Introducing Data Literacy in the Classroom using Sound Exploration Tools

Nagarajan Akshay
Department of Cognitive Science
University of California San Diego
La Jolla, USA
anagaraj@ucsd.edu

Victor Minces
Department of Cognitive Science
University of California San Diego
La Jolla, USA
vminces@ucsd.edu

2023 IEEE World Engineering Education Conference (EDUNINE) | 979-8-3503-2050-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/EDUNINE57531.2023.10102886

***Abstract*—Data literacy is becoming increasingly important for many professions due to the growing reliance on data-driven systems in the workplace. This means that children in school today need to be able to understand and work with data in order to be prepared for their future careers. However, many schools are having difficulty finding good resources for teaching data science. The authors of this paper describe the online tools and a workshop they developed to introduce students to working with data in a hands-on way, using sound and multimodal interaction. The initial feedback from participants indicated that the workshop helped improve their understanding of the practical applications of data and statistical operations.**

***Index Terms*—data science, education, multimodal learning, STEAM, online learning**

I. Introduction

The ubiquity of data science methods utilized in various fields makes it an indispensable skill set in the new world. The term data scientist describes those with experience in organizing and analyzing data and writing code to manipulate it, distinguishing them from those with more theoretical expertise [1]. The job roles of the future creates high demand for data scientists who have the ability to derive novel insights using data representations and its manipulation [2]. This demand for data science expertise in diverse domains makes data literacy an essential literacy of the 21st century workplace.

Jeannette Wing defines data science as “the study of extracting value from data” [3]. Freely accessible online datasets provide real world data that can be leveraged readily to educate students in value generation using data driven methods. Data science education in the classroom can also extend beyond canned data sets to include data collection, processing and utilization in value added contexts [4].

Kandel et al. describe the term ”data wrangling” as the various complex transformations a data scientist makes to bring data into a “credible and usable” form [5]. Data wrangling includes various various operations, data acquisition, cleaning, filtering, and conversion, among others [6]. There is demand for inclusion of data literacy in middle and high education in different regions of the world [7] [8]. An assimilation of data science concepts within a narrow definition can be a detriment. Oner way to avoid this is to create curriculum that promotes engagement with real world data.

Traditional data science programs are designed to serve undergraduate or early graduate students at the university level. These data science programs can be easily adapted to a school level data literacy curriculum by choosing the appropriate methodology and extending them across different disciplines [9]. Yet, very few approaches have leveraged on the arts and constructionism to teach data literacy [10] [11]. There is gap to be addressed here in inclusion of data literacy programs that allow for expression and engagement in school education.

The data literacy workshops described in the paper explores a data driven curriculum that offers a broad introduction utilizing STEAM-based activities. The workshops has been introduced in India in grades 9 through 12 utilizing activity-rich approaches for data wrangling. Various activities such as sound exploration, hands-on making, image manipulation and data visualizations were introduced as part of the data literacy curriculum. A data literacy assessment was conducted before and after the above mentioned activities which indicated participants had positive perception and increased comprehension of data operations and its real world applicability after the workshop.

II. Related Work

The foundations for being well-equipped to thrive in a data-driven society lies in inclusion of data literacy in school. Being data literate requires gaining a broad set of skills to comprehend, manipulate and utilize data to solve real world problems. Different programs to introduce data literacy to children around the world focus on various facets of working with data. Data literacy can be situated in an inquiry-based approach to probe data and determine good questions that can be posed for a given dataset [6]. Others like TuvaLabs provides access to large datasets, and use graph-based visualizations and computational models to aid data driven experimentation [12]. The MAINE daata literacy project is another example that provides a data literacy framework using online data sets to teach statistical concepts like central tendency, variability etc. [13]. The focus here is to introduce data visualizations and working with complex datasets. The Urban Data School takes a different approach to foster young makers who can use utilize data, in this context from a smart city, to create solutions like phone apps to solve urban problems [14].

All the above mentioned approaches focus on providing an understanding of what a real-world problem is and how it can be addressed to solve problems. Although this is an impactful way to introduce data literacy, its centered more on children who are driven to problem solve, and is not inclusive of those who enjoy self expression in other ways. That said, like the program described in this paper, they all leverage the power of experience before explanation situating these approaches in a maker-centric environment that foster more effective learning [15].

The data literacy workshop described in this paper attempts to leverage the joy of self expression with sound exploration and other multi-modal tools. Sound forms an engaging and creative medium for students to explore data concepts and gain an intuitive understanding of various operations.

Fig. 1. Children explore the Spectrogram music tool (example from later study conducted in Tumsong village, Darjeeling)

III. Methodology

The hypothesis we were testing was that students at middle and high school level will score higher on a data literacy test after exposure to our data literacy workshops involving multimodal data driven activities.

To test this hypothesis, a pilot workshop was conducted in a middle school in south India. The workshop was 18 hours long, spread over 3 days and divided into four sections covering various aspects of data operations described below. The introductory session focused on a discussion with the students on their understanding of data science and how data operation manifested in their daily activities. It was not surprising that the students responses indicated that they associated data operations with number systems and math. At first, they were confused when asked to identify sources of data around them. The question prompted an active group exploration on what data means and what sources of data can be identified by the students.

The philosophy of this workshop was to create broad definitions for data and its manifestations that would allow students to appreciate the various ways in which data driven methods influence our lives. The data sources were divided into categories based on the way they were perceived: auditory (heard), visual (seen), tactile (touched), and language (spoken or written).

A. Auditory Data Representation: Spectrogram

In the first part of the auditory portion of the lesson, the students were asked to make a humming noise by feeling the vibrations in their throats as they hummed. This activity was followed up with a question prompting a discussion about the differences in pitches, volumes, and temporal variations produced by different individuals. To further illustrate this concept, a demonstration was given on how the frequency of vibrations can be modulated by thin and thick elastic bands. The students also watched a video on how the frequency of the human voice can be modulated by inhaling thin and thick gas. The conversation then turned to the topic of how we hear sounds, the mechanics behind it, and the audible range of human perception. The students also learned about the incredible hearing capacity of the great wax moth that has evolved to counter predators such as insect eating bats. The workshop included activities that involved exploring data through sound using an educational program and freely available online tools focusing on the science of sound and music [16] [17]. These activities used sound as a way to interact with data operations. As an initial exploration of sound, the students could strike pipes of different sizes held in one hand to produce sounds. It was observed that a sustained sound would only be produced when the pipes were suspended between their fingers positioned at roughly 22.5% of the length of the pipes where the sound vibration nodes of the pipe are located. This intrinsic physical property of pipes puzzled the students.

Later the students were introduced to the Spectrogram online tool created by Minces et al. This tool and the Musicscope described below are freely accessible through the www.listeningtowaves.com website [18]. The Spectrogram offers a visual monitor that displays the various frequency components of the recorded sound over time (in our case the student's voice) as a spectrum that can be explored by multitouch interactions to replay the individual component frequencies back to the student. This is a powerful tool that creates deeper curiosity in the child to engage actively with the data frequencies in their own voice and provides a distinct visual representation of real-life time series data.

B. Data Manipulation: Sound Wrangling

From the previous session, the students had learned about the various parameters of sound, yet they did not fully understand the nuanced effects manipulating frequencies, amplitudes and time scales created in the overall sound. It must be noted that it is essential to bridge the gap in understanding between raw signals(sound) and meaningful data representations(music). To provide a foundation for the activity that was to follow, the students watched how Yosi Horikawa, a creative music composer used soundscapes recorded in nature and edited them down by modifying sound parameters to create music [19].

Fig. 2. Students exploring frequency and temporal variations and producing music mashups on the musicscope

Fig. 3. Data conversion from light to sound using photo-resistors

The students were then introduced to Musicscope, an online oscilloscope that can visualizing recorded audio in real-time [18]. The students quickly became adept at using the tool and were soon recording audio bytes, manipulating the signals, and discovering the complex relationship between various sound parameters, shown in figure 2. A group of students got creative and created a sound clip with claps, table taps and beat-boxing. This inspired other groups who also created new sound compositions. The students then had an idea to set up a makeshift recording station to record their creations on a mobile application. The students gained a sound appreciation for data manipulation through these enjoyable activities. This session also set the stage for an exploration of data conversion and signal transduction.

C. Data Conversion: Sound-light transduction

In this activity, students learned how to build electrical circuits. They had little prior knowledge of circuits and only a basic understanding of electric current and battery sources. During this activity, they followed a circuit diagram to make a breadboard photoresistor circuit that outputs audio signals, see figure 3. When light shines on the photoresistor, it produces sputtering sounds that can be heard on headphones connected to the audio output of the circuit. From this hands-on activity, students learned how analog data can be transmitted via different mediums and can be converted from light into sound and vice versa.

D. Data Smoothing: Handling Noise and Outliers

Data smoothing is a function that is utilized to smooth out any noise effects and outliers in a system. Students were introduced to the concept of central tendency that can be calculated for a dataset. An intuitive example to illustrate central tendency is group biological behaviors by species such as fish or birds. Birds in flight and swimming schools of fish have a localized measure of the central tendency of the global population and use this to maintain their shape in spite of large numbers. When the group shifts directions, a few group members stray from the group and take time to reunite with the group. In this moment, they form the outliers in the group. In a similar way, when a group of people vocalize in unison in response to a baseline pitch, they will drift towards the central tendency of the sound. Data smoothing was introduced as one way to reduce noise and eliminate outliers.

The next activity was to explore tactile just noticeable difference (JND) of the human body. Students were grouped in pairs of experimenters and perceivers. With the perceiver's eyed closed, the experimenter touched the perceiver's arm with two pencils placing them at different distances from each other in each trial. The perceiver made a guess of how many pencils were touching them in each trail. For each trial, the experimenter noted down the number of pencils that touched the perceiver, the approximate gap between the touched points, and the perceived number of touches. After collecting data from n instances, the roles were switched and data was collected again. At the end of the experiment, the students coded their correct and incorrect responses with positive and negative scores, respectively. The averages of the scores were then entered into a spreadsheet and plotted. The plot was used to discuss the outliers and central tendency in the class, and the possible causes of these outliers were discussed.

E. Data Encryption

To explore data encryption, a visual medium approach was taken. The class discussed human vision and how visual information is processed by the brain. We used video aids to help explain the process and provided examples, such as the differences in the range of colors that humans and mantis shrimp can perceive. As an activity, we introduced the concept of steganography, a method of encrypting one image within another to conceal it. We used a free online tool to demonstrate how to perform steganography and decrypt the hidden image.

F. Data Visualization

During this session, the students learned about the importance of visualizing data to understand patterns and highlight specific aspects. They had previously learned how to create pie and bar charts, but did not fully understand when to use each type. In this session, they learned about various types of data visualizations, including pie, column, scatter, histogram, map, and trend plots, and how they can be used to represent not only quantitative data, but also spatial and temporal variations.

As a hands-on activity, the students were given pouches of M&M candies and tablulated with the count of candies of each color. The mean and standard deviation of the taabulated data was calculated and students explored different types of representations to visualize the data. The goal was to determine the ideal type of data plots that carried most information.

G. Data Relationships

As the students continued to explore data and relationships through visualizations, the topic of causality arose. They discussed how it is sometimes possible to infer causal relationships from data, but it is important to be cautious and not assume causation based on correlation alone. The concept of ”Correlation does not imply causation” was introduced to help emphasize the need to consider other factors that may be influencing the relationship between variables. Examples were discussed to illustrate how it is possible to mistakenly assume causation based on correlations that are found in data.

H. Data Literacy Test

The data represented below is from a workshop conducted for a class of 36 students aged 14 to 15 years old, studying in 9th grade in a rural school in India. Among the participants, 27 students (18 girls and 9 boys) agreed to participate in the data literacy test. The assessment included a pre-questionnaire before the first workshop session and a post-questionnaire filled out after the last session. The assessment tool used was the Data Literacy Test [20], which included 10 questions data topics such as: the test taker’s general relationship with data, understanding of data concepts and the utility of visualizations. The results of the assessment are based on the responses of the 27 students.

The data literacy test questions had the following themes(for specific questions go to [20]):

1) Confidence working with graphs
2) Extend of data use
3) Frequency of data use
4) Relevance of data as a literacy
5) Ability to tell stories with data
6) Comprehending statistical measures: central tendency
7) Comprehending statistical measures: outliers
8) Does correlation imply causation?
9) Self-perception around utilizing data visualizations
10) Relevance of utilizing predictive analytics

IV. Results and discussion

Before the workshop, the students had a limited understanding of and relationship with data, according to the results of a data literacy test. They has a vague understanding of data concepts such as outliers and central tendency, and only a few of them had dabbled with any data visualization. The post-test plot suggest that the students either improved or maintained their understanding of data after the workshop, with low levels of standard deviation, as shown in Figure 4.

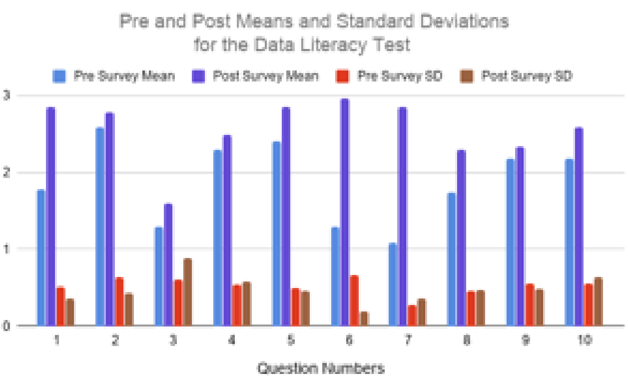

Fig. 4. Graph chart that shows the means and standard deviations of the responses to the data literacy test

According to the average of response to the first question, the workshop seems to have increased the students’ interest and confidence in working with data visualizations. In the post-test, most of the students responded that they build data visualizations, but do not feel as skilled as they would like to be. One student even said they want to improve their skills. In contrast, the common response in the pre-test indicated a mix of excitement and fear with the subject. This suggests that the workshop had a positive impact on the students’ attitudes towards data visualizations.

