Data for Healthy Communities:
Development of a Supplemental High School Course

Jessica T. Kulp

THE OHIO STATE UNIVERSITY
COLLEGE OF EDUCATION AND HUMAN ECOLOGY
TRANSLATIONAL DATA ANALYTICS INSTITUTE

OVERVIEW

"Data for Healthy Communities" is a 3-week supplemental course for high school students that is currently in development. This course will introduce students to the emerging field of **Public Interest Technology (PIT)** by exploring how data science may be applied in the service of social good. It will cover real-world uses of data in public health work and address common gaps in public health workers' data literacy.

Fig 1. Larger context of the project.

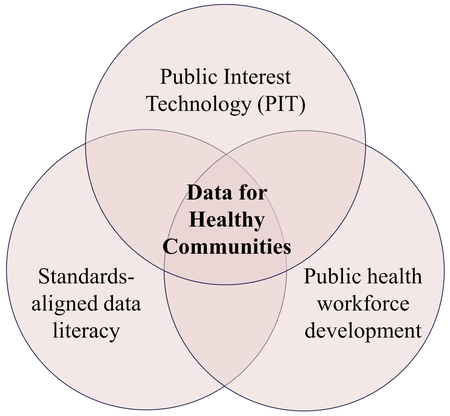
BACKGROUND

- High school students should understand how data is integrated in a wide variety of professional fields and areas of everyday life, including public health.
- The development of the public health workforce is vital as the world grapples with global health threats like pandemics, environmental hazards, and the mental health crisis (Jonas et al., 2014).
- Tying these two priorities together, PIT describes the use of tech (data) for social good (Lusk, 2022).
- Integrating data science education and public health will enable students to engage in meaningful, context-rich, data-driven problem solving while promoting civic engagement (Gebre, 2022). The hope is that this will stimulate interest to explore careers in public health while developing data literacy.

OBJECTIVES

1. **Develop course focus and objectives**
   - Aligned with the mission of PIT, data science learning standards, and real-world uses of data in public health work
2. **Plan multi-purpose course evaluation**
   - Pre-post-test with attitudinal and cognitive items
3. **Design first learning activity**
   - Introduces the data lifecycle (implicitly)
   - Uses topical, local public health data

Fig 2. Key motivations for the project.



SETTING

The Data for Healthy Communities course will be offered at a public-serving STEM high school in Columbus, Ohio during their 3-week winter term in 2024. Courses offered in this term are ungraded. It will be primarily taught by a math teacher at the school and may include students in 9th to 12th grade with a range of prior math, data science, and computing experience.

COURSE FOCUS AND OBJECTIVES

External evaluation reports and training materials from early stages of the Community Health Worker training program were analyzed to identify common data uses and gaps in data literacy for public health workers. **Visualization, interpretation, decision-making, and communication with data** were most vital to these public health professionals' work.

The math teacher was consulted to select five key Ohio Data Science Foundations learning standards. This included a mixture of high school mathematics and computer science standards.

Box 1. Selected data science standards.

1. Create and graph equations to relate two quantities.
2. Represent data with two quantitative variables on scatter plots and describe the relationship.
3. Interpret the slope and intercept of a linear model.
4. Distinguish between correlation and causation.
5. Analyze the benefits and limitations of data visualizations, artifacts, and tools.

EVALUATION MATERIALS

A pre-post-test containing attitudinal and cognitive items was developed for student assessment as well as evaluation of course outcomes.

The attitudinal component of the test contains 14 items on a 5-point Likert scale, probing students' self-efficacy (6 items), utility value (4 items), interest value (3 items), and identity in data science (1 item). Self-efficacy items focused on students' beliefs about their ability to *learn* (as opposed to *do*) data science, as many students will have minimal prior experience. Several items were adapted from an instrument published by Dichev & Dicheva (2017)

Box 2. Sample attitudinal items.

