Evaluation of a Financial Portfolio Visualization using Computer Displays and Mixed Reality Devices with Domain Experts

Kay Schroeder
HDILab, Zuyd University of Applied Sciences
Heerlen, Netherlands
kay.schroder@zuyd.nl

Batoul Ajdadilish
Zuyd University of Applied Sciences
Heerlen, Netherlands
betty.ajdadilish@zuyd.nl

Alexander P. Henkel
Center for Actionable Research of the Open University (CAROU), Open University
Heerlen, Netherlands
alexander.henkel@ou.nl

André Calero Valdez
HCIC, RWTH Aachen University
Aachen, Germany
calero-valdez@comm.rwth-aachen.de

ABSTRACT
With the advent of mixed reality devices such as the Microsoft HoloLens, developers have been faced with the challenge to utilize the third dimension in information visualization effectively. Research on stereoscopic devices has shown that three-dimensional representation can improve accuracy in specific tasks (e.g., network visualization). Yet, so far the field has remained mute on the underlying mechanism. Our study systematically investigates the differences in user perception between a regular monitor and a mixed reality device. In a real-life within-subject experiment in the field with twenty-eight investment bankers, we assessed subjective and objective task performance with two- and three-dimensional systems, respectively. We tested accuracy with regard to position, size, and color using single and combined tasks. Our results do not show a significant difference in accuracy between mixed-reality and standard 2D monitor visualizations.

Author Keywords
information visualization; mixed reality displays; HoloLens; user study; UX study.

INTRODUCTION
Financial services, particularly investment banking, is a high-stakes industry. Highly trained humans need to apprehend and process masses of data in split seconds and make far-reaching investment decisions. To make these individual investment decisions, financial professionals are supported by technology. However, so far, the visualization technology deployed in financial services merely provides a limited benefit to professionals, as the important underlying relationships are high dimensional and require novel interaction techniques.

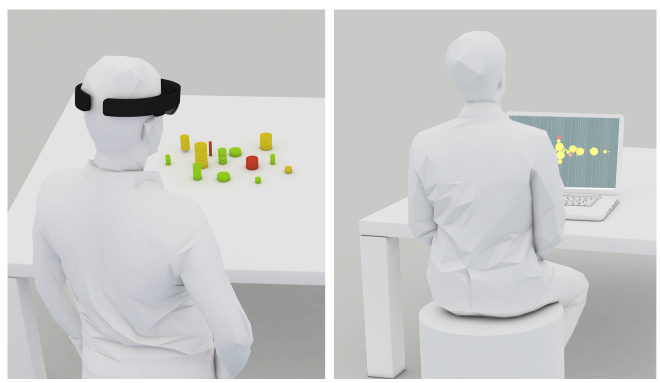

Figure 1. Experiment: Mixed reality and conventional display

The proliferation of mixed reality devices has given rise to questions among researchers and practitioners regarding how these technologies can be utilized to interact with information in a spatial context [21]. Prior research has shown that the three-dimensional visualization of data can outperform two-dimensional representations in terms of accuracy and performance under specific conditions, especially when the information has spatial characteristics. However, there is a long-standing debate within the research community about the

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

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376556

added value of three-dimensional visualizations of quantitative data [25, 21, 18].

Early [6], and recent [12] studies have shown that a spatial position is the most effective way to map information within a 2D space. However, the extension of the two-dimensional space with a third dimension used to plot additional data attributes has proven to be problematic on two-dimensional displays (in the literature also described as 2.5D) [21, 18].

One of the primary reasons for this drawback according to perception research, is the lack of depth cues, which are playing a major role in interpreting three dimensional (3D) objects and visualizations. The presence of these cues not only provides a complete and more perceptible representation of spatial information, but it can also increase the performance and accuracy of 3D visualizations [26]. Technical advances in recent years have greatly improved the quality of stereoscopic displays with the effect that prior findings may no longer apply.

In 1999, in his survey, Tegarden [22] briefly discussed the issues of applying data visualization to business problem solving. While giving an overview of information visualization and available technologies at that time, he highlights that these technologies potentially enhance the task of decision making in the business domain/financial sector. To provide more evidence, he suggests that more research should be conducted to address business problems by means of corresponding data visualization.

After 20 years, the development and availability of virtual reality glasses and HoloLens technologies is drawing attention of researchers across disciplines to understand the potential benefits of applying data visualization on inherently complex financial data. Applied visual analytics in financial decision processes have been discussed extensively in the article by Savikhin [20]. He suggests that interactive visual analytic tools can play an appropriate role as a decision support system. He proposes to use experimental methodology to investigate the effectiveness of visual analytics in financial decision-making tasks and risk management.

Today—almost 20 years since the emergence of 3D presentation of data on computer screens—the debate about the use of 2D vs. 3D is still going on, now from different perspectives across different disciplines [7]. A recent survey by Ko et al. [14] shows that when it comes to financial data analysis, using 3D visualization is preferred compared to 2D visualization. And now, the recent resurgence of (vastly improved) display technologies takes the debate to the next level. Researchers try to understand and compare the standard 3D visualization on regular screens (2.5D) with 3D visualization empowered by virtual and augmented reality technologies.

In their position paper *Immersive Analytics*, Chandler et al. [4] challenge researchers to understand the usability and design issues of new interfaces and display technologies and how these technologies can be used to provide more immersive data analysis.

Our Contribution

In response to the call for further research on *Immersive Analytics* [1], and motivated by emerging technologies, our study is an attempt to understand the effectiveness of immersive analytic applications in financial data decision-making. In this paper, we discuss the step-wise development of the application, the related user-centered design process as well as a within-subject comparison. We did not find significant difference regarding the accuracy between the 2D and 3D devices, but users perceived the use of the application slightly more enjoyable in mixed reality. Whilst we developed the tool mainly for application in financial asset management, it can be applied in any visual analytics setting with a comparable data structure.

RELATED WORK

Our research contributes to the question whether the financial industry can benefit from deploying immersive analytics into their daily based decision-making. To the best of our knowledge, there is a gap in the published literature regarding the application of immersive analytics on financial data decision-making. But, does it even make sense to approach financial data visualization using AR?

Research in the financial sector

A comprehensive overview of visualization techniques and visual analytic tools that have been applied to financial data was provided by Ko et al. [14]. In their extensive survey, they found that despite the availability of several visualization systems, financial market participants tend to apply conventional visualization techniques in their routine tasks. They articulate one possible explanation that there is not much collaboration between researchers and market participants due to **privacy issues**. On one hand it is not easy for researchers to get access to real-world financial data, on the other hand, companies are not willing to potentially expose their data to competitors. From this aspect, our research is unique and it has been conducted through a close collaboration with a business partner with real-world data. We provide an experimental comparison between a 2D desktop visualization/application and a 3D immersive/augmented visualization using Microsoft Hololens.

Benefits of 3D Visualizations

An early insight comparing data exploration using 2D vs. VR has been provided by Millais et al. [17]. They designed two VR visualizations parallel two 2D representation of the same data from a self-tracking study (including: mood, productivity, sleep, music listening, and physical activity). After introducing the 'think-aloud protocol'and no time limit on exploration to the participants, they evaluated the users' perceived workload using the NASA TLX Questionnaire and compared the results with its counterpart 2D visualizations. By applying an insight-based evaluation methodology, they found that the participants were **more successful and satisfied** to explore data using VR. Interestingly, the participants reported more accurate insight when using VR visualization.

In their work, Lugmayr et al. [15] give an overall overview about various projects exploring the capability of 3D visualization to enhance effective decision making in the financial data (Australia's Energy sector). By developing a prototype, adding a third dimension, they conclude that immersive virtual environments (may) provide support for cognitive functions and may make the data **more understandable**.

Butcher et al. [3] developed a framework for immersive analytics on the web. For their preliminary evaluation, they used a 3D bar chart analytical task with non-expert participants. They report that although VR visualization are **more engaging**, on average it takes more time to complete the task.

Benefits of 2D Visualizations

Smallman et al. [21] conducted an interesting experiment and claimed that when it comes to comparing 2D vs. 3D visualization, information availability and representation plays a larger role than display format. Challenging the previous studies on this topic, they designed an experiment for visual search tasks (i.e., air-traffic control console). By controlling the different information coding schemes, they found that the 2D display format provides **faster access** to the information. They claimed that the benefits of utilizing the third dimension can be easily surpassed in a well-designed 2D display format. They suggest for tasks which deal with **exact spatial judgments**, it is better to use 2D visualization.

Combinations of 2D and 3D

Following up on their previous work on cognitive difficulties, Tory et al. [23] set up a series of experiments to compare task performance in 2D, 3D and Combined 2D/3D displays—a display that provides at least one 2D view and one 3D view at the same time. Over the three experiments, 40 students from computer science and the engineering domain were exposed to the generalised tasks such as estimating position, relative orientation, and volume of interest tasks. The idea behind the chosen tasks were that they were expected to benefit from both displays, 2D and 3D. They found that whereas strict 3D displays were effective for relative position estimation and orientation, the combined 2D/3D displays were **better for precise orientation** and positioning tasks. To measure learning effects on participants and to have a better control over variation of spatial ability of individuals, a within-subject study was suggested.

Immersive Analytics and Mixed-Reality

To address the issue of handling depth information in urban visual analytics, Chen et al. [5] came up with a novel solution using HoloLens. They developed an optimization algorithm in response to some general principles of creating exploded views, such as transition, intra-layer, and inter-layer. Although the model was not suitable for a dynamic user's perspective (different viewpoints), the result of comparing this method with the two other well-known methods (in the domain of) for exploded views, have gained a **more promising performance** for immersive urban analytics.

In contrast with the major research trend on immersive analytics, Zielasko et al. [32] prototyped an application to evaluate the usability of the integration of new display technologies into conventional desktop workflows. They believe by simulation of user's desktop while providing hands-free interaction, they **maximize the feeling of immersion**. They also proposed a method to prevent the major issue of VR display—cybersickness.

METHOD

To identify how the display technology affects decision making in real-life financial settings, we developed two visualization prototypes. The requirement for the visualization approaches were collected in qualitative semi-structured interviews, trying to identify the aims and scope of the visualization within the related working process from a user perspective, the data supplier perspective as well as the characteristics of the underlying data.

In our case the data supplier is a pension fund managing investments in large portfolios. The central aim of visualization was to support decision making in identifying stocks that are vulnerable to changes in climate policies. For this purpose stock portfolios were to be analyzed with regard to current CO_2 output, relative changes in carbon emissions compared to the last year and weight in the overall portfolio.

Data was made available from the company with the additional requirement to anonymize stocks to ensure that during the evaluation additional knowledge of the participants—who were all portfolio managers—did not substantially influence decision making. As a means for anonymization we used pseudonyms to replace actual stock names and replaced numerical values with values from representative distributions.

First, we developed the two-dimensional prototype within a user-centered design process (UCD). In a step-wise, iterative process we tested and optimized different visual encodings and evaluated them together with portfolio managers within a realistic scenario. To avoid an influence on test subjects of the final experiments, we realised the UCD process with portfolio managers from another subsidiary and no direct contact to the final test group but with the same demographic background and comparable working experience.

To develop and test the three dimensional approach in an early stage we started working with physical representations of the data. Within the first iteration we started with wooden blocks of basic shapes to evaluate their effectiveness within different scenarios. During several iterations we found out that a stacked tubes as visual encoding of the given data led to positive feedback.

Visualization design

The resulting D3-based visualization is a combination of a bubble chart and a bee-swarm plot (see Fig. 3 and Fig. 5). The position on the horizontal axis represents the variable value, while the vertical axis shows the distribution of the data. This visualization was chosen as it addresses both of the biggest challenges in the domain task. Analysts need to **identify** specific individual shares by multiple dimensions, while at the same time relate them to the **overall distribution** of the portfolio. Early evaluations in our user-centered design process showed that many classical visualization types (e.g., histograms, scatterplots) lead to significant problems in selecting individual shares in the visualization. This was caused by the resulting small size of the visual marks, which led to

Figure 2. User centered design process with wooden blocks

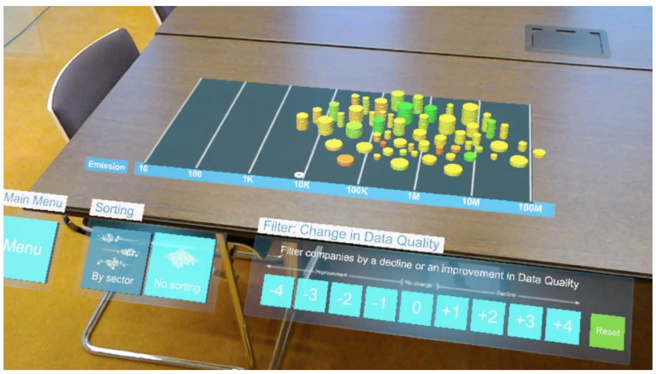

Figure 3. Visualization as seen in the Microsoft HoloLens

the idea of bubble charts and bee-swarm plots. Our approach shows the overall distribution while also providing individual marks of sufficient size without overplotting.

The following performance indicators from the data were employed in order of importance: portfolio weight was encoded in the bubble size (i.e., radius); amount of total carbon emitted was encoded in the horizontal position; relative changes in carbon emissions within the last year was encoded as color (red = increase, green = decrease). The color was selected from a range that was predetermined by the corporate identity and thus sadly not perceptually ideal for decision tasks. Furthermore, as the code of the visualization is proprietary, we have decided to release all the necessary spatial positions and color information of the visualization in the supplemental materials. This allows other researchers to replicate our study with a different framework.

Procedure of the experiment

Our experiment followed a one-factorial within-subjects design with two conditions: Microsoft HoloLens Version 1 (3D) and a predefined dataset vs. a MacBook Pro (2D) with an additional predefined dataset. A within-subject design was chosen to allow for a small sample size, as the availability of participants in our case is limited by the department size involved in the study. Both datasets had the same level of complexity but different values.

In a first step, users were introduced to the procedure of the experiment, how to use the specific device, the visual encoding as well as assignment and questions. All participants conducted both a 2D (see Fig. 4) and a 3D (see Fig. 3) task. To ensure that in the second task data had to be reevaluated, two different portfolios where shown to each participant. Both, order of the device and order of portfolios were counterbalanced across participants in randomized order to prevent ordering effects. No or only very weak effects of ordering are found in the data (see suppl. material). Each participant thus took part in 14 (2×7) tasks in total, which were grouped in 3 categories: A, B, and C.

