Bringing Students' Lives Into Data Science Classrooms

David Weintrop[1] Rotem Israel-Fishelson[1]

[1]College of Education, University of Maryland, College Park, Maryland, United States of America

Column Editor's Note: Data science is fast becoming a topic of interest not just for university students, but at the high school level as well, as we have highlighted in the [Minding the Future](#) column over the years. In this article by Weintrop and Israel-Fishelson, the authors describe one way to identify data sets of relevance to the lives of high school students and to integrate them into a course curriculum. Such integration can help to enhance students' excitement about data science, as they see how the field allows them to answer questions about subjects near to their hearts, be it music, sports, film—or something else altogether. Teachers: Have you explored new ways of getting your students excited about data science? If so, consider sharing your experiences with others!

Keywords: data science education, interest-driven learning, API can code, high school, curriculum

Introduction

Data is everywhere. Growing up in a world saturated with data, students are constantly creating, using, and being impacted by the data that surrounds them [(Ng, 2012)](#). From the moment they wake up and glance at their smartphones to the choices they make throughout the day, data influences and shapes the experiences of today's youth. These influences are both visible and invisible, as the effects of data are not always clear. Further, the way data impacts students is not always fair or equitable [(Carmi et al., 2020)](#). For example, college admissions often use data-driven algorithms to screen applicants, but these algorithms can reflect and reinforce biases related to students' socioeconomic status and school quality [(O'Neil 2016)](#). This underscores the importance of all students gaining a fundamental understanding of data science so that they can understand the role of data in their lives [(Gould, 2021)](#). Data science empowers students to understand, respond to, and succeed in a world shaped by data. The goal of data science education is to equip students with the skills and knowledge to make informed decisions, advocate for fair data practices, and contribute to a more just and inclusive digital world [(Biehler et al., 2022)](#).

Students' experiences with data can serve as fertile grounds for data science education. By linking data science instruction with students' interests and the real-world data they encounter, we can produce dynamic, engaging, and equitable learning experiences [(Brooks et al., 2021)](#). Crafting instructional materials that integrate thoughtfully chosen data sets and activities that reflect students' passions, values, and unique voices can deepen their engagement with data science and establish authentic and meaningful connections with the field [(Lee et al., 2021](#); [Wilkerson & Polman, 2020)](#).

Incorporating data sets relevant to students' lives allows teachers to illustrate the practical applications of data science and how it intersects with the activities and communities the students participate in. For instance, a data science class could have students analyze their own social media data, investigate trends of music they love, discover characteristics about their neighborhood, or explore their favorite sports team's statistics. Given

all the potential data explorations students could pursue, giving them the ability to choose the data they want to explore can both increase interest and help them understand the relevance of data science in their lives. Letting students choose their own data sets makes data science more personalized and meaningful and can help them grasp the importance and relevance of data science (Lee et al., 2021).

In this article, we advocate for incorporating these data science learning experiences that draw on students' interests and identities. These interest-driven data sets can lay the groundwork for engaging and purposeful data science education. In doing so, we can empower today's students to succeed in an increasingly data-driven world.

Centering Students' Interests in High School Data Science Instructional Materials

Creating an interest-driven instructional sequence starts with investigating the interests of the students you plan on teaching. To that end, we ran a series of participatory design (PD) sessions with 28 high schoolers from a city in the Mid-Atlantic region of the United States (Israel-Fishelson et al., 2024). Participatory design is a strategy for making the design process inclusive, ensuring that the voices, values, and ideas of the users, or in this case students, are reflected in outcomes (DiSalvo, 2016). In our work, this meant providing opportunities for students to share their experiences with data, interests, and concerns through discussions and hands-on design activities (e.g., Coenraad et al., 2019). Our analysis revealed that students are highly interested in topics such as music, video games, TV shows, movies, sports, animals, art, and design. These topics then informed the design of classroom data science learning activities.

