Hazards Within Our Homes: An Analysis of Indoor Air Pollution In the Bay Area In Partnership With OCOB

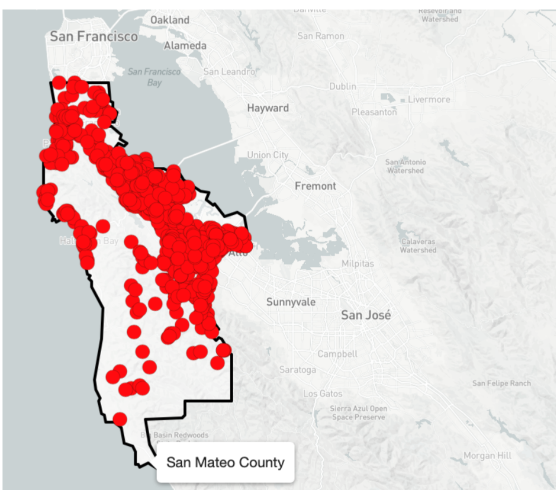
CEE218Z: Shaping The Future of the Bay Area

June 2022

Awoe Mauna-Woanya, Heyaojing Huang, Jinpu Cao, Spencer Zhang, and Isabelle Pilson

Project Description

This spring quarter (2022), the CEE 118/218Z Air Quality team partnered with Our Communities, Our Bay (OCOB) to study the quality and impact of local indoor air. As a whole, OCOB represents a multi-faceted research project that aims to better understand the climate hazards that Bay Area communities face and, in doing so, empower communities to improve their resilience and well-being.[1] As a student group, we focused on the organization's recent pilot study centered around the installation and data collection of indoor air sensors. The initiative uses Purple Air (PA) sensors to measure air quality; PA monitors collect hyper-local air pollution information and – when not restricted by research agreements – share it in a public manner. With 26 sensors currently deployed throughout Belle Haven, North Fair Oaks, and East Palo Alto, the project ultimately intends to expand to approximately 400 sensors in the region, providing new insights on how both indoor and outdoor pollution sources influence the quality of the air in homes. These indoor sensors join an existing fleet of air quality monitors in San Mateo County, California that includes 657 outdoor sensors and 409 indoor ones.



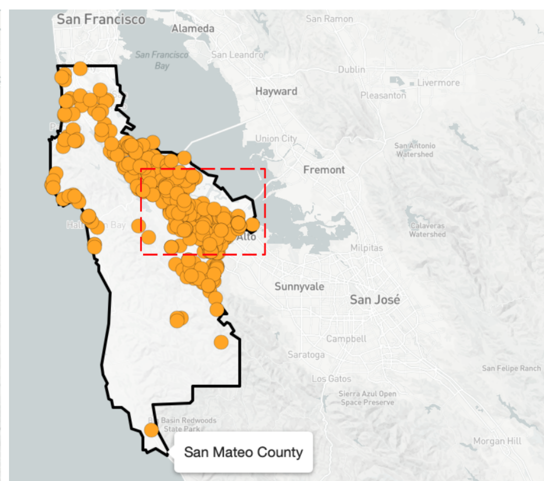


Figure 1 (Left). Map of Existing Outdoor Sensors in San Mateo County (Red)

Figure 2 (Right). Map of Existing Indoor Sensors in San Mateo County (Orange) (With Boundary of New Study in Red)

[1] Our Communities, Our Bay. (2022). What is "Our Communities, Our Bay?" Our Project. https://www.ourcommunitiesourbay.org/project.

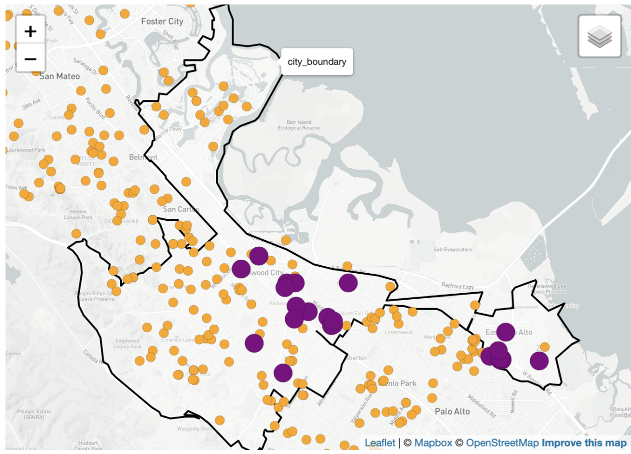


Figure 3. Map of New OCOB Indoor Sensors (Purple) In Relation to Existing Indoor Ones (Orange)

Over the course of ten weeks, the CEE team explored a wide range of topics – from the connectivity and distribution of the new sensors to the relationship between indoor and outdoor air quality, the impact of highway proximity and rush hour times on outdoor air quality, the power of home insulation in mitigating the infiltration of poor outdoor air, the relationship between a building's characteristics and its indoor air quality, and the design of end-of-month homeowner report.

Key Takeaways

At a high level, **our work demonstrates the importance of analyzing and improving indoor air quality, as indoor air quality is worse – and thus more dangerous – than outdoor**. Further, as our main findings and deliverables, we found there is a high relationship between indoor and outdoor air quality and identified a distinct 40-50 minute lag period between spikes in outdoor air pollution and subsequent changes in indoor air pollution levels; gained clarity on how highway traffic affects air quality; and incorporated community feedback to create a straightforward and comprehensive air quality report for homeowners. In this paper, we will detail our methods, findings, and recommendations for areas of future research in hopes of informing community leaders and members of the critical issue of indoor air quality and inspiring attention and effort around the matter.

Background Information

As contextual information, AQI stands for Air Quality Index, measuring the level of air pollution and categorizing the degree of health concern. According to the Environmental Protection Agency, AQI takes into account five major air pollutants regulated by the Clean Air Act – ground-level ozone, particulate matter (PM2.5 and PM10), carbon monoxide, sulfur dioxide, and nitrogen dioxide.[2] Generally-speaking, an AQI value below 100 indicates that the air is relatively safe; as AQI climbs above 100, air quality becomes increasingly more dangerous – first for sensitive populations and then for all individuals. More specifically, Air Quality Index is divided into six categories ranging from Healthy to Hazardous. Scales differ slightly; for our work, we followed the EPA's standards:

AQI Basics for Ozone and Particle Pollution

Daily AQI Color	Levels of Concern	Values of Index	Description of Air Quality
Green	Good	0 to 50	Air quality is satisfactory, and air pollution poses little or no risk.
Yellow	Moderate	51 to 100	Air quality is acceptable. However, there may be a risk for some people, particularly those who are unusually sensitive to air pollution.
Orange	Unhealthy for Sensitive Groups	101 to 150	Members of sensitive groups may experience health effects. The general public is less likely to be affected.
Red	Unhealthy	151 to 200	Some members of the general public may experience health effects; members of sensitive groups may experience more serious health effects.
Purple	Very Unhealthy	201 to 300	Health alert: The risk of health effects is increased for everyone.
Maroon	Hazardous	301 and higher	Health warning of emergency conditions: everyone is more likely to be affected.

Figure 4. EPA AQI Standards and Categorization (EPA)

Research on air pollution and health impacts shows that high exposure to poor air quality is associated with "cancer, cardiovascular disease, respiratory diseases, diabetes mellitus, obesity, and reproductive, neurological, and immune system disorders."[3] In addition, the risk is most severe for those living in urban areas, those exercising in high AQI conditions, children, pregnant women, the elderly, and individuals suffering from respiratory and cardiovascular conditions[4]. AQI can quantify both outdoor and indoor air quality, and sources of air pollution exist both inside and outside of our homes. Common outdoor sources of air pollution include

---

[2] AirNow.Gov. (2022). AQI Basics for Ozone and Particle Pollution. Air Quality Index (AQI) Basics. https://www.airnow.gov/aqi/aqi-basics/

[3] National Institute of Environmental Health Sciences. (2022). Air Pollution and Your Health. https://www.niehs.nih.gov/health/topics/agents/air-pollution/index.cfm#:~:text=Air%20pollution%20can%20affect%20lung,are%20linked%20to%20chronic%20bronchitis.

[4] *Ibid*.

weather events (such as wildfires), traffic, and emissions from industry.[5] On the other hand, common indoor sources of air pollution are smoking and vaping, cooking, cleaning, and home heating.[6,7] On average, Americans spend approximately 90% of their time indoors, illustrating the importance of indoor air quality as well as its potential impact on human well-being.[8] Both indoor and outdoor sources contribute to indoor AQI, as outdoor air can infiltrate into buildings through open windows and doors as well as smaller crevices. Studies suggest that – in California – people of color and of low socioeconomic status are exposed to worse air quality, and that air quality sensors are disproportionately found in affluent neighborhoods, leading to gaps in data and information.[9] Additionally, conversations with stakeholders reveal that the prevalence of cardiovascular disease in East Palo Alto and similar cities is disproportionately high; to date, researchers have focused their attention on the impact of outdoor AQI.[10] Taking on a novel perspective and turning from the problem to viable solutions, policy as well as household practices can improve *indoor* air quality, in turn mitigating its negative consequences on human health.[11] Given the significance of indoor air quality, relevant environmental justice concerns, lack of previous research, and the opportunity to enhance well-being here, we focused our work on the sources, characteristics, mechanisms, and communication of indoor AQI in San Mateo County.

Our Analysis

Below, we outline four analyses that show our analysis methodology as well as the potential insights we can gain from the project's sensors in San Mateo County. These include an

---

[5] Government of New South Wales. (2021). Outdoor Air Pollution. Air Quality. https://www.health.nsw.gov.au/environment/air/Pages/outdoor-air-pollution.aspx

[6] Government of New South Wales. (2021). Indoor Air Pollution. Air Quality. https://www.health.nsw.gov.au/environment/air/Pages/indoor-air-pollution.aspx

[7] Liang, Y., et al. (2021). Wildfire smoke impacts on indoor air quality assessed using crowdsourced data in California. PNAS, 118(36). https://www.pnas.org/doi/10.1073/pnas.2106478118

[8] Roberts, T. (2016). We spend 90% of our time indoors, says who? Building Green. https://www.buildinggreen.com/blog/we-spend-90-our-time-indoors-says-who

[9] Boyd-Barrett, C. (2019). People of Color and the Poor Disproportionately Exposed to Air Pollution, Study Finds. California Health Report. https://www.calhealthreport.org/2019/02/08/people-of-color-and-the-poor-disproportionately-exposed-to-air-pollution-study-finds/

[10] Personal Communication with Sustainability Officer of East Palo Alto, May 2022.

[11] EPA. (2022). Improving Indoor Air Quality. Indoor Air Quality. https://www.epa.gov/indoor-air-quality-iaq/improving-indoor-air-quality

indoor and outdoor air quality comparison, a highway analysis to determine the impact proximity to highways impacts local air pollution, a time-series analysis to identify key usage patterns in homes, and an investigation of the relationship between AQI and socio-demographic characteristics of homeowners. We end on a discussion with a presentation of our draft of an end-of-month summary report for OCOB homeowners that aims to inform and empower the local community.

The PurpleAir (PA) sensors report one PM2.5 measurement every ten minutes. In most air quality evaluation studies, raw PM2.5 is converted to air quality index (AQI). AQI is further categorized into six levels as described by the EPA ("Good", "Moderate", "Moderately Unhealthy/Unhealthy for Sensitive groups", "Unhealthy", "Very Unhealthy", and "Hazardous")[12] for intuitive understanding when communicating air quality situations to the public. Therefore, all PM2.5 measurements reported by PA sensors in your analysis are converted to AQI for discussion in the present report as well. Importantly, for the analysis, we removed data based on sensor connectivity and performance to ensure that only highly-connected and properly functioning sensors (with realistic data for AQI and other values) were included and studied.

Indoor vs. Outdoor Air Quality

Throughout our research, we analyze the 26 indoor Purple Air sensors recently installed in the neighborhoods of East Palo Alto, Belle Haven, and North Fair Oaks – an area historically lacking air quality sensor coverage before the new study (Figure 3). In doing so, we looked at indoor air quality data in this neighborhood that was never reported elsewhere and deeply inspected the reported AQI values for each individual indoor sensor for the month of April, 2022.

General Comparison

As discussed earlier in this report, indoor air quality can be associated with adjacent ambient outdoor air as well as specific indoor activities. To represent "adjacent" ambient outdoor air relevant to our participant sensors, we filtered outdoor sensors to those within 0.5 mile of each indoor sensor and averaged their reported AQI (Figure 5).

[12] EPA (2022). Technical Assistance Document for the Reporting of Daily Air Quality the Air Quality Index (AQI). https://www.airnow.gov/sites/default/files/2020-05/aqi-technical-assistance-document-sept2018.pdf

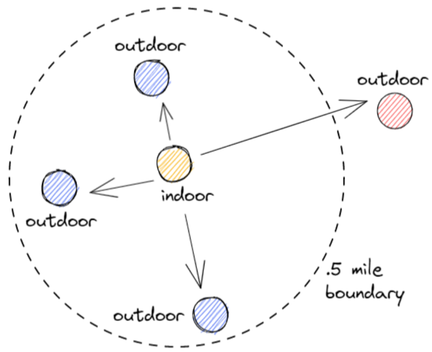


Figure 5. Schematic of Technique Used to Define Relevant Outdoor Sensors In Relation to Indoor Sensor

The PA sensors report one data point every 10 minutes. The comparison was conducted for various time granularities – namely, 10 minute increments, hourly averages, and daily averages. Figure 6 shows a representative daily comparison result of indoor and outdoor AQI for the 26 sensors.

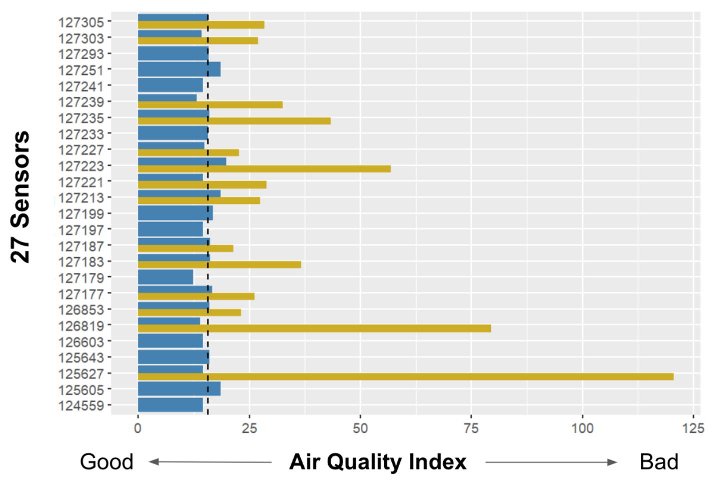


Figure 6. Comparison of Average Indoor (Blue) and Outdoor (Green) AQI by Sensor, April 2022

Surprisingly and significantly, the indoor average AQI values reported by the 26 sensors are consistently higher than its adjacent ambient average outdoor AQI. This indicates the presence and influence of indoor air pollution sources.

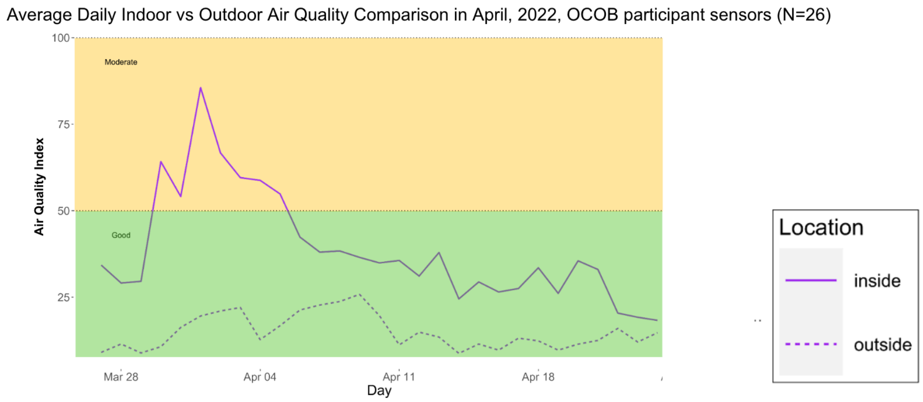


Figure 7. Comparison of Indoor and Outdoor AQI Daily Average Values from All 26 Sensors, Mid-March to End of April

While the focus in the Bay Area has traditionally been on outdoor air pollution, its sources, and its relation to cardiovascular health concerns, our analysis suggests that attention may want to be shifted toward indoor AQI, instead.

Individual Household Analysis

We also compared each individual participant sensor's reported AQI to those of similar households. Similar households are defined as the average performance of the 26 participant sensors at the given timestamp. The indoor air quality presents high variation among different households. For instance, the Household ID 126603 reports "Hazardous" indoor air conditions over 60% of the time (Figure 8), six times higher than the average household (Figure 8). Differences in indoor air quality could be the result of differences in household activities, infiltration of outdoor air, and/or sensor miscalibration/malfunction.

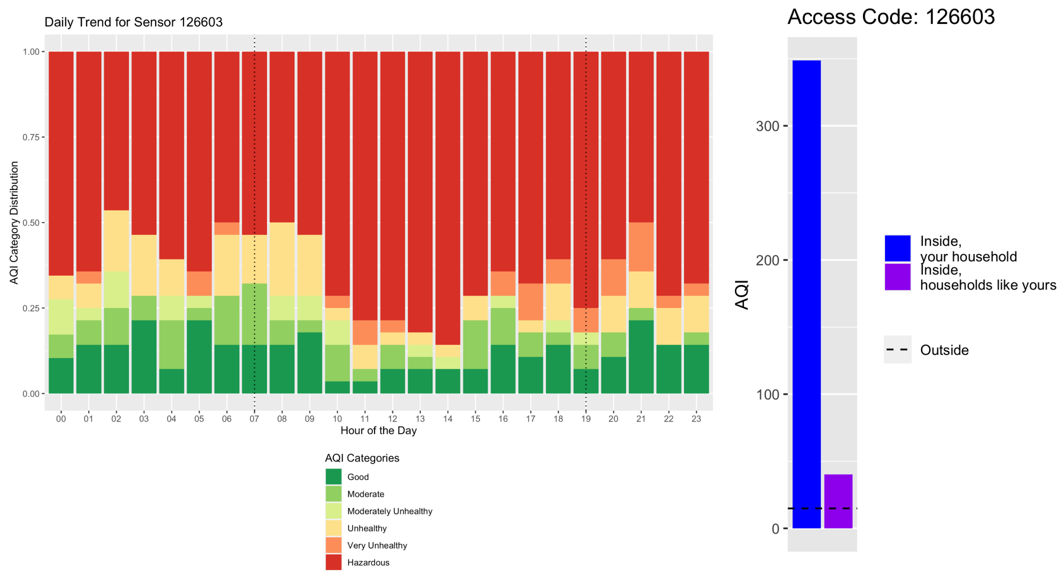


Figure 8. Average hourly AQI categories for April 2022 for Household 126603 (left). Daily indoor air quality comparison between 126603 and similar households in OCOB community on one unspecified day in April, 2022; shown in blue is the indoor AQI reported by 126603, shown in purple is the similar households average (right).

Year-round Comparison

As introduced by the General Comparison of indoor and outdoor air quality, the finding that indoor air quality is worse than outdoor conditions in the OCOB community is alarming and deserves attention and further investigation. For years, the air quality effort in the Bay Area has been focused on wildfire and related outdoor pollution sources. However, we may be getting the effort on the wrong cause if the results reported in OCOB participant sensors are universal across the bay. The general population may be spending their daily life in “unhealthy” air, when the air quality efforts are soaring high about “sealing your windows,” “insulating your house” against very occasional wildfires, when indoor pollution sources are just a finger away.

Therefore, we zoomed out a bit, choosing to focus on the entire community of San Mateo County (SMC) to understand if it still holds true that the general indoor air quality is worse than adjacent outdoor air quality on a bigger scale. Additionally, we investigated how the level of indoor pollution compares to the level of outdoor pollution when wildfire events occur. As mentioned earlier, SMC holds 657 outdoor PA sensors and 407 indoor sensors. We randomly selected a hundred of each category to represent the general conditions in SMC. We investigated the time period of July 16, 2020 to October 1, 2020, with the CZU Lightning Complex fires event from August 16 to September 22, the biggest fire event SMC has seen since 2000.[13] We separated the time frame into "wildfire season" and "non-wildfire season," to compare with the OCOB results. The reported PM2.5 from all sensors are once again converted to AQI, and grouped into AQI categories. The results are shown in Figure 9.

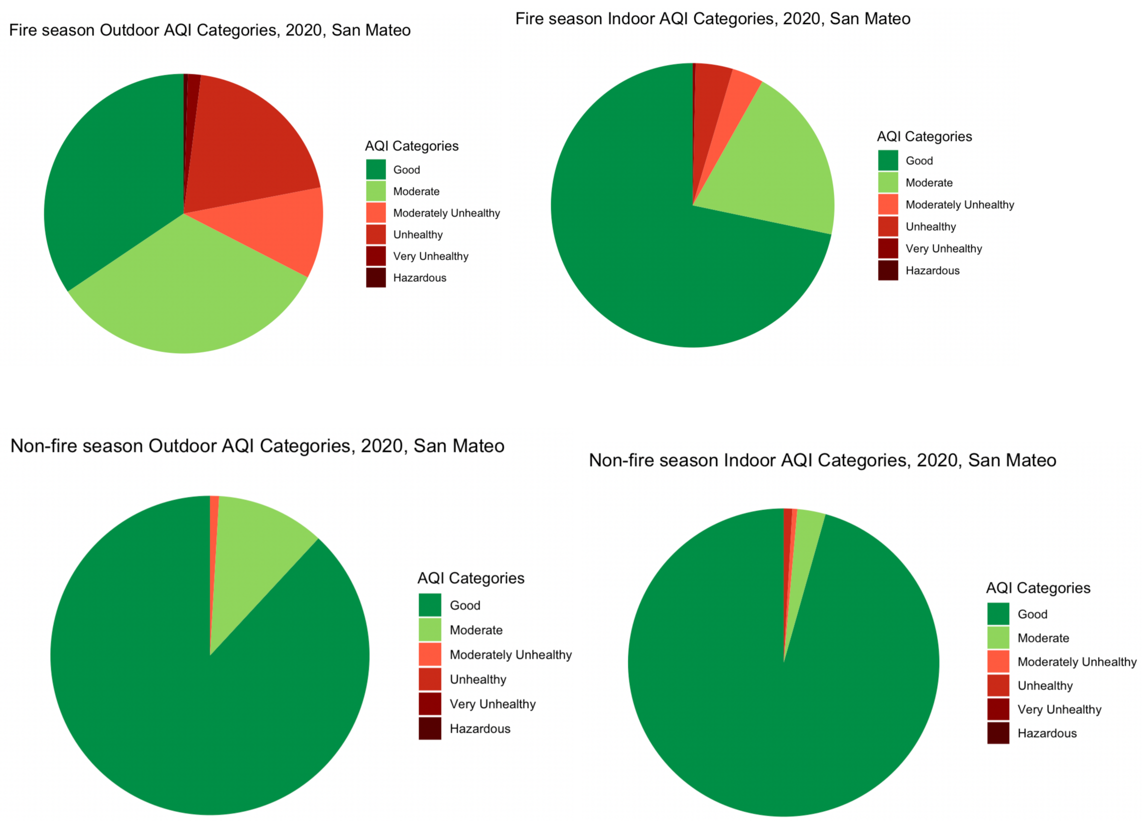


Figure 9. Average SMC indoor and outdoor air quality performance during fire season in 2020. Upper left: SMC outdoor AQI performance during fire season; Upper right: SMC indoor performance during fire season; Lower left: SMC outdoor AQI performance during non-fire season; Lower right: SMC indoor AQI performance during non-fire season. Note the indoor performance generally corresponds to the amount of outdoor pollution in the same area.

Surprisingly and fortunately, the results shown in SMC refute the hypothesis that indoor air quality is consistently worse than adjacent outdoor conditions in the community, disagreeing with the results in OCOB participants. In both fire and non-fire seasons, the indoor air quality reported in SMC is at least represented by its outdoor conditions, and most of the time lower than the amount of outdoor pollution. Such results lead us to a more unexpected hypothesis: the

---

[13] CA.Gov, (2022). Cal Fire. https://www.fire.ca.gov/incidents/

OCOB community stands in a unique position of vulnerability to indoor air pollution with distinct indoor air pollution sources that are contributing to severe indoor air pollution. However, this hazard has been overlooked due to historical lack of data, i.e. low sensor installation coverage here compared to other communities in the Bay.

The amount of *daily* indoor pollution the OCOB community faces is equivalent to the average SMC community *during severe fire conditions*, shown in Figures 9 and 10. The amount of fire pollution needed to cause the level of indoor pollution shown in OCOB is shown in Figures 9 and 10. However, the OCOB outdoor condition is generally very good, Figure 10. This indicates severe indoor air pollution sources, which may be cooking facilities, smoking, vaping, and etc.. Further analysis is needed to pinpoint the indoor sources of such pollution. Given the high amount of reported respiratory disease cases in the community, we believe further efforts surrounding the indoor pollution sources – both research-based and homeowner behavior-focused – might help mitigate the public health issue that is specific to this underrepresented area.

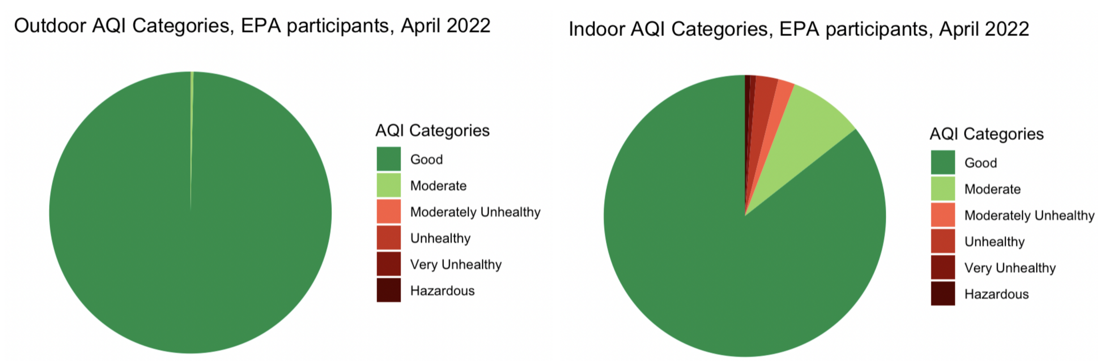


Figure 10. Average outdoor(left) and indoor(right) AQI performance in OCOB community for the month of April, 2022. Note the amount of indoor pollution is equivalent to Figure 10-1-Upper right, however the amount of outdoor pollution is significantly lower compared to Figure 10-1-upper left.

Indoor and Outdoor AQ Cross-Correlation

Many studies indicate that indoor air quality is influenced by outdoor air (Yocom et al., 1971[14]; Poupard et al., 2005[15]; Kuo and Shen, 2010[16]; Meadow et al., 2014[17]). In the section, the quantitative relationship between indoor air quality and outdoor air quality will be discussed through correlationship analysis and the Granger causality test.

Cross-correlation is a method by which the degree of similarity between two sets of numbers can be quantified.[18] Given two time-series x_t and y_t, we can delay x_t by T samples. The *cross-correlation* function between the pair of series is defined as follow:

$$r_{xy}(T) = \frac{\sigma_{xy}(T)}{\sigma_x \sigma_y}$$

Here, σ_x and σ_y are the variances of each series and the function $\sigma_{xy}(T)$ is the *cross-covariance* function, which is defined as follow:

$$\sigma_{xy}(T) = \frac{1}{N-1}\sum_{t=1}^{N}(x_{t-T} - \mu_x)(y_t - \mu_y)$$

In the equation above, μ_x and μ_y are the means of each time series, and there are N samples in each.

Before the cross-correlation function is computed, the indoor AQI induced by indoor sources need to be processed. If the indoor AQI at some point is greater than the maximum of its ambient outdoor AQI in the past 6 hours, we regarded it as an outlier since, the indoor sources play a considerable role at this point and the impact of outdoor AQI cannot single-handedly explain this abnormally high value. The outlier is replaced by the maximum of its ambient outdoor AQI in the past 6 hours since cross-correlation function would make more sense if the

---

[14] Yocom, J. E., Clink, W. L., & Cote, W. A. (1971). Air Quality Relationships. Journal of the Air Pollution Control Association, 21(5), 251-259. https://www.tandfonline.com/doi/abs/10.1080/00022470.1971.10469525

[15] Poupard, O., Blondeau, P., Iordache, V., & Allard, F. (2005). Statistical analysis of parameters influencing the relationship between outdoor and indoor air quality in schools. Atmospheric Environment, 39(11), 2071-2080. https://www.sciencedirect.com/science/article/pii/S1352231005000166

[16] Kuo, H. W., & Shen, H. Y. (2010). Indoor and outdoor PM2. 5 and PM10 concentrations in the air during a dust storm. *Building and Environment*, *45*(3), 610-614. https://www.sciencedirect.com/science/article/pii/S0360132309001966

[17] Meadow, J. F., Altrichter, A. E., Kembel, S. W., Kline, J., Mhuireach, G., Moriyama, M., ... & Bohannan, B. J. (2014). Indoor airborne bacterial communities are influenced by ventilation, occupancy, and outdoor air source. *Indoor air*, *24*(1), 41-48. https://onlinelibrary.wiley.com/doi/full/10.1111/ina.12047

[18] Derrick, T., & Thomas, J. (2004). Time series analysis: the cross-correlation function. https://dr.lib.iastate.edu/entities/publication/7a1756d8-b106-4ad8-b0ac-d6cb1fb76bc0

time-series is continuous. It should be pointed out that the replacement is not an exact approximation of outdoor AQ's contribution to indoor AQ. In further research, splitting the indoor AQI series into pieces by these outliers and computing cross-correlation function in each piece may lead to a more robust analysis and more accurate or nuanced findings. Figure 11 shows the comparison of AQI before and after processing indoor AQI outliers for Participant 126853. From the boxplot in the figure, we can see that the range of processed indoor AQI lies within the range of outdoor AQI, which means the outdoor AQ becomes the main source for indoor AQ.

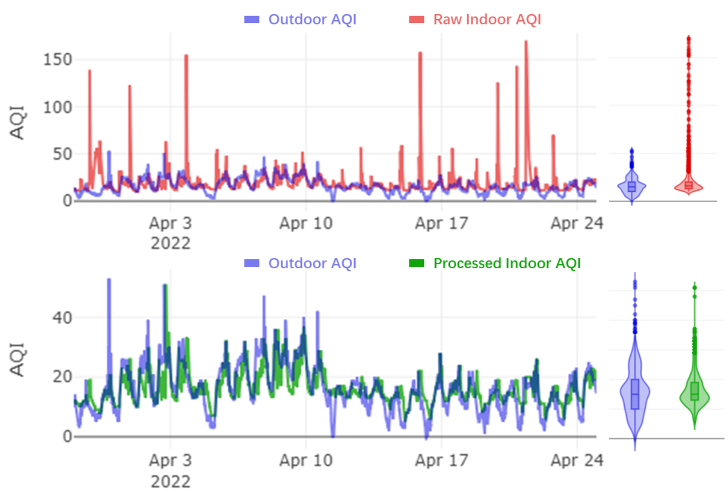


Figure 11. Comparison of AQI before and after Processing Indoor AQI outliers for Participant 126853

Figure 12 shows cross-correlation function curves of indoor and outdoor AQI for 10 participants whose data is relatively complete. It can be seen that the trends of these curves are similar except for the participant 127305. The cross-correlation will be relatively high during the first one hour and will start descending after that. The correlation between indoor and outdoor AQ reaches its bottom when there is a 15-hour time lag between them. The correlation begins to increase when the time lag is greater than 15 hours.

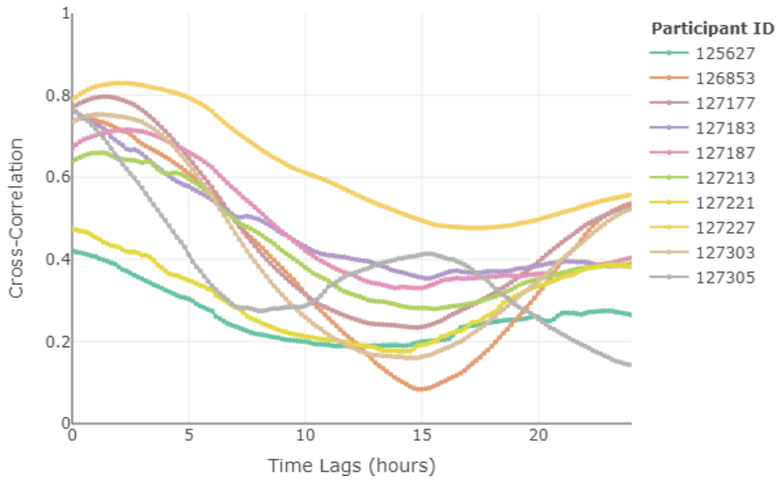


Figure 12. Cross-Correlation Curves of Indoor and Outdoor AQI for 10 Participants

Figure 13 presents the maximums of indoor and outdoor AQI cross-correlation function and the corresponding time lags for 10 participants. From the figure we can find that most correlations are greater than 0.6 and almost all the time lags are within 2 hours. Thus, from the perspective of cross-correlation function, we can have a preliminary conclusion that there is relatively high correlation between indoor and outdoor AQ, especially when the time lag between them is no more than 2 hours.

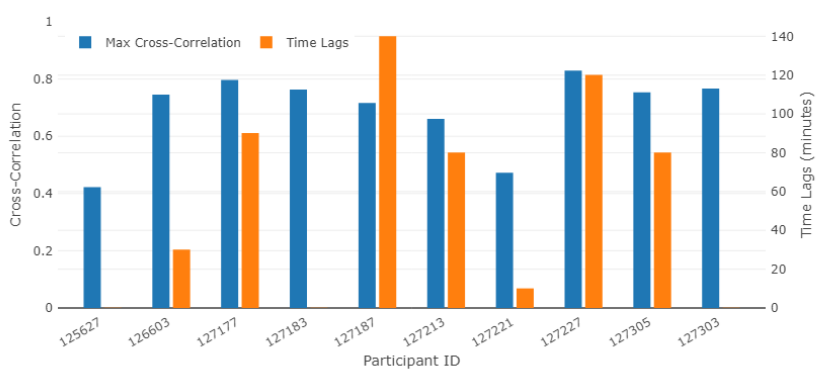


Figure 13. The Maximum of Indoor and Outdoor AQI Cross - Correlation Function and the Corresponding Time lags for 10 Participants

Granger Causality Test for Indoor and Outdoor AQ

After noting the high relationship between indoor and outdoor AQ by conducting cross-correlation analysis, it can be inferred that knowing the past outdoor AQ seems helpful for forecasting the future indoor AQ. In the section, a statistical hypothesis test - Granger causality test is used to test this conjecture. The Granger causality test is used for determining whether one time series is useful in forecasting another, first proposed in 1969[19]. When time series X Granger-causes time series Y, as the Figure 14. shown, the patterns in X are approximately repeated in Y after some time lag. Thus, past values of X can be used for the prediction of future values of Y.

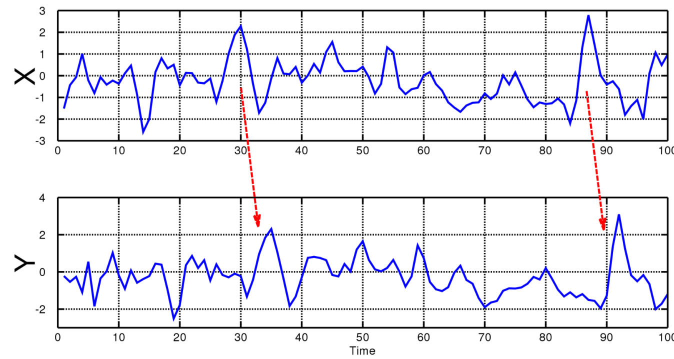


Figure 14. Granger Causality Test Example

We assume indoor and outdoor AQI are both stationary time series and conduct Granger Causality test for 11 participants indoor and outdoor AQI. The results show that past outdoor AQI provides statistically significant information about the future indoor AQ. Figure 15 shows the optimal time lags for forecasting future indoor AQI. From the perspective of the Granger Causality test, in general, the past 6 hours' outdoor AQI are useful for forecasting future indoor AQI.

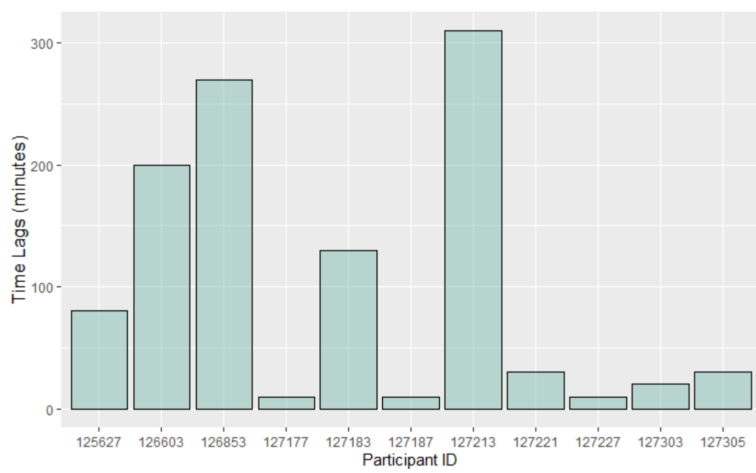


Figure 15. Optimal Time Lags for Forecasting Future Indoor AQQ

In addition, a placebo test is conducted in order to make the results of Granger Causality test more persuasive. Take Participant 127305 as an example. From the Granger Causality test

[19] Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica: journal of the Econometric Society, 424-438. https://www.jstor.org/stable/1912791?seq=1

we know that the past half hour's outdoor AQI will provide statistically significant information about the future indoor AQI. In the placebo test, the outdoor AQI is shifted forward one day. We use yesterday's corresponding past half hour's outdoor AQI as Granger-cause time-series and conduct the similar Granger Causality test. As we expected, the results show that in this case, outdoor AQ does not provide statistically significant information about the future indoor AQI.

Air Quality Diffusion

In the General Comparison section, it is shown that the indoor air quality is worse than outdoor air quality. As described in our Background Information, indoor activities such as cooking, smoking and cleaning are the main within-home causes of poor indoor air quality. However, Jones et al. (2000) conducted a chemical composition analysis and found that fine lead and sulfate particles, which are part of PM2.5, penetrate into the indoor environment from the outdoors.[20] This conclusion highlights that outdoor air pollutants must also be considered when analyzing indoor air environments. Johnson et al. (2004) found that air-exchange rate (AER), a key factor that reflects how fast outdoor sources influence indoor air quality, is related to outdoor and indoor pollutant concentration, ventilation conditions, wind speed, temperature and relative humidity.[21] It is difficult to measure how each factor individually influences the penetration/diffusion rate from outdoor sources to indoor environments. Additionally, one can determine the 'Diffusion Interval' for the impact of outdoor AQI on indoor air. According to Leung (2015), mechanical ventilation, natural ventilation, and infiltration are three main mechanisms that allow outdoor air to enter and affect indoor environments.[22] In our Air Quality Diffusion research, we work with Purple Air data and take all penetration factors into consideration in order to estimate the penetration/diffusion rate of a specific house during a particular period and quantify a 'Diffusion Interval' between outdoor and indoor AQI peaks.

---

[20] Jones, N. C., Thornton, C. A., Mark, D., & Harrison, R. M. (2000). Indoor/outdoor relationships of particulate matter in domestic homes with roadside, urban and rural locations. Atmospheric environment, 34(16), 2603-2612. https://www.sciencedirect.com/science/article/pii/S1352231099004896
[21] Johnson, T., Myers, J., Kelly, T., Wisbith, A., & Ollison, W. (2004). A pilot study using scripted ventilation conditions to identify key factors affecting indoor pollutant concentration and air exchange rate in a residence. Journal of Exposure Science & Environmental Epidemiology, 14(1), 1-22. https://www.nature.com/articles/7500294
[22] Leung, D. Y. (2015). Outdoor-indoor air pollution in urban environment: challenges and opportunity. Frontiers in Environmental Science, 2, 69. https://www.frontiersin.org/articles/10.3389/fenvs.2014.00069/full

In an overarching sense, to measure how fast the outdoor sources can cause a significant impact on indoor air quality, we first clean the data – removing AQI peaks that meet certain conditions from the time-series, as will be discussed in more detail in the next paragraph – and then analyze the relationship between subsequent outdoor and indoor AQI peaks. As Figure 16 demonstrates, generally, there will be at least one small peak in an indoor AQI curve right after each significant peak in the corresponding outdoor AQI curve. In this case, a peak in an indoor AQI series with prominence greater than a threshold can be regarded as a significant impact caused by outdoor sources. In this analysis, the average time interval between all pairs of peaks (an outdoor peak adjacent to an indoor peak) approximately reflects the penetration speed. The interval is defined as the 'Diffusion Interval' and takes into account mechanical ventilation, natural ventilation, and infiltration.

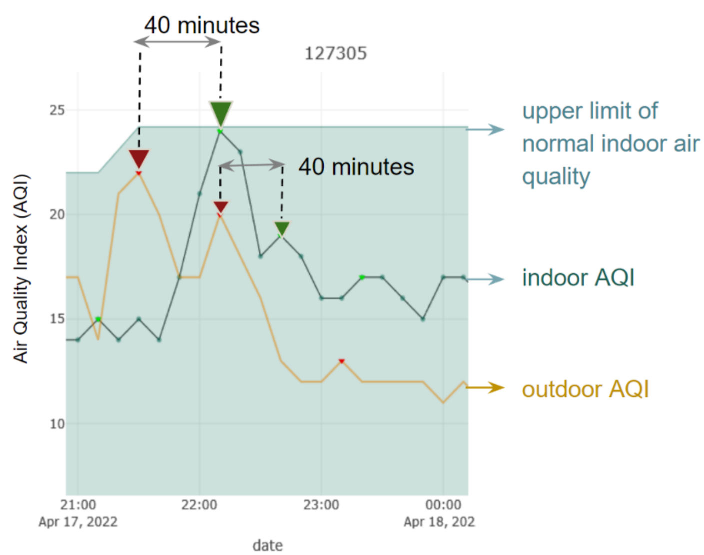


Figure 16. Diffusion Interval between indoor and outdoor air quality

Specifically, the MatLab function – *findpeaks* – is used for extracting peaks from indoor and outdoor AQI series. Table 1 shows the detailed configuration parameters used in the process. In the table, 'Outdoor Impact Time Interval' refers to the 120 minutes after there is a peak in its outdoor AQI series, in which the indoor AQI peaks are searched and analyzed. The 120 minutes is an empiric value based on the correlation analysis between indoor and outdoor AQI series. In addition, before the 'Diffusion Interval' is computed, the indoor AQI peaks induced by indoor sources need to be filtered out. If the indoor AQI at some point is greater than the maximum of its ambient outdoor AQI in the past 6 hours, it would be regarded as an outlier since obviously, the indoor sources play a considerable role at this point and the impact of outdoor AQI cannot single-handedly explain this peak.

Table 1. Configuration Parameters in Extracting Peaks from Indoor and Outdoor AQI

Parameters	Values
Outdoor Minimum Peak Prominence	1
Outdoor Minimum Height Difference	1
Intdoor Minimum Peak Prominence	0.1
Outdoor Impact Time Interval	120 minutes

Figure 16 shows the 'Diffusion Interval' between indoor and outdoor AQI for participant 127305. From the figure we can see that generally, it seems to take about 40 minutes for outdoor peaks to create a similar peak signal in indoor environments. It is worth noting that the peaks in the figure are generated near midnight, when indoor sources can be negligible. Therefore, the Diffusion Interval of 40 minutes can be a reasonable and comprehensive reflection of the house's resistance against ambient outdoor air pollutants in this particular outdoor environment condition (particular air pollutant concentration, temperature, wind speed, etc.). Figure 17 shows the diffusion intervals of 11 houses whose data is relatively complete among all 26 homes with OCOB sensors. The average of these houses' Diffusion Intervals is 43.7 minutes. From the figure, we can tell that – except for participant 126603 – others' Diffusion Intervals are very close to the average. Based on our observations, Participant 126603's indoor air quality is always much worse than its ambient outdoor air quality, interfering with the calculation of the Diffusion Interval due to the high number of outlying data points that must be removed prior to computation. Therefore, the reliability of its Diffusion Interval is not as high as other participants.

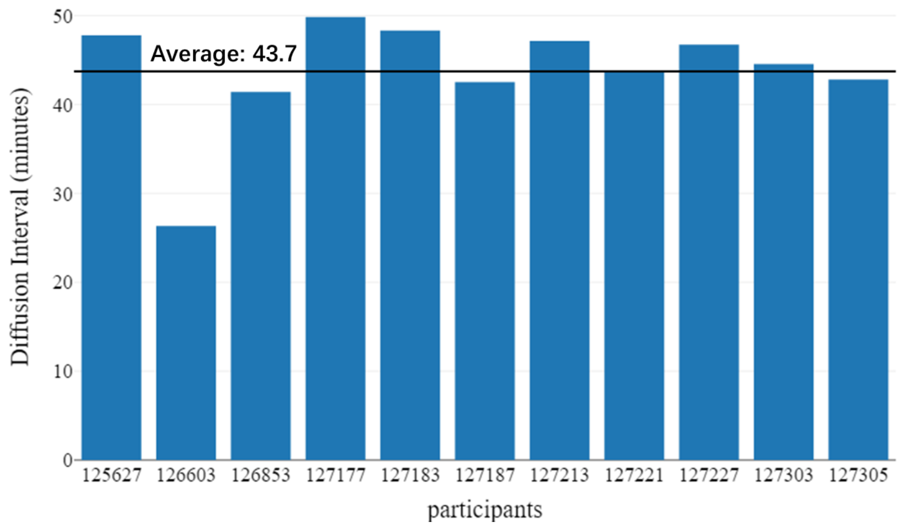


Figure 17. Diffusion Intervals of 11 Houses

The distribution of these 11 houses are shown in Fig. 18. The greener and larger the point is in the figure, the more resistant against outdoor air pollutants the house is – or, the longer the Diffusion Interval is. In the mapping, we can see that Participant 127221's sensor is located very close to that of Participant 125627. The outdoor air pollutant concentration, wind speed, and temperature can therefore be regarded as the same. On these grounds, it can be assumed that the difference between two houses' diffusion intervals lies in housing insulation conditions – as the impact of indoor activities has already been accounted for. Figure 19 presents a comparison between two houses' facility and insulation conditions. From the figure, we can see that Participant 125627 has better insulation than Participant 127221, since it has sealed windows. This may explain the higher Diffusion Interval of Participant 125627; this home is better protected against the infiltration of poor outdoor air.

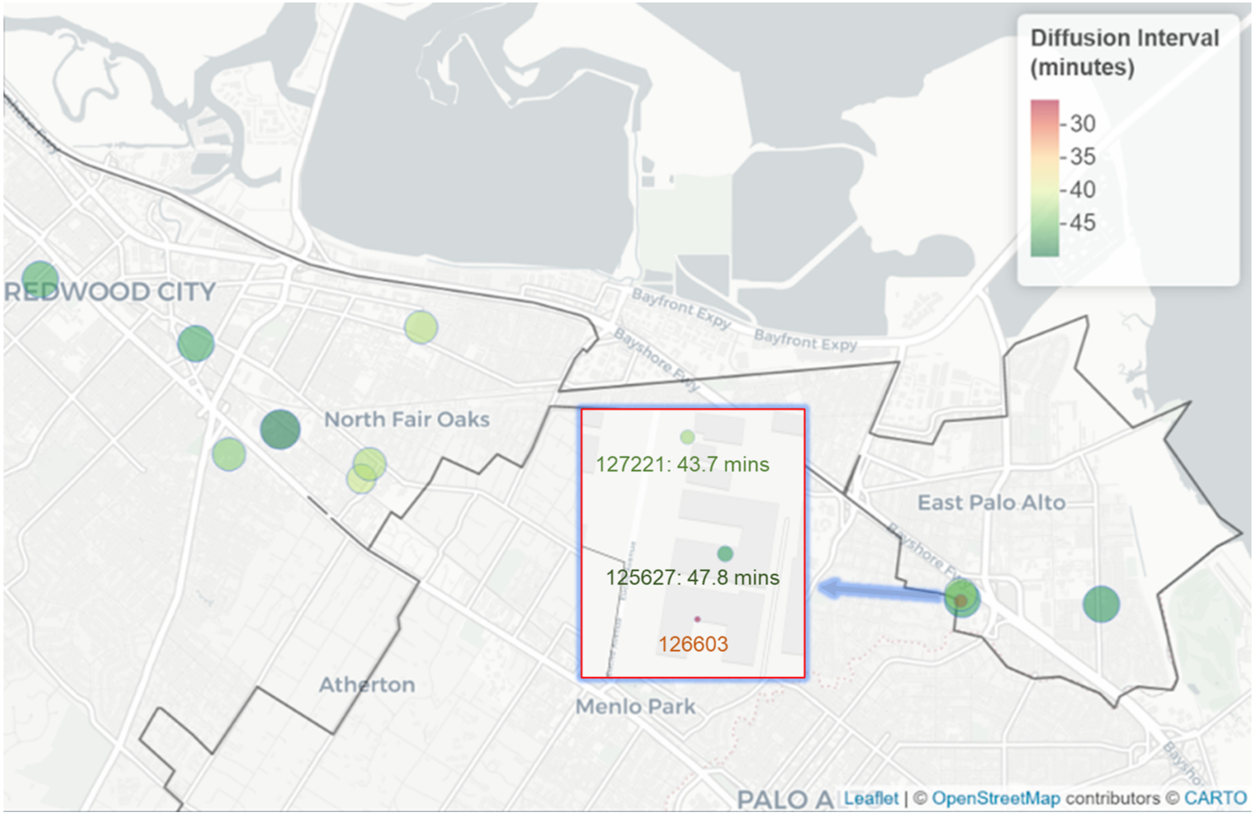


Figure 18. Mapping of 11 Houses and Their Diffusion Intervals

Figure 19. Home Facility and Insulation Comparison between Participant 127221 and 125627

To summarize, the methodology of exploring indoor and outdoor AQI peaks' properties in the section can be used to not only provide guidance to homeowners, but also measure a house insulation condition, as mentioned before, and give insights as to whether or not insulation steps are proving effective. Forty minutes of average Diffusion Interval time could inform how homeowners arrange the working time of the indoor purifier. For example, when the homeowner realizes that the concentration of outdoor air pollutants is relatively high, the homeowner does not need to turn on their purifier immediately, instead, one could wait for 15 minutes. In addition, the average duration of an indoor AQI peak induced by an outdoor peak can also be computed, with which the homeowner can estimate how long does their purifier need to work. Lastly, if a home shows high infiltration – thus low insulation – values in this investigation but has a variety of insulation measures installed, one could speculate that such techniques may be out-of-date or not functioning properly. With this information, the homeowner could work on replacing the window seals, for instance.

Time Series Analysis

Time-Series Decomposition

Time-series decomposition involves breaking down time-series data into three components: seasonality, trends, and random fluctuation. Seasonal variability refers to a repeating pattern over a fixed period of time. The trend underlies the time-series; for instance, the series may be increasing or decreasing. And, the random residual encompasses the noose that remains once the first two components are accounted for. The technique centers around three key steps: extracting the trend, finding the "average" season, and looking at the residual left-over noise. Decomposition can be done in additive or multiplicative sense; additive decomposition is best for data in which the seasonal variation does not change as the time-series values do, while multiplicative models should be used when the magnitude of the time-series data changes with the seasonal variation[23]. In quantitative terms, additive and multiplicative decomposition can be described as follows:

[23] Anomaly (2015). Extracting Seasonality and Trend from Data. Decomposition Using R. https://anomaly.io/seasonal-trend-decomposition-in-r/index.html

Figure 20. Additive Vs. Multiplicative Decomposition (Anomaly)

In our research, we experimented with both additive and multiplicative techniques and ultimately found that the multiplicative approach to be most realistic, as the additive model showed negative AQI values in its seasonal variation which is impossible.

When detecting the trend, we tried out a variety of different seasonality periods, ranging from weekly to daily and different hourly segments. The analysis shown in the report is based on a period of one day. Additionally, we used the DECOMPOSE() and STL() functions in R to assist with the data manipulation. Figure 21 shows the decomposed time-series data for a single participant's sensor, illustrating the observed data, the underlying trend, the seasonal variability, and the random noise.

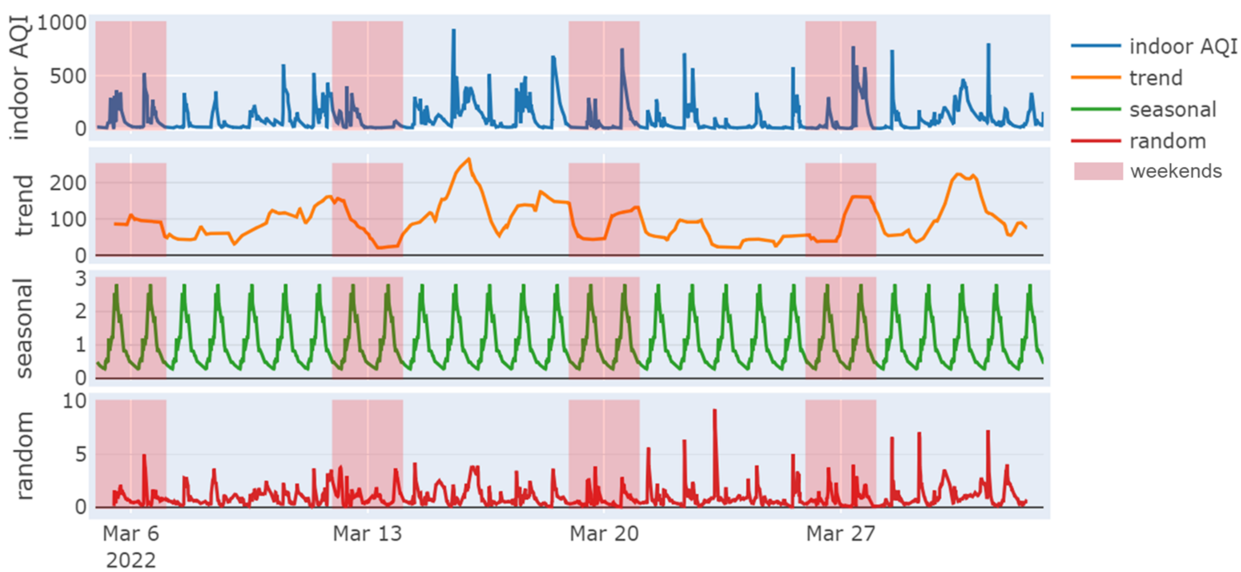


Figure 21. Indoor AQI Multiplicative Decomposition of Participant 125627 (Frequency = 1 Day)

For Participant 125627, the seasonal component has a main peak near noon (from 12:00pm to 2:00 pm), which might be because the homeowner always makes lunch at this time. In addition, there tends to be a small peak around 9:00am in the seasonal components, which could result from preparing breakfast. In terms of the trend and random components, they do not show significant regularity. And there is no obvious difference between these two components in terms of weekdays versus weekends, which can be seen by the highlighted red blocks in Figure 21. In summary, the time-series decomposition method may give us a general and rough picture of the daily indoor air quality pattern within a household by checking the seasonal component. However, the reliability of the pattern is still subject to the variability of the original indoor AQI from household to household. For 9 participants, their seasonal components are shown in Figure 22.. There does not seem to be a unified daily pattern for all participants. Therefore, the seasonal component decomposed from the original indoor AQI series can be a reflection of each homeowner's habit and deserves further attention and investigation.

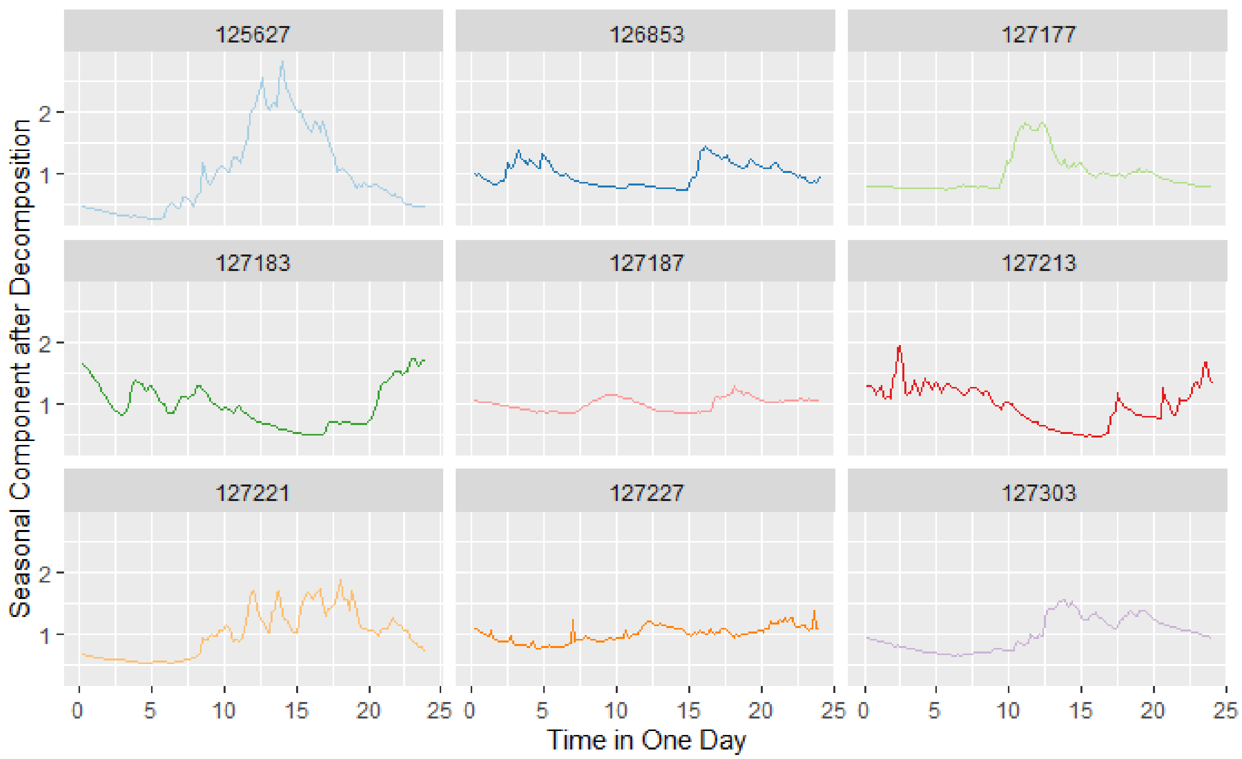


Figure 22. Seasonal Components of Indoor Air Quality Index for 9 Participants

Clustering

Once seasonal decomposition is performed, it is often useful to do a clustering analysis in order to identify a number of representative data points for the whole dataset, summarizing and distilling the patterns in the data. There are many popular methods for clustering with the most popular being the K-means approach. The essential concept of clustering can be seen in the following figure:

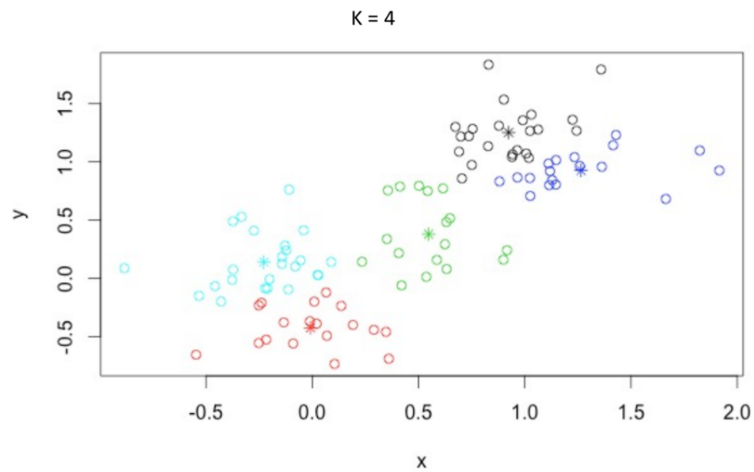


Figure 23. Basic Schematic of Clustering

The technique assigns each data point to its closest cluster center, where the intra-cluster distance is the average distance between member points and the cluster center. Inter-cluster distance on the other hand is that between the different cluster centers.[24] We experimented with varying numbers of clusters and time intervals for breaking down the data. We explored looking for patterns based on days of the week, 4 hour time intervals, and two hour time intervals. In the present analysis, we focus on four clusters and a time interval of 2 hours. Our findings and visualization of them in terms of the clustering analysis are quire preliminary and raw, but one can gain insights from looking at the following link which shows the results from a single household (participan 125627) in the OCOB study: https://j-i-n-p-u.github.io/jinpu.github.io/Clustering.html. Here, it is important to note that the time intervals refer to hours in the data, wherein 0-2 corresponds to 12 am-2 am, 2-4 to 2 am-4 am, and so on and so forth. The bar charts show the distribution of samples in each cluster, and the graphs demonstrate the individual AQI trends as well as the average trend for the cluster. Overall, the results suggest that healthy AQI may dominate at night, while more dangerous AQI values arise between 12 pm and 8 pm. However, the analysis currently only looks at the data of one participant at a time, and therefore is not representative of a broader scale phenomenon.

[24] Rajagopal, R. Time Series Clustering. Lecture 4: ED III. https://drive.google.com/drive/u/0/folders/1rbexope0nWtPz--LPF61IYGxfyCpYNx3

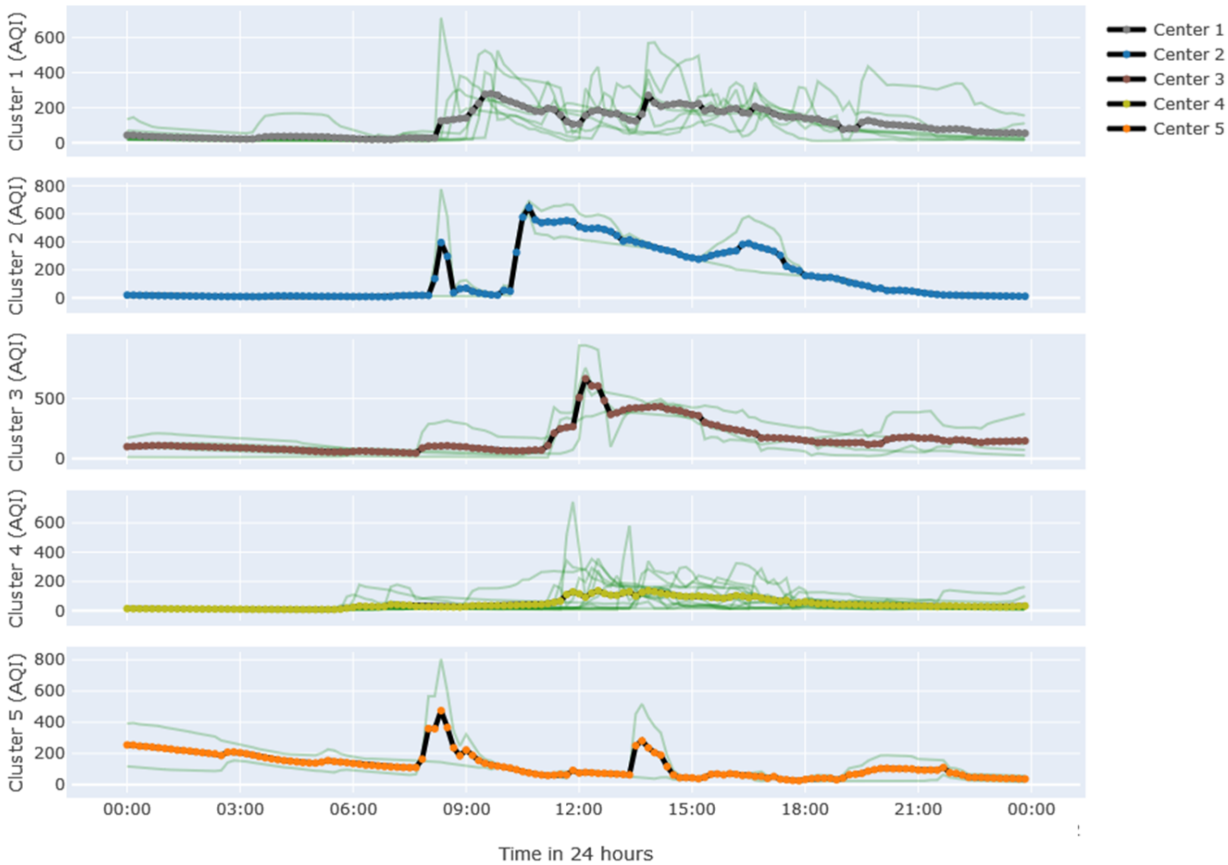


Figure 24. Five Daily Patterns for Participant 125627

Figure 24 shows five daily patterns of indoor AQI over one month (from March 5, 2022 to April 2, 2022) for Participant 125627, generated by K-means clustering. Compared to the daily pattern, i.e., seasonal component, generated in Time-Series Decomposition section (Figure 22), these patterns can show more details about this participant's indoor AQI during this period. Just from the Cluster 4 we can infer that sometimes, the homeowner does not prepare breakfast in his home. From the Cluster 1 we can find the indoor AQI tends to be relatively high from 9:00 am to 8:00 pm, and there seems to be no obvious peaks, which might indicate that, sometimes, the homeowner stays at home all day and is involved in some activities that lead to poor indoor AQ, such cleaning, or that ventilation in the home is poor, leading to long periods of elevated indoor AQI. In brief, currently, with limited participants (only 10 participants with relatively complete AQ data in this month), clustering is more suitable for analyzing a single participant's activity habits related to indoor AQ. However, with more data, one could look at larger, multi-household trends. Important next steps regarding the clustering techniques would be to use a moving time window, add more households, and flag extreme AQI values as a way of focusing on peaks rather than the general trend.

Highway Analysis

Studies show that air quality can be negatively impacted by proximity to roadways.[25] Given the presence of major roadways suchs Highway 101 and I-280, we investigated the air quality impacts relative to highway proximity. Using PurpleAir's outdoor sensor data in San Mateo County, we were able to perform an analysis that consisted of comparing air quality near highways to that far from highways. Our buffer for near or far from highways was ½ mile, meaning sensors within this buffer were tagged as "near", and otherwise they were "far". We limited the analysis to sensors within five miles of major roadways to minimize confounding variables, such as the influence from adjacent highways or smaller roadways. The map below shows highways, their half mile buffer, and outdoor sensors (green for near and red for far).

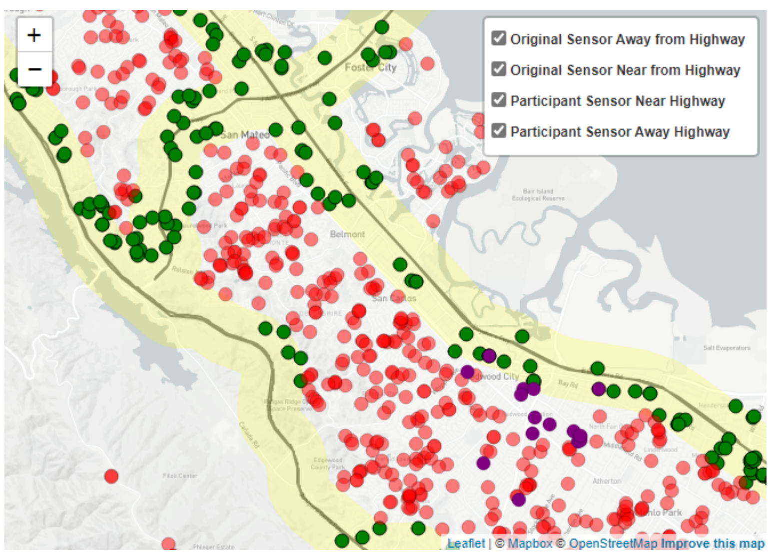


Figure 25. Highway Analysis map showing highways, their half mile buffer, and outdoor sensors.

Calculating the daily AQI average did not show any noticeable difference in AQI between near and far homes. We could explain this lack of difference because we are averaging

[25] Karner, A., et al. (2010). Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data. *Environ. Sci. Technol.* 44, 14, 5334–5344. *https://pubs.acs.org/doi/pdf/10.1021/es100008x?src=getftr.*

over the course of a day, meaning the impact of being near a highway is overshadowed by some other factors during the day which regresses the average back to a uniform mean.

To reduce the effect of averaging each day's AQI, we narrowed the analysis by averaging our month's data in 10-min intervals, enabling us to understand how AQI changed over the course of a day, instead of just one daily average. With this, we were able to focus on rush hour time intervals, understanding that that is when we'd expect a greater traffic volume. Here, we developed the figure below.

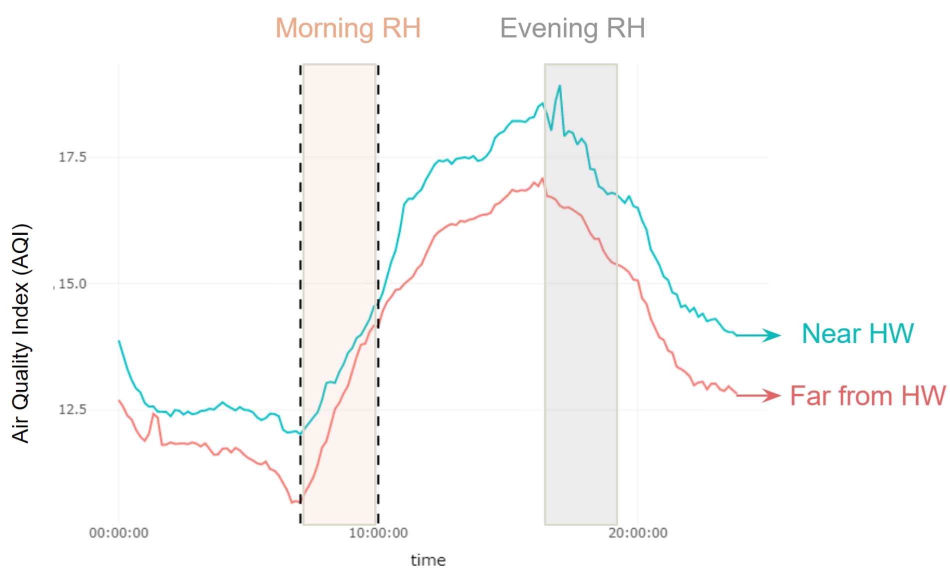


Figure 26. Average 24-hour AQI filtered by proximity to highway, overlaid by rush hours

Over 24 hours, we see that sensors closer to highways have worse AQI than those further away from it. Focusing on rush hour times as highlighted in the figure, for the morning rush hour (7-10am), the ambient outdoor AQI becomes worse quickly, as evidenced by the steep slope. But beyond our exceptions, the AQI far from the highway does so as well and actually tends to increase faster than that near the highway. It might be because some factors, such as plant emissions or industry emissions also play an important role in this period. For the evening rush hour (4-7pm), we expect to see a significant peak during this period. There is a peak in outdoor AQI near the highway. However, it is only a small peak, which lasts only 30 minutes. Confounding variables such as building emissions, the portion of electric vehicle users, or even just accounting for traffic volumes might mask the true effect of transportation/highways on AQI.

As a follow-up study/analysis, we suggest minimizing confounding variables as much as possible. This can look like incorporating wind data and focusing on sensors along wind lines. Another tweak could involve looking at data a year apart if electric vehicle ownership increased since they don't directly produce emissions. One could use a difference-in-difference approach here.

Questionnaire Analysis

In addition to installing sensors in participants' homes, the OCOB study also asks the homeowner a variety of questions in a questionnaire. The questionnaires provide key research information about the following 4 aspects: 1) Home and Household Description, including tenure, household members, air purification equipment/ventilation system, community involvement, building quality, home expenditure, etc; 2) Awareness / Experiences With Negative Environmental Events, including attitudes and awareness towards air pollutions, action taken under the situations of bad air quality, experiences with wildfires, smoky air, etc; 3) Health Description, including mental and physical health that may be related to air quality, symptoms that are currently suffering from, etc; 4) Demographic Characteristics, including gender, income, race, language, etc.

Our CEE 218Z student team was able to work with the questionnaire responses of 23 participants, and all Spanish responses were translated into English. Three participants did not have corresponding responses that could be used for analysis.

1) Home and household description

In terms of household description, these questions and responses enable us to undertake statistical analysis to determine whether equipment such as ventilation systems or certain types of insulation have significant impacts on indoor air quality. Thus, we subsequently may be able to provide useful suggestions about how to improve indoor air quality.

Figure 27. Chart of Presence or Lack of Insulation/Ventilation By Participant

The plot above shows the home insulation/ventilation facilities of 23 participants. 52% (12 out of 23) of homes have sealed windows, and 35% have weatherstripped doors – both of which intend to prevent the leakage of pollutants from the outside into the interior. For air purification facilities, only one household has an air conditioner, and two households have a HEAP purifier. These two purifying technologies can actively reduce indoor air pollutants and ensure good indoor air quality.

However, as we tried to move forward to linear regression analysis, we found that the correlation between indoor AQI and the facilities mentioned in the above plot was not significant. We see two possible reasons for this: 1) the sample size is too small to generate confident conclusions; 2) even if these households have insulations or air purification facilities, if they open their doors and windows when the outdoor air quality is poor, air pollutants can still penetrate indoors. Thus, increased sample size or the removal of confounding variables would improve the analysis.

In order to further explore the second reason for the above assumption, we studied the phenomena observed in the participants' households that might be related to insulations and indoor air quality. For insulations, one questionnaire question asks whether homeowners find their homes to be "cold in the winter", be "hot in the summer", have"yellow stains on walls," have a "musty smell", or "need dust[ing] more." We consider these attributes relevant to both insulation and indoor air quality. For instance, if a home is cold in the winter and hot in the summer, there may be high infiltration of outside air into the indoor environment and lack of ventilation/air conditioning. Additionally, if the home needs consistent dusting, the air quality may be poor and particles are settling as a result.

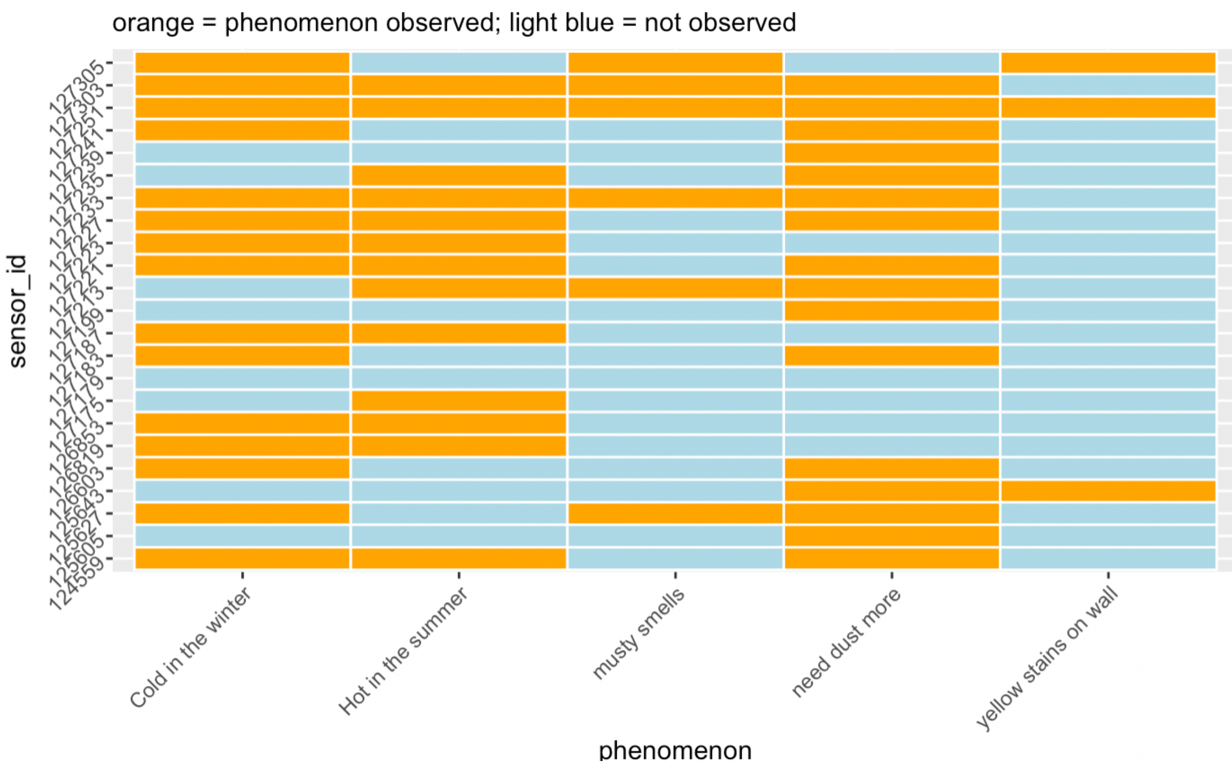


Figure 28. Chart of Home Occurrences/Phenomena By Participant

Once again, due to too few data points (only 23 participants), we believe that the conclusion of linear regression is not valid, even if its model output is statistically significant. Any findings may be a coincidence caused by lack of data and should not be given much confidence. From a qualitative point of view, we make the following reasonable assumption: households with more facilities observe fewer negative home phenomena that could indicate poor air quality and insulation. Further information is needed to validate this hypothesis. If the hypothesis holds, we can provide more specific recommendations for improving indoor air quality, such as adding air purification facilities to households where more air pollutants come from indoor activities, and adding insulation to households where more air pollutants come from outdoor penetration.

2) Awareness / Experiences with negative environmental events

In terms of experiences and awareness, we focus on 2 aspects – the knowledge and attitudes towards air quality and the experiences with negative events such as wildfires, smoky air, etc.

Public knowledge, behavior, and attitudes about air quality have major impacts on people's sentiments, risk perception, and exposure to health hazards regarding air pollution and the willingness to change in the name of mitigation. Previous studies have shown that a segment of the population does not actively understand AQI in their daily lives, and rarely checks the AQI of their working/living environment on a regular basis. However, according to Hilal K Al-Shidi, more than 90% of the people are concerned about air quality.[26] This also shows that people are willing to have a more intuitive and quantitative understanding of the air quality of their working/living environment, but due to insufficient education about air quality measurement indicators, it may be hard for people to link AQI values with air quality. Increased education could lead to promising results.

The following two pie charts show the AQI perceptions and actions taken by participants in the pilot study. Less than 34% (8 out of 23) of the participants have looked at an air quality index, and only 17% (4 out of 23) of the participants have checked the index on a regular basis (all the time & sometimes).

[26] Al-Shidi, K. H., et al. (2021). Public awareness, perceptions and attitudes on air pollution and its health effects in Muscat, Oman. J Air Waste Manag Assoc. 2021 Sep;71(9):1159-1174. https://pubmed.ncbi.nlm.nih.gov/33989134/

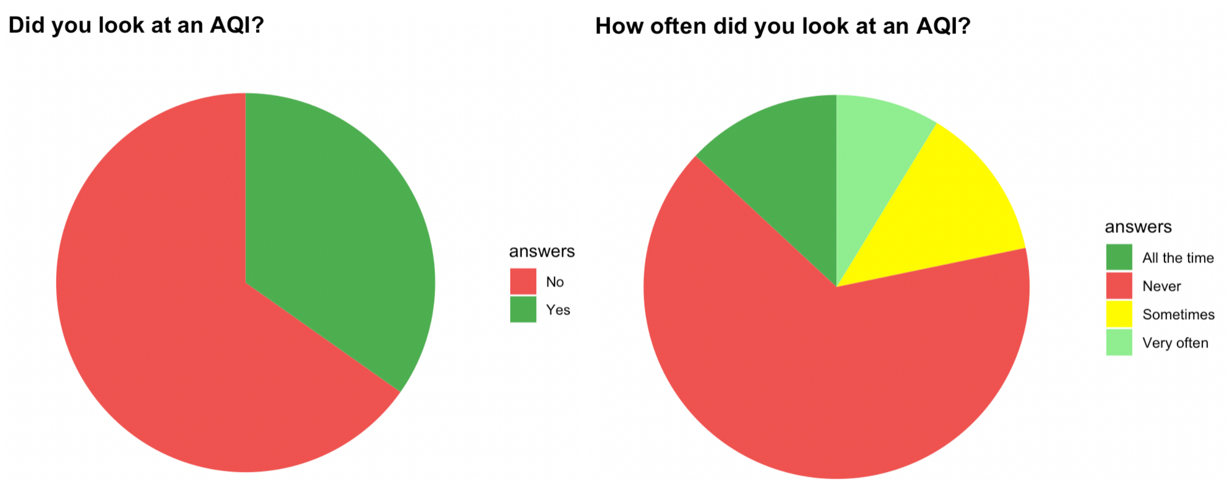


Figure 29. Pie Charts of Frequency of Checking the Air Quality Index

Therefore, when drafting the End-of-month Summary Report, we used a more vivid way to describe the AQI. For example, when the AQI value is 301-500, it belongs to the "Hazardous" category, which is equivalent to smoking 11-28 cigarettes per day. This insight allowed us to better understand our intended audience and tailor information to it. More information about the report will be provided in the sections that follow the Questionnaire Analysis.

Another aspect studied is the participants' experiences with negative events regarding air pollution. The questionnaire asks whether a series of statements related to air quality sound similar to the participants' experiences.

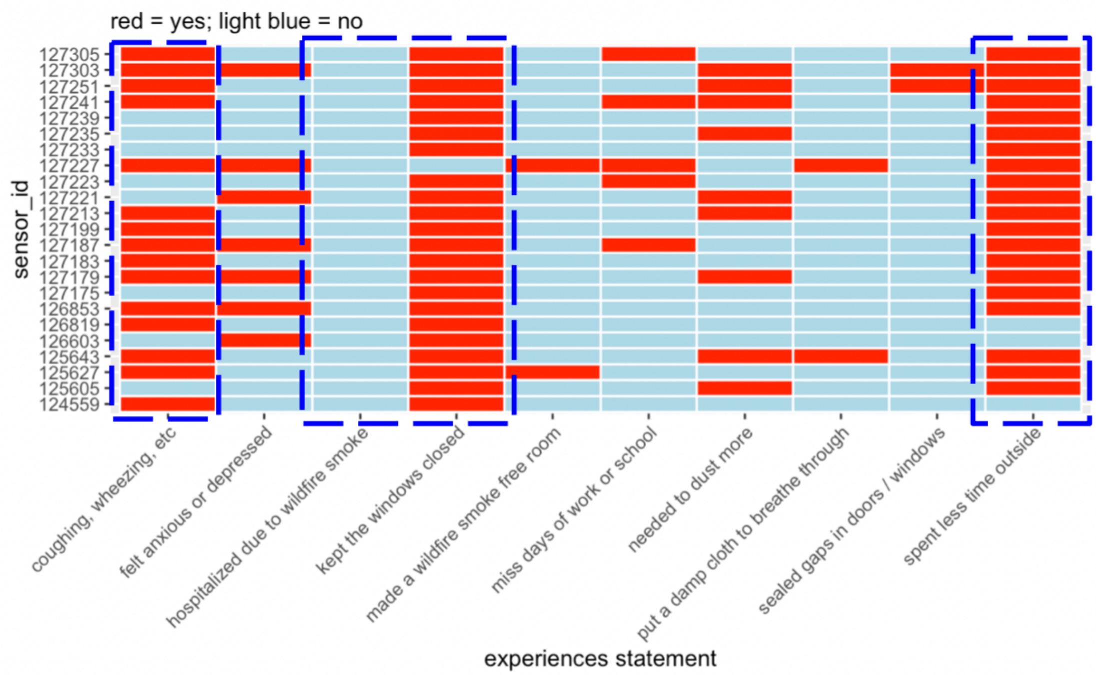


Figure 30. Chart of Experiences by Participant

As is shown in the above plot, although none of the participants were hospitalized due to wildfire smoke, nearly 83% (19 out of 23) of participants spent less time outside during negative events, and 96 (22 out of 23) of them kept their windows closed in their homes. 65% of the participants experienced coughing, wheezing, or headaches. Such results confirm that negative events such as wildfires do have a negative impact on air quality and human health. Put simply, residents can strongly perceive the existence of air pollution.

The questionnaire also asks participants about their perceptions of negative events such as wildfires. One of the questions asked was "Do you agree or disagree that smoky air that lasts for days from wildfires has been made worse by climate change?." 78% (18 out of 23) of the participants strongly agreed or agreed with this statement, while 21% (5 out of 23) of the participants strongly disagree with it. The diversity of thought may represent a challenge when working toward behavior change and increased resilience. It is suggested that in the follow-up promotion of sensors, the OCOB team could strengthen its education and outreach efforts, clarifying the sources of indoor and outdoor air pollutants, helping participants understand the air quality of their living/working environments, and guiding them to take actions to improve air quality.

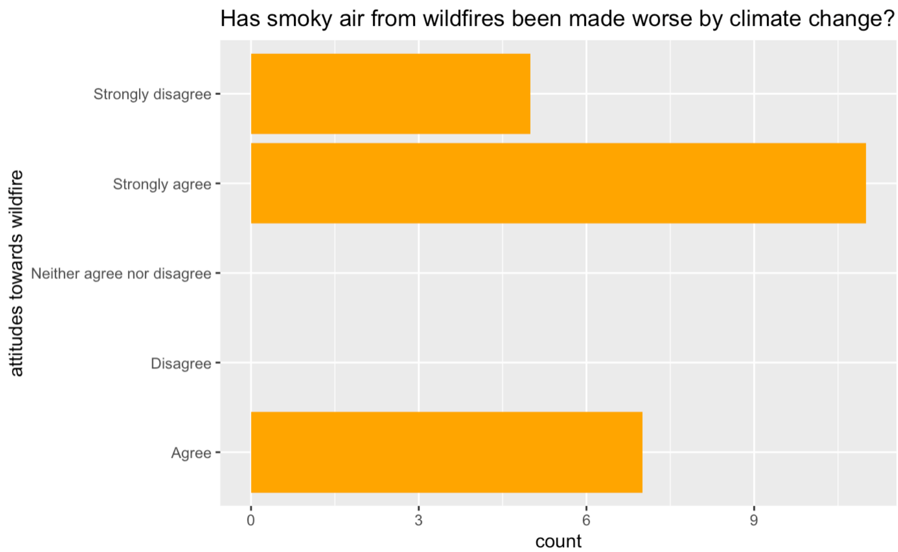


Figure 31. Participant Counts Corresponding to Particular Attitudes About Wildfires

Moving to the next chart below, regarding how people felt about smoky air that lasts for many days or wildfire events, we can see that people's attitudes vary from household to household. But the only thing that is certain is that people tend to more or less have negative emotions due to wildfires, such as stress, helplessness, etc. Therefore, we can see that – in this area – wildfires and air pollution (indoor and outdoor) in general may contribute to negative mental health outcomes.

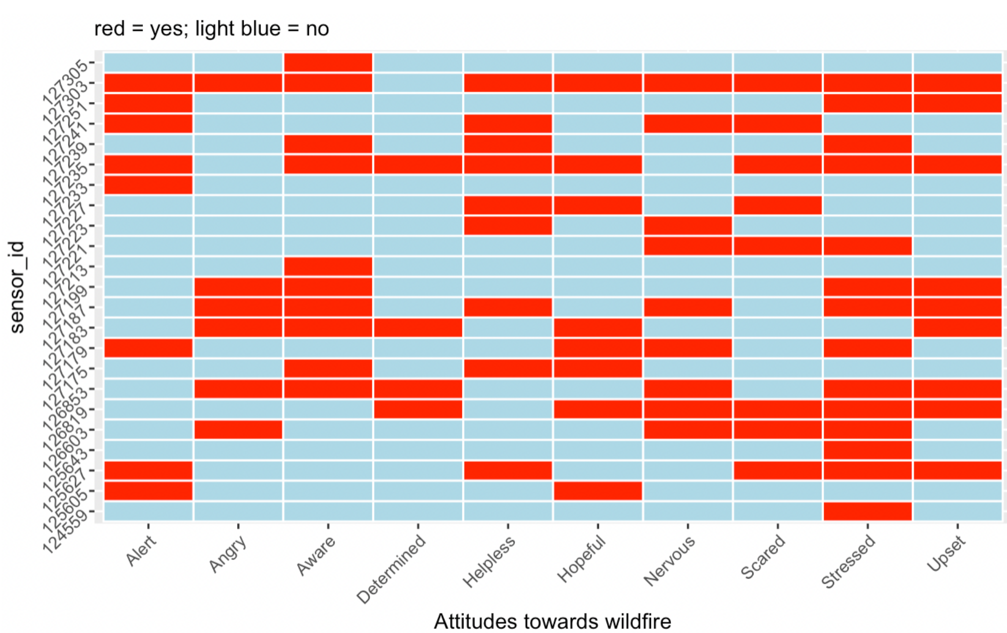


Figure 32. Chart of Attitudes About Wildfire by Participant

3) Health description

In this part, we explore the symptoms that may be related to poor air quality. According to CCOHS, when indoor air is not adequate, there may be issues such as increased health issues (e.g., eye and respiratory irritations) and, in rare cases, more serious health issues (e.g., carbon monoxide poisoning).[27] Symptoms are often linked to poor indoor air quality, and air quality issues do not affect everyone in the same way. When it is an issue, it is common for people to report one or more of the following symptoms: coughing, trouble breathing, wheezing, asthma attacks, or similar; stinging eyes, scratchy throat, or similar; runny or stuffy nose, irritated sinuses, or similar. The below plot shows the health description of the 23 participants. Symptoms associated with poor air quality are marked with blue dashed boxes.

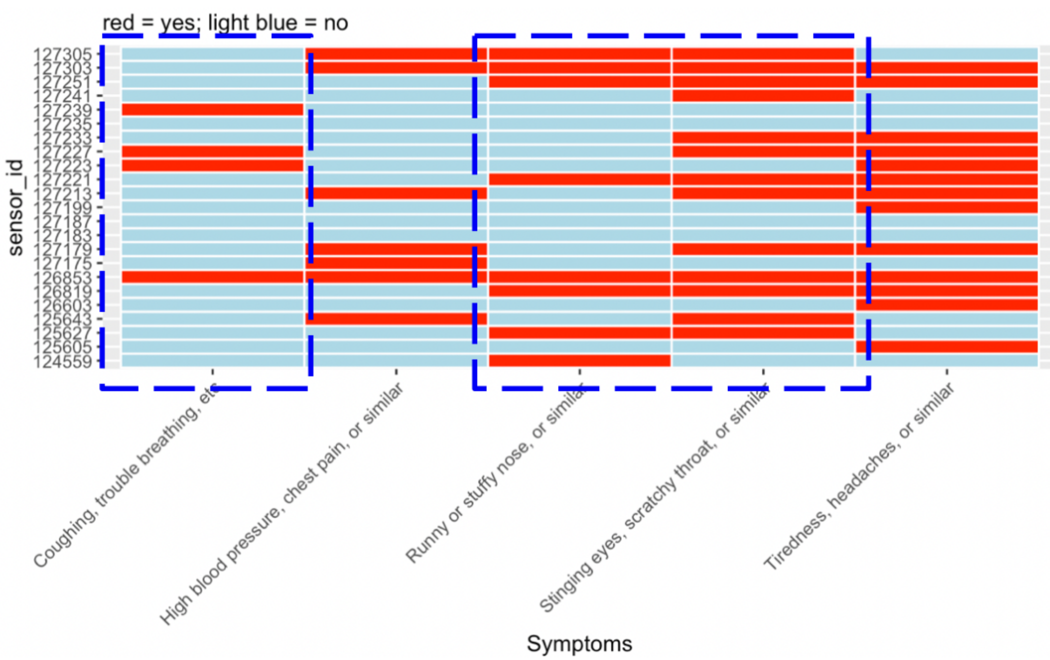


Figure 33. Chart of Air Quality-Related Symptoms by Participant

Although there is insufficient evidence to suggest that these diseases/symptoms are caused by the current poor indoor air quality, it is reasonable to infer that poor air quality can

[27] Canadian Centre for Occupational Health and Safety. (2022). Indoor Air Quality - General. https://www.ccohs.ca/oshanswers/chemicals/iaq_intro.html

contribute to the exacerbation of such symptoms. Therefore, the health conditions should be accounted for when assessing the situation. For the End-of-month Summary Report, we can also make recommendations to help participants improve their health by focusing on air quality. Further analysis could look at whether or not the prevalence or severity of symptoms are related to the insulation of the home, the degree of outdoor air pollution, the types of indoor activities, etc.

4) Demographic characteristic

In this part, we could consider gender, age, race, education level, etc. Due to the small number of participants (less than 30), it was extremely difficult to explore the relationship between demographic characteristics and air quality/attitudes towards air quality. As a common theme throughout this part of our project, very small samples make it difficult to generate convincing conclusions.

However, looking toward the future, equity analyses represent worthwhile next steps, when larger samples are available. For example, we can study the differences in the perception of air quality by different educational groups, the home facilities of different income groups, the differences in measures taken against air pollutants by different gender/age groups, and so on. Such investigations could generate rich insights and inform policy and public communication.

5) Questions that would be helpful to ask on the questionnaire

In our research, we found the following information to be potentially helpful in understanding the causes of indoor and outdoor air pollution, as well as possible measures to improve air quality. The information may better allow us and the OCOB team to provide valuable suggestions on a household-basis in the End-of-month Summary Report. The questions below are ideas for future inquiry and additions to the questionnaire. It is important to note that the framing and wording is likely not well-tuned, yet we believe the underlying concept to be useful.

- <u>Home and Household Description</u>
    - Do you participate in any of the following activities:

        - cooking on a gas stove, frying, smoking, vaping, vacuuming, etc. (actions that tend to harm indoor air quality)
        - opening the windows for ventilation on a clear day, turning on an air purifier, keeping carpets clean, changing the AC filter, etc. (actions that tend to improve indoor air quality)
        - If so, when and how often?
    - Have you ever done any other things indoors that you believe could lead to change in the AQI of your home environment? If so, when and how often are these actions performed?
- <u>Awareness / Experiences with Negative Environmental Events</u>
    - To what degree do you believe that you have the power to improve your indoor air quality? (number rating scale for responses)
    - In what ways do you think can improve indoor air quality?
    - What actions have you taken to improve air quality?
    - What potential air pollutants have you identified in your home (e.g. cooking, smoking, etc...)?
    - What potential outdoor air pollutants have you identified (e.g. industry, transportation, etc...)?
    - To what extent do you see indoor air pollution as hazardous to you and your family? (number rating scale for responses)
    - To what extent do you see outdoor air pollution as hazardous to you and your family? (number rating scale for responses)
- <u>Health Description</u>
    - What symptoms have you had (or currently have) that you think are strongly related to poor air quality?
    - Which symptoms do you think are strongly related to poor air quality?

Socioeconomic (Tax & Parcel & Deed) data analysis

As mentioned, the above questionnaire analysis only involved a small number of participants in the pilot study. We learned about the behaviors, attitudes, and household descriptions of these participants in relation to air quality, which will be helpful in providing suggestions for further improving air quality and the promotion of sensors. Additional research can be done as the project scales toward 400 indoor sensors.

In an effort to continue this type of analysis with a larger sample size at the present moment, we looked toward readily-available socioeconomic data. To gain a high-level understanding of the relationship between air quality and building properties, we studied the air quality across San Mateo County through the Tax, Parcel and Deed datasets.

1) Property value (Tax data)

Assumption: Buildings with higher property values have better average indoor air quality.

Preliminary experiment: AQI is derived from the Purple Air API, with 10-minute intervals, over the last month. After removing extremely high property values, we show the relationship between property value and average indoor AQI. While we can't jump to conclusions, as the graph below shows that buildings with a property value of less than 100,000 may have worse indoor air quality than homes with higher property values. A high property value may mean that the residents have higher income, stronger perception of air quality, better air purification or insulation equipment, or the building is located in a better local environment.

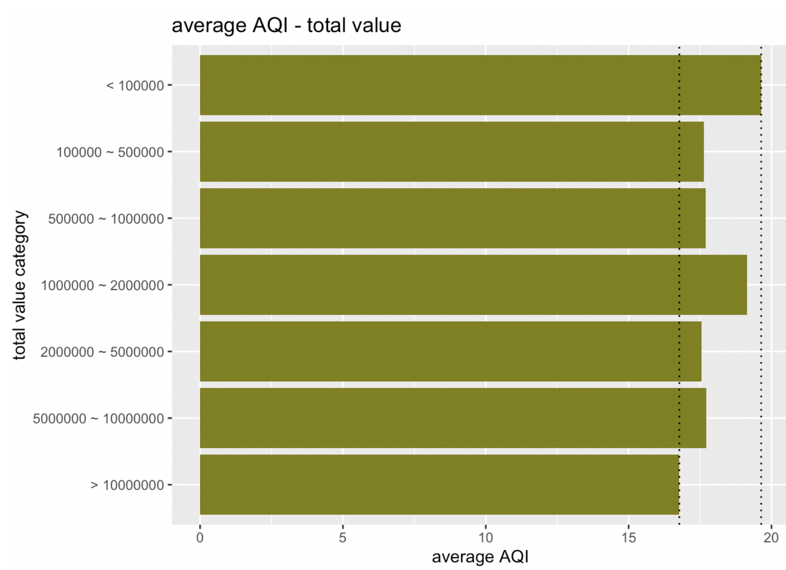


Figure 34. Average Indoor AQI vs. Total Property Value

The two maps below show the distribution pattern of property value and average indoor AQI in SMC. Darker colors means a higher value. i.e. Darker green indicates a higher property value, and darker red represents a higher average indoor AQI. As a whole, from visual inspection, it appears that the darker green circles – representative of higher property values – exist where the lighter red circles – representative of healthier AQI – do. And, vise versa, suggesting issues of environmental justice associated with indoor air pollution and wealth.

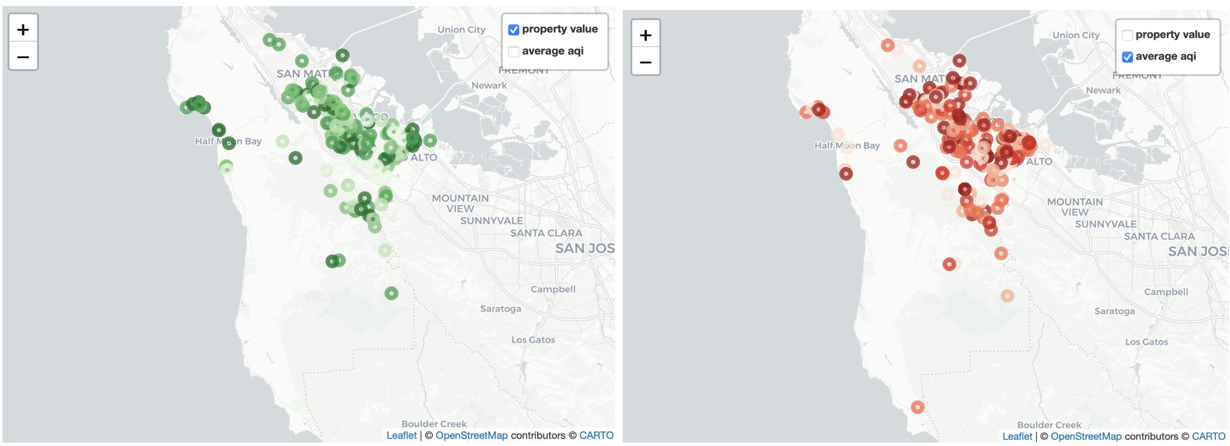


Figure 35. Maps of Property Value (Left) and Average Indoor AQI (Right). Darker hues indicate higher values for both property value and AQI.

2) Residential buildings VS non-residential buildings (Deed data)

Assumption: Residential buildings' average indoor AQI is different from that of the non-residential buildings.

Preliminary experiment: We compared average indoor air quality between residential and non-residential buildings. The graph below shows that the average daily indoor air quality may not be significantly different between the two types of buildings. We suspect that if indoor air quality differs between the two types of buildings, the difference is mainly due to conditions like cooking. In the future, we propose that researchers explore the differences in average indoor air quality between residential and non-residential buildings during peak cooking hours and in conjunction with knowledge of within-building activities.

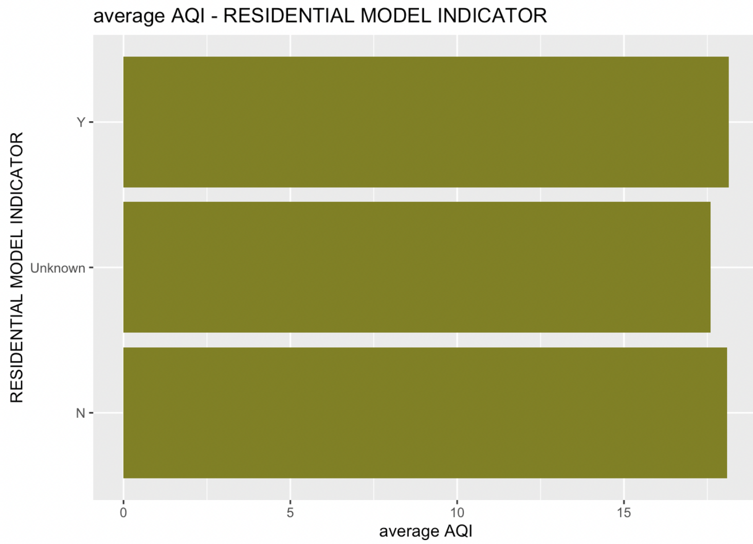


Figure 36. Average AQI By Building Type (Y = Residential, N = Not)

3) Year built (Parcel data)

Assumption: There are differences in average indoor air quality between older and later buildings. Year-built has a relationship with average indoor AQI.

Preliminary experiment: As shown in the graph below, we were surprised to find that there was no clear linear relationship between year built and the average indoor air quality, and even buildings that were built earlier have better indoor air quality. This is different from our assumption. We consider the following reasons to explain the situation:

1) The geographical location of the building. If indoor air quality is more affected by outdoor air quality than indoor activities, then different locations have higher impacts on indoor air quality, and thus the effects of building characteristics on air quality are weakened.

2) Regular renovation of the buildings. If the old building is renovated on a regular basis, it may have the same or better ventilation facilities than the new building, and the built year would not have an excessive impact on the air purification and insulation function of the building.

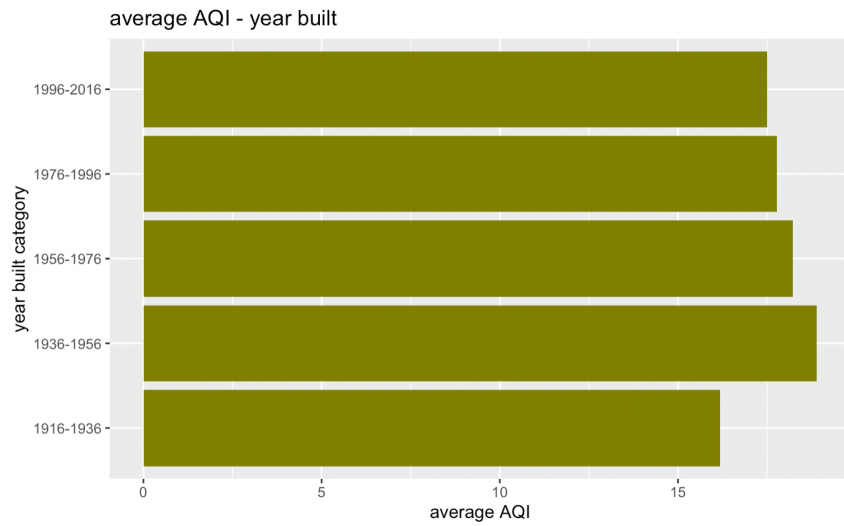


Figure 37. Average AQI By Year Built

Next Steps For Questionnaire and Socioeconomic Data Analyses

Given the sparsity of the current available data and the small sample size, we provide potential methods for subsequent comprehensive analysis. One can explore the factors that affect indoor and outdoor air quality by building a multiple regression model and understand which factors have the greatest impact on air quality.

Further, for the data from the questionnaire, we could assign different weights to all the questions in order to better analyze the data we collected, and we determine the score for each household by computing the weighted average. These scores include facility score, sanitary score, and insulation score, etc. The main reason for this is that the information we acquire from the questionnaire is binary data, which is difficult to evaluate alongside quantitative data, so we use a weighted average to convert these binary data components into quantitative data. Next, we may combine these generated scores with the parcel data, deed data and tax data for analysis, and these scores allow us to conduct a comprehensive analysis of the data more intuitively. Finally, we could set the above-mentioned factors as inputs and AQI as the output to study the linear relationship between the input and the output through the method of multiple regression. In conclusion, this method may be helpful to get a better understanding on "what lead to poor air quality" and "what is the most critical action that we need to take in order to improve the air quality."

End-of-month Summary Report

Moving from data analytics to public communication, we also created a draft of an end-of-month summary report for homeowners. In theory, this report would be sent to households with OCOB sensors, informing them about the significance of air quality, ways to improve air quality within the home, the indoor air quality of their own home this month, and the overall indoor air quality of all sensors within the OCOB project this month. The development process included high-level conceptualization, data manipulation and visualization in R, as well as feedback from classmates, our professor, and two members of the local community. On this note, we met with East Palo Alto's Sustainability Officer and an individual who spearheaded the 2019 effort to install outdoor air quality monitors in East Palo Alto. With deep understandings of the realities and thought processes of community members, the two provided invaluable insights and sparked critical revisions. Over the 10 week quarter, our student team created three versions of the summary report.

Through its organization, text, and plots, the report operates two-fold, aiming to both educate and empower homeowners. With regard to education, we focused on maximizing homeowner comprehension, while on the topic of empowerment, we wanted homeowners to feel informed and equipped to take control of their indoor air quality – rather than stressed or helpless. The final iteration of the end-of-month summary report is the three-page document depicted below.

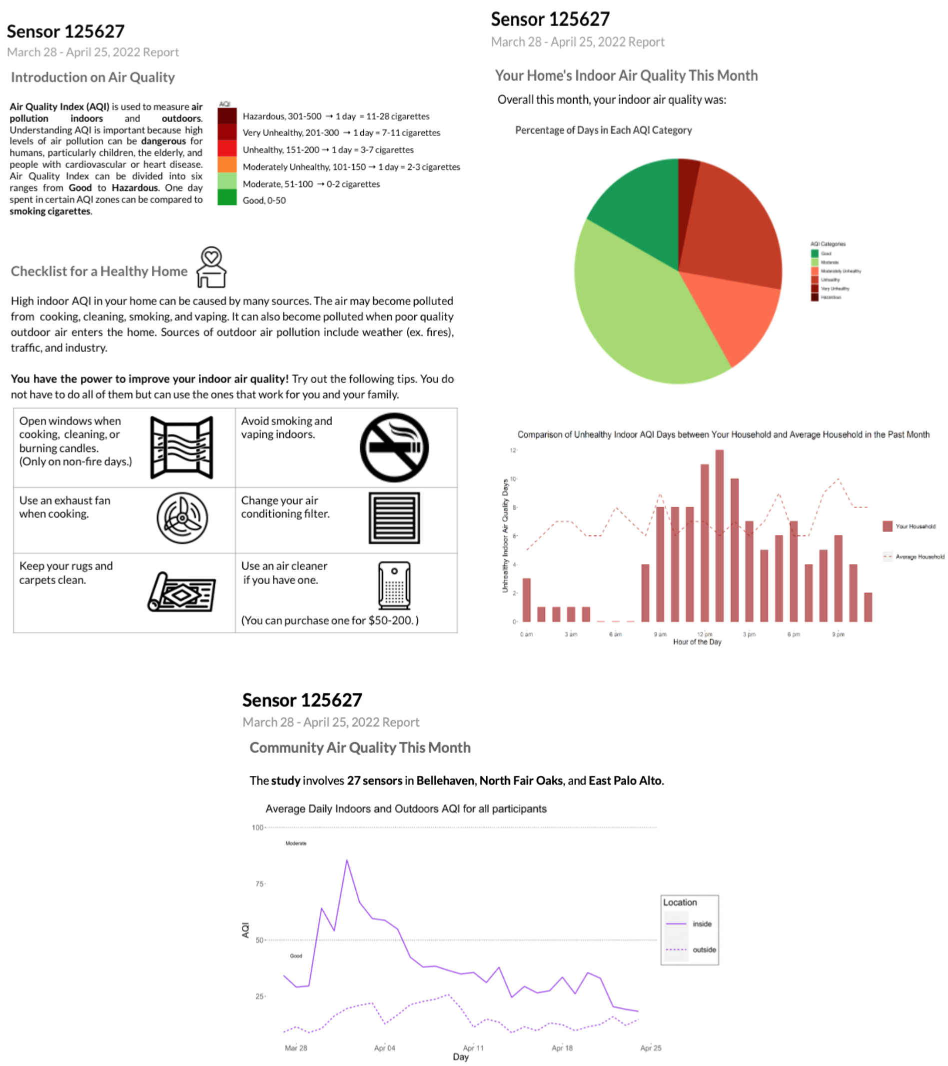


Figure 38. Draft of End-Of-Month Summary for OCOB Homeowner

On all pages, the header states the Sensor ID and time period referenced in the report. The first page begins with a brief description of AQI as well as the scale for differentiating between air quality levels and the degree of health concern. Significantly, we adjusted the color scale here to more accurately reflect the level of danger, switching to an orange/red hue when the AQI exceeds 100. Further, we compared a 24-hour exposure to the different AQI conditions to smoking "X" number of cigarettes. This choice reflects the fact that homeowners likely are more familiar with the health implications of smoking, making the analogy relevant and effective. Below the introductory section lies the "Checklist for a Healthy Home" which serves to provide accessible ways in which homeowners can mitigate indoor air pollution. Techniques range from opening one's windows to keeping carpets clean and buying an air quality monitor. Here, we incorporated visuals to convey our message and worked to diversify the tips such that any family could find one or two that fit their lifestyle and level of resources. In addition, we chose to lead with the checklist – before showing any data – to give the homeowner a sense of agency and hope.

On page two, we describe the recent air quality conditions within the singular home. First, as a main takeaway, we categorize the air for that month, labeling it as "Healthy," "Moderate," "Moderately Unhealthy," "Unhealthy", "Very Unhealthy," or "Hazardous." (This is not shown in Figure 38, but the code has been written to generate this characterization and can easily be added to the design.) If homeowners scan the report for a few seconds, we want to ensure that they at least gain this overarching insight. Below, one finds a pie chart that reflects the percentage of days spent in each AQI category over the course of the month. The plot provides a more nuanced view of indoor air quality than the earlier categorization. In the graph below, the homeowner learns about how their indoor AQI compares to that of the other home's in the study. It depicts only the time with an average AQI value above 100. For this reason, we chose to make all bars and lines red, indicating unhealthy quality. Further, the axes clearly reflect the time of day and day count, helping readers understand the information. In the sample summary report, we notice that indoor air quality is at its worst midday around 12 or 1 pm and is worse than that of the average home in the study, potentially suggesting that cooking techniques at lunch time contribute to the pollution.

The final page zooms out to analyze the quality of indoor air across the community of OCOB homeowners with the plot illustrating the difference in indoor and outdoor air quality daily average trends. We chose purple here as to not convey a message of health or danger.

Looking forward, we see a few areas of improvement or continued work. First, the report should be converted into a Shiny dashboard, allowing users to interact with information more readily. In addition, a script that auto-generates the report each month would be very useful for disseminating the summaries. As for smaller changes, one could include the address at the top of the report, in the case that the homeowner does not know his or her Sensor ID number, and could carry the analogy between AQI and cigarettes smoked throughout the report. The initial description could show certain quantities of smoking icons to demonstrate the number smoked, and the axes of the plots could be changed from AQI to cigarettes smoked. On the final page, the plot could more clearly depict the quality of the air and level of health concern by having colors in the background that correspond to the AQI category.Ultimately, we hope that our mock-up communicates the sensor data, importance of indoor air quality, and opportunity to reduce in-home air pollution to those participating in the OCOB project.

Conclusion

In partnership with OCOB on their new project studying indoor air quality in East Palo Alto, Bellehaven, and North Fair Oaks, our student research team explored the sources, mechanisms, impact, and communication of indoor air quality based on the data of 26 new indoor air sensors. Overall, we found indoor air quality to be a significant hazard in this region, emphasizing the importance of continuing to study and educate about the phenomenon. We conducted many forms of analysis detailed above. Notably, our results show indoor air quality to be worse than outdoor air quality – particularly for homes in the OCOB study and likely due to the impact of indoor activities. Next, we found there is a high relationship between indoor and outdoor air quality by conducting cross-correlation analysis and Granger Causality test. Further, our effort focused on outdoor air penetration found a 40 min lag time between outdoor and subsequent indoor peaks. Additionally, we created a mock-up of an end-of-month summary report for homeowners that aims to empower and inform those with sensors. We also conducted time-series, clustering, highway, questionnaire, and socioeconomic data analyses. All methodology can be built upon, and we anticipate more nuanced, robust results from them with a larger sample size. This report summarizes our work and lays the foundation for continued investigation.

References

AirNow.Gov. (2022). AQI Basics for Ozone and Particle Pollution. Air Quality Index (AQI) Basics. https://www.airnow.gov/aqi/aqi-basics/

Anomaly (2015). Extracting Seasonality and Trend from Data. Decomposition Using R. https://anomaly.io/seasonal-trend-decomposition-in-r/index.html

Boyd-Barrett, C. (2019). People of Color and the Poor Disproportionately Exposed to Air Pollution, Study Finds. California Health Report. https://www.calhealthreport.org/2019/02/08/people-of-color-and-the-poor-disproportionately-exposed-to-air-pollution-study-finds/

CA.Gov, (2022). Cal Fire. https://www.fire.ca.gov/incidents/

Derrick, T., & Thomas, J. (2004). Time series analysis: the cross-correlation function.
EPA (2022). Technical Assistance Document for the Reporting of Daily Air Quality the Air Quality

EPA. (2022). Improving Indoor Air Quality. Indoor Air Quality. https://www.epa.gov/indoor-air-quality-iaq/improving-indoor-air-quality

Government of New South Wales. (2021). Indoor Air Pollution. Air Quality. https://www.health.nsw.gov.au/environment/air/Pages/indoor-air-pollution.aspx

Government of New South Wales. (2021). Outdoor Air Pollution. Air Quality. https://www.health.nsw.gov.au/environment/air/Pages/outdoor-air-pollution.aspx

https://dr.lib.iastate.edu/entities/publication/7a1756d8-b106-4ad8-b0ac-d6cb1fb76bc0

Index (AQI).
https://www.airnow.gov/sites/default/files/2020-05/aqi-technical-assistance-document-sept2018.pdf

Johnson, T., Myers, J., Kelly, T., Wisbith, A., & Ollison, W. (2004). A pilot study using scripted ventilation conditions to identify key factors affecting indoor pollutant concentration and air exchange rate in a residence. Journal of Exposure Science & Environmental Epidemiology, 14(1), 1-22. https://www.nature.com/articles/7500294

Jones, N. C., Thornton, C. A., Mark, D., & Harrison, R. M. (2000). Indoor/outdoor relationships of particulate matter in domestic homes with roadside, urban and rural locations. Atmospheric environment, 34(16), 2603-2612.
https://www.sciencedirect.com/science/article/pii/S1352231099004896

Karner, A., et al. (2010). Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data. Environ. Sci. Technol. 44, 14, 5334–5344.
https://pubs.acs.org/doi/pdf/10.1021/es100008x?src=getftr.

Kuo, H. W., & Shen, H. Y. (2010). Indoor and outdoor PM2. 5 and PM10 concentrations in the air during a dust storm. Building and Environment, 45(3), 610-614.
https://www.sciencedirect.com/science/article/pii/S0360132309001966

Leung, D. Y. (2015). Outdoor-indoor air pollution in urban environment: challenges and opportunity. Frontiers in Environmental Science, 2, 69.
https://www.frontiersin.org/articles/10.3389/fenvs.2014.00069/full

Liang, Y., et al. (2021). Wildfire smoke impacts on indoor air quality assessed using crowdsourced data in California. PNAS, 118(36).
https://www.pnas.org/doi/10.1073/pnas.2106478118

Meadow, J. F., Altrichter, A. E., Kembel, S. W., Kline, J., Mhuireach, G., Moriyama, M., ... & Bohannan, B. J. (2014). Indoor airborne bacterial communities are influenced by ventilation, occupancy, and outdoor air source. Indoor air, 24(1), 41-48. https://onlinelibrary.wiley.com/doi/full/10.1111/ina.12047

National Institute of Environmental Health Sciences. (2022). Air Pollution and Your Health. https://www.niehs.nih.gov/health/topics/agents/air-pollution/index.cfm#:~:text=Air%20pollution%20can%20affect%20lung,are%20linked%20to%20chronic%20bronchitis.

Our Communities, Our Bay. (2022). What is "Our Communities, Our Bay?" Our Project. https://www.ourcommunitiesourbay.org/project

Personal Communication with Sustainability Officer of East Palo Alto, May 2022.

Poupard, O., Blondeau, P., Iordache, V., & Allard, F. (2005). Statistical analysis of parameters influencing the relationship between outdoor and indoor air quality in schools. Atmospheric Environment, 39(11), 2071-2080. https://www.sciencedirect.com/science/article/pii/S1352231005000166

Rajagopal, R. Time Series Clustering. Lecture 4: ED III. https://drive.google.com/drive/u/0/folders/1rbexope0nWtPz--LPF61IYGxfyCpYNx3

Roberts, T. (2016). We spend 90% of our time indoors, says who? Building Green. https://www.buildinggreen.com/blog/we-spend-90-our-time-indoors-says-who

Yocom, J. E., Clink, W. L., & Cote, W. A. (1971). Air Quality Relationships. Journal of the Air Pollution Control Association, 21(5), 251-259. https://www.tandfonline.com/doi/abs/10.1080/00022470.1971.10469525