2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
ASEE
Paper ID #37606

Board 29: Compiling Census Data and Atmospheric Repository Data to Infer Socio-Environmental Trends

Dr. Joe Woo, Lafayette College

Work-in-Progress: Compiling Census Data and Atmospheric Repository Data to Infer Socio-Environmental Trends

Joseph L. Woo
Lafayette College

Abstract:

In engineering coursework, it can be difficult to find real-world datasets that convey meaningful, correlative relationships between measurable phenomena and relevant social issues. With the recently-completed 2020 US Census, a set of up-to-date, publicly-available and geospatially-distributed population demographic information can be compared against atmospheric pollutant datasets.

Students selected census data for a minimum of five zip code tabulation areas (ZCTAs) near their homes. Students extracted relevant census data and compiled their findings against one year of historical NO, NO_2, and ozone concentration measurements from EPA Air Quality monitors in the same ZCTA. As they find trends in their results, students develop a deeper understanding of the physical drivers behind air quality and the computational skills necessary to align, clean, and process their data. The open-ended nature of this project, combined with the direct connection between the students' home neighborhoods and the data being collected, fosters student investment and curiosity in their analysis.

Keywords

Air quality, modeling

Motivation

In chemical engineering curricula, it can often be difficult to identify relevant and meaningful examples that relate beyond the realm of traditional process engineering. Environmentally-oriented coursework often employs examples within the natural world's subsystems (i.e., the atmosphere, hydrosphere, lithosphere, and biosphere) to demonstrate principles of mass transport, material and energy balances, and chemical kinetic phenomena. Notably, climate and atmospheric systems have provided a consistently topical and well-documented source of information from which inspiration for case studies in undergraduate courses can be developed.

As the impacts of climate change have continued to evolve and manifest over the past few decades, there is also a growing need to develop more nuanced and expansive discourse around environmental topics. [1] Due to their complexity, the social, ethical, and justice elements of environmental issues often take a secondary role to more economic or policy-based motivations (loss of product, emission/release standards, etc.) in these discussions, which may result in the unintentional erasure or lack of apparent attention to the socially disadvantaged groups whom are disproportionately affected. [2]–[4] As such, when creating new materials for environmentally-focused chemical engineering coursework, it is desirable to keep these factors in mind from the

conception stages of case studies so that they can appropriately capture these topics without appearing superfluous or unrelated in scope.

The interplay between social and engineering issues lends itself well to project-based learning approaches of assessment, which enable a deeper and more longform analysis of an individual topic compared to exams. While many classes that discuss atmospheric systems will focus on broader regional- or global-scale climate as motivation for research questions in such projects, conversations around public health in outdoor air systems provide a smaller-scale but equally important context from which atmospheric phenomena can be observed. As more granular and publicly available atmospheric datasets become more available, other questions exploring the differences in ambient air quality between differently populated areas can be more easily brought to the classroom.

The United States Census provides a uniquely comprehensive and geographically-distributed snapshot of the demographics across the nation. Information is broken down by ZIP code tabulation area (ZCTAs), which roughly match with United States Postal Service ZIP codes but with some adjustment for infrequently-used or sparsely populated regions.[5] While some geographic holes do exist in this dataset, the Census can be used to identify an extensive array of information about the population in a given populated area, ranging from distributions of race, age, salary, and education to information about resident occupancy, commute time, and sector of business.

Clark et al. recently demonstrated the ability to assess statistically significant effects between pollutant concentrations and different subsections of census information, demonstrating differences in nitrogen dioxide (NO_2) concentrations for non-white, below-poverty, and/or low-education areas. [6] While this study took a comprehensive approach across the entire 2000 Census dataset and a satellite-based land-use regression model for NO_2 concentrations [7], the general premise of identifying trends between census data and pollutant concentrations is generalizable enough to apply to any set of publicly available atmospheric data. Given the nontrivial data manipulation to get such datasets into compatible formats, and the wide range of design space for interesting-yet-straightforward research questions, the conception of hypotheses about the connection between air pollution and population groups is a feasible and culturally relevant project for undergraduate students exploring atmospheric data.

