Cultivating data visualization literacy in museums

Kylie Peppler, Anna Keune and Ariel Han
Department of Informatics, University of California Irvine, Irvine, California, USA

Received 15 April 2020
Revised 11 January 2021
Accepted 11 January 2021

Abstract

Purpose – This paper aims to explore what design aspects can support data visualization literacy within science museums.

Design/methodology/approach – The qualitative study thematically analyzes video data of 11 visitor groups as they engage with reading and writing of data visualization through a science museum exhibition that features real-time and uncurated data.

Findings – Findings present how the design aspects of the exhibit led to identifying single data records, data patterns, mismeasurements and distribution rate.

Research limitations/implications – The findings preface how to study data visualization literacy learning in short museum interactions.

Practical implications – Practically, the findings point toward design implications for facilitating data visualization literacy in museum exhibits.

Originality/value – The originality of the study lays in the way the exhibit supports engagement with data visualization literacy with uncurated data records.

Keywords Data visualization, Constructionism, Data visualization literacy, Out-of-school, Science museum, Science museums

Paper type Research paper

Introduction: a need for data visualization literacy

People generate over 2.5 quintillion bytes of data daily, a number that is accelerating exponentially through the pervasive use of smart devices, social media, mobile technologies, online platform services and more (Forbes, 2018). As these data production trends continue, the ability to read, analyze and visualize data sets – a set of skills referred to as *data literacy* (Wilkerson and Polman, 2020; Wise, 2020) – is growing in increasing importance across sectors (Bhargava *et al.*, 2015; Konold *et al.*, 2015; Acker and Bowler, 2018; Roberts and Lyons, 2020; Rubin, 2020). Within educational settings, data literacy is typically incorporated as part of data science approaches (Finzer, 2013; Lee, 2019). Prevalent among these efforts is the use of *data visualization* as a way to articulate, narrate and critique data sets (Bhargava *et al.*, 2015; Philip *et al.*, 2016). Data visualization encompasses how a data set is visually rendered (e.g. list, scatter graph, map), overlaid with other data variables (e.g. age group) and encoded (e.g. mapping data variables onto graphic variables; Börner, 2015).

Despite the preponderance of personal data in people's daily lives and the myriad tools for visualizing and tracking it, data visualizations in educational settings too often rely on

Funding: This material is based upon work supported by the National Science Foundation under Grant No. 1713567 awarded to PI Katy Börner, co-PI Joe Heimlich, co-PI Bryan Kennedy, and co-PI Kylie Peppler. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Information and Learning Sciences

2398-5348
DOI 10.1108/ILS-04-2020-0132

normative, prepared data sets that have been designed so that specific insights can be easily consumed (Konold *et al.*, 2015). In real-world settings, it may not always be transparent how and by whom data is produced (Philip *et al.*, 2016), which kind of patterns can be interpreted from it, and whether the data is accurate. Science museums present a unique opportunity for increasing data visualization literacy because they can produce emerging data sets through real-time and sensor-based visitor-generated data records, offering the possibility for visitors to engage with personally meaningful data visualization (Hardy *et al.*, 2020; Mallavarapu *et al.*, 2019; Roberts and Lyons, 2020; Stornaiuolo, 2020). In an effort to engage the public in activities to cultivate such data visualization literacy, we ask: *What aspects of museum exhibit design invite visitors to engage meaningfully with data visualization literacy?*

To answer this question, this paper follows Börner's (2015) definition of data visualization literacy, which is the ability to read and construct visual representations to make meaning of data and to support the understanding of datasets through data visualization types (e.g. scatter graph, geo map), data variables (i.e. qualitative, quantitative) and graphic variable types (e.g. shape, size, color). Fusing data visualization literacy with constructionist approaches to learning (Papert, 1980, 1993), this qualitative analysis investigated visitor engagement with the *Walk* exhibit at an urban midwestern science museum. The Walk exhibit gave visitors the opportunity to capture their time as they traversed a walking path, and compare their time against other visitors in the museum, along with the data each user entered about themselves (i.e. icon and favorite color, height in inches, favorite activity, age group and zip code). Researchers captured visitor engagement with the exhibit through video observations and semi-structured interviews. From a total of 74 observed groups, researchers conducted an in-depth analysis of 11 visitor groups, the groups that engaged with the exhibit for longer periods and included youth. First, thematic content analysis across data sources supported the creation of brief narratives about group engagement. Iterative and thematic analysis of semi-structured interviews and an analysis of interactions with the Walk exhibit identified what design aspects informed engagement with data visualization literacy and how. Findings present three prominent design aspects to support the engagement with data visualization literacy based on our analysis of visitor engagement with the Walk exhibit:

- enabling reading and constructing data visualizations across different visualization types can support identifying single data records, comparisons of data records and (underrepresented) data patterns;
- opportunities to question the accuracy of visualized data can invite critical engagement with data visualization literacy; and
- encouraging real-time and physical data entry can support sensing data and correcting faulty reading of data visualizations.

The design aspects and the engagement with data visualization literacy have implications for the design of museum exhibits that are intended to support data visualization literacy and present starting points for studying data visualization literacy learning in science museum settings.