After identifying areas of interest, the first steps in developing an instructional sequence are to define the learning objectives, identify technologies, and design learning activities that invite learners to engage with data sets that represent these interests. Drawing insights from the PD sessions, we took a student-centered approach and developed "API Can Code," an instructional sequence that introduces students to the computational foundations of data science by having them explore and query authentic and meaningful data sets using publicly available application programming interfaces (APIs). Using an API search engine, students can find data sets on topics they are passionate about. API Can Code is grounded in the interest development theory, which suggests that students learn best when they can explore, interact with, and derive meaning from the subject they are interested in (Michaelis & Weintrop, 2022; Renninger & Hidi, 2015). In the context of our instructional sequence, interest development theory informs two key aspects of the instructional sequence. First, we incorporate data sets based on the students' interests as gathered during the PD research. Second, students have agency throughout the learning activity to pose questions and query data sets that align with their interests beyond the provided data sets. As part of API Can Code, students work through the data science cycle (International Data Science in Schools Project Curriculum Team, 2019). The cycle starts with having students formulate questions based on their passions, identify relevant data sets that can shed light on their questions, explore and analyze the data, and present insights and answers to the questions.
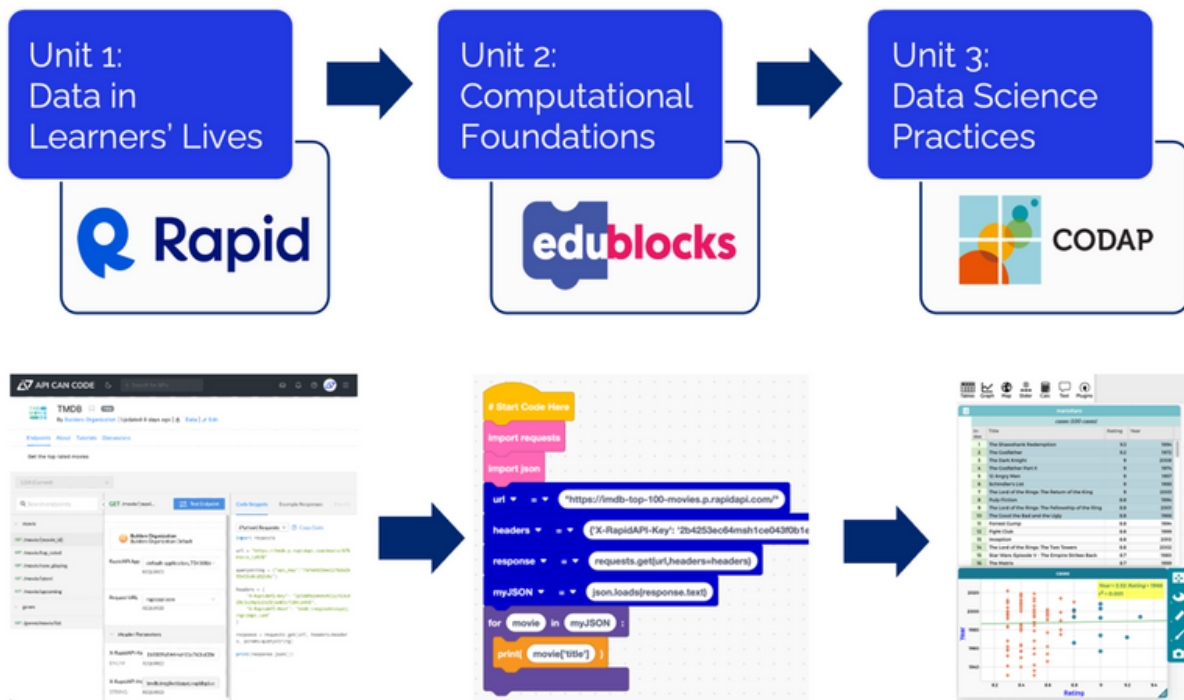
Figure 1. The API Can Code instructional units, aligning with three emerging technologies.