Approach / Methods

Course

The project presented in this work took place in a four-credit-hour, junior-level chemical engineering elective in the Spring 2022 semester, *Atmospheric Engineering and Science*. While familiarity with material and energy balances, transport phenomena, and chemical kinetics are useful for deeper discussion regarding the underpinnings of atmospheric phenomena, relevant equations and concepts are reintroduced or reframed from previous chemical engineering courses for the context of the material at hand. Further, this course serves as a cross-disciplinary elective across multiple programs, with previous enrollment having included students from civil/environmental engineering, mechanical engineering, and integrative engineering.

The Spring 2022 instance of *Atmospheric Engineering and* Science comprised of 3 seniors, 8 juniors, 8 sophomores, and one first-year student. With the exception of two junior students, whom were environmentally-focused integrative engineering students, students in this instance of the course majored in or planned on majoring in chemical engineering. Though first-year students are broadly discouraged from taking this course, the first-year student enrolled in this instance of the course was also taking part in undergraduate research with a professor whom focused on atmospheric science and therefore had sufficient background to maintain parity with their classmates.

The intrinsic intersectionality of this work calls upon a variety of different skill sets across multiple programs, necessitating a survey-style approach to the material being covered. In light of this diversity of background, more recent iterations have implemented a series of four data-driven projects that build upon the theory delivered in lecture and enable students to better leverage their varied experiences and skillsets into the content of this course. Students are encouraged to use any form of quantitative software that they are familiar with to perform the analyses necessary for their projects. While underclassmen primarily utilize Microsoft Excel, upperclassmen have used a combination of MATLAB, R, Python, and/or Minitab, depending on the language(s) employed in courses they have previously taken. For students lacking any background working with statistics, basic tutorials in statistics and MATLAB were provided offline during office hours. These tutorials consisted of small-group meetings with the instructor in which the instructor introduced fundamental statistical tests (e.g., ANOVA, t-tests, etc.) and/or coding principles (e.g., data visualization functions, scripts, etc.) While the tutorials did not exceed more than 3-4 hours total in the Spring 2022 instance of this course, larger sections or populations of students that are less trained in these concepts would likely instead be referred to self-guided tutorials in the form of pre-recorded primer videos that cover similar concepts.

The projects of *Atmospheric Engineering and Science* aim to improve analytical skills based on larger, noisier data sets than students may be previously familiar with. While in laboratory classes, students collect data and perform regressions to infer trends, the fundamentally simple nature of the experiments and low number of collected data points necessarily result in relatively easy analysis and data that did not necessitate any pre-processing before his analysis. The most recent iteration of this course, which took place in the Spring 2022, replaced one project, which previously focused on estimated surface temperatures based on location-based solar irradiance, with one that focuses on a simplified version of the regressions performed in Clark et al. that connected pollutant concentration to census subset data. A description and summary of the project follows.

Project Description

The project described in this work takes place roughly halfway through the semester, as the second major deliverable of the course. The first project in the course, which required a deep dive into the local air quality of each student's hometown, was used as a means of familiarizing students with baseline values of pollutant concentration and meteorological trends. After extracting historical air quality index (AQI) and weather data from a variety of online resources (the EPA Air Quality Collection [8], NOAA Weekly Weather Maps [9], and the PurpleAir AQI

Database [10]), students visualized trends in AQI, temperature, and humidity based off of time of day and month of the year. Students were able to identify both large-scale trends correlating to local climate, as well as specific idiosyncrasies that could explain outliers or shifts from expected trends, such as local events or businesses that existed near monitoring stations. Furthermore, the different resources' varied formatting and time resolution in their dataset files required that students become familiar with the aggregation and/or retiming of temporal data into a uniform hourly format. As a result, students developed familiarity and confidence, both in understanding the behaviors that their hometown air quality should exhibit and the necessary mechanical skills to work with time-resolved atmospheric data.