Background

Data literacy, the ability to make meaning from data, is growing in increasing importance (Wilkerson and Polman, 2020; Wise, 2020; Rubin, 2020). One way to support making meaning from data is through data visualizations (Bhargava *et al.*, 2015). Data visualizations are used to help people understand data relationships, as opposed to focusing on individual data records (Konold *et al.*, 2015; Lee and Dubovi, 2020). Where recent work in the learning sciences focused on data literacy toward equitable and socially just engagement (Wilkerson and Polman, 2020; Wise, 2020), making meaningful interpretations of data using data visualizations requires its own set of literacies (Börner, 2015). These skills include selecting a visualization, such as a scatter graph or geo maps, choosing which data variables (e.g. age) to gather, identifying how to visualize them on the scatter graph or map and considering what graphic elements to use to show the data on the visualization (e.g. shapes and color).

Data visualization literacy (DVL) refers to the ability to read, analyze and visualize complex datasets, through the possibility to:

- select a visualization type (e.g. list, scatter graph, map);
- overlay data (e.g. geolocate); and
- visually encode data (e.g. mapping data variables onto graphic variables; Börner, 2015).

A data visualization literacy framework (DVL-FW) supports the visual encoding and decoding of complex data toward effective meaning-making. In this research, we drew on Börner's DVL-FW, which consists of a typology and a process model. The typology defines seven types relevant for designing effective data visualizations: insight need, data scale, visualization, graphic symbol, graphic variable and interaction. It offers a comprehensive way to break apart and help construct data visualizations with the aim to communicate and interpret information and trends embedded within complex data sets. The framework provides a way to understand how people's engagement with data visualization literacy leads to understanding data visualizations and use them for personal meaning-making.

Science museums are unique educational settings for engaging with data visualization literacy because these spaces can make it possible for visitors to contribute personal data through sensors to a larger publicly emerging data set. This makes it possible for a larger contextual and personally meaningful data set to be produced over longer periods of time and across multiple visitor groups. Within science museums, data visualization has been used to augment engagement with the scientific ideas behind museum exhibits (Mallavarapu *et al.*, 2019; Roberts and Lyons, 2020). Data visualization techniques have further been used to visualize visitor flow and other forms of exhibition engagement, framing data visualization as an analytical tool for museum workers (Strohmaier *et al.*, 2015; Schettino, 2013) also to support the design of future exhibits (Gwilt *et al.*, 2019). An interesting example of the use of data visualization as an underlying design aspect of a science museum exhibit is the interactive exhibit EMDialog, supporting exploration of an interactive presentation related to artist Emily Carr, which found the utility of supporting shared visitor group engagement and engaging with data visualization in multiple ways (Hinrichs *et al.*, 2008).

Beyond science museums, important aspects of engaging with data visualization relate to understanding data collection methods and selection processes of data (Stornaiuolo, 2020; Wilkerson and Polman, 2020; Wise, 2020), as well as recognizing patterns and aggregates (Konold *et al.*, 2015; Lee and Dubovi, 2020). According to constructionist theories of learning, it is particularly the engagement with personally meaningful contexts and personally created data sets that can support engagement with data visualization (Acker and Bowler, 2018; Stornaiuolo, 2020). Further, supporting learners with opportunities to actively produce personal data records can provide opportunities to develop a deep understanding of data infrastructures (Hardy *et al.*, 2020; Kahn, 2020). These studies point to design aspects of how to support data visualization in museum exhibits. Yet, within science museums it would be purposeful to explore what design aspects can foster meaningful engagement with data visualization literacy.

Constructionist approaches to engaging with data visualization literacy

Across our work, we seek to involve learners in coming to understand complex concepts such as data visualization through the design and reflection on externalized artifacts. This is in line with constructionist approaches to learning, which suggest that the design of a personally meaningful externalized artifact is particularly important to the learning process (Papert, 1980, 1993). Piaget's constructivist learning theory argues that learning needs to engage the "plane of activity" before engaging the "plane of language," the formalized speech or written text that describes the experiences (Inhelder and Piaget, 1958). Importantly, when these two stages are inverted, they seem to have immediate and long-term negative consequences for learning (Schneider *et al.*, 2013).

Papert's constructionist theory of learning (Papert, 1980) postulates that being part of a creation process positions the learner as an active agent, rather than as a passive recipient of materials designed for the learner. Designing an artifact (and in the case of this article, designing a data visualization), is what Papert (1980) called an "object-to-think-with" (p. 23) and involves externalizing one's current mental model and iterating on it throughout the creation process (e.g. see iterative visualization design cycle). Designing an artifact creates several conditions that are ideal for learning. First, having to explain through an externalized artifact or in words what you think you understand necessarily requires a reorganization of that idea into a different format. Second, the creation of an externalized representation and observations or reflections on that design create an opportunityto receive formative feedback. Learners can be engaged in building any number of artifacts in this process, including creating a sculpture, composing a song or creating a visualization, and it is important that learners are actively engaged in creating something that is meaningful to themselves and to others around them (Resnick, 2002).

High-quality museum exhibits can be educationally meaningful because participants learn by creating their own artifacts (Papert, 1993). Exhibits have also demonstrated significant promise for engaging youth with STEM topics in personally meaningful ways (Case, 1996). It is the engagement with personal data, such as seeing and visualizing one's own data in relation to a larger pool of data records, that could contribute to a sustained and potential deep engagement with data visualization literacy.

Methods

This study investigated design aspects to support data visualization literacy through qualitative observations of the Walk exhibit at the Center of Science and Industry (COSI), a science museum in Columbus, Ohio. Exhibits at COSI invite over one million visitors annually, including school, out-of-school and family groups to engage with temporary exhibits and permanent activities. The Walk exhibit was set up in a second-floor hallway to provide an accessible yet secluded space to foster prolonged engagement with the exhibit (Colleagues and Authors, under review).

