Be the Data: A New Approach for Immersive Analytics

Xin Chen, Jessica Zeitz Self, Leanna House, and Chris North
Virginia Tech
{chenxin, jzself, lhouse, north}@vt.edu

ABSTRACT

In an era with exploding amounts of data, educating students about data analysis techniques becomes increasingly important. However, many students find it challenging to understand complex analytical methods and in turn have an unenthusiastic attitude about learning from them. This paper describes *Be the Data*, an immersive analytical approach for teaching abstract data analytical concepts, such as dimension reduction. In particular, we present the design and development of a novel system to engage students in exploring alternative 2D projections of high dimensional data. In our system, each student embodies a virtual data point visualized in a collocated physical space. The coordinates of students on the floor represent coordinates on a 2D projection of the high-dimensional data they embody. Students can explore alternative projections by physically moving themselves, and hence the corresponding data points, in the space. Students receive visual feedback about the data variables that would produce their projection. Therefore, students can pose hypotheses about the data and further explore and understand it. Our goal is to encourage and foster students' interest in data analysis by engaging them in an immersive experience.

Keywords: immersive analytics, dimension reduction, multi-dimensional scaling.

Index Terms: H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces

1 INTRODUCTION

Big Data. Big Data. Big Data. In the news, online, and at work, we are constantly hearing the buzz phrase, "Big Data". With the advances in technology, the amount of analyzable data is growing rapidly at low cost. Within these large datasets is information that we hope to derive scientific discoveries. However, as noted in the book Illuminating the Path [1], datasets are just tables of numbers without humans to discover, process, reflect, and communicate information in the data.

There is a clear need to promote education in knowledge discovery from big data. In practice, learning from data requires comprehensive critical thinking skills which (1) extend beyond the application of quantitative statistical or computational methods and (2) include qualitative forms of thought, such as formalizing potential biases, communicating personal judgment, exploring multiple solutions, assimilating new information with old, and assessing implications of discoveries. Unfortunately, current approaches in teaching data analytics focus primarily on its quantitative aspects to train students to master quantitative theory and methods. Students without strong math prerequisites may be excluded from the analytical classes. Even worse, the complexity of quantitative aspects scare students away from learning. Students normally have an unenthusiastic attitude towards learning data analytics if they do not have a strong mathematical background [2].

Immersive data exploration has the potential to motivate and reinforce quantitative and qualitative aspects of data analyses. To promote STEM outreach and attract students to learn data analytical skills, we designed and developed a novel combination of physical, virtual, and social worlds for immersive data exploration. Specifically, we propose a novel concept, *Be the Data*, which means an individual person embodies a unique virtual data point in a high-dimensional data set. As a proof of concept, we developed a system that immerses students as data points in a physical space (Figure 1). In our system, students enter a physical space to become individual data points, and the room becomes the low-dimensional projection. For example, if we consider a high-dimensional dataset about animals (Table 1), each student becomes an animal data point. Their positions on the ground represent the two-dimensional projection of the high-dimensional data. That is to say, coordinates in the room are coordinates in a two-dimensional plane to which the high-dimensional data are projected.

Be the Data is developed to promote interactive data exploration under an immersive and collaborative learning context. In a collocated physical space, students are able to explore alternative projections by physically moving themselves, and hence their data points, in the space. Therefore, they can collaboratively experiment on the various structures of data by rearranging themselves to discover hidden relationships in the data. This environment facilitates active participation in exploring data, critical thinking about alternative projections, and multiple perspectives on the same data. It helps foster both the quantitative and qualitative aspects of data analysis.

Table 1: A segment of the animal dataset (20 animals X 31 dimensions). The table only shows 4 animals and 5 dimensions.

Name	Flys	Hands	Lean	Smelly	Speed
Bat	95.9	5.28	45.69	30.33	80.96
Giraffe	4.44	0	68.82	22.09	31.86
Gorilla	0	61.5	12.68	42.6	37.06
Skunk	64.86	8.33	16.88	100	30.21

2 BACKGROUND AND DESIGN MOTIVATION

2.1 Immerse In the Data

In the presence of large datasets, research in immersive analytics is devoted to facilitating the comprehension of data by bringing data and the data analysis process into the physical world. We are seeing immersive interfaces develop quickly to immerse analysts in the data for natural methods of data exploration and collaboration.

