Sensing Lungs

Breathing urban realities

A Project By:
Steven Shi, Jui Shih, and Erma Swartz

Urban Sensing & Data
Professor: Anthony Vanky
Spring, 2025
MS Urban Planning, GSAPP

Contents

EXECUTIVE SUMMARY

Sensing Lungs is an interactive environmental sensing project investigating disparities in ozone pollution levels and ozone exposure's physiological effects in urban environments. Developed by Steven Shi, Jui Shih, and Erma Swartz as a way to better understand the embodied realities of inhabiting often polluted urban environments, the project combines hardware prototyping, three-dimensional design, environmental science, and statistical modeling to measure how urban residents embody the realities of urban living.

Motivated by hypothesized disparities in air quality between Harlem, the Columbia University campus, and the North Wood of Central Park, the team wanted to know what the body of a student living in Harlem and studying at Columbia experienced throughout the average day. Does walking the dog, walking to school, going out in the neighborhood, or studying at home have different effects on the pollutants that a subject is exposed to? What are those differences, and what does the body do to respond to them?

By designing and building a set of human lungs and enclosing real "breathing" sensors within them, the team set out to collect ozone pollution levels and uncover what relationship between ozone exposure had on the target variable of the heart rate. The lungs had two chambers from which to breathe, one containing a Pulse Sensor Amped to monitor heart rate and an MQ-131 sensor to measure ambient ozone concentrations. These sensors feed data to an Arduino microcontroller, with results exported to SD cards for analysis.

Once the data was collected and cleaned, the analysis part of the process began. The team attempted several techniques to try and uncover the true relationship between ozone exposure and bodily reaction, including several machine learning models. This includes linear regression, polynomial regression, random forest, and ordinary least squares (OLS), and was done to examine the relationship between ozone levels and BPM.

While initial models suggested a weak to moderate correlation, deeper investigation of the data had better results, with the OLS regression performing most successfully. These high results were in such contrast with the original linear regression results that our team felt the need to dive deeper into our findings. Once we explored the information more clearly, ozone appeared to have an impact on heart rate when grouped by location. Rather than answering our questions, this created clearer and more crucial questions for us to answer. If we controlled for environment and movement, would the results be as strong?

When we re-calibrated our analysis, it became clear that it was never ozone impacting heart rate; rather, it was purely movement and environment that did so. The true drivers of heart rate variability, not ozone exposure. Once location-based confounders were controlled for, the direct correlation between ozone and heart rate diminished significantly.This finding underscores the complexity of environmental health monitoring and the necessity of accounting for contextual variables in sensor-based studies. It also highlights the potential of wearable and spatially aware technologies for generating granular environmental insights, especially when paired with robust data analysis.

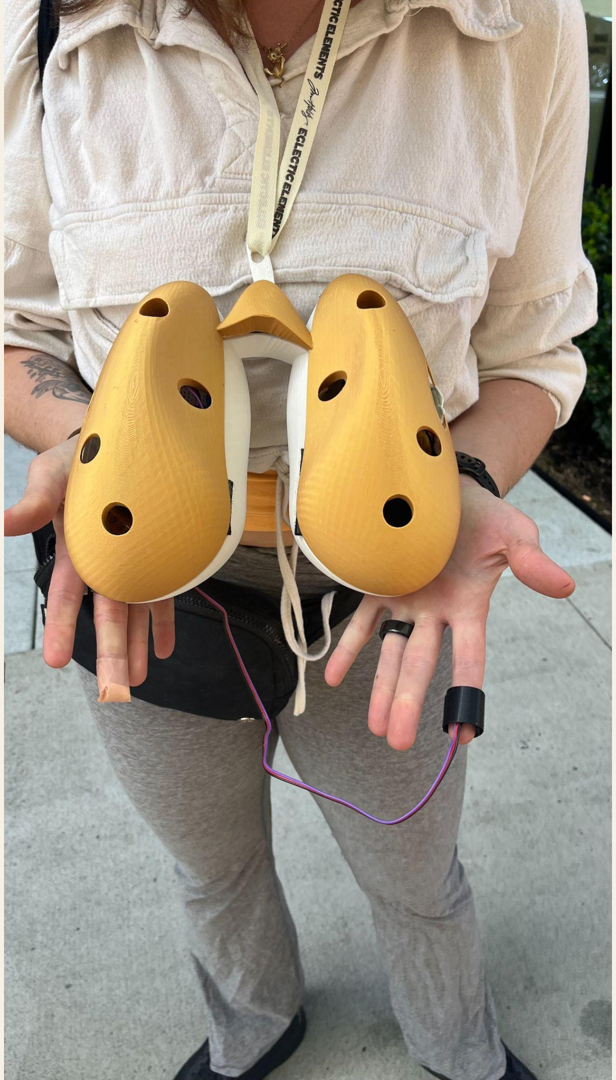
While our experiment was successful at proving a null hypothesis, we believe that with more time and more data collected in the real world, we may have found different results. In addition, with more variance in location visited and spent in one place outdoors, without walking, we may have uncovered different, still significant findings. In the future, we envision refining the prototype, improving sensor calibration, and potentially expanding the study to include additional pollutants, biometric indicators, or real-time mapping functions. The project serves as both a critique of simplistic environmental-health correlations and a proof of concept for community-based sensing tools in public health research.

INTRODUCTION

Overview & Motivations

Our project is called the “Sensing Lungs”. It is a lung-shaped model measuring heart rate and ozone level simultaneously. The model is designed to have a sensor on each side of the “lungs” to measure either variable: heart rate or ozone level. Specifically, when facing the front of the model, the left side of it has the sensor measuring heart rate, while the right side has the sensor measuring ozone level. We ask our users to manually initiate both sensors inside the “lungs” at the same time, connect the heartbeat sensor to their finger or ear, and start walking around. The model has holes on the surface, which allows users to observe the circuit inside and make sure the model is working while walking. Whenever the users feel like stopping the model, they can simply plug the power cord out of the two sensors on the both sides. There is a SD card connected to each sensor, which is saving heart rate or ozone level data while the sensors are working. After the sensors are stopped, users can take the SD card out and export the data to their own computers and conduct analysis.

Our motivation for this project comes from the fact that air quality keeps being a concern in the urban and rural contexts globally. New York City has specific concerns about ozone pollution levels, and there are disparities on ozone pollution levels between different parts of the city. We therefore wish to explore the potential difference on ozone pollution between Columbia University and its neighboring communities, such as Harlem and Central Park North. From literature, we are also aware of the impact of heart rate on how much air pollution can impact a person, so we also want to sense heartbeat to manifest the impact of ozone pollution on human body and health and to explore the potential correlations between these two variables.