The second project builds upon the first and focuses on connecting information from different databases to infer social trends on overall air quality. The findings of Clark et al., focused on the atmospheric concentrations of nitrogen dioxide (NO_2), one of the six EPA criteria air pollutants and a known marker of high amounts of vehicular traffic and power plant emissions. [6], [11] However, the process of drawing trends between pollutant concentration can be extended to any of the pollutant datasets available. First, students identify five EPA monitoring stations on the EPA Air Quality Collection that are located near their hometowns, such that yearly data for ozone, NO_2, and nitric oxide (NO) can be downloaded.

The first part of this assignment is a more traditional (i.e., non-socially related) calculation-based exercise, in which students identify the latitudinal/longitudinal locations of their five sites, then calculated values of photochemical rate constants as dictated by their location using the NCAR Tropospheric UV Model. [12] Based on the ambient concentrations of NO2 and NO from their EPA data, students are tasked with calculating the apparent steady-state concentration of ozone in the atmosphere using a simplified chemical kinetic calculation of the nitrogen oxides (NO_x) cycle in the atmosphere. It is expected that students will see a large discrepancy between the concentrations of ozone calculated via this method and the ambient values noted through the monitoring stations. This discrepancy encourages students to think reflectively upon the other atmospheric phenomenon taking place in the troposphere; for instance, volatile organic compounds, which are not taken into account in the NOx cycle, contribute heavily to the overall tropospheric ozone balance, necessarily resulting in a difference between what was observed and what was estimated. The discussion taking place here is meant to prime students for the subsequent component of the project, in which it is likely that unexpected dependences may factor into incongruities between estimated and observed trends.

Students are then asked to reverse geocode the locations of the students' EPA monitoring stations so that their approximate address, and subsequently their US Postal Service ZIP code, could be inferred. Using these ZIP codes as ZCTA values, which were found to match entirely in the Spring 2022 instance of this course due to the students selecting populated areas for their projects, 2020 Census data for the five areas around each of the students' monitoring stations can be downloaded from the US Census Dataset. Upon downloading the datasets, students were asked to look through the list of US Census Tables and posit a minimum of three sets of demographic factors that might lead to differences in ambient pollutant concentrations. When

possible, their specific knowledge of the layout and population of the area is encouraged to be included in their reasoning.

Trends of pollutant concentration versus the demographic properties are then regressed using simple linear fits for the five sets of census and pollutant data, with confidence intervals for slope, and intercept to determine a statistically significant dependence between the phenomenon. Upon completion, students then remarked on their correlations to explain why their analyses did or did not align with their expectations, either from a statistical (e.g., noise-to-signal ratio or effect size) or causal standpoint (e.g. proximity to industrial structures or high-traffic roadways, population density, etc.)

The learning outcomes for this project were for students to be able to:

- Draw information from a variety of online models and databases,
- Estimate atmospheric pollutant concentrations given limited information, and validate against existing datasets for model accuracy,
- Develop substantive hypotheses regarding potential causal societal factors for pollutant concentrations, and
- Use a statistically appropriate method to infer trends, or lack thereof.

Students were allowed to present their results in any form of summary that they deemed appropriate: while most employed a more typical laboratory report style structure to their reports, some employed PowerPoint, or slide style presentations to emphasize the graphical results that they had attained in their analysis. This open-ended structure received was anecdotally met with mixed reviews from students. Underclassmen who were less familiar with laboratory report structures struggled with the lack of specific structural points in the assignment, while upperclassman who were more familiar with a wide range of different communication skills appreciated the flexibility in the presentation. This being said, as the flexibility in presentation format is not critical to the learning outcomes of this project, institutions that emphasize uniformity and consistency across instruments of assessment, a singular mode can be specified without otherwise altering the scope of work performed. While students were expected to submit assignments individually due to the differences in their hometowns (and by extension, the locations they were finding data for), they were also encouraged to work together to identify commonalities in their findings and to assist one another through their analyses.

