INFO 247: Final Project Writeup, Video, and Thumbnail

Project Title: Air pollution is not "out there." It could be right under your nose.

Team: Amber Jiang, Monica Paz Parra, Aigerim Kurmanbekova

Demo Link: https://amber-jiang.webflow.io/unlisted-air-asthma

Project Goals:

Our project aims to make air pollution data personal and accessible. Many people think of pollution as distant, but in reality, it very well could be "under your nose." By enabling users to search their ZIP code and interact with data visualizations, we help current and future California residents understand their local air quality, how it may affect asthma rates, and how pollution intersects with factors like poverty and race.

This interface supports a variety of exploratory tasks, including looking up the current AQI for a specific ZIP code, comparing county-level pollution and asthma rates over time, and investigating how social and environmental factors relate to health. The primary goal is to raise awareness and promote action through personalized and contextualized data exploration.

Discussion:

Our visualizations are designed to make pollution data understandable at both macro and micro levels:

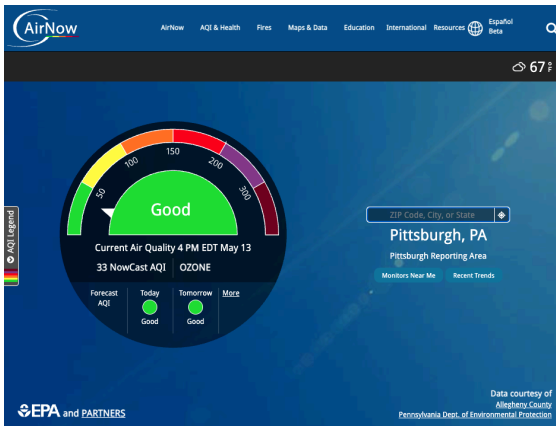
Part 1: Introduction: Do you know what is under your nose?
In this section, we wanted to introduce our users to their local air quality index (AQI) and help them better understand how it connects to their environment and health issues like asthma. We focused on visualizations that helped make the data feel more personal, connected to real life, and easier to understand.

- **ZIP Code AQI Lookup Feature**
  To make air quality feel personal and immediately relevant, we implemented an AQI lookup feature that allowed users to input their ZIP code and see their air quality in real time. Inspired by the AQI search feature on AirNow.gov (https://www.airnow.gov/), we designed our version to be visually simpler. Users

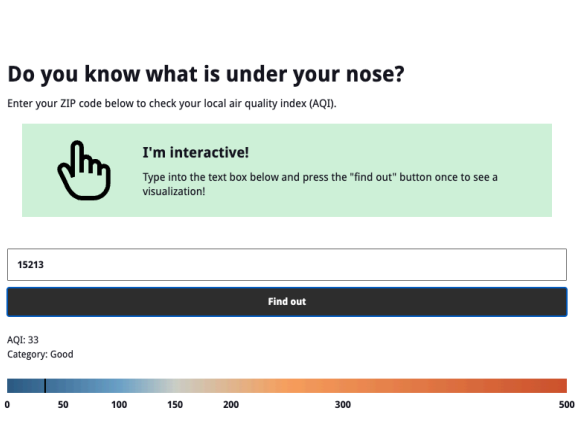
can enter any US-based ZIP code and receive their current AQI value, the category it falls into (e.g., Good, Unhealthy for Sensitive Groups, Hazardous), and where it lands on a color gradient AQI scale from 0 to 500.

To improve accessibility, we used a blue-to-orange color scale instead of the traditional green-to-dark red scale to accommodate users with red-green color blindness. We pulled real-time AQI data from the AirNow API (https://docs.airnowapi.org/webservices), which provides air quality data directly from EPA monitoring stations. This allowed us to make sure our feature was giving accurate and updated information.

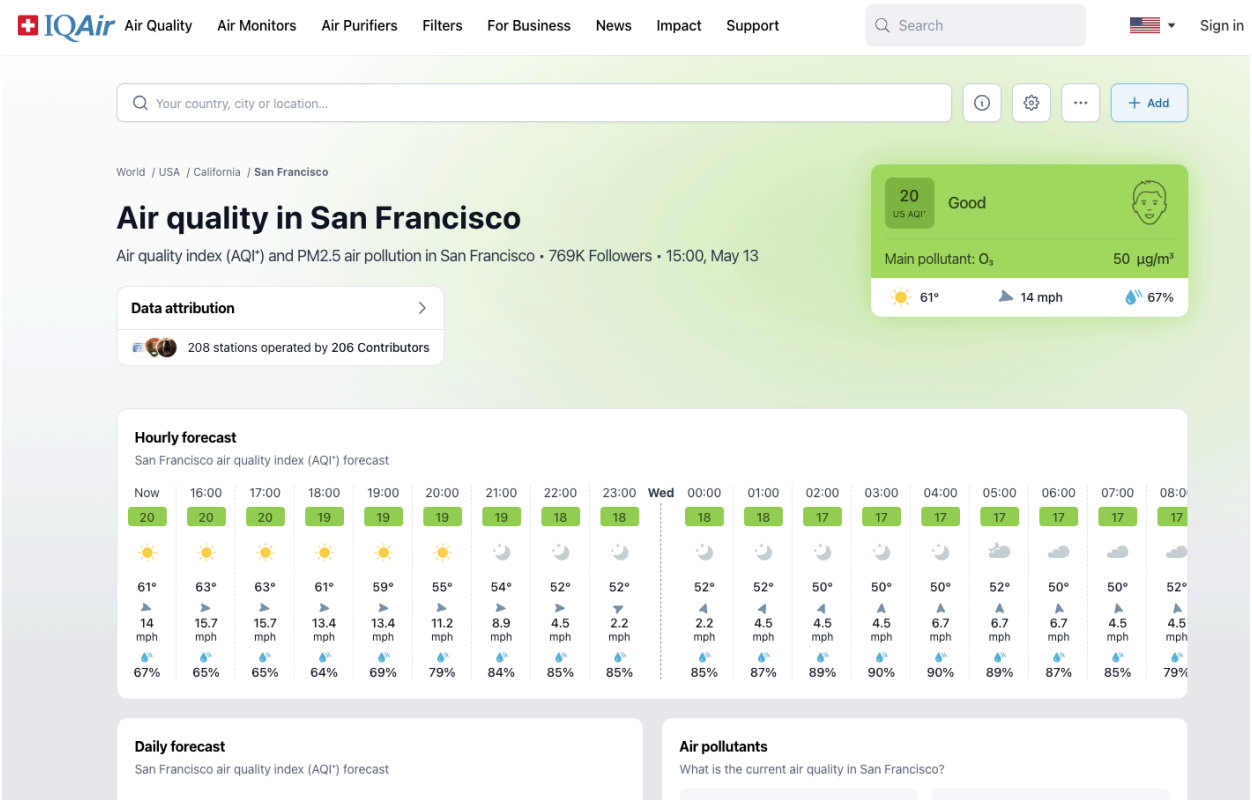
Original Color Scale



Our Color Scale



- **AQI Visual Progression**
  Along with our AQI Feature, we included a visual progression showing what different AQI levels look like in real environments. This idea was partially inspired by IQAir's air quality tool (https://www.iqair.com/us/), where if you search for a city like San Francisco, it provides detailed AQI data along with color-coded sections and icons that indicate air quality categories and other metrics.



While their presentation of information is data rich and visually organized, we found that it lacks a real-world visual reference to show how these AQI levels actually appear outside. Because of this, we decided to add an image sequence below our AQI ZIP code feature to help users visualize the impact of different AQI levels, ultimately bridging the gap between data and real life.

What do the categories mean? According to AirNow[1], a partnership of the U.S. Environmental Protection Agency, National Oceanic and Atmospheric Administration (NOAA), National Park Service, NASA, Centers for Disease Control, and tribal, state, and local air quality agencies, each category is defined as follows.

1. **Good:** Air quality is satisfactory, and air pollution poses little or no risk.

2. **Moderate:** Air quality is acceptable. However, there may be a risk for some people, particularly those who are unusually sensitive to air pollution.

3. **Unhealthy for Sensitive Groups:** Members of sensitive groups may experience health effects. The general public is less likely to be affected.

4. **Unhealthy:** Some members of the general public may experience health effects; members of sensitive groups may experience more serious health effects.

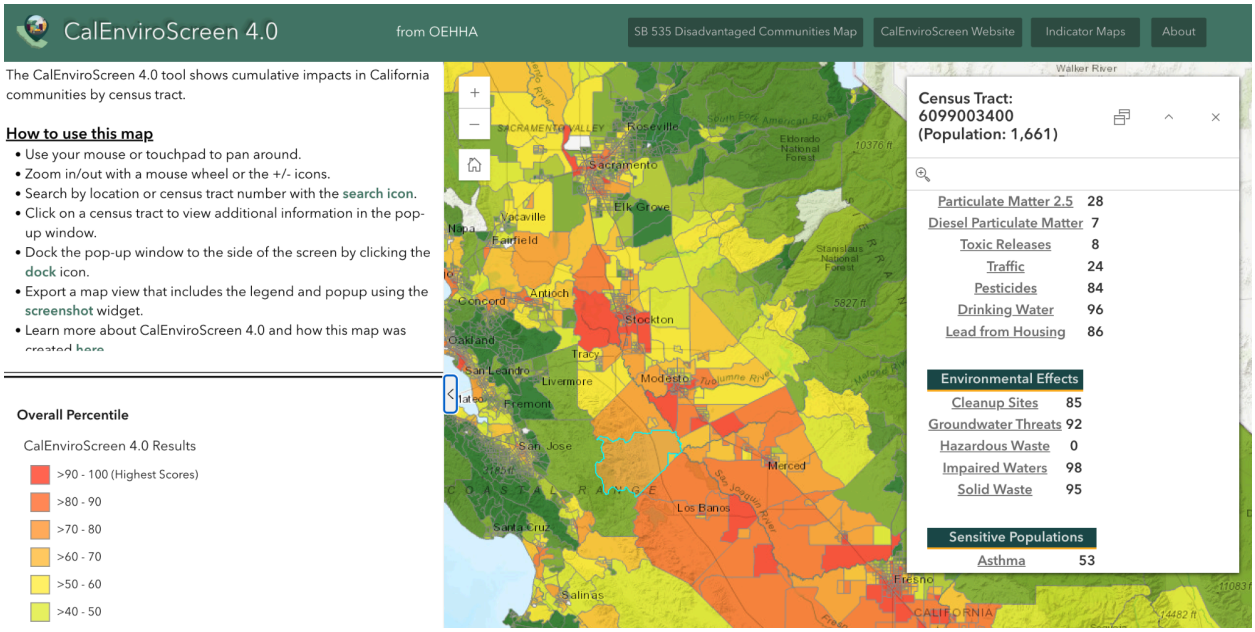
5. **Very Unhealthy:** Health alert: The risk of health effects is increased for everyone.

6. **Hazardous:** Health warning of emergency conditions: everyone is more likely to be affected.

The more severe the air quality and the more exposed you are to polluting facilities, the more susceptible you are to respiratory conditions like asthma. While many of these conditions can be genetically linked, the environment can and does affect severity.

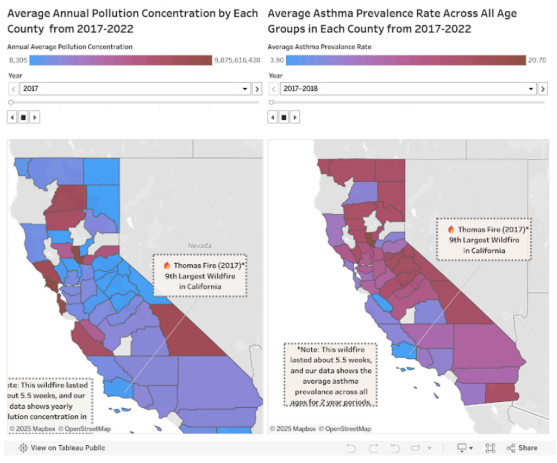
This is how areas across California compare in asthma rates and pollution levels.

- **Side-by-side Choropleths of Pollution and Asthma**
  To help users explore overall trends between air pollution and respiratory health, we created two choropleths: one showing the two-year average for air pollution concentration (in µg/m3) and the other showing the two-year average asthma prevalence rate (in %) for all age groups across California counties from 2017 to 2022. This design was inspired by the CalEnviroScreen 4.0 (CalEnviroScreen 4.0 Results), which maps cumulative environmental factors that impact California communities by census tract. We extended this idea by allowing users to compare two datasets side by side, something CalEnviroScreen 4.0 doesn't currently support, making it easier to see high-level trends between air pollution and asthma rates across counties.

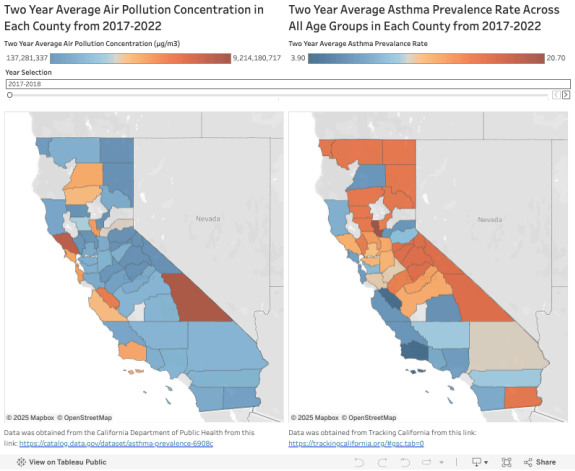


Our early version of this visualization had visual inconsistencies, especially when comparing data across different time periods. Through usability testing, we learned that the misalignment in year ranges created unnecessary cognitive load. Because of this, we reformatted both datasets into consistent two-year groupings and added a shared year slider that filters both maps simultaneously. These changes improved usability and encouraged users to explore more deeply how asthma and air pollution vary over time and across communities.

Early Version



Current Version



Here is a link to the reformatted data we used for this visualization and where we originally obtained the data from:

- Pollution Map Reformatted Data + Original Data Source
- Asthma Map Reformatted Data + Original Data Source

Part 2: The Link Between Air and Asthma

- In Part 2 of our website, we aimed to highlight the primary air pollutants in California and their impact on respiratory health. Initially, we used data from Tracking California (trackingcalifornia.org) as mentioned in our proposal and exploratory data analysis (EDA). However, specifically for part 2, this dataset did not meet our needs, as it lacked information about specific pollutant types present in California and included general AQI level only.

  To address this, we conducted additional research and found a more comprehensive dataset from the California Environmental Protection Agency (calepa.ca.gov). Because the website did not provide a direct download option, we manually extracted the relevant data. The dataset included information on a wide range of pollutants, which we cleaned and filtered to focus on the most significant ones affecting air quality and respiratory health in the state.

  Link to the dataset: V1_Raw EPA Air Quality Dataset

  Cleaned dataset was used in the final Observable Notebook. It's combined with the Asthma Dataset, but we decided to shift Astma to Part 3 for a smooth transition in our narrative, so Asthma Variables weren't used.

Visualizations

- **Pollutant Overview Bar Chart:**
  This bar chart shows PM2.5 and NO_2, the two key pollutants affecting respiratory health. Normalized average concentrations from 2015–2022 show PM2.5 slightly edges out NO_2 statewide. This helps users to get an overall understanding of pollution severity across California and contextualize later data.

  A bar chart was chosen as a visualization that is easy to understand, as discussed in the class. We differentiated the colors to show the difference between the two main pollutants in California.

- **PM2.5 and NO_2 Bubble Charts by County (2022):**
  Counties are grouped into low, medium, and high pollution tiers. Bubble color reflects concentration, while size emphasizes severity. This grouping makes regional disparities immediately visible. The bubbles are draggable and interactive,

making the experience more tactile, exploratory, and even a bit fun, inviting users to rearrange the data and discover patterns visually.

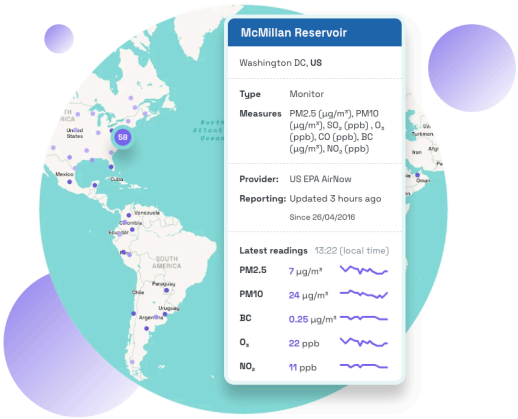
We got inspired by OpenAQ's (https://openaq.org/) bubble metaphor for our visualizations by county, using size and color to encode average pollution levels. However, we expanded on their approach by introducing categorical grouping (Low, Medium, High) to help users more easily compare regions. Also, we made the bubbles interactive and draggable, turning a static map into a more playful, exploratory interface. This interactivity encourages users to physically engage with the data and draw their own comparisons (something OpenAQ's implementation doesn't offer). Our goal was not only to visualize pollution but to spark curiosity through direct manipulation.

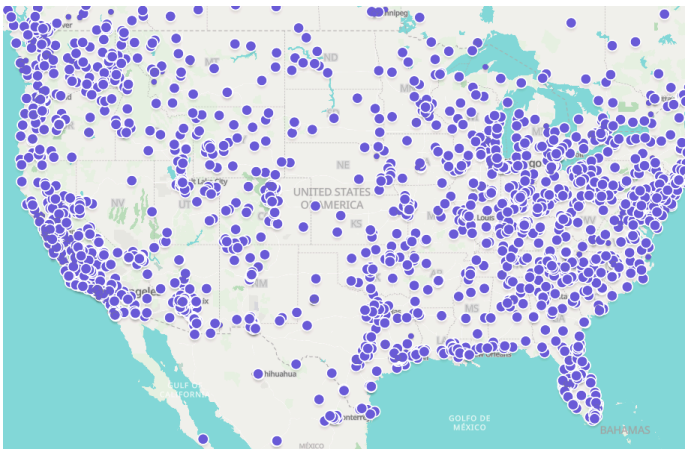
Together, these visualizations help users:

- Understand pollutant differences
- Identify geographic inequalities

Inspirations

Our early design was inspired by real-time environmental monitoring platforms such as OpenAQ and Water Data for the Nation, which use map-based interfaces and circular markers to represent environmental conditions across broad geographies. These visualizations effectively use spatial distribution and color to communicate local severity and accessibility of environmental data.





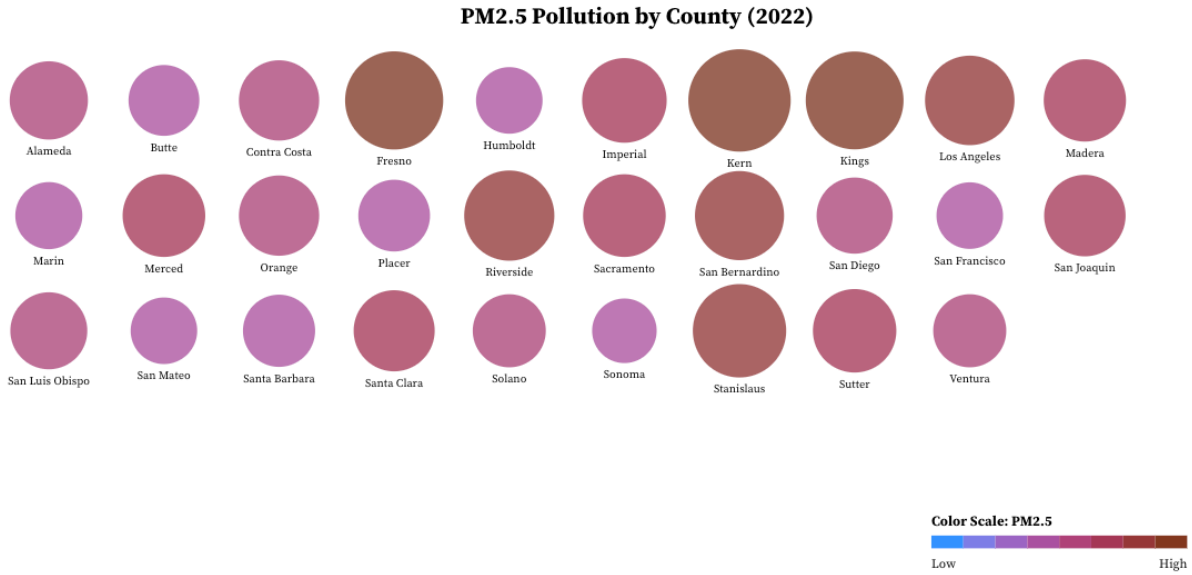
Initial Design (screenshots of current designs are provided in the next section)

Bubbles

The first iteration of our bubble design displayed (in this example PM2.5) pollution level by county in California using a grid layout rather than a map. This was a design choice

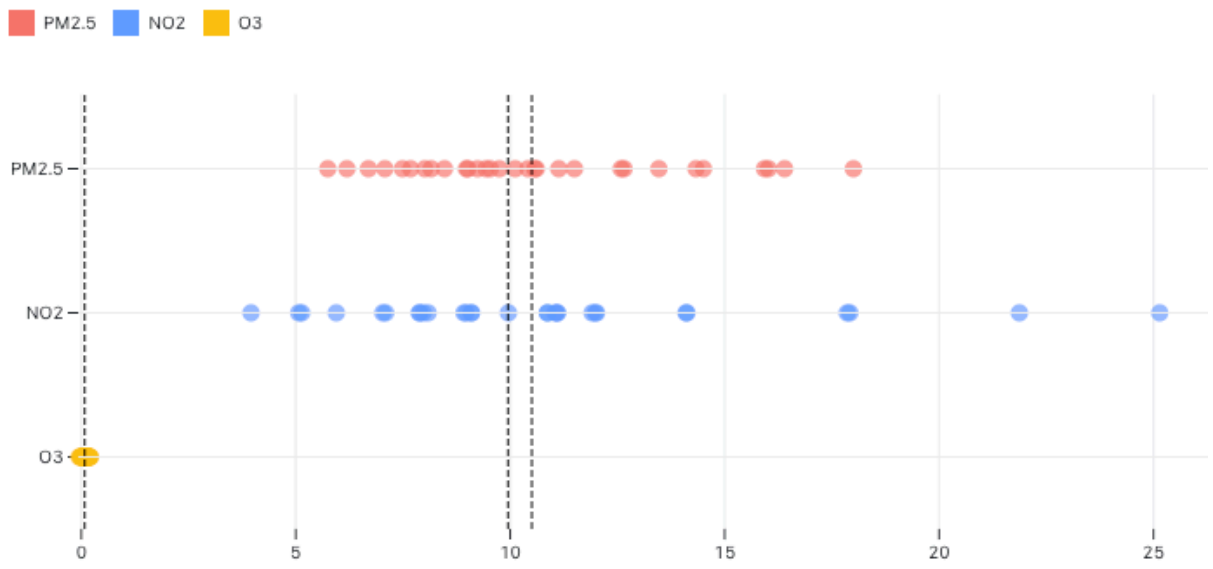
intended to simplify comparison between counties and draw attention to relative severity without relying on spatial proximity. Each circle represented a county, and we used a single-variable color scale (light to dark) to represent a pollution concentration.

While the concept helped emphasize differences in pollution levels, it lacked interactivity and context, leading us toward more engaging, draggable bubble charts with integrated legends, grouping by pollution tiers, and multi-dimensional encodings like size and color. These changes were driven by a desire to support exploratory learning and comparative reasoning more effectively.



Bar Chart: Alternative Visualization of Pollutant Levels

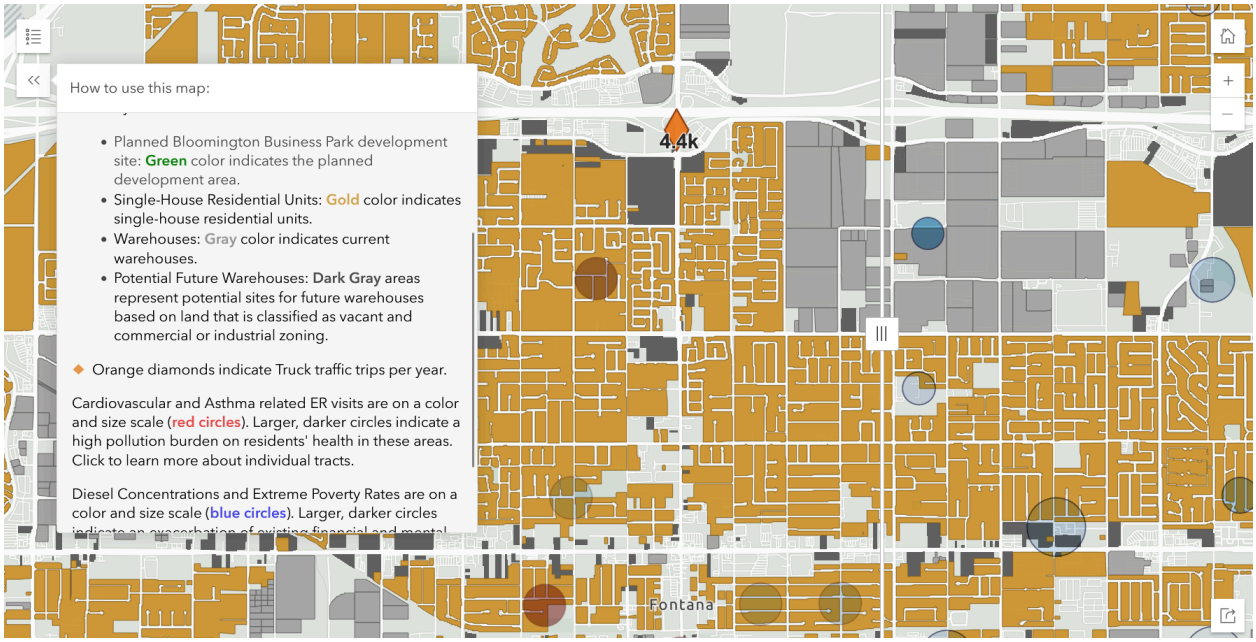
This simplified dot-based bar chart presents three key pollutants - PM2.5, NO_2, and O_3 - using individual data points across counties. Each pollutant is color-coded for clarity, and dashed vertical lines indicate key threshold values (e.g., EPA health-based standards). We chose this bar chart format for its simplicity and readability, allowing users to quickly compare pollutant distributions across regions without overwhelming visual complexity. Also, we removed O_3 because it existed in a small amount. This version addresses prior user feedback by making individual values easier to interpret while maintaining a clear structure.



Part 3: Zooming In: A Case Study on the Crisis of the Inland Empire

The third part of our website strives to contextualize pollution and asthma more in the context of our society, demonstrating that they are not issues existing in a vacuum. To paint the most accurate picture of reality and our society, I wanted to demonstrate how different identities can face different pressures.

- **Residential Proximity to Railyards:**
  Railyards and high vehicle activity are a major source of air pollution. Proximity to such polluters are often linked to many socioeconomic factors. *Mapping Black California* Project Manager Alex Reed has created a map that visualized areas of high vehicular traffic in Bloomington, overlaid with health impacts.



While this map is thorough and comprehensive, it's not easy to interpret. We were inspired to find similar data in order to create an easier-to-understand visualization that more simply showed the relationships between vehicular activity, race, and asthma.

- **Correlations Between Poverty, Asthma, and Healthcare:**
  Poverty and healthcare are other factors that relate strongly to asthma. A publication in the *Allergy and Asthma Proceedings* journal has revealed that besides race, low-income makes a large difference on asthma severity. However, this being a scientific publication meant that it was filled with advanced jargon and was strenuous to read. Because accessibility is one of our main goals, we wanted to visualize a similar message in a form that is comprehensible to most adults. We compiled multiple sources from the government and organizations to show a correlation plot between poverty and asthma.

  Furthermore, the National Health Council has discussed that poverty "hinder[s] adequate access to health care and resources". The Council, however, does not provide any visualizations as to what this decrease in access looks like. We understand that it is more impactful on a reader to be able to see a demonstration of what decreased access means, so we believed it to be worthwhile to show which areas suffer from poverty, and also where healthcare centers exist. And because healthcare is linked to poverty, and poverty is linked to asthma, healthcare is therefore linked to asthma as well. We've overlaid the factors in a couple of

choropleths.

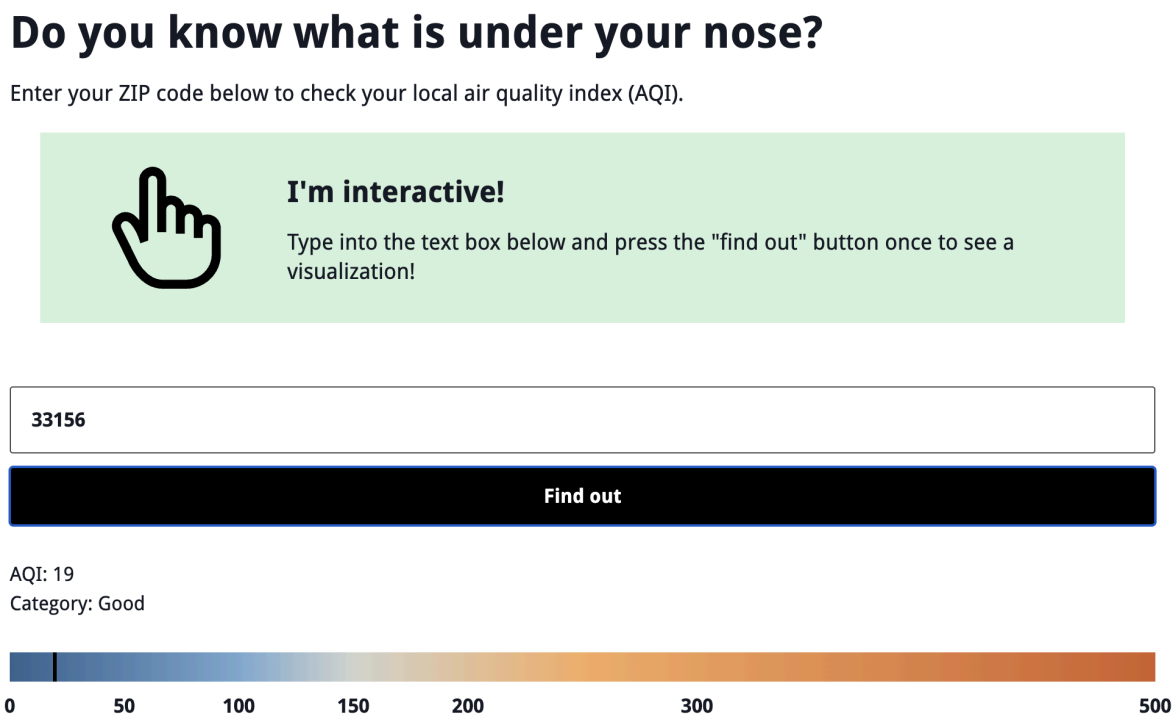
Description of Visualizations

1. **Main Title Images:**

   The project title and introduction are paired with a series of real images from different regions in California to show a range of air quality conditions, from clear coastal skies to smog-covered urban areas. These visuals emphasize that pollution impacts communities across the state and help set the tone of the project by grounding environmental data in familiar places. Our goal with this introduction is to reinforce that air quality is a local and immediate concern.

2. **AQI Feature:**

   With this interactive tool, users can enter any US-based ZIP code to check their current location's Air Quality Index (AQI). Once submitted, the AQI value and its corresponding category, ranging from Good to Hazardous, are displayed along with a color-coded scale from 0 to 500. This visual scale helps users quickly understand where their local air quality falls within the broader spectrum. This feature offers users a fast and personalized snapshot of their local environmental conditions while serving as an introduction for the rest of the project.



Video Link: AQI Look Up Feature.mov

3. **AQI Visuals:**

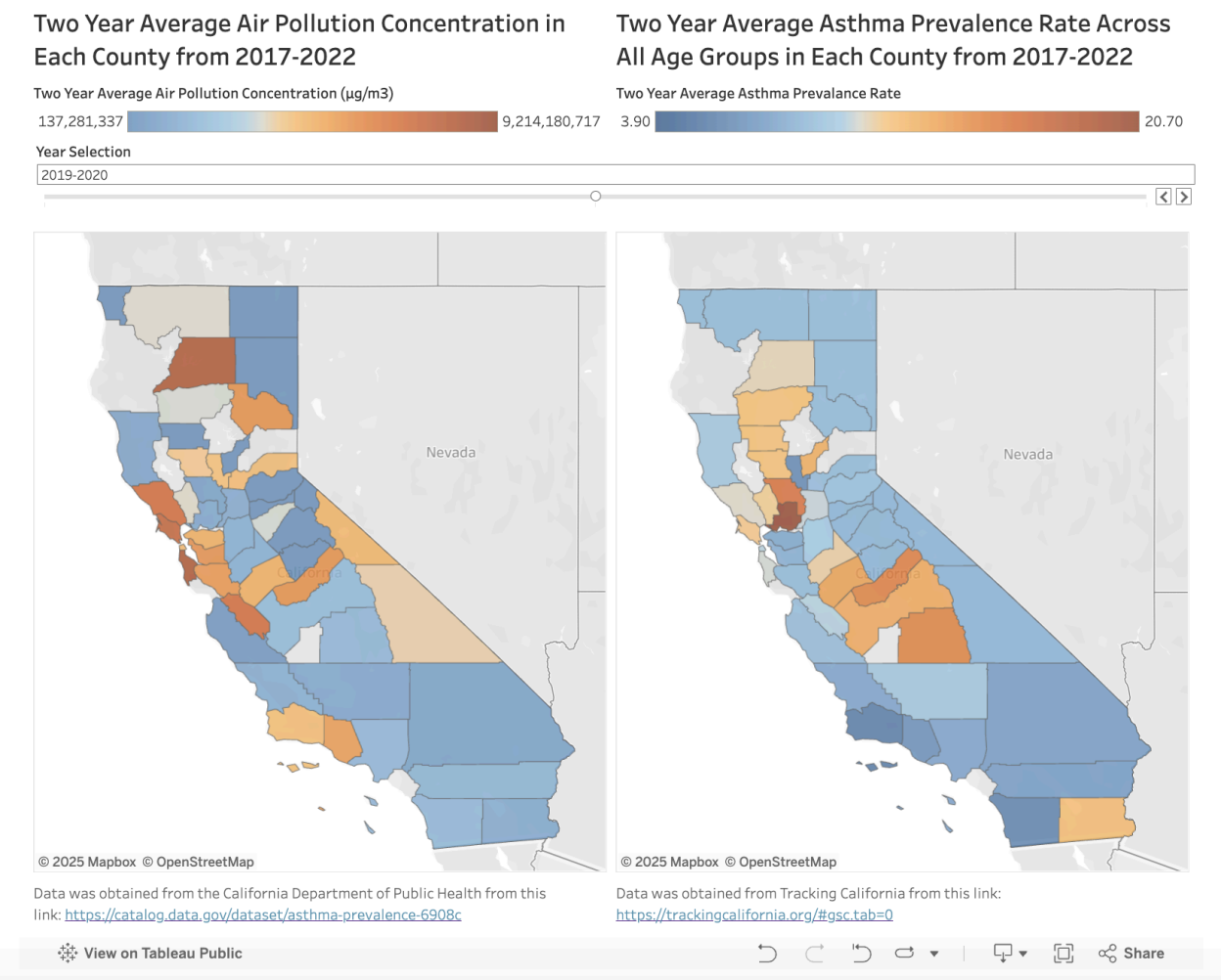
   This visual shows how air quality conditions change across different AQI levels, using real-world images to represent increasing pollution severity. From clear skies in Berkeley at 20 AQI to heavy smog internationally at 400 AQI, these images help users visualize what these values mean in context. The worsening progression highlights how air pollution impacts visibility and daily life, reinforcing the importance of understanding environmental data and the associated health risks, such as asthma. This section helps bridge the gap between numerical AQI values and their real-life application.

The more severe the air quality and the more exposed you are to polluting facilities, the more susceptible you are to respiratory conditions like asthma. While many of these conditions can be genetically linked, the environment can and does affect severity.

This is how areas across California compare in asthma rates and pollution levels.

4. **Side by Side Choropleths: Air Pollution and Asthma Prevalence**

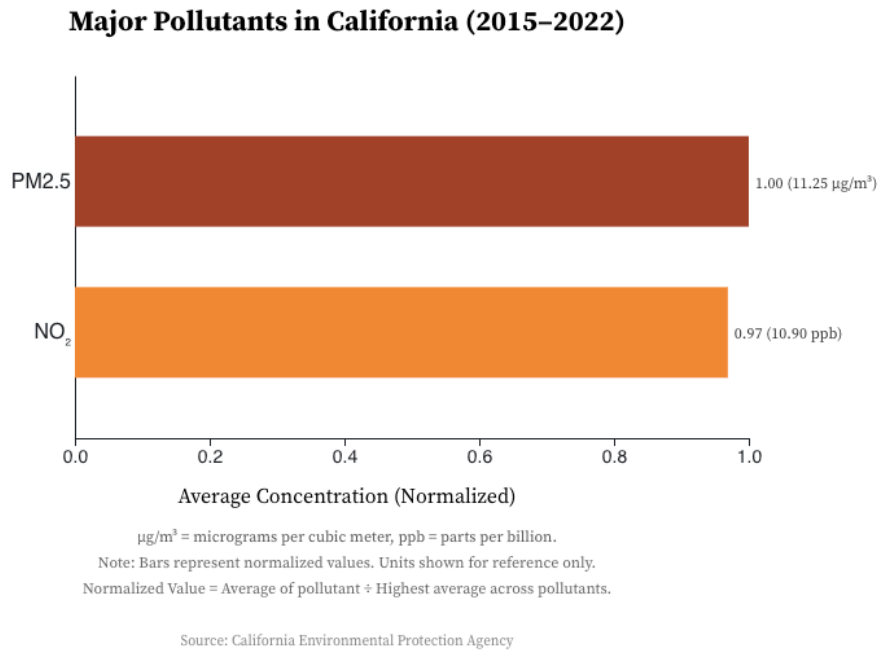
   These choropleth maps show the two-year average for air pollution concentration (in $\mu g/m^3$) and asthma prevalence rate (in %) across all age groups in each California county from 2017 to 2022. Both datasets in this visualization were aggregated into two-year periods to match the reporting format of the asthma data, making it easier for users to compare spatial patterns over time. An interactive slider was added to allow users to explore how pollution and asthma rates have evolved across regions. Additionally, the diverging color scale made it easier to identify counties with lower or higher asthma or air pollution levels. This visualization was designed to support high-level, quick comparisons of pollution and asthma trends, encouraging users to explore potential geographic relationships between the two factors.



Video Link: Exploring California Air Pollution and Asthma Over Time.mov

5. **Bar Chart: Major Pollutants in California**
    a. This bar chart compares the average concentrations of PM2.5 (fine particulate matter) and NO_2 (nitrogen dioxide) over a seven-year period in California. The values are normalized, with PM2.5 serving as the baseline (1.00). PM2.5 emerges as slightly more prevalent, with a real average of 11.25 $\mu g/m^3$ compared to NO_2's 10.90 ppb. This visualization was chosen for its clarity and accessibility, helping users quickly compare the severity of two key pollutants linked to respiratory health issues like asthma. The visualization is animated.

Pollutant	Unit	Type of Pollutant	Health Impact Focus
PM2.5	µg/m³	Fine particulate matter	Penetrates lungs, affects heart and lungs
NO_2	ppb	Gaseous pollutant	Irritates airways, causes inflammation

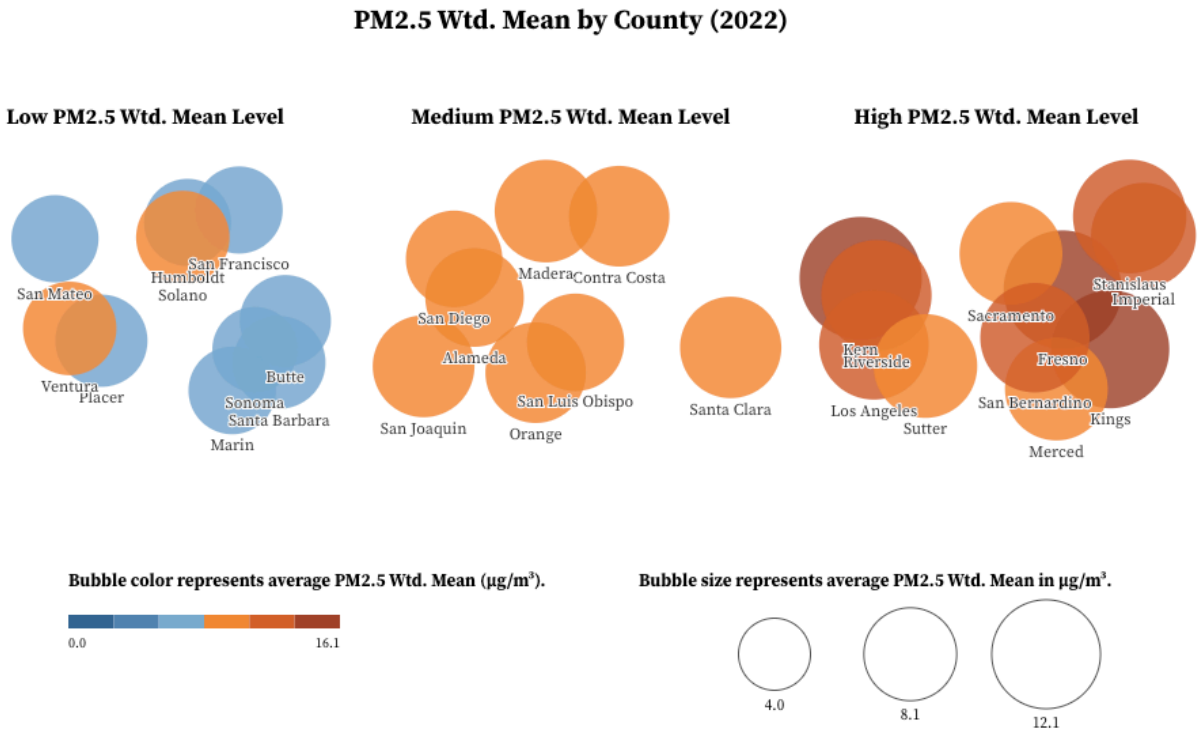


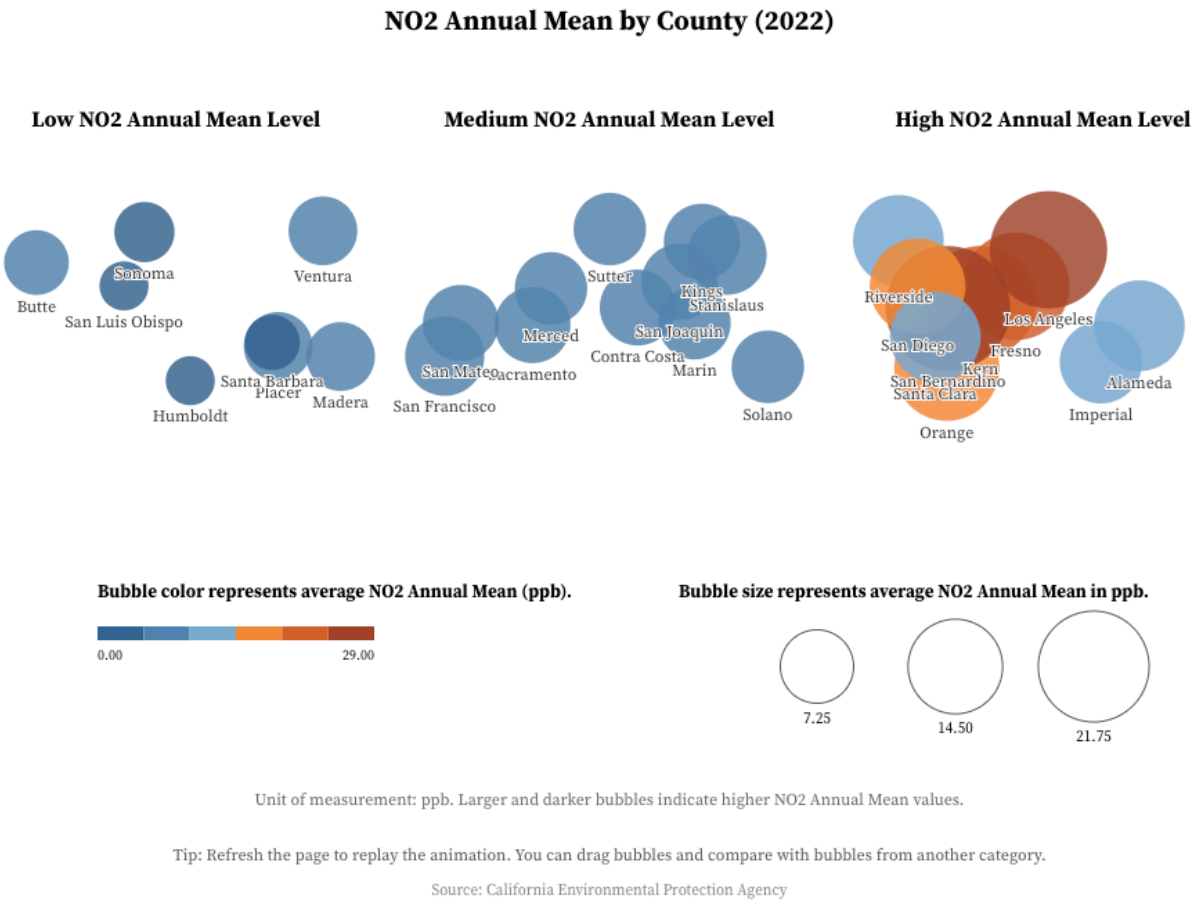
6. **Bubble charts: PM 2.5 and NO_2**
    a. Bubble charts visualize annual pollutant levels across California counties in 2022 for NO_2 (nitrogen dioxide) and PM2.5 (fine particulate matter) - two pollutants strongly linked to respiratory health issues. Each chart groups counties into low, medium, and high exposure tiers to facilitate quick comparisons.

       Both charts use bubble color to represent average pollutant concentration (darker shades = higher levels) and bubble size to show the relative magnitude of pollution. The visualizations are interactive: users can drag and rearrange the bubbles, making the data feel tactile and encouraging deeper engagement with patterns and outliers.

       Together, these charts make regional disparities in air quality easy to understand. High-risk counties like Los Angeles, Kern, Fresno, and San

Bernardino emerge as particularly impacted by both pollutants, showing the environmental and public health challenges faced by certain communities.



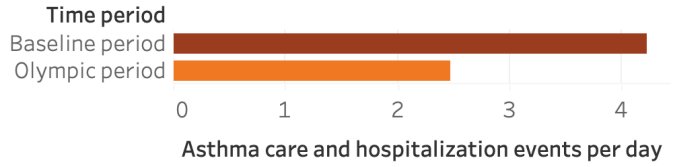


Video Link: Bar Chart and Bubbles Animation.mov

7. **Asthma care and hospitalization events in the 1996 Olympics**

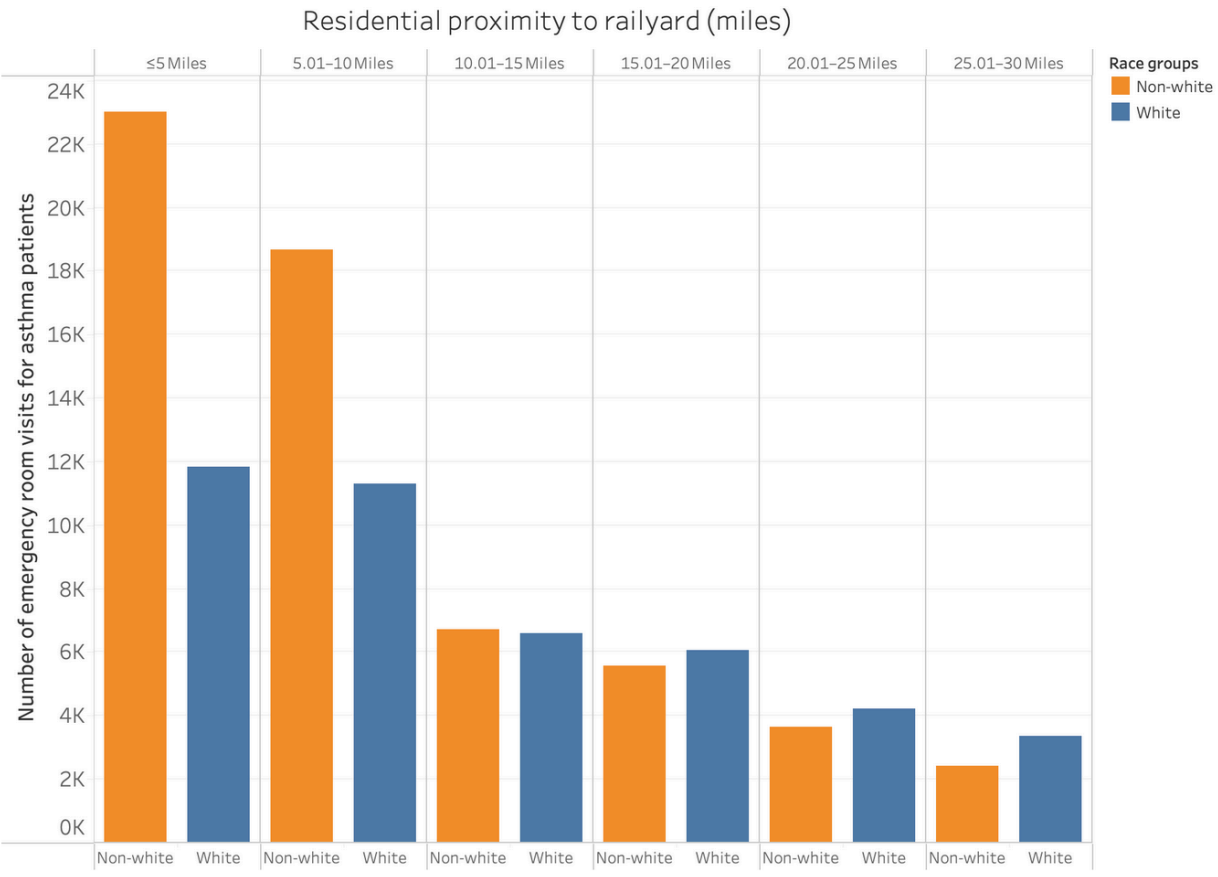
   This image is a simple horizontal bar graph that compares daily hospitalization events for asthma during the Olympic period in Georgia during 1996 versus a typical Georgian period. What becomes apparent here is that during the Olympic period, hospitalization events have decreased. The surrounding narrative (not pictured) explains how pollutants decreased drastically during this time. While not making a direct claim about causation, it allows readers to theorize and see that asthmatic events can vary quickly within short timeframes.

Number of asthma emergency care visits and hospitalizations among children aged 1 to 16 from Medicaid claims file database (Friedman, 2001)



8. Emergency room visits bar chart

This visualization is a bar graph with their heights being the number of ER visits for asthma patients. On the x-axis represents white and non-white populations, faceted by their proximity to a railyard across California. Two trends become apparent. One, the closer residents live to railyard, the more ER events occur for asthma patients. Two, for residents that live in closer proximity, more ER events are from non-white patients than white patients.



9. Asthma Emergency Department Visits vs Poverty Rate

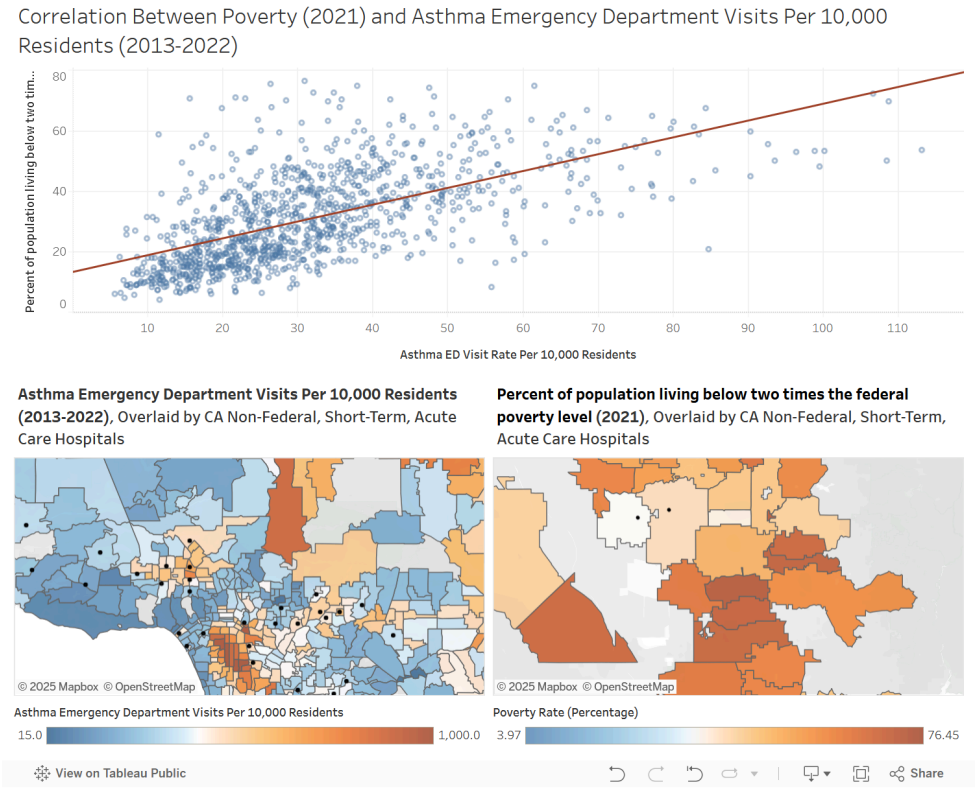
Video: Poverty and Asthma Correlation.mp4

This graph shows three charts. The first, on top, is a scatter plot with asthma emergency department (ED) rates per 10,000 people on the x-axis and poverty (defined by percent of population living two times below two times of the federal poverty level). Overlaid on top is a trend line to more clearly show the relationship. Hovering the line indicates the correlation strength as well as the analysis confidence (p-value). Hovering a dot shows the exact poverty value, the asthma rate, and the ZIP code for which the data corresponds to.