There was a significant increase in the mean responses to questions about outliers and central tendency in the post-test. The post workshop report suggests that students (now perceive that they) utilize central tendency regularly. From pre-test results, a majority of students did not know what outliers were, but in the post-test, the mean student response indicated that they had found outliers to be a valuable source of insight that helped them delve deeper into data. This suggests that the workshop effectively taught the students about central tendency and outliers.

V. Conclusion

Before participating in the data science workshop, the students had limited understanding of data representation, only knowing how to create basic plots. This limited their ability to apply their data literacy skills in their daily lives and reduced their appreciation for its usefulness. The data literacy workshops provided a broad definition of data, covering sensory sources such as hearing, vision, and touch. Interactive and constructionist activities such as the ones used in the workshop can help children form a positive association with data science.

The data literacy workshops is a good introduction for students interested in artificial intelligence, which involves creating AI models to classify real-world data. It can also improve reasoning and understanding of data and influence better decision-making for the individual.

VI. Acknowledgments

We would like to acknowledge the students and administrators at Amrita Vidyalayam schools in Kerala, India and the people of Tumsong village in Darjeeling, India for their participation in the program.

References

[1] R. A. Irizarry, "The role of academia in data science education," 2020.
[2] M. Shields, "Information literacy, statistical literacy, data literacy," *IASSIST quarterly*, vol. 28, no. 2-3, pp. 6–6, 2005.
[3] J. M. Wing and D. Banks, "Highlights of the inaugural data science leadership summit," 2019.
[4] K. Williamson, V. Bernath, S. Wright, and J. Sullivan, "Research students in the electronic age: Impacts of changing information behavior on information literacy needs," *Communications in Information Literacy*, vol. 1, no. 2, p. 4, 2008.
[5] S. Kandel, J. Heer, C. Plaisant, J. Kennedy, F. Van Ham, N. H. Riche, C. Weaver, B. Lee, D. Brodbeck, and P. Buono, "Research directions in data wrangling: Visualizations and transformations for usable and credible data," *Information Visualization*, vol. 10, no. 4, pp. 271–288, 2011.
[6] A. Wolff, D. Gooch, J. J. C. Montaner, U. Rashid, and G. Kortuem, "Creating an understanding of data literacy for a data-driven society," *The Journal of Community Informatics*, vol. 12, no. 3, 2016.
[7] E. Deahl, "Better the data you know: Developing youth data literacy in schools and informal learning environments," 5 2014.
[8] J. Vanhoof, G. Verhaeghe, P. V. Petegem, and M. Valcke, "Improving data literacy in schools: Lessons from the school feedback project," in *In Data-based Decision Making in Education*. Springer, Dordrecht, 2013, pp. 113–134.
[9] S. Kross, R. D. Peng, B. S. Caffo, I. Gooding, and J. T. Leek, "The democratization of data science education," *The American Statistician*, vol. 74, no. 1, pp. 1–7, 2020.
[10] R. K. E. B. G. C. Bhargava, Rahul and C. D'Ignazio, "Data murals: Using the arts to build data literacy," *The Journal of Community Informatics*, vol. 12, no. 3, 2016.
[11] C. Matuk, K. DesPortes, A. Amato, M. Silander, R. Vacca, V. Vasudevan, and P. J. Woods, "Challenges and opportunities in teaching and learning data literacy through art," in *Proceedings of the 15th International Conference of the Learning Sciences-ICLS 2021.* International Society of the Learning Sciences, 2021.
[12] L. Reiten and S. Strachota, "Promoting statistical literacy through tuva," *The Mathematics Teacher*, vol. 110, no. 3, pp. 228–231, 2016.
[13] W. M. Schlager, *Assessing Students' Understanding of Variability and Graph Interpretation Through an Authentic Science Investigation*. The University of Maine, 2017.
[14] A. Wolff, J. J. C. Montaner, and G. Kortuem, "Urban data in the primary classroom: bringing data literacy to the uk curriculum," *The Journal of Community Informatics*, vol. 12, no. 3, 2016.
[15] S. Papert and I. Harel, "Situating constructionism," *constructionism*, vol. 36, no. 2, pp. 1–11, 1991.
[16] V. Minces, J. Iversen, G. Musacchia, C. Zhao, and A. Chiba, "Music, cognition and education," pp. 167–177, 2019.
[17] V. Minces, A. Khalil, I. Oved, C. Challen, and A. Chiba, "Listening to waves: Using computer tools to learn science through making music," 07 2016, pp. 3844–3852.
[18] V. Minces and A. Khalil, "Oscilloscope listening to waves." [Online]. Available: https://listeningtowaves.com/oscilloscopeinfo
[19] Y. Horikawa, "Bandcamp." [Online]. Available: https://yosihorikawa.bandcamp.com/
[20] "The data literacy project- online assessment." [Online]. Available: https://thedataliteracyproject.org/