Integrating Data Science Practices into Informal Learning: A STEM Summer Camp Approach

By Marc T. Sager, Southern Methodist University; Saki Milton, Southern Methodist University; Candace Walkington, Southern Methodist University; and Anthony J. Petrosino, Southern Methodist University

Within this article, we introduce a curriculum designed to integrate data science practices into an informal learning environment. We implemented this curriculum as part of a week-long residential science, technology, engineering, and mathematics (STEM) summer camp tailored for middle school-aged girls. The camp's core objective is to enhance participants' data literacy and proficiency in data science through hands-on activities focused on socio-scientific topics (Kang et al., 2018; Milton et al., 2023; NGSS Lead States, 2013).

Students engage collaboratively in structured activities that build data science skills while working towards a culminating project: creating infographics that address campaign-related issues. The learning objectives include understanding the connections between secondary education, higher education, and career pathways; developing effective written and verbal communication skills for various audience needs; fostering collaborative professionalism among peers and adults; and using technology proficiently to inform or persuade audiences (NGSS Lead States, 2013; Petrosino, 2023; Sager, 2024; Wilkerson & Polman, 2020).

This curriculum offers a practical and adaptable framework for educators seeking to integrate data science into STEM curricula. It equips students not only with foundational technical skills in data analysis and visualization but also with critical awareness around ethical considerations and real-world applications. This dual focus prepares learners to navigate and contribute meaningfully to the evolving field of data science, making the curriculum relevant and scalable across diverse educational settings.

Connecting Data Science to Environmental Education

For teachers and practitioners, this curriculum offers a way to seamlessly integrate data science into environmental science education. By using real-world data, students are able to see the immediate relevance of STEM concepts to their own lives. The use of data analysis tools, such as CODAP (Common Online Data Analysis Platform; The Concord Consortium, n.d.), allows students to engage with the data in a meaningful way, gaining insights that go beyond traditional textbook learning.

Furthermore, this curriculum serves as a model for integrating data science into various scientific disciplines, encouraging educators to adopt a cross-curricular approach. For example, educators can adapt it to study other environmental issues such as air quality or water contamination, making it flexible and applicable to a range of topics within both informal and formal educational settings.

Context: A STEM Summer Camp for Girls

Middle school aged girls at a week-long STEM summer camp in the U.S. Southwest explore mathematics, science, and data science through hands-on learning. Since 2010, the camp has rotated across universities, expanding access for underrepresented minority girls. Recruitment efforts include schools, community organizations, cultural centers, and social media, with a nonprofit partnership supporting outreach to all-girls schools.

Experienced educators guide campers through grade-level STEM concepts aligned with state standards, while the curriculum intentionally emphasizes data literacy. Working in teams, girls apply data science skills to create infographics on socio-scientific issues, strengthening their communication, critical thinking, and collaboration. The camp also helps them see connections between education and career pathways as they use technology to inform and persuade.

Expectations and the *Next Generation Science Standards*

Our idea of data science education for this curriculum is consistent with the *Next Generation Science Standards* (NGSS) practices: a) Asking questions and defining problems; b) Analyzing and interpreting data; c) Using mathematics and computational thinking; d) Engaging in argument from evidence; and e) Obtaining, evaluating, and communicating information (MS-ESS3). Tools in this curriculum include the use of data, mathematical and computational representations, graphs, diagrams, and data analysis software (CODAP and Google Sheets). Throughout this curriculum, we focus on environmental-related data to help students analyze and interpret data to help inform the use of technology to mitigate the effects of environmental issues in their communities.

Overview of the Curriculum

This curriculum, as summarized in Figure 1, follows the 5E instructional model (Engage, Explore, Explain, Elaborate, Evaluate), which provides a structured framework for inquiry-based learning in science education. The curriculum spans four 65-minute class periods, with each period corresponding to a different phase of the 5E model. This approach helps middle school students develop critical inquiry skills by systematically moving from initial engagement with environmental issues to data analysis and community presentation of their findings. Each subsection below describes the specific activities and learning objectives for one class period, building toward the ultimate goal of students creating and presenting data-driven infographics about local environmental hazards to a community audience.

