# There's Data all around you: Improving data literacy in high schools through STEAM based activities

Nagarajan Akshay
Ammachi Labs
Amrita Vishwa Vidyapeetham
Amritapuri, Kerala, India
akshay.nagarajan@ammachilabs.org

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

Anu Vazhayil
Ammachi Labs
Amrita Vishwa Vidyapeetham
Amritapuri, Kerala, India
anu.vazhayil@ammachilabs.org

Vaishnavi Gopalasamy
Dept. of Computer Science and Engineering
Amrita Vishwa Vidyapeetham
Amritapuri, Kerala, India
vaishnavigopalasamy@gmail.com

Rao R. Bhavani
Ammachi Labs
Amrita Vishwa Vidyapeetham
Amritapuri, Kerala, India
bhavani@ammachilabs.org

## ABSTRACT

Data literacy has taken a front seat in present day conversations on education reform primarily due to the need for education on disruptive technologies such as Artificial Intelligence and Internet-of-Things that are rapidly transforming the future of work and life. School systems worldwide have already included data literacy several years ago in their curriculum, still the definition of data and the activities utilized to teach data handling are verily outdated and seek change to reflect the new relationship we are starting to form with data. This paper discusses a workshop conducted for data literacy education in schools. The hands-on activity based approach taken in the workshop seeks to offer a broad definition to data along the lines of real world application in terms of our human sensory perception of audition, vision, and haptics.

## CCS CONCEPTS

• **Applied computing** → **Mathematics and statistics**; **Collaborative learning**; Sound and music computing; • **Mathematics of computing** → *Exploratory data analysis.*

## KEYWORDS

Data literacy, sound production, data conversion, data manipulation, tactile perception, data visualization, statistical literacy

**ACM Reference Format:**
Nagarajan Akshay, Victor Minces, Anu Vazhayil, Vaishnavi Gopalasamy, and Rao R. Bhavani. 2020. There's Data all around you: Improving data literacy in high schools through STEAM based activities. In *Proceedings of Fablearn Asia 2020.* , Bangkok, Thailand, 4 pages.


*Fablearn Asia 2020, January 10- 12, 2020, Bangkok, Thailand*


## 1 INTRODUCTION

The pervasiveness of data science has deemed it the new currency of the age. Data-driven job roles of the future demand forming a new relationship with its representation and manipulation in the digital age[6]. Children of this century are born into a time in which a plethora of data is available at their fingertips mainly due to social media and mobile applications. Yet the definition of data as described in the digital age has inevitably boxed away its analog real-world existence. Inclusion of digital fabrication and electronics based activities, opens up avenues for children to gain an appreciation and intimate experience of real-world data collection, it's processing and utilization in meaningful contexts[9].

When looked at as a science, data and its operations have several pragmatic dimensions. The various aspects of data science namely, data handling, acquisition and cleaning, filtering, conversion, data manipulations and statistics to name some[10]. In the context of high school education, school curricula worldwide are witnessing an inclusion of Artificial Intelligence(AI) and data literacy in the school curriculum[3][8]. A narrow or weak comprehension of data science can hamper the children's understanding of even other subjects such as AI. Furthermore, data literacy education in schools should also be delivered with close association with handling real-world data. Few approaches have found creative in-ways to offer data literacy in creative ways such as a project that explores teaching data literacy through data murals[2].

The intention of the data science workshop described in the paper is to explore if a data science curriculum that broadens the school textbook definition of data science and introduces data handling through STEAM activities, can help school students form a meaningful connection between the theory of data science and its applicability. The workshop was conducted for grade 9 students, offering a broad definition for data and utilized an activity-rich approach such as data comprehension in the context of sound production and music composition, creating electronic circuits to convert light to sound, image-based data encryption and plotting data visualizations. A data literacy test was conducted before and after the workshop and results indicate an increase in appreciation and understanding of data and its operations.

## 2 METHODOLOGY

The workshop lasted for 3 days(18 hours) and was divided into four thematic sections described below. The workshop started with a discussion on what the students understood by the term and their association with it in their daily lives. Not surprisingly, the students struggled to associate anything beyond numbers and math with data and looked puzzled when asked to point at sources of data in their immediate surroundings. This led to the intended discussion on data sources. A broad classification was placed by the instructors as a suggestion in terms of the body's sensory perception sources, namely, auditory data, visual data, tactile data, and language. The students discussed a few examples of data types from these sources and concluded that this definition of data and its types can be accepted.

