Bringing Open Data in the Classroom to Foster Student Engagement

Erika Martin PhD MPH$^{1,2,3}$ and Natalie Helbig PhD$^4$

$^1$ Nelson A. Rockefeller Institute of Government-State University of New York; $^2$ Department of Public Administration and Policy, Rockefeller College of Public Affairs and Policy, University at Albany-State University of New York; $^3$ Department of Health Policy, Management, and Behavior, School of Public Health, University at Albany; $^4$ Center for Technology in Government, University at Albany

Although new curricula are sprouting around the use of Big Data, students do not have to become data scientists in order to understand the fundamental principles of using quantitative data to understand and solve problems. The ease of browsing, visualizing, and downloading data from new Open Data portals make it possible to bring data into the university classroom at all levels, from freshman undergraduates to doctoral students. Instructors can take advantage of Open Data to develop in-class activities, assignments, and term projects that challenge students to ask critical questions about the world in which they live and develop data skills to address important social problems. We provide some basic background on student engagement, explain how student engagement can occur around Open Data, and list tangible ideas for incorporating Open Data into the university classroom. Our goal is not to provide an exhaustive overview of specific open data classroom activities or theories on student engagement. Rather, we aim to provide preliminary ideas for a continuing dialogue about how to take advantage of these resources in higher education to make students more critical consumers and producers of data.

Student Engagement: Why It Matters and Guiding Principles$^2$

Effective student engagement is the process that coaxes — and even requires — students to act in ways that make their thinking visible, thereby turning them into critical thinkers. When students act on their thinking and generate consequences, they are no longer passive observers of the discipline but instead become independent agents of disciplinary thinking. Effective student engagement can allow students to increase their traction in the discipline, and learn to be active users and evaluators of disciplinary knowledge. By comparison, traditionally formatted courses are designed primarily around providing opportunities for students to learn (via lectures, office hours, a bibliography, etc.) without the framework that impels or compels students to invest effort and take more responsibility for their learning. The traditional, opportunity-based approach will work for some students, but not for all. Those with less rich academic preparation or academic acculturation may need more than open opportunities to be successful. They will need a structure, strategy, and process that implicate them more directly in the disciplinary content.

---


$^2$ The material in this section is adapted with permission from presentations and unpublished writings by Bill Roberson and Billie Franchini of the University at Albany Institute for Teaching, Learning, and Academic Leadership.
There are a few guiding principles behind fostering student engagement.

**First, students do not need to know everything before they can do anything.** Although “research” is often reserved for advanced electives or graduate courses, students of any skill level are capable of using basic research principles to learn about their discipline and challenge assumptions.

**Second, students’ naïveté about a discipline is not an obstacle for instructors.** Rather, it is an opportunity to encourage them to adopt an inquisitive attitude about their discipline and environment. Structuring learning tasks that force them to question their assumptions and determine how to assemble evidence to understand a problem can promote critical thinking and excitement about the discipline. For example, a traditional learning approach might encourage students to make judgments and decisions after receiving extensive background information; assignments are designed to be applications, and speculation is discouraged. A critical thinking classroom approach could force students to make decisions, judgments, and hypotheses prematurely, creating frustration and a perceived need for information. This allows students to start to inquire about their discipline at a higher level.

**Third, there are several ways to fuel student curiosity.** These include:

- The **expectation to discover something** (Are there geographical patterns in childhood obesity rates? Who suffers most from type 2 diabetes? How do patterns of recreational drug use differ across regions?);

- The **desire to resolve a weighing or intriguing doubt** (To what extent can racial disparities in childhood obesity rates be explained by features of the built environment? If type 2 diabetes is so closely related to socioeconomic status, what does that say about the role of personal choice in health behaviors? Are suburban communities really safer than big cities?); and

- The **anticipation of solving a problem** (Would building additional safe play spaces reduce childhood obesity rates? How can health departments promote healthier lifestyle choices? In what ways do we need a more nuanced approach to address the causes and consequences of substance use disorders?).

