South Gate CEHAT

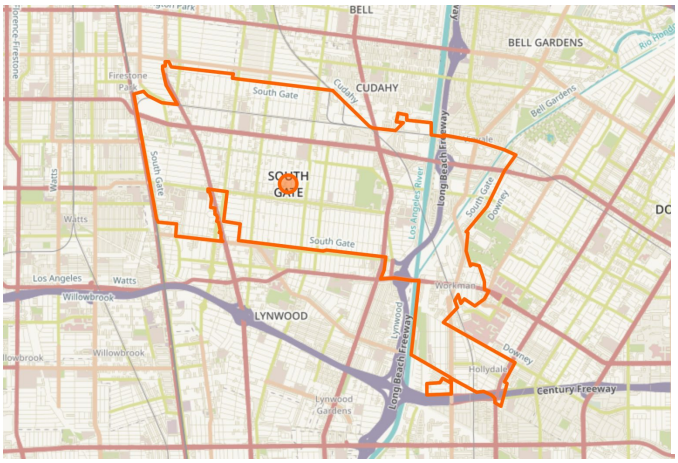
Harvey Mudd College Math Clinic

2020-2021

Hi everyone, welcome to our final presentation. We are the South Gate CEHAT Clinic team advised by Professor Medero. Our liaisons are Samir Patel, Liz Ruiz, and Victor Ferrer.

[SWITCH SLIDE]

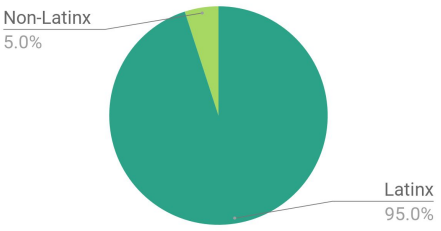
The City of South Gate



https://www.openstreetmap.org/#map=13/33.9382/-118.1938&layers=H

98,633
residents

Demographics



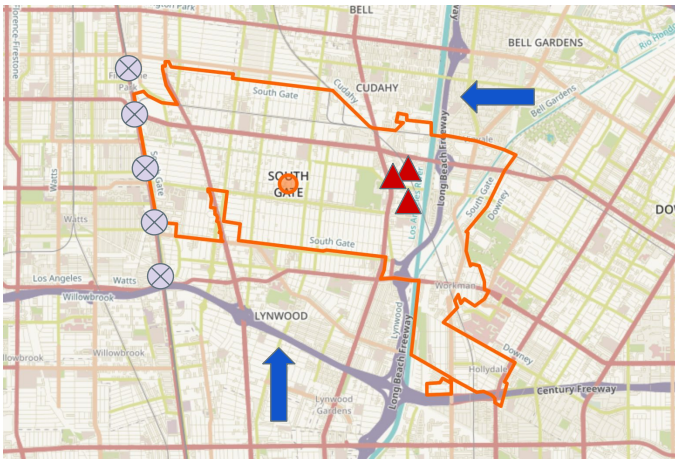
https://www.cityofsouthgate.org/258/City-Statistics

Our clinic project is community based, so it's important that we offer some background to the community we're working with.

The city of South Gate is in Southern Los Angeles near the intersections of the 710 and 105 freeways. There are just under 100 thousand residents, of which 95% are Latinx. It's important to note the demographics of the community not only because it affects the users of our product, but also because of environmental justice concerns.

[SWITCH SLIDE]

The City of South Gate



https://www.openstreetmap.org/#map=13/33.9382/-118.1938&layers=H

EPA Superfund Site

Alameda Freight Railroad

South Gate City Limits

Major Highway

https://www.cityofsouthgate.org/258/City-Statistics

A historically industrial city, South Gate is subject to higher levels of pollution from industrial plants. Within the city limits there are three superfund sites, which are locations that the EPA has designated as among the most contaminated land in the country. Furthermore, South Gate is home to two oil and gas wells, whose toxic products and byproducts also pose a risk to residents.

In addition to contaminated sites, South Gate residents are also subject to high levels of pollution from the transportation industry. The 710 freeway alone brings over 20,000 diesel trucks through the city every single day. Train traffic is also a source of pollution, as 30-40 freight trains pass through every day by means of the Alameda Corridor, the freight network that connects the ports of Long Beach and Los Angeles, which are the two busiest ports in the country.

On top of all that, being in Southern California, South Gate is subject to intense fire seasons as well as pollution from the greater Los Angeles region. All of these factors combined puts the community of South Gate at high risk for environmental health consequences.

[SWITCH SLIDE]

Community Environmental Health Action Team

- Founded in 2013
- Cross-functional community group
- Take action to make South Gate a healthier place to live
- Current focus: air quality

This leads me to introducing our sponsor, the Community Environmental Health Action team, otherwise known as the CEHAT. Founded in 2013, the South Gate CEHAT is a community organization comprised of city residents, business owners, community activists, and members of state, local, and federal government.

The CEHAT researches environmental health concerns in the city and works with the EPA and regulators to improve conditions in South Gate. Another large focus of the CEHAT is education and outreach with the local community to bring awareness to these issues and inform community members how they can stay safe.

The CEHAT is currently focused on the poor air quality in the city.

[SWITCH SLIDE]

Air Quality

- $PM_{2.5}$ - Particulate matter 2.5 microns in diameter
- $PM_{2.5}$ concentration → directly measured by sensors

- AQI - Air Quality Index (National standard)
- AQI → computed from $PM_{2.5}$ concentration

Air quality is affected by compounds such as ground level ozone and nitrogen oxides, as well as particulate matter (which ranges from 1 to 10 micrometers in size). PM 2.5, which stands for particulate matter 2.5 microns in diameter, is the largest contributor to air quality measurements and is of particular interest to the CEHAT. This size of particulate matter is common in emissions from transportation and power plants and is particularly dangerous because, at 30 times smaller than the width of the average human hair, it has the ability to travel deep into our respiratory tract and lead to health consequences.

The concentration of PM2.5 can be directly measured, and then used to calculate an AQI, which stands for Air Quality Index. The Air Quality Index is the U.S. national standard for measuring air quality. We use EPA formulas for converting from PM 2.5 to AQI.

[SWITCH SLIDE]

PurpleAir Sensors

- Affordable (~$250 vs $15,000+)
- Accessible (only need WiFi and power source)
- Report PM_1, $PM_{2.5}$, PM_{10}

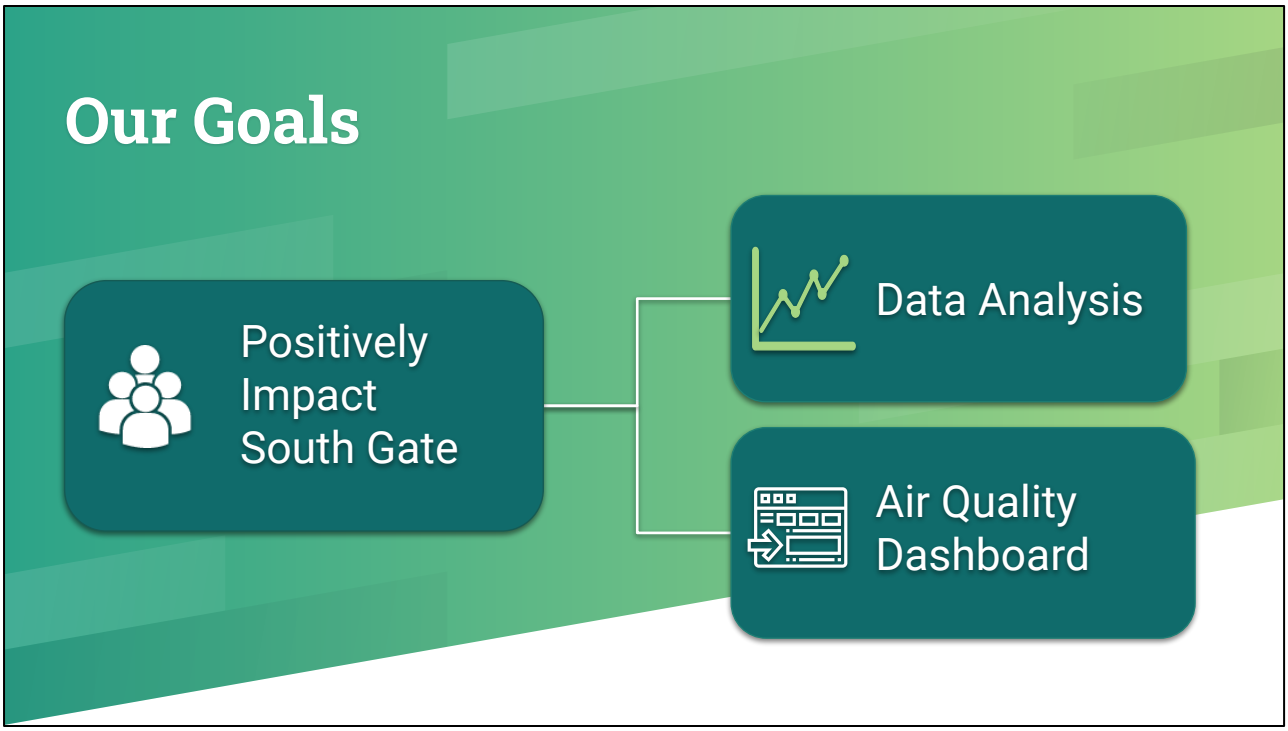
PurpleAir

purpleair.com

In 2019, the South Gate CEHAT received a grant for 20 PurpleAir sensors, which are an affordable air quality sensor option, costing around $250. These sensors are great for the CEHAT because they are easy to install and maintain as they only need WiFi and a power source. Despite their low price, they actually measure PM 2.5 concentration with reasonable accuracy. The EPA has even started integrating readings from PurpleAir sensors into its own fire and smoke map.

These sensors are essential to our project since there is no EPA-regulated sensor within the South Gate city limits, so residents currently have no local data to inform their health decisions.

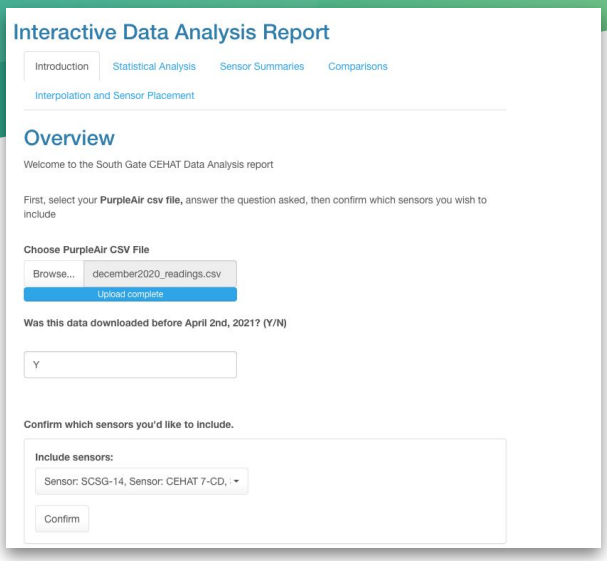
[SWITCH SLIDE]

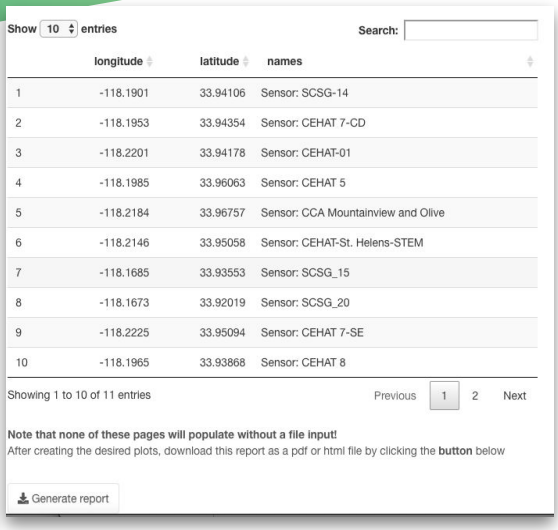


The goal of our project is to positively impact residents of South Gate by providing information, analysis, and education on air quality in the city. We hope to help residents make more informed decisions about their own health and to provide regulators with concrete information to use when creating policy.

Our work falls into two categories: an online air quality dashboard and reproducible data analysis for the CEHAT. I'll now turn it over to Hillary to talk more about our analysis work. **[SWITCH SLIDE]**

Data Analysis Report





Thanks Amber! For our data analysis work, we wanted to give the CEHAT members ways to visualize and analyze the Purple Air data. So, we decided to make an interactive report using R shiny so that the CEHAT members can reproduce the results for future data. Here's our report:
[SWITCH SLIDE]

http://127.0.0.1:4360 | Open in Browser | Publish

Interactive Data Analysis Report

Introduction | Statistical Analysis | Sensor Summaries | Comparisons | Interpolation and Sensor Placement

Hourly Readings for the Week up to the Selected Date:

Select the date:

2020-12-31

Hourly Readings for the Month up to the Selected Date:

Select the date:

2020-12-31

Select date ranges for the density plot:

Select the date range:

2020-12-01 to 2020-12-31

Compare 8-hour Averages for Two Different Time Periods

Select the date range:

2020-12-01 to 2020-12-05

Select the date range:

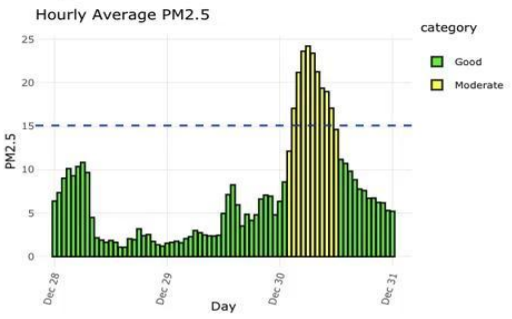
2020-12-08 to 2020-12-12

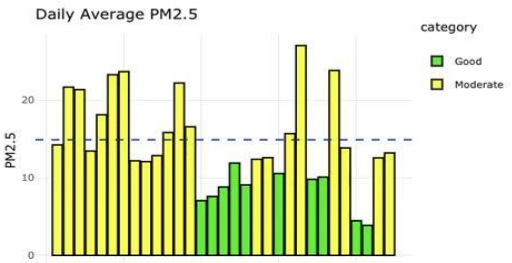
Compare Diurnal Patterns to a Shorter Time Period

Select the date range:

2020-12-01 to 2020-12-05

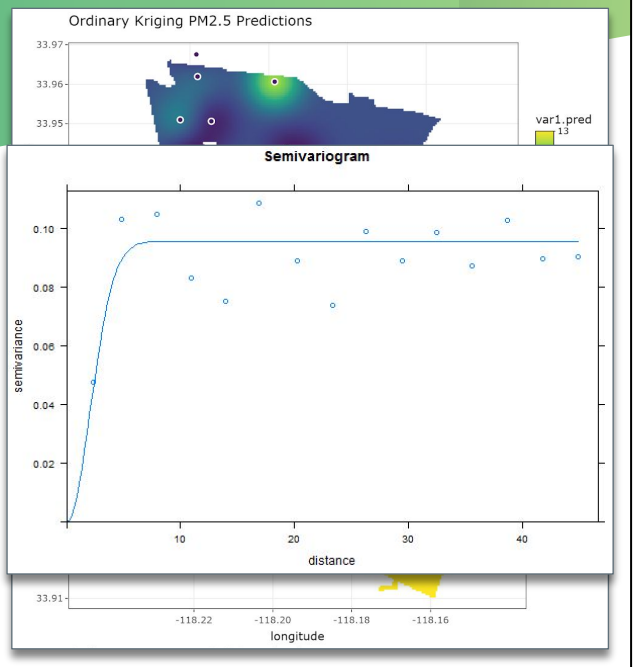
Summary





Optimizing Sensor Placement

- To maximize the usage of our sensor network, we interpolate the data via **kriging**
- The effect is two-fold: prediction and knowledge of where calculations are weakest



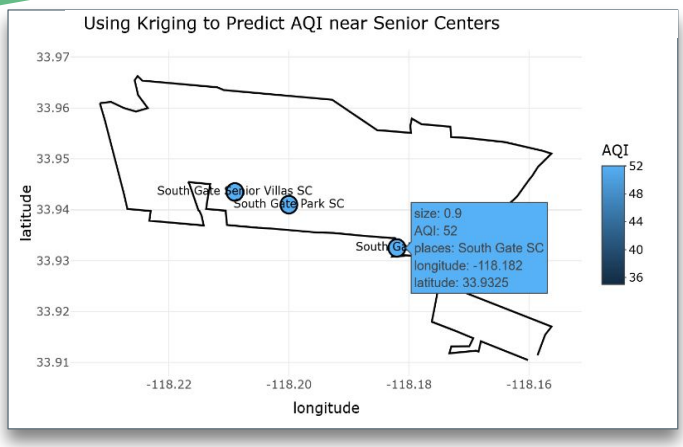
Thank you Hillary,
As Hillary mentioned, the final tab of the analysis application displays the interpolation calculations for our network of sensors for a given timestamp. The interpolation method that we've implemented in our app is known as ordinary kriging. This method of spatial modeling invokes a semivariogram, which is a kind of model that indicates spatial correlation within our data. In short, a semivariogram models the relationship among sensor locations to indicate the variability of the measure in relation to the physical distance of separation.

In our case, **[CLICK TWICE]** it shows the variance in readings in relation to the calculated distance between all pairs of sensor locations. In our app, ordinary kriging takes the semivariogram model and uses it to apply weights to the measured values in the data and estimates values at locations without sensors.

Each set of ordinary kriging calculations, provides ***two*** important pieces of information: predicted PM2.5 values across South Gate, and the variance and/or standard deviation of those values. Thus, CEHAT gets additional, albeit estimated coverage, as well as a measure of where they should deploy more sensors, since better coverage reduces spatial variance.
[NEXT SLIDE]

Sensitive Locations

- Interpolation calculations along with information about sensitive locations better informs CEHAT about areas that need better coverage



The interpolation features of our app provides CEHAT with the ability to determine the typical air quality at sensitive locations, like schools, parks, medical clinics, and senior centers where they may not be allowed to place sensors directly.

Though the version of the app that Hillary demonstrated did not show this, the app will include detailed descriptions on each page. On the interpolation page, these descriptions will explain how CEHAT can use the variance and prediction calculations to either restructure their network around these sensitive locations, or to advocate for more sensors.

I'll now pass it off to Adam on the web development side to talk about how we've structured the website and tackled technical accessibility.
[NEXT SLIDE]

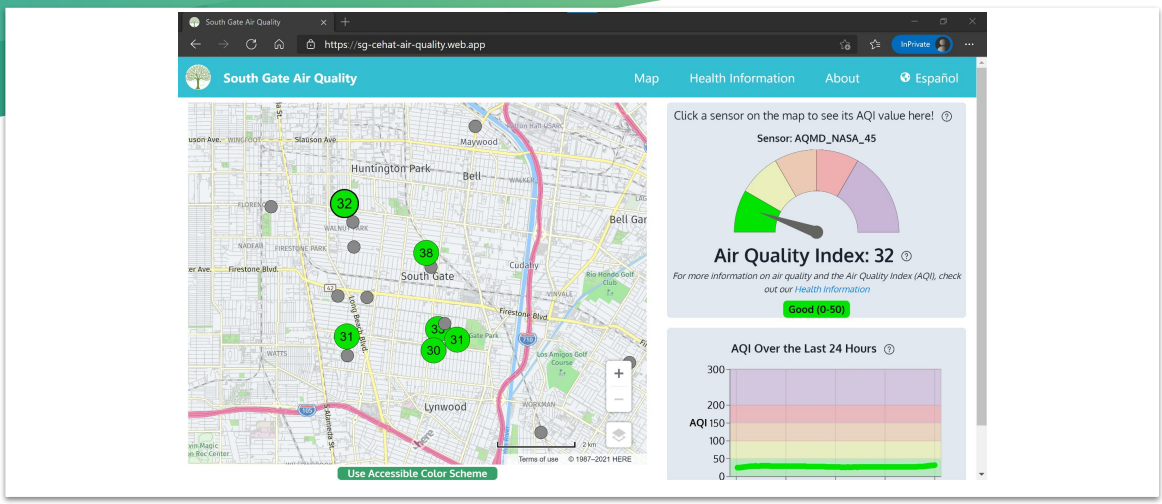
for questions?: It is important to note that our ordinary kriging calculations do not take into account wind patterns, or historical data, as there was no way to automate the influence of this data on our model.

Air Quality Dashboard

https://sg-cehat-qir-quality.web.app/

Community oriented project -> major way community interacts is our website
Visit using the link or QR code visible on screen
Built using React framework
[Just like analysis application, available open source on the Hippocratic License 2.1 for other communities to build off of]

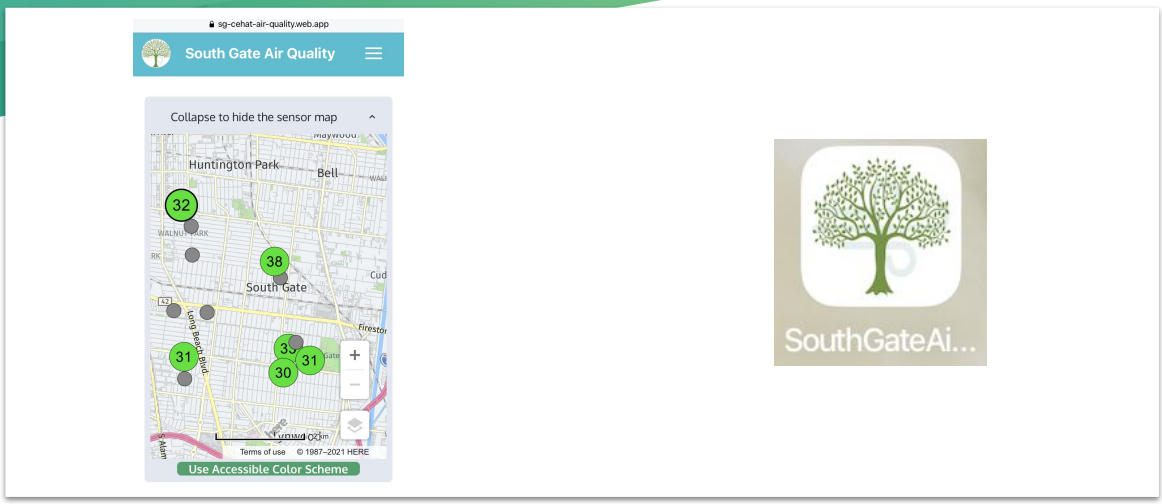
Sensor Data



Biggest feature of the website is communicating the air quality in a clean and user friendly way

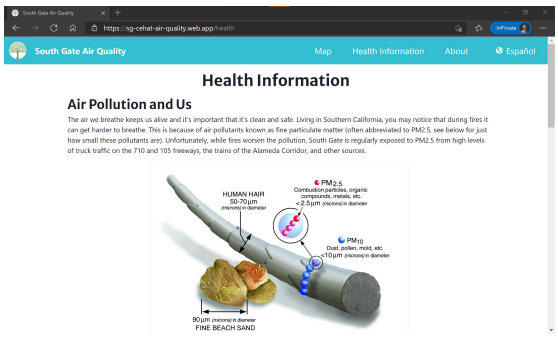
Presented with a map of South Gate annotated with sensors and their readings
Clicking on a sensor puts the current AQI in context using a gauge to provide a sense of scale and
Graph of air quality over the past 24 hours

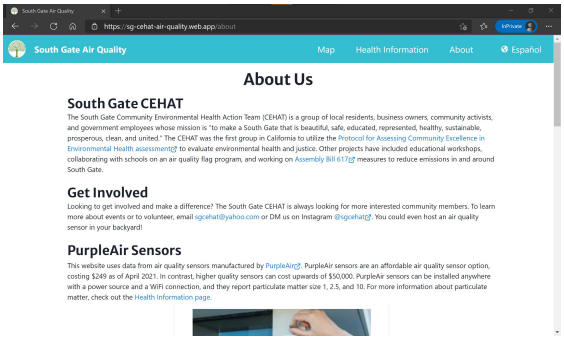
Air Quality On The Go



On the go, working outside -> have air quality info wherever you are
Always thinking about mobile experience
User friendly mobile web experience
Can install as an application (outside of the browser) thanks to the Progressive Web Application framework

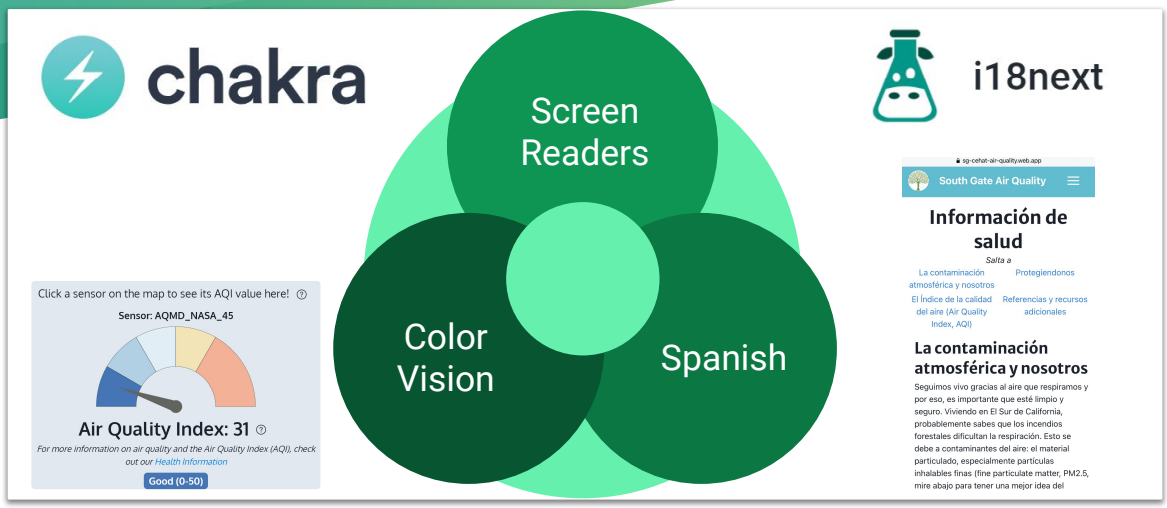
Learn About Air Quality





Air Quality can be a complex topic
For those arent as familiar with, we provide health info
why important, what it means, how to protect
even provide definition of AQI in tooltips throughout the website to make it clear (feedback from user testing)
Learn more about CEHAT get involved

Accessibility

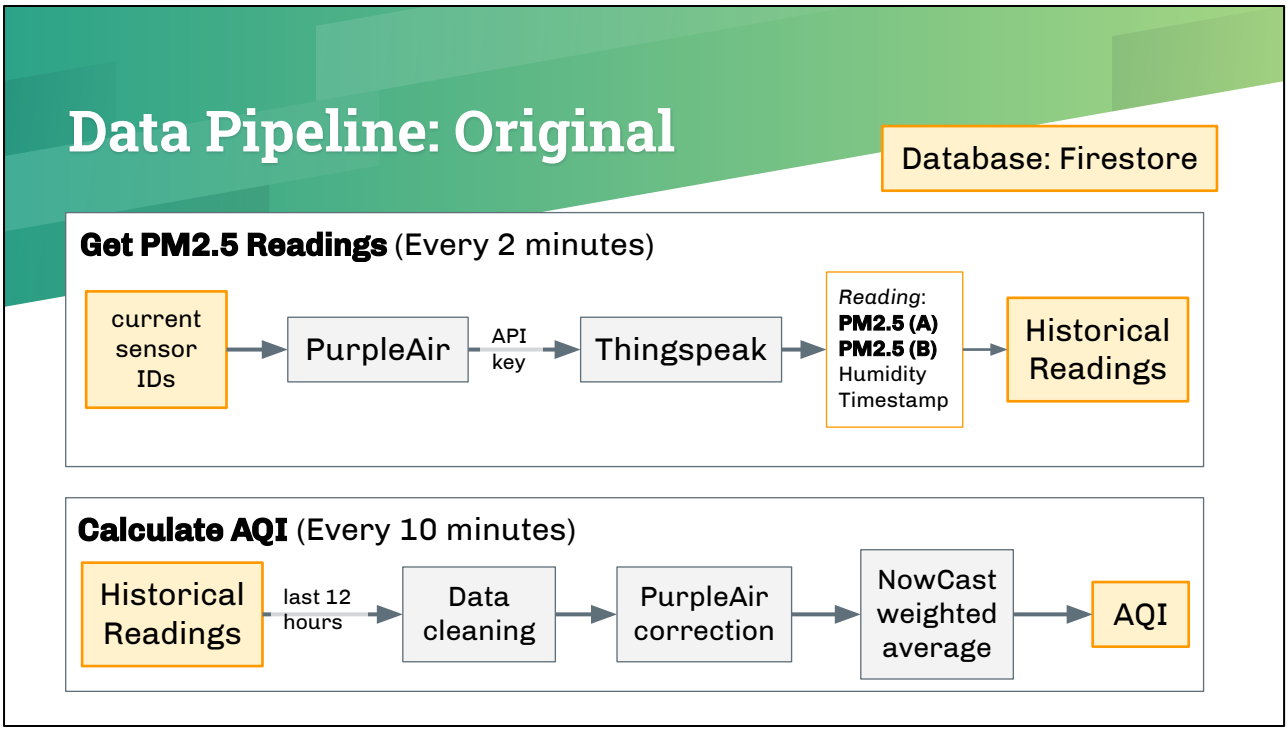


Community oriented project, make sure everyone can use
Broad definition of accessibility

Typical sense, low vision screen readers, accessible UI framework Chakra
Keyboard navigation, alternative text
Significant strides tested using screen readers, unfortunately mapping software isn't super screen reader friendly, room for additional improvement

Color vision difficulties like myself, epa color schemes not friendly (GYORPB), opt into an alternate color scheme more colorblind friendly

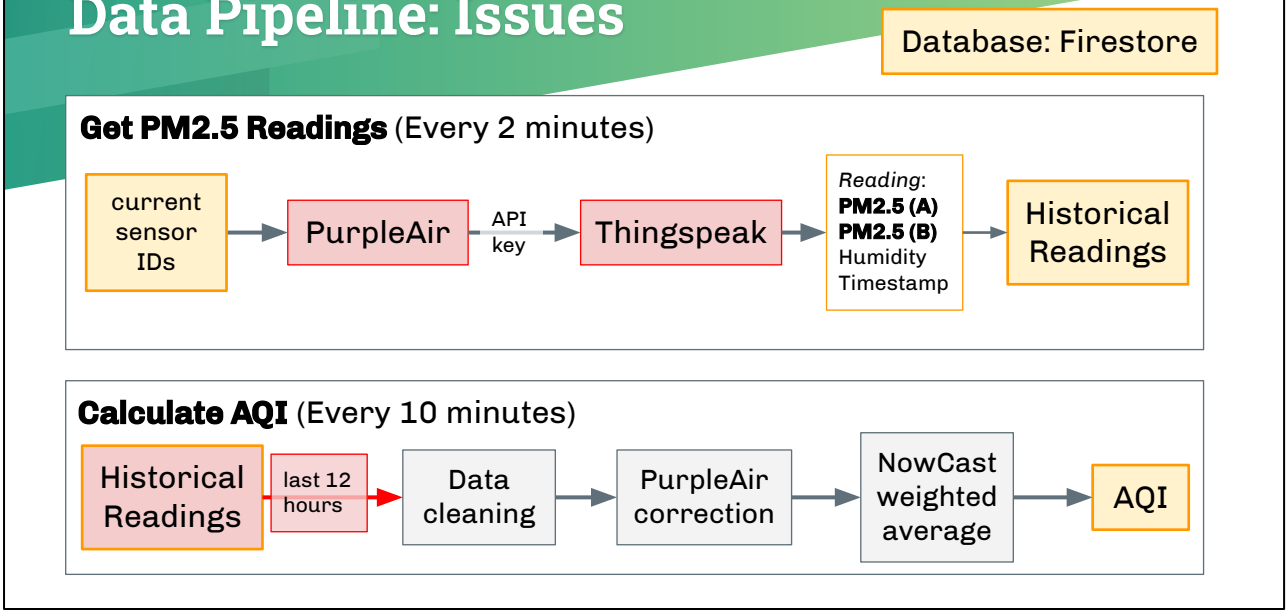
Large spanish monolingual community, using the i18next library to provide spanish version of website
How get translations? Many of us have learned spanish or are bilingual, multi stage review process internally and with liaisons



Thanks Adam! Here's a quick overview of our original data pipeline

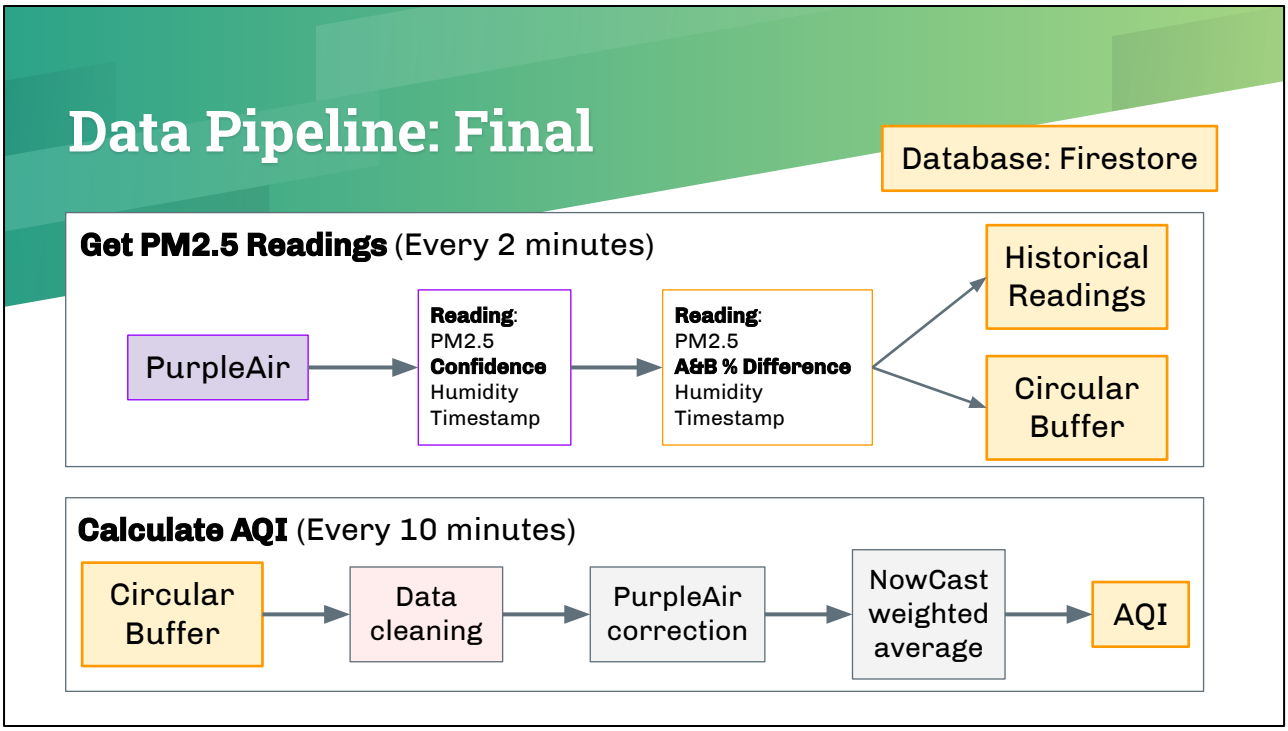
- We got our sensor readings every 2 minutes using two different APIs, and we stored each of these readings in our historical database.
- Then every 10 minutes, we re-calculate the AQI for each sensor, using the last 12 hours of readings.
    - Each PurpleAir sensor has two particulate matter sensors called Channel A and Channel B. When data cleaning, we check that the two PM2.5 readings for each sensor have not diverged, which can indicate sensor malfunction.
    - We then apply the EPA provided correction value for PurpleAir sensors since PurpleAir sensors tend to overestimate PM2.5 values, and the correction factor depends on the current humidity
    - Finally, we calculate an average of the readings that weights the more recent hours heavier before calculating the NowCast AQI
    - This calculated AQI is the number that is displayed on the website

Data Pipeline: Issues

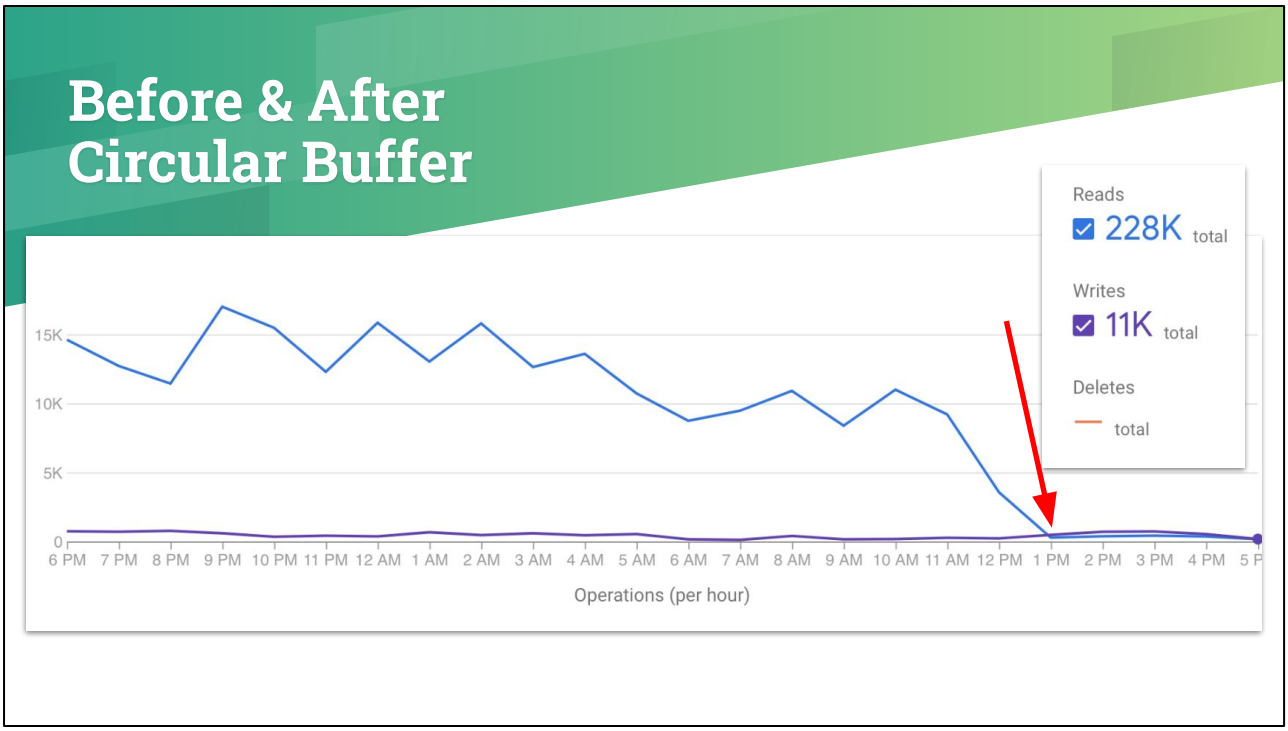


Unfortunately, we ran into 3 problems this spring

- First, PurpleAir deprecated their original API rather suddenly
- Second, we were sending too many requests to Thingspeak, which was causing us to get back errors and miss some sensor readings.
- Finally, our database reads in the calculate AQI step were too high and costly, and since we wanted to make this website as low-cost as possible, this was a problem we needed to address.

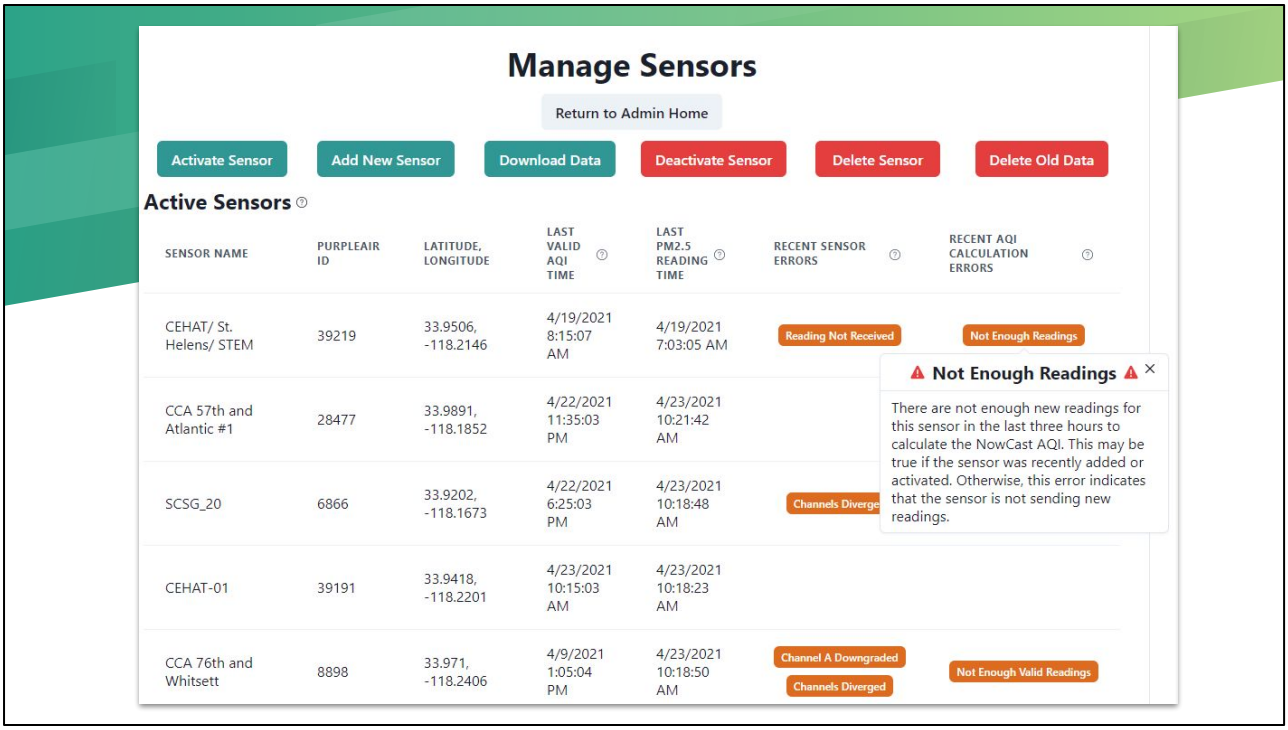


- To address the first two problems, we transitioned to a new PurpleAir API. Overall, this was a positive change that made the entire process more scalable and efficient.
- However, this introduced a new problem—a different data format. Instead of including the readings for Channel A and Chanel B for each sensor, we can only access a “confidence” value that PurpleAir provides but does not document.
    - This was a problem because we used these 2 readings in the data cleaning process.
    - After much correspondence with PurpleAir, we were able to determine how to replicate most but not all of the data cleaning process from before, a concession we had to make
- Finally, we also started using a circular buffer to store the most recent 12 hours of readings. Essentially, this means that every 10 minutes we read the database once per sensor instead of reading 360 times per for each AQI calculation

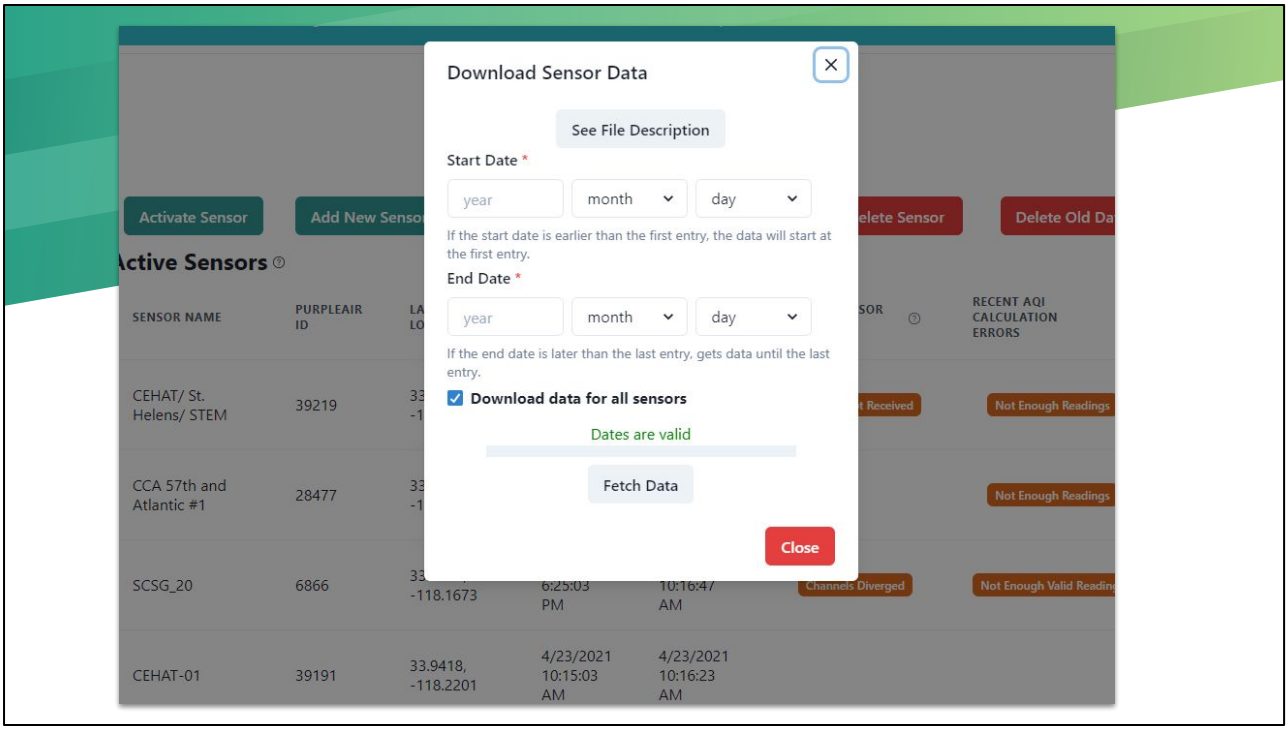


As can be seen from this graph, using a circular buffer was a huge success. Our writes were doubled but our reads were dramatically reduced, which significantly reduced our database costs each month

[15 seconds]



- Finally, I'll show the website admin interface, which allows CEHAT to manage the sensors and the historical readings.
- From this panel, admins can see the current status of the sensors on the website, including the last sensor reading time and any errors that a sensor is having. There are many reasons that a sensor might not be functioning, so this helps admin users determine if they need to reach out to the sensor host to reconnect the sensor to WiFi or if the sensor has malfunctioned and they need to reach out to PurpleAir to replace a part.



- Additionally, from this page admins can download historical data in the format expected by the math analyses discussed by Hillary and Lotenna
- This will allow the CEHAT to continue to run these analyses with new air quality data, which is especially important given the timeline of our project. We have been gathering data from our sensors since October, but air quality is often much poorer in the summer months, which is after our project ends. Since the website will continue to collect sensor data, they will be able to use the tools developed my Hillary and Lotenna to analyze the air quality of the summer months.
- Additionally, through this panel admins can delete sensor data since data storage is not cheap and we're collecting new data every 2 minutes. One of our goals for this website was to make the cost as minimal as possible for the CEHAT, so this function is critical.
- Overall, this admin panel allows the CEHAT to continue this work as our clinic project ends.

Project Extensions

- Features to improve upon the website:
  - Interpolation integration on map
  - Crowd-sourcing
  - Alerting
- Features to improve upon interactive report:
  - Correlations/Comparisons with other data sets
  - Forecasting

Thank you for that Carson,

Though we've completed the main components of this project with all the functionality that we hoped to provide, there are a variety of areas in which this project could have been extended, and more features could have been added had we had more time.

On the web development air quality dashboard side, CEHAT was interested in integrating the interpolation calculations, which were demonstrated in the data analysis app by Hillary, into a map of South Gate, so that they would have real-time access to these calculations.
CEHAT was also interested in a crowd-sourcing feature which would handle complaints/reports from community members about odorous airborne pollutants and public events that affect local air quality, like controlled fires. Additionally, CEHAT wanted the ability to alert the city about extreme air quality, via

texts to residents and banners on the website.

On the data analysis side, CEHAT and our team was interested in being able to forecast future PM2.5 values from the data, identify correlations in the data, and compare the air quality in South Gate to other locations. Of particular interest were correlations between rates of asthma and other respiratory diseases and correlations between PM2.5 and other pollutants like ozone or sulfur oxides. Similarly, comparing the PM2.5 in residential areas to non-residential areas, comparing PM2.5 across census tracts (focusing on income level), and comparing PM2.5 in green spaces versus industrial areas. Handling other sets of data would be key to implementing this in the data analysis app.