Inquiry Learning with Data and Visualization in the STEM Classroom

Samantha W. Michalka [1]
James Dalphond [2]
Fred G. Martin [2]
[1] Machine Science; [2] University of Massachusetts Lowell

Abstract: Understanding data is a crucial 21st century skill for students in STEM education. At the core of the latest science and math standards is the principle that students should be able to apply real world science and engineering principles to work with and understand data. Here, we discuss three styles of integrating collaborative data analysis and exploration in the classroom using a web-based data visualization system (iSENSE) that was purpose-built for middle through high school use. We discuss each of these styles/strategies, potential benefits and reasons for use, and give concrete examples of classroom use by teachers.

Introduction

Frameworks for science education in the United States call for students to learn data analysis methods, including visualization techniques at all levels of education (Schweingruber et al. 2012). From an early age, standards require students to learn how to collect and think about data to draw conclusions. Since 2006, the Massachusetts Department of Education has required that high school students learn Scientific Inquiry Skills (Massachusetts Department of Education 2006). These skills include designing and conducting scientific investigations and analyzing resulting data to formulate and test hypotheses (Chang et al. 2003).

The National Research Council, the National Science Teachers Association, and the American Association for the Advancement of Science with representatives from 26 states have worked together to develop the Next Generation of Science Standards (NGSS 2013). The purpose of these standards is to promote the idea of inquiry and investigation of data and to prepare students for college and careers in industry. The NGSS have been adopted by 14 states with others in the process of adoption (Academic Benchmarks 2015).

In parallel with these science standards, new math standards promote the idea of using data to learn core concepts. The Common Core math standards released in 2010 were designed to prepare students for success globally and were voluntarily adopted by 42 states (Common Core 2015a). The core standards for high school math focus on modeling and interpretation of concepts including representations of relationships in data. Statistics, probability, geometry, algebra, and other high school mathematics courses are affected by these standards. In probability and statistics classes, the use of technologies is specifically encouraged to generate plots, regression functions, and simulations to test predictions (Common Core 2015b). Through these standards, there is a clear theme of teaching students the skills they need to understand data.

Understanding of how to work with data is also recognized as a central part of computational thinking. The College Board's Computer Science (CS) Principles project has identified fluency with data as one of seven "Big Ideas" (Astrachan and Briggs 2012). In the CSTA/ACM K-12 Computer Science Standards (2011), data representation and working with data are part of the computational thinking strand.

The iSENSE Web-Based Data Visualization System

iSENSE is a web-based data visualization system that allows students to collect, visualize, and share data-centric projects. The iSENSE website offers a shared repository of user-contributed data collection activities, such as engineering projects, environmental studies, classroom science experiments, and surveys, together with the data produced by those activities (Dalphond et al. 2013, Lee et al. 2014, Willis et al. 2015, Cogger 2015). The system

enables users to contribute their own activities, upload data, and explore dynamic data visualizations. During exploration, visualizations can be saved and shared with collaborators. iSENSE is a National Science Foundation-funded research project, and its technology is open and free to use.

Work on iSENSE began at University of Massachusetts Lowell in 2008 with the goal of building an end-to-end system where students could collect data and share it with classmates or across schools. Since then, iSENSE has been iteratively developed with direct feedback from teachers and students. iSENSE is made up of 4 main components: projects, fields, data sets, and visualizations. A project is the framework on which to build upon for a specific activity. In order to contribute data to a project, it must contain one or more fields (i.e., variables), which describe the structure of the data being collected. Once a project is created and its fields defined, users may contribute data sets, which contain data for all or a subset of the fields. After data has been contributed to a project, users can explore their data dynamically using the visualization tools.

iSENSE has been designed to streamline the creation of projects and reduce the number of steps required to visualize data. To create a project, a teacher enters a relevant project name for the activity. This creates the project with that name and provides the tools to fill in other information and related materials. Field creation can be done in two ways. The first way to create fields is to create them manually on the website. The second method is to simply upload a spreadsheet containing data and the system will determine the fields and their types automatically. To allow students to contribute data to the project, teachers may have their students create accounts, or they may simply add a "contributor key" that students can then use to enter data. Data may be entered manually on the website, uploaded from a file, or uploaded through the API or mobile phone apps. As of this writing, over 600 users have created 1,200 projects and over 10,000 data sets.

