2024

# Investigating Factors Contributing to Differences in Air Pollutants Between Metropolitan Areas in the US

Ellen Hu

# Investigating Factors Contributing to Differences in Air Pollutants Between Metropolitan Areas in the US

A Thesis Presented

by

Ellen Hu

To the Keck Science Department

Of Claremont McKenna, Pitzer, and Scripps Colleges in partial fulfillment of

the degree of Bachelor of Arts

Senior Thesis in Environmental Analysis

4th of December 2023

**Table of Contents**

**Abstract**

Densely populated regions around the United States of America exhibit varying concentrations of air pollutants that can impact the health of urban populations and surrounding ecosystems. Understanding why these differences exist can be important in addressing health issues as the human population continues to increase and communities across the US move to accommodate these trends. To investigate the impact of natural processes and anthropogenic sources on pollutant concentrations, daily data from the Environmental Protection Agency's (EPA's) publicly available database was downloaded for fine particulate matter ($PM_{2.5}$), nitrogen dioxide ($NO_2$), and ozone ($O_3$) pollutant concentrations from seven metropolitan regions across the US for the years 2018-2022 with considerations for population density and pollution rankings. Metropolitan locations of study include San Francisco (SF), Los Angeles (LA), New York (NY), Houston, Chicago, Atlanta, and Phoenix. A Spearman Rank-Order Correlation test was conducted between air pollution data and weather data, including precipitation, wind speed, and average temperature, from National Oceanic and Atmospheric Administration (NOAA). Further Spearman Rank-Order Correlation tests were conducted between air pollution data and vehicle miles travelled data. While precipitation is weakly correlated with all pollutants, generally strong positive correlations between $O_3$ and temperature in addition to a strong negative correlation between wind speed and $NO_2$ was observed for all regions. Many outliers were observed for $PM_{2.5}$ concentrations in SF, possibly the result of wildfire burning in California during the fall months. Further inquiry into the impact of anthropogenic sources utilized vehicle miles travelled (VMT) data acquired from Streetlight Data LLC for January 1, 2020 through June 30, 2020 in SF and LA. A Spearman-Rank Order correlation test found a relatively strong positive correlation with $NO_2$ for SF, but generally weak or nonsignificant correlations between vehicle travel and pollutants in LA. Overall, low r-values for all correlations indicate that other factors may be contributing to differences in pollutant concentrations between metropolitan regions.

## 1. Introduction

### *1.1 Metropolitan and Urban Expansion*

While metropolitan areas make up less than 20 percent of US land area, approximately 80 percent of the US population resides in these regions (Auch et al., 2004). They consist of a city proper and additional counties, often suburbs, that have high economic contributions to the inner city. The expansion of metropolitan areas can be traced back to the end of World War II where improved highway systems and transportation networks led to a rise in suburban neighborhoods. Extensive movement occurred during this period with US census information from the 1970s indicating that more of the American population lived in the suburbs than in central urban areas (Auch et al., 2004). Railways and transportation accessibility in the present day have maintained an important role in urban expansion, but cheaper living expenses further away from urban centers have incentivized relocation where commuting is possible (Kheyroddin & Ghaderi, 2023). This has been the trend since the year 2000, as suburban population growth has outpaced the national overall population growth and urban center population growth (Fry, 2020). Changes in air quality rise alongside this urban growth – increased human activity leads urban areas to exhibit higher concentrations of air pollutants when compared to natural environments or areas that are not as developed (Ling et al., 2011).

### *1.2 Air Pollution*

Previous studies have extensively investigated the harmful impacts of air pollution on both ecosystem and human health. Within an urban context, plants and other agricultural crops experience increased selective stress that can reduce genetic diversity (Bell et al., 2011). This reduction in biodiversity can have implications for local ecosystem health and the availability of

ecosystem services for human populations such as erosion management and pathogen regulation (Quijas et al., 2010). While these ecological impacts contribute to human health, exposure to air pollution also has more direct impacts on population health. Thorough research has been conducted on the area of study in recent years and it is estimated that approximately 6.7 million premature deaths are associated exposure to air pollution (*Household Air Pollution*, n.d.). Many results from these studies have drawn strong connections between increased risk of mortality in addition to heart and lung diseases through blood contamination, resulting in a long-term mild hypoxia state (Zhao & Ma, 2021). The characteristics of individual pollutants result in varying impacts on human health, although their combined presence can contribute to a higher overall risk of declining health over time.