### 2.1 Auditory Data Production

The auditory section started with a discussion of how data is produced by humans, for this, we asked the students to produce a loud humming sound while touching their throats and feeling the vibrations. They were then asked to take a deep breath and hum loudly together. A discussion ensued on what they were doing, as a group, and with the individual differences, leading to responses on variable pitches, higher or lower volumes, and temporal variations being produced by different individuals. This led to a demonstration of frequency modulation during vibrations of thinner and thicker elastic bands. This knowledge was further explained with a video on how inhaling a thicker and thinner gas can modulate the frequency of the human voice. Now the conversation shifted to: How do we hear sounds and in what form does data flow to the brain? followed by videos and discussions on the mechanics of hearing and the range of sound audible to humans and the incredible hearing adaptation of the great wax moth that evolved this capacity to evade its predator, the insect bat. Discussion on auditory forms of data production was interspersed with activities designed by Listening to Waves, an education program on the science of music lead by Dr. Victor Minces, at the University of California in San Diego [4][5]. In the first set of activities, students were given pipes of different diameters and were asked to hold them in one hand and strike them to create a sustained sound. A sustained sound is produced only when the pipes are held at the vibration nodes, which are located at 22.5% of the length of the pipes. This physical phenomenon baffled students.

### 2.2 Data manipulation: Exploring the Musicscope

Although the students learned about the parameters that are used to define Sound, they still lacked the distinctive understanding of the effects of each parameter, and the audio variations that can be heard on manipulating varying time scales, frequencies, and amplitudes. As a precursor to the next activity, a video on the popular music producer Yosi Horikawa was shown to the students where they watched how Yosi used natural sounds recorded from his surroundings to produce imaginative music scores [11].

Musicscope is an online oscilloscope developed by researchers at the University of California, San Diego for visualizing the audio signals captured using a microphone [7]. This tool is freely available through the listeningtowaves.com website. The students picked up the tool very easily and in no time were busy recording audio bytes, manipulating the signals, and were able to figure out the intricate relationship between time scales and frequencies as seen in figure 1. One team get so involved that they started recording beats by clapping and tapping on the table coupled with background voices. Other teams observed them and also started creating their musical renditions. An idea cropped up to run a recording studio side by side during this activity, and a makeshift "studio" was set up in another room where students now scampered to record their sound productions on a mobile phone recording app. This fun-filled activity shaped up their understanding of sound and data manipulation and prepared them for the conversation relating to statistical literacy.

**Figure 1: Students exploring frequency and temporal variations and producing music mashups on the musicscope**

**Figure 2: Students learn data conversion by making photoresistor circuits to convert light to sound**

## 2.3 Data Conversion: Producing Sound from light using a Photoresistor circuit

The next activity involved circuit building. The students had no prior exposure to electrical circuits and possessed a cursory understanding of direct currents and battery sources. They learned to use a breadboard and to follow a circuit diagram to make connections to build a simple photoresistor circuit that connects to an audio jack on its output end as shown in figure 2. When light is shown to the photoresistor it creates sputtering sounds that play over a speaker connected to the audio jack. This simple yet powerful example of light converting to sound became a good experiment to demonstrate signal conversion and analog data transmission to the students.

## 2.4 Central Tendency and Outliers

The next session sought to introduce statistical literacy through experiments in which students have demonstrated the utility of central tendency in the natural world and man-made systems. The students were shown a video on swarming behaviors exhibited by animals, such as by a large school of fish or the murmuration of birds in flight. It was discussed as to how these animals maintained their formations even when in such large numbers, the answer is an indication of implicit measurement of central tendency. Similarly, when the swarm changes direction, some members get left behind and scamper to catch up much later, these at that moment constitute the outliers in the swarm.

*2.4.1 Tactile Just Noticeable Difference experiment.* The students were introduced to the concept of tactile just noticeable difference (JND) and then asked to pair up for this activity, one as an experimenter and the other as the perceiver. The experimenter was asked to probe the perceiver's forearm with pencils while the perceiver sat with their eyes closed. The experimenter would touch the perceiver's forearm with the shaved end of either one or two pencils closer or farther from each other. The perceiver with closed eyes had to guess the number of pencils touching their skin. For each instance, the experimenter was to record in a table the number of pencils used for probing, the distance between the pencil points, and the response on the number of pencils felt by the perceiver. After recording 10 data points, they had to swap roles and record data again. Once the class had completed the experiment, they had to compute the average score of the collected data by coding their right and wrong response with causal scores. these averages were then entered on an excel and plotted to demonstrate the central tendency of the class. Further, the outlier perceivers in the class were also noted and what could have led to these outliers was discussed.

## 2.5 Visual perception and Steganography

This session focused on the sensory aspect of vision, and a discussion similar to that with hearing was taken up with the students with video aids to try to describe the process of visual data perception and transmission to the brain. The range of visual perception for humans and in contrast the unusually large range of colors perceived by the mantis shrimp was offered as an example. The activity in this session was based on steganography, a data encryption technique that encrypts one image within another to hide it. This encrypted image can then be decrypted by the receiver to view the original hidden image. We utilized a free online tool to carry out the steganography activity provided by the popular mobilefish.com tutorial website.

## 2.6 Data Visualizations

This session focused on teaching students about the utility of visualizing data to comprehend data, identify unobvious patterns and to emphasize specific aspects of interest within the data or its meaning. The students had prior experience plotting a pie chart and a bar chart on spreadsheets but were able to articulate in what situations each type of data visualization could be best utilized. They then learned how visualizations can be useful to represent not just quantitative value variations in data using pie, column, and scatter plots, but also can be used to visualize spatial and temporal variations using other types of data visualizations such as histograms, maps, trends.

As an activity to practice creating visualizations, pouches of MM candies were given to each student group and they were asked to first create a table containing the number of all the colors of candies they had in their pouches. Once the tables were ready, they explored the central tendency of the data by calculating means and standard deviations and then proceeded to try out different types of charts to plot the data. The task was to determine what they found was the best type of chart to represent their data and then present why they picked that chart to the class.

## 2.7 Data Relationships: Correlations and Causation

Further exploration of data and their relationships as observed in their visualizations naturally led to a discussion on the causal relationship that is sometimes apparent from the data and under what conditions can these causalities be accepted as true. Various examples were discussed on causation that is often hurriedly and incorrectly implied from correlations that are found in data, and the statistical phrase "Correlation does not imply causation" was discussed.

## 2.8 Data Literacy Assessment

A total of 36 students of ages between 14-15 years studying in grade 9 participated in the workshop, and among them, 27 students (18 girls, 9 boys) willingly participated in the data literacy assessment. The pre-questionnaire was administered at the beginning of the workshop and the post-questionnaire was administered after all the activities during the concluding session. The results report the findings measured from the 27 students' responses to the Data Literacy Test which was the tool chosen for this assessment [1]. The questionnaire posed 10 questions in broadly four categories, on the general relationship of the test taker with data, on the understanding of central tendency and outliers, on comprehension of data visualizations, and causation and correlation.

The questions in the data literacy test were as follows:

- How would you feel if you were confronted by this scenario? Which of the following best describes your experience with data?
- How often do you feel overwhelmed with data?

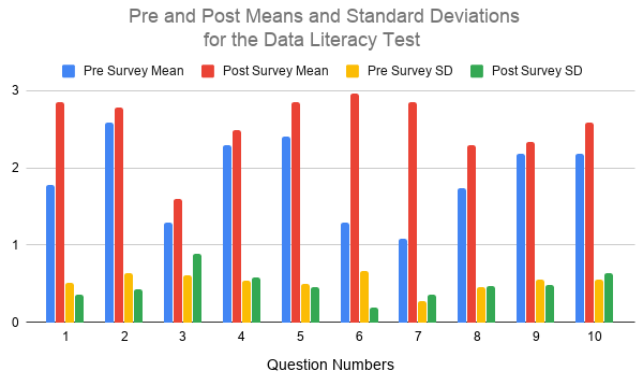


**Figure 3: Plot of Means and Standard deviations of the Pre and Post response on the Data Literacy Test**

- Do you agree that understanding data is as vital as reading or writing?
- How important is the ability to tell stories with data?
- Central tendency is utilized to measure a central or typical value in a data-set. Do you feel comfortable utilizing it?
- Outliers can be used to analyze data differently. How often do you use them?
- How often does correlation mean causation with data?
- What are your feelings on data visualization?
- Do you agree that predictive analytics allows us to see into the future by finding patterns and trends?

