Air Quality Data Mining: Mining the US EPA AirData website for student-led evaluation of air quality issues

Mary A. Williams[1,2], Katherine Barry[1] and Deena Wassenberg[1]*

[1]University of Minnesota, College of Biological Sciences, Department of Biology Teaching and Learning, Minneapolis, MN 55108

[2]Current affiliation: Minnesota Pollution Control Agency, Environmental Analysis and Outcomes Division, St. Paul, MN 55155

Abstract

Air pollution directly affects human health endpoints including growth, respiratory processes, cardiovascular health, fertility, pregnancy outcomes, and cancer. Therefore, the distribution of air pollution is a topic that is relevant to all, and of direct interest to many students. Air quality varies across space and time, often disproportionally affecting minority communities and impoverished neighborhoods. Air pollution is usually higher in locations where pollution sources are concentrated, such as industrial production facilities, highways, and coal-fired power plants. The United States Environmental Protection Agency manages a national air quality-monitoring program to measure and report air-pollutant levels across the United States. These data cover multiple decades and are publicly available via a website interface. For this lesson, students learn how to mine data from this website. They work in pairs to develop their own questions about air quality or air pollution that span spatial and/or temporal scales, and then gather the data needed to answer their question. The students analyze their data and write a scientific paper describing their work. This laboratory experience requires the students to generate their own questions, gather and interpret data, and draw conclusions, allowing for creativity and instilling ownership and motivation for deeper learning gains.

Learning Goal(s)

- Students will understand the impacts of air quality as it relates to human health, public health, and environmental justice.
- Students will demonstrate an understanding of how publicly available air quality data can be used to examine questions about environmental health, public health and environmental justice.

Learning Objective(s)

At the end of the activity, students will be able to:
- Describe various parameters of air quality that can negatively impact human health, list priority air pollutants, and interpret the EPA Air Quality Index as it relates to human health.
- Identify an air quality problem that varies on spatial and/or temporal scales that can be addressed using publicly available U.S. EPA air data.
- Collect appropriate U.S. EPA Airdata information needed to answer that/those questions, using the U.S. EPA Airdata website data mining tools.
- Analyze the data as needed to address or answer their question(s).
- Interpret data and draw conclusions regarding air quality levels and/or impacts on human and public health.
- Communicate results in the form of a scientific paper.

INTRODUCTION

Some students are aware that race and socioeconomic factors influence individual exposures to environmental pollution. However, many are not familiar with specific examples of environmental justice issues and the extent to which exposure to hazardous air quality can be determined by racial and socioeconomic factors (1,2). Our goal in developing this lesson was to inspire students to explore spatial and temporal trends of air pollution in the context of a local example of disparities in air quality and resulting legislation.

In this lesson, run over two laboratory periods, students learn about issues of environmental justice and air pollution and then use the EPA AirData database to pose their own questions regarding special or temporal trends in air quality. This module was developed for non biology-major undergraduates in the laboratory portion of Biology 1055 (Environmental Biology: Science and Solutions). The laboratory course is 1 hour and 55 minutes once a week. Each section has 24 students who generally work in either pairs or groups of four. This laboratory activity has been run for four semesters in 10 laboratory sections. While the EPA AirData website has previously been suggested as a resource for educators wanting to provide students with an opportunity to analyze real scientific data (1,2,3,4), to the best of our knowledge, this is the first structured activity using this resource to be published.

The importance of providing education about environmental justice issues has been highlighted in several publications (1,2,3,4,5,6). The use of local examples and student empowerment has been suggested as important to creating engagement within a student population (1,2,3,4,5,6,7,8).

SCIENTIFIC TEACHING THEMES

Active learning

Students begin this lesson by discussing an article (9) about the relationship between environmental justice and air pollution. After a short tutorial on how to use the EPA website, students work in pairs to ask their own questions and mine the EPA database to address their question.

Assessment

Formative assessment occurs through performance on pre-lab questions and the in-class discussion. As the summative assessment, students individually write a scientific paper about their work. This assessment currently counts for 80 percent of the topic's laboratory grade. The student handout contains the grading rubric for the paper.

Inclusive teaching

The reading for this lesson highlight the impact that race and socioeconomic factors play in people's exposures to air pollutants (9). To include students with different learning preferences, the activities include different ways of engaging with the material, including reading, interactive discussion, and independent, hands-on computer work. By working in pairs, students can work with students that have distinct strengths than their own.

LESSON PLAN

Table 1 (page 3) shows a typical timeline for this activity.

Before Class

1. Learn to use the EPA Data Website: Begin preparation for the lab by becoming familiar with the EPA data website (10) and the datasets available there. Ensure that you understand how to navigate the interactive map on the website. For example, at the time of this writing, the interactive map requires the Google Earth plugin. We were successful in getting this plugin to work when we used the Firefox browser. We contacted the EPA site administrator with questions about problems with the plugin and he informed us that, by summer of 2015, the website should no longer require the plugin.

