

Second Year Paper Summary Statistics

Aaron Watt

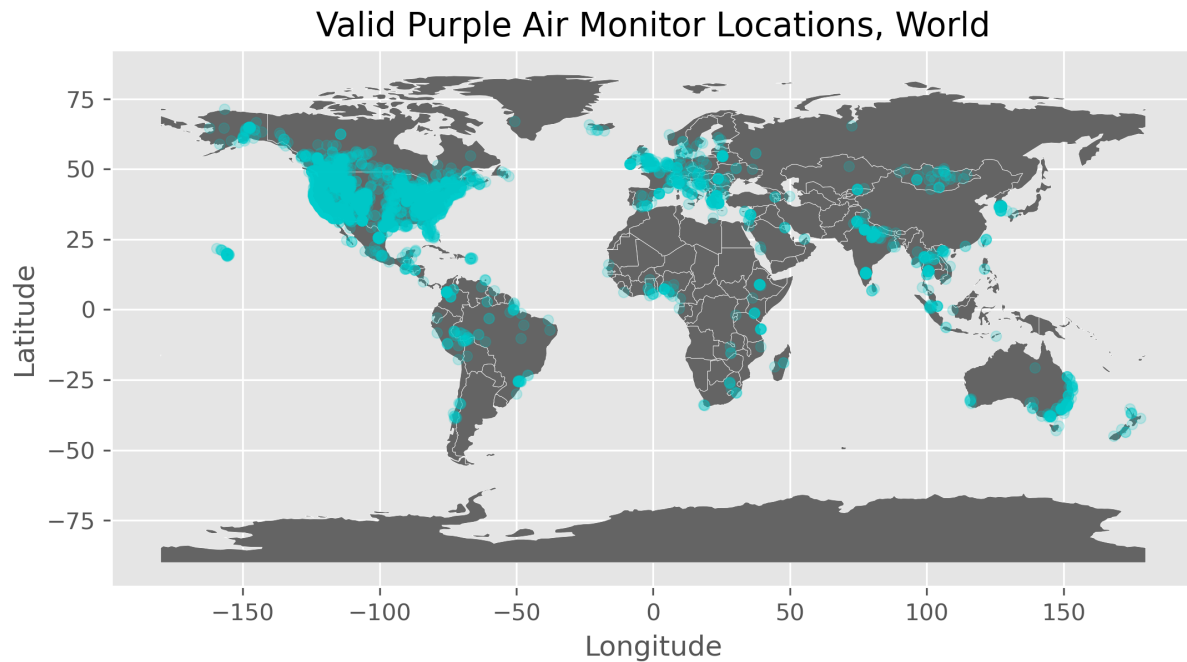
2021-11-08

Contents

Geographical Distribution of Purple Air Sensors	1
Distribution of adoption over time	4
Data plan	5

Geographical Distribution of Purple Air Sensors

We can see from the figure below that the majority of Purple Air sensors are currently concentrated in the United States and Europe.



Note: These location data were collected in 2021, and I am unsure if there is a way to see the historical location of existing sensors or information on sensors that dropped out of the network.

In the US, the sensors are concentrated on the coastlines, with the majority in California.

