Data for Healthy Communities

A Public Interest Pilot Course Designed to Develop K-12 Data Literacy

Emily Nutwell
The Ohio State University
Columbus, Ohio, USA
nutwell.1@osu.edu

Kelsey Badger
The Ohio State University
Columbus, Ohio, USA
badger.60@osu.edu

Jessica T. Kulp
The Ohio State University
Columbus, Ohio, USA
kulp.95@buckeyemail.osu.edu

ABSTRACT

Data analytics skills are in high demand in a wide variety of professions and data is increasingly present in our everyday lives. Considering this, educators should be equipped to prepare future data scientists, but also data literate citizens. This project engages high school students in the real-world practice of data analytics to promote the public interest. The research team is partnering with a local public STEM high school to design, evaluate, and publicly share an enrichment course that introduces students to data analytics in the context of public health. Integrating data science education and public health will enable students to engage in meaningful data-driven problem solving while promoting civic engagement. A key challenge in introducing data skills in K-12 is addressing barriers to access. To cope with these barriers, the course design incorporates the use of spreadsheets to engage students in computational thinking. This poster will present preliminary assessment results from the 3-week enrichment course conducted in January 2024. This project contributes to the ongoing discussion of K-12 data science education by demonstrating the value of contextualizing data science as a key competency for 21st century public service.

ACM Reference format:

Emily Nutwell, Kelsey Badger, and Jessica T. Kulp. 2024. Data for Healthy Communities: A Public Interest Pilot Course Designed to Develop K-12 Data Literacy. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education (SIGCSE 2024), March 20–23, 2024, Portland, OR, USA.* ACM, New York, NY, USA, 2 pages.
https://doi.org/10.1145/3626253.3635589

1 INTRODUCTION & BACKGROUND

As data becomes more prevalent in all aspects of society, introducing students to data literacy concepts in the high school years is becoming increasingly important. Data literacy can be understood as the ability to work with data through an inquiry process to address real-world issues, including knowledge and skills related to collection, preparation, analysis, visualization, communication, and critique of data and results [11]. Educators need to prepare students to be proficient in applying these skills across all aspects of their personal and professional lives, including as engaged citizens [4, 10].

Preparing a data literate citizenry also ensures a workforce that can apply data-driven problem solving to social and cultural problems, which are complex and typically involve several interconnected issues and constraints. Since data is present in every field, it is imperative that students looking to choose careers or majors for college understand how data is integrated in all professions, including public service roles that may not be historically associated with data or computing. Public interest technology is an emerging field of interdisciplinary technologists focused on applying data and technology in the service of social good [6]. The development of a workforce adept at understanding social problem contexts and data-driven approaches is increasingly urgent as the world grapples with global health threats such as pandemics, environmental hazards, the mental health crisis, and antibiotic resistance [5].

However, there are significant barriers preventing equitable access to data education across the country. Many of these relate to an ongoing debate about when and how data skills should be introduced in the high school curriculum. Due to the use of computational tools like Python and R in many popular data science curricula, exposure to data education is often contingent on the availability of computer science courses. Ohio ranks near the middle in the country in computer science course availability, with 48% of public high schools offering a class in computer science [2]. Yet only 3.6% of Ohio high school students took a computer science class during the 2021-2022 school year [3]. To improve access, it is necessary to both increase computer science course availability and to embed foundational data concepts across the curriculum.

Open educational resources (OERs) provide teachers a streamlined approach to experimenting with new material in existing courses. However, there is a dearth of OERs that do not rely on the use of advanced computational tools to introduce data concepts. One way to lower this barrier is to develop data science modules using spreadsheets, a more universal tool that can be used to teach fundamental computational concepts such as basic sequences, functions, and variables [7]. Spreadsheets offer a compelling opportunity to introduce data science concepts across the curriculum while eliminating a steep learning curve required for more specialized tools.

2 COURSE OVERVIEW

The Data for Healthy Communities (DHC) course is a pilot project of spreadsheet-based modules that combines the need for

SIGCSE 2024, March 20–23, 2024, Portland, OR, USA

ACM ISBN 979-8-4007-0424-6/24/03.
https://doi.org/10.1145/3626253.3635589

more accessible data science curricula with the values of public interest technology. We aim to broaden student horizons regarding the use of data for social good and to create open educational resources that introduce data science using approachable, spreadsheet-based methods.

DHC is a three-week enrichment course with approximately 15 instructional hours. It was developed in close collaboration with a diverse STEM high school whose student body is drawn from over two dozen Central Ohio school districts. It will be taught in January 2024 during a period where students are offered the opportunity to explore extra-curricular interests through an array of short elective offerings. The course will be evaluated using a pre/post-test that incorporates both cognitive assessment and attitudinal measures of student interest and perceived self-efficacy towards data science.