Results

Students were able to conceive a wide range of potential sociological factors that could affect air quality. While many chose median income or race as metrics for discussion, others took more nuanced subsets of data for the purposes of their discussion, many of which were informed by their own identities (as people of color or first-generation college students), as well as by knowledge of the neighborhoods in which the monitoring stations were located:

- *“One of the factors that could impact the correlation between travel time to work and atmospheric emissions is the mode of transportation workers are taking to work. I specifically chose means of transportation to work by vehicle available because I think this*

classification speaks to the likelihood of an individual to drive themselves to work versus taking public transportation. The more individuals who drive themselves to work versus taking public transportation the more positive correlation will exist between travel time to work and atmospheric emissions because there are more emissions being generated per person." –A junior chemical engineering major, on selecting Travel Type as a census metric in the DC/Maryland/Virginia area.

- *"It is no secret that marginalized groups of people are disproportionately affected by environmental issues and concerns. We see recent examples like the Flint, Michigan water crisis and past examples like Uranium mining, and tailings pond pollution in Navajo Nation in the middle and later half of the 20th century. We want to explore and see if atmospheric conditions and pollutant concentration seem to disproportionally affect marginalized groups by race and socio-economic status. The race and income tables are obvious in this but the reason for focusing on language as well is due to a flaw in the census data. The Race Census table does not distinguish between white and white of Hispanic or Latino origin. DFW and Texas in general is home to lots of Hispanic immigrant communities and so the Language Tables will be a better representation of that group of people."* – A junior integrative engineering major, on selecting percentage Spanish-speaking population as a census metric in Texas.

Examples of student-calculated regressions and discussions are shown in Figures 1 and 2. Upon compiling the EPA pollutant data with the census data subsets, students largely found that measurable and statistically significant trends in NO_2 concentration as seen in Clark et al., and far fewer or weaker trends with respect to NO and ozone. For students from more remote or rural regions, where zip codes encompass a large geographic area, negligible trends were observed across all pollutants and social factors, limiting conclusions that could be drawn apart from low population density.

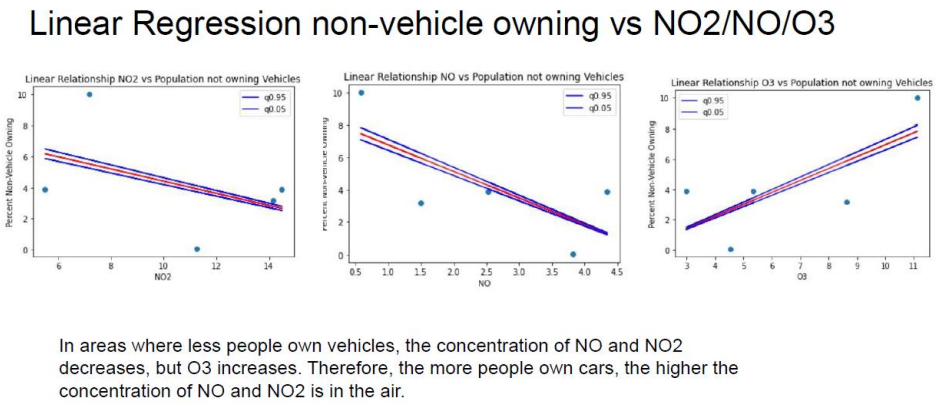
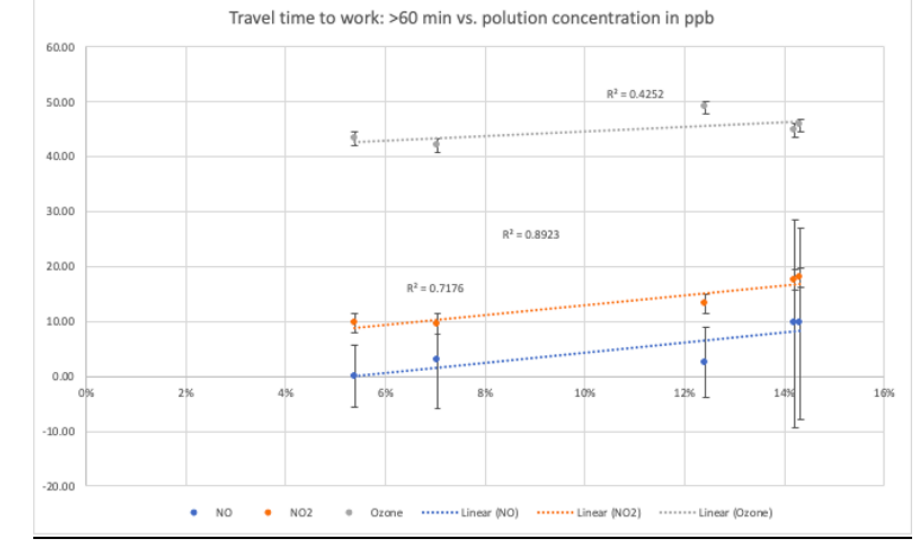