### *1.3 Air Pollutants Studied*

The study of air pollution often prioritizes a set of six principal pollutants, deemed criteria pollutants, that are identified by the Environmental Protection Agency (EPA) as most dangerous and important to monitor (Axelrad, et al., 2013). This study analyzes three pollutants from the EPA's list that are particularly prevalent within urban environments due to anthropogenic activity: nitrogen dioxide ($NO_2$), fine particulate matter ($PM_{2.5}$), and ozone ($O_3$).

Nitrogen oxides ($NO_x$) are a family of highly reactive gasses that primarily take the form of either NO as it is emitted through the combustion of fossil fuels, or the resulting $NO_2$ produced when NO reacts with free radicals in the atmosphere. The pollutant is primarily derived from vehicle emissions and other industrial activity, and as a result concentrations tend to be higher in more populated regions or areas adjacent to highways. Classified as a primary pollutant due NO's direct derivation from combustion, $NO_x$ can react in the lower atmosphere to form

ground level ozone or $PM_{2.5}$. This process can lead to a buildup of $NO_x$ in the atmosphere. It has been linked to a positive correlation with the presence of mental health disorders and long-term exposure can lead to increased risk of cardiovascular and respiratory diseases, asthma, and lung cancer (Shaw & Van Heyst, 2022). These health impacts can be further exacerbated by $NO_2$'s contributions to the formation of ozone.

Tropospheric ozone, also known as ground-level ozone, is a secondary pollutant of $NO_x$ formed through a series of chemical reactions in the lower atmosphere. This process involves the photodegradation of $NO_x$ in a reaction with volatile organic compounds (VOCs), so the presence of heat and sunlight are essential. Unlike stratospheric ozone that filters the sun's ultraviolet (UV) rays, ground-level ozone can lead to a myriad of harmful health impacts. As a principal component of photochemical smog, high concentrations of ozone can inhibit visibility and act as an irritant for the eye and nose (Zhang et al., 2019). Additionally, past studies have indicated that ozone can cause oxidative damages to cells leading to increased risk of asthma and other immune-inflammatory responses (Zhang et al., 2019). Further contributions to these health impacts come in the form of particulate matter that can exist alongside the chemically formed ozone.

Fine particulate matter, often written as $PM_{2.5}$, is classified as airborne particles that are 2.5 μm or smaller in diameter in either a liquid or solid form. These particles are particularly dangerous because their small size can lead to direct entry into the bloodstream when inhaled. As a result, they are closely monitored by the EPA due to the variety of associated cardiovascular and cognitive health concerns. Exposure to $PM_{2.5}$ has been shown to contribute towards mortalities related to strokes, acute lower respiratory infections, ischemic heart disease, lung cancer, and chronic obstructive pulmonary disease (Song et al., 2017). Furthermore, long-term

exposure to $PM_{2.5}$ has been linked to a decline in cognitive function in adults above 50 years of age (Ailshire & Crimmins, 2014). These particles are generated through a variety of anthropogenic and natural sources. Major anthropogenic sources include vehicle activity, dust, industry, biomass burning, and coal combustion (Lv et al., 2016). Yet, biomass burning can also occur naturally in the event of wildfires. Shifts in climatic conditions due to global climate change has led to a 10-fold increase in the amount of area burned within the US since the 1980s alongside increased frequency of wildfire occurrences (Wang et al., 2023). It is often acknowledged that $PM_{2.5}$ pollution is one of the largest invisible threats to human health on a shorter time scale, and as a result, government agencies around the world have heavily focused on reducing this form of pollution.

As mandated by the Clean Air Act, the EPA assigns each of these pollutants a primary and secondary standard which aim to protect against the previously discussed human and environmental health concerns (Axelrad, et al., 2013). The annual standards for $PM_{2.5}$, $NO_2$, and ozone are 15 $\mu g/m^3$, 53 ppb, and 0.7 ppm respectively (US EPA, 2016; US EPA, 2020; US EPA, 2016). In an attempt to analyze and maintain concentrations below these standards, the government agency has established a network of air quality monitor stations across the US equipped with air quality monitoring instruments. The number and type of recorded pollutant differ amongst these stations, but all recorded data is made publicly available through an online database. While the database provides outputs for analysis, differences in air pollution mixtures across regions of the US highlight how it is difficult to analyze pollutants across such a wide geographic range (Requia et al., 2019). It is generally agreed that weather impacts air pollution, but there is a lack of extensive research into the mechanisms of these relationships and the extent to which they impact one another. Previous studies have examined weather factors including

temperature, humidity, and wind with findings that support temporal and spatial differences in pollutant concentrations (Pleijel et al., 2016). Yet, local differences in this weather, topography, industrial activity, and many other competing factors contribute towards complexities in the area of study.