Instructors can use quantitative Open Data to structure exercises that encourage students to ask these broader questions.

**Open Data: A Springboard for Student Engagement**

The availability of free, easily accessible, and downloadable data makes it increasingly easier to design activities that force students to use quantitative data to develop preliminary hypotheses, test them empirically, critically reflect on their findings, consider how further evidence could be used to support or refute their hypothesis, and identify potential solutions to address social
problems. These data can allow instructors to set the stage for students to implement the scientific method in order to gain knowledge about the discipline, become more invested in their own learning, and move towards becoming critical thinkers.

Open Data are in a variety of formats and exist in many areas besides public health, making it relevant for almost any discipline. As we are in the early phases of Open Data, these portals will continue to expand their catalogues of available data, providing opportunities to refresh examples, create new examples, or use students to explore and find new examples. The open access policy eliminates the need to go through paperwork-intensive processes to request the datasets from agencies and adhere to strict data use agreements.

Open Data-based assignments can teach critical thinking skills about data use. Basic data management and statistical skills are similar to those needed to analyze traditional datasets. However, Open Data requires additional critical thinking skills about how to use data appropriately. Because Open Data are a work in progress, the data are rarely in a desirable format for direct statistical analysis. Many of these datasets are raw, and not cleaned to create tidy variables or data layouts, and they may have a very coarse-grained granularity (e.g., aggregated to the school district or county level) to protect confidentiality. To work with these data, students need to be aware of data quality concerns, including understanding what data quality means and being able to discriminate between good and poor quality data. This will force students to identify ways to resolve these issues technically (e.g., using data management skills to reformat the data, check for errors, and resolve discrepancies) and analytically (e.g., including other variables in the analysis to address potential omitted variable bias and running additional specification checks). Students may also encounter different visualizations created by the Open Data platforms and infographics used in newspapers and magazines. To become sophisticated data users and consumers, students will need to think critically about how different visualizations may inappropriately bias readers, and how to display data appropriately.

Finally, Open Data is becoming a part of many disciplines. Workforce expectations are changing, and students will increasingly be expected to know about these resources and how to use them. Journalists, policymakers, and advocacy groups are already using these data. As part of their discipline, students should be able to think critically about how and when these data can be used, and to question assumptions about the results of an Open Data-based analysis.

**Practical Tips: Ideas for Fostering Student Engagement Using New York State Open Data**

In 2013, Governor Andrew Cuomo signed Executive Order 95,³ which asks all New York State executive branch agencies to make government data open. In addition, local governments across the state are encouraged to submit data. New York has multiple Open Data sources, including the main portal from the governor’s office (Open NY, https://data.ny.gov/), a dedicated health data portal at the Department of Health (Health Data NY, https://health.data.ny.gov/), and local government data for New York City (NYC Open Data, https://nycopendata.socrata.com/). Below are some specific ideas for ways to incorporate these data into courses in order to encourage students to become more engaged in the production of knowledge:

³ [http://www.governor.ny.gov/executiveorder/95](http://www.governor.ny.gov/executiveorder/95)
- **Web-based communications**: Ask students to use the Visualize tool to create a map displaying geographical differences in the variable. Students will write a news story that uses the data to explain what is happening in both the state and local community, and use the Embed tool to add a live link to the map. Have a class discussion on possible sources of bias in the data, and how to describe data to the public.

- **Data management**: Create an assignment where students use the Export tool to download multiple datasets as .csv files, and import into their statistical software package (e.g., SAS, Stata, SPSS). Students then merge files using a common identifier, clean the data, and make new variables. Ask the class to discuss data quality concerns and how to resolve them, and how analytic decisions on operationalizing variables may lead to bias.

- **Introductory statistics**: Task students with using the Export tool to download the data as an .xls file. Students subsequently use Excel’s basic summary statistics functions to describe univariate characteristics of variables, such as means, medians, percentages, and ranges. Discuss the advantages and disadvantages of using various summary statistics to describe a population, and possible sources of bias in the data.