## 3 RESULTS AND DISCUSSION

The data literacy test conducted before the workshop suggests a weak understanding of and relationship with data among the students. Students reported not knowing what central tendency or outliers were, and only a few among them had an appreciation of data visualizations. The results as shown in figure 3 indicates that there is either an increase or equal response in the post-test when compared with the pre-test with low levels of stand deviation.

From the mean response to question 1, is it seen that students reported more interest and confidence in handling data visualizations after the workshop? Majority of students had chosen in the post-test to respond that, "I build them, but I'm not as good as I want to be" or "I'd really like to advance my skills", as compared the pre-test in which the common response was "I am torn between excitement and dread".

The maximum increase in mean responses in the post-test was for questions 6 and 7 on Central tendency and outliers. After the workshop, the students now report that they perceive that they "use central tendency on a daily basis". For the question on outliers, the mean student response indicates that "Outliers have been a great source of insight, allowing me to dig deeper", as opposed to in the pre-test where the mean response was "I have no clue what outliers mean".

## 4 CONCLUSION

The experience from the workshop indicates that although a primer to data science is taught in schools, they remain to a large extent limited to plotting numerical values that do not lend meaning to the students. This prevents the students from being able to apply this knowledge of data literacy to their daily lives and has also been observed to diminish their appreciation for its utility. The workshop offered a very broad definition of data and its handling touching upon its global sensory sources of audition, vision, and haptics. Through several hands-on activities, the students have reportedly gained a deeper understanding and finer relationship with various aspects of data and its processing. This work is also intended as a precursor to teaching students a curriculum on artificial intelligence, where they collect data from the real world and learn to create AI models to classify the data according to their needs. Having gained a broad view of data and its sources prepare the students better for learning AI and other technologies that are pervading our future lives.

## 5 ACKNOWLEDGMENTS

This project is a product of a deeply stirring conversation with Mata Amritanandamayi Devi, the Chancellor of Amrita University and an inspiring humanitarian who is popular all over the world. For her insights and inspiration, we owe her our gratitude. We also thank the principal, teachers, and students at Amrita Vidyalayam Puthiyakavu high school for supporting the conduct of this workshop.

## REFERENCES

[1] [n.d.]. The data literacy project. https://thedataliteracyproject.org/
[2] Ricardo Kadouaki Emily Bhargava Guilherme Castro Bhargava, Rahul and Catherine D'Ignazio. 2016. Data Murals: Using the Arts to Build Data Literacy. *The Journal of Community Informatics* 12, 3 (2016).
[3] Erica Deahl. 2014. Better the Data You Know: Developing Youth Data Literacy in Schools and Informal Learning Environments. https://doi.org/10.2139/ssrn.2445621
[4] Iversen J Musacchia G Zhao C Minces, V and A Chiba. 2019. Music, cognition and education. (2019), 167–177.
[5] Victor Minces, Alexander Khalil, Iris Oved, Cate Challen, and Andrea Chiba. 2016. Listening to Waves: Using Computer Tools to Learn Science Through Making Music. 3844–3852. https://doi.org/10.21125/edulearn.2016.1919
[6] Milo Shields. 2005. Information literacy, statistical literacy, data literacy. *IASSIST quarterly* 28, 2-3 (2005), 6–6.
[7] Khalil V, Minces A. [n.d.]. Oscilloscope Listening to Waves. https://listeningtowaves.com/oscilloscopeinfo
[8] Goedele Verhaeghe Peter Van Petegem Vanhoof, Jan and Martin Valcke. 2013. Improving data literacy in schools: Lessons from the school feedback project. In *In Data-based Decision Making in Education*. Springer, Dordrecht, 113–134. https://doi.org/10.1007/978-94-007-4816-3_7
[9] Kirsty Williamson, Vivienne Bernath, Steven Wright, and Jen Sullivan. 2008. Research students in the electronic age: Impacts of changing information behavior on information literacy needs. *Communications in Information Literacy* 1, 2 (2008), 4.
[10] Daniel Gooch Jose J. Cavero Montaner Umar Rashid Wolff, Annika and Gerd Kortuem. 2016. Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics* 12, 3 (2016).
[11] Horikawa Yosi. [n.d.]. Bandcamp. https://yosihorikawa.bandcamp.com/