Figure 1. Student work (junior year, integrative engineering major) demonstrating relationships between airborne pollutants and vehicle ownership in five areas the Pittsburgh region using a PowerPoint slide format, performing analysis in R.

Common mistakes in this project largely occurred with the interpretation of statistics that were inferred from their students' findings. Since both the census data and Atmospheric pollutant readings were gathered from online databases students were occasionally confused with respect to which of the two data sets were the factor and which were the response; as a result, several students reported figures with atmosphere pollutant concentration in the x-axis and social demographic information in the y-axis, further muddying already noisy trends in their regressions. Other students applied common-practice statistical significance thresholds of $p < 0.05$ in this project, without considering the inherent variance to the census and atmospheric data used, resulting in some trends that were apparent in their data to be dismissed unnecessarily. Conversely, other students identified trends with statistical significance but an extremely small magnitude of effect. Most of these points of confusion can be attributed to the intrinsic noisiness in this data, implying that for future iterations of this project more than five data points would be necessary to achieve stronger goals.



The graph above shows the correlation between the number of people that take more than an hour to travel to work and the pollution concentration. The first time that we observe trustworthy results that align for all three pollutants. We can conclude that as the percentage of people increased, the concentrations of NO, NO2 and ozone all increased. Significant R-values and consistent error bars suggest that there is a strong linear relationship between the two quantitative variables and that the trendline is indeed a great fit. A possible explanation for this trend may be because of exhaustion by automobiles. As the number of people that travels to work for more than an hour increases the number of cars on the road increases which leads to higher exhaust emissions.

Figure 2. Student work (sophomore year, chemical engineering major) demonstrating relationships between airborne pollutants and percentage of residents with longer than a 60-minute commute to work in the Los Angeles area using a lab report format, performing analysis in Microsoft Excel.

Student Reception and Discussion

Likert-scale responses from a standard, College-wide course evaluation for the three total instances of the course *Atmospheric Engineering and* Science are reported in Figure 3. The Spring 2019 instance of the course utilized exams and homework as means of assessment. The Fall 2020 and Spring 2022 instances both used a four-project model; three of the four projects were identical, the only change being the replacement of the census project described in this work.

The responses to the question, "I learned a great deal in this course" yielded comparable results across all three instances of this course, with average scores of 4.56 in Fall 2019, 4.75±0.48 in Spring 2020, and 4.50±0.62 in Spring 2022. The relative proximity of these scores compared to the variance observed implies that the implementation of these projects, as well as the newer census-based project, did not appreciably enhance or disrupt the students' learning experience.

When asked if "the tests/assignments provided [students] adequate opportunity to convey the learning expected in the course," responses rose from an average score of 4.05 in Spring 2019 to 4.75±0.48 in Fall 2020 and 4.38±1.00 in Spring 2022. Further, in the Fall 2020 and Spring 2022 semesters, nearly all students agreed or strongly agreed with the question, with a single outlying student in Spring 2022, who expressed strong disagreement.