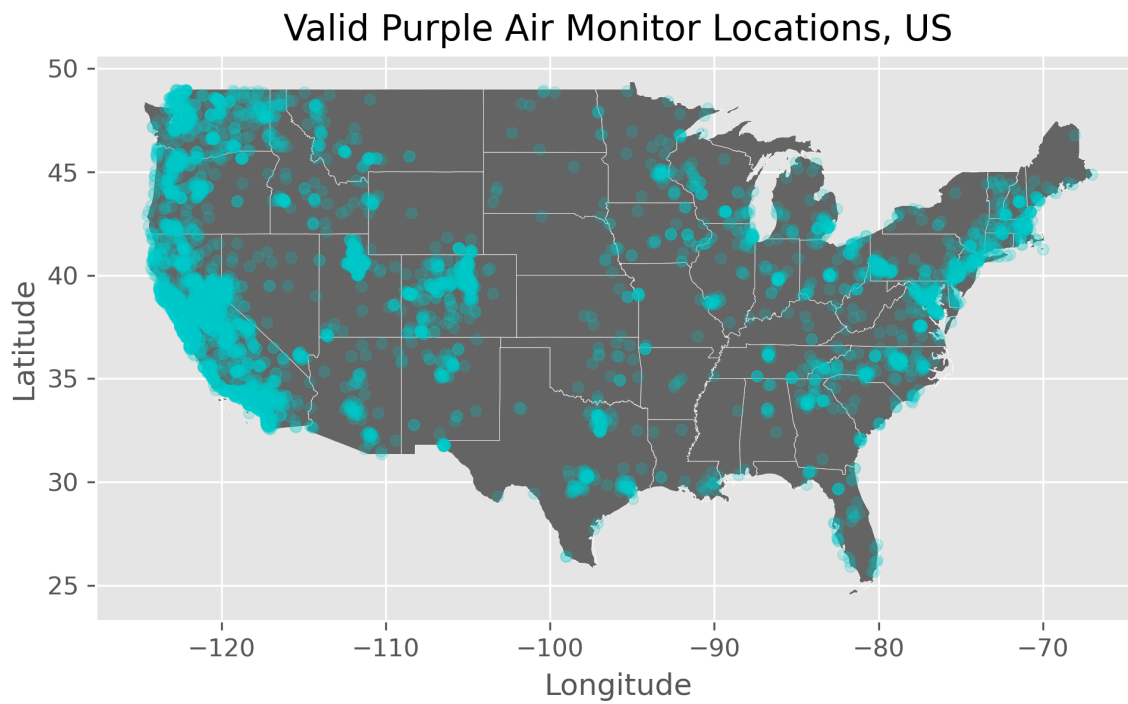


Figure 1: Valid Purple Air Monitor Locations, Contiguous United States

Within California, we can see the sensors are clustered around the Bay area and LA.

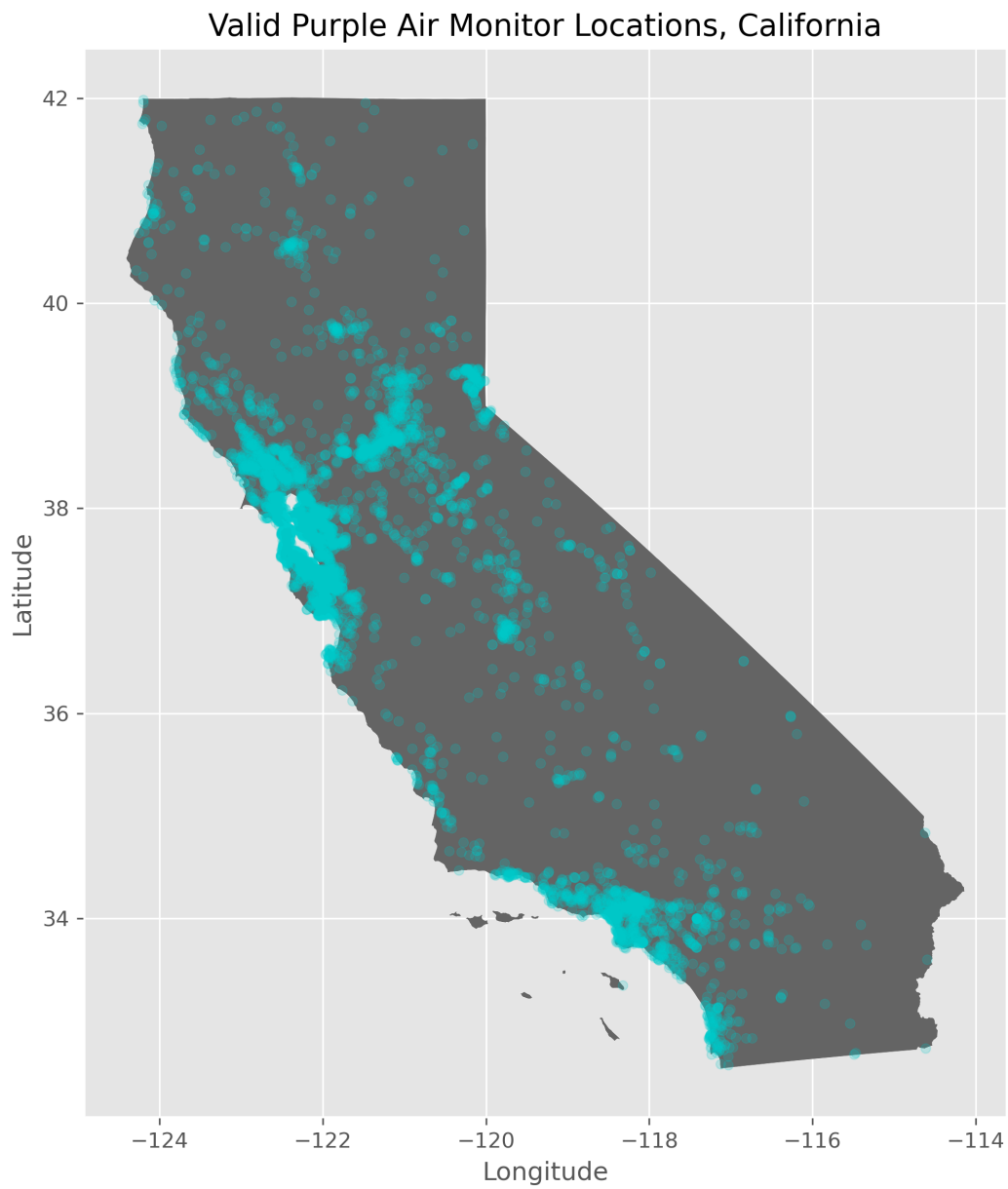


Figure 2: Valid Purple Air Monitor Locations, California

Distribution of adoption over time

For California, we can see from the below graph that adoption of Purple Air pollution monitors (in blue) increased steadily until summer of 2020, when adoption went up dramatically. There were particularly bad smoke days in populated areas of California that summer (e.g., “Mars Day”). However, these data are biased toward recent increases in adoption because I only have access to the sensors that are currently in the network – I have no data on the sensors that dropped out, therefore the increases pre-2020 are likely understated.