- **Intermediate statistics**: Create an assignment where students import one or more datasets into a statistical software package, run multivariate analyses to evaluate relationships among variables and confounding factors, and compare the regression-based results to the maps showing geographical differences in the variable. Discuss whether the statistical analysis can fully explain these geographical differences.

- **Geography and urban planning**: Ask students to use the Export tool to download multiple datasets as .csv files, and import them into ArcGIS, apply their data management skills to clean the data and create new variables, and overlap multiple variables to display visual relationships. Discuss whether the location of hospitals, clinics, or other services matches the distribution of the target population, and how to convey this information to policymakers.

- **Software development**: Use the data for a term project in computer science. Task students with either selecting their own application to see how creative they can get with the data, or else specify an industry-relevant priority task with a final contest of what works best for the industry’s specific purposes. The project could be organized as a Code-a-thon, with a prize to the team that develops the best application. Invite practitioners to be guest judges. Here the data serve as a test bed for the application of new concepts.

- **Data systems design**: Design a project-based lab class in which students design systems that utilize the data. Turning this into a contest with external judges could boost the stakes and improve students’ engagement. Class discussions could push students to identify different types of data (e.g., administrative, survey, electronic health records) and how to adapt data systems to meet the data’s unique features and possibly diverging needs of end users.
• **Political research:** Assign students to use the Open Legislation of the NY Open Senate application (http://www.nysenate.gov/legislation) to analyze voting patterns or bill content. Ask students to compare their analysis to the rankings of legislatures developed by public interest groups and to hypothesize reasons for discrepancies. Some of this information would have previously been available through subscription-only databases such as Westlaw.

• **Budgeting:** Have students use Open Budget to locate appropriations, budgeted and actual spending, and historical information. Task students with using those data to answer various questions: What is the total size of the budget? How large are the budget deficits and surpluses? How do the budget size and deficits change over time? What are the major sources of revenues and expenditures, as a proportion of the whole budget? Use the results to start a broader discussion about the factors driving these changes, and potential consequences for population health and other social outcomes.

• **Public administration:** Ask students to compare data from Open Budget (budget portal from the Governor’s office) and Open Book New York (budget portal from the Comptroller’s office, see http://www.openbooknewyork.com/). Describe similarities and differences in the ways two agencies report information, and reporting incentives. Discuss agencies’ motivations for reporting data, ways in which data collection is influenced by political factors such as formula-based funding, and how to overcome potential data biases in research.

• **Criminal justice:** Design an activity where students explore relationships among ZIP code-level crime rates, race/ethnicity, and income. Use these empirical findings to start a discussion about the causes of disparities and how to reduce them. Discuss the limitations of existing data and how to improve the measurement of these factors.

### Citing Open Data

Although Open Data are available free of charge to the public, it is helpful for government agencies to know where, when, and how the data are used. Citing data is also good practice for replication, and gives credit to data repositories. Instructors and students should cite all data sources appropriately in teaching and research materials.

### Other Teaching Inspirations

The School of Data (http://schoolofdata.org/) provides free courses on how to use and be inspired by data. This nonprofit organization aims “to teach people how to gain powerful insights and create compelling stories using data.” The website has some interesting uses of data and ideas for bringing data into the classroom.

The Open Knowledge Foundation provides a list of Open Data catalogues from government agencies around the world (http://opengovernmentdata.org/data/catalogues).
Browse past and current Code-a-thons (http://www.health2con.com/devchallenge/about/) for real-world challenge programs for developers to create innovative technologies using data. A class project could be organized around a similar model. Depending on the timing of the next challenge, the class could work together on a collaborative entry.

**Other Ideas?**

If you have other ideas on using Open Data in the classroom, including success stories, please share them with us at emartin@albany.edu and nhelbig@ctg.albany.edu.