* email address

Interactive surfaces enable direct interaction with data which is easier and more intuitive than using a mouse on a desktop display [3]. These interfaces are also increasingly developed into large high-resolution displays (or high-pixel tiled display walls) [4, 5]. The combination of higher pixel count and multi-touch interaction allow embodied physical navigation that outperforms virtual navigation, especially when dealing with large datasets [6].

Emerging stereoscopic 3D display technologies, complemented with virtual reality techniques, immerse users in computer-generated scenes. For example, the design of CAVE2 [7] allows users to "walk through" and "fly-over" the hybrid reality scenes. Users in the AVIE interface are able to walk inside a 360-degree stereographic interface to manipulate digital archives in forms text, photos, images, and sound [8]. With tools embracing multi-modalities (e.g., audio, haptic, gestural), immersive analysis is not only about a visual experience, but becomes an integrated multi-sensory experience [9, 10, 11].

In addition to bringing users to the hybrid reality environments, attempts are being made to bring the virtual data into the physical world. Digital data are now made accessible in graspable and manipulative artifacts whose physical attributes (e.g., geometry, materials) encode data [12]. For instance, Professor Richard Burdett presented his Maps of City Population in wooden 3D models in which height property encodes population density [13]. It was an engaging way to represent mass statistical information that invited people to explore. As digital fabrication technologies made it possible and easy to create physical representations of the data (even large datasets), researchers increasingly investigate how to leverage a human's perception skills in exploring data in physical forms [14, 15].

By means of touching virtual data on an interactive surface, walking inside computer-generated scenes of data, or exploring physical representations of data, existing immersive analytical approaches place users in their data. We call it "*Be In the Data*".

2.2 Be the Data – A New Perspective

We propose a new facet of immersive data analytics that seeks to take immersion to the extreme. Unlike existing approaches that immerse users in the data, we immerse users to actually become data points. We call this new perspective "*Be the Data*".

Be the Data shares many similarities as *Be In the Data*. Users navigate and explore a physical-virtual hybrid space to analyze data. Users take advantages of collaborative work in a 3D space. However, *Be the Data* differs from *Be In the Data* in the perspective that the user is taking. Instead of looking into the data points, users are the data points. In our system, after students embody data points, they are able to take an egocentric role in conjecturing various relationships among the data. For example, for being a skunk, I may naturally separate myself from other animals because obviously I am very smelly. Students are able to discuss/negotiate with their neighbors to determine the positions of themselves based on their prior knowledge about animals and based on the context of specific problems. Data exploration naturally becomes a social process of users collaboratively reorganizing themselves in the room.

Research in virtual environments suggests that learning benefits from embodiment of an avatar. The activities performed by avatars inside virtual worlds render situated or authentic learning experiences as users would solve problems contextualized in real life situations. Similar to the idea of an "avatar", here users become a "datatar" as we focus on data. We seek to explore if and how people interact with data in embodied ways through the "datatar" could render an engaging, collaborative, and effective experience, which could lead to deep insights about data and analytical processes.

Be the Data is an extension of a desktop-based application called Andromeda [16] that explores high dimensional-data in a professional manner. The desktop-based platform has its educational limitations. It is difficult to imagine that a novice student would engage in such an advanced interface. Moving data points on a screen could turn into a tedious and meaningless task. Also, it is challenging to conceptualize an abstract mapping from the virtual data to the virtual visualizations. The key concepts and insights are veiled behind small screen portals and simplistic interaction mechanics suggested by mouse and keyboard. Therefore, we invented *Be the Data* for immersive analytics. Within the physical space, students have an intuitive and egocentric spatial perception to judge the physical distance: walking toward people that seem similar to me while staying away from people that seem different from me. The physical metaphor "near is similar" matches the conceptualization of the underlying mathematical model. The concrete experience provides a physical medium for students to reason about the abstract Euclidean distance. Moreover, the shared space brings multiple learners for collaboration and the sharing of ideas.