The Walk exhibit (Figure 1), designed by Professor [BLINDED]'s research team at [BLINDED] University in conjunction with the Exhibit Design group at the Science Museum of Minneapolis, stretches 80-feet long and includes three areas:

- a data entry station for visitors to enter data (i.e. an icon and color, height in inches, favorite activity from a list of options and zip code);
- a walkway that visitors stroll along to capture walking times; and
- the Make-A-Vis (MAV) screen version 1.0 that displays museum visitor data and invites visitors to create personalized data visualizations.

In the exhibit, participants first enter their data at the data entry station, then, after a verbal countdown, walk across the walkway. Once they reach the finish line, a motion sensor captures the time in seconds it took the visitor to walk across; this data is then added to the personal data that visitors entered prior to walking.

The Make-A-Vis screen version 1.0 (Figure 2) displays the 50 most recent visitor data records across three DVL framework topology types:

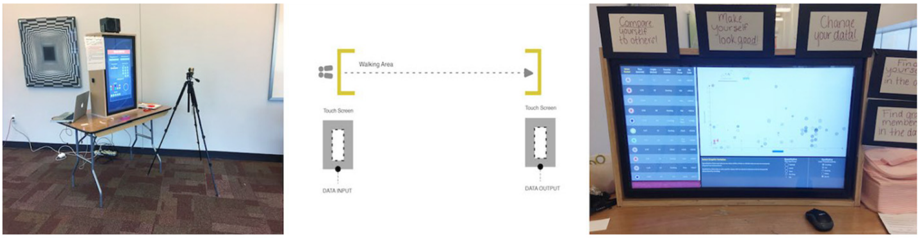
(1) visualization types;

(2) data variable types; and

(3) graphic variable types.

With the MAV, visitors are invited to engage with and make sense of contextualized data that includes their personal data records. Additionally, with the aim to support and scaffold visitor engagement with the MAV screen, tasks are displayed around the MAV screen that are aligned to the data visualization framework. The tasks include:

- find yourself in the data;
- compare yourself to others;



Notes: Data entry station (left), floorplan (center), and Make-A-Vis screen version 1.0 for data visualization creation (right)

Figure 1.
Walk exhibit setup at COSI

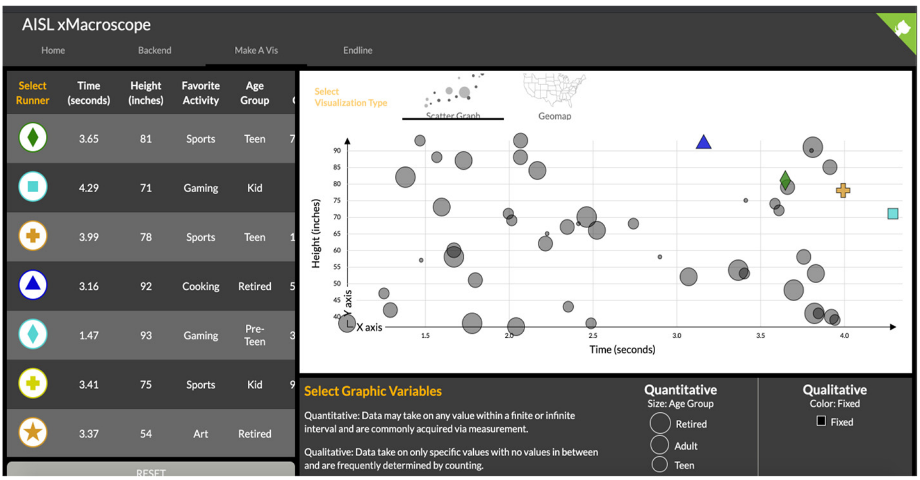


Figure 2.
Make-A-Vis (MAV) screen version 1.0

- find a group member in the data;
- make yourself look good; and
- change your data.

Data visualization types include a data table, a scatter graph and a geo map. The data table is always visible on the left of the screen and shows the data variable types that participants can personalize. Visitors can scroll through the list to view all data records and select one data case at a time to be highlighted on the scatter graph or geo map visualization types. When the visitors engage with the scatter graph, they can manipulate the axes by dragging the data variables and dropping them into each axis on the scatter graph. This makes it possible for participants to explore relationships between variables. In the COSI installation of the walk exhibit, the geo map visualization type presented the Ohio state map as a default and, from here, visitors could select to change views to represent data records on the USA as well as the world map. The map visualization type showed zip code data variables.

The MAV screen included five data variables:

- Runner, a visitor selected icon (e.g. star, diamond, square, triangle and cross) that was paired with the favorite color they selected (i.e. yellow, purple, red, blue and green);
- the time in seconds that it took participants to cross the walkway (i.e. the physical data entry);
- the height in inches that the participants entered;
- the favorite activity that visitors selected at data entry from a list (i.e. sports, art, gaming, cooking);
- the age group that they chose from a list (i.e. kid, teen, adult, retired); and
- the zip code of the location the visitors were, presumably, from.

Beneath both the scatter graph and the geo map visualization type, the MAV also showed qualitative and quantitative graphic variable types (i.e. size for quantitative data variables such as height and color for qualitative data variables such as age group). Visitors could drag and drop data variables onto graphic data variables to explore possibilities as not all data variables could be represented with all graphic variables.

Participants

Museum visitors dominantly explored the museum exhibit in family or friend groups. This meant that a wide age range of people engaged with the exhibit, also frequently together. For instance, a family group in which two adults supported two younger children in their engagement with the exhibit, such as entering data and exploring the MAV screen, were common. Observations could not consider individuals' interactions with the exhibit as separated from their whole family engagement. Thus, to study data visualization literacy engagement with typical visitors at the museum, we focused our observations on groups and their engagement with the exhibit, which naturally included people of a wide range of age groups. We observed a total of 74 groups, which ranged between one and eight people with an average of three people per group (total of 195 individuals). Of all participants, 8% (n = 16) were Latines. Most participants were white (74%, n = 145). Overall, 11% (n = 21) of participants reported more than one racial identity, 10% (n = 20) were black, 3% (n = 5) were Asian, and 1% (n = 2) selected "other" when reporting their racial background. Some visitors did not provide race information (1%). Of all participants, 53% (n = 103) identified as female, 44% (n = 85) as male, and 3% (n = 5) as gender non-binary. Two participants (1%) did not disclose gender-related information. Of the visitors, over half were adults (52%; n = 101), followed by 21% (n = 40) of youth age 7–10, 11% (n = 22) age 14–16, 10% (n = 19) below age 7 and 6% (n = 11) age 11–13. Two visitors (1%) did not provide age information. The age distribution was unsurprising, especially in light of the typical family visitor group structures that frequented the museum and included multi-generational members, such as children, parents and grandparents. These led to adults dominating the groups in terms of numbers.

Visitor groups engaged with the exhibit for an average of 7 min, with a maximum engagement time of 19 min and a minimum engagement of one minute. This is above the average for museum exhibits (typically one minute; see Dancstep and Gutwill, 2019) and, thus, frames the observations as suitable for study. Among the 74 visitor groups, the 11 groups that engaged with the exhibit for more than 8 min included youth (e.g. family groups). This presented a substantial length of engagement of typical multigenerational visitor groups at the museum. Thus, we chose to focus in depth analysis on these 11 groups.

Research question

To answer the research question – *What aspects of museum exhibit design invite visitors to engage meaningfully with data visualization literacy?* – we investigated the particular features of the Walk exhibit that were designed toward the engagement with data visualization literacy and the type of engagement with data visualization literacy these aspects led to.

To observe visitor engagement and to identify aspects of the design that supported data visualization literacy across the exhibit, we recorded three different sets of data, covering the three main areas of the exhibition:

(1) video-recorded data entry;

(2) video-recorded walks; and

(3) video-recorded facilitated engagement with the MAV screen.

Table 1 presents an overview of the data sources and the analytical techniques that correspond to each data source.

Data sources

To observe visitor engagement and to identify aspects of the design that supported data visualization literacy across the whole spatial setup of the exhibit, our data sources included the following steps: First, *video-recorded data entry* captured visitors' data input. Although

Data source	Analytical technique
Video-recorded data entry	Spreadsheet to match data entries with visitor groups and individuals, as well as recognize whether visitors engaged with their own data records on the screen
Video-recorded walks	Frequency of walks and thematic analysis of social activities related to walks
Video-recorded facilitated MAV engagement	Thematic content analysis of visitors' emergent engagement patterns with data visualization literacy

Notes: *Research Question*: What aspects of museum exhibit design invite visitors to engage meaningfully with data visualization literacy?

Table 1. Overview of data sources and analytical techniques

it was possible to download all data entries through the exhibition system, it was not possible to match the individual data records with particular individuals and visitor groups. The video of visitors entering digital data was captured using a GoPro camera that was positioned just behind the visitors, focusing on the screen to enable capturing the visitors' hands as they entered data. This camera also made it possible to capture multiple data entries of the same visitors. As visitors joined the Walk exhibit in groups, we recorded the videos as a record of the data entry of each individual visitor. The recorded video data made it possible to capture personalized digital data entries and to match these data entries with individuals and groups. This supported the analysis of whether visitors recognized their own data entry on the MAV screen visualization.

Second, *video-recorded walks* produced a data record of physical data entry and the personalized manner in which it happened. The data showed visitors walking in different manners from the starting line to the finish line, as well as unfacilitated comments that visitors made about the exhibit. The video of visitors walking across the walkway was captured using a camera with a wide-angle lens that was positioned at the end of the walkway facing toward the data entry to capture the whole walkway. This data made it possible to capture whether and when visitors, especially young ones, performed a repeated number of walks and whether they were captured by the exhibit and visualized on the MAV screen. Additionally, the video captured interactions of visitor groups with aspects of the walkway in addition to walking itself, including the motion sensor that captured the visitors' walking times in seconds.

Third*, video-recorded and facilitated MAV engagement* captured visitors while directly working with data visualizations, including data visualization types and data variable types. While visitors were engaging with the screen, we facilitated their engagement with questions that asked visitors to reflect on their experience with the screen, explain their engagement process, and articulate what they learned from their engagement with the visualizations. We recorded the facilitated MAV engagement using an iPad that was positioned on a tripod behind the visitors. The video, thus, captured the visitors from behind while recording their speech and focused on how they engaged with the screen through the provided mouse.

Analytical techniques

The collected data served as sources to analyze engagement with data visualization literacy across the full setup of the exhibit, the digital and physical personalization of data entry, the data entry station and the walkway, as well as the reading and writing of data visualizations at the MAV screen.

First, we *analyzed the video-recorded data entry* by entering the digital data that visitors added to the system (e.g. name, age, favorite color, etc.) into a spreadsheet. To analyze the data, we first matched visitors' data entries with particular visitors and visitor groups as well as how frequently individual visitors entered data and who in the group entered data (e.g. only the adults or mostly the children in multigenerational visitor groups). This kind of shared engagement was particularly interesting to analyze because the exhibit was conceptualized by the design team as an individual experience; individual group members had to take turns entering data. Being placed in a museum setting, where visitors engaged in groups, this design aspect was interesting to explore in relation to data visualization literacy engagement. The spreadsheet of data entries matched with particular visitor groups and individuals in a group further supported the analysis of the video-recorded and facilitated MAV engagement, helping to see whether visitors engaged with their personal data entries.

Second, we *analyzed the video-recorded walks* by capturing frequency counts of walks as well as thematic analysis of social activities related to the walks. Analyzing the physical data entry of the walks supported understanding how this design aspect of physical personalized data entry supported engagement with data visualization literacy, in addition to the digital data entry. Focusing the analysis on the way visitor group members engaged with the physical space of the exhibit further helped identify whether and how the design aspects of the exhibit related to physical data entry, including the motion sensor, performing walks that were not recorded and walking styles (e.g. running, strolling etc.), supported engagement with data visualization literacy in the museum setting.

Third, the *analysis of the video-recorded facilitated MAV engagement* focused on whether and how visitors read and write data visualizations in relation to data visualization types, data variables and graphic variables available through the MAV screen. The facilitation of the engagement made it possible to thematically elicit those visitors' engagement processes with the MAV screen that led to the production of personalized data visualizations as well as engagement with data visualization literacy. Combined with the other two data sources, it was possible to analyze when and how participants engaged with their personal data entries and those of other visitors, both members of their own visitor group as well as prior visitors. In this analysis, we paid close attention to how visitors reflected on their engagement process. Additionally, combined with other data sources, the analysis of the video-recorded and facilitated MAV engagement were used to produce narratives, brief descriptions of how groups engaged with the exhibit. The narratives were then analyzed using a thematic content analysis to characterize emergent engagement patterns that were linked to personalized digital and physical data entry as well as reading and writing data visualizations at the MAV.

Findings

Across the data analysis, focusing on the 11 groups who engaged over 8 min and included youth, we identified four themes that supported engagement with reading and writing of data visualizations. First, the engagement with the MAV screen supported the reading of data visualizations to identify single data records as visitors spotted their own data record as well as those of other group members by producing different visualizations of the data record by changing data visualization types and graphic variable types. Second, the MAV screen engagement further supported reading of data visualization to identify data patterns, including those patterns that were not included. This became particularly possible through the geomap visualizations as visitors flipped between State map, country map and world map visualizations, recognizing that data clustered around the state of Ohio and did not include entries beyond the USA. Third, physical data entry by walking supported the identification of mismeasurements, data that was located at the edges of the graph, and considerations of producing a reliable data record. Lastly, the physical data entry made it possible to sense personal data, which supported visitors in correcting common misunderstanding about the visual distribution of values on a graph. In the following, we elaborate these themes.

Reading data visualizations to identify single data records

Our analysis of the video of the facilitated MAV engagement in combination with the spreadsheet of data entries showed that museum visitors engaged with the reading of data visualizations by finding their own data records or those of other group members and by comparing their own data records to others across all three visualization types (i.e. data table, geo map, scatter graph). Constructing data visualizations was predominantly related to adding new data records to the MAV and to dragging data variables into graphic variable types.

Identifying their own data record among the many data records shown on the MAV screen was the most frequent engagement of visitor groups with the MAV data visualization screen (n = 58, 78%). Across data visualization types (i.e. list, map, graph), the video analysis showed that visitor groups identified their own personal data cases predominantly through the geometric shape and color they entered at data entry. For instance, one youth visitor navigated to the Ohio map and stated while looking at a nearby adult in their group, "(I can) see where I live." The visitor pointed at their icon on the map that was surrounded by a cluster of other visitors' data records.

Finding their own data record on the MAV was a productive reading activity related to data visualization literacy because it required the engagement with data visualization types as well as data variable types. Our analysis of the video of facilitated MAV engagement further showed that finding personal data records was closely followed in frequency by acts of comparing one's own data record with those of others, especially the data records of other members of the same visitor group (n = 35, 47%). For example, when several visitors selected a similar combination of shapes and colors at the data entry, visitors flipped back and forth between data visualization types and used a wider number of data variable types (i.e. height and favorite activity in addition to icon and color) to distinguish one's own data record from that of others.

Our observations of visitor groups with data visualizations through engagement with the MAV screen pointed to the utility of including several data visualization types and several data variable types in combination with entering personalized data records. This range of ways to read data can support museum visitors in finding their own data records, comparing data records across visualizations and distinguishing their own data records from those of others, especially those that look similarly across several data variables (e.g. icon and color). Multiple ways to read data combined with the possibility to enter personalized data when entering the exhibit as well as through walking across the walkway made it possible for visitors to add multiple data records to the exhibit. In this way, visitors could contribute to the construction of data visualizations as well as take part in producing overall trends in the data set (e.g. by entering many data records with the same zip code).

Reading data visualization to identify (unrepresented) data patterns

The analysis of the data entry spreadsheet in combination with the analysis of the video of facilitated MAV engagement showed that reading data visualizations by actively comparing data records also led to identifying larger data patterns. For example, a group of three teenage girls compared data records on the geo map data visualization type, flipping back and forth between tvhe Ohio state map, the USA country map, and the world map. Daniela, one of the girls in the group observed when navigating from the Ohio state map to the US country map and then the world map: "Woah. It's the whole United States, but most of it is in Ohio [. . .]. [clicks on world map] And then the world is just in North America". Daniela was surprised that much of the data records clustered around Ohio on the US map and around the USA on the world map. Daniela had expected that zooming out on the map through three different geo map visualization types (from State to country to world map) would gradually reveal more and more data records. Instead, Daniela was presented with a country map in which most data clustered around one state, Ohio, and one country, the USA, where the exhibit and the museum were located.

The whitespace on the data visualization surprised Daniela and she used the whitespace on the map to support her claim that "not many people do it (use the exhibit). Like see." She selected the US map and pointed to the cluster of data entries that centered around Ohio with only very few data entries located around the other states of the country. The exploration of the different geo map visualization types opened up something new to her and she could use the information to support her reasoning about aggregate data records. Daniela's sensitivity to the whitespace on the map may have been related to her own personal migratory history. She later disclosed that she and her family immigrated to the USA from Central America. Other visitor groups with migratory and international backgrounds also pointed to the whitespace as indicatory of what and who was not represented in the exhibit. Whitespace on data visualizations made it possible to reason about data aggregates and usage patterns.

The exploration of data patterns can empower people to think that they can look at a data visualization and know that they can learn something beyond their own data record. In the MAV, whitespace opened up opportunities to question patterns within the data, including who was and who was not represented, how and why. In the case of Daniela and international visitors of the museum, whitespace became a form of erasure.

Physical data entry to identify mismeasurement

Our analysis of the video of facilitated MAV engagement showed that when reading walking times on the scatter graph, visitors pointed at mismeasurements within the data visualization (e.g. data points that were below a certain time), isolated points at the edges of the graph that seemed impossible. To make sense of seemingly impossible data points, visitors first sought to find explanations of how these data points could have been produced before discounting them as mismeasurements that should be removed from the larger visitor-generated aggregate. For example, one of the youth visitors pointed out a walk time of 0.37 s and stated, "maybe this person was on a pogo stick or a skateboard." It was the mismeasurements that highlighted to visitors that the data visualization treated all of the physical data entries for walking times equal although, as they noted, people engaged in different ways of walking. Visitors explained impossibly fast waking times with different walking styles, the direction of the motion sensor and the timing of the initiation of a run and considered that subsequent design iterations of the exhibit should capture walking styles. For example, visitors suggested that "one question should be [related to] style of walking to make sense of correlation, keep running trends, to see trends per state." This observation and suggestion to ask for recording a particular walking style at data entry suggested that the mismeasurements in the data visualizations led to considerations about reliability of the data records and visualizations as well as the importance of a consistent form of collecting data to make meaning from the visualizations.

Additionally, when looking at the video of walks, we saw that, as visitors walked slowly across the walkway or ran as fast as they could, some were intrigued by the motion sensor and wanted to learn more about the technical set-up. Participants noticed that the timer occasionally stopped before visitors crossed the finish line and that this depended on the sensor's position (Figure 3).

For example, a youth visitor began to wonder whether the sensor of the exhibit was set-up properly as the timer stopped before he physically crossed the finish line. Closely examining the sensor, he explained "You learn facts you didn't know about things". The technological transparency of the exhibit made the motion sensor visible to visitors and transparently communicated how data was collected. The exploration of the mode of data collection via the motion sensor led to inquiries about input/output relationships as visitors observed the accuracy of its reading. Engagement with the input and output settings of the exhibit highlighted a different way of engaging with data visualization literacy, paying attention to how data can be collected.

Our observations showed that visitors identified mismeasurements related to the physical data entry on the data visualizations of the MAV screen as well as inconsistent forms of data collection based on the form of physical data entry. Identifying challenges with the physical data entry led to reasoning about why mismeasurements were captured (e.g. different walking styles) as well as suggestions of how to produce more reliable data. Visitors explored the transparently available motion sensor to reason about how physical movements were translated into digital data points that were then part of the data visualization. The prototypical set-up of the exhibit invited design recommendations by visitors that led to critical engagement with data visualization literacy, especially related to the purpose, reliability and audience assumed of the data visualizations in context of the museum setting. Considering the transparent presentation of capturing physical data can support the reasoning about how data is being collected as well as the accuracy of this form of data collection. It seemed as though the possibility to physically feel the data that was being shown on the visualization screen supported a more personally meaningful engagement with the data and the nuances of its representation meant.

Physical data entry to identify distribution rate

Looking across the video of walks and that of facilitated MAV engagement showed that embodying physicalized data also led to engagement with distribution rate. After walking, visitors often considered the data records that were positioned on the far-right side of the *x*-axis as faster, although these data records were located further away from 0, the origin point of the axis. Within the Walk exhibit, and particularly the MAV, axes of the scatter graph visualization type could not be turned off. The scatter graph always displayed two axes. Yet, feeling the walk data through physical data entry with the motion sensor and seeing how such data records got translated into representations on the scatter graph in combination with other personal data variables made it possible for visitors to correct their reading of the scatter graph.