2. Ensure students can access the internet during lab. Completion of the lesson requires that each pair of students has access to a computer connected to the internet.

3. Update the lesson to reflect local geography. This lesson features Twin Cities Metro data, which is relevant for University of Minnesota students. This local report and data appeared to be compelling for many students, especially those who were familiar with these issues or had lived in or near the neighborhoods in question. Of course, a local example will likely be more relevant to students living elsewhere. Thus, we advise instructors to use examples from the EPA website that are relevant to your institutional location and/or to reflect student demographics. The student handout and class slides should then be modified to reflect the instructor's choice of the geographical locations of from which the students will obtain their air pollution data.

4. Assign students reading to complete before the lab. A week before the lab, assign the entire Environmental Health news article (9) to the entire class. If doing the optional jigsaw activity, every student should additionally read the abstract and the "Brief History" sections (pages 8-9) and the "Moving Forward" section (page 21-22) of the Stanek paper (11). Then each group will read a different, assigned portion of the Stanek paper. If students generally have assigned groups of four, we suggest assigning each person in their 'home' group to a different portion of the paper. During class, students who have been assigned the same portion of the paper will meet in their "Jigsaw Group."

5. Jigsaw Group 1 reads the section titled "Early Toxicology and its Impact on Public Sentiment Toward Air Pollution" (pages 9-10). Group 2 reads the sections on particulate matter (pages 11-13). Group 3 read the sections titled "Toxicology Establishes Biological Plausibility for Epidemiologic Findings," "Cardiovascular Effects," and "Potential Systemic Effects Revealed by Toxicological Research" (pages 13-15), and Group 4 reads the section on ozone (pages 15-17). Please see Table 2 (page 4) for a layout of the Jigsaw assignments.

Class Period -Week 1

Review Background Material (20 minutes)

The student handout (S1) presents an introduction to the material that the students should have read prior to the start of class (1,2,3,4,5,6,7,8,9,11). The TA/Instructor can start the lab by presenting a short review lecture to refresh the students on this background material describing air pollution and its effects on human health. Potential discussion slides are included in the Air Quality Slide Deck (S5), which includes a recent report showing dramatic differences in life expectancy by neighborhood within the Twin Cities metro

Table 1: Air Quality-Teaching Timeline

Activity	Description	Time
Week 0 (week before this exercise)		
Assign to students	Assign students to read the laboratory handout for the next week's work (provided in Supplementary material; "Laboratory Handout"), and the assigned readings. Optional: Assign students different sections of the Stanek et al. article.	
Assign to students	"Check yourself" questions in the laboratory handout –to be ready to be turned in at the start of lab Week 1	
Instructor preparation	Prepare a set of lecture slides to facilitate in-class discussion (example lecture slides provided in supplementary material; "Air Quality Power Point")	
Instructor preparation	Become familiar with the EPA data website3 and the datasets available there. Check for technical issues involved with accessing interactive map.	
Week 1		
Instructor-led group discussion	Instructor leads the class with a discussion of air pollution and environmental justice issues. Instructor also orients students to their task for this laboratory activity.	20 min
Jigsaw activity (Stanek et al. paper)	Students participate in a jigsaw activity based on their assigned portion of the Stanek et al. paper.	30 min (optional)
Getting to know AirData	Students proceed through a guided inquiry that helps them get familiar with the AirData website.	15 min
Brainstorming	Students brainstorm questions they can address using the AirData website. Students are instructed to come to class with a question picked out.	10 min
Week 2		
Instructor discussion	If desired, instructor can lead a class discussion in which student share the questions they have chosen.	15 min
Independent work	Students work independently gathering data, analyzing data and creating figures. Instructor circulates helping students with their questions and interpretations.	55 min
Instructor announcements	they can continue to use these tools outside of class if they need more time. They are reminded of the laboratory report due the following week and that there is a template for this report in their laboratory handout.	5 min
Week 3		
Assignment Due	Student laboratory reports are due. Instructor may lead an optional discussion inviting students to share what question they asked and what they concluded.	15 min

Table 2: Air Quality-Jigsaw Assignments

Assignments	Jigsaw Group 1	Jigsaw Group 2	Jigsaw Group 3	Jigsaw Group 4
Environmental Health News	Yes	Yes	Yes	Yes
Stanek et al., Abstract and Brief History	Yes	Yes	Yes	Yes
Stanek et al., Moving Foward	Yes	Yes	Yes	Yes
Stanek, Early Toxicology (pp 9-10)	Yes			
Stanek, Particulate Matter (pp 11-13)		Yes		
Stanek, Biological Plausibility, Cardiovascular System Effects (pp 13-15)			Yes	
Stanek, Ozone (pp 15-17)				Yes

Williams, M.A., Barry, K. and Wassenberg, D. Air Quality Data Mining: Mining the U.S. EPA AirData website for student-led evaluation of air quality issues. *CourceSource*.

area (1,2,3,4,5,6,7,8,9,11,12,13).( )Other national reports of similar differences are available (14-16).

