Touch and Beyond: Comparing Physical and Virtual Reality Visualizations

Kurtis Danyluk, Teoman Ulusoy, Wei Wei, and Wesley Willett

Abstract—We compare physical and virtual reality (VR) versions of simple data visualizations and 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 examines differences in how viewers interact with physical hand-scale, virtual hand-scale, and virtual table-scale visualizations and the impact that the different forms had on viewer's problem solving behavior. A second study examines 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 virtual reality representations and hint at the potential of interactive annotation and filtering tools in VR visualizations.

Index Terms—Human-Computer Interaction, Visualization, Data Visualization, Virtual Reality, Physicalization.

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 [1] 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 real-world tasks. Meanwhile work on data physicalization [2] has highlighted the potential of representing data via real objects in physical spaces. However the trade-offs associated with presenting data using these 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 [3] and that viewers may find them more memorable than visualizations on paper [4]. While current VR and AR tools are not able to support this kind of tactile feedback and manipulation, they 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 interaction and manipulation, allowing viewers to reach through visualizations or dynamically change their form and behavior, while still retaining many of the characteristics of physicalizations.

In this paper we are interested in determining if VR versions of physical visualizations can convey some of the benefits of physicalizations. To answer this question we conducted two experiments which compare physical 3D bar charts against virtual copies at two different scales, and then test whether simple annotation tools can replicate some of the affordances lost in the transition from physical to virtual.

- K. Danyluk, T. Ulusoy, Wei W., and W. Willett are with the Department of Computer Science, University of Calgary, Calgary, AB, T2N 1N4. E-mail: {kdanylu, ttulusoy, wei.wei2, wesley.willett}@ucalgary.ca

Manuscript received ...; revised ...

We designed this initial examination of the potential of immersive virtual visualizations by extending Jansen et al.'s studies of physicalizations [3] to VR environments. We first describe a new version of Jansen et al.'s original experiment in which we recreated their physical visualizations out of Lego and compared them against virtual versions at hand-sized and table-sized scales. We then explore the addition of new interactive annotation and filtering tools enabled by the move to VR. Our results highlight a clear preference among participants for physical versions of the visualizations, as well as enthusiasm for the kinds of interactive tools introduced in the VR versions. 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.

2.1 Virtual Reality Information Visualization

Virtual reality (VR) is by no means a new field of research, with the first system created by Sutherland [5] 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 [6], [7].

Early investigations of abstract data visualizations in VR typically used either "fishtank" VR or CAVE systems which rely on head-tracking and stationary displays [8], [9]. As early as 1993, Arthur et al. [8] examined participants' ability to trace tree structures using a fish tank VR setup and found considerable speed and accuracy benefits compared to its 2D, on-screen, counterpart. Later work by Demiralp et al. [9]

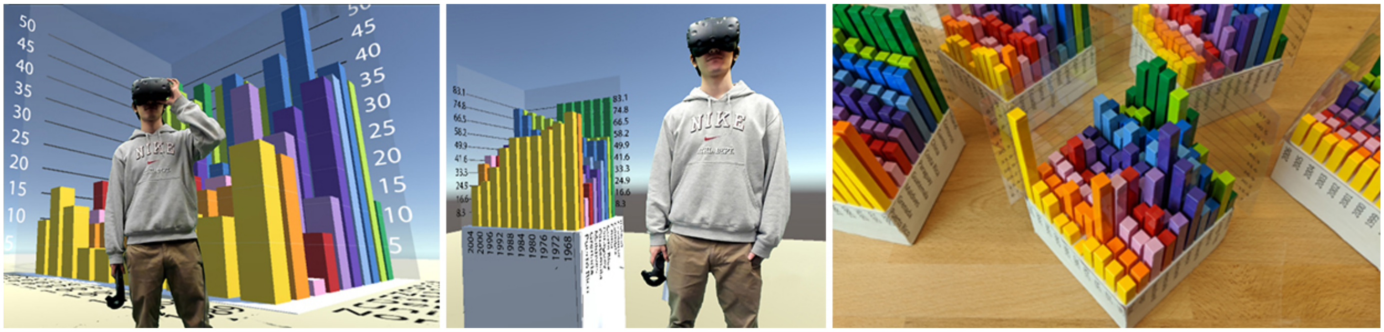

Fig. 1. (From left to right) A virtual room-scale visualization (6.4 m diameter across with 1.7 m person for context), a virtual table-scale visualization (64 cm diameter across), and physical hand-scale versions (10 cm diameter across).

further explored the impact of visualizations of different scales using both fish tank VR and CAVE VR systems. Their findings highlighted the potential of VR visualization generally, while noting that fish tank 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 [6] which adapt traditional abstract visualizations like scatterplots to shared 3D spaces, as well as more complex tools like Cordeil et al.'s ImAxes [10] 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 [11]. Experiments have also shown that visualizations displayed on HMDs compare favorably against much more costly CAVE systems [12]. Similarly, work on immersive unit visualization [13] has 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.

2.2 Physical Visualizations

Meanwhile, work on data physicalization has identified a variety of benefits for immersive physical instantiations of data [2]. 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 and emotional presence [14] – traits which VR and AR tools are increasingly able to approximate. As such, our experiments are heavily inspired by fundamental work by Jansen et al. [3] which 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.

Studies by Berard and Louis. [15] have also begun to explore the interstitial space between physical and virtual systems, examining novel "handheld perspective-corrected displays" which can stereoscopically project complex interactive puzzles and other objects onto simple volumetric props. Interestingly, participants in Berard et al.'s studies were able to solve complex 3D puzzles faster and more accurately when using projected virtual objects than when using physically printed ones – likely because the virtual objects did not suffer from the poor contrast, occlusions, and other shortcomings of the physical materials. However, other recent studies have highlighted some challenges related to viewers' perception of physicalizations. For instance, Jansen and Hornbæk [16] have shown consistent biases in viewers' perception of physicalizations that use size as a physical variable (reminiscent of similar biases in 2D and 3D on-screen representations). Similarly, Sauvé et al. [17] have shown that the orientation and layout of a physicalization can drastically change how viewers interpret it.