API Can Code includes three instructional units (Figure 1), each consisting of six 90-minute lessons. The objective of the first unit is to help students better understand the data surrounding them and its impact on their lives. In the first unit, students learn what data is, who collects their data, and the data-information-knowledge-wisdom (DIKW) model (Rowley, 2007), which explains how data is transformed from its raw form into valuable insights. Moreover, the students learn about the 5Vs model (Sinha, 2020), which defines the key attributes of data: volume (amount of data), velocity (data recency), variety (types of data), veracity (accuracy and quality of data), and value (ability to extract meaningful insights). The students evaluate data sets using the 5Vs model and discuss the effects of biases on data. The objective of the second unit is to help students develop foundational computational skills to retrieve and manipulate data. More concretely, students learn how to query APIs aligning with their interests through the RapidAPI platform, an API repository, and then manipulate or organize the returned data. To do so, students write short programs in EduBlocks, a block-based programming environment that uses the Python programming language. The lessons in the units will employ the Use→Modify→Create structure (Franklin et al. 2020) in which they first use an existing code, continue to modify one, and finally write a code from scratch. The objective of the third and final unit is to teach students data science practices, emphasizing the analysis and visualization of data to extract meaningful insights. Students learn to use CODAP (Common Online Data Analysis Platform), a free, user-friendly data visualization, analysis, and exploration tool, to perform data analysis, create and interpret a variety of summary plots, and perform basic statistical tests (CODAP, 2022). Each of the three units is anchored with data

exploration activities that align with students' interests as identified in the PD sessions. Additionally, each lesson concludes with a final assignment to review and test their understanding of the content taught. More details about each unit, its learning objectives, and activities can be found here.[1]

API Can Code culminates with a final project where students go through the entire data science cycle, starting with coming up with a question they want to be answered or a topic they want to learn about. From there, students identify a relevant data set, then programmatically retrieve the data set, manipulate it as needed, and then visualize it in CODAP as a way to answer their question or learn new things about their topic of interest. The next section presents an example of a student's final project.

API Can Code in Action

We piloted the API Can Code materials with two computer science classes at a U.S. high school. Each class completed the full instructional sequence. At the end of the three units, each student completed the final project independently under the guidance of their teacher. First, the students chose a topic of interest to them and found an interesting data set from the RapidAPI repository, evaluated it using the 5Vs model, and formulated questions that could be answered using the data. The students chose APIs on various topics, including sports, music, television programs, computer games, and animals. Students then wrote programs in EduBlocks to retrieve, filter, and organize the data relevant to their questions and then created visualizations in CODAP, which brought their data to life, presenting answers to their research questions.

To provide a clearer sense of what this process looked like, here we present a final project by Janeese, one of the students in the class, who decided that her project would be about music. She created an EduBlocks program to retrieve data about her favorite singer, Jhené Aiko, from the Deezer API (Figure 2), a French music streaming service.
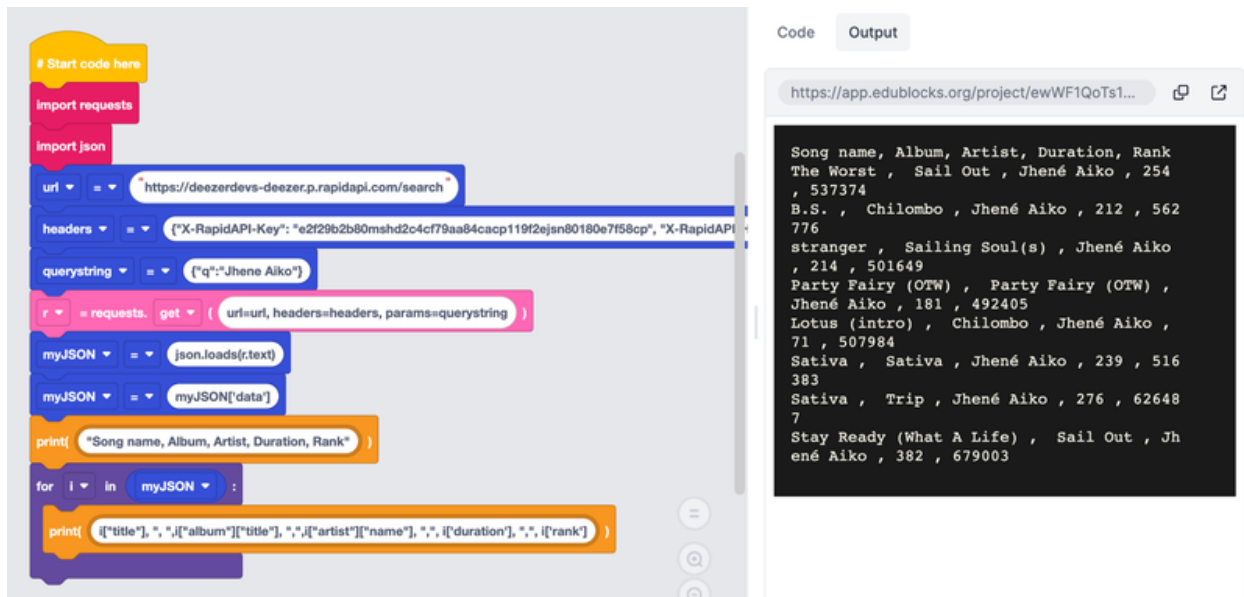
Figure 2. The EduBlocks program created by Janeese to retrieve data about Jhené Aiko.