For example, a family visitor group of three – a father, Ben, and two youths, Robert and Justin – repeatedly entered data records and compared these data records on the scatter graph. First, both youths arrived at the MAV after entering two separate data records, one after the next. The MAV screen displayed a scatter graph with height on the *x*-axis and walking time on the *y*-axis. The facilitator (Author 2) asked which their data records

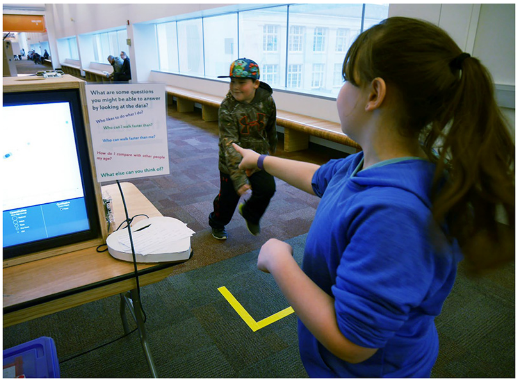


Figure 3.
One visitor looked at the motion sensor while another pointed out where the motion sensor tracked walking times

made a faster time, and Robert responded: "I (am) the top one", before both darted off to enter another data record. Although Robert's claim suggests that he considered his walking time as faster compared to that of his brother and that the faster data record should be located higher up on the *y*-axis. However, faster the walk time, the lower it should be positioned on the *y*-axis as the bottom of the *y*-axis marks zero and the values increase toward the top.

The brothers entered another data record, this time racing against each other simultaneously. Both youths turned to the MAV to see their latest shared data entry, which was located lower on the *y*-axis than both of their previous data records. In sync, both youths called: "I won.", implying that this time, they were faster. Ben, their father, who had followed them closely, asked: "Is that lower?", implying that the boys' reading of the walk times was incorrect. However, the youth used the data visualizations to support their claim. Robert stated: "I won, see?" while pointing at the latest data record that was located below any of the prior records (Figure 4).

Knowing that the speedily recorded entry, in which the brothers ran, must have been faster than the data entry in which the two slowly strode across the walkway provided the boys with sensory information about how to interpret the visualization. This sensory information seemed to support the older boy in correcting their previous misunderstanding of where on the *y*-axis faster times are located.

The possibility to enter physical data led to not only identifying mismeasurements but also embodying data toward engagement with data distribution across two axes. The possibility to enter multiple data records through physical data entry afforded visitors to feel their data, to physically compare their data records over time, and to see how these different data records were translated into visual representations on the MAV. Combined, it opened up possibilities to engage with distribution of data in a different way.

Discussion

The findings point toward design aspects that can support sense making of uncurated and continuously updating data within museum settings. The design aspects of the exhibit (i.e. reading data visualizations and physical data entry) led to engagement with data visualization literacy, namely, to identify single data records, (unrepresented) data patterns, mismeasurements and distribution rate.

The way participants engaged underscored the need to design personally meaningful entry points for engaging with data and showcases challenges in helping museum visitors move from individual data points to investigating larger trends in the data (i.e. moving from locating their own or other single data records to recognizing aggregates and outliers within larger trends). Together these findings point toward the need to scaffold engagement with

Figure 4. Youth points at their own data point on the MAV in support of their argument

data visualization literacy (i.e. visualization types, data variable types and graphic variable types) through different tasks. In the Walk exhibit, tasks included finding personal data records, comparing data records and adding data records. It would be interesting to explore the utility of these tasks in other data visualization literacy contexts.

Designing for entering personal data multiple times combined with arranging this data across multiple data visualization types and data variable types can support engagement with data visualization literacy, especially identifying single data points (e.g. one's personal data and that of other visitor group members). Additionally, our observations highlighted the *utility of whitespace for engaging with data visualization literacy*, especially aggregates, trends and patterns, in the museum setting. Yet, our observations indicated the importance of supporting museum visitors to enter data records beyond one country to produce data records that reflect the international backgrounds of the museum visitors and to counter possibilities of omitting groups of people from being represented and from engaging.

Designing for physical data entry led to pinpointing mismeasurements (e.g. too fast walking times), to consider categorizing data (e.g. through walking styles), and questioning data collection methods (e.g. with a motion sensor). *Physical data entry along with transparent presentation of how this data is being captured* may provide opportunities for engaging critically with data visualizations, especially the need to carefully consider how data is captured. These aspects could be further facilitated in subsequent museum designs.

Conclusions

Returning to the research question related to the aspects of museum exhibit design that invite visitors to engage meaningfully with data visualization literacy, our findings point to the utility of reading and constructing data visualizations as well as physical forms of data entry. Beyond that, the work highlights design recommendations for supporting data visualization literacy engagement within museums in connection with these larger aspects:

- The opportunity to arrange and rearrange data across a range of visualizations can support identifying single data records as well as comparisons of data records.
- Facilitating opportunities to question data patterns, including mismeasurements and who is not represented in data visualizations, can facilitate critical engagement with data visualization literacy.
- Making it possible to sense data together with others can create a social and shared engagement toward correcting faulty understanding of reading data visualizations.

Our study highlights that data visualization literacy can be a physical and social activity. The shared reading and creation of data visualizations suggests the need to design for group data entry alongside personal data entry to support intergenerational engagement in data visualization literacy. In conjunction, opportunities for physical data entries seems to support clarifications of non-intuitive data readings. Overall, the findings inform the design of data visualization literacy related science museum exhibits that are less facilitated than the Walk exhibit. Together, the design recommendations could support the creation of science museum exhibits that support the meaningful engagement data visualization literacy.

References

Acker, A. and Bowler, L. (2018), "Youth data literacy: teen perspectives on data created with social media and mobile devices", in *51st Hawaii International Conference on System Sciences*, pp. 1923-1932.

Bhargava, R., Deahl, E., Letouze, E., Noonan, A., Sangokoya, D. and Shoup, N. (2015), "Beyond data literacy: reinventing community engagement and empowerment in the age of data (data-pop alliance)", White paper, Harvard Humanitarian Initiative, MIT Media Lab, and Overseas Development Institute, Boston, 1 September.