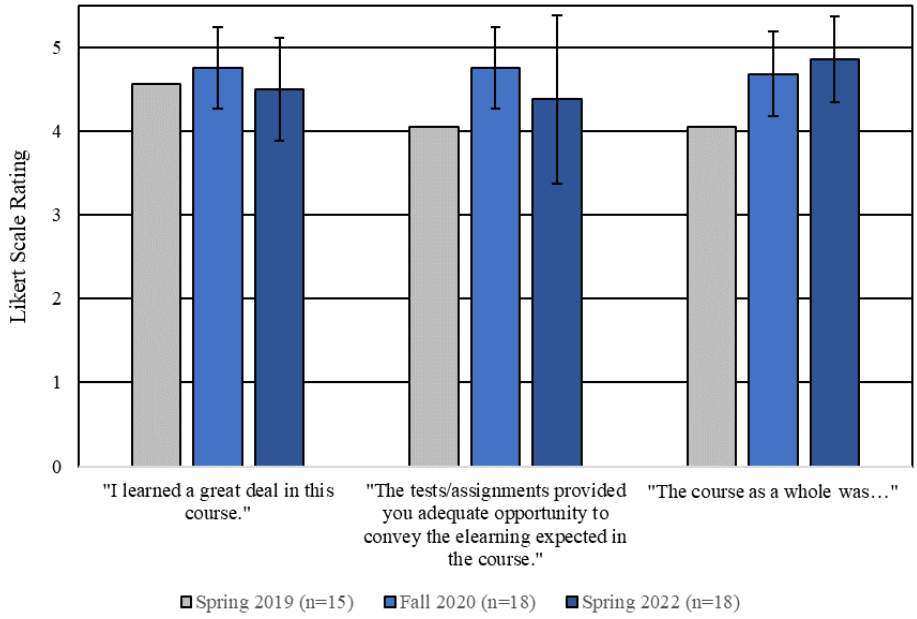


Figure 3. Semester-end course evaluation scores for *Atmospheric Engineering and Science*. The Spring 2019 semester used homework-and-exam forms of assessment compared to project-based assessment in the other two semesters; the projects between Fall 2020 and Spring 2022 were identical with the exception of the census-based project discussed in this work. Error bars denote one standard deviation of Likert scale scores; Spring 2019 used a different institutional format of aggregated course evaluation with less granular response data, but identical questions.

In qualitative responses, students were asked "How did the various components of the course contribute to your learning?" to qualify the previous question. Students pointed to the usage of

real-world datasets and the connection to their homes as the primary strengths in the course overall, observed examples of the Spring 2022 response data below:

- *"Although [the projects] were time consuming, they were an opportunity for us to apply what we were learning to our own environment and atmosphere in our hometowns. I also think that they were good opportunities for learning more about programming with MatLab [sic] and Excel as well as how to present research."*
- *"I found that the projects contributed the most to my overall learning in this course. I felt that they were a great way to get hands-on experience working with atmospheric data and they challenged me to put thoughtful analysis into why I was seeing the trends I was seeing and incorporate knowledge we had been taught in class."*
- *"The lecture allowed me to not only learn general principles about the atmosphere, but also get answers to specific questions. The projects helped to reinforce this, but more importantly, they allowed me to become more familiar with working with large data sets."*
- *"The lectures and projects for this course coincided with one another very effectively. The knowledge from the lecture could be carried over to help with the projects and the topics built off of one another. Also, the projects enabled me to connect the concepts to real-life scenarios as well as become more comfortable working with and analyzing large data which will be very beneficial for my future engineering career."*

This question also received a single negative comment in Spring 2022, which corresponded with the singular outlying strong disagreement to the question asking about tests and assignments.

- *"In this class, we had lecture and projects. I learned a great amount from both of these opportunities. However, lecture and projects were completely unrelated and I found that very frustrating."*

When asked about the course as a whole, however, every student in both the Fall 2020 and Spring 2022 instances responded with "Very Good" or "Excellent," with a sharper difference between the Spring 2019 data (4.05 in Spring 2019, compared to 4.68±0.51 in Fall 2020 and 4.86±0.51 in Spring 2022. As such, it can be inferred that student reception to the course was broadly improved with the implementation of projects as a mode of assessment, with neutral or positive effects from the addition of the census-based project.