Team: Tackling Urban Issues, One Dog Walk at a Time

Meet Jui

Meet Steven

Meet Erma

Meet Goose

Anecdotes

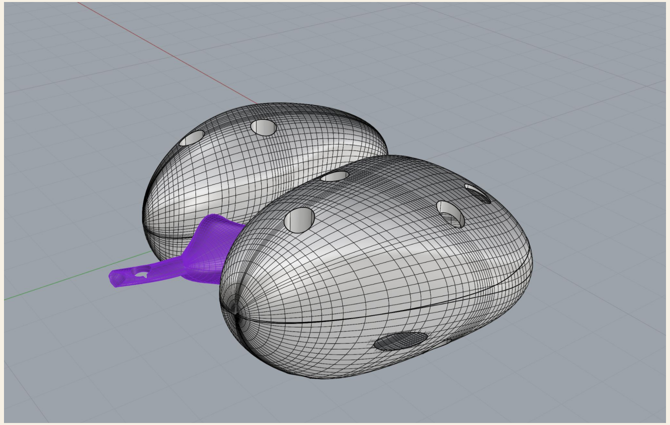
This project began with a simple observation: our bodies respond in real time to the environments we move through. At the individual level, we wanted to investigate measurable changes in heart rate linked not only to physical activity but the the urban realities of pollution. Knowing that our physiological responses rely on several variables, we set out to try and build an experiment that might help us isolate the embodied experienced of an individual urban dweller.

Under time and weather constraints, we wanted our work to be applicable to our peers and to help work toward proving or disproving our hypothesis. While stationary, small levels of exposure and potential elevated stress levels triggered heart rate increases, suggesting that internal, embodied responses can serve as proxies for the invisible burdens of urban life, but that filtering these independent variables out of the process requires sometimes non-obvious approaches,

These micro-level signals become powerful when placed in dialogue with macro-level patterns. By grounding biometric data in specific localities like Harlem or Columbia's campus, our anecdotal experience reveals how exposure to environmental stress varies not just by activity but by geography, infrastructure, and inequity.

The relationship between the body and the city is not new, but it is often overlooked in the design of public health interventions and environmental sensing. Our project surfaces the gap between lived experience and existing monitoring systems, which rarely account for how individuals encounter air quality differently across neighborhoods.

In asking what it would mean to embed biometric sensing into community health tools, we're not only proposing a new method but we are also challenging systems to recognize embodiment, subjectivity, and spatial injustice of urban realities. In that sense, our anecdote is not an outlier, but a lens for understanding how health is shaped at the intersection of place, policy, and perception.



Local Interactions

At the local scale, whether it's 1 meter, 10 meters, or 100 meters—the sensor installation reveals just how much micro-scale environments influence both exposure and experience. Within just a few meters, ozone levels can shift dramatically: from a courtyard to a sidewalk, from under a tree canopy to beside a bus exhaust. The 1-meter scale matters when the sensor is worn on the body; it's capturing real-time exposure at breathing height, reflecting exactly what the wearer inhales. At 10 meters, you're dealing with threshold transitions, stepping from inside to outside, crossing a street, or walking past an idling vehicle. These movements through the urban system making data interpretation more complex, and introduce key features that have an impact on the target feature we are working to understand.

By 100 meters, you begin to see neighborhood-scale variation, such as how street orientation, building density, or green space influence pollutant behavior and dissipation. In the first chapter of our experiment, movement revealed changes in heart rate, but spatial variation did demonstrate important variance in the proportion of the pollutant itself.

While causal realities at those scales are not obvious or seen in the data, and the embodied portion of pollutant exposure may not have become clear, even remaining in the same place did demonstrate that ozone, just as the urban resident, moves. And therefore movement remains key in understanding local urban experience.

Therefore, the installation isn't passive, it requires movement to function. It doesn't just record data; it invites a mode of investigation that's embodied. You become aware of how place, activity, and pollution intersect. Walking past traffic or pausing in shaded areas takes on new meaning when your heart rate responds and the ozone sensor jumps. In this way, the sensor instigates interactions—not only between human and device, but between body and environment. It reframes everyday movement as data-rich investigation, and encourages reflection on the conditions we normalize and the exposures we rarely feel until we measure them.

Individual Scale

We noticed that when remaining in the same place, but experiencing stress, the heart rate of our subject changed significantly. Conversely, when we moved around and it was hot, the subjects heart rate increased

The Local Scale

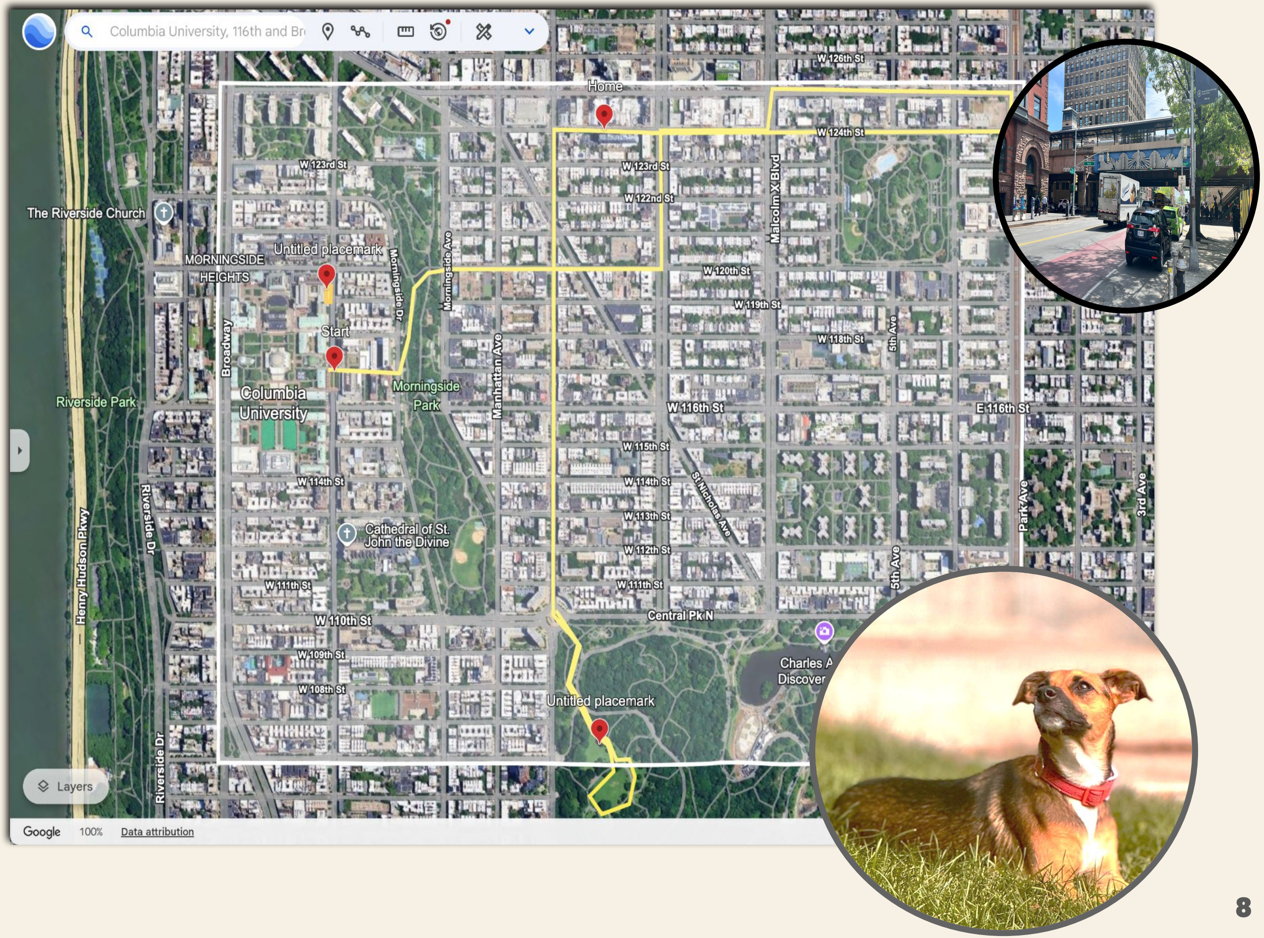
Harlem has much higher traffic rates and busy, louder streets than Columbia's campus or the North Wood of Central Park. We thought this locality would prove our hypothesis

