Goal: Explore cognitive, technical, and pedagogical issues in bringing large scientific data sets into the high school classroom. How can we use student interest in topics addressed by these data sets to teach data and statistics ideas, as well as a deeper understanding of the science?

Compelling Data

Exoplanet Transit Light Data
From http://www.superwasp.org/wasp_planets.htm

Despite lots of variability in individual measurements, large data sets allow us to see a small periodic drop in stellar magnitude indicating a planet in orbit around a distant star.

How do students make sense of the increased precision of having lots of data? What tools can help them deal with the periodic/time series nature of these data?

Opportunities, Challenges & Questions

• Large data sets may help students focus on the whole/aggregate rather than individual points—a key cognitive shift.
• Large amounts of data increase precision, allowing people to see even small differences that are statistically significant (e.g., Exoplanet Transit Data). How do students think about “statistical” vs. “practical” significance? Can they imagine “shrinking a data set” to explore how few data points would be needed to still see this (to get at the Law of Large Numbers in reverse)?
• How do students’ quasi-proportional understandings of sample play out when data are very complex?
• How much rigor vs. intuitive/informal understanding of statistical ideas should we aim for?
• How do we address issues of data cleaning, formatting, and structure? How can large data sets be usable in classrooms?
• Which classroom contexts (science, social studies, math classes) offer the most potential for working with large data sets? What curricular, technical, policy, knowledge, and other obstacles will teachers face?
• Which representational tools support student exploration of large complex data sets, including relationships in time and space?

Compelling Data

Website Usage Data
This graph shows the effects of statistical control in examining use of an educational website. Parallel lines depict relationship of “Video Content” to “Logins” at 3 levels of the “Friends” variable (colored). Slope is shallower than when “Friends” variable is excluded.

What representations will help students see and understand complex relationships among multiple attributes in data?

Sea Surface Temperatures in the Equatorial Pacific
from http://www.data.gov/erclassroom/content/el-nino/

El Niño Event — Dec 16, 1997
Normal Temperatures — Dec 16, 2005

Color Coded Map Data: Overall view but imprecise

Equatorial Temperatures in Quartiles with Group Means: Quartile variability seems smaller in El Niño group. All mean differences are statistically significant, maybe not practically significant.

What tools will allow students to explore data with a geographic component? How will they think about variability, or about statistical significance vs. practical significance?

Compelling Data

Products

• Report on the current state of educational uses of large scientific and social scientific data sets in secondary schools, and the statistical, cognitive and learning issues for students exploring them.
• Design recommendations for software tools, curriculum materials, and professional development for teachers to support secondary students’ learning about statistics and science/social science using large data sets.

Papers/Presentations