Janeese was curious to learn more about Jhené and her hit songs, and to try and figure out with data whether or not Jhené was a 'star.' To answer her question, she defined the fields she needed (song name, album, artist, duration, and rank) so that she could work with them in a tabular form within CODAP (Figure 3). Janeese created two visualizations as part of her final project; in her own words, this is how Janeese presented her visualizations: "The first graph [Figure 4 (Left)] represents artists who are featured in Jhené Aiko's songs; it helped me answer my question by giving me insight on how many artists wanted to collaborate with Jhené and how many songs were created by Jhené Aiko herself. The second [Figure 4 (Right)] represents Jhené Aiko's music career." Additionally, she stated that the table itself (Figure 3) represents the most popular songs by Jhené Aiko.

index	Song name	Album	Artist	Duration	Rank	Year
1	Stay Ready (What A Life)	Sail Out	Jhené Aiko	382	679003	2013
2	Sativa	Sativa	Jhené Aiko	239	516383	2017
3	The Worst	Sail Out	Jhené Aiko	254	537374	2015
4	Bed Peace	Sail Out	Jhené Aiko	256	578451	2013
5	Sativa	Trip	Jhené Aiko	276	626487	2017
6	While We're Young	Trip	Jhené Aiko	236	463802	2017
7	B.S.	Chilombo	Jhené Aiko	212	562776	2020
8	stranger	Sailing Soul(s)	Jhené Aiko	214	501649	2011
9	Party Fairy (OTW)	Party Fairy (OTW)	Jhené Aiko	181	492405	2020
10	Lotus (intro)	Chilombo	Jhené Aiko	71	507984	2020
11	living room flow (Bonus)	Sailing Soul(s)	Jhené Aiko	262	414266	2011

Figure 3. The CODAP table created by Janeese after manipulating the data in EduBlocks.
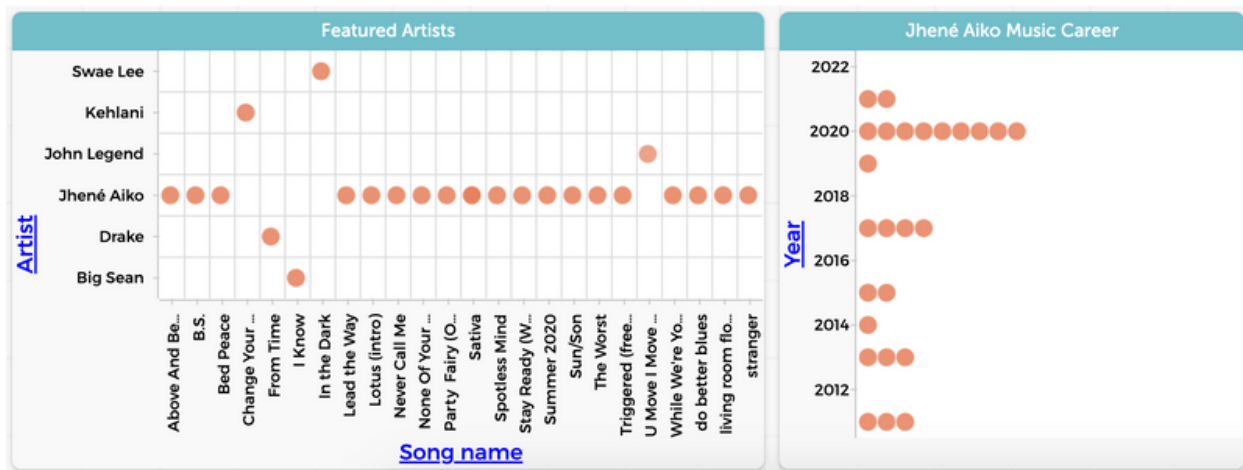
Figure 4. Two visualizations created by Janeese in CODAP: Jhené Aiko's collaborations with featured artists (Left) and her career timeline based on the years her songs were released (Right).

When presenting her reflections on the project in class, Janeese concluded, "The clear conclusion I received from the data was that Jhené Aiko is very popular and has made a lot of hits over the years. I have found out that Jhené Aiko was more popular than I thought she was." When considering the learnings related to data science, Janeese emphasized that the project enabled her to discover new information and use the data to answer her research questions; "The data that I've collected was very helpful and useful; it helped me gather more information and determine if I was correct about my theories."

Discussion

In the ever-evolving landscape of technology and the rapid expansion of the role of data in our lives, it is important to prepare today's students to be informed and empowered data-literate citizens. The ambitious approach presented here seeks to achieve this goal by situating high school data science instruction in the interests and lived experiences of today's students. Our work shows how students' passions can serve as an engaging context for introducing them to data science. Furthermore, we present a novel instructional approach that uses innovative tools and professional platforms as a way to grant learners access to authentic data sets to ground their data science learning experiences. In doing so, API Can Code is authentic, interest driven, and can help learners understand the data that surrounds them.

This approach draws on interest development theory, both within and beyond computing education (Azevedo 2013; Kafai & Peppler, 2011). The interest-driven nature of the inquiry approach taken here not only provides essential context and meaning to the data field (Makar & Ben-Zvi, 2011; Pfannkuch, 2011) but also showcases the potential of data science instruction by grounding it in students' interests and real-world applications. In this article, we hope to connect these academic findings with practical ideas and strategies that can be used in classrooms.

The hands-on experience with authentic, meaningful data makes learning more engaging and data science more relevant. It also equips students with skills applicable beyond the classroom. As the findings suggest, when students are allowed to choose projects based on their interests, they become more invested in the learning process, gaining valuable insight into the role data plays in their lives. Engaging with the data they choose provides a unique opportunity for students to pose meaningful questions, critically analyze data, and develop problem-solving skills to gain insights into topics they genuinely care about. This sense of ownership not only boosts motivation and engagement but also leads to a deeper understanding of data science principles.

Collectively, with this work, we hope to inspire and support educators to think creatively on how to introduce students to data science in ways that align with their experiences living in a data-filled world. In doing so, we can better prepare them to thrive in the data-rich world that awaits them.

Disclosure Statement

References

Azevedo, F. S. (2013). The tailored practice of hobbies and its implication for the design of interest-driven learning environments. *Journal of the Learning Sciences*, *22*(3), 462–510. https://doi.org/10.1080/10508406.2012.730082

Biehler, R., Veaux, R. D., Engel, J., Kazak, S., & Frischemeier, D. (2022). Research on data science education. *Statistics Education Research Journal*, *21*(2), Article 2. https://doi.org/10.52041/serj.v21i2.606

Brooks, C., Quintana, R. M., Choi, H., Quintana, C., NeCamp, T., & Gardner, J. (2021). Towards culturally relevant personalization at scale: Experiments with data science learners. *International Journal of Artificial Intelligence in Education*, *31*(3), 516–537. https://doi.org/10.1007/s40593-021-00262-2

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, *9*(2), 1–22. https://doi.org/10.14763/2020.2.1481

CODAP. (2022). *CODAP - Common Online Data Analysis Platform*. Concord Consortium. https://codap.concord.org/

Coenraad, M., Palmer, J., Franklin, D., & Weintrop, D. (2019). Enacting identities: Participatory design as a context for youth to reflect, project, and apply their emerging identities. In J. A. Fails (Ed.), *IDC '19:*

Proceedings of the 18th ACM International Conference on Interaction Design and Children (pp. 185–196). ACM. https://doi.org/10.1145/3311927.3323148

DiSalvo, B. (2016). Participatory design through a learning science lens. In J. Kaye & A. Druin (Eds.), *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4459–4463). ACM. https://doi.org/10.1145/2858036.2858405

Franklin, D., Coenraad, M., Palmer, J., Eatinger, D., Zipp, A., Anaya, M., White, M., Pham, H., Gökdemir, O., & Weintrop, D. (2020). An analysis of Use-Modify-Create pedagogical approach's success in balancing structure and student agency. In A. Robins, A. Moskal, A. J. Ko, & R. McCauley (Eds.), *ICER '20: Proceedings of the 2020 ACM Conference on International Computing Education Research* (pp. 14–24). ACM. https://doi.org/10.1145/3372782.3406256

Gould, R. (2021). Toward data-scientific thinking. *Teaching Statistics*, *43*(S1), S11–S22. https://doi.org/10.1111/test.12267

International Data Science in Schools Project Curriculum Team. (2019). *Curriculum frameworks for introductory data science*.

Israel-Fishelson, R., Moon, P. F., Pauw, D., & Weintrop, D. (2024). Exploring interest-driven data science through participatory design. In R. Lindgren, T. I. Asino, E. A. Kyza, C. K. Looi, D. T. Keifert, & E. Suárez (Eds.), *ICLS 2024: Proceedings of the 18th International Conference of the Learning Sciences* (pp. 1159–1162). International Society of the Learning Sciences. https://doi.org/10.22318/icls2024.793415

Kafai, Y. B., & Peppler, K. A. (2011). Youth, technology, and DIY: Developing participatory competencies in creative media production. *Review of Research in Education*, *35*(1), 89–119. https://doi.org/10.3102/0091732X10383211

Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A call for a humanistic stance toward K–12 data science education. *Educational Researcher*, *50*(9), 664–672. https://doi.org/10.3102/0013189X211048810

Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning*, *13*(1–2), 1–4. https://doi.org/10.1080/10986065.2011.538291

Michaelis, J. E., & Weintrop, D. (2022). Interest development theory in computing education: A framework and toolkit for researchers and designers. *ACM Transactions on Computing Education*, *22*(4), Article 43. https://doi.org/10.1145/3487054

Ng, W. (2012). Can we teach digital natives digital literacy? *Computers & Education*, *59*(3), 1065–1078. https://doi.org/10.1016/j.compedu.2012.04.016

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, *13*(1–2), Article 1–2. https://doi.org/10.1080/10986065.2011.538302

Renninger, K. A., & Hidi, S. (2015). *The power of interest for motivation and engagement*. Routledge.

Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, *33*(2), 163–180. https://doi.org/10.1177/0165551506070706

Sinha, S. (2020). Big data analysis: Concepts, challenges and opportunities. *International Journal of Innovative Research in Computer Science & Technology* (IJIRCST), *8*(3). https://doi.org/10.21276/ijircst.2020.8.3.29

Wilkerson, M. H., & Polman, J. L. (2020). Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences*, *29*(1), 1–10. https://doi.org/10.1080/10508406.2019.1705664

Footnotes

1. The full curriculum as well as accompanying materials can be accessed at https://apicancode.umd.edu. ↩

References

- Azevedo, F. S. (2013). The tailored practice of hobbies and its implication for the design of interest-driven learning environments. *Journal of the Learning Sciences*, *22*(3), 462–510. https://doi.org/10.1080/10508406.2012.730082

  ↩

- Biehler, R., Veaux, R. D., Engel, J., Kazak, S., & Frischemeier, D. (2022). Research on data science education. *Statistics Education Research Journal*, *21*(2), Article 2. https://doi.org/10.52041/serj.v21i2.606