3 SYSTEM DESCRIPTION

We exploit a multi-media physical cube to implement *Be the Data*. Relying on advanced interactive technologies for physical-virtual cross-overs, our system is comprised of a collocated physical space (Cube), a motion tracking system, several trackable hats, a backend software layer, and an overhead large display.

3.1 Motion Tracking System

Be the Data uses an OptiTrack motion tracking system, which includes 24 Oqus cameras, reflective markers, and the Qualisys Track Manager (QTM) software. QTM is used to collect and process motion capture data from the cameras. We retrieve data from the QTM server over a UDP/IP connection in real-time by following the Open Sound Control (OSC) protocol.

Figure 1: In our system, students become individual data points. A bird's eye view of their locations in the room is displayed on the large display above their head.

3.2 Trackable Hats

To simultaneously track multiple individuals and differentiate them from each other, we made our trackable hats (Figure 2a). Each hat is a rigid body that has its own particular and definite space. It is defined by a particular placement of 4-6 reflective markers (Figure 2b). Each rigid body in 3D space has six degrees of motion freedom (Figure 2c). We determine the 2D coordinates of individuals in the room by streaming the x and z values of the rigid body in real time. The current implementation of the tracking system and the trackable hats allow for accurate tracking and differentiation of more than 50 objects.

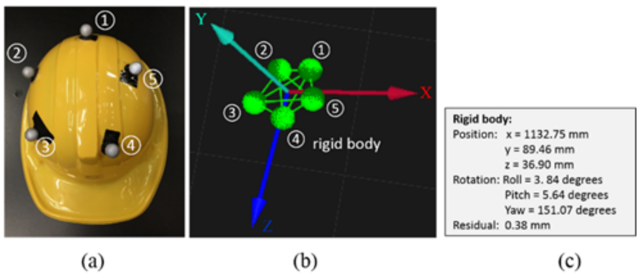

Figure 2: (a) A trackable hat. Its unique structure is defined by the placement of reflective markers on it. (b) A rigid body presentation of the hat in 3D views. (c) The two-dimensional coordinates are from x and z values in the rigid body.

3.3 Backend Software Layer

Be the Data is supported by the backend software layer called Andromeda [16], a desktop-based application for professional data analysis. By applying Visual to Parametric Interaction (V2PI) [17], Andromeda allows users to communicate their ideas about the high-dimensional data by manipulating data points in the visualization, which is a 2D projection of the high-dimensional data. For example, users can drag data points to change the pairwise Euclidian distances among them. Users convey the judgment that data points are similar by pulling them closer and data points are different by pushing them further apart. In turn, the system runs the inverted MDS algorithm to provide visual feedback: a set of weights that describe the visualization.

V2PI shields users from the technicalities of mathematical models so that users may focus on exploring data based on what they know, hypothesize, or learn from the data in an iterative way. *Be the Data* integrates the Andromeda software to immerse users as movable data points in a physical immersive environment. With *Be the Data*, users employ their whole body as portable input that works from any location within the defined area in the Cube. The inputs to the system are users' positions in the Cube captured in real time via the trackable hats. The outputs are interactive visualizations as described in the next section.

3.4 Interactive Visualization

The interactive visualization for *Be the Data* includes two essential parts: (1) a WMDS plot and a dimension chart, organized left and right respectively on the large display (Figure 3), and (2) a dynamic clustering plot on the top of the WMDS plot (Figure 4).