Figure 4. Visualization as seen on a regular screen (Portfolio 17 on top, 25 at the bottom)

- **A** - Three tasks measured decision quality on *one* individual output dimensions (i.e., horizontal position, radius, and color).
- **B** - Two tasks measured conjoint decisions tasks based on *two* dimensions (i.e., horizontal position $\times$ radius and radius $\times$ color).
- **C** - One task measured the conjoint decision task on all *three* dimensions at the same time.

Initially, an additional conjoint decision task measuring decision quality on all four dimensions was included. However, as the visual dimensions were not comparable in this case (i.e., filter vs. height), we omit this task from further analysis, yet report it here for completeness.

After each task-group (i.e., A, B, C), participants were asked to complete subjective performance assessments using the respective tool. Subjective performance was measured as perceived task completion speed (two items assessing whether using the tool for the task was fast / time-consuming), perceived precision (two items assessing whether using the tool for the task was precise / accurate), and perceived usefulness (two items assessing whether using the tool for the task was useful / not helpful). We also assessed perceived effort put into the task with two items (I tried very hard at this task; I put a lot of effort into this task) [13]. All items were assessed on a 7-point scale from *Strongly Disagree* to *Strongly Agree*.

Upon completion of all tasks, participants were also asked to indicate their task enjoyment with the respective tool. Following prior research, we assessed task enjoyment with an intrinsic motivation index (e.g., [11]). The index comprised five items on whether participants perceived the task as fun / boring / interesting / absorbing / enjoyable, which were assessed on 7-point scales from 1 (*Strongly Disagree*) to 7 (*Strongly Agree*).

After all seven tasks for each device were completed we assessed six-item TAM-based acceptance scales for each device for both perceived ease of use (e.g., "Learning to operate this application would be easy for me") and perceived usefulness (e.g., "Using this application would make it easier to do my job"), respectively [9]. All items were assessed on a 7-point scale from *Extremely Unlikely* to *Extremely Likely*. This was used to compare subjective differences with objective differences in our study.

In addition, all experimental sessions were audio and video recorded. To minimize the effect of current limitations of the used technology, tasks were short to minimize discomfort and user position was fixed inside a bounding box to account for field-of-view-limitations. This minimal range of movement also allowed that the resolution of the mixed-reality device was high enough to ensure readability.

Stimulus Overview

Since data was not artificially created, it is necessary to determine what decision quality is. Decision quality was measured on relative visualization data. We did not assess root mean square error (RMSE) or a comparable metric on the real data, as too many aspects play a role in mapping the visualization to real world data. Instead we analyze the quality of decisions relative to the respective visualization dimension. This means we pick all horizontal, radius, and color positions and normalize them to a range between 0 and 1. For color we used the CIE 2000 color space to determine the distances between two different colors to map to real perceptual distances.

Using real-life data, this might lead to distributions that are sub-optimal for comparing performance on different devices. Ideally, normally distributed data would be used. To see whether our data matches this, we first look into the distribution of the individual tasks and variables to investigate whether all tasks are 'fair' and comparable between devices and settings.

To clarify, in 2D the pixels presented on a screen provide a hard limit for the maximal resolution of data. In the HoloLens on the other hand, users may move around and thus increase the resolution ad infinitum. This holds true for both radius and horizontal position. Furthermore, color ranges between input devices differ perceptually. To account for this we look at relative values within each individual portfolio and task device. Differences in distributions are unavoidable. We try to

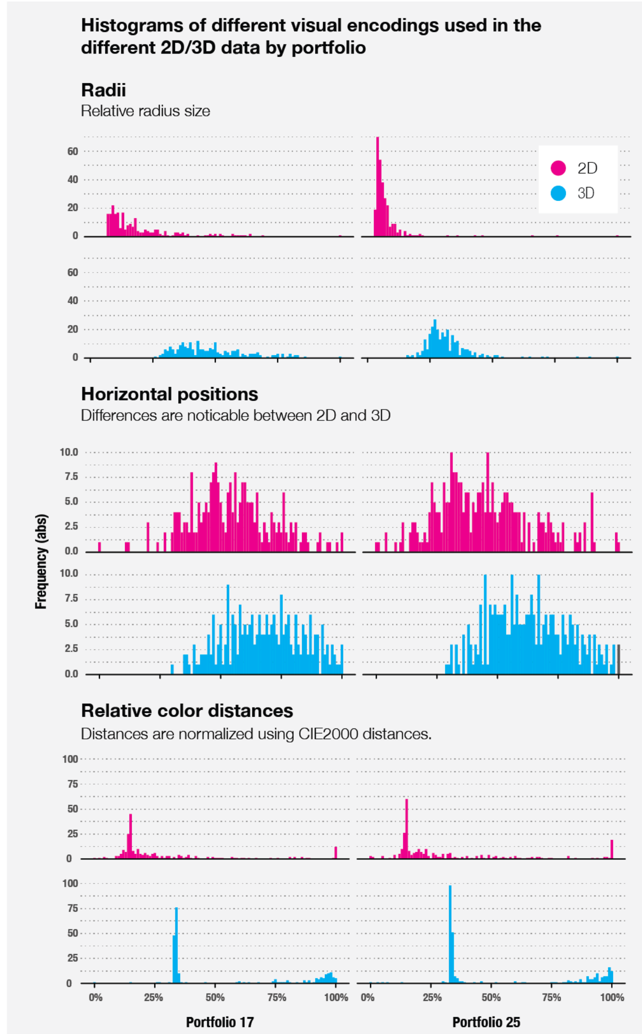

Figure 5. Comparison of different visual encodings between devices and portfolios

account for this by asking multiple answers and normalizing against the optimal solution (e.g., three best choices) in each setting.

