Data for Healthy Communities:

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

THE OHIO STATE UNIVERSITY

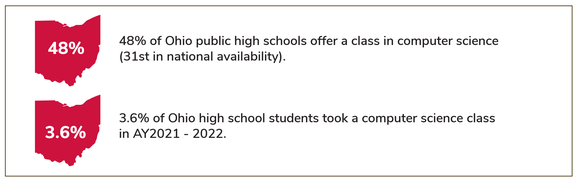
Overview

The Data for Healthy Communities (DHC) course is a pilot project of spreadsheet-based modules that combines the need for more accessible data science curricula with the values of public interest technology. This 3-week enrichment course was developed in partnership with a public STEM high school in Columbus, Ohio and was taught in January 2024.

Background

Many readily available data science curricula rely on the use of computational tools such as Python and R. This approach requires access to computer science courses which may be unavailable or underutilized.

Figure 1: Availability and participation in Ohio high school CS courses. Source: Code.org [1] and Duffey, Wiseman, & Evans Gaier [2]



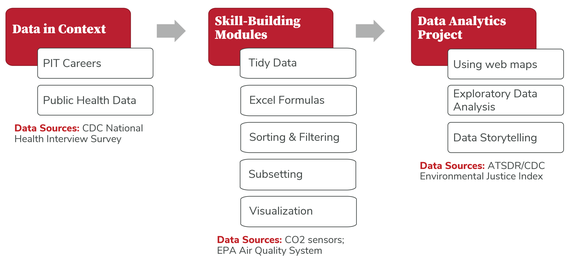
One way to lower this barrier is to develop data science curricula that use spreadsheets, a more universal tool that can be used to teach fundamental computational concepts such as basic sequences, functions, and variables [4].

Spreadsheet-based lessons can also be embedded across the curriculum to create new opportunities for students to encounter data literacy concepts. It is imperative that students understand how data is integrated into all professions, including public service roles that may not be historically associated with data or computing. One strategy for increasing this awareness is to contextualize data science education as part of public interest technology (PIT), an interdisciplinary field that unites technologists and humanists in the service of social good [3].

Course Overview

The course situates data science in the context of public health and is comprised of a series of skill-building modules that collectively model the technical and problem-solving skills needed for a class project. The project challenges students to use a real-world dataset of local indicators as evidence to create an argument for how neighborhood-level public health can be improved with a community garden.

Figure 2: Data for Healthy Communities course outline



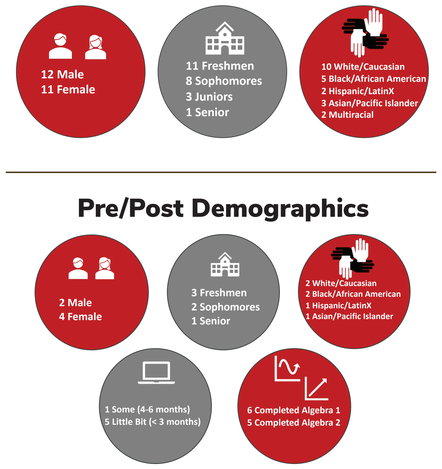
Results

A pre/post evaluation was developed comprising 14 attitudinal questions, 14 cognitive questions (10 multiple choice), and 5 demographic questions. The evaluation measured interest and knowledge in:

- Public interest technology
- Public health and advocacy
- Select Ohio Math Standards and Spreadsheet skills

The class enrollment was 23 students, with 6 students providing the necessary parental permission and student assent for the pre/post assessment.

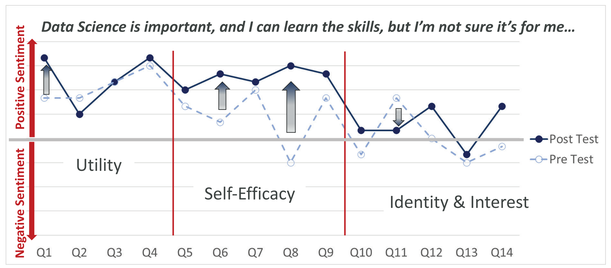
Figure 3: Demographic information for the full class and pre/post population sample



Attitudinal Responses

Attitudinal questions were posed as both positively and negatively worded statements. Responses were summarized by converting all questions to positively worded sentiments and averaging the Likert scale responses.

Figure 4: Average responses of pre/post attitudinal questions



Utility:
Students understand the utility of data science, with the largest positive change that data is important in our society (Q1).

Self-Efficacy:
Comfort using spreadsheets (Q8) and perceived ability to learn data science (Q6) showed a marked increase in the post-test.

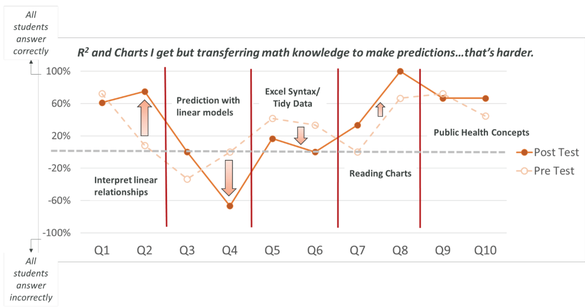
Identity & Interest:
There is a shift from slightly negative to slightly positive sentiment, apart from student intention to minimize their college math courses (Q13).
A decline in career interest (Q11) is attributable to two students who responded with a neutral sentiment in the pre-test responding negatively in the post-test.

Cognitive Responses

Multiple choice responses were scored 1 point for fully correct, -1 point for fully incorrect, and fractionally for partially correct responses on questions with multiple correct answers. Scores in Figure 4 are averaged among the students.

Students spent 20% - 50% less time completing the post-test compared to the pre-test.

Figure 5: Response correctness for pre/post cognitive multiple-choice questions



Interpretation of linear relationships:
There was an improvement in student interpretation of the linear coefficient R2 (Q2). A highly structured class exercise asked students to complete the following sentence for a given scatter plot with R2 provided:

[Indicator 1] predicts __% of the variance in [Indicator 2]

Prediction of Linear Models:
Students did not readily connect prior Algebra 1 knowledge about the equation of a line within the context of data science (Q3-4).

Excel Syntax and Tidy Data:
Despite slightly lower performance, students selected at least partially correct responses (Q5-6). Notably, the attitudinal results demonstrate that students have confidence that they can learn appropriate Excel syntax.

Contributions & Future Work

The DHC course contributes to the ongoing debate about when and how data skills should be introduced in the high school curriculum. The course evaluation offers several lessons:

- Spreadsheet-based curricula promote confidence even over short instructional periods
- Transferring prior math knowledge to the context of data analysis may be more of a challenge for students than interpreting visualizations
- Contextualizing data science education using real-world data from areas such as public health creates opportunities to discuss the importance of data across all professions.

Coming Soon:
Check out our lesson plans and curated real-world datasets!

go.osu.edu/dhc-oers

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

Acknowledgements

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] Code.org. 2023. Support K-12 Computer Science Education in Ohio. Promote Computer Science.
[2] Mike Duffey, John Wiseman, and Kelly Evans Gaier. 2022. Report of the State Committee on Computer Science. Ohio Department of Education.
[3] Tara Dawson McGuinness and Hana Schank. 2021. Power to the public: The promise of public interest technology. Princeton University Press, Princeton, NJ.
[4] 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.

Translational Data Analytics Institute
University Libraries