The bottom left graph shows asthma in a choropleth, where each ZIP code is color-coded based on the intensity of asthma ED visits. Overlaid on top, with black dots, is the location of a specific type of hospital: non-federal, short-term, acute care. This does not include small clinics, just large scale hospitals.

The bottom right graph is the same as the bottom left, but instead of asthma ED visits, it shows poverty level.

What is demonstrated with this visualization is that there's a strong relationship between poverty and asthma. Additionally, healthcare access is unequally distributed across the state.



Data Used:

- AQI Feature: https://docs.airnowapi.org/webservices
- Air Pollution Concentration Map: https://trackingcalifornia.org/#gsc.tab=0
- Asthma Prevalence Rates Map: https://catalog.data.gov/dataset/asthma-prevalence-6908c
- The Link Between Air and Asthma Visualization Data - Pollution data by type: https://calepa.ca.gov/
- Inland Empire Case Study Visualization Data - Asthma Emergency Department Visits Data: https://data.ca.gov/dataset/asthma-emergency-department-visit-rates
- Poverty Data: https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-2022-zip-code-data-soi
- Hospital Data: https://www.ahd.com/states/hospital_CA.html
- Railyards and hospital visits: https://pubmed.ncbi.nlm.nih.gov/30533348/
- Olympic 1996 transportation and asthma hospitalizations: https://pubmed.ncbi.nlm.nih.gov/11180733/

Tools Used:

- **Tableau:** To create visualizations used on the website
- **Observable**: To create visualizations used on the website
    - **D3** in Observable to create animations
- **Webflow:** To build our project website
- **Microsoft Excel:** To clean and maintain our data
- **VSCode:** To build the prototype + final version for the AQI feature
- **Google Drive:** To organize project documents and datasets
- **Slack:** Used for our team communication
- **Google Calendar:** To schedule weekly team meetings and user interviews
- **Google Meets:** To conduct user interviews and hold weekly team meetings
- **Python:** To perform pivots, clean data, and merge datasets (previous expertise)

Results Obtained:

To evaluate the usability and interpretability of our visualizations, we conducted a structured study with three participants familiar with California's geography. Participants were recruited through convenience sampling, with selection based on availability and prior or current residency in California. Each participant completed a series of guided tasks across three sections of our website focused on local AQI lookup, pollutant and asthma comparisons, and a case study of the Inland Empire. In every session, we asked

our participants to describe what they were doing, noticing, or confused by as they interacted with each feature.

Key Findings:

- Participants responded positively to interactivity, especially the personalized ZIP-code-based AQI lookup tool. This feature was described as fun, intuitive, and engaging, and we plan to retain and enhance it.
- The correlation scatterplot between poverty and asthma stood out as the most impactful visualization. Users described it as clear, compelling, and emotionally resonant. It effectively highlighted structural inequities and left a lasting impression.
- Despite placeholder text, participants were still able to orient themselves in the topic space, showing that the visual design successfully communicated key themes.

Areas for Improvement mentioned in the interviews were:

- Participants struggled to interpret visual metaphors (e.g., blue for clear skies vs. blue as a high value). Colors lacked consistent meaning across the interface. Based on feedback, our users wanted to see a more intuitive green–orange–red scale to represent "good to bad" outcomes, but we decided to use a blue-orange scale for red-green color blindness accessibility. We also included clear legends for all visualizations.
- Axis labels (e.g., "poverty") caused uncertainty: users weren't sure if it referred to income levels, population proportions, or another metric. In our final version, we made labels more descriptive and defined variables in context. Additionally, users expressed a desire to assess data validity for themselves, so we included direct links to our data sources to support transparency and trust.
- Some users expected maps to be clickable or dynamically responsive (e.g., popups updating with the timeline or choropleths filtering by year). Enhancing interactivity and responsiveness will be a priority for future iterations.

Design Implications
Based on user feedback, we:

- Unified our color scale across visualizations and clearly explained it with legends and inline context.
- Improved axis labels and tooltip content for clarity and contextual depth.
- Expanded the narrative framing to better support key plots.

- Preserved and improved engaging elements like ZIP-code lookup and drag-and-drop bubble charts.
- Added granularity to maps and timelines to support deeper temporal and spatial comparisons.

Additional Feedback from Final Showcase:

During our final project showcase, we received valuable feedback from classmates, class staff, and general reviewers. Those points included:

- At the top, users requested a single-year slider to control both maps simultaneously in the first side-by-side choropleth maps. Because of this, we implemented a year control for both plots in our final project website.
- Multiple users suggested cutting down the text in some sections, especially the "So What?" and "Limitations" parts, by making the content more visually digestible.
- People loved being able to drag the bubbles around (interactive and fun) as well as the AQI API (personalized)
- One user pointed out that the bar chart comparing pollutant types may not have been the most effective choice, since the two pollutants (PM2.5 and NO_2) appeared nearly identical in value. They suggested that instead of using a chart, it might be more impactful to simply highlight this similarity.

Overall, the usability study and final project showcase gave us invaluable insights into how our interface supports exploratory data interpretation. While users were able to extract meaning and engage with the visualizations, their feedback pointed us toward critical improvements in visual encoding, labeling, and interactivity that strengthened the clarity and impact of the final product.

Distribution of Work

Team Members	**Parts Done**	**Proportion**
Amber	• Gathered, cleaned, and organized data sets for asthma rates, poverty data, hospital data, railyard data, Olympics data; merging on ZIP code and pivoting data fields • Constructed the website, embedding and implementing materials from other teammates; formatted text, compiled pictures, and hosted the site	33%

	• Wrote the narrative for the project and defined the necessary data/visualizations • Developed visualizations for the "Zooming In: A Case Study…" portion of the site (along with the Olympics visualization) • Standard participation in general background research, presentations, weekly discussions, group assignments, etc.	

Monica	• Cleaned and reformatted datasets found and preprocessed by my classmates for asthma rates and air pollution rates.
	• Designed and implemented the Tableau visualizations in Part 1 - "Do you know what is under your nose?" section
	  ○ Side-by-side choropleths
	• Prototyped and designed the AQI ZIP code feature for Part 1 to embed into Webflow
	• Conducted two User Interviews and gathered feedback for our first round of visualizations
	  ○ Figured out the color scheme for us to use post feedback
	• Helped with general research and structuring and completing group assignments
	• Attended weekly check-ins
	• Cowrote the "So What? Here's Why You Need To Know This Information" section and references section on the website
		33%

Aigerim	• Researched air pollution in California and identified relevant datasets across multiple sources • Cleaned, merged, and organized datasets related to general air quality and pollutant-specific data, including integration with asthma prevalence	33%

	data. • Designed and implemented visualizations for the Part 2 - The Link Between Air and Asthma ○ Interactive Bubble and Animated Bar Charts • Conducted a User Interview and gathered feedback on our Version 1 visualizations. • General research, weekly check-ins, group assignments, etc.	

Thumbnail Link: Thumbnail.png

Software/materials Links:

- Observable: https://observablehq.com/d/18ed024b41ad6f0c
- Tableau files will be submitted with this assignment.
  Attachments for Final Report
- Code files for the API Feature will be submitted with this assignment and this file:
  Attachments for Final Report
- Python code processing is also shared here Attachments for Final Report