When comparing radii between 2D and 3D, we see the hard limit of the pixel based resolution on the 2D screen. A circle in the visualization will never be smaller than 2 pixels, while the circles in the HoloLens may be smaller than 2 pixels from the standard viewing range. As the user may move around a little, smaller circles become more visible (see Fig. 5). Horizontal positions are mostly normally distributed in both portfolios and both display settings (see Fig. 5).

When it comes to evaluating the different colors and their impact on visualization, we used CIE 2000 to calculate the relative distance to the target stimulus. We then normalize the distances by dividing through the largest possible distance. As we are interested in performance (1 is best, 0 is worst), we subtract the relative distance from 1 to get a metric that measures how close the user is to the target (see Fig. 5).

Participants

Participants ($n = 28$) were recruited as employees from the asset management department of a Dutch pension fund. The mean age of our participants was 33.8 years ($SD = 7.7$) and they on average had 7.5 years of work experience ($SD = 7.2$). Gender was equally distributed among men and women. The average height was 180cm ($SD = 10$). All participants had completed a university masters degree and used the HoloLens for the first time. Two users had some form of color blindness (i.e., Deuteranomaly) and 16 of the participants required to have vision aids during the tasks.

Used Devices

We used a HoloLens (1. Gen) with 2.3 megapixel widescreen see-through holographic lenses with a weight of 579g for the three-dimensional visualization. Test subjects used the HoloLens clicker and gestures as interaction modes. The two-dimensional visualization was presented on a MacBook Pro with a 15inch Retina display and a resolution of 2880x1800 pixel. All interactions were carried out with the integrated trackpad.

Statistical Methods

In order to test whether differences exist between the individual factors with regard to both decision quality and subjective evaluation, we conduct analyses of variance to determine the effect of our independent variable on the dependent measures. We use a level of significance of $\alpha = .05$ for our null-hypothesis significance testing. We further use boot-strapped 95% confidence intervals in our plots for visual inspection of differences.

From the sample size, and an expected correlation of within subjects measurements of approx. $r = 0.7$, a α-error level of .05 and an expected power $(1-\beta)$ of .95, we may achieve a (post-hoc) sensitivity level that allows us to detect effects above a critical F-Value of $F > 4.23$ and an effect-size of $f > 0.63$. Any effects smaller than this are therefore considered as non-existent, although using a larger sample size smaller effects could potentially be discovered.

RESULTS

To understand how the visualization device impacted the performance of our participants we used two-way ANOVAs to determine the impact of the device and the chosen portfolio on the quality of the decision (see the additional materials for the full ANOVA tables).

Performance results

In the so called A tasks only performance regarding individual encodings were analyzed. When looking at the individual dimensions we can see that neither the different portfolios nor the devices show large differences in performance (see Fig. 6), when looking at radii ($R^2 = 0.09, F(3,51) = 1.77$, $p = 0.165$), horizontal positions ($R^2 = 0.21$, $F(3,51) = 4.53$, $p = 0.007$: here a minute difference between devices is statistically significant $p < .05$), and color distance ($R^2 = 0.01$, $F(3,51) = 0.13$, $p = 0.941$).

In the so-called B tasks, two encodings were evaluated at the same time with equal weight. Here, first differences become

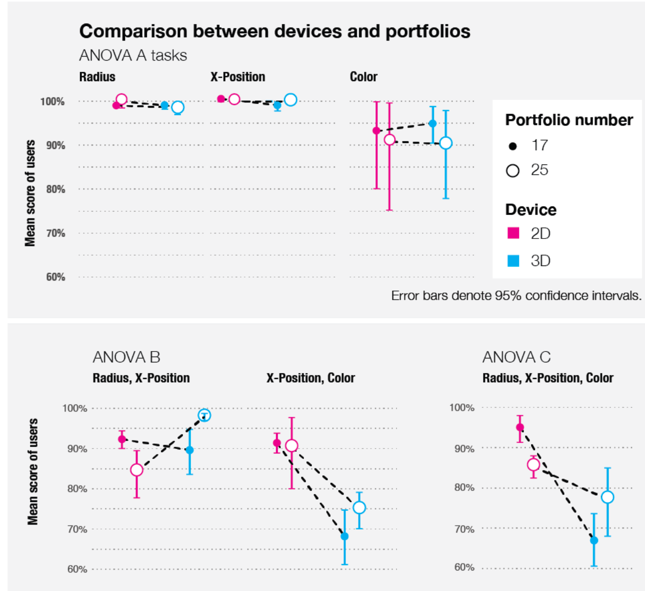

Figure 6. Visual comparison of the ANOVA tasks

apparent. When looking at radius and the horizontal position at the same time (see Fig. 6), 3D outperforms 2D visualizations ever so slightly ($R^2 = 0.25$, $F(3,51) = 5.81$, $p = 0.002$). However, when the horizontal position and the color are to be evaluated, 3D is trumped by 2D visualization ($R^2 = 0.42$, $F(3,51) = 12.12$, $p =< .001$). While in the first case, being able to move freely enhanced resolution, bringing color into the mix made users perform less well.

This last finding holds, when all three dimensions need to be evaluated at the same time. Picking the best options is achieved less well in 3D than in 2D (see Fig. 6 ANOVA C), irrespective of the portfolio used in the visualization ($R^2 = 0.47$, $F(3,51) = 15.2$, $p =< .001$).

Summarizing, we can say that possibly due to bad (or at least worse) color representation or fidelity, determining the best options in the HoloLens is achieved less well than through a regular (retina) Macbook screen, when multiple dimensions need to be evaluated. However, the differences are not very large.

Subjective Evaluations

To go beyond these metrics that are based on pure effectiveness, we asked participants about their user experience during the experiments. All participants deal with portfolio management on a day to day basis and are thus well aware of the requirements in comparing items on multiple dimensions at the same time.

Moreover, subjective measures—especially hedonic ones—play a role in everyday work. To capture this dimension, we asked the participants how absorbing, boring, enjoyable, fun, and interesting, working with this visualization was to them. Similarly to previous results, differences are not large (see Fig. 7), but follow an interesting trend. The HoloLens visualization is more fun ($R^2 = 0.47$, $F(3,51) = 15.2$, $p =< .001$),

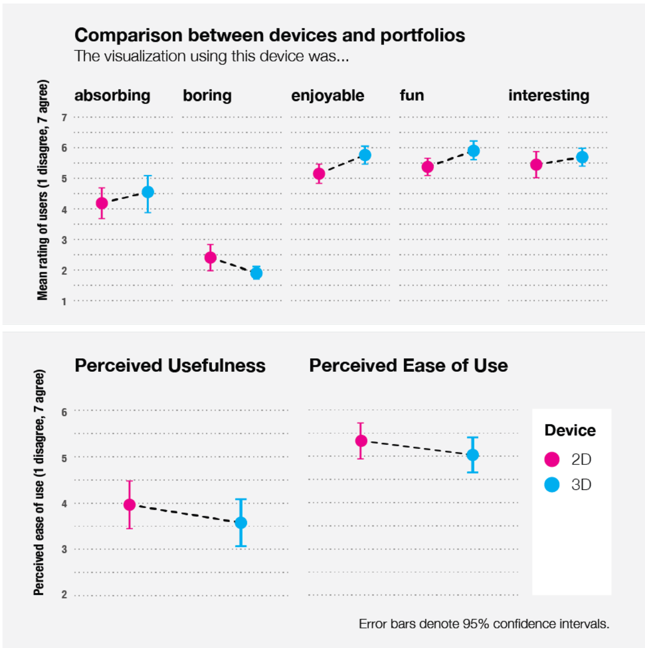

Figure 7. Comparison between devices and portfolios

equally boring ($R^2 = 0.47$, $F(3,51) = 15.2$, $p =< .001$), equally absorbing ($R^2 = 0.02$, $F(3,52) = 0.34$, $p = 0.798$), equally enjoyable ($R^2 = 0.11$, $F(3,52) = 2.07$, $p = 0.116$), and equally interesting ($R^2 = 0.02$, $F(3,52) = 0.41$, $p = 0.743$).

Technology acceptance

In line with the findings from the previous sections there is little difference between the perceived ease of use and the perceived usefulness in our sample. A slightly smaller perceived ease of use and usefulness (see Fig. 7) can be attested to the 3D visualization, albeit one that does not play a large role from the perspective of the general target audience.

DISCUSSION

Nowadays, portfolio managers face the challenge to evaluate increasingly complex relationships of different factors. One of the major drivers in this field is the need to consider new ambiguities in a highly dynamic environment (e.g., carbon footprint vs. financial benefits). Therefore, the major challenge is to evaluate multiple dimensions of a portfolio at the same time, as the experts in our user-centered design (UCD) process have indicated.

As traditional displays lack depth cues, which improve observational accuracy of visual dimensions [26], 3D visualizations become an increasingly viable option. Our UCD process yielded a working two-dimensional visualization and our initial attempts with wooden blocks indicated its suitability in three dimensions as well (see Figure 2). In our study, we set out to determine whether current technology can be utilized to adapt our visualization to three dimensions with sufficient accuracy, even when an additional dimension is used.

Interestingly, our users pointed out that a spatial representation might improve the ability to reason and discuss portfolios in teams. They liked that a mixed-reality based solution allowed them to communicate directly with the experimenter. As our results show no significant difference between the performance in a 2D-based visualization and a 3D mixed-reality visualization, it seems promising to conduct further research with mixed-reality visualizations that focus on the meaningful integration of the additional channel and the effective combination of 2D and 3D visualizations.

However, the used mixed-reality technology is not yet fully developed. The users complained about the weight as well the limited field of view during the experiment. We assume that this influences the user experience as well as the task effectiveness. Furthermore, the color rendering in the displays may be different (less fidelity in the Holo Lens must be assumed, thus the true differences are probably lower). But this is caused by the quality of the current device. . The small differences are—to our belief—explainable from the different color resolutions of the devices. Further research, e.g., using just-noticeable difference setups should be incorporated to measure precise perceptual differences between the two display technologies from a modern perspective.

Future generations of such devices are likely to overcome some of these challenges. Moreover, the users also mentioned that increased portability of a mixed-reality device in the future could allow working in more mobile settings without sacrificing screen real estate. During the UCD process we realized that the natural interaction, such as gestures, pointing, and moving objects, was perceived as much more intuitive and convenient for discussions. Integrating both 2D visualizations with 3D representations within mixed-reality could enable portfolio managers to cope with the increasing need to consider multiple dimensions at the same time.

CONCLUSION

This work represents a first step towards a better understanding of task performance and user satisfaction while using two dimensional and three dimensional displays. In our study we developed a first prototype that was optimized to fulfil a specific task within a financial decision making context.

Our data suggests that mixed reality devices may not perform significantly less accurate than standard 2D monitor visualizations. Yet, increasing complexity resulted in a slightly negative effect regarding the accuracy. Users experienced the three-dimensional representation in general as more enjoyable. Further research is required to determine in how far these findings are generalizable to other tasks, visual encodings and data types.

ACKNOWLEDGMENTS

We would like to thank the reviewers for their extensive feedback that substantially improved our article. We would also like to thank Lilian Kojan for her support in transcribing paper based-assessments to digital data. Thank you to Johannes Nakayama, Nils Plettenberg, Stefano Bromuri, Visara Urovi, Ellen Pijpers, Kenny Jeurissen for additional feedback on earlier versions of this manuscript. Thank you to Rolf Habets, Joost Slabbekoorn, Auke Hoevenberg, René Rateischak, Mart Fijten, Rick de Boer, Marielle Heuts, and Sten Martens.

We would like to thank the developers of the following r-packages as their open tools were used to create this article and the accompanying open-data website (knitr [29], tidyverse [27], logger [8], spacesXYZ [10], rmdformats [2], DT [30], kableExtra [31], scales [28], psych [19], CHI20Rmarkdown [16], rmdtemplates [24]).

REFERENCES

[1] Benjamin Bach, Raimund Dachselt, Sheelagh Carpendale, Tim Dwyer, Christopher Collins, and Bongshin Lee. 2016. Immersive Analytics: Exploring Future Interaction and Visualization Technologies for Data Analytics. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces (ISS '16)*. Association for Computing Machinery, New York, NY, USA, 529–533. DOI: http://dx.doi.org/10.1145/2992154.2996365

[2] Julien Barnier. 2019. *rmdformats: HTML Output Formats and Templates for 'rmarkdown' Documents*. https://CRAN.R-project.org/package=rmdformats R package version 0.3.5.

[3] Peter W. S. Butcher, Nigel W. John, and Panagiotis D. Ritsos. 2019. VRIA - A Framework for Immersive Analytics on the Web. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, USA, Article LBW2615, 6 pages. DOI: http://dx.doi.org/10.1145/3290607.3312798

[4] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, and E. Wilson. 2015. Immersive Analytics. In *2015 Big Data Visual Analytics (BDVA)*. IEEE, Hobart, TAS, Australia, 1–8. DOI: http://dx.doi.org/10.1109/BDVA.2015.7314296

[5] Zhutian Chen, Yifang Wang, Tianchen Sun, Xiang Gao, Wei Chen, Zhigeng Pan, Huamin Qu, and Yingcai Wu. 2017. Exploring the design space of immersive urban analytics. *Visual Informatics* 1, 2 (2017), 132 – 142. DOI: http://dx.doi.org/10.1016/j.visinf.2017.11.002

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

[7] Grégoire Cliquet, Matthieu Perreira, Fabien Picarougne, Yannick Prié, and Toinon Vigier. 2017. Towards HMD-based Immersive Analytics. In *Immersive analytics Workshop*. IEEE VIS 2017, Phoenix, AZ, USA, hal–01631306.

[8] Gergely Daróczi. 2019. *logger: A Lightweight, Modern and Flexible Logging Utility*. https://CRAN.R-project.org/package=logger R package version 0.1.

[9] Fred D Davis. 1985. *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. Ph.D. Dissertation. Massachusetts Institute of Technology.

[10] Glenn Davis. 2018. *spacesXYZ: CIE XYZ and some of Its Derived Color Spaces*. https://CRAN.R-project.org/package=spacesXYZ R package version 1.0-4.

[11] Andrew J Elliot and Judith M Harackiewicz. 1996. Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of personality and social psychology* 70, 3 (1996), 461.

[12] Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, ACM, Atlanta, Georgia, USA, 203–212.

[13] Debra Steele Johnson, Daniel B Turban, Kalen F Pieper, and Yin Mei Ng. 1996. Exploring the Role of Normative-and Performance-Based Feedback in Motivational Processes 1. *Journal of Applied Social Psychology* 26, 11 (1996), 973–0992.

[14] Sungahn Ko, Isaac Cho, Shehzad Afzal, C. Yau, Junghoon Chae, A. Malik, K. Beck, Y. Jang, W. Ribarsky, and David Ebert. 2016. A Survey on Visual Analysis Approaches for Financial Data. *Computer Graphics Forum* 35 (06 2016), 599–617. DOI: http://dx.doi.org/10.1111/cgf.12931

[15] Artur Lugmayr, Yi Juin Lim, Joshua Hollick, Joyce Khuu, and Felix Chan. 2018. Financial Data Visualization in 3D on Immersive Virtual Reality Displays: A Case-Study for Data Storytelling and Information Visualization of Financial Data of Australia's Energy Sector. In *Enterprise Applications and Services in the Finance Industry (Lecture Notes in Computer Science (LNCB))*, Brahim Saadouni (Ed.). Springer-Verlag, Cham, 118–120. [ACCEPTED].

[16] Ulrik Lyngs. 2019. chi20-papers-rmarkdown. https://github.com/ulyngs/chi20-papers-rmarkdown. (2019). DOI: http://dx.doi.org/10.5281/zenodo.3451422

[17] Patrick Millais, Simon L. Jones, and Ryan Kelly. 2018. Exploring Data in Virtual Reality: Comparisons with 2D Data Visualizations. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA, Article LBW007, 6 pages. DOI: http://dx.doi.org/10.1145/3170427.3188537

[18] Tamara Munzner. 2008. Process and pitfalls in writing information visualization research papers. In *Information visualization*. Springer, Cham, 134–153.

[19] William Revelle. 2018. *psych: Procedures for Psychological, Psychometric, and Personality Research*. https://CRAN.R-project.org/package=psych R package version 1.8.10.

[20] Anya C. Savikhin. 2013. *The Application of Visual Analytics to Financial Decision-Making and Risk Management: Notes from Behavioural Economics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 99–114. DOI:http://dx.doi.org/10.1007/978-3-642-32232-7_5

[21] H. S. Smallman, M. St. John, H. M. Oonk, and M. B. Cowen. 2001. Information availability in 2D and 3D displays. *IEEE Computer Graphics and Applications* 21, 5 (July 2001), 51–57. DOI: http://dx.doi.org/10.1109/38.946631

[22] David P. Tegarden. 1999. Business Information Visualization. *Commun. AIS* 1, 1es, Article 4 (Jan. 1999), 37 pages. http://dl.acm.org/citation.cfm?id=374109.374113

[23] M. Tory, A. E. Kirkpatrick, M. S. Atkins, and T. Moller. 2006. Visualization task performance with 2D, 3D, and combination displays. *IEEE Transactions on Visualization and Computer Graphics* 12, 1 (Jan 2006), 2–13. DOI:http://dx.doi.org/10.1109/TVCG.2006.17

[24] André Calero Valdez. 2019. *rmdtemplates - an opinionated collection of rmarkdown templates*. https://github.com/statisticsforsocialscience/rmd_templates R package version 0.3.2.0.

[25] Jarke J Van Wijk and Edward R Van Selow. 1999. Cluster and calendar based visualization of time series data. In *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis' 99)*. IEEE, IEEE, Burnaby, BC, Canada, 4–9.

[26] Colin Ware and Peter Mitchell. 2008. Visualizing Graphs in Three Dimensions. *ACM Trans. Appl. Percept.* 5, 1, Article 2 (Jan. 2008), 15 pages. DOI: http://dx.doi.org/10.1145/1279640.1279642

[27] Hadley Wickham. 2017. *tidyverse: Easily Install and Load the 'Tidyverse'*. https://CRAN.R-project.org/package=tidyverse R package version 1.2.1.

[28] Hadley Wickham. 2018. *scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales R package version 1.0.0.

[29] Yihui Xie. 2019. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. https://CRAN.R-project.org/package=knitr R package version 1.25.

[30] Yihui Xie, Joe Cheng, and Xianying Tan. 2018. *DT: A Wrapper of the JavaScript Library 'DataTables'*. https://CRAN.R-project.org/package=DT R package version 0.5.

[31] Hao Zhu. 2018. *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra R package version 0.9.0.

[32] Daniel Zielasko, Martin Bellgardt, Alexander Meißner, Maliheh Haghgoo, Bernd Hentschel, Benjamin Weyers, and Torsten Kuhlen. 2017. buenoSDIAs: Supporting Desktop Immersive Analytics While Actively Preventing Cybersickness. In *Workshop on Immersive Analytics*. IEEE, Phoenix, AZ, USA.