Site Selection

Walking the route from Columbia's campus to Central Park through Morningside and East Harlem, you cross multiple environmental and social thresholds that shape exposure, often in ways that aren't visible until you're measuring the air.

On the west side, the area around Columbia and Morningside Heights is buffered by tree-lined streets, limited traffic, and academic infrastructure. Ozone still spikes here—especially near buildings and courtyards—but the environment is relatively controlled. As you move east, the landscape changes. Along 125th Street and Malcolm X Blvd, you pass wider intersections, dense bus traffic, and heat-retaining surfaces. This is where the built environment shifts—from institutional to infrastructural—and with it, the likelihood of sustained ozone exposure increases. Fewer trees, more pavement, more combustion.

By the time you hit East Harlem and approach Central Park, you're in a historically under-resourced neighborhood with higher asthma rates, greater proximity to highways and waste transfer stations, and fewer health buffers. These aren't just environmental differences—they're structural. Who gets clean air is shaped by who has historically had power to shape spaces. So when we trace ozone exposure along this path, we're not just tracking molecules in the air—we're tracing a spatial record of inequality that becomes visible, sensor by sensor, meter by meter.



Technologies – Hardware: interior

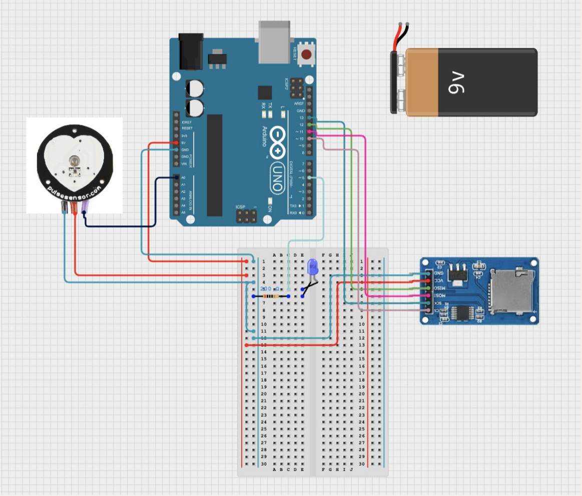
We aimed to measure two key indicators: human physiological response and environmental air quality.

Specifically, we tracked heart rate using a Pulse Sensor, a lightweight optical sensor that detects changes in blood volume through photoplethysmography.

To assess ambient ozone levels, we employed the MQ-131 Gas Sensor, which is sensitive to low concentrations of ozone and suitable for real-time monitoring.

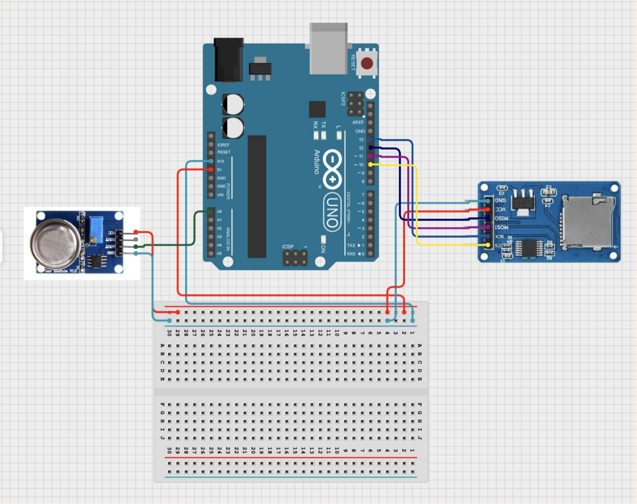
Both sensors were connected by circuits onto Arduino boards and powered with 9V batteries, creating a portable device suitable for field deployment. To capture and store the data, we connected an MD disk drive and saved all readings onto a mini SanDisk for later analysis.

This whole setup allowed us to take the device on-site for real-world testing, enabling simultaneous collection of physiological and environmental data to explore potential correlations between bodily stress responses and air quality.



Pulse Sensor

We picked this sensor for it plug and play manner. We added an LED light to help Visualize heartbeat with a blinking LED.

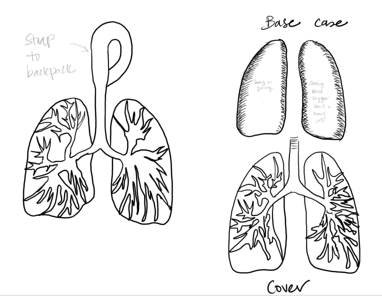


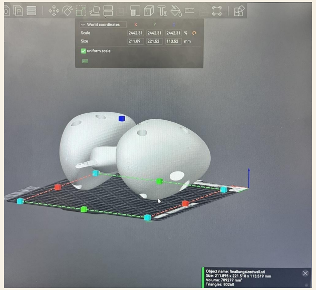
MQ-131 Sensor

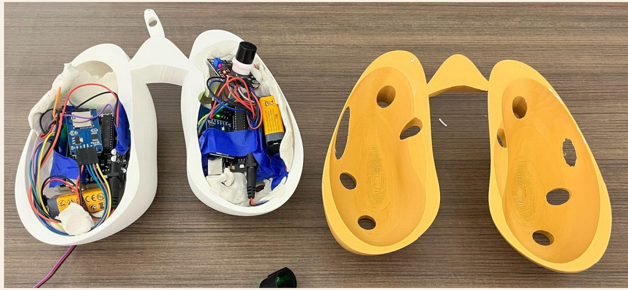
We picked this sensor as it is one that detects and measures ozone (O3) concentration in the air. We were curious about the pollution in the Harlem area as our subject lives and goes to school in the neighborhood. Additionally, this is an area with a lot of traffic and bad quality air.