### *1.4 Project Aims*

This study aims to utilize air pollution, weather, and vehicle data to address how a variety of factors contribute to differences in air pollutant concentrations between US metropolitan areas. The analysis seeks to uncover possible correlations that can point towards further areas of study rather than provide an extensive review that accounts for confounding variables and insight into causal relationships. Pollutant concentrations between seven metropolitan regions with high population density in the US will be compared across years and seasons as to address how a large portion of the US population will be affected. Following this initial identification of trends, correlations between pollutant concentrations and weather from each region will be analyzed to examine the impact of natural hydrologic cycles. Anthropogenic patterns are introduced through correlation analysis between vehicle data and pollutants for two regions in California.

These findings may have important implications as population density increases. It is expected for the human population to grow by approximately two billion people over the next 30 years, and it is likely that continued urban expansion will follow (United Nations, n.d.). Understanding the impact of different factors on air pollutants can have implications for policy change regarding public health. This can be essential in metropolitan areas where dense populations can be exposed to hazardous conditions.

## 2. Materials and Methods

### *2.1 Selection of Metropolitan Areas*

For this study, seven metropolitan areas from across the United States were selected and analyzed (Table 1). The areas include San Francisco (SF), Los Angeles (LA), New York (NY), Houston, Chicago, Atlanta, and Phoenix (Fig. 1). Each area was chosen due to a combination of considerations including population density according to the 2022 US Census, geographic coverage, and the American Lung Association's rankings of the worst-polluted cities (*Most Polluted Cities | State of the Air*, n.d.). Counties recognized as members of the selected metropolitan regions of study were identified using US Census information through Census Reporter (Census Reporter: https://censusreporter.org) and the number of monitor stations for each pollutant within each county were recorded (Table 2). This information was utilized to download air pollution, weather, and VMT data for the study.