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 [2] and may also encourage greater memorability [4] and engagement [18]. However, physicalizations can be complicated, difficult, and impractical to construct – especially as their scale and degree of interactivity increases. Even relatively simple tabletop systems like EMERGE [19] and inFORM [20], for example, required long-term, concentrated engineering efforts to develop and maintain. Meanwhile the few examples of even larger room- and building-scale visualizations 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 kinds of spatially-embedded 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 examine the potential for VR interfaces that build on the kinds of simple chart designs and interactions that already show promise in the physical world. Specifically, we use virtual reality prototypes to recreate 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 Experiments

We based both our visualizations and our experiment designs on those used by Jansen et al. [3] in their early studies of physicalization use. In these studies, participants used small physical 3D bar charts as well as 2D and 3D on-screen versions 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 physicalizations likely related to participants' ability to manipulate and inspect the objects while simultaneously 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 the 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.

4.1 Visualization Designs

To examine these tradeoffs further, we created a variety of virtual and physical charts which mirror the bar charts created by Jansen et al. (Figure 2). Like the originals, our charts (Figure 1) featured a 10 $\times$ 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 VR charts were 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

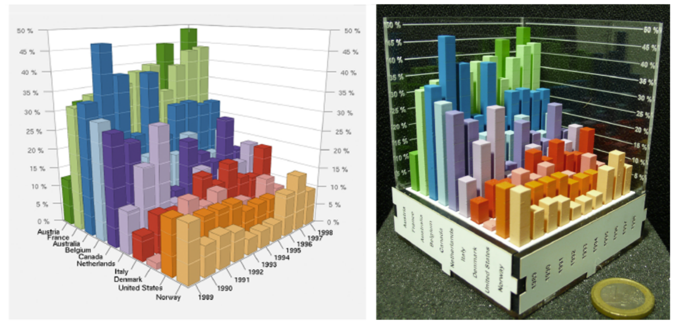

Fig. 2. Jansen et al.'s 3D on-screen visualization and physicalization [3].

from Gapminder [21] organized by country. We opted to use percentages for all axes, rather than raw counts or intervals, to reduce the potential for confusion.

We initially created three different versions of these charts in VR to examine the impact of visualization scale. The smallest of these virtual charts were **hand-scale**, measuring roughly 10 cm across. While the overall form of the visualization mirrored of Jansen et al.'s original stimuli, we increased the dimensions by 25% (from 8 cm to 10 cm) in order to ensure the legibility of labels in VR. These resulting charts are also similar in scale to the VR small-multiple bar charts used in recent work by Liu et al. [22]. Next, we created **table-scale** versions which measured 64 cm across, similar to the size of tabletop bar-chart displays like EMERGE [19] (Figure 3a) and shape-changing displays like Relief [23], Tangible Cityscape [24] and inFORM [20]. We placed these table-scale visualizations atop a virtual plinth with a default height of 1 m. We also created **room-scale** versions of the visualizations, which measured 6.4 m to a side. This scale was inspired by large-scale installations like Richard Burdett's population-density models of major cities [25], the walkable age pyramid (Figure 3b) created by Atelier Brükner [26], and the eCLOUD [27] and airFIELD [28] sculptures – all of which allow viewers to explore data by physically walking through, under, and around it. Finally, we created **physical hand-scale** charts with the same 10 cm dimensions as our VR versions. While Jansen et al. constructed their original charts using laser-cut and painted acrylic, we built ours out of Lego bricks with custom 3D-printed baseplates. This allowed us to construct new charts more quickly, while also precisely matching the dimensions of our 10 cm virtual hand-scale charts. We excluded 2D versions of the charts, which have already been examined extensively in Jansen et al.'s work, and instead focused explicitly on comparisons between physical charts and their VR counterparts. We provide more detailed descriptions of the specific designs used in each of our experiments below.

4.2 Virtual Environment

We conducted the VR component of our experiments using a test environment that we implemented using Unity which supports a variety of 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,

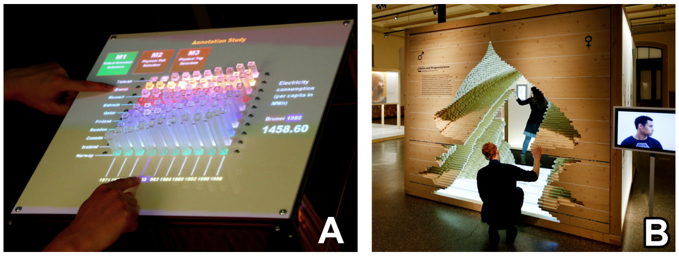

Fig. 3. (A) EMERGE tabletop display [19]. (B) Walkable age pyramid (Exhibition Design: ATELIER BRÜCKNER, Photography: Michael Jungblut) [26].

including gestural hand-tracking [29], [30]. However, we chose to use Vive controllers and Vive trackers for input, 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 [31]. Participants held a controller or tracker in each hand at all times during the studies. In the virtual environment, the tracker appeared as a hand-scale visualization and the controller appeared as a 30 cm virtual wand in the virtual hand-scale condition or as a 30 cm ruler in the virtual table-scale condition.

4.3 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.

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 4) attached to one of the controllers for the virtual conditions. In VR, participants could summon or dismiss the question board as needed. When using larger visualizations the question board appeared to the right of one of the rulers. For the smaller hand-scale visualizations the board appeared in front of the virtual wand, so as not to obscure the chart. Participants were free to choose either hand to hold the board. In the physical hand-scale condition we asked the questions out loud. 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 through tasks. Timing started as soon as the experimenter started reading the question, and ended when the participant uttered the word "confirm". We calculated error using the same method as Jansen et al. For range tasks, the error was the average absolute difference between the participant's min and max values and the true values, divided by the total axis range. For order tasks, the error was the normalized Kendall Tau distance (the number of pairwise disagreements) between the answer and the true order. For compare tasks, the error was 0 or 1, based on correctness.

