Academic Report
INFOSCI 301
Team: Jiean Zhou & Nazirjon (NJ) Ismoiljonov
Prof. Luyao Zhang

Air Quality and Health Disparities: A Multidimensional Analysis

Abstract:

This visualization explores the potential relationships between air quality, income levels, and asthma prevalence with data visualization. By integrating air quality, income, and asthma prevalence, we want to uncover patterns that are not visible with isolated datasets. Using visualization techniques, we plan to show how these variables affect asthma prevalence. This contributes to public health by allowing for data supported decisions, and it aligns with the third UN Sustainable Development Goal of Good Health and Well-Being. Our visualization can guide public health decisions and raise awareness of environmental health risks. However, there is no clear correlation between AQI, asthma, and income.

Background and Motivation:

Asthma is a chronic respiratory condition that affects millions globally and has no known cure. While it is assumed that asthma could be influenced by environmental and socioeconomic conditions, the relationship between air quality, income levels, and asthma prevalence is unclear and remains an assumption. Do communities that have lower incomes and worse air quality consistently experience higher rates of asthma? Or are there more factors at play?

Data visualization plays an important role here, allowing us to explore multidimensional relationships but also communicating the findings clearly for both academic and casual audiences. It allows us to use integrated datasets and visualizations allow us to uncover hidden relationships.

Real World Inspiration:

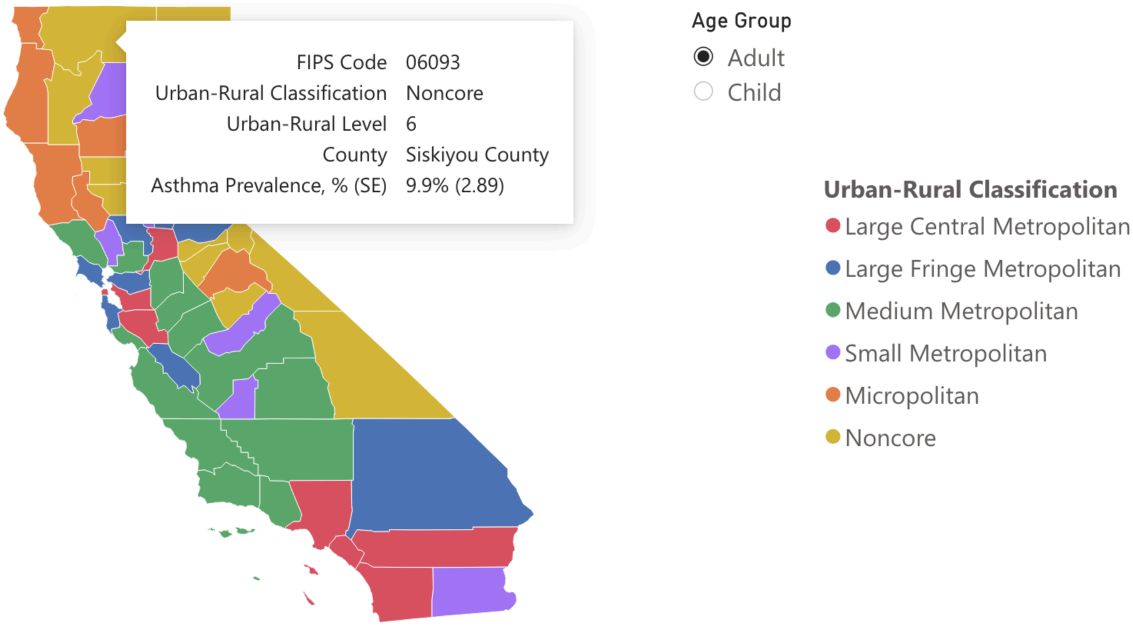

Figure 1: Visualization from CDC showing asthma prevalence
https://www.cdc.gov/asthma/national-surveillance-data/asthma-prevalence-state-classification.htm

This visualization inspired our group to include the hover mechanic in our visualization to give more details to the viewer and to make our visualization interactive. Our project is better with this feature because it allows us to maximize our use of multiple datasets.

Research Questions:

How do patterns of air pollution exposure align with reported respiratory illnesses, and are these patterns influenced by socioeconomic status?

- Can visualizations uncover overlooked regions where lower air quality and health overlap?
- Are there patterns between air quality and low income or minority neighborhoods?
- Can these patterns reveal systematic disparities or suggest areas for policy intervention?
- How does the number of days with poor air quality (AQI) relate to adult asthma prevalence?
- What role does average household income play in state asthma rates?

- Are CO_2 emissions or fine‑particle ($PM_{2.5}$) exposure significant predictors of asthma?