Börner, K. (2015), *Atlas of Knowledge: Anyone can Map*. MIT Press, Cambridge, MA.

Case, R. (1996), "Changing views of knowledge and their impact on educational research and practice", in Olson, D.R. and Torrance, N. (Eds), *The Handbook of Education and Human Development: New Models of Learning, Teaching and Schooling*, Blackwell, Malden, MA, pp. 75-99.

Dancstep, T., Gutwill, J.P. (2019), *Towards disrupting the status quo of exhibit design: A data reanalysis to support future research*. [Viewed on 10 April 2020], available at: www.exploratorium.edu/sites/default/files/pdfs/DisruptingExhibitDesign_DancstepGutwill_FINALMay2019.pdf

Finzer, W. (2013), "Technology innovations in statistics education",*Technology Innovations in Statistics Education*, Vol. 7, pp. 1-9.

Forbes (2018). *How data visualization supports communication*. [Viewed on 10 April 2020], available at: www.forbes.com/sites/evamurray/2019/01/28/how-data-visualization-supports-communication

Gwilt, I., McEntaggart, P., Levick-Parkin, M., Wood, J. (2019), "Enhancing museum visits through the creation of data visualisation to support the recording and sharing of experiences", *Communities, Archives and New Collaborative Practices*, Vol. 123.

Hardy, L., Dixon, C. and Hsi, S. (2020), "From data collectors to data producers: shifting students' relationship to data", *Journal of the Learning Sciences*. Vol. 29 No. 1, pp. 104-126.

Hinrichs, U., Schmidt, H. and Carpendale, S. (2008), "EMDialog: bringing information visualization into the museum", *IEEE transactions on visualization and computer graphics*, Vol. 14 No. 6, pp. 1181-1188.

Inhelder, B. and Piaget, J. (1958), *The growth of logical thinking from childhood to adolescence: an essay on the construction of formal operational structures*, Basic Books, New York, NY.

Kahn, J. (2020), "Learning at the intersection of self and society: the family geobiography as a context for data science education", *Journal of the Learning Sciences*, Vol. 29 No. 1, pp. 57-80.

Konold, C., Higgins, T., Russell, S. and Khalil, K. (2015), "Data seen through different lenses", *Educational Studies in Mathematics*, Vol. 88 No. 3, pp. 305-325.

Lee, V.R. (2019), "On researching activity tracking to support learning: a retrospective", *Information and Learning Sciences*, Vol. 120 Nos 1/2, pp. 133-154.

Lee, V.R. and Dubovi, I. (2020), "At home with data: family engagements with data involved in type 1 diabetes management", *Journal of the Learning Sciences*, Vol. 29 No. 1, pp. 11-31.

Mallavarapu, A., Lyons, L., Uzzo, S., Thompson, W., Levy-Cohen, R. and Slattery, B. (2019), "Connect-to-connected worlds: piloting a mobile, data-driven reflection tool for an open-ended simulation at a museum", in *2019 CHI Conference on Human Factors in Computing Systems*, pp. 1-14.

Papert, S. (1980), *Mindstorms: Children, Computers, and Powerful Ideas*, Basic Books, New York, NY.

Papert, S. (1993), *The Children's Machine: Rethinking School in the Age of the Computer*, Basic Books, New York, NY.

Philip, T.M., Olivares-Pasillas, M.C. and Rocha, J. (2016), "Becoming racially literate about data and data-literate about race: data visualizations in the classroom as a site of racial-ideological micro-contestations", *Cognition and Instruction*, Vol. 34 No. 4, pp. 361-388.

Resnick, M. (2002), "Rethinking learning in the digital age", in Kirkman, G. (Ed.), *The Global Information Technology Report: Readiness for the Networked World*, Oxford University Press, New York, NY, pp. 32-37.

Roberts, J. and Lyons, L. (2020), "Examining spontaneous perspective taking and fluid self-to-data relationships in informal open-ended data exploration", *Journal of the Learning Sciences*, Vol. 29 No. 1, pp. 32-56.

Rubin, A. (2020), "Learning to reason with data: how did we get here and what do we know?", *Journal of Learning Sciences*, Vol. 29 No. 1, pp. 154-164.

Schettino, P. (2013), "Emotions, words and colors: a strategy to visualize and analyze patterns from visitors' narratives in museums", in *2013 17th International Conference on Information Visualisation*, *IEEE*, pp. 551-554.

Schneider, B., Wallace, J., Blikstein, P. and Pea, R. (2013), "Preparing for future learning with a tangible user interface: the case of neuroscience", *IEEE Transactions on Learning Technologies*, Vol. 6 No. 2, pp. 117-129.

Stornaiuolo, A. (2020), "Authoring dataStories in a media makerspace: adolescents developing critical data literacies", *Journal of the Learning Sciences*, Vol. 29 No. 1, pp. 81-103.

Strohmaier, R., Sprung, G., Nischelwitzer, A. and Schadenbauer, S. (2015), "Using visitor-flow visualization to improve visitor experience in museums and exhibitions", *Museums and the Web (MW2015)*.

Wilkerson, M.H. and Polman, J.L. (2020), "Situating data science: exploring how relationships to data shape learning", *Journal of the Learning Sciences*, Vol. 29 No. 1, pp. 1-10.

Wise, A.F. (2020) "Educating data scientists and data literate citizens for a new generation of data", *Journal of the Learning Sciences*, Vol. 29 No. 1, pp. 165-181.

Corresponding author
Kylie Peppler can be contacted at: kpeppler@uci.edu