Figure 1

Overview of the Curriculum

1. **Objective**: Students will examine the quality of soil and water samples, identifying potential pollutants and analyzing their findings through the lens of data science. Using CODAP, they will work with existing data, ask questions about the data, and develop an infographic to present their results.
2. **Materials**:
    - Soil samples from local environments (e.g., parks, gardens, school grounds)
    - Water samples from nearby water sources (e.g., lakes, rivers, fountains)
    - pH test kits for basic soil and water quality testing
    - CODAP software on laptops or tablets for data analysis
    - Graphing tools (e.g., rulers, graph paper)

Safety Considerations

- **Precautions**: Ensure students use gloves when handling soil and water samples to avoid contamination. Students should follow standard lab safety procedures during testing.
- **Outdoor Testing**: Be mindful of the environment and safety when gathering samples from outdoor locations.

Procedure

1. **Contextualizing the Problem**: Before collecting samples, students are introduced to the concept of environmental pollutants in soil and water. They are encouraged to consider what questions they have about these pollutants and how they might explore those questions through data.
2. **Data Collection**: Students will collect soil and water samples from various locations around their community. These samples are used to identify potential pollutants and examine their impact on the local environment.
3. **Data Analysis with CODAP**: Students will upload their collected data into CODAP, where they can manipulate, analyze, and visualize the information. This process will help them answer their own questions about pollutants, environmental quality, and community health.
4. **Developing Infographics**: After analyzing the data, students will create an infographic that summarizes their findings. The infographic should include graphs or charts that visualize the data, as well as a narrative that explains the analysis and connects the findings to real-world environmental concerns.

Community Presentation

In a culminating event, students will present their infographics to a community audience, such as local residents, environmental groups, or school peers. This community-focused presentation allows students to share their insights, demonstrate their understanding of environmental issues, and practice communicating data-driven findings in a meaningful way.

Discussion and Reflection

In small groups or as a whole class, students will reflect on the activity:

- What patterns did you observe in the soil and water samples you tested?
- How did your findings help answer the questions you had about pollutants and their impact?
- In what ways did your analysis help you understand how communities can be affected by environmental issues?

This reflection period encourages students to think critically about their work and to see the real-world relevance of data science in environmental science.

Engage: What are local environmental hazards or pollutants that may impact health? (65-minute class period)

During this Engage session, we introduce students to fundamental concepts of data science and environmental science, focusing on how pollutants in soil and water can be detected and their potential hazards. Students engage in hands-on exploration of environmental testing methods (Figure 2 shows students performing soil and water quality tests) and community-based discussions to build awareness of environmental-health connections. This session establishes the foundation for the curriculum's main goal of helping students develop critical inquiry skills by generating their own questions, using existing environmental data, analyzing that data with CODAP, and presenting their findings in a community-based context. This initial session focuses on introducing basic data science concepts that will be expanded upon in subsequent modules, while students will gain comprehensive experience in data collection, analysis, and visualization throughout the complete curriculum.

Figure 2

Students Testing Water and Soil Quality

In the context of investigating local environmental hazards and pollutants, we introduce the campers to the basics of data science and the ethical considerations essential in data analysis. Students explore the significance of data science in analyzing environmental data, such as soil and water quality measurements. For example, students examine real datasets showing pH levels and heavy metal concentrations in local water sources, learning to identify different data types and their significance for environmental health. We place emphasis on ensuring data integrity and accuracy while collecting and interpreting environmental data. Students learn to recognize biases and limitations in their analysis, such as how testing water quality only during dry seasons might miss important pollution from stormwater runoff, or how sample size matters when comparing soil contamination across different neighborhood locations, which is crucial for making informed and responsible decisions about environmental health. Understanding these ethical considerations helps promote the responsible use of data in real-world decision-making.