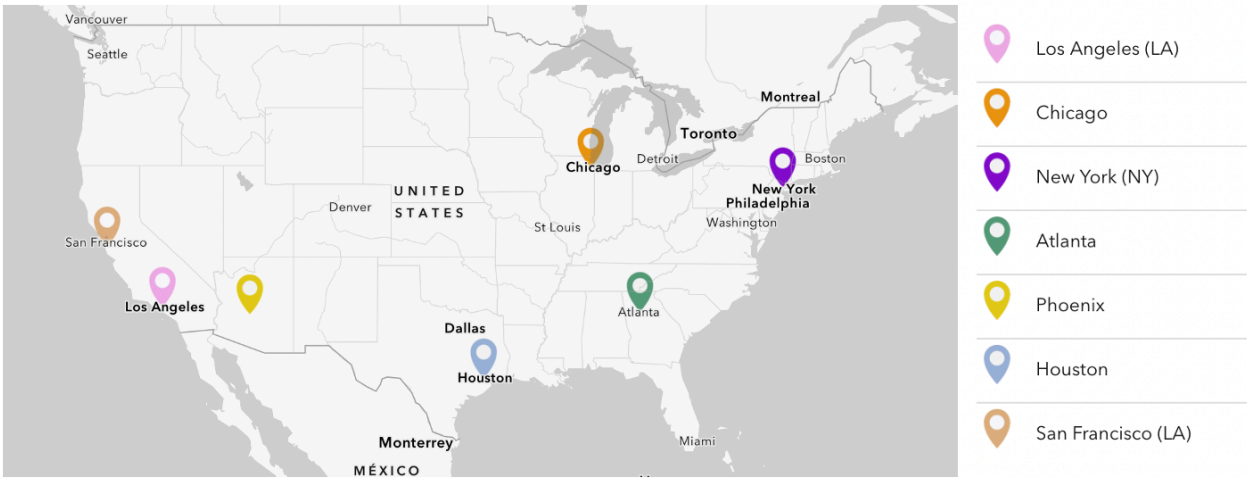


**Figure 1.** Map of 7 metropolitan areas within the US selected for study.

**Table 1. Demographic information for 7 metropolitan areas of study.** N/A indicates no associated information.

| Metropolitan Area | State(s) | Number of Counties | Area (miles$^2$) | Population (millions) | American Lung Association Rankings |
|---|---|---|---|---|---|
| **SF** | CA | 9 | 2,470.3 | 4.58 | #11 by Ozone<br>#7 by Year-Round Particle Pollution<br>#6 by Short-Term Particle Pollution |
| **LA** | CA | 3 | 4,852.1 | 12.87 | #1 by Ozone<br>#4 by Year-Round Particle Pollution<br>#9 by Short-Term Particle Pollution |
| **NY** | NY, NJ, CT | 12 | 6,684.4 | 19.62 | #12 by Ozone |
| **Houston** | TX | 9 | 8,268.8 | 7.34 | #9 by Ozone<br>#15 by Year-Round Particle Pollution |
| **Chicago** | IL, IN, WI | 14 | 7,194.9 | 9.44 | #23 by Year-Round Particle Pollution |
| **Atlanta** | GA | 11 | 8,685.7 | 6.22 | N/A |
| **Phoenix** | AZ | 2 | 14,568.7 | 5.02 | #5 by Ozone<br>#7 by Year-Round Particle Pollution<br>#13 by Short-Term Particle Pollution |

**Table 2. Counties and data availability in 7 metropolitan areas of study.** Crosses (x) indicate available data, N/A showing non-available data between 2018-2022 for air pollutants and the impacting factors, and italicized values indicate available stations with significant gaps in data.