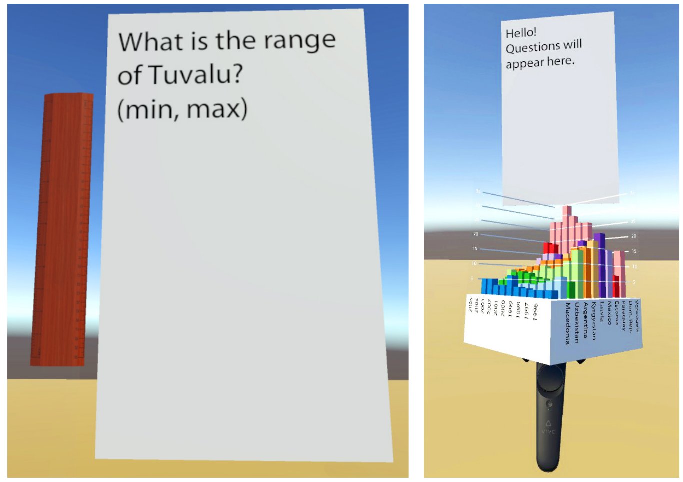

Fig. 4. Question board design and location relative to the ruler tool (left) and hand-scale visualization (right).

4.4 Pilot Study

To test the viability of the three different visualizations scales, we first ran a VR-only pilot study in which we asked 9 participants to use and compare virtual hand-scale, virtual table-scale, and virtual room-scale visualizations. Each pilot participant completed one task of each type (range, order, compare) at each scale, then completed an exit questionnaire in which they discussed their experience and impressions of using each scale.

We found that for all scales and tasks participants in our study performed more slowly than participants in Jansen et. al.'s original experiment [3]. We also found that while there was no clear difference in performance between the virtual hand-scale and virtual table-scale visualizations, the virtual room-scale condition was considerably slower than the other two. Participants' feedback echoed this, with all participants reporting that they found either the virtual hand-scale or virtual table-scale visualizations the easiest to use, with the majority preferring virtual table-scale. Participants responded more negatively to the virtual room-scale condition, noting that while the large sizes of the bars made it easier to differentiate very similar values, navigating and examining the chart was considerably more difficult. Across all three conditions, much of the feedback we received from participants reflected a desire for better tools. Four participants specifically asked for the ability to mark and select bars, while another suggested filtering tools to hide rows or columns that obscured their view.

5 Experiment One – Physical vs. Virtual

In our first full experiment, we compared the physical hand-scale charts against their virtual counterparts at both hand- and table-scales (Figure 5). This allowed us to examine how changes in scale and tangibility impacted participants' ability to perform basic chart reading and analysis tasks and how it changed their overall experience. While we include a virtual table-scale condition, practical challenges associated with constructing enough table-scale physical charts to properly counterbalance our study conditions prevented us from including a physical table-scale condition.

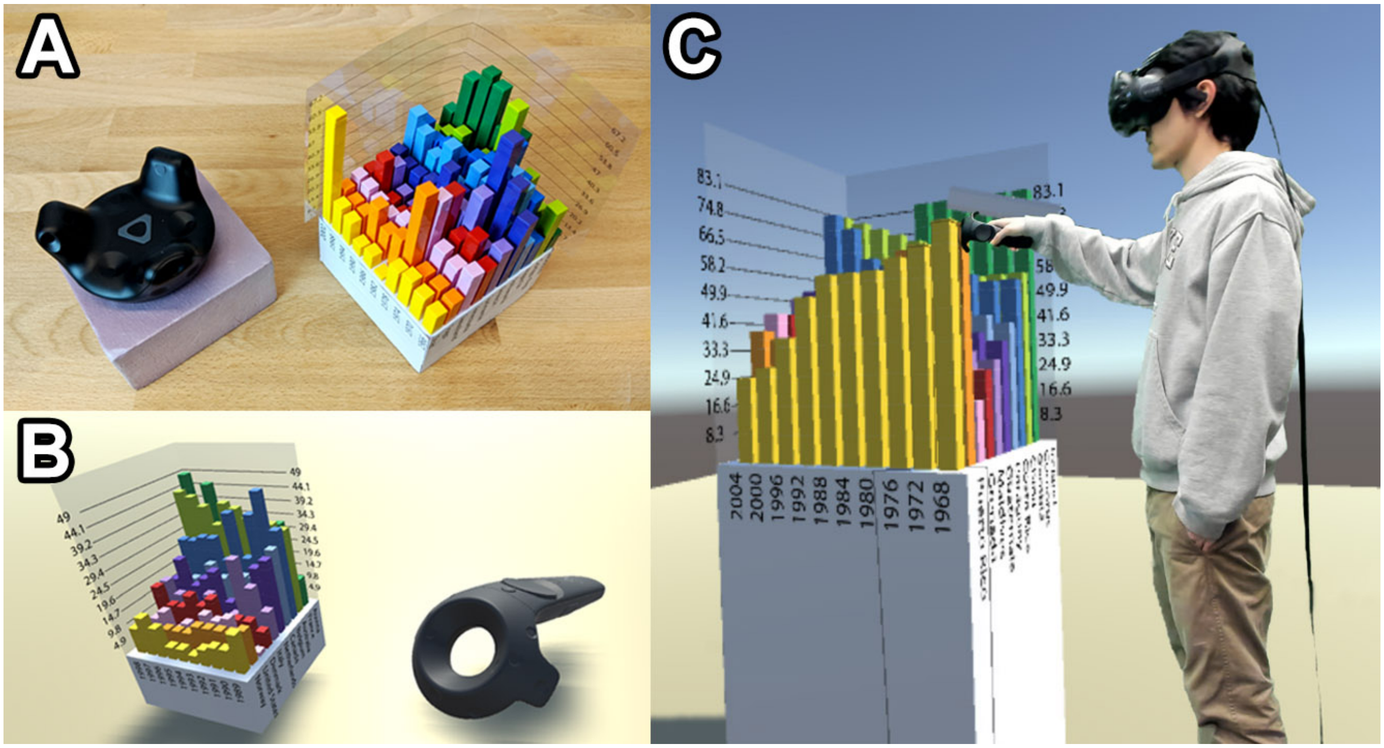

Fig. 5. Experiment 1 stimuli: (A) physical hand-scale chart, (b) virtual hand-scale chart, and (C) virtual table-scale (with 1.7m person).

5.1 Visualizations

We modeled our **physical hand-scale** visualizations after Jansen et. al.'s. Each featured a 10×10 array of bars created out of 1×1 Lego bricks topped by 1 × 1 Lego plates to a maximum height of 10 bricks. We 3D printed custom 10 × 10 white Lego base plates that had the same relative spacing intervals as the charts created by Jansen et. al. We then mounted these charts on 5 cm thick foam blocks and attached printed labels to them. We printed the axes on transparencies and attached these to the back of the foam. The final physical charts measured 10 cm across and weighed between 152 g and 300 g depending on the size of the bars (somewhat less than Jansen et al.'s smaller 8 cm charts, which weighed between 270 g and 350 g). Because of the fixed thickness of Lego plates, our charts have coarser vertical resolution than Jansen et al.'s. Additionally, we spaced the tick lines for the data set at intervals 1 Lego plate thick. As a result, we adapted our task prompts to ensure that no tasks were ambiguous or overly simple given this rounding of the bar heights. To preserve the comparability between the physical and virtual charts, we adapted the size, appearance and functionality of the **virtual hand-scale** visualizations in this study to match the Lego versions. This entailed adjusting the bar colors, bar heights, and tick marks to match the coarser vertical resolution of Lego charts and the colors of the Lego bricks as well as restricting virtual functionality to operations that could also be performed on the physical charts to reduce confounds. To more closely approximate the experience of holding a physical chart, we attached a Vive Tracker to a foam block with the same dimensions as the ones we used for the Lego versions (Figure 5a). This allowed participants to use the foam block to hold, rotate, and manipulate the virtual chart. The block and Vive Tracker together weighed 103 g, and its balance and heft felt similar to the lighter physical charts.

For the **virtual table-scale**, we used the same 64 cm width as in the pilot and again adjusted the bar heights, tick marks, colors, and spacing to match the physical charts. Before each study we 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.

Using these three templates we created matching sets of charts for each condition using 6 different Gapminder datasets, plus a unique training chart used for each condition. Due to the **virtual room-scale** visualizations' poor performance and negative feedback in the pilot study, we did not include it in the full experiment.

5.2 Procedure

Our independent variables included the 3 conditions (physical hand-scale, virtual hand-scale, and virtual table-scale) and 3 task types (range, order, compare). Each participant completed 6 blocks of tasks (two in each condition) each using a different chart. Within each block, participants completed two tasks of each type, for a total of 36 trials. We permuted conditions and tasks using a balanced Latin square. We also permuted the order in which we presented charts using a second independent Latin square. Participants saw each chart for only one block.

We included a training block at the beginning of the study where we asked participants to perform each task type once in each different chart condition. During this training block we helped participants adjust the VR equipment to ensure a good fit and appropriate interpupillary distance, adjusted the table-scale visualization to a comfortable height, and explained how to perform each of the task types. We allowed participants to take a break and remove the HMD after every 3 tasks (half of a block) as well as between blocks. Afterwards we asked participants to complete a post-study questionnaire. Most participants took approximately 1 ½ hours to complete the study.

We recruited a total of 18 participants through internal university email lists and via snowball sampling. The majority of participants reported some prior experience with data visualization (14/18). Most also had some VR experience prior to the study (13/18). Eleven reported experience both with data visualization and VR. In total we recorded 648 trials: 18 users $\times$ 3 main conditions $\times$ 2 blocks $\times$ 3 tasks $\times$ 2 repetitions.

5.3 Analysis

We preregistered both our study design and analysis before beginning the experiment. Our registration is available at: https://osf.io/vsz8m and Jupyter notebooks containing our complete time and error analyses are included in our supplementary materials.

We analyze our results using estimation techniques and and report results using confidence intervals (CIs) as is consistent with recent APA guidelines [32]. All confidence intervals, both in the charts and in the text are 95% bootstrap confidence intervals. This use of pre-registered studies and confidence-interval comparison in lieu of null hypothesis significance testing reflects emerging best practices in a wide variety of fields [33], [34].

5.4 Results

Error rates were low across all tasks, conditions, and participants. As a result, our quantitative results focus predominantly on timing. We also report qualitative observations as well as results from our post-study questionnaire.

Error Rate. Across all conditions the mean error rate for both *range* and *order* tasks (*Range*= 0.09, *Order*=0.04) was very low. We saw higher error rates for the *compare* tasks (*Compare*=0.1037) though this is likely due to the binary nature of the questions.

Time on Task. We computed average time-on-task by participant for each task and condition (Figure 7). For *range* tasks, all conditions produced similar times, with the physical hand-scale chart being marginally faster than both virtual conditions (Figure 8-top). However, there was a clear difference in performance in both *order* tasks (Figure 8-center) and *compare* tasks (Figure 8-bottom), where the physical hand-scale chart was markedly faster than either virtual condition. For *order* tasks, participants were more than 20s faster on average with the physical hand-scale chart (43.3s, CI = 39.1s, 47.3s) than both the virtual hand-scale (74.7s, CI = [64.8s, 85.1s]) and the virtual table-scale (69.5s, CI = [61.8s, 77.2s]). This pattern persisted for the *compare* tasks, with even larger differences between the physical hand-scale chart (33.6s, CI = [28.6s, 38.9s]) and the virtual hand-scale (56.6s, CI = [49.2s, 65.1s]) and virtual table-scale (65.7s, CI = [56.6s, 76.4s]) versions.

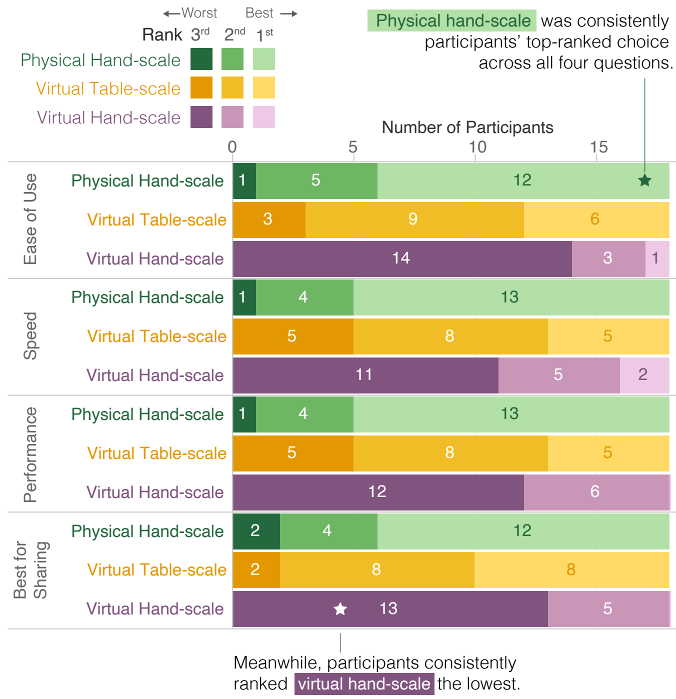

Fig. 6. **Experiment 1.** Post-study survey results. Participants ranked the three chart types based on their perceived ease of use, speed, which they felt they performed best with, and which they would share with others. Ties were possible, and several participants ranked both physical hand-scale and virtual table-scale as the best.

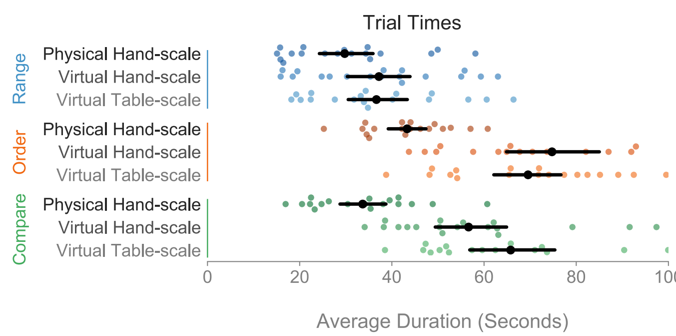

Fig. 7. **Experiment 1.** Time-on-task for each condition $\times$ task combination. Error bars show 95% CIs.

5.5 Feedback and Observations

We found that all participants chose to hold both the physical and the virtual hand scale visualizations with their non-dominant hand so that they could point at bars with their dominant hand. This is consistent with prior work with VR props and worlds-in-miniature [35] and human motor behavior studies [36], in which users often used their non-dominant hand as a reference frame while using their dominant hand to perform finer-grained operations. When using the physical hand-scale charts, all participants actively touched and tracked bars using the fingers on their dominant hand, much like participants in Jansen et al.'s study [3]. Across all scales, participants seldom used rulers for measuring, and instead used them as a pointing device to help track and recall specific bars. Five participants were comfortable with clipping through the virtual table-scale visualization with their headset and body as it helped them

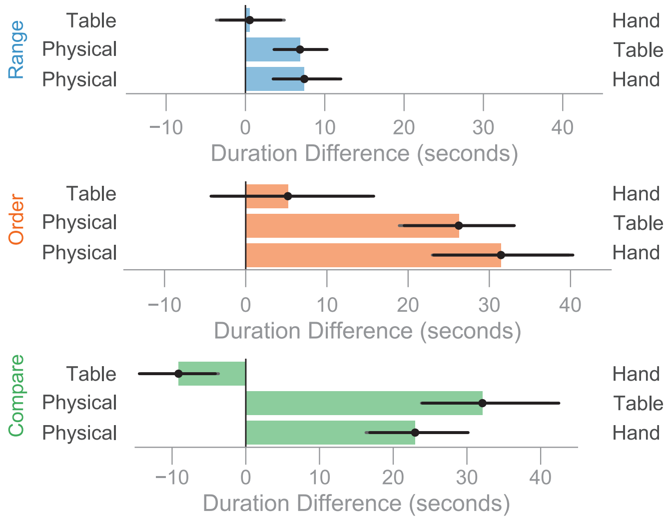

Fig. 8. **Experiment 1.** Pairwise comparisons of time-on-task for each condition $\times$ task combination. Error bars show 95% CIs.

see bars that were obscured by taller bars. No participants showed any concern about clipping through the bars with the virtual ruler.

Overall, 9 participants voiced a strong preference for the physical hand-scale visualizations over the other conditions, with several noting that they preferred the ability to physically touch and manipulate it with their hands. However, participants also noted some drawbacks, including that the visualizations' relatively small size made it difficult to see bars near their center. In contrast, participants expressed a common dislike for the virtual hand-scale, and no participant preferred it over other scales. Six participants preferred the virtual table-scale with 2 participants citing the ability to more easily examine and read the central bars as the primary benefit. However, 7 participants disliked that the larger size required more physical movement in order to accurately answer the questions. Finally, 2 participants did not list any specific preference, with both stating that each chart may do well in different contexts.

Similarly, in their post-study feedback, the majority of participants ranked the physical hand-scale condition as the easiest to use and fastest of the three (Figure 6). A majority also ranked the physical hand-scale charts as the one that they performed the best with and as their preferred version for sharing with others. For all questions at least two thirds of participants rated physical hand-scale as the best. On the other hand, most participants rated virtual hand-scale as the worst for each question.

Fourteen of the participants specifically expressed a desire for better tools for manipulating the VR visualizations. Six suggesting adding highlighting or marking tools that would allow them to keep track of bars, while 7 asked for mechanisms that would let them filter or hide bars they were not interested in.

6 Experiment Two – Annotation & Filtering

Based on strong feedback requesting additional tools, we also conducted a second experiment to explore how virtual

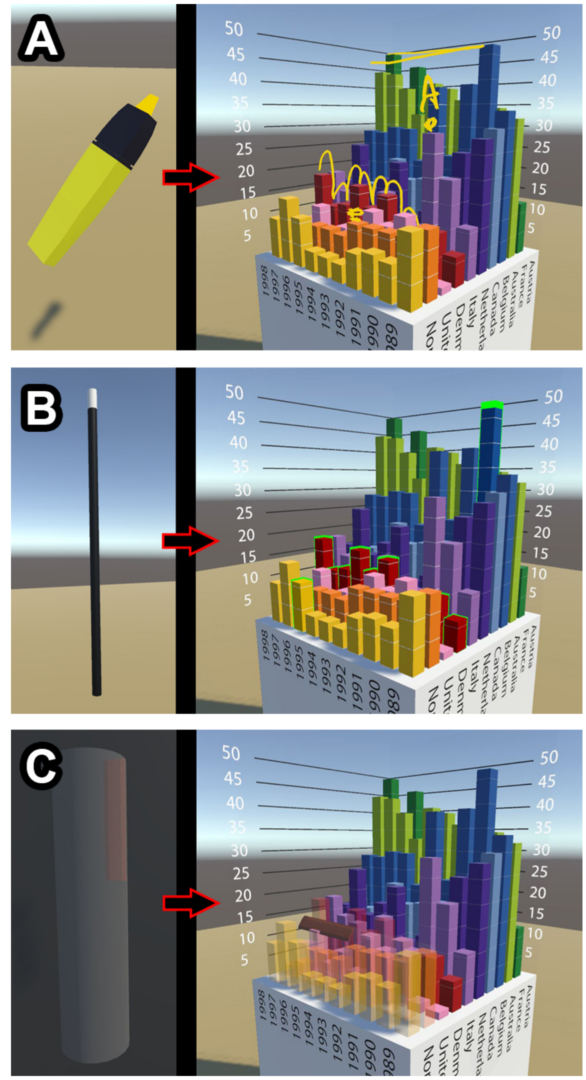

Fig. 9. The drawing stylus (top), highlighting wand (middle), and filtering volume (bottom).

annotation and measuring tools might alter the experience of using VR visualizations. Current VR tools, even experimental ones, lack the precise haptic feedback mechanisms necessary to enable the kinds of manual exploration, comparison, and marking with the fingers that are possible with physical charts. 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 pilot 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.

6.1 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 [37]. 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 more minimalist tool 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 off highlights 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 which 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. We incorporated filtering volumes into all of virtual tools, including the drawing stylus, highlighting wand, and virtual ruler. For virtual rulers, we aligned the volume with the tool's left edge. This allowed 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. For the drawing stylus and highlighting wand we included a filtering volume around the annotation tool, 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 helped viewers to more precisely understand the extent of the volume and how it would behave.

6.2 Procedure

Our second experiment 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. To support more precise highlighting and annotation, we also increased the size of the virtual hand-scale charts to 24 cm and allowed participants to move and manipulate them using a Vive controller instead of the foam block and Vive tracker we used in the first experiment. These changes provided more space to use the tools and also helped reduce wrist strain. Because we did not include a physical condition, we did not constrain bars to discrete Lego height intervals, instead using the full vertical resolution of the bars. We also added more visible tick marks to both the virtual hand-scale and virtual table-scale charts.

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

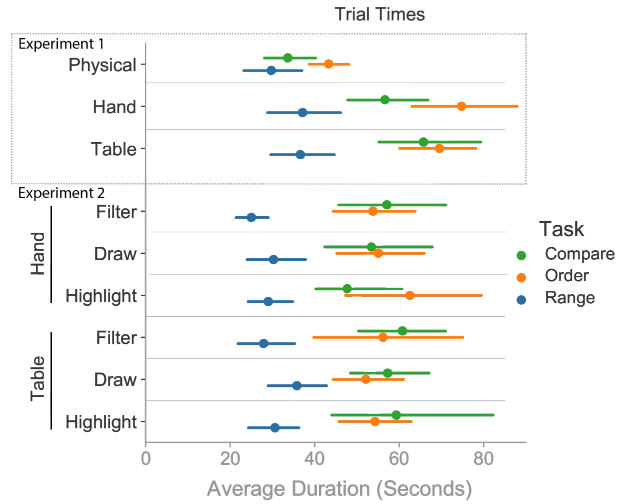

Fig. 10. Time-on-task for each scale (hand, table) × tool (filter, draw, highlight) × task (compare, order, range) combination. Results from Experiment 1 are shown for context (top). Error bars show 95% CIs.

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 to 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. In total we recorded 216 trials: 12 users × 3 tools × 2 scales × 3 tasks.

6.3 Results

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

Time on Task. As in the previous study, *range* tasks were markedly faster than the *order* or *compare* tasks (Figure 10). However, we saw little discernible difference between the six conditions, and neither drawing or highlighting seems to lead to a systematic increase or decrease in task time. Overall, results for the majority of conditions and tasks tended to be faster than those from participants in Experiment 1, suggesting that the addition of these new tools did not distract from or otherwise complicate the tasks.

Survey Results. Responses from the post-study survey showed that a majority of participants (9 of 12) preferred the virtual table-scale visualization over the virtual 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

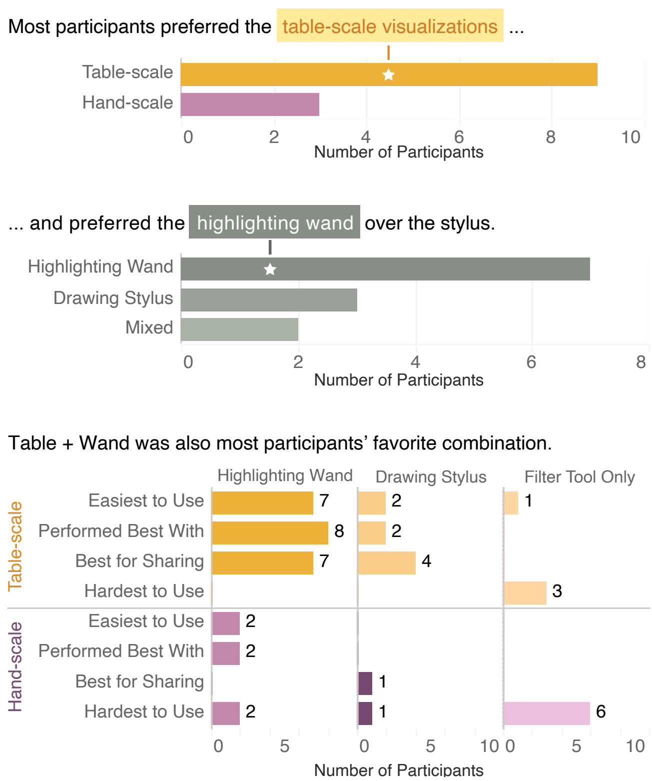

Fig. 11. **Experiment 2.** Post-study survey results by scale and tool.

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 handscale 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.

6.4 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 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 Experiment 1 (as well as 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 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 directly 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, virtual 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 either the physical charts or the larger table-sized virtual ones.

The poor performance and mixed responses to the bigger room-scale visualizations in our pilot also reflect the underlying challenges associated with exploring and manipulating visualizations at large scales and over physical distances [38]. Furthermore, the larger depth and height as well as diminished reachability of visualizations at this scale also limit the annotation, filtering, and data manipulation tools that can be used with them. However, VR and AR visualizations that support transitions between multiple scales [13] may have the potential to balance the trade-offs between both large- and small-scale approaches.