A key component of this class period involves students engaging with a case study on data ethics to illustrate the ethical dilemmas and best practices in the field of data science as it applies to environmental studies, enhancing their ability to conduct investigations responsibly and effectively. By the end of this period, students develop initial questions about local environmental hazards and understand the ethical framework they will use throughout their investigation.

Explore: Exploring the topic and identifying relevant data (65-minute class period)
This session establishes the foundation for students' campaign projects while developing essential skills in research planning, team collaboration, and project management within an environmental science context. In this class period, students embark on an engaging project where they delve into the world of politics and presidential campaigns. They work in teams to design compelling campaign materials, such as flyers or infographics, aimed at raising awareness and inspiring action on key environmental issues. The primary focus is to create an informative and engaging infographic on a chosen environmental topic, such as air quality or soil pollution. This project provides hands-on experience in understanding campaign strategies, effectively communicating environmental issues, and using data to support their arguments.

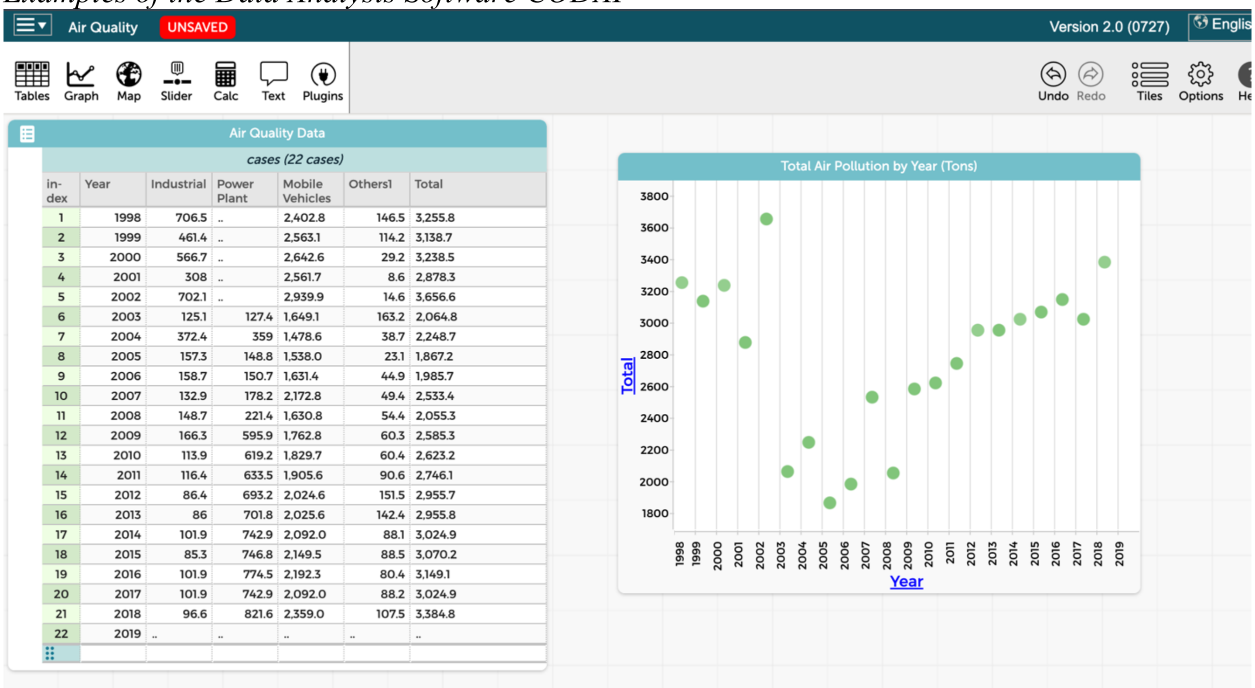
Students began by brainstorming environmental topics of interest and then generate ideas based on their individual and group interests. They develop campaign names, taglines, and key research questions (e.g., How can a flood affect soil pollution?) to guide their project. An important aspect of this session involves defining team roles, including Researcher, Data Analyst, Graphic Designer, Content Writer, and Project Manager, ensuring that each team member has clear responsibilities and that all aspects of the project are covered.

Explain: Data analysis and visualization workshop (65-minute class period)
In this class period, students engage in a workshop focused on data analysis and visualization using CODAP. CODAP, developed by the Concord Consortium, is a free, web-based tool designed to facilitate data exploration through an intuitive interface, allowing users to manipulate and analyze a wide range of data types. It supports the creation of graphs, tables, and maps, helping users uncover patterns and insights while fostering data literacy and critical thinking skills.

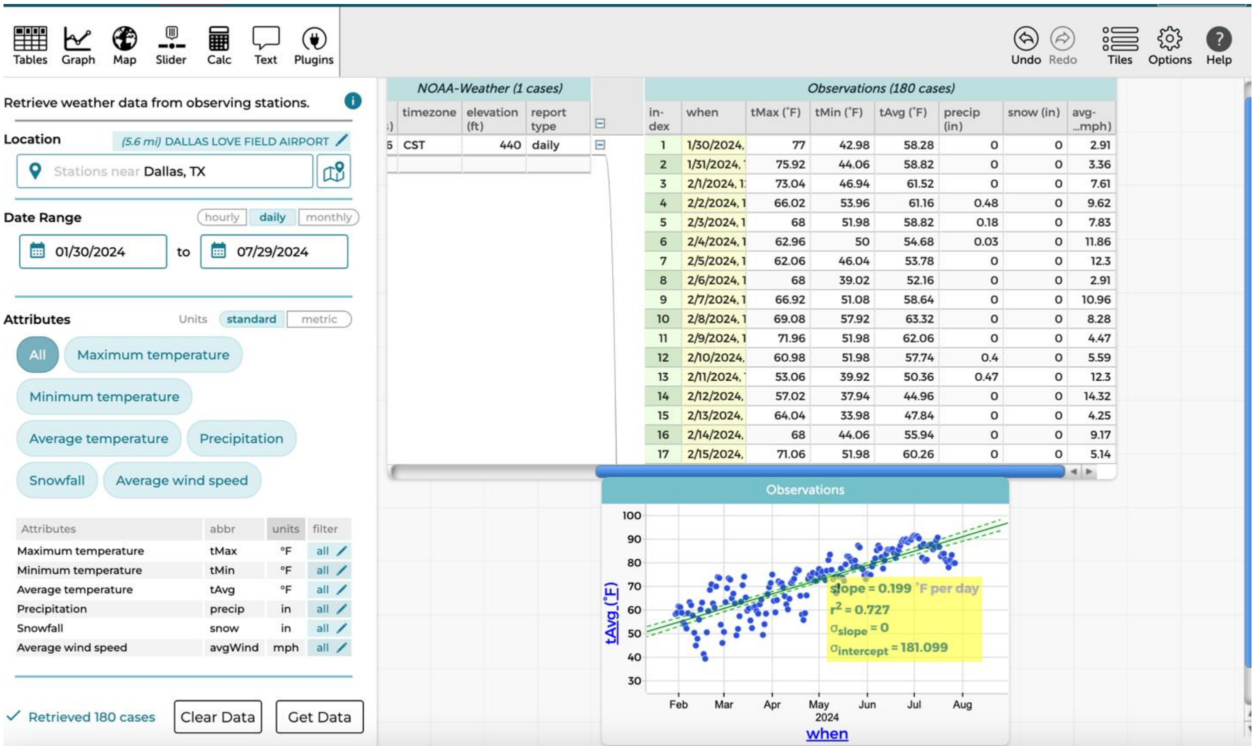
Students participate in activities that involve reading graphs, interpreting data visualizations, and understanding the context of data. They then use a supplementary dataset to explore CODAP's features, gaining practical experience with the tool (see Figure 3, this is an example of a preloaded dataset within CODAP so the students can become familiar with the functionality of the analysis software). This hands-on experience allows them to consider how they could apply CODAP to analyze and visualize data for their campaign launch, enhancing their ability to create compelling and data-driven campaign materials.