Application Scenarios:

- Public health departments could use visualizations to highlight high-risk groups
    - Find areas with high asthma prevalence and poor air quality to have clean energy investments and health outreach programs
- Schools and journalists could use it as a tool for promoting environmental justice
    - Highlight the dangers of bad air
    - Even short term exposure to PM2.5 can trigger asthma attacks
- Urban planners can use air quality and income to design building layouts
    - could show how highways or industrial plants are placed disproportionately, and can create better decisions and emission regulations

Relevant Disciplines

- Public Health
- Public Policy
- Urban Planning
- Environmental Health

Methodology:

<u>Visualization Techniques</u>

- Marks and Channels
    - We used points in scatter plots to show individual data, areas in choropleth maps, and lines in our bar chart. These marks are paired with channels like position, color, and length to showcase variables like asthma prevalence, AQI, income, and PM2.5 levels.

- Arranging Tables, Spatial Data, and Networks
    - Spatial arrangement is a key feature in our choropleth maps. They allow viewers to see relationships in a geographical context. The sorted tables allow the comparison of numbers to be read in an understandable way at first glance.

- Data Abstraction and Task Abstraction

- We abstracted real concepts like air quality, income, and health factors into quantifiable and measurable ways like median income, average AQI, and asthma prevalence.

- Color Theory and Design
    - We used color scales from light to dark to make information more digestible and intuitive of high to low values.
    - Colors make income levels readable at a glance.
    - Text is contrasted from the background to ensure readability, and the layout avoids visual clutter.

These design choices work well because they make our complex data easier to understand and highlight the relationships in a clear way. Viewers are able to answer their own questions on their own.

<u>DataSets</u>

1. Data descriptor or benchmark paper Open AQ:
    a. OpenAQ: https://explore.openaq.org/?parameter=o3&active=false
    b. Papers:
    https://ui.adsabs.harvard.edu/abs/2023AGUFMIN41C0602R/abstract
    https://ehp.niehs.nih.gov/doi/10.1289/isee.2016.3150?
    c. Open source platform that aggregates global air quality from governments and research grade monitors. Includes real time data and historical data. Fully compliant with Fair principles

2. United States Census:
    a. https://data.census.gov/table/ECNBASIC2017.EC1700BASIC?g=050XX00US24003,24001&tid=ECNBASIC2017.EC1700BASIC
    b. Shows indicators like income, race, education level, and poverty rates
    c. gives context for understanding health outcomes

3. CDC Places:
    a. https://www.cdc.gov/places/tools/data-portal.html
    b. gives health information on areas around the US, like asthma prevalence

4. United States Energy Information Administration:
    a. https://www.eia.gov/environment/emissions/state/
    b. information on emissions levels for each state

The combination of these datasets allows us to create visualizations and analyze how air quality affects health, and the addition of the US census data allows us to ensure that our data is not skewed by outside factors. It shows a more multidimensional and complete story about how air quality affects us. All of these datasets are open access and from reputable sources.

Advanced Tools

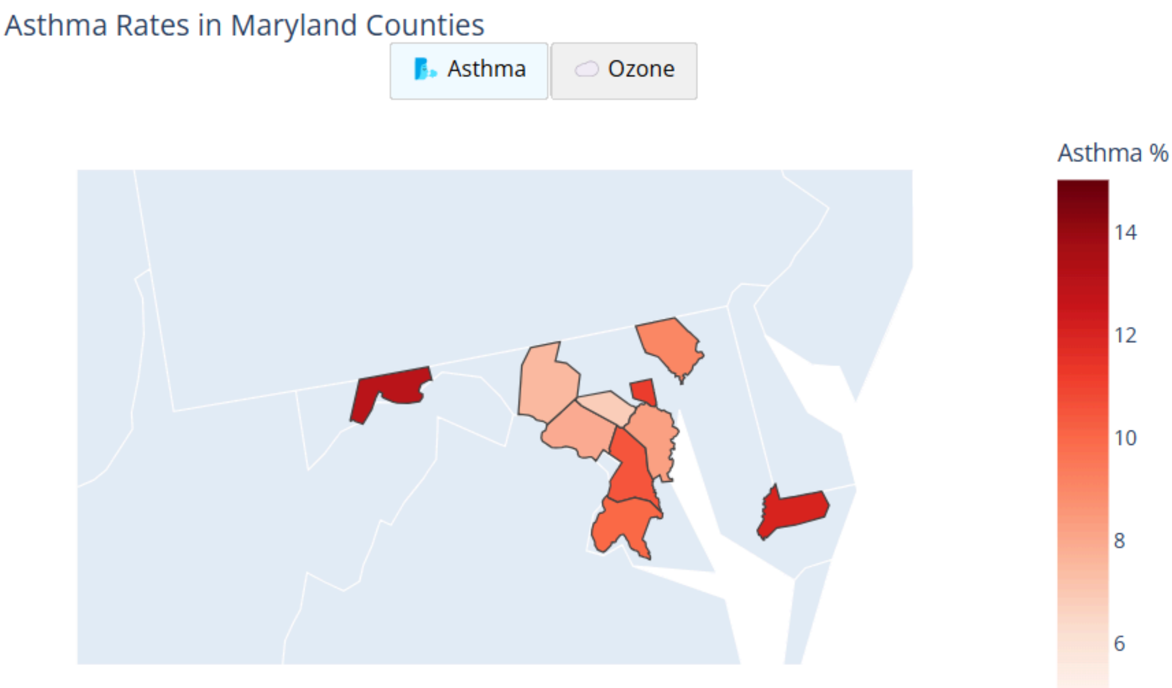

Figure 2: Preliminary Visualization showing asthma rates in Maryland's Counties

Technologies Used: Google Colab, Python (plotly), Streamlit (dashboard)
Interactive Features: Hovering for more information