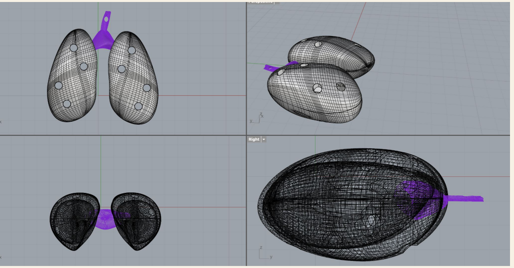
Technologies – Hardware: exterior

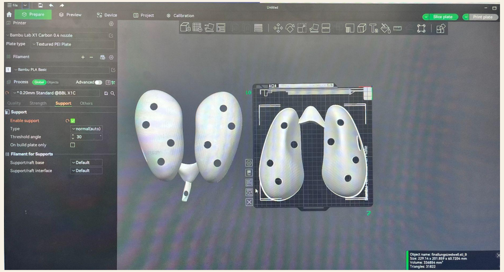
To house our sensors securely while maintaining a strong visual metaphor, we designed a custom case shaped like a pair of lungs—symbolizing the very functions we were measuring: air quality and respiratory response.











Sketch & digital model on Rhino

Set up on 3D printing Slicer software. Scale, and cut our model to fit with our needs

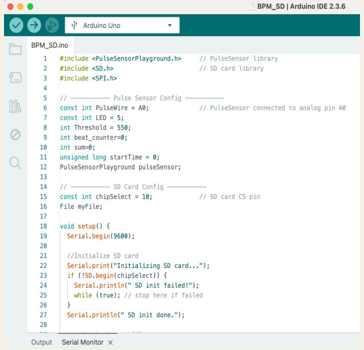
PLA filament printed 3D models, designed for adequate space & airflow for the sensors. The holes allow for ventilation and wiring.

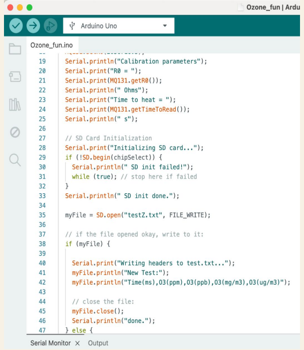
Technologies – Software

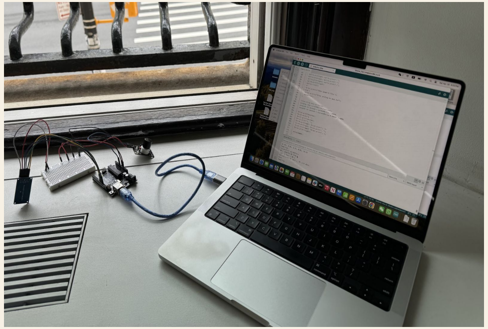
We wrote the programs for the sensors in Arduino IDE. For both of the sensors, we utilized open-sourced packages and libraries and made modifications in order to adapt the sample codes for our sensors and model. Since we implemented the 2 sensors in 2 separate arduino uno boards, we also wrote 2 separate programs for our sensors.

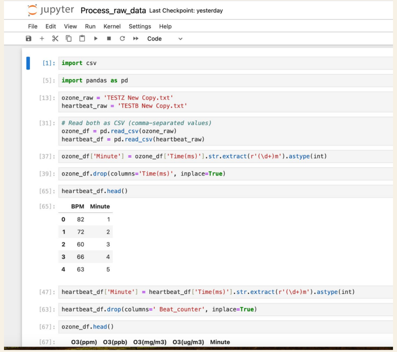
Because design of our experiment requires us to bring the model and walk outside, we were not able to monitor the data collected while conducting the experiment. The last thing we wanted is to find out the data wasn't saved properly after bring the model and walk for 3 hours. As a result, we had a lot of considerations when writing the programs in order to prevent and accommodate potential errors. For example, we implemented a LED light to the heartbeat sensor board such that the LED light will lit and indicate the program is working properly when conducting the experiment. We also put the SD initialization as the first part of both programs so that both of the sensors will only run if SD cards are initialized properly.

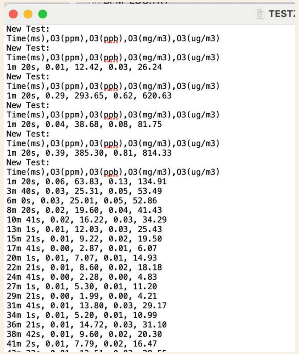
We also tested the sensors a lot of times before we actually brought them into the model and conducted the experiment. These considerations and testings helped us monitor the status of the sensors during the experiment and greatly reduced the chances of errors and failures.











Technologies – Data

The heartbeat sensors senses every 1 second, while the ozone sensor has be heat up first and then sense, which can take up to 3 minutes. In order to integrate the two sets of data together, we aggregated the data about heart rate so that we only save the BPM average for every minute. After the experiment, we joined the heartbeat data with ozone level data on the common variable which is minute. Changes in locations were manually recorded during the experiment and added to the dataset afterwards. Other notes such as climbing hills, sitting, walking faster than usual are also recorded. During our testing prior to the experiment, we also found out that the ozone sensor needs 10-15 minutes to heat up and sense the correct ozone pollution level. This is also reflected in our data cleaning process, which we discarded the data when the sensor was still heating up.

Minute	O3(ppm)	O3(ppb)	O3(mg/m3)	O3(ug/m3)	BPM	Location Change	Notes
1	0.45	451.96	0.96	955.22	82	Erma's building	Heating up sensor
3	0.03	31.52	0.07	66.63	60	Erma's building	Heating up sensor
6	0.03	31.13	0.07	65.8	68	Erma's building	Heating up sensor
8	0.04	39.19	0.08	82.83	71	Erma's building	Heating up sensor
10	0.05	50.69	0.11	107.14	73	Erma's building	Heating up sensor
13	0.06	55.98	0.12	118.31	65	Erma's building	Heating up sensor
15	0.05	54.4	0.11	114.97	77	Erma's building	Heating up sensor
17	0	2.46	0.01	5.19	81	Left Erma's building	
20	0.01	5.15	0.01	10.89	112	125th st. to the east	
22	0	2.46	0.01	5.19	105		
24	0	2.22	0	4.7	81		
27	0	2.35	0	4.96	119		
29	0	1.18	0	2.5	97	125th st. & 5th Ave	
31	0	0.58	0	1.22	104		
34	0	2.39	0.01	5.05	70	125th st. & Park Ave	
36	0	1.62	0	3.41	106	124th st.	
38	0	2.26	0	4.79	102		
41	0	3.71	0.01	7.85	124	124th st. & Park Ave	