7.1 VR vs. Physicalization

In comparison to their physical counterparts, participants generally performed more quickly when using physical charts. 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 onscreen 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 [15]. 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 [39]. 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 fatigue [40].

Finally, our results echo Jansen et al.'s initial results which suggest that support for direct touch and physical manipulation were likely the biggest advantage of their physical prototypes [3]. Haptic displays or shape-changing interfaces capable of replicating this tactility for VR and AR visualizations remain a very distant prospect. While realistic real-time visual inspection using head tracking is already possible in VR, physical and tactile interactions with hand-scale visualizations remain less convincing. For example, current VR hand-tracking and haptic systems still lack the feedback necessary to support the active bi-manual manipulation of complex objects. We think this disparity may at least partially explain participants' strong preference for table-scale views (where VR tools can already provide a more satisfying approximation of their real-world counterparts) over hand-scale views where the gap remains wider.

However, 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 seen with physicalizations is promising. These findings suggest that tools and interactions which enable viewers to inspect, manipulate, and externalize their thought process visually on top of VR and AR visualizations could eventually 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 [12], are also encouraging. Ultimately, static 3D barcharts like those we used are unlikely to outperform either 2D or 3D visualizations that include interactive features like dynamic filtering or tooltips. However future physical, virtual, and augmented reality visualization tools have the potential to close that gap by incorporating more dynamic and interactive controls.

VR and AR visualizations also allow for interactions that are much more difficult, or impossible with physicalizations. For example, filtering and data reorganization are trivial in a virtual context, in stark contrast to the extensive physical implementation required for systems like Jacques Bertin's reorderable physical matrices [41]. Physical system like Microsoft's Tenison road charts, which represented community data using live pie charts [42], also often require complicated mechanical components to represent changes in data. Virtual representations of these same systems are comparatively much simpler to realize.

8 Conclusion & Future Work

Our results highlight several advantages that physicalizations can offer over their virtual counterparts, and emphasizes the potential value of tangible manipulation for simple chart reading tasks. Within the space of virtual charts we find clear advantages for virtual table-scale 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 XR visualization are still emerging fields. Consequently, considerable additional work is needed to understand the advantages and disadvantages of these techniques for more complex and realistic visualizations, interaction techniques, tasks, and scales. With that in mind, VR visualization tools represent an opportunity to leverage many of the benefits of the physical world, while also transcending its limitations.

Acknowledgments

The authors wish to thank our participants as well as the members of the Data Experience Lab for their help constructing the Lego charts. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN2016-04564] and a Mitacs Globalink Internship.

References

[1] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, and E. Wilson, "Immersive Analytics," in *2015 Big Data Visual Analytics, BDVA 2015*, 2015.

[2] Y. Jansen, P. Dragicevic, P. Isenberg, J. Alexander, A. Karnik, J. Kildal, S. Subramanian, and K. Hornbæk, "Opportunities and Challenges for Data Physicalization," *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, 2015.

[3] Y. Jansen, P. Dragicevic, and J.-D. Fekete, "Evaluating the Efficiency of Physical Visualizations," *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, p. 2593, 2013. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2470654.2481359

[4] S. Stusak, J. Schwarz, and A. Butz, "Evaluating the Memorability of Physical Visualizations," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, 2015.

[5] I. E. Sutherland, "A Head-Mounted Three Dimensional Display," *Proceedings of the Fall Joint Computer Conference*, pp. 757–764, 1968.

[6] 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," *2014 IEEE International Conference on Big Data Immersive*, pp. 609–614, 2014.

[7] E. Olshannikova, A. Ometov, Y. Koucheryavy, and T. Olsson, "Visualizing Big Data with Augmented and Virtual Reality: Challenges and Research Agenda," *Journal of Big Data*, vol. 2, no. 22, 2015.

[8] K. W. Arthur, K. S. Booth, and C. Ware, "Evaluating 3D Task Performance for Fish Tank Virtual Worlds," *ACM Transactions on Information Systems*, vol. 11, no. 3, pp. 239–265, 1993.

[9] Ç. Demiralp, C. D. Jackson, D. B. Karelitz, S. Zhang, and D. H. Laidlaw, "CAVE and Fishtank Virtual-Reality Displays: A Qualitative and Quantitative Comparison," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 3, pp. 323–330, 2006.

[10] M. Cordeil, A. Cunningham, T. Dwyer, B. H. Thomas, and K. Marriott, "ImAxes: Immersive Axes As Embodied Affordances for Interactive Multivariate Data Visualisation," *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pp. 71–83, 2017.

[11] O. H. Kwon, C. Muelder, K. Lee, and K. L. Ma, "A Study of Layout, Rendering, and Interaction Methods for Immersive Graph Visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 7, pp. 1802–1815, 2016.

[12] M. Cordeil, T. Dwyer, K. Klein, B. Laha, K. Marriott, and B. H. Thomas, "Immersive Collaborative Analysis of Network Connectivity: CAVE-style or Head-Mounted Display?" *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 441–450, 2017.

[13] A. Ivanov, K. Danyluk, C. Jacob, and W. Willett, "A Walk Among the Data: Exploration and Anthropomorphism in Immersive Unit Visualizations," *IEEE Computer Graphics and Applications*, vol. 39, no. 3, pp. 19–28, May 2019.

[14] Y. Wang, A. Segal, R. Klatzky, D. F. Keefe, P. Isenberg, J. Hurtienne, E. Hornecker, T. Dwyer, and S. Barrass, "An Emotional Response to the Value of Visualization," *IEEE Computer Graphics and Applications*, vol. 39, no. 5, pp. 8–17, Sep. 2019, conference Name: IEEE Computer Graphics and Applications.

[15] F. Berard and T. Louis, "The Object Inside: Assessing 3D Examination with a Spherical Handheld Perspective-Corrected Display," *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 4396–4404, 2017.

[16] Y. Jansen and K. Hornbæk, "A Psychophysical Investigation of Size as a Physical Variable," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 479–488, Jan. 2016.

[17] K. Sauvé, D. Potts, J. Alexander, and S. Houben, "A Change of Perspective: How User Orientation Influences the Perception of Physicalizations," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, ser. CHI '20. Honolulu, HI, USA: Association for Computing Machinery, Apr. 2020, pp. 1–12. [Online]. Available: https://doi.org/10.1145/3313831.3376312

[18] T. Hogan and E. Hornecker, "How Does Representation Modality Affect User-Experience of Data Artifacts?" *Proceedings of Haptic and Audio Interaction Design*, pp. 141–151, 2012.