Three Models for Working with Data in the Classroom

In our work with teachers and students using our data visualization system, we have developed three models of typical usage. As illustrated in Fig. 1, each has its own considerations and benefits. The three models are:

(1) Using student-collected data sets. Here, students (individually or in small teams) gather data (conduct experiments) and contribute it to a project in the system. Having the students' data in a central, structured location allows the classroom participants to gain meaning from each others' contributions.

(2) Working with pre-existing data organized into "curated" data sets. Here, teachers prepare data sets from a trusted source, and students make sense of the data in an inquiry-based process. Also, students can be engaged in creating their own curated data sets. This usage model allows students to gain or reinforce domain or content expertise while developing data analysis skills.

(3) A hybrid model where students contribute experimental data to a previously curated data set. In this model, the trusted curated data set is used to help students find meaning in the data that they gather experimentally.

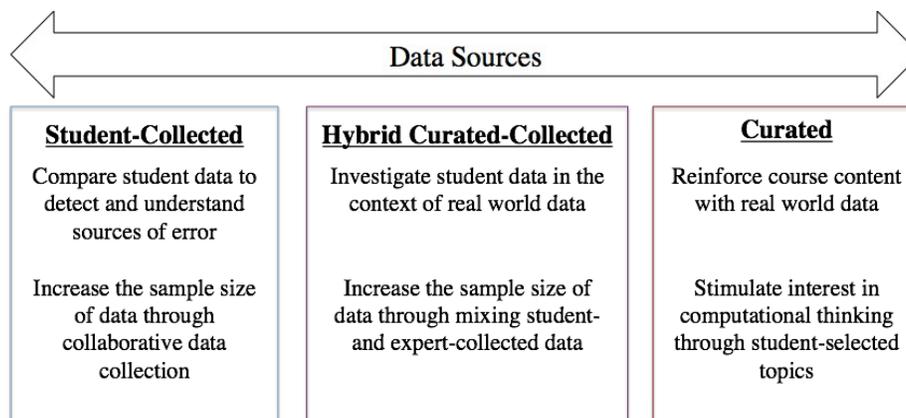The following sections present details about each of these usage models.

Figure 1. Three approaches to working with data.

Using Student-Collected Data Sets

In the student-collected model, individual students or groups gather and share their own data. This typically occurs in some form of a laboratory experiment or simulation but sometimes extends to surveys or researching outside sources. In one common use case, the entire class reviews the procedure and hypothesis and then individual students or teams conduct the experiment and record the appropriate data. Students or teams upload their own data to iSENSE and independently assess their findings. Students then compare their own data to data submitted by their peers (either in the same class or a larger group). This strategy of individual collection and investigation allows students to build a concrete understanding of the experiment and variables. The subsequent combining of data with other students offers two primary benefits: 1) detecting sources of error and 2) increasing the amount of data available for interpretation.