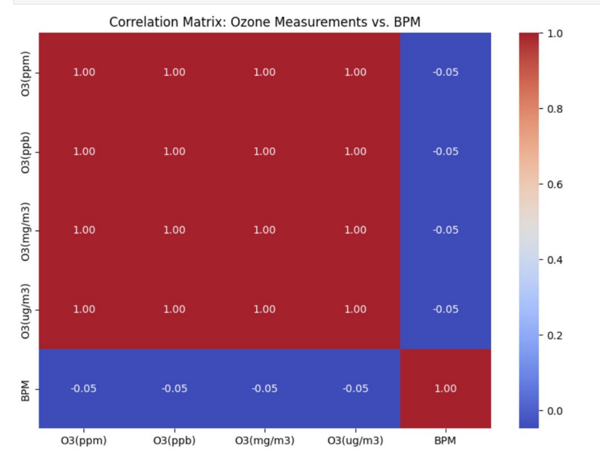
Method For Data Analysis

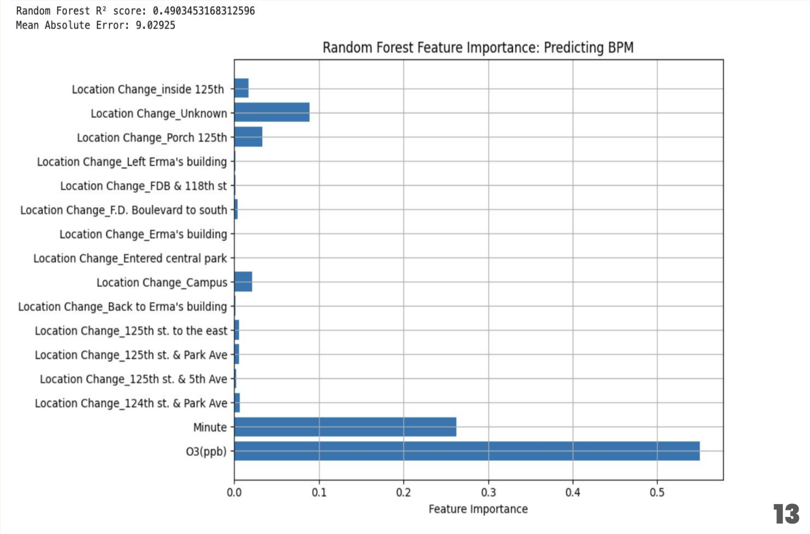
We started with a basic question: is there any direct relationship between ozone and heart rate? A quick correlation test said no. The value was close to zero ($r = -0.05$), and a linear regression backed it up. Ozone alone wasn't doing anything to explain changes in BPM.

Still, we tested whether the relationship might be non-linear. A polynomial regression gave us a better fit ($R^2 \approx 0.26$), but it overreacted to extreme values and didn't hold up to scrutiny. So we moved to a Random Forest model. That worked better, with an R^2 of about 0.63. It showed there was a pattern, but likely one tied to movement or environment rather than a direct physiological response.

To explore that, we built an OLS model with interaction terms: ozone by location and ozone by time. The results improved, especially for people who were moving ($R^2 \approx 0.49$), and dropped off for stationary cases ($R^2 \approx 0.11$). It became clear the model was picking up on patterns in behavior, like where and when people were walking, rather than any consistent effect of ozone on heart rate.

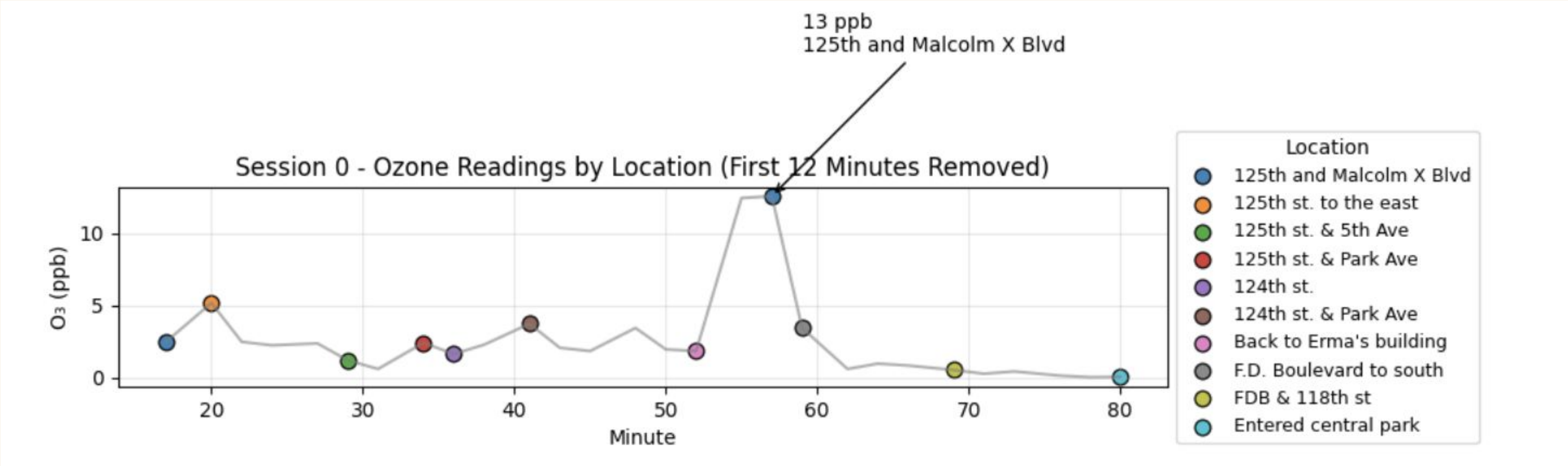
So we tested that idea directly. We tagged certain locations as stationary and split the data. In those still points, the relationship between ozone and BPM completely disappeared. When people were moving, we saw a small negative trend, probably tied to exertion or environment. The takeaway is that ozone tracks with behavior, not biology. What really matters is context: movement, place, and time. That's what the models are responding to.





Spatial Findings

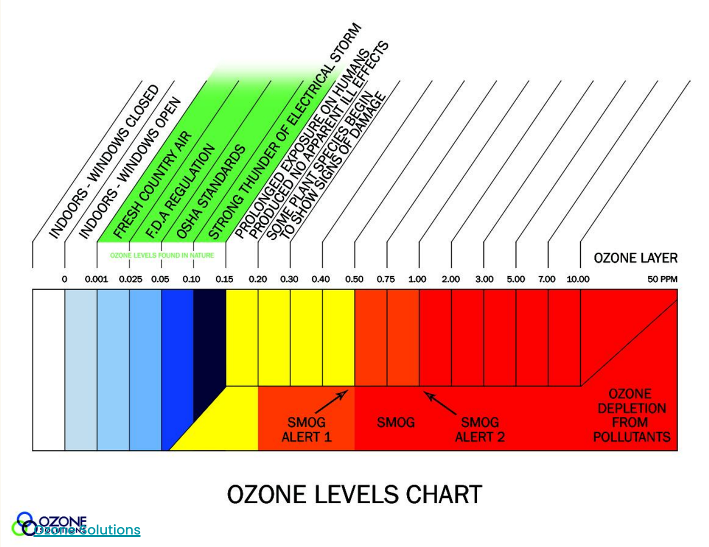
This chart shows ozone levels over time during Session 1, the session we spent walking around the neighborhoods. Each point is color-coded by location. A clear peak of ~15 ppb occurs at minute 57 as we are heading are one of the busiest streets in harlem, 125th and Malcolm X (Lenox). Shortly after as we turn back away from east harlem and toward central Harlem, ozone levels drop sharply and remain low across subsequent locations, including areas like 125th Street and F.D. Boulevard, which is west of Erma's building and heads closer to central park and Columbia's Morningside campus. Both of these locations have lower traffic rates and significantly more green space. The trend line highlights this rapid decline, suggesting a strong shift in environmental conditions based of neighborhood. Suggesting embodied racial environmental injustice,

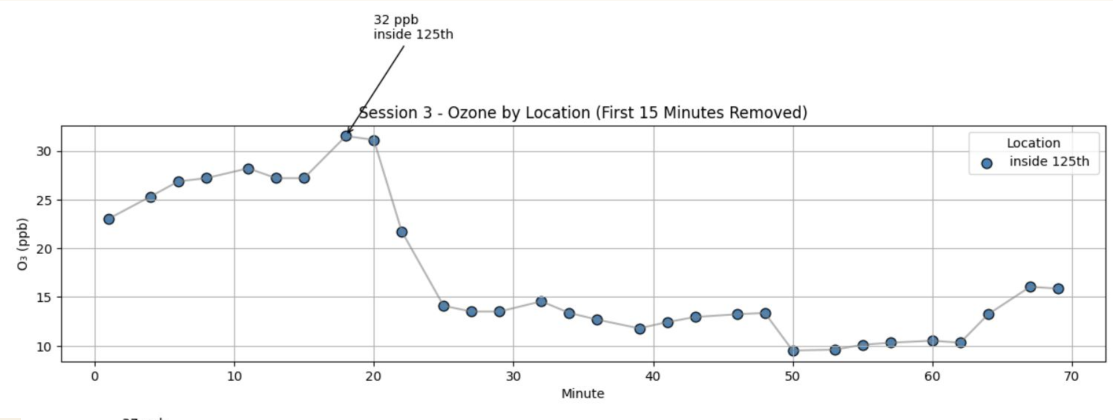


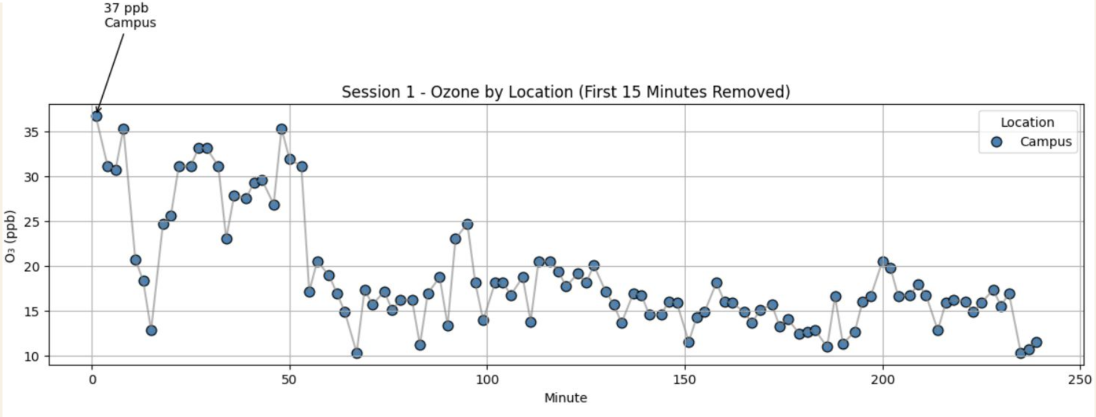
Ozone and Location

When our subject sat in the same location, it because clear that ozone fluctuates in the air rapidly over time.

Interesting the other thing that became very obvious, is that when the rain began to fall in both instances, the levels began to fall significantly.





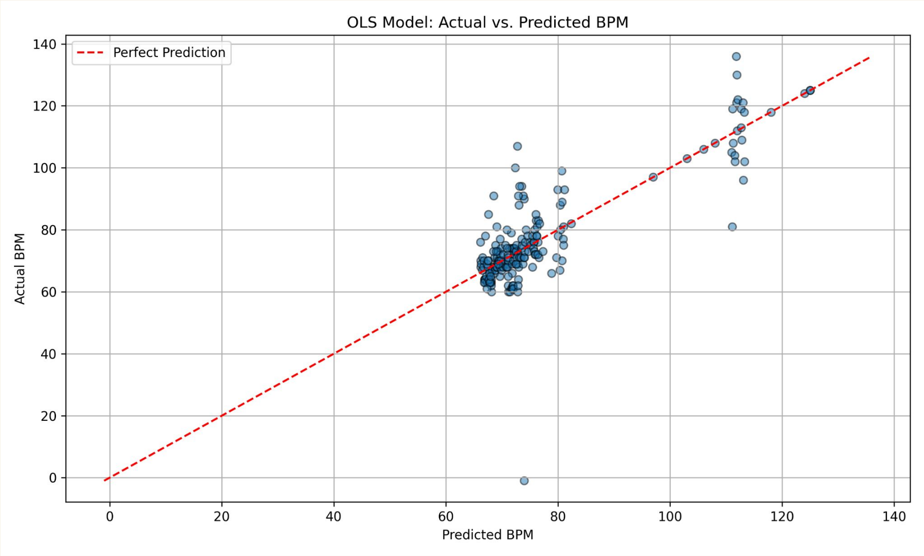


Initial Bodily Interaction Findings

This scatter plot compares the actual heart rate (BPM) to the predicted BPM generated by an Ordinary Least Squares (OLS) regression model. Each point represents a single observation, and the dashed red line indicates a line of perfect prediction, where predicted and actual BPM values would be equal.

The clustering of points around this line suggests the model is moderately accurate, especially within the 60–90 BPM range where most data points fall.

However, some deviation from the perfect prediction line is visible, particularly at higher BPM values where the model begins to underpredict actual heart rates. This indicates the model may struggle with extreme or outlier readings. But it also made us wonder, could this indicate that something more complex is going on with the data?



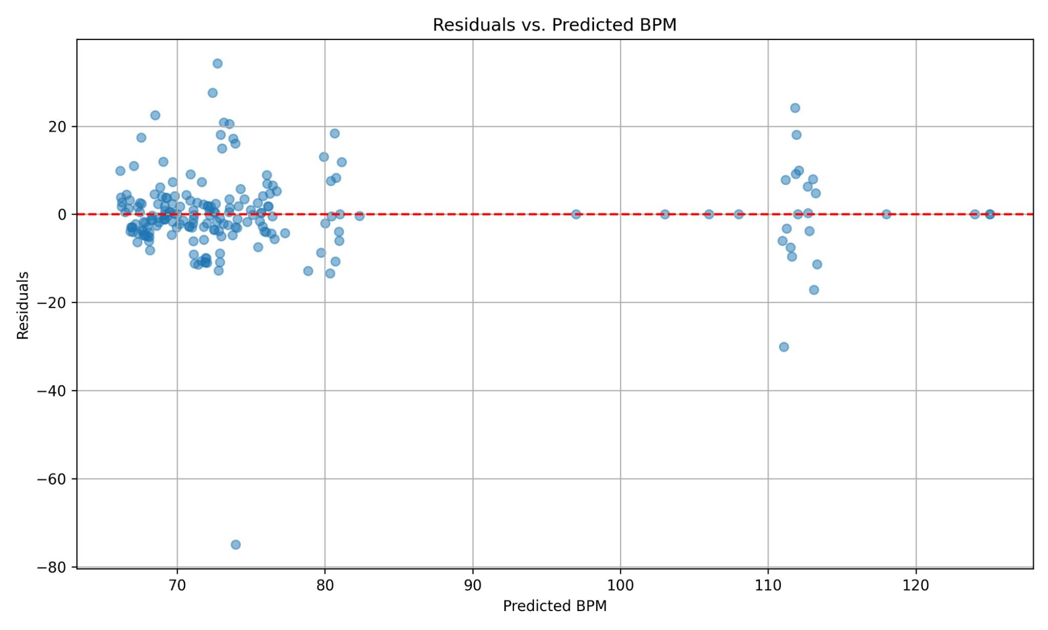
Error Analysis

So we decided to look at the errors occurring in the model.

This residuals plot shows the difference between the actual BPM and predicted BPM from the OLS model, plotted against the predicted BPM values. Each point represents one observation, and the red dashed line at zero represents perfect predictions—where the residual (error) is zero.

What we observe here is that for most predictions between 65 and 80 BPM, the residuals are fairly centered around zero, indicating reasonably accurate predictions. However, as predicted BPM increases beyond 100, the residuals begin to fan out and skew, suggesting that the model becomes less reliable at higher heart rates. Additionally, there's a non-random pattern in the spread, especially a dense cluster of underestimated and overestimated points around 70–80 BPM, which may indicate heteroscedasticity (i.e., the variance of errors changes with the prediction) or missing explanatory variables that influence BPM differently at high vs. low ranges.

This plot reveals that while the model performs adequately in the middle range of BPM, it struggles to generalize well at the extremes, hinting that refinements or additional features (like activity state or environmental lag) could improve predictive accuracy.

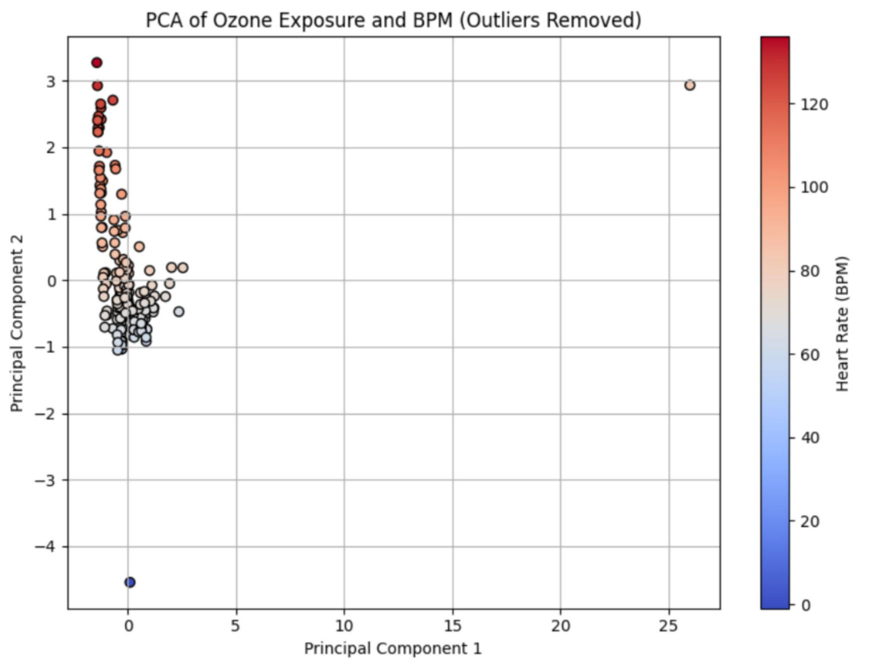


PCA Analysis

This PCA plot shows how ozone exposure and heart rate (BPM) vary together, with outliers removed for clarity. Each point is an observation, colored from blue (low BPM) to red (high BPM), and the axes represent the two main patterns of variation in the data.

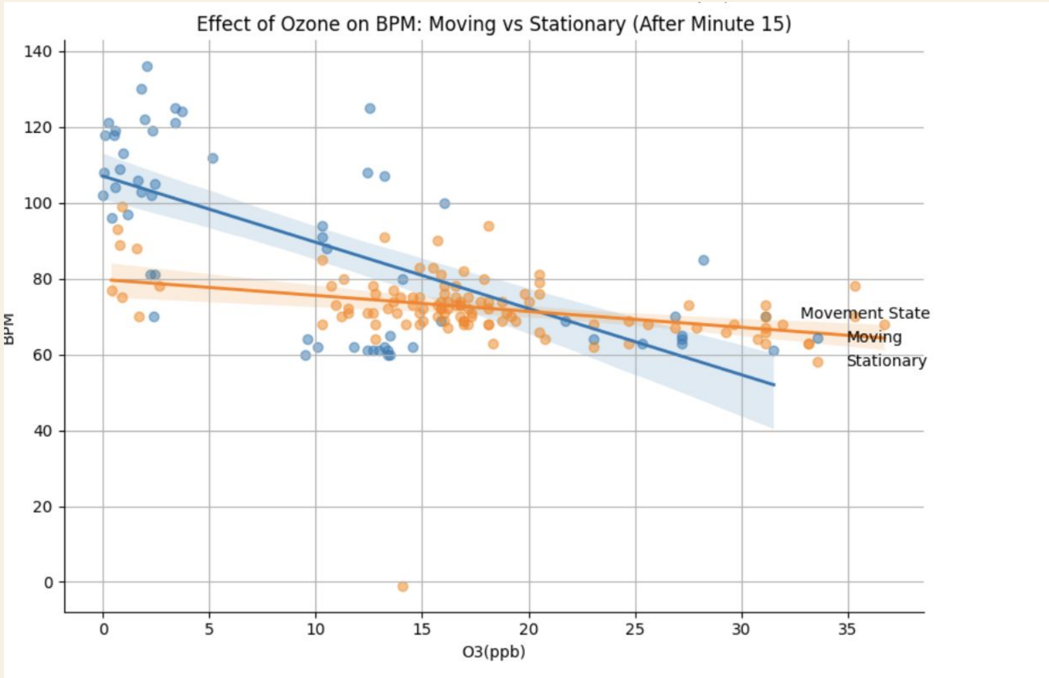
Most points cluster in the lower left, suggesting similar ozone–BPM profiles across observations. A few outliers, especially one far right, reflect rare combinations like unusually high BPM or atypical ozone.