  ↩

- Brooks, C., Quintana, R. M., Choi, H., Quintana, C., NeCamp, T., & Gardner, J. (2021). Towards culturally relevant personalization at scale: Experiments with data science learners. *International Journal of Artificial Intelligence in Education*, *31*(3), 516–537. https://doi.org/10.1007/s40593-021-00262-2

  ↩

- Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, *9*(2), 1–22. https://doi.org/10.14763/2020.2.1481

    ↩

- CODAP. (2022). CODAP - Common Online Data Analysis Platform. Concord Consortium. https://codap.concord.org/

    ↩

- Coenraad, M., Palmer, J., Franklin, D., & Weintrop, D. (2019). Enacting identities: Participatory design as a context for youth to reflect, project, and apply their emerging identities. In J. A. Fails (Ed.), *IDC '19: Proceedings of the 18th ACM International Conference on Interaction Design and Children* (pp. 185–196). ACM. https://doi.org/10.1145/3311927.3323148

    ↩

- DiSalvo, B. (2016). Participatory design through a learning science lens. In J. Kaye & A. Druin (Eds.), *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4459–4463). ACM. https://doi.org/10.1145/2858036.2858405

    ↩

- Franklin, D., Coenraad, M., Palmer, J., Eatinger, D., Zipp, A., Anaya, M., White, M., Pham, H., Gökdemir, O., & Weintrop, D. (2020). An analysis of Use-Modify-Create pedagogical approach's success in balancing structure and student agency. In A. Robins, A. Moskal, A. J. Ko, & R. McCauley (Eds.), *ICER '20: Proceedings of the 2020 ACM Conference on International Computing Education Research* (pp. 14–24). ACM. https://doi.org/10.1145/3372782.3406256

    ↩

- Gould, R. (2021). Toward data-scientific thinking. *Teaching Statistics*, *43*(S1), S11–S22. https://doi.org/10.1111/test.12267

    ↩

- International Data Science in Schools Project Curriculum Team. (2019). *Curriculum frameworks for introductory data science*.

    ↩

- Israel-Fishelson, R., Moon, P. F., Pauw, D., & Weintrop, D. (2024). Exploring interest-driven data science through participatory design. In R. Lindgren, T. I. Asino, E. A. Kyza, C. K. Looi, D. T. Keifert, & E. Suárez (Eds.), *ICLS 2024: Proceedings of the 18th International Conference of the Learning Sciences* (pp. 1159–1162). International Society of the Learning Sciences. https://doi.org/10.22318/icls2024.793415

↩

- Kafai, Y. B., & Peppler, K. A. (2011). Youth, technology, and DIY: Developing participatory competencies in creative media production. *Review of Research in Education*, *35*(1), 89–119. https://doi.org/10.3102/0091732X10383211

↩

- Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A call for a humanistic stance toward K–12 data science education. *Educational Researcher*, *50*(9), 664–672. https://doi.org/10.3102/0013189X211048810

↩

- Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning*, *13*(1–2), 1–4. https://doi.org/10.1080/10986065.2011.538291

↩

- Michaelis, J. E., & Weintrop, D. (2022). Interest development theory in computing education: A framework and toolkit for researchers and designers. *ACM Transactions on Computing Education*, *22*(4), Article 43. https://doi.org/10.1145/3487054

↩

- Ng, W. (2012). Can we teach digital natives digital literacy? *Computers & Education*, *59*(3), 1065–1078. https://doi.org/10.1016/j.compedu.2012.04.016

↩

- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

↩

- Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, *13*(1–2), Article 1–2. https://doi.org/10.1080/10986065.2011.538302

↩

- Renninger, K. A., & Hidi, S. (2015). The power of interest for motivation and engagement. Routledge.

↩

- Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, *33*(2), 163–180. https://doi.org/10.1177/0165551506070706

↩

-

↩

- Wilkerson, M. H., & Polman, J. L. (2020). Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences, 29*(1), 1–10. https://doi.org/10.1080/10508406.2019.1705664

↩