| Region | County | $PM_{2.5}$ | $NO_2$ | $O_3$ | VMT | Temperature | Rainfall | Wind |
|---|---|---|---|---|---|---|---|---|
| **SF** | Alameda County | 6 | 6 | 5 | × | 4 | 32 | 3 |
| | Contra Costa County | 2 | 4 | 4 | × | 6 | 30 | 1 |
| | San Francisco County | 1 | 1 | 1 | × | N/A | 7 | N/A |
| | San Mateo County | 1 | 1 | 1 | × | 4 | 27 | 1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Marin County | 2 | 1 | 1 | × | 3 | 28 | N/A |
| | Santa Clara County | 3 | 2 | 4 | × | 3 | 36 | 2 |
| | Solano County | 1 | 1 | 3 | × | N/A | 13 | 1 |
| | Sonoma County | 1 | 1 | 2 | × | 2 | 83 | 1 |
| | Napa County | 2 | 2 | 2 | × | N/A | 10 | 1 |
| **LA** | Los Angeles County | 15 | 16 | 15 | × | 23 | 75 | 11 |
| | Orange County | 2 | 3 | 3 | × | 2 | 45 | 2 |
| | Ventura County | 5 | 2 | 5 | × | 6 | 21 | 3 |
| **NY** | Kings County | 3 | N/A | N/A | | N/A | 1 | N/A |
| | Queens County | 3 | 2 | 1 | | 2 | 10 | 2 |
| | New York County | 5 | N/A | 1 | | N/A | 5 | 2 |
| | Bronx County | 4 | 2 | 2 | | N/A | 1 | N/A |
| | Richmond County | 2 | N/A | *1* | | N/A | 3 | N/A |
| | Westchester County | 1 | N/A | *1* | | N/A | 13 | 2 |
| | Bergen County | *2* | 1 | *1* | | N/A | 16 | 1 |
| | Hudson County | 2 | 2 | 1 | | N/A | 5 | N/A |
| | Passaic County | *1* | N/A | *1* | | N/A | 15 | N/A |
| | Putnam County | N/A | N/A | *1* | | N/A | 5 | N/A |
| | Orange County | 1 | N/A | *1* | | N/A | 10 | 1 |
| | Rockland County | 1 | N/A | *1* | | N/A | 3 | N/A |
| **Houston** | Austin County | N/A | N/A | N/A | | N/A | 16 | N/A |
| | Brazoria County | N/A | 2 | 2 | | 2 | 17 | 3 |
| | Chambers County | N/A | N/A | N/A | | 1 | 4 | N/A |
| | Fort Bend County | N/A | N/A | N/A | | N/A | 20 | 1 |
| | Galveston County | 1 | 1 | 1 | | 1 | 44 | 1 |
| | Harris County | 10 | 14 | 17 | | 2 | 80 | 3 |
| | Liberty County | N/A | N/A | N/A | | 2 | 10 | N/A |
| | Montgomery County | 1 | 1 | 1 | | 2 | 50 | 1 |
| | Waller County | N/A | N/A | N/A | | N/A | 2 | N/A |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Chicago** | Cook County | 12 | 6 | *10* | | 1 | 105 | 3 |
| | DuPage County | 1 | N/A | *1* | | 1 | 57 | 2 |
| | Lake County, IN | 7 | 1 | 2 | | N/A | 36 | N/A |
| | Lake County, IL | N/A | N/A | *1* | | N/A | 44 | 1 |
| | Will County | *2* | N/A | *1* | | N/A | 66 | 1 |
| | Kane County | 2 | N/A | *1* | | N/A | 44 | 1 |
| | McHenry County | 1 | N/A | *1* | | N/A | 36 | N/A |
| | Kenosha County | 1 | *1* | *2* | | N/A | 19 | 1 |
| | Porter County | 1 | N/A | 2 | | 1 | 40 | 1 |
| | Kendall County | N/A | N/A | N/A | | N/A | 21 | N/A |
| | DeKalb County | N/A | N/A | N/A | | N/A | 16 | N/A |
| | Grundy County | N/A | N/A | N/A | | N/A | 15 | N/A |
| | Jasper County | N/A | N/A | N/A | | N/A | 18 | N/A |
| | Newton County | N/A | N/A | N/A | | N/A | 18 | N/A |
| **Atlanta** | Cherokee County | N/A | N/A | N/A | | N/A | 8 | N/A |
| | Clayton County | 1 | N/A | N/A | | 1 | 4 | 1 |
| | Cobb County | 1 | N/A | *1* | | N/A | 21 | 1 |
| | DeKalb County | 1 | 2 | 1 | | N/A | 22 | 1 |
| | Douglas County | N/A | N/A | *1* | | N/A | 4 | N/A |
| | Fayette County | N/A | N/A | N/A | | N/A | 8 | N/A |
| | Forsyth County | N/A | N/A | N/A | | N/A | 12 | N/A |
| | Fulton County | 3 | *1* | *1* | | N/A | 22 | 1 |
| | Gwinnett County | *1* | N/A | *1* | | N/A | 26 | 1 |
| | Henry County | 1 | N/A | *1* | | N/A | 9 | N/A |
| | Rockdale County | N/A | N/A | *1* | | N/A | 3 | N/A |
| **Phoenix** | Maricopa County | 12 | 7 | 24 | | 1 | 247 | 4 |
| | Pinal County | 4 | N/A | 6 | | 1 | 49 | N/A |

*2.2 Data Acquisition: Air Pollution Data*

Air pollution data for each of the selected areas were downloaded from the EPA's public online database for the years 2018-2022 (EPA: https://www.epa.gov/outdoor-air-quality-data/download-daily-data). This database, known as the Air Quality System (AQS), is composed of data collected from government-funded air quality monitors around the country with gaps in available daily pollutant data filled by AirNow monitors (AirNow: https://www.airnow.gov/). EPA monitors contributing data to the AQS operate on large-scales due to the quality and capabilities of the filtering instruments, yet the same filtration data can be collected through small-scale operations which will be outlined in the appendix (Section III). Particulate matter ($PM_{2.5}$), nitrogen dioxide ($NO_2$), and ozone ($O_3$) were available for all regions of study and were downloaded for all available counties. Downloaded data were identified as the following: daily mean $PM_{2.5}$ concentration ($\mu g/m^3$), maximum 1-hour $NO_2$ concentrations reported daily (parts per billion, ppb), and maximum 8-hour ozone concentrations reported daily (parts per million, ppm). Daily concentrations for each of the pollutants within the given years of study were downloaded for all available monitoring stations for the identified counties. Missing or incomplete data from counties was recorded. Mean daily pollutant concentrations were calculated for each metropolitan region by averaging the concentrations across monitoring stations from each county for the corresponding area of study in R Studio. These singular pollutant concentration values associated with each day from January 1, 2018 to December 31, 2022 were input into a CSV file that allowed for further analysis.