Figure 3

Examples of the Data Analysis Software CODAP



in-dex	Year	Industrial	Power Plant	Mobile Vehicles	Others1	Total
1	1998	706.5	..	2,402.8	146.5	3,255.8
2	1999	461.4	..	2,563.1	114.2	3,138.7
3	2000	566.7	..	2,642.6	29.2	3,238.5
4	2001	308	..	2,561.7	8.6	2,878.3
5	2002	702.1	..	2,939.9	14.6	3,656.6
6	2003	125.1	127.4	1,649.1	163.2	2,064.8
7	2004	372.4	359	1,478.6	38.7	2,248.7
8	2005	157.3	148.8	1,538.0	23.1	1,867.2
9	2006	158.7	150.7	1,631.4	44.9	1,985.7
10	2007	132.9	178.2	2,172.8	49.4	2,533.4
11	2008	148.7	221.4	1,630.8	54.4	2,055.3
12	2009	166.3	595.9	1,762.8	60.3	2,585.3
13	2010	113.9	619.2	1,829.7	60.4	2,623.2
14	2011	116.4	633.5	1,905.6	90.6	2,746.1
15	2012	86.4	693.2	2,024.6	151.5	2,955.7
16	2013	86	701.8	2,025.6	142.4	2,955.8
17	2014	101.9	742.9	2,092.0	88.1	3,024.9
18	2015	85.3	746.8	2,149.5	88.5	3,070.2
19	2016	101.9	774.5	2,192.3	80.4	3,149.1
20	2017	101.9	742.9	2,092.0	88.2	3,024.9
21	2018	96.6	821.6	2,359.0	107.5	3,384.8
22	2019	..	..	..	..	..



Elaborate: Feedback, revisions, and finalizing presentations (65-minute class period)
During this class period, students focus on refining their campaign materials and preparing for their community presentation. They began by using a progress tracker, a tool that allows them to monitor their completion of key tasks and milestones throughout the project, to plan, and to draft their work. For example, students check off steps like selecting a dataset, conducting initial analyses, and drafting infographic elements, which helps them stay organized and focused on their goals. Students brainstorm the key messages they want to convey about their environmental topic and create a sketch outline for their presentation board. This helps them organize their thoughts and plan the layout of their infographics. Next, they use CODAP to incorporate their selected datasets, focusing on clarity and effectiveness in their visualizations to ensure their graphs, tables, and maps communicate their message effectively.

Following this, students engage in a peer review session, evaluating each other's work based on criteria such as message clarity and design principles. To support students in giving useful feedback, we provide a review checklist with guiding questions and model examples of specific, respectful, and actionable comments. This structure ensures that the peer review process produces targeted suggestions for improvement. Based on this feedback, students revise their infographics, paying close attention to identified areas for enhancement.

Finally, students develop a script for their community presentation scheduled for the next day. They create an initial draft and utilize Generative Artificial Intelligence to help revise and refine their script, using prompts such as: "Rewrite this script to be clear for a middle school audience while keeping the tone enthusiastic and persuasive." This helps ensure the final version is engaging and communicates their key messages effectively. This comprehensive approach enables students to polish their campaign materials, enhance their data visualizations, and prepare a compelling presentation.

Community Presentations

During the community presentation, students showcase their final infographics to an audience of family members, camp staff, and invited community guests. We structure the event as an open gallery walk, where each student group displays their infographic at a table or poster station. Community members can circulate freely, view the infographics up close, and engage in one-on-one or small-group conversations with the students.

Students take turns explaining the environmental issue they investigated, the data they analyzed, and the visual and narrative choices they made to convey their findings. These exchanges give students opportunities to answer questions, clarify complex points, and connect their work to real-life experiences shared by community members. Many presentations spark discussions about local environmental challenges and potential solutions, allowing students to see the relevance of their work beyond the camp setting.

In addition to informal conversations, some groups deliver short, structured overviews of their project to the whole audience, ensuring everyone has a chance to hear the main messages. The mix of open interaction and brief formal presentations create a dynamic setting where students can share their data-driven stories, receive encouragement and feedback, and practice communicating their ideas to a diverse audience.

Assessment

As practitioners, it is important to assess students' understanding both during and after the activity. In this activity, formative assessment can take the form of class discussions, peer feedback sessions, and the students' ability to apply their knowledge in their final infographics and presentations. For a more detailed evaluation, teachers can use rubrics, such as the one shown in Table 1, to assess the clarity of students' messages, the effectiveness of their visualizations, and their use of data science tools.

Table 1
Rubric

Criteria	Excellent (4)	Good (3)	Fair (2)	Needs Improvement (1)
Clarity of Message	The infographic effectively communicates the main message or key findings. Information is clear, concise, and easy to understand.	The infographic mostly communicates the main message or key findings. Information is clear but may be slightly verbose or cluttered.	The infographic partially communicates the main message or key findings. Some information is unclear or difficult to understand.	The infographic does not effectively communicate the main message or key findings. Information is confusing or disorganized.

Visual Appeal	The infographic is visually appealing and engaging. Graphics, colors, and layout are used effectively to enhance readability and interest.	The infographic is visually appealing but may lack some consistency or cohesion in design elements. Graphics, colors, and layout are generally effective.	The infographic is somewhat visually appealing but may be overly simple or lack creativity in design. Graphics, colors, or layout could be improved.	The infographic lacks visual appeal and may appear cluttered or unprofessional. Graphics, colors, or layout are ineffective.
Data Analysis and Visualization	Data analysis is thorough and insightful, supporting the main message of the infographic. Visualizations are clear, accurate, and effectively convey key information.	Data analysis is mostly accurate and relevant to the main message of the infographic. Visualizations are clear but may lack some detail or depth of analysis.	Data analysis is limited in scope or depth and may not fully support the main message of the infographic. Visualizations are present but may be difficult to interpret.	Data analysis is minimal or inaccurate and does not effectively support the main message of the infographic. Visualizations are unclear or misleading.
Use of Data Science Tools	Demonstrates proficient use of data science tools (e.g., ArcGIS, CODAP) to analyze and visualize data. Tools are used effectively to enhance the quality and depth of analysis.	Shows adequate use of data science tools to analyze and visualize data. Some aspects of tool usage may be overlooked or underutilized.	Demonstrates basic use of data science tools but may lack sophistication or depth in analysis and visualization.	Shows limited or ineffective use of data science tools. Analysis and visualization are rudimentary or incomplete.

Reflections and Recommendations

Reflecting on the effectiveness of the unit, students demonstrated significant growth in data literacy, problem-solving, and collaborative skills. The peer review process encouraged critical thinking and refinement of their work, while the community presentation offered a real-world opportunity to communicate findings to a broader audience. These experiences helped build both technical competencies and important 21st-century skills such as communication, teamwork, and ethical decision-making.

The activity effectively engaged students by immersing them in a real-world scenario—designing a campaign to address soil pollution—which made learning relevant and impactful. Students developed skills in clear communication through infographics, visual design, and data analysis techniques, supported by the use of accessible tools like CODAP. Integrating math and science concepts allowed students to apply theoretical knowledge practically, demonstrating real-world relevance.