3.4.1 WMDS Plot

The WMDS plot reflects the current physical layout in the Cube (from a bird's eye view). The dimension chart lists the dimensions in alphabetical order and reveals their current weight values. To interpret the plot, the relative distances between data points reflect their similarity or difference: near suggests relatively similar while far suggests relatively different in the dimensions that are emphasized (i.e., variables that are weighted more). All weights are set equal and ordered alphabetically in the default layout (Figure 3a). As users change the layout by rearranging themselves in the room, the weights get updated to explain users' choice of positions (Figure 3b). The length of the dimension bar reflects its relative weight as compared to other bars: longer means a higher weight. For example, as demonstrated in Figure 3a and 3b, the Tiger moves closer to the Pig, thus the Tiger is now considered more similar to the Pig than the remaining animals in the up-weighted dimensions, such as Flipper, Hibernate, and Size.

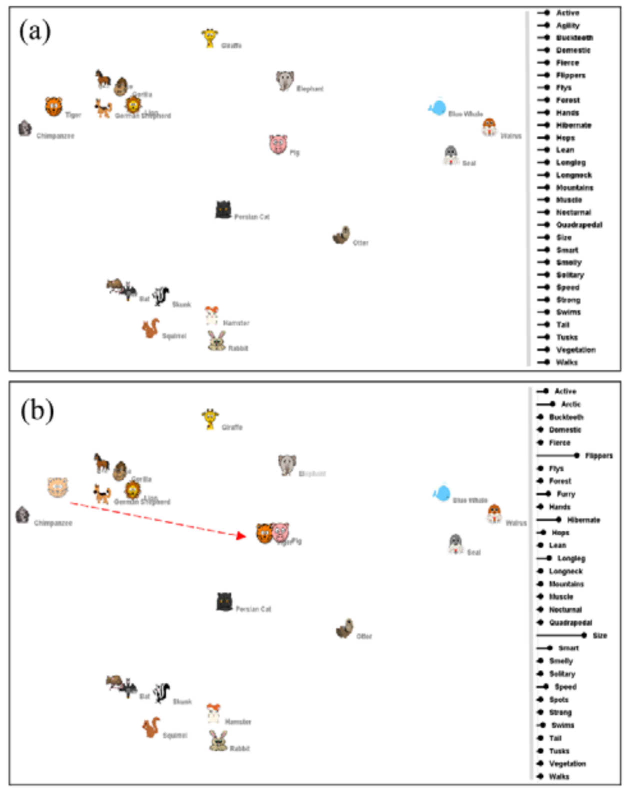

Figure 3: (a) A clear image shown on the overhead large display to visualize students' locations in the room. (b) When students move in the room, they are changing the two-dimensional coordinates of the WMDS plot and relative weights of dimensions.

The underlying algorithm of the WMDS plot relies on Weighted Multi-Dimensional Scaling (WMDS) [18] to visualize high-dimensional data on a two-dimensional plane. WMDS plots a low-dimensional spatialization of the data in 2D Euclidean space to represent how the data spread in the high-dimensional space. The 2D layout is determined by weight parameters of p dimensions $\omega = [\omega_1, \omega_2, ..., \omega_p]$, which reflects the relative importance of each dimension in a visualization. The coordinates *r* of a WMDS plot for high-dimensional data *d* is determined by minimizing a stress function:

$$r = \min_{r_1,\dots,r_n} \sum_{i=1}^{n} \sum_{j>i}^{n} | dist_L(r_i, r_j) - dist_H(\omega, d_i, d_j) |,$$

where *n* is the number of data points, $dist_L(r_i, r_j)$ is a distance between 2D points r_i and r_j, and $dist_H(\omega, d_i, d_j)$ is a distance measured between high-dimensional points d_i and d_j.

The system employs the inverse WMDS algorithm [17] to map layout changes to new values for weights. That is, the inverse algorithm solves weight ω given adjusted two-dimensional coordinates r*,

$$\omega = \min_{\omega_1,\ldots\omega_p} \sum_{i=1}^{n}\sum_{j>i}^{n} | dist_L(r_i^*, r_j^*) - dist_H(\omega, d_i, d_j) | .$$

Because the algorithm considers the relative distance, not the absolute distance between data points, the size of the Cube does not affect the performance of the algorithm. The inverted algorithm runs fast enough to get results in real time.

In *Be the Data*, students adjust their low-dimensional coordinates by rearranging themselves in the Cube. In turn, they are provided with new weights for the dimensions that explain their choice of locations. When students move several times, they are effectively exploring data from multiple perspectives that is defined by different 2D projections and the updated weights. This is a clear advantage of our system that students are shielded from the technicalities of mathematical models and may focus on exploring and learning from data based on their domain-specific questions.

3.4.2 Dynamic Clustering

Although the WMDS plot reveals up-weighted dimensions that characterize users' choice of grouping, it does not show information about how groups distribute on these dimensions. Therefore, we implement dynamic clustering to visualize clusters of data on top of the WMDS plot (Figure 4a). That is to say, given the projected coordinates on the two-dimensional plane, the system automatically reveals clusters of data points based on their Euclidean distance in real time. We focus clustering on the 2D view space, not the high-dimensional space.

Dynamic clustering is calculated by an optimized *k-means*: the number of clusters (*k*) is determined at scene. We apply the heuristic elbow method [19] to automatically refine *k* to improve the quality of clustering. The elbow method plots an error measure (also called percentage of variance) against *k*. The error measure decreases as the number of clusters *k* increases; but starting with some *k*, the decrease suddenly flattens and the appropriate *k* is the one that hits this "elbow".

Centralized cluster values (i.e. the mean value for a given dimension of all data points in a cluster) are calculated. We show relative centralized values on the top highest weighted dimensions. For example in Figure 4a, cluster 2 ranks highest on the Swims dimension, suggesting that cluster 2 differentiates with other clusters because animals in this cluster tend to be good swimmers. With the dynamic clustering feature, the dimension chart is set to be sorted based on the weights. Therefore, users are able to identify cluster distributions on the most up-weighted dimensions that characterize the clustering.

Label switching (if cluster 1, 2, 3 change their color encoding from Figure 4a to 4b) affects users to track their cluster characteristics on the dimension chart. We let clusters appropriately restore the color encoding from the previous clustering. For example, from Figure 4a to Figure 4b, the German Shepard moves from cluster 1 to 3, the Skunk and Chimpanzee move away from their original clusters to form cluster 4, and the Bat becomes cluster 5. We see the current clusters 1, 2, 3 in Figure 4b preserve their original colors from Figure 4a. The German Shepard changes to blue as it merges into cluster 3. Clusters 4 and 5 are assigned new colors.