[19] F. Taher, J. Hardy, A. Karnik, C. Weichel, Y. Jansen, K. Hornbæk, and J. Alexander, "Exploring Interactions with Physically Dynamic Bar Charts," *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 3237–3246, 2015.

[20] S. Follmer, D. Leithinger, A. Olwal, A. Hogge, and H. Ishii, "inFORM: Dynamic Physical Affordances and Constraints through Shape and Object Actuation," *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pp. 417–426, 2013.

[21] "Gapminder Foundation." [Online]. Available: https://www.gapminder.org/data

[22] J. Liu, A. Prouzeau, B. Ens, and T. Dwyer, "Design and Evaluation of Interactive Small Multiples Data Visualisation in Immersive Spaces," in *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. Atlanta, Georgia, USA: IEEE, 2020.

[23] D. Leithinger and H. Ishii, "Relief: A Scalable Actuated Shape Display," *Proceedings of the ACM conference on Tangible, Embedded, and Embodied Interaction (TEI)*, pp. 221–222, 2010.

[24] S. Tang, Y. Sekikawa, D. Leithinger, S. Follmer, and H. Ishii, "Tangible CityScape," 2013. [Online]. Available: http://tangible.media.mit.edu/project/tangible-cityscape

[25] R. Burdett, "Global Cities: Elevation Maps of City Population," 2007. [Online]. Available: http://dataphys.org/list/city-populations/

[26] Exhibition Design: Atelier Brückner and Photography: Michael Jungblut, "Living the Future – The Demographic Opportunity," 2013. [Online]. Available: https://www.atelier-brueckner.com/en/projects/living-future-demographic-opportunity

[27] D. Goods, N. Hafermaas, and A. Koblin, "eCloud," 2007. [Online]. Available: http://www.ecloudproject.com/

[28] J. Barlow, D. Goods, and N. Hafermaas, "airFIELD – Atlanta International Airport," 2012. [Online]. Available: http://airfieldsculpture.com/

[29] R. P. Theart, B. Loos, and T. R. Niesler, "Virtual Reality Assisted Microscopy Data Visualization and Colocalization Analysis," *BMC Bioinformatics*, vol. 18, no. 64, 2017.

[30] S. Han, B. Liu, R. Wang, Y. Ye, C. D. Twigg, and K. Kin, "Online optical marker-based hand tracking with deep labels," *ACM Transactions on Graphics*, vol. 37, no. 4, pp. 166:1–166:10, Jul. 2018. [Online]. Available: https://doi.org/10.1145/3197517.3201399

[31] E. Gusai, C. Bassano, F. Solari, and M. Chessa, "Interaction in an Immersive Collaborative Virtual Reality Environment: A Comparison Between Leap Motion and HTC Controllers," in *New Trends in Image Analysis and Processing – ICIAP. Springer Lecture Notes in Computer Science*, vol. 10590, 2017.

[32] *Publication Manual of the American Psychological Association*, Sixth ed. Washington, D.C.: American Psychological Association, 2017.

[33] G. Cumming, "The New Statistics: Why and How," *Psychological Science*, vol. 25, no. 1, pp. 7–29, 2014.

[34] P. Dragicevic, F. Chevalier, and S. Huot, "Running an HCI Experiment in Multiple Parallel Universes," in *ACM CHI 2014 Extended Abstracts*. ACM, 2014, pp. 607–618.

[35] R. Stoakley, M. J. Conway, and R. Pausch, "Virtual Reality on a WIM: Interactive Worlds in Miniature," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, ser. CHI '95. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1995, pp. 265–272. [Online]. Available: http://dx.doi.org/10.1145/223904.223938

[36] Y. Guiard, "Asymmetric Division of Labor in Human Skilled Bimanual Action," *Journal of Motor Behavior*, vol. 19, no. 4, pp. 486–517, Dec. 1987. [Online]. Available: https://doi.org/10.1080/00222895.1987.10735426

[37] Google, "Tilt Brush by Google." [Online]. Available: https://www.tiltbrush.com/

[38] C. Andrews, A. Endert, B. Yost, and C. North, "Information Visualization on Large, High-Resolution Displays: Issues, Challenges, and Opportunities," *Information Visualization*, vol. 10, no. 4, pp. 341–355, 2011.

[39] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence–Accommodation Conflicts Hinder Visual Performance and Cause Visual Fatigue," *Journal of Vision*, vol. 8, no. 33, 2008.

[40] G. Kramida, "Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 7, 2016.

[41] C. Perin, J.-D. Fekete, and P. Dragicevic, "Jacques Bertin's Legacy in Information Visualization and the Reorderable Matrix," *Cartography and Geographic Information Science*, vol. 46, no. 2, pp. 176–181, 2019.

[42] S. E. Lindley, A. Thieme, A. S. Taylor, V. Vlachokyriakos, T. Regan, and D. Sweeney, "Surfacing Small Worlds through Data-In-Place," *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)*, vol. 26, no. 1, pp. 135–163, Apr 2017.

Kurtis Danyluk is a Computer Science graduate student with the University of Calgary, actively exploring camera control techniques for digital maps and virtual reality applications. His interests span human-computer interactions, and information visualization. He is a member of the Interactions Lab.

Teoman Ulusoy graduated from the University of Calgary with a BSc in Computer science, concentrating in HCI. His central research interests include human-computer interaction, data visualization, and virtual reality interactions. He is currently an instructor in the IT Software Developer program at the Southern Alberta Institute of Technology.

Wei Wei is an undergraduate student at Southwest Jiaotong University in Sichuan, China majoring in Computer Science. In 2019 he completed a Mitacs Globalink research internship at the University of Calgary. He is interested in human-computer interaction, data visualization and human-robot interaction.

Wesley Willett is an Assistant Professor of computer science with the University of Calgary where he holds a Canada Research Chair in Visual Analytics. He leads the Data Experience Lab, an interdisciplinary research group whose research spans information visualization, human-computer interaction, and data science and focuses on pairing data and interactivity to support collaboration, learning, and discovery.