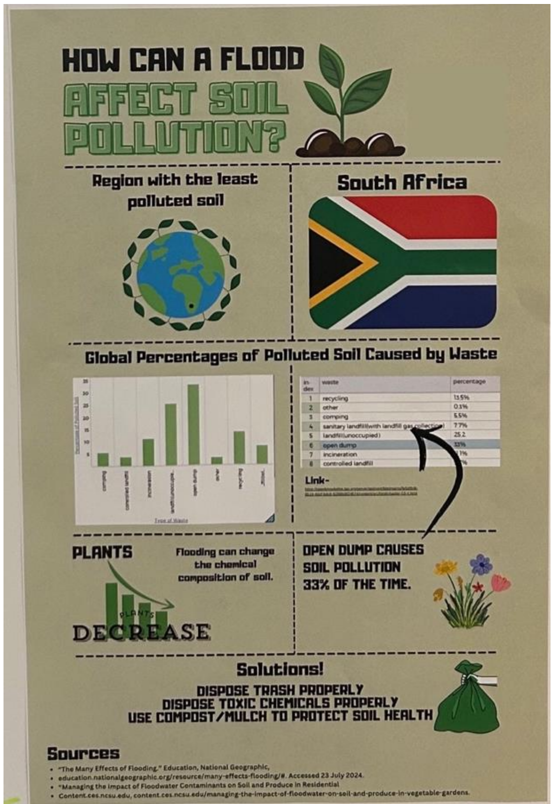
The peer review and feedback sessions fostered critical thinking and collaboration, improving students' abilities to evaluate and enhance their work constructively. The focus on a tangible issue, like soil pollution, made the activity relevant to current environmental concerns, showing students the impact of their work on real-world problems.

Examples of student products showcased the effectiveness of the activity. In Figure 4A, the campers successfully highlighted key aspects of soil pollution, such as sources, impacts, and solutions, with clear and accurate visualizations. Its appealing design and effective use of data science tools demonstrated the students' ability to convey complex information clearly. In contrast, Figure 4B, while visually appealing and focused on health effects of soil pollution, showed areas for improvement in design clarity and depth of data analysis. Feedback from peers was instrumental in refining this infographic, resulting in a more impactful final product. Overall, the activity effectively developed key skills, connected academic learning to practical applications, and addressed a relevant environmental issue through engaging and collaborative work.

For educators interested in adopting this approach, we recommend selecting a locally relevant socio-scientific issue to increase student engagement, providing structured peer review tools such as checklists and model feedback examples to ensure feedback is constructive, and incorporating accessible data science tools like CODAP to lower barriers to entry. Additionally, explicit scaffolding to help students translate complex data insights into clear, actionable messages should be incorporated, with guidance on audience awareness and narrative development. We also recommend including an authentic audience, such as family members, community leaders, or subject-matter experts, to make the final presentation meaningful and allowing time for iterative improvement so students can apply feedback and refine both their data products and presentation materials.

Lessons learned from this implementation include the importance of giving explicit guidance on both data analysis and visual communication, providing early scaffolding in topic selection and dataset exploration to keep projects focused, and embedding ethical discussions about data collection and representation to encourage critical thinking about how student work might be used or interpreted.

Figure 4
Examples of Students' Produced Infographics



HOW CAN A FLOOD
AFFECT SOIL POLLUTION?
Region with the least polluted soil
South Africa
Global Percentages of Polluted Soil Caused by Waste
index
waste
percentage
1 recycling 13.5%
2 other 0.3%
3 comping 5.5%
4 sanitary landfill(with landfill gas collection) 7.7%
5 landfill(unoccupied) 25.2
6 open dump 33%
7 incineration
8 controlled landfill
Link-
PLANTS
Flooding can change the chemical composition of soil.
DECREASE
OPEN DUMP CAUSES SOIL POLLUTION 33% OF THE TIME.
Solutions!
DISPOSE TRASH PROPERLY
DISPOSE TOXIC CHEMICALS PROPERLY
USE COMPOST/MULCH TO PROTECT SOIL HEALTH
Sources
"The Many Effects of Flooding." Education, National Geographic,
education.nationalgeographic.org/resource/many-effects-flooding/#. Accessed 23 July 2024.
"Managing the Impact of Floodwater Contaminants on Soil and Produce in Residential
Content.ces.ncsu.edu, content.ces.ncsu.edu/managing-the-impact-of-floodwater-on-soil-and-produce-in-vegetable-gardens.