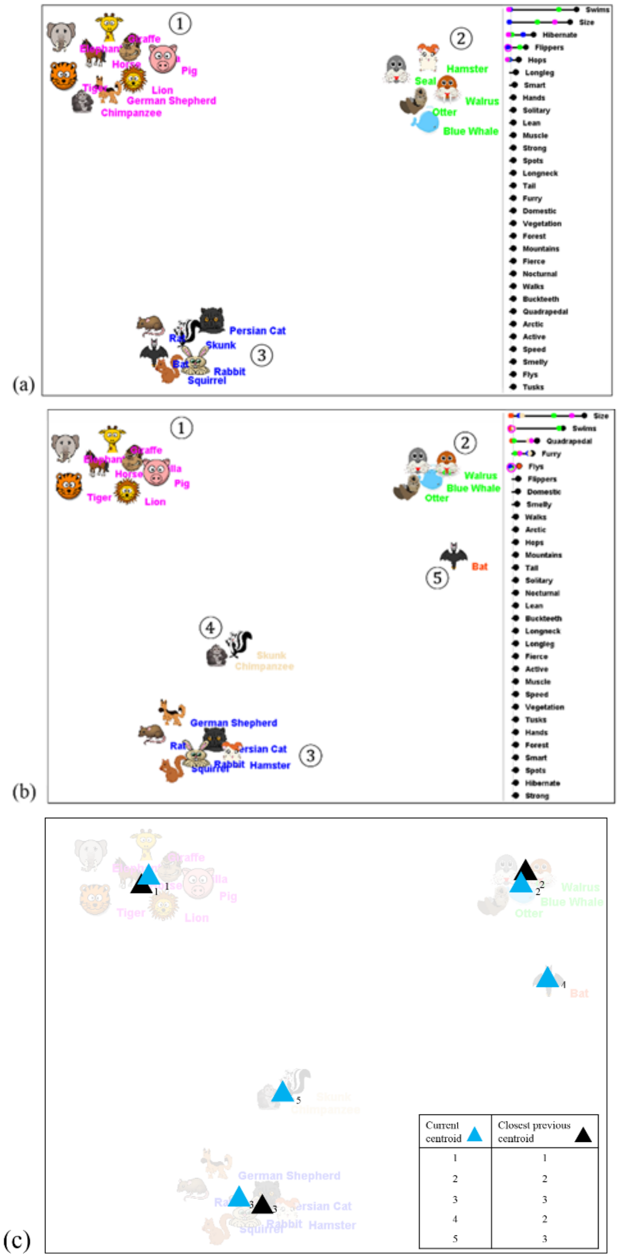

Figure 4: (a) Dynamic clustering of 2D points. (b) Cluster 1, 2, 3 preserve their colors while cluster 4, 5 are assigned new colors. (c) Color preserved by comparing current cluster centroids to previous centroids.

We preserve colors from the previous clustering by comparing the centroids (centers) of current and previous clusters. Specifically, for each current cluster, we iterate its centroid over previous centroids, from which we find one located most closely to the current centroid. If more than two clusters share the same closest centroid, the cluster that appears closest to this previous centroid inherits its color. Figure 4c illustrates how colors are restored from Figure 4a to 4b. In Figure 4c, blue triangles are the centroids of current clusters as mapped in Figure 4b while black triangles are centroids of previous clusters as mapped in Figure 4a. Both current cluster 3 (the blue triangle 3) and cluster 5 (blue triangle 5) have the previous centroid 3 (the black triangle 3) as their closet centroid. Because cluster 3 is closer than cluster 5 to the previous centroid 3, cluster 3 preserves the color from the previous cluster 3 while cluster 5 is assigned a new color.

Dynamic clustering helps students reveal cluster distributions on important dimensions. It also provides an opportunity to verify themselves within and outside of a cluster. We strive for simplicity in our algorithms for linear algorithmic time complexity. Cluster detection is performed real time.

4 Data Exploration Case Study

We presented *Be the Data* in several STEM outreach workshops, as invited by various organizations, including the Center for Human-Computer Interaction, the Association for Women in Computing, the Center for the Enhancement of Engineering Diversity, and the Student Transition Engineering Program which is a summer orientation for incoming freshmen to the College of Engineering.

The goal of our workshop was to encourage and foster further interest in data-related disciplines. We reached over 100 students, ranging from 7th grade middle school, through pre-college, undergraduate, and graduate students. The majority of students participating in our workshops were new to high-dimensional data analysis. They had not learned the MDS algorithm before, with the exception of a few graduate students. We began with enabling students to explore high-dimensional data about animals (Table 1). Each student embodied one animal in the Cube. Students worked collaboratively to explore the data with the system. A sub-group of students congregated in the space (clustering themselves) to discover virtual feedback about what made their data points similar to each other. Some students wandered away from others to identify what made her/him unique.

Through this bi-directional process of posing queries via proactive movement and understanding results through reactive movement, students understood numerous complex and latent relationships in the animal data. They collectively answered many questions about the data, such as "what make some animals good to eat?", "what makes animals more attractive to humans?", "what differentiates predators, prey, neither or both", "how are vegetarians, carnivores, and omnivores different and similar?".

Students exploited embodiment to analyze data. For example, in answering the question "What make some animals good to eat", the student who embodied "skunk" immediately separated herself from the group because she was definitely not edible. Next, all students clustered themselves in two groups as they were edible or not. However, a student who embodied the rat did not feel herself belonging to either of the groups (Figure 5). She explained that although rat was normally not edible, it was good to eat in some countries. Therefore, she moved away from the non-edible group. She then stood between the edible and non-edible groups to identify what made her unique. While dimensions of "domestic, furry, hops, lean, and smelly" contributed the differentiation of edible and non-edible groups, the "buckteeth" dimension further identified the rat out of the two groups. In addition to this example, we found that students were able to produce 3-4 different visualizations for each question. It suggested that they gained a deep appreciation for the many tradeoffs that could be weighed during data analysis.