1. Data is important in our society.
2. I rarely encounter data in my everyday life.
3. I feel confident that I can learn data science.
4. Pursuing a career that involves data is a possibility for me.
5. I'm not the type of person who does data science.
6. Using data to better the world is exciting to me.
7. Data science is boring.

The cognitive component of the test contains 14 items (4 open-ended, 10 multiple choice), aligned with the course objectives. Some items from the Programme for International Student Assessment (PISA) (OECD, 2023) and the Data Analysis Research Experience (Wilder et al., 2016) were adapted to address the selected standards and adjusted to a public health context. Additional items were developed to address all course objectives, such as those related to spreadsheet skills and the relationship between data, public health, and PIT.

Fig 3. Sample cognitive item: Spreadsheet formulas.

	A	B
1	Activity	GPA
2	77	3
3	45	2.7
4	89	3.3
5	94	3.8
6	80	3.1
7	81	3.7
8	86	3.6
9	56	2.9
10	90	4
11	32	2.5

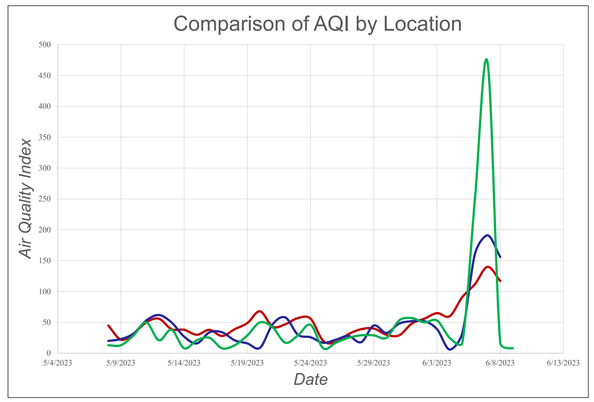
Question 6. Suppose you want to calculate the average GPA in the sample of 10 students shown in the spreadsheet below. What would you write into the Excel formula bar to calculate this? Select all that apply.

a. =MEAN(A2:A11)
b. =SUM(A2:A11) / 10
c. =AVERAGE(B2:B11)
d. =SUM(B2:B11) / 10

ACTIVITY: LOCAL AIR QUALITY

The first activity developed for this course guides students through each step of a simplified 'data lifecycle', based on the Cross-Industry Standard Process for Data Mining (CRISP-DM; Chapman et al., 1999), as they explore air quality data in three locations (Franklin County, OH; Queens County, NY; Ottawa, Ontario) during the Canadian wildfires in June 2023. Students can engage with relevant, meaningful, real data through an inquiry process, learn about the role of data in public health for social good, and work towards state data science learning standards (2, 3, and 5 in Box 1).

Fig 4. Visualization to be created by students.



REFERENCES

1. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (1999). *CRISP-DM 1.0*. SPSS Inc.
2. Dichev, C., & Dicheva, D. (2017). Towards Data Science Literacy. *Procedia Computer Science, 108*, 2151–2160.
3. Gebre, E. (2022). Conceptions and perspectives of data literacy in secondary education. *British Journal of Educational Technology, 53*(5), 1080–1095.
4. Jonas, O., Parry, I., Chrisholm, D., Banatvala, N., & Laxminarayan, R. (2014). Global Health Threats of the 21st Century. *International Monetary Fund Finance & Development, 51*(4).
5. Lusk, K. (2022). Public Interest Technology University Network: Understanding the state of the field.
6. Ministry of the Environment, Conservation, and Parks [MECP] (2023). *Ottawa Downtown: Hourly Fine Particulate Matter Readings.*
7. Organization for Economic Cooperation and Development [OECD] (2023). Programme for international student assessment.
8. United States Environmental Protection Agency [US EPA] (2023). [Daily air quality data for PM2.5 in Queens-NY and Franklin-OH].
9. Wilder, E. I., Vianna, E., & West, R. W. (2021). Data analysis research experience (DARE) project student pre-assessment.

ACKNOWLEDGEMENTS

This project is supported with funding from the New Venture Fund as part of the Public Interest Technology University Network 2022 Challenge Awards.