A)

SAVE
OUR SOIL
SAVE OUR SOIL:
CULTIVATE A SUSTAINABLE FUTURE
What has been done to decrease pollution levels?
• Clean Air Act
What is soil pollution?
The contamination of soil with large concentrations that is harmful to the body.
Common Pollutants:
Particles
Ozone
Lead
Carbon monoxide
Nitrogen dioxide
Sulfur dioxide
How does soil pollution impact the state of our health?
• Cadmium
• Arsenic
Annual Pesticide Used by Continent in 2017 (Tons)
How does soil pollution affect our environment?
• Healthy soil vs Polluted Soil
• Circulatory
Arable Soil in The United States
Toxins seep into the soil from various sources
The toxins reach the plant and become integrated in the roots

B)

Conclusion

Integrating data science practices into informal learning environments, particularly for middle school girls, offers significant benefits. This curriculum not only imparts essential technical skills but also fosters critical thinking, teamwork, and an understanding of ethical considerations in data science applications. By centering learning around real-world issues like soil pollution and utilizing tools such as CODAP for data analysis and visualization, students engage in hands-on experiences that prepare them for future educational and career opportunities in data science.

Reflecting on the learning trajectory, the iterative process of student inquiry, data analysis, and public presentation proves to be a key strength of the curriculum. Students not only build confidence in handling data but also develop ownership over their projects by posing their own questions, analyzing existing datasets, and making sense of their findings in meaningful ways. The peer review process further strengthens their ability to critically evaluate data representations and refine their communication strategies. While students demonstrate strong engagement and growth in data literacy, future iterations of the curriculum can provide additional scaffolding to help translate data insights into clear, actionable messaging, particularly for audiences unfamiliar with data science concepts.

This approach aligns with existing research emphasizing the importance of early exposure to data science for skill development and incorporates ethical considerations, ensuring students are aware of the broader implications of their work (Sager, 2024; Wilkerson & Polman, 2020). Moreover, it offers a practical, adaptable framework for educators to integrate into STEM curricula, equipping students with both the technical expertise and the critical awareness needed to navigate and contribute meaningfully to the field of data science. The emphasis on ethical data practices ensures that students not only learn how to work with data but also understand the responsibility that comes with it.

The Campaigning for Data Literacy lesson overview by Sager and Milton is available online. A more detailed version of the lesson materials is available upon request.

References

Kang, H., Calabrese Barton, A., Tan, E., Simpkins, S. D., Rhee, H-y., & Turner, C. 2019. How do middle school girls of color develop STEM identities? Middle school girls' participation in science activities and identification with STEM careers. *Science Education* 103:418–439.

Milton, S., Sager, M.T., & Walkington, C. 2023. Understanding racially minoritized girls' perceptions of their STEM identities, abilities, and sense of belonging in a summer camp. *Education Sciences* 13(12), Article 1183.

NGSS Lead States. 2013. *Next Generation Science Standards: For states, by states*. The United States: National Academies Press.

Petrosino, A.J., Walkington, C., & Ekberg, D. 2023. *Frameworks for integrated project-based instruction in STEM disciplines.* The United States: Information Age Press.

Sager, M.T. 2024. "It's hard to quantify community togetherness": Exploring the evolution of data science practices and uncovering critical tensions. *Teaching and Learning Theses and Dissertations,* Article 24.

Sager, M.T. & Milton, S. 2024. *Campaigning for data literacy*. The GEMS Camp.

The Concord Consortium. N.D. *Common Online Data Analysis Platform (CODAP*). *https://codap.concord.org*

Wilkerson, M.H. & Polman, J.L. 2020. Situating data science: Exploring how relationships to data shape learning [Special Issue]. *Journal of the Learning Sciences* 29(1):1-10.