As the discussion progresses about air pollution and the air quality index, the instructor can use a think-pair-share strategy to encourage students to discuss air pollution or air-quality-index issues. The class then regroups as a whole to share proposed answers.

A few example questions are:
- "For the month of July 2012, Los Angeles, California reported an average daily Air Quality Index (AQI) of 158. For this same month, Chicago, Illinois reported an average daily AQI of 142. Which city appears to have the greater amount of air pollution? What do you think might be the cause of this difference?"
- "How is air quality linked to health? What effect do you think air quality has on quality of life in affected areas?"
- "Do you think the legislation regarding the Phillips Neighborhood in Minneapolis is justified? Should current air quality be taken into account when choosing where to locate new sources of pollution?"
- "How should researchers investigate discrepancies in air quality across racial or socioeconomic lines?"

Students should have enough background from the reading assignment and review lecture to discuss and answer these types of questions.

Jigsaw Activity (Optional, 30 minutes)

For 10 minutes, students in each Jigsaw Group will meet to discuss their common assigned reading and compile a document containing five to ten take-home messages from their assigned reading (their assigned portion of the Stanek paper). Students will then return to their lab or 'home' group, which will include students who have read each of the different portions of the paper. For 10 minutes, each student will share what they have learned about their section of the paper and come up with their reflections about what is known about the biological effects of air pollution. In the final 10 minutes, the instructor will lead a general class discussion about the biological effects of air pollution, including information about known health disparities that exist in communities that have higher exposure to air pollution.

Complete the Tutorial (20-40 minutes)

At this point, the students working in pairs will go to the EPA AirData website and start the tutorial exercise outlined in their handout. This tutorial will help them to learn the basic data mining procedures on the EPA website. The interactive map and some of the visualization tools on the website provide the students with a tool that can spark their creative ideas about how they want to query the data.

While the students are working, there are reading questions in the handout that the TA/Instructor can check and note their answers in her gradebook. As the student groups work through the tutorial, the TA/Instructor can walk around the room and answer questions. The TA/Instructor can join in on conversations about places or times that groups are discussing and can help the groups focus their ideas on questions that they can ask and comparisons that they can make of the data. In our experience, some tech savvy student groups will complete the tutorial within 10 to 15 minutes and will start developing their ideas and questions. Other groups will need more time.

Question Brainstorming (15 minutes)

After the tutorial, students brainstorm questions about air quality/pollution that they can answer using the data. The laboratory handout contains links to other databases that might inform a student's question (for example U.S. Census data, 17). The students may notice that there are some data visualization tools on the AirData website. Our experience is that these tools in present form are not adequate to address the questions students have posed, but some students may find that exploring these tools may spark some ideas for questions that they want to explore further. Examples of student inquiries from our experience include: examining sulfur dioxide concentrations near a refinery close to a student's hometown before and after public attention to pollution surrounding that refinery; comparing air quality in cities with high overall health measures to the air quality in cities with lower health measures; and comparing particulate matter in areas with known coal mining activities to areas without mining activity.

By the end of the first lab, all student groups should have discussed their ideas and questions with the lab instructor. They can interact with the lab instructor over the subsequent week to finalize their question.

Class Period - Week 2

During the week two lab, students will have time to pull data from the AirData website that addresses their question and begin generating graphs and working on their analysis. Supporting File S3 (Graphing in Excel) includes a short overview of how to graph in Excel, should the students need help in that area. Alternatively, if two full lab periods are not available, portions of the data mining and data analysis can be done outside of class.

TEACHING DISCUSSION

The goals of this module were to introduce students to the concepts of 1) spatial and temporal trends in air quality and 2) data mining as it relates to answering questions about air quality. In our class, other skills such as general scientific method, generation of hypotheses and testable questions, which are needed to perform this activity, are introduced earlier in the semester. For courses in which the timing does not work out that way, additional time should be spent on these aspects of the activity. While we developed this activity to fit into two 115-minute laboratory sessions, this activity could be done in a traditional college or high school classroom setting if small groups of students could share Internet access via laptop or other device. This module can be condensed to fit into a shorter time period or extended to achieve additional learning outcomes.

Based on student feedback and student assessment, we feel that this lab module met our goals in an engaging, active-learning environment. Furthermore, we felt that requiring the students to develop their own questions about the data deepened their learning and sense of ownership in the activity. Several students listed this lab as one of their favorites of the class, noting the association of the lab with a real-world societal problem.

In general, students seemed to understand the connections between air quality and quality of life in communities. Also, through the local example of the air quality issues in

the Minneapolis Phillips Neighborhood, they were able to connect these concepts with the concept of environmental justice. Several students even volunteered their experiences and connections with environmental justice issues.

Some students noted that the lab helped them develop data analysis skills that would be valuable in other contexts; others were challenged by the data-mining component. We feel that this experience helped the students to realize that, although the process of going to a website to gather data might seem like an easy task, data mining activities can be complex and require specific skills.

SUPPORTING MATERIALS
- S1. Air Quality-Student Handout
- S2. Air Quality-Answers to the Check Yourself Questions in Student Handout
- S3. Air Quality-Graphing in Excel
- S4. Air Quality-Writing the Materials and Methods sections for the final paper
- S5. Air Quality-Air Quality Slide Deck

ACKNOWLEDGMENTS

REFERENCES

1. Clark LP, Millet DB, Marshall JD. 2014. National Patterns in Environmental Injustice and Inequality: Outdoor NO2 Air Pollution in the United States. PLoS ONE 9(4): e94431. doi: 10.1371/journal.pone.0094431

2. Hajat A, Diez-Roux AV, Adar SD, Auchincloss AH, Lovasi GS, O'Neill MS, Sheppard L, Kaufman JD. 2013. Air pollution and individual and neighborhood socioeconomic status: evidence from the Multi-Ethnic Study of Atherosclerosis (MESA). Environ Health Perspect 121:1325-1333; http://dx.doi.org/10.1289/ehp.1206337

3. Trundle, K. 2007. Acquiring online data for scientific analysis. Technology in the Secondary Science Classroom. Ch 6. Ppg 53-61.

4. Yang, Xiusheng. "GIM3-A visual and interactive contaminant transport simulator for regulatory and educational applications." Environ Manage 83: 44-55.

5. Andrzejewski, J., Baltodano, M. P., & Symcox, L. 2009. Social justice, peace, and environmental education: Global and Indivisible. Social justice, peace, and environmental education: Transformative standards, 1-16. New York,, NY: Routledge.

6. Corcoran, P. B., & Wals, A. E. 2004. Higher education and the challenge of sustainability. Dordrecht: Kluwer Academic Publishers.

7. Theobald, E. J., Crowe, A., HilleRisLambers, J., Wenderoth, M. P., and Freeman, S. 2015. Women learn more from local than global examples of the biological impacts of climate change. Frontiers in Ecology and the Environment 13(3): 132-137.

8. Dimick, A. S. 2012. Student Empowerment in an Environmental Science Classroom: toward a framework for social justice science education. Science Education 96(6):99--1012.

9. Gammon, C. 2012. Pollution, Poverty and People of Color: Asthma and the Inner City. Environmental Health News. www.environmentalhealthnews.org/ehs/news/2012/pollution-poverty-people-o.... Accessed January 13, 2014.

10. United States Environmental Protection Agency. 2013. AirData Website. www.epa.gov/airquality/airdata/. Accessed December 14, 2013.

11. Stanek, L., Brown, J., Stanek, J., Gift, J., and Cotsta, D. 2011. Air Pollution Toxicology - A Brief Review of the Role of the Science in Shaping the Current Understanding of Airpollution Health Risks. Toxicol Sci. 120 (S1) S8-S27

12. Wilder Research. 2012. Health inequities in the Twin Cities. www.wilder.org/Wilder-Research/Publications/Studies/Forms/Study/docsetho.... Accessed December 14, 2013.

13. MPR news. 2014. In the Twin Cities, where you call home can say a lot about how long you may live. www.mprnews.org/story/2014/02/03/news/twin-cities-life-expectancy-map. Accessed January 13, 2014.

14. The Atlantic Cities. 2011. Place, not race may better explain America's health disparities. www.theatlanticcities.com/neighborhoods/2011/10/health-disparities/268/. Accessed January 13, 2014.

15. Young et al. 2012. Differential Exposure to Hazardous Air Pollution in the United States: A Multilevel Analysis of Urbanization and Neighborhood Socioeconomic Deprivation. Int J Environ Res Public Health 9(6):2204-2225.

16. Center for Disease Control. 2011. Fact sheet: health disparities in unhealthy air quality. www.cdc.gov/minorityhealth/CHDIR/2011/FactSheets/AirQuality.pdf. Accessed December 14, 2013.

17. United States Census Bureau. 2014. Census data mapper. www.census.gov/geo/maps-data/maps/datamapper.html. Accessed January 13, 2014.