All scientific experiments involve some degree of error, both from systematic and random sources. When trying to generalize to abstract concepts and relationships from experimental data, errors can add a challenge for student learning. For example, if a student accidentally omits a negative sign, they may interpret their recorded data correctly, but incorrectly learn the relationship between to two variables. When students compare their data to that of their peers, they are able to identify sources of error (either independently or with teacher guidance). In a recent classroom laboratory, students calculated density of clay by measuring mass and volume. When the students compared their data (see Fig. 2), one group quickly noticed that their volume measurement and their density calculation were both very different from all other groups. The group was able to revisit their data and find their calculation error (dividing volume by mass instead of mass by volume) and their error in recording the volume. The teacher commented that measurement and calculation errors such as these had likely been happening throughout her years of teaching but that combining and comparing student data highlights these errors and provides students with immediate feedback.
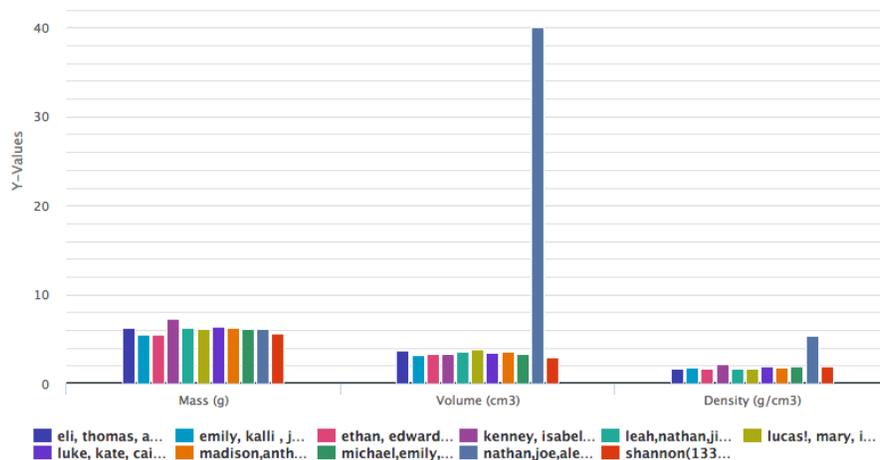
Figure 2. Bar plots showing the mass, volume, and density of pieces of clay for 11 student groups. Comparing group data highlights the volume and density errors made by one group.

Combining student-collected data allows a class to efficiently collect data while still providing each student with a hands-on data collection experience. This is especially beneficial when manipulating an independent variable over a large range or when a large about of data is needed to observe a subtle effect. We recently conducted a class experiment in which students measured the decrease in temperature during a vinegar and baking soda reaction. The amount of vinegar was held constant, but students were randomly assigned different amounts of baking soda. Each student measured the decrease in temperature over 60 seconds and contributed their data to the project (see Fig. 3). This design allowed the class to cover a range of baking soda amounts while minimizing class time spent on data collection. In this example, multiple students use the same amount of baking soda to facilitate a discussion about measurement error, but we could have alternatively extended the range of baking soda amounts.
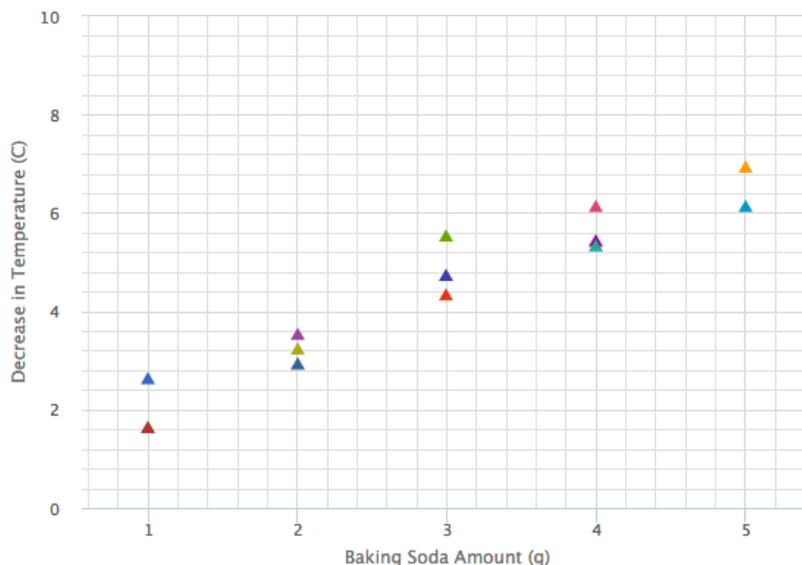

Figure 3. Scatter plot showing the relationship between the amount of baking soda adding to vinegar and the decrease in temperature. Each colored triangle represents data from an individual student.

Collaborative data collection increases the amount of data available for exploration and interpretation; this is especially important in probability and statistics and in experiments when the effect size is small. In one collaborative experiment, students each rolled a pair of dice 10 times and recorded the dice values and their sum for each roll. Fig. 4 shows the histogram of dice sums for a single student (left), 15 students (middle), and approximately 70 contributors (right). Notably, the dice sum distribution for the single student does not reflect the

theoretical triangle distribution; however, the sum distribution approaches the theoretical distribution as the number of contributors increases.
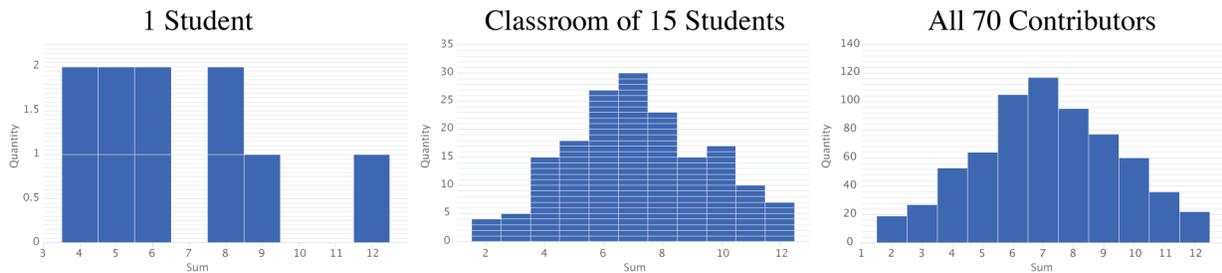

Figure 4. Histograms of the sum of rolling two dice. Data shown from 1 student conducting 10 rolls (left), from 15 students (middle), and 70 students (right). Note the different scales of the y-axes.

Using Curated Data Sets

We use the term "curated" to refer to data sets that have been edited and prepared for student use with the assumption that the data sets contained in them are accurate. The original source of the data may be published reports from scientists or government organizations, data sets published by other groups (e.g., sports statistics), or well-known facts. We have encouraged teachers: (1) to create projects that contain data sets that are relevant to their instructional goals, and create assignments that support their students in exploring these data sets; (2) to encourage their students to explore existing data sets on the site, choosing subjects that pique their interest, and using this as the basis for investing in understanding the data; and (3) to support their students in creating original curated-data projects, sharing their interests with others and learning how to structure data for narrative purposes in the process.

We seeded this work with a variety of general-interest data sets. For example, we imported data published by Major League Baseball that includes business statistics like home attendance, payroll, and home field capacity, plus sports performance data such as team hits, runs, batting average, and the newer "on base plus slugging" percentage. Using a heat map view, it's easy to see which teams have the most money to spend, and how this correlates to other statistics of the team performance. For fans of baseball, their domain knowledge allows them to represent complex relationships among the sports data.
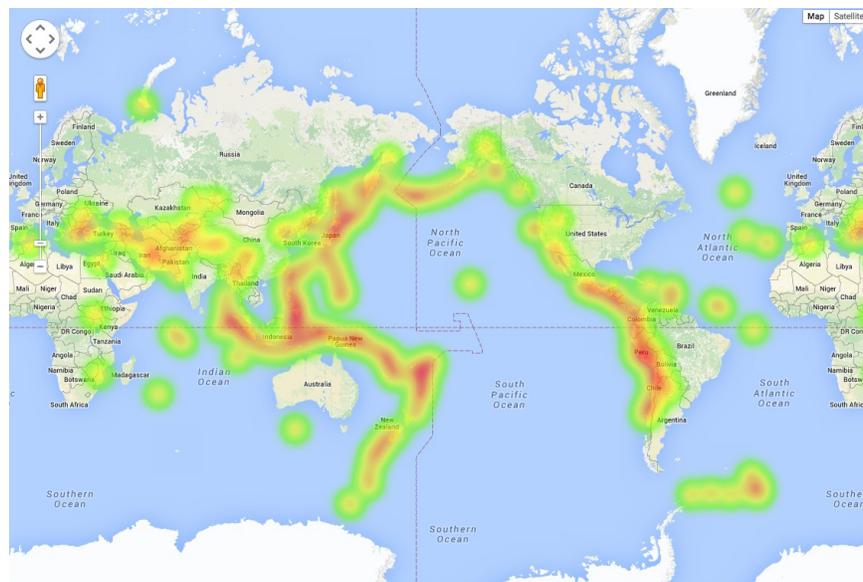

Figure 5. Heatmap showing location and magnitude of earthquakes of 7.0+ magnitude since 1973, using data from the USGS.

We created a series of environmental data sets, including hurricane data (showing relationships between wind speed and air pressure, and wind speed and declared class of hurricane); atmospheric carbon dioxide levels (showing periodic increases and decreases over the last 650,000 years, followed by a contemporary spike); and earthquake data (showing magnitude, location, and time of earthquakes). Fig. 5 shows a map visualization of the latter.

Teachers have built a variety of curated data sets for their students' use, including projects showing: statistics on cyber-bullying, using data from the National Center for Education Statistics that shows relative proportions of different types of bullying, and rates from four regions of the country; earth science content matter (minerals, their hardness and other properties, and their countries of origin); solar system content matter (relationships among planetary distance from the sun, mass, period of revolution, and period of rotation); and world population (most populous countries in 2015, and projected most-populous in 2050). The teachers' instructional goals were for their students to simultaneously learn about the subject matter while learning to work with data.

An alternative use of curated data sets is to teach students about research and data analysis through topics that they find interesting. This strategy was recently used in a science fair preparation mini-course. Students learned about finding and citing reliable online sources of data and how to organize data in a spreadsheet. Students then created their own projects about a topic they found interesting by finding a reliable data source, formatting the data appropriately, creating a project on iSENSE, and exploring the data. Students then composed a short report about their project, including key data visualizations. Student-selected topics included goals scored by star soccer plays over several years, mortality rates of Alzheimer's disease and Parkinson's disease by state, gun violence by state, and height by country. Here, the teacher's instructional goals were for students to learn about online research and appropriate graph use.

Using Hybrid Curated-Collected Data Sets

Student-collected and user-curated data offer different advantages for learning with data and sparking student interest; the process of collecting one's own data can create a sense of ownership and curated data from external sources can connect content knowledge to the world outside of the classroom. In some cases, these two strategies can be combined into a hybrid curated-collected model.

In one example of the hybrid curated-collected strategy, students learn about clean water through a mix of data exploration and a hands-on water testing field trip. Prior to the field trip, students are introduced to several water quality measurements (e.g., temperature, pH, dissolved oxygen) through class discussion and an activity in which they explore a large expert-collected data set from a local river. During the field trip, students test water samples from the local river and compared their data to historical expert-collected data from the same water source. Students assessed whether their measured temperature, pH, dissolved oxygen, and phosphate levels fell within expected ranges. The timeline shown in Fig. 6 displays historical data from a previous year as a blue line and the student-collected data as three separate points (one color for each group). Teams of students presented their findings to their peers, using iSENSE as a visual aid. In this hybrid curated-collected format, students were able to gain concrete understanding of the water quality measurements through hands-on experience while gaining context from the larger expert-collected data set.
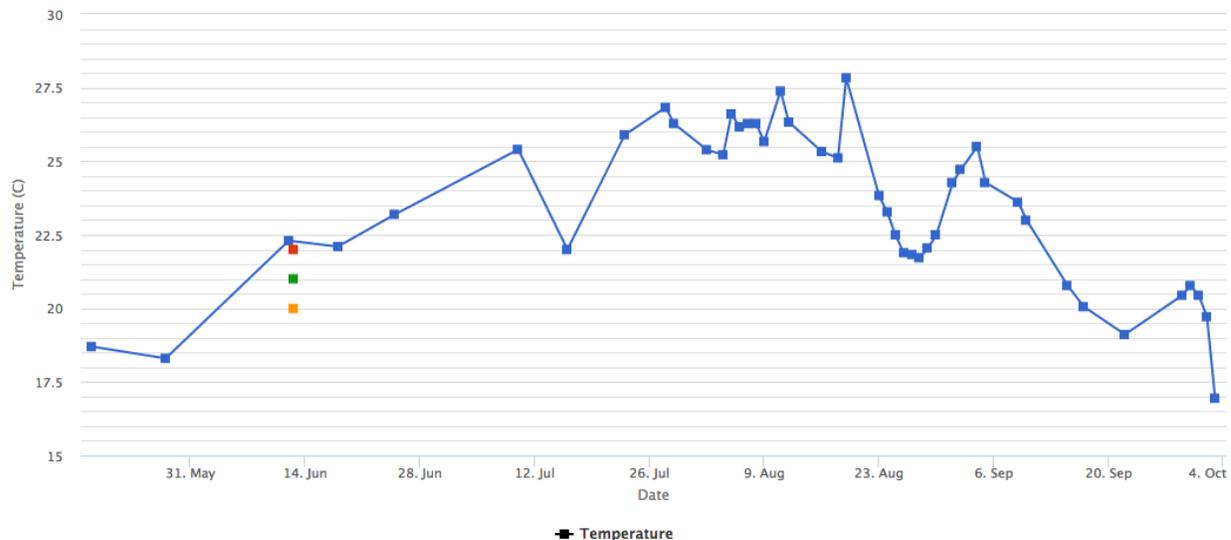
Figure 6. Temperature of Merrimack River during the summer/fall of 2012, with three
student group temperature measurements taken on June 12, 2015.

Conclusions

In a large study that focuses on how schools and school systems are integrating technology into students' learning experiences, the Organization for Economic Cooperation and Development reported that "adding 21st century technologies to 20th century teaching practices will just dilute the effectiveness of teaching" (OECD 2015, foreword). The report makes it clear that investments in teachers' own learning are necessary, so they can develop uses of technology that improve students' learning.

Our work has demonstrated how teachers can use a web-based data visualization platform to support their students' learning with use of data. Collaboration among learners—sharing their data sets, and examining one's own data in the context of others'— is particularly effective in supporting this sense-making. These examples highlight innovative teaching with innovative technology, which brings learning into the 21st century.

References

Academic Benchmarks (2015). Next Generation Science Standards Adoption Map. *http://academicbenchmarks.com/next-generation-science-standards-adoption-map/*

Astrachan, O., & Briggs, A. (2012). The CS principles project. *ACM Inroads*, 3(2), 38-42.

Chang, K.-E., Sung, Y.-T., and Lee, C.-L. (2003). Web-based collaborative inquiry learning. Journal of computer assisted learning, 19(1):56–69.

Cogger, S. (2015). Doing the Data Walk. The Science Teacher, 82(2), 43.

Common Core (2015a). About the standards. *http://www.corestandards.org/about-the-standards/*

Common Core (2015b). Common core state standards for mathematics. *http://www.corestandards.org/wp-content/uploads/Math_Standards.pdf*

Dalphond, J., Martin, F., Rudnicki, I., and Scribner-MacLean, M. (2013) Transforming Science and Engineering Classrooms with Online Collaborative Tools. International Society for Technology in Education Conference. http://www.isteconference.org/uploads/ISTE2013/HANDOUTS/KEY_80762835/ISTE_2013_iSENSE_RP.pdf

Lee, I., Martin, F., & Apone, K. (2014). Integrating computational thinking across the K–8 curriculum. ACM Inroads, 5(4), 64-71.

Massachusetts Department of Education (2006). Massachusetts science and technology/engineering curriculum framework. *http://www.doe.mass.edu/frameworks/scitech/1006.pdf*

NGSS (2013). The next generation science standards executive summary. *http://www.nextgenscience.org/next-generation-science-standards*

OECD (2015), Students, Computers and Learning: Making the Connection, PISA, OECD Publishing, Paris.

Schweingruber, H., Keller, T., Quinn, H., et al. (2012). A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. National Academies Press.

Seehorn, D., Carey, S., Fuschetto, B., Lee, I., Moix, D., O'Grady-Cunniff, D., ... & Verno, A. (2011). CSTA K–12 Computer Science Standards: Revised 2011.

Willis, M. B. N., Hay, S., Martin, F. G., Scribner-MacLean, M., and Rudnicki, I. (2015). Probability with Collaborative Data Visualization Software. Mathematics Teacher, 109(3), p. 194–199.