

Designing and Exploring a Model for Data Science Learning for Middle School Youth

Award Numbers: DRL-1742255 and DRL-1741989 (collaborative grant)

PIs: Andee Rubin, TERC (andee_rubin@terc.edu) and Jan Mokros, Science Education Solutions (jmokros@scieds.com)

Partners: Gulf of Maine Research Institute, Maine; Program Evaluation Research Group (PERG) at Endicott College, Mass.; Oxford Hills School District, Maine (summer camp program); Malden, Mass. YMCA (summer camp program); Girls Inc., Lynn, Mass. (after-school program)

Project Overview:

What aspects of data science are accessible to middle school youth, given appropriate technology, datasets, and materials? We are developing and implementing out-of-school data science modules with community partners in urban Massachusetts and rural Maine communities. We target middle school participants from populations under-represented in STEM, including girls, youth of color, economically disadvantaged, and rural youth. Youth participate in 10-hour modules after school or in summer camp. They use the Common Online Data Analysis Platform (CODAP) as a data analysis tool, in conjunction with large, pre-existing datasets and data that they collect themselves. Each module focuses on a high-interest scientific or social science topic, selected with input from youth advisory groups. The project is developing three modules, a survey measure of data science dispositions, and an interview assessment of students' ability to ask and answer questions with data. The project integrates computer science, the mathematics of data, and scientific topics.

Goals for Youth:

- Appreciate the ubiquity of data and the potential for learning from data
- Be aware of the complexities of measurement and look at data through these complexities
- Understand the case/attribute structure of data
- Have experience with describing distributions and distributional shape
- Examine relationships of attributes within a dataset
- Understand how data representations are constructed by mapping attribute values to representational elements, both on the computer and off
- Learn to address questions about data through a series of “data moves” with technology
- Engage with social science and scientific data in ways that are intellectually and personally satisfying

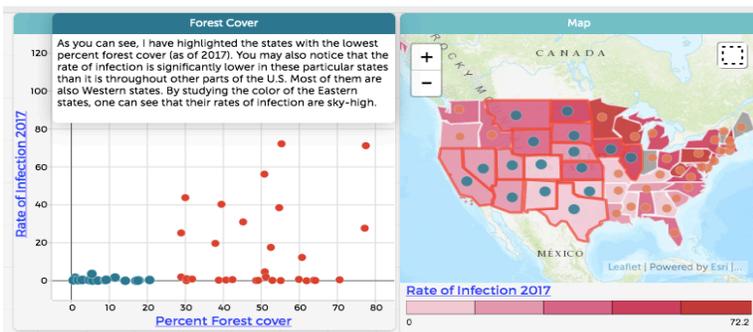


Image: Example of a student's data representation in CODAP from the Ticks and Lyme module. Rate of forest cover is compared to the rate of Lyme in each state. Green dots represent states with a low percentage of forest cover. Red dots represent states with medium to high forest cover.

Accomplishments/Outcomes:

We have developed and implemented two modules, including accompanying data sets, on the topics of: **Teens and Time** (using Pew Research Institute and Census at School datasets) and **Ticks and Lyme Disease** (using CDC and NOAA data sets). A consistent finding from this work is: *youth can develop iterative questions about the datasets and easily use CODAP to investigate their questions, especially about relationships between attributes.*