The fact that BPM doesn’t vary clearly along either axis implies that heart rate is shaped by more than ozone alone, reinforcing the idea that movement or context plays a larger role.



Correlation not Causation Discovered

- For moving individuals (blue), there's a clear negative slope, as ozone levels increase, BPM tends to decrease. This may reflect behavioral or environmental factors (e.g., slowing down due to discomfort or location), not necessarily a physiological ozone response.
- For stationary individuals (orange), the regression line is almost flat, indicating no meaningful relationship between ozone and BPM when people aren't moving.
- The confidence bands around the moving line are wider, reflecting more variability in that state—likely due to movement intensity, location changes, or effort level.
- R2 score = 0.445

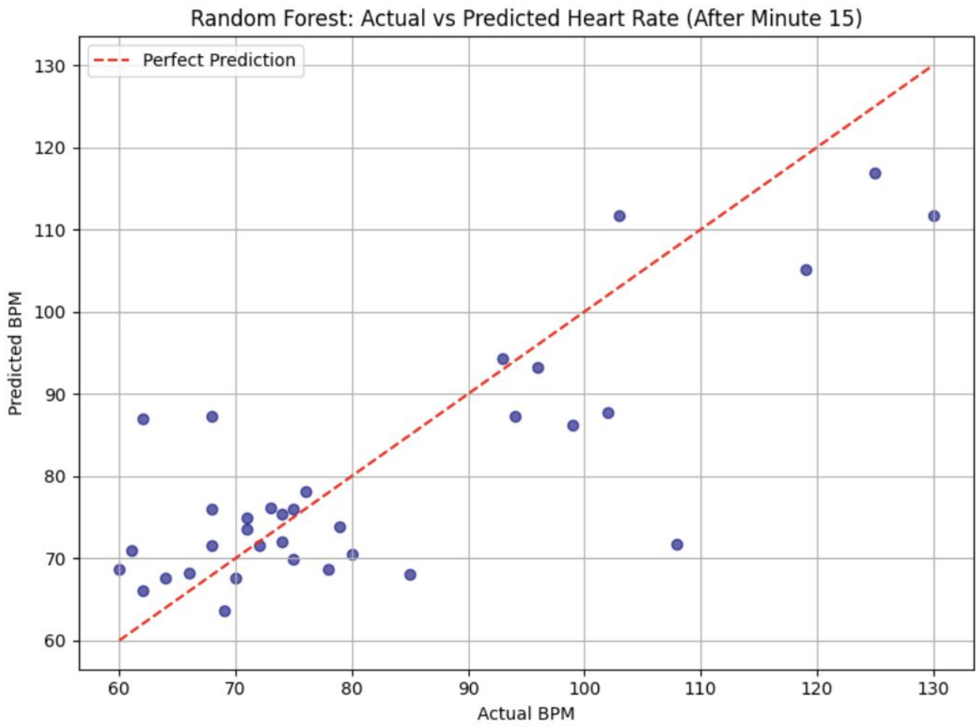


Wait wait wait... there is a potential issue here...

When we filtered out data to disclude the ozone ‘heat up’ period (15 minutes), the Random Forest model all of a sudden started performing better than the OLS model.

Is this because the random forest model is muting all the other variables?

now our model is showing there is a pattern between heart rate and ozone... BUT It’s more likely that ozone is correlated with behavior or context, rather than directly causing BPM changes. That’s still important! It means ozone exposure and physical state co-occur in a meaningful, mappable way.



RMSE: 11.22688143472466
R^2 Score: 0.633986570903252

Experiment Conclusion

Our experiment set out to explore how ozone pollution affects heart rate in real-world, embodied urban settings. Through a wearable sensor system, we collected and analyzed environmental and biometric data across multiple locations and movement states. What we found wasn't a clear cause-and-effect link between ozone and BPM, but something more nuanced.

Heart rate was influenced less by ozone itself and more by *where* and *how* people moved. When we controlled for movement and location, the ozone, BPM relationship disappeared. This revealed that environmental context, not ozone in isolation, is what truly shapes physiological response. Our project doesn't just refute a hypothesis; it reframes the question. The body is a sensor, and it tells us more when we listen across time, place, and motion.

Urban Interactions & Forward Looking

If ozone and heart rate sensors like these were deployed pervasively across a city, we could move beyond static air quality maps to create a dynamic, real-time understanding of how environmental conditions affect people in motion. Instead of relying solely on fixed monitors or average pollution levels, we'd gain insight into how individuals experience exposure differently depending on where they go, how they move, and when they're active. This would allow cities to visualize health impacts as lived, embodied experiences—not just concentrations on a map.

The key difference this approach highlights is the gap between ambient exposure and personal exposure. While citywide monitors may report air quality as "moderate," wearable data reveals how certain micro-environments, like building courtyards, subway entrances, or street canyons, can produce sudden, harmful spikes in ozone or other pollutants. At the same time, heart rate data shows that these exposures don't impact everyone the same way; physiological response varies depending on movement, stress, or even just where someone happens to stand for a few minutes.

The opportunity here is profound: policymakers could target interventions more precisely, improve urban design to minimize high-exposure zones, and empower individuals with real-time feedback to make healthier decisions. But challenges remain, data privacy, calibration across different devices, and making sense of vast, context-dependent datasets are real hurdles. Still, the imagined future is one where environmental health becomes personal and actionable, with cities shaped not just by where pollution is, but by how people feel it.

The Urban Scale

We pose the question: What if biometric sensing was built into community health monitoring? Could healthy equity be significantly improved?

Thank You

Steven Shi, fs2860@columbia.edu
Jui Shih, js6676@columbia.edu
Erma Swartz, es4284@columbia.edu