: Issues, challenges, and opportunities. *Information Visualization*. https://doi.org/10.1177/1473871611415997
2. Kevin W. Arthur, Kellogg S. Booth, and Colin Ware. 1993. Evaluating 3D task performance for fish tank virtual worlds. *ACM Transactions on Information Systems* 11, 3: 239–265. https://doi.org/10.1145/159161.155359
3. Francois Berard and Thibault Louis. 2017. The Object Inside: Assessing 3D Examination with a Spherical Handheld Perspective-Corrected Display. *CHI '17 Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*: 4396–4404. https://doi.org/10.1145/3025453.3025806
4. Atelier Brükner. 2013. Walkable Age Pyramid. Retrieved from http://dataphys.org/list/walkable-age-pyramid/
5. Richard Burdett. 2007. Global Cities: Elevation Maps of City Population. Retrieved from http://dataphys.org/list/city-populations/
6. Tom Chandler, Maxime Cordeil, Tobias Czauderna, Tim Dwyer, Jaroslaw Glowacki, Cagatay Goncu, Matthias Klapperstueck, Karsten Klein, Kim Marriott, Falk Schreiber, and Elliot Wilson. 2015. Immersive Analytics. In *2015 Big Data Visual Analytics, BDVA 2015*. https://doi.org/10.1109/BDVA.2015.7314296
7. Maxime Cordeil, Benjamin Bach, Yongchao Li, Elliott Wilson, and Tim Dwyer. 2017. Design space for spatio-data coordination: Tangible interaction devices for immersive information visualisation. *IEEE Pacific Visualization Symposium*: 46–50. https://doi.org/10.1109/PACIFICVIS.2017.8031578
8. Maxime Cordeil, Andrew Cunningham, Tim Dwyer, Bruce H. Thomas, and Kim Marriott. 2017. ImAxes: Immersive Axes As Embodied Affordances for Interactive Multivariate Data Visualisation. *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology - UIST '17*: 71–83. https://doi.org/10.1145/3126594.3126613
9. Maxime Cordeil, Tim Dwyer, Karsten Klein, Bireswar Laha, Kim Marriott, and Bruce H. Thomas. 2017. Immersive Collaborative Analysis of Network Connectivity: CAVE-style or Head-Mounted Display? *IEEE Transactions on Visualization and Computer Graphics* 23, 1: 441–450. https://doi.org/10.1109/TVCG.2016.2599107
10. Çağatay Demiralp, Cullen D. Jackson, David B. Karelitz, Song Zhang, and David H. Laidlaw. 2006. CAVE and fishtank virtual-reality displays: A qualitative and quantitative comparison. *IEEE Transactions on Visualization and Computer Graphics* 12, 3: 323–330. https://doi.org/10.1109/TVCG.2006.42
11. Ciro Donalek, S. G. Djorgovski, Alex Cioc, Anwell Wang, Jerry Zhang, Elizabeth Lawler, Stacy Yeh, Ashish Mahabal, Matthew Graham, Andrew Drake, Scott Davidoff, Jeffrey S. Norris, and Giuseppe Longo. 2014. Immersive and Collaborative Data Visualization Using Virtual Reality Platforms. *2014 IEEE International Conference on Big Data Immersive*: 609–614. https://doi.org/10.1109/BigData.2014.7004282

12. Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. 2013. inFORM: Dynamic Physical Affordances and Constraints through Shape and Object Actuation. *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*: 417–426. https://doi.org/10.1145/2501988.2502032
13. Dan Goods, Nikolaus Hafermaas, and Jamie Barlow. airFIELD. Retrieved from http://airfieldsculpture.com/
14. Elisa Gusai, Chiara Bassano, Fabio Solari, and Manuela Chessa. 2017. Interaction in an Immersive Collaborative Virtual Reality Environment: A Comparison Between Leap Motion and HTC Controllers. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-70742-6_27
15. David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. 2008. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*. https://doi.org/10.1167/8.3.33
16. Trevor Hogan and Eva Hornecker. 2012. How Does Representation Modality Affect User-Experience of Data Artifacts? *Proceedings of Haptic and Audio Interaction Design*. https://doi.org/10.1007/978-3-642-32796-4_15
17. Alexander Ivanov, Kurtis Thorvald Danyluk, and Wesley Willett. 2018. Exploration & Anthropomorphism in. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*: 1–6. https://doi.org/\url{10.1145/3170427.3188544}
18. Yvonne Jansen, Pierre Dragicevic, and Jean-Daniel Fekete. 2013. Evaluating the efficiency of physical visualizations. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*: 2593. https://doi.org/10.1145/2470654.2481359
19. Yvonne Jansen, Pierre Dragicevic, and Jean-Daniel Fekete. 2013. Evaluating the efficiency of physical visualizations. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*: 2593. https://doi.org/10.1145/2470654.2481359
20. Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornbæk. 2015. Opportunities and Challenges for Data Physicalization. *Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/2702123.2702180
21. Aaron Koblin, Nikolaus Hafermaas, and Dan Goods. eCloud. Retrieved from http://www.ecloudproject.com/
22. Gregory Kramida. 2016. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*. https://doi.org/10.1109/TVCG.2015.2473855
23. Oh Hyun Kwon, Chris Muelder, Kyungwon Lee, and Kwan Liu Ma. 2016. A study of layout, rendering, and interaction methods for immersive graph visualization. *IEEE Transactions on Visualization and Computer Graphics* 22, 7: 1802–1815. https://doi.org/10.1109/TVCG.2016.2520921
24. Daniel Leithinger and Hiroshi Ishii. 2010. Relief: a scalable actuated shape display. *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*: 221–222. https://doi.org/10.1145/1709886.1709928
25. Ekaterina Olshannikova, Aleksandr Ometov, Yevgeni Koucheryavy, and Thomas Olsson. 2015. Visualizing Big Data with augmented and virtual reality: challenges and research agenda. *Journal of Big Data* 2, 1. https://doi.org/10.1186/s40537-015-0031-2
26. Simon Stusak, Jeannette Schwarz, and Andreas Butz. 2015. Evaluating the Memorability of Physical Visualizations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. https://doi.org/10.1145/2702123.2702248
27. Ivan E. Sutherland. 1968. A head-mounted three dimensional display. *Proceedings of the December 9-11, 1968, fall joint computer conference, part I on - AFIPS '68 (Fall, part I)*: 757. https://doi.org/10.1145/1476589.1476686
28. Faisal Taher, John Hardy, Abhijit Karnik, Christian Weichel, Yvonne Jansen, Kasper Hornbæk, and Jason Alexander. 2015. Exploring Interactions with Physically Dynamic Bar Charts. *Proc. CHI'15*: 3237–3246. https://doi.org/10.1145/2702123.2702604
29. S.K. Tang, Y. Sekikawa, D. Leithinger, S. Follmer, and H. Ishii. 2014. Tangible CityScape. Retrieved from http://tangible.media.mit.edu/project/tangible- cityscape
30. Rensu P. Theart, Ben Loos, and Thomas R. Niesler. 2017. Virtual reality assisted microscopy data visualization and colocalization analysis. *BMC Bioinformatics* 18, Suppl 2: 1–16. https://doi.org/10.1186/s12859-016-1446-2
31. Tilt Brush by Google. Retrieved from https://www.tiltbrush.com/