*2.3 Data Acquisition: Weather Data*

Weather data for each of the selected areas was downloaded from the US National Oceanic and Atmospheric Administration's (NOAA) public online database, Climate Data Online (CDO) for the years 2018-2022 (CDO: https://www.ncei.noaa.gov/cdo-web/search;jsessionid=DE94D380A5ABAF961A0159D5BD07B18F). Daily average wind speed (m/sec), precipitation (mm), and average temperature (degrees Celsius) were available for all regions of study and downloaded for all available counties. Missing or incomplete data from counties was recorded. Mean daily values for each factor were calculated for each metropolitan region by averaging the recorded values across monitoring stations from each county for the corresponding area of study in R Studio (Section IV). Once a singular value associated with each day from January 1, 2018 to December 31, 2022 was calculated for each of the weather factors, it was input into a CSV file that allowed for further analysis.

*2.4 Data Acquisition: VMT Data*

Vehicle miles travelled (VMT) data were acquired for the two California metropolitan areas, LA and SF, from Streetlight Data LLC. The company uses machine learning alongside vehicle monitors to calculate vehicle travel per county on a daily basis. Thus, the provided travel data is calculated at a county-level from January 1, 2020 to June 30, 2020. Mean daily values for LA and SF were calculated by averaging the provided VMT values across corresponding counties for the overall metropolitan area in R Studio. Once a singular value associated with each day from January 1, 2018 to June 30, 2020 was calculated for both regions, it was input into a CSV file that allowed for further analysis.

*2.5 Data Analysis*

R Studio (V346.421) was used to treat $PM_{2.5}$, $NO_2$, and $O_3$ concentrations across all regions for the study period. The initial stages involved calculations for the daily average mean, as previously discussed. This process involved averaging daily concentrations for each pollutant across all monitoring stations within the identified counties associated with the area of study. Averages were calculated for all seven metropolitan areas. The code used to calculate mean daily pollutant concentrations are available in the appendix (Section IV). While most metropolitan areas had at least one monitoring station that measured $PM_{2.5}$, $NO_2$, and $O_3$ for each county, several metropolitan areas did not have any monitoring stations for these pollutants. Most notably, Houston, Atlanta, and Chicago were missing coverage of five, three, and five counties respectively. In Houston, these counties include Austin, Chambers, Fort Bend, Liberty, and Waller. In Atlanta, these counties include Cherokee, Fayette, and Forsyth. In Chicago, the missing counties include Kendall, DeKalb, Grundy, Jasper, and Newton. $NO_2$ measurements were the least available while $O_3$ and $PM_{2.5}$ were fairly well covered. It is important, however, to note that several $O_3$ datasets were missing concentrations for the winter months. This is likely due to a greater concern about the pollutant during warm summer months when conditions are ideal for the $O_3$ photochemical formation process to occur. When available datasets were identified as unusable due to extensive missing data, they were marked and removed from the analysis (Table 1). All mean daily pollutant concentrations, their corresponding dates, and their corresponding metropolitan areas were input into a CSV file that was used for further analysis.

Mean daily pollutant concentrations were used to run an analysis of variance (ANOVA) test and create significance boxplots comparing overall pollutant concentrations between each of the seven metropolitan areas to identify initial trends of interest. The code was run for each type

of pollutant to create three different boxplots (Section IV). This data was used to create additional significance boxplots plotting annual concentrations of each pollutant for each region (Section IV). Significance boxplots plotting seasonal concentrations for each metropolitan area were created as well. Mean daily pollutant concentrations were assigned to seasons based on the months that fall under the generally accepted meteorological seasons: March, April, and May were assigned to spring; June, July, and August were assigned to summer; September, October, and November were assigned to fall; and December, January, and February were assigned to winter. The code to create regional annual pollutant concentration plots and regional seasonal concentration plots was run for each metropolitan area for each pollutant (Section IV). Finalized boxplots were imported into Microsoft PowerPoint where full figure formatting was arranged, and appropriate axes were added.

Weather and VMT significance correlation plots were configured using the Spearman Rank-Order Correlation test (corroplot package) in RStudio. NOAA daily weather data was averaged for each factor across all monitoring stations within the identified counties associated with the area of study. Averages were calculated for all seven metropolitan areas (Section IV) and calculated means were added to the previously