Conclusions

The usage of this new project has demonstrated that issues regarding environmental justice can be implemented into data-driven engineering courses with comparable amounts of student learning, and neutral to improved perception of a course on a whole. The inclusion of real-world both atmospheric and demographic data provides a direct connection of these environmental phenomena to a student's personal context, their hometown area, which enables their lived experiences to be reflected in the analyses performed in this course. It should be noted that the Fall 2020 instance of this course, which contained similar projects except for the census project, was delivered via synchronous, online lecture instead of in-person due to the COVID-19 pandemic. However, it is not expected that the change in format resulted in any significant

difference in student perception of course or project materials. More importantly, the diversity of students capable of taking a course of this nature is likely to result in different strengths and weaknesses in project performance. While the cohort of students in the Fall 2022 instance of the course were primarily chemical engineers, who have similar curricular experiences and exposure to statistics, coding, and data visualization, the historical enrollment of the course containing students from mechanical and civil engineering implies that future instances of this project may necessitate more or less varied amounts of support to achieve similar outcomes.

References

[1] M. J. Martin *et al.*, “The climate is changing. Engineering education needs to change as well,” *Journal of Engineering Education*, vol. 111, no. 4, pp. 740–746, 2022, doi: 10.1002/jee.20485.

[2] E. Aoki, E. Rastede, and A. Gupta, “Teaching Sustainability and Environmental Justice in Undergraduate Chemistry Courses,” *J. Chem. Educ.*, vol. 99, no. 1, pp. 283–290, Jan. 2022, doi: 10.1021/acs.jchemed.1c00412.

[3] A. Hajat, C. Hsia, and M. S. O’Neill, “Socioeconomic Disparities and Air Pollution Exposure: a Global Review,” *Curr Envir Health Rpt*, vol. 2, no. 4, pp. 440–450, Dec. 2015, doi: 10.1007/s40572-015-0069-5.

[4] L. Schweitzer and A. Valenzuela, “Environmental Injustice and Transportation: The Claims and the Evidence,” *Journal of Planning Literature*, vol. 18, no. 4, pp. 383–398, May 2004, doi: 10.1177/0885412204262958.

[5] United States Census Bureau, “Census Bureau Data.” https://data.census.gov/ (accessed Feb. 26, 2023).

[6] L. P. Clark, D. B. Millet, and J. D. Marshall, “National Patterns in Environmental Injustice and Inequality: Outdoor NO2 Air Pollution in the United States,” *PLoS ONE*, vol. 9, no. 4, p. e94431, Apr. 2014, doi: 10.1371/journal.pone.0094431.

[7] E. V. Novotny, M. J. Bechle, D. B. Millet, and J. D. Marshall, “National Satellite-Based Land-Use Regression: NO2 in the United States,” *Environ. Sci. Technol.*, vol. 45, no. 10, pp. 4407–4414, May 2011, doi: 10.1021/es103578x.

[8] United States Environmental Protection Agency, “Air Data: Air Quality Data Collected at Outdoor Monitors Across the US,” Jul. 08, 2014. https://www.epa.gov/outdoor-air-quality-data (accessed Feb. 26, 2023).

[9] National Oceanic and Atmospheric Administration, “Weekly Daily Weather Map PDF Files.” https://www.wpc.ncep.noaa.gov/dailywxmap/pdffiles.html (accessed Feb. 26, 2023).

[10] PurpleAir, “Real-Time Air Quality Map | PurpleAir.” https://map.purpleair.com/?mylocation (accessed Feb. 26, 2023).

[11] B. Beckerman, M. Jerrett, J. R. Brook, D. K. Verma, M. A. Arain, and M. M. Finkelstein, “Correlation of nitrogen dioxide with other traffic pollutants near a major expressway,” *Atmospheric Environment*, vol. 42, no. 2, pp. 275–290, Jan. 2008, doi: 10.1016/j.atmosenv.2007.09.042.

[12] National Center for Atmospheric Research, “National Center for Atmopsheric Research Quick TUV Calculator,” *ACOM: Quick TUV*. https://www.acom.ucar.edu/Models/TUV/Interactive_TUV/ (accessed Feb. 26, 2023).