The course is designed to introduce high school students to data science concepts in the context of public health. These topics, such as high infant mortality, poor air quality, and the opioid epidemic are all prevalent problems in today's society. We hope to engage these students in understanding the connections between their own communities and technology in the public interest, helping them envision educational paths and career opportunities where they are empowered to address long-standing inequities through data-driven problem solving.

A backwards course design process was used to develop a series of modules, each of which introduces a skill tied to one or more learning objectives [9]. These objectives were created in part through evaluation of public health workforce development programs and in collaboration with a high school teacher. They are also mapped to one or more of the Ohio Department of Education's standards, including those for Data Science Foundations, Statistics and Probability, and Algebra. This breadth of mappings is intended to facilitate future transfer of the curriculum to use during the regular school year.

The first modules of the course have students explore public interest technology careers and consider how data can be used to explore health inequities in their communities. The remainder of the course modules guide students through the process of the data lifecycle using various sources of air quality data [1]. Students will collect readings using sensors and organize their data according to tidy data principles [8]. They will then be introduced to more complex, real-world data from the Canadian wildfires during the summer of 2023. Students will practice using spreadsheet functions and features for selecting, summarizing, and visualizing data.

Collectively, the skill-building modules demonstrate how an argument can be supported with data. This provides students an example to apply in the completion of a course project. The project features a fictional case-study about the use of county funds to improve public health through investment in a community garden. Students will be asked to recommend which neighborhood should be prioritized for the garden, using a real-world dataset of local indicators as evidence. To highlight the importance of effective visualizations, they will be asked to support their argument with a chart or graph. Following group presentations, the instructors will guide students through the implications of using data to persuade, rather than as definitive proof of a single correct response.

3 CONTRIBUTIONS & FUTURE WORK

This project fills an important gap in data science open educational resources (OERs) that are taught using accessible spreadsheet-based methods. Using this approach, teachers across the curriculum can engage students with data literacy concepts and basic computing skills outside of computer science courses.

It is also unique in its emphasis on problem-solving using open government data. Experience with messy real-world data develops needed familiarity with the challenges of authentic problem-solving. However, a key challenge is curating datasets such that they are accessible for secondary education while still retaining the complexity of the real-world. To expand this educational model, further work is needed to curate data appropriate to various public interest technology contexts. The hope is that the framework presented here will contribute to the development of a citizenry and workforce that is prepared to use data to solve complex societal challenges.

ACKNOWLEDGMENTS

This work is based upon work supported by New Venture Fund under NVF-PITU-The Ohio State Univers-Subgrant-018811-2022-10-01. We also gratefully acknowledge our colleagues Chris Orban and Wonjung Hwang who collaborated on course design.

REFERENCES

[1] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rüdiger Wirth. 1999. *CRISP-DM 1.0.* SPSS Inc.

[2] Code.org. 2023. *Support K-12 Computer Science Education in Ohio.* Promote Computer Science.

[3] Mike Duffey, John Wiseman, and Kelly Evans Gaier. 2022. *Report of the State Committee on Computer Science.* Ohio Department of Education.

[4] Engida Gebre. 2022. Conceptions and perspectives of data literacy in secondary education. *British Journal of Educational Technology* 53, 5 (2022), 1080–1095. DOI:https://doi.org/10.1111/bjet.13246.

[5] Olga Jonas, Ian Parry, Dan Chrisholm, Nick Banatvala, and Ramanan Laxminarayan. 2014. Global health threats of the 21st century. *Finance & Development* 51, 4 (December 2014). International Monetary Fund.

[6] Tara Dawson McGuinness and Hana Schank. 2021. *Power to the public: The promise of public interest technology.* Princeton University Press, Princeton, NJ.

[7] Marcos Román-González, Juan-Carlos Pérez-González, and Carmen Jiménez-Fernández. 2017. Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior* 72 (2017), 678–691. DOI:https://doi.org/10.1016/j.chb.2016.08.047.

[8] Hadley Wickham. 2014. Tidy data. *Journal of Statistical Software* 59, 10 (Sep. 2014), 1–23. DOI:https://doi.org/10.18637/jss.v059.i10.

[9] Grant Wiggins and Jay McTighe. 1998. *Understanding by design.* Association for Supervision and Curriculum Development, Alexandria, VA.

[10] Alyssa Friend Wise. 2020. Educating data scientists and data literate citizens for a new generation of data. *Journal of the Learning Sciences* 29, 1 (2020), 165–181. DOI:https://doi.org/10.1080/10508406.2019.1705678.

[11] Annika Wolff, Daniel Gooch, Jose J. Cavero Montaner, Umar Rashid, and Gerd Kortuem. 2016. Creating understanding of data literacy for a data-driven society. *The Journal of Community Informatics* 12, 3 (2016), 9–26. DOI:https://doi.org/10.15353/joci.v12i3.3275.