

Make with Data: Challenging and contextualizing open-source data with personal and local knowledge

Nathan Holbert, Teachers College, Columbia University, holbert@tc.columbia.edu
Cassie Xu, Earth Institute, Columbia University, cassie@ei.columbia.edu

Abstract: In this poster we describe Make with Data, a two-year project that invites teachers and students from public high schools to work with professional data scientists and open-source data to explore issues important to their local community. While the negotiation of the personal and the quantitative resulted in tensions, Make with Data students found their personal experiences a useful tool for adding context and complexity to the phenomena being studied.

Purpose and significance of the work

In the Make with Data project described here, we engage learners age 16-17, from groups underrepresented in STEM domains, to work together with educators and data scientists in an after-school setting to identify and understand a local community challenge. Leveraging the constructionist design paradigm and research on project-based service learning, the project investigates how framing data science practices as a means of contributing to and improving one's community can impact students' understanding of data and data science practices. By encouraging learners to identify challenges that impact their local community, and supporting them as they use open-source data in collaboration with educators, professional data scientists, and other STEM professionals, students developed a rich understanding of how data can be used to address local needs, which in turn positively impacted their interest in data science related careers.

However, because students have personal experiences related to these local challenges, their exploration of the data inevitably involves negotiating their own perspectives with what they see in the data (Rubel et al., 2017; Wilkerson & Laina, 2018). In this poster we examine how students leveraged their personal experiences in their local community to further elucidate the context and complexity of trends found in open-source data.

Methodology

In the first year of the Make with Data project, teachers were recruited from four boroughs of New York City. Participating teachers were mostly science educators with one pair being social studies teachers. These teachers then selected one to two students from their school to be "student leaders" for the duration of the project.

During this first year, all participants jointly worked through the identification of a locally relevant issue, and in smaller groups set out to explore and find open-source datasets to better understand the issue and propose potential solutions. The teacher and student teams jointly attended approximately 9 sessions with the other teacher-student teams, the project leaders, and professional data scientists. Additional conference calls and meetings between teams and data scientists occurred throughout the year as needed. Full team meetings took participants through a process of brainstorming possible project ideas, learning how to access relevant open-source data, exploring the use of data analysis tools, tinkering with data to explore possible trends, and developing representations and presentations to communicate their findings.

The following year, teacher and student teams implemented a new Make with Data-type project in their own school. Each teacher and student team recruited new students from their school to join their group, developed their own meeting schedules, and—following a structure similar to the one used during the first year of the project—worked with these students to identify and explore locally relevant community challenges. In total, the team (made up of the project leads, data scientists, and students and educators) grew to approximately 50 individuals distributed across four schools.

In this poster we describe the two-year data explorations of two Make with Data students. These cases illustrate the way in which participants' personal experiences with their chosen topic drove their search for data and impacted their interpretation of discovered datasets.

Findings

As expected, students in the Make with Data project took on and explored topics relevant to the experiences and needs of their local community. However, as we will show below, the personal nature of these topics led to an important back and forth between the data, the original research question, and the students' preconceptions.

In the first year of the project, we invited teacher and student participants to consider their own experiences, and those of their community to identify a category of challenges that they would like to explore further. The team from the Bronx chose to focus on student mental health. Speaking about the mental health

challenges faced by her and her classmates, Bronx-S (student) stated, “I’ve experienced mental breakdowns, and I think mental health for students is really important [...] It has affected me because I’ve had anxiety and stress over school. And that has affected me at home too, my attitude just changes.”

The Bronx team recognized that the geographic layout of their particular borough and school limited access to public transportation so using GIS software, they mapped out the location of mental health facilities, schools throughout the Bronx, and public train and bus lines. However, rather than rely solely on the GIS Data, Bronx-S read her own experiences into the data asserting, “the map and data doesn’t tell the whole story” (Van Wart et al., 2020). For example, while there was one mental health facility relatively close to her school, this lone facility served a large number of schools in the area. Furthermore, getting to this facility required overcoming many geographic obstacles—such as moving through areas of high crime—as well as personal ones—like the expectation to come home after school to take care of younger siblings.

Together these life experiences led the Bronx team to build a board game (based on the classic board game “Life”) where players try to make it to the local mental health facility while random events and continually building depression slow their progress. In their project the Bronx team illustrated that access to mental health support is more than just the measure of distance on a map. Rather, their game narrates how a complex interaction of factors—that includes things that can be observed in public datasets as well as factors that are personal, hidden, but no less relevant—impact how students navigate access to mental health support (Wilkerson & Laina, 2018).

Another student, Staten-S, chose to examine marijuana use among teens. While she quickly found relevant datasets, each came with unique challenges and contradictions. For example, when looking at arrest data, Staten-S realized that the NYPD no longer tracks the possession of small amounts of marijuana meaning this data might underrepresent marijuana possession. In Staten-S’s words, “not everyone gets caught.” This led Staten-S to examine data on “Stop and Frisk.” She found that incidents of stop and frisk overwhelmingly occurred on Black males (10 times as often as their white counterparts) and showed a miniscule number of incidences resulting in possession charges. Believing this dataset to indicate racist policing policy, Staten-S sought out an alternate dataset, eventually finding a survey on substance abuse. In this survey, more white students reported using cannabis than Black and Latino students. As these results were different than her expectations—expectations that Staten-S admitted were based on problematic stereotypes—she wondered if self-report on a survey could be trusted?

When presenting this work at the end of the second year, Staten-S described the extent of her search and argued that while she had tried to look at this issue from multiple angles using a number of different datasets, “the trends weren’t strong enough to come to a specific conclusion.” In the end Staten-S believed the data she was exploring obfuscated things like racist policing practices and data collection methods she wasn’t sure she trusted leaving her with “more questions” than answers (Wilkerson & Laina, 2018).

Conclusions

The Make with Data project’s central hypothesis is that interest in STEM for underrepresented youth can be substantially impacted by connecting the practice of science with students’ personal and communal values and experiences. Through the creation of a learning environment that situates data science concepts and practices in local phenomena and challenges, these two cases show how personal experiences and knowledge about the local context can support data science novices as they make meaning out of complex datasets. While the negotiation of the personal and the quantitative can result in tensions, Make with Data students found their personal experiences a useful tool for adding context and complexity to the phenomena being studied.

References

- Rubel, L. H., Hall-Wieckert, M., & Lim, V. Y. (2017). Making Space for Place: Mapping Tools and Practices to Teach for Spatial Justice. *Journal of the Learning Sciences*, 26(4), 643–687. <https://doi.org/10.1080/10508406.2017.1336440>
- Van Wart, S., Lanouette, K., & Parikh, T. S. (2020). Scripts and Counterscripts in Community-Based Data Science: Participatory Digital Mapping and the Pursuit of a Third Space. *Journal of the Learning Sciences*, 29(1), 127–153. <https://doi.org/10.1080/10508406.2019.1693378>
- Wilkerson, M. H., & Laina, V. (2018). Middle school students’ reasoning about data and context through storytelling with repurposed local data. *ZDM*, 50(7), 1223–1235. <https://doi.org/10.1007/s11858-018-0974-9>

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1759224