A more interesting red graph would probably be population-distance-weighted fires per-month or a population-weighted measure of visibility.

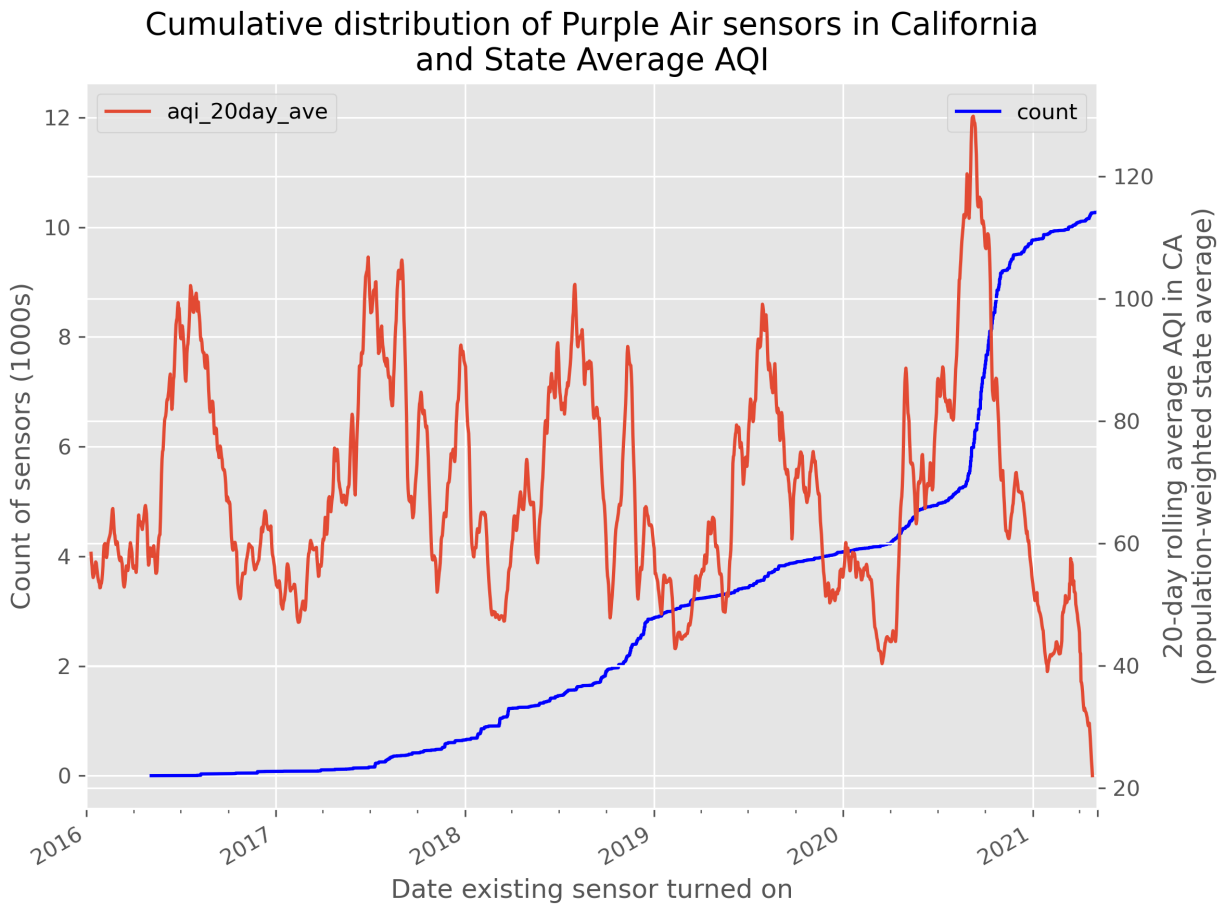


Figure 3: Purple Air Adoption over time vs. Air Quality in California

Data plan

I am planning to limit my second year project to California and focus on comparing Purple Air PM2.5 data to EPA monitor data. I have not yet finished collecting PM2.5 data from Purple Air monitors but plan to collect years 2018-2021 for all California sensors via the Purple Air API. I am also in the process of collecting EPA PM2.5 data via the EPA API.

In order to calibrate a selection model into Purple Air adoption, I plan to collect Census Block Group level demographic data and calibrate the model on the census block group level over 2018-2020 (inclusive). I also plan to collect weather data from PRISM.

To model measurement error of sensors (EPA sensors vs. Purple Air sensors), I plan to use pollution distribution data generated for IPCC climate models. I've been told these data are available to download and are probably the best measure of "ground truth" that I have access to.