Students were active participants in their learning. In our workshop, 90% of or more of the workshop time was spent with students actually exploring, discovering, and experiencing data analysis. A middle school teacher commented "I have never seen my students being so engaged". In addition, the improvement from pre-workshop test to post-workshop test indicated that students gained knowledge about WMDS related concepts.

Figure 5: The student who embodied the rat placed herself between the edible group (lower left) and the non-edible group (upper right).

5 Discussion and Conclusion

The goal of *Be the Data* is to promote interactive data exploration for STEM education and outreach. Attracting students to learn data analytical skills is an important national need. We can imagine that on a desktop computer, moving data points around would be tedious. But with *Be the Data*, data points are students who move themselves, either by their own volition or based on instructions from a collaborator. It engages students in an otherwise potentially boring data exploration. Moreover, for students who have no analytical experience, the concrete engaging experience of being a data point makes an abstract data analytical concept such as WMDS approachable to them, instead of scaring them away.

Be the Data has the potential to benefit understanding and learning. It exploits users' embodied skills as they physically interact with the data. It is more nuanced than interacting with symbolic objects. A large body of evidence from embodied learning suggests that people unconsciously apply bodily experience (e.g., distance perception, gesturing, body orientation) to support the cognitive process [20, 21, 22]. We expect that moving around would exploit students' spatial cognitive capabilities [23] and in turn aid understanding in spatial organizations of the high dimensional data.

Be the Data facilitates a collaborative environment which can be extremely beneficial for data analysis tasks [24, 25, 26]. Our collocated space invites students to work together in a social context [27]. Students deploy natural social interactions to communicate with the data model and with each other. They had the authority to determine their own position from their perspectives. They were able to stand out from the group to identify themselves. They also converse, discuss, and negotiate alternative hypotheses to explore data from different aspects.

There are many ways *Be the Data* can be improved and extended. Learned from previous user studies, showing length changes in dimension bars tells students that data are clustered based on up-weighted dimensions, but fails to provide enough information about how clusters distribute on the dimensions. For example, "is my cluster high or low in a particular dimension?" Therefore, we implemented the dynamic clustering feature. In addition, we may give students access to value details of data points and weight parameters with a hand-held device. We can further provide parametric interactions with hand-held devices. Students may tune weights or modify choices of distance metrics to see how it affects the projection. We may further project data on the floor, and students chase their data points as the mathematical model updates.

The idea of being a data point can be applied to other data analytical problems, such as factorial design, classification, and clustering where participants have their features and move into different groups. There are more open-ended questions stemming from *Be the Data*. For example, does physical interaction improve the collaborative understanding of information over purely virtual interactions? What and how can other bodily actions (e.g., pointing, waving, jumping) be applied to interact with data?

ACKNOWLEDGMENTS

This work was supported by NSF grant DUE-1141096.

REFERENCES

[1] J. Thomas. Illuminating the path: the research and development agenda for visual analytics. IEEE Computer Society, 2005.

[2] P. Valero-Mora and R. Ledesma. Using interactive graphics to teach multivariate data analysis to psychology students. *Journal of Statistics Education*, 19(1):1-19, 2011.

[3] P. Isenberg, T. Isenberg, and T. Hesselmann. Data visualization on interactive surfaces: A research agenda. *IEEE Computer Graphics and Applications*, 33(2):16–24, 2013.

[4] C. Andrews and C. North. Analyst's workspace: an embodied sensemaking environment for large, high-resolution displays. In *IEEE Conference on Visual Analytics Science and Technology (VAST),* pages 123–131. IEEE, 2012.

[5] C. Andrews, A. Endert, and C. North. Space to think: large high-resolution displays for sensemaking. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 55–64. ACM, 2010.

[6] R. Ball, C. North, and D. a Bowman. Move to improve : promoting physical navigation to increase user performance with large displays. In *Proceedings of the Conference on Human factors in Computing Systems*, pages 191–200. ACM, 2007.

[7] A. Febretti. CAVE2: a hybrid reality environment for immersive simulation and information analysis. In *IS&T/SPIE Electronic Imaging*, pages 864903–864903. International Society for Optics and Photonics, 2013.

[8] J. Bennett. *T_Visionarium: A User's Guide*. Karlsruhe & Sydney: ZKM/UNSW Press, 2008.

[9] J. Thompson and J. Kuchera-Morin. The Allobrain: an interactive, stereographic, 3D audio, immersive virtual world. *International Journal of Human-Computer Studies*, 67(11): 934–946, 2009.

[10] Y. Tanaka, H. Yamauchi, and K. Amemiya. Wearable haptic display for immersive virtual environment. In *Proceedings of the JFPS International Symposium on Fluid Power*, pages 309–314. 2002.

[11] R. Kehl and L. Van Gool. Real-time pointing gesture recognition for an immersive environment. In *Proceedings of International Conference on Automatic Face and Gesture Recognition*, pages 577–582. IEEE, 2004.

[12] Y. Jansen and P. Dragicevic. Opportunities and challenges for data physicalization. In *Proceedings of the Conference on Human factors in Computing Systems*, pages 3227–3236. ACM, 2015.

[13] Global cities at Tate Modern | Creative Review. http://www.creativereview.co.uk/cr-blog/2007/july/global-cities-at-tate-modern/.

[14] F. Taher, J. Hardy, and A. Karnik. Exploring interactions with physically dynamic bar charts. In *Proceedings of the Conference on Human factors in Computing Systems*, pages 3237–3246. ACM, 2015.

[15] S. Stusak, A. Tabard, and F. Sauka. Activity sculptures: exploring the impact of physical visualizations on running activity. *IEEE Transactions on Visualization and Computer Graphics,* 20(12):2201–2210, 2014.

[16] J. Self, L. House, S. Leman, and C. North. Andromeda: observation-level and parametric interaction for exploratory data analysis. 2015. http://infovis.cs.vt.edu/sites/default/files/jzself-tech-report.pdf

[17] S. Leman, L. House, D. Maiti, A. Endert, and C. North. Visual to parametric interaction (v2pi). *PLoS One*, 8(3):e50474, 2013.

[18] J. Kruskal and M. Wish, *Multidimensional scaling*. Sage, 1978.

[19] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer, Berlin: Springer series in statistics, 2001.

[20] S. Bakker, E. van den Hoven, and A. Antle. MoSo tangibles: evaluating embodied learning. In *Proceedings of the International Conference on Tangible, Embedded, and Embodied Interaction*, pages 85–92. ACM, 2011.

[21] T. Martin and D. Schwartz. Physically distributed learning: adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive science,* 29(4):587–625, 2005.

[22] M. Howison and D. Trninic. The Mathematical Imagery Trainer: from embodied interaction to conceptual learning. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 1989–1998. ACM, 2011.

[23] S. Klemmer, B. Hartmann, and L. Takayama. How bodies matter: five themes for interaction design. In *Proceedings of the Conference on Designing Interactive systems*, pages 140–149. ACM, 2006.

[24] J. Heer and M. Agrawala. Design considerations for collaborative visual analytics. *Information Visualization*, 7(1):49–62, 2008.

[25] P. Isenberg, D. Fisher, and S. Paul. Co-located collaborative visual analytics around a tabletop display. *IEEE Transactions on Visualization and Computer Graphics*, 18(5): 689–702, 2012.

[26] P. Keel. Collaborative visual analytics: inferring from the spatial organization and collaborative use of information. In *IEEE Symposium on Visual Analytics Science And Technology*, pages 137–144. IEEE, 2006.

[27] Chen, X., Fang, Y., & Lockee, B. (2015). Integrative review of social presence in distance education: Issues and challenges. *Educational Research and Reviews*, 10(13):1796-1806