We drew inspiration from A Computational Design Process to Fabricate Sensing Network Physicalizations by S. Sandra Bae, Takanori Fujiwara, Anders Ynnerman, Ellen Yi-Luen Do, Michael L Rivera, and Danielle Albers Szafir. The article discusses how physical sensors can be used for meaningful visualizations and emphasizes integrating real-world sensor data. We took this and used real sensor data like OpenAQ's air monitors that track real time to make invisible problems like air quality understandable to allow future action.

Results:

Anticipated Insights

- We expected correlations between air quality and asthma prevalence, where the lower the air quality, the higher the asthma prevalence.
- We also expected areas of lower income to have more asthma prevalence.

Observed Patterns

- We noticed no correlation between air quality and asthma prevalence.
- There was a correlation between income and asthma prevalence that suggests higher income groups have lower asthma rates, although it did not have significance ($p > 0.4$).
- The emissions of each state varied widely but also had no significance to asthma prevalence ($p > 0.2$).

Intellectual Merit and Practical Impacts:

Academic Contribution

- This project contributes to the academic field of data visualization and public health by exploring how environmental, health, and economic data can reveal patterns. We use data abstraction, spatial encoding, and color theory to visualize the relationship between air quality, income, and asthma prevalence and communicate the results in a clear way. We contribute to the research on environmental justice and public health with our visualization.

Practical Impacts

- Our project supports the United Nations Sustainable Development Goal 3 of Good Health and Well-being by uncovering environmental and social harms to respiratory health. By seeing how air quality and income affect asthma prevalence, we can create more research and drive community organizations to create better health. We also raise awareness for environmental health risks and promote healthier environments.

Reflection on Growth and Learning:

From the individual InfoVis Redesign project, we became familiarized with the process of creating visualizations and its supporting resources like posters, reports, and creating a github. We shared our strengths and weaknesses from the project to give ourselves the best roles to create the best visualization for our final project. The in-class activities improved our understanding, and gave practice for our visualization and gave ideas of inspiration for our final project. The final session with feedback was also very helpful. Group 2 told us the importance of color scheme and clarity, and this was something we kept in mind throughout our final process. The symposium improved our presentation skills and provided valuable feedback that we can take with us to our final project. From the Zhou Zhuang Mysterious Life Museum, our group saw that visualizations could be used in a real case and the importance of visualizations in the day to day and in academic fields.

Supplementary Materials

All data and code will be made public on GitHub with documentation and steps for replicability.
Github Link: https://github.com/jiiean/301-Final-Project
Dashboard Link: https://asthma-dashboard.streamlit.app/

Team Contribution Statement:

- **Jiean**: Data sourcing, literature review, writing of academic report and GitHub
- **NJ**: coding interactive visualizations, dashboard creation, poster creation

Acknowledgements:

Special thanks to Professor Luyao Zhang, Professor Ming-Chun Huang, David Schaff, and Dongping Liu

References:

Bae, S. S., Fujiwara, T., Ynnerman, A., Do, E. Y.-L., Rivera, M. L., & Szafir, D. A. (2023). A Computational Design Process to Fabricate Sensing Network Physicalizations. *IEEE VIS 2023*.

OpenAQ. (n.d.). *Explore OpenAQ*. Retrieved April 22, 2025, from https://explore.openaq.org/?parameter=o3&active=false

Centers for Disease Control and Prevention. (2023). CDC PLACES: Local Data for Bette Health. Retrieved April 22, 2025, from https://www.cdc.gov/places/tools/data-portal.html

OpenAQ. (2016). *OpenAQ: Open Air Quality Data for Research and Action*. *Environmental Health Perspectives, 124*(7), A118–A119. https://doi.org/10.1289/isee.2016.3150

Ramanathan, A., Wilkins, R., Murray, L., Kilaru, V., & Goshorn, M. (2023, December). *Crowdsourcing Air Quality: Supporting Community Science and Advocacy with OpenAQ* [Abstract]. AGU Fall Meeting 2023. https://ui.adsabs.harvard.edu/abs/2023AGUFMIN41C0602R/abstract

U.S. Energy Information Administration. (n.d.). *State carbon dioxide emissions data*. U.S. Department of Energy. Retrieved May 7, 2025, from https://www.eia.gov/environment/emissions/state/

U.S. Census Bureau. (2017). *2017 Economic Census: EC1700BASIC*. Retrieved April 22, 2025, from https://data.census.gov/table/ECNBASIC2017.EC1700BASIC?g=050XX00US24003,2001&tid=ECNBASIC2017.EC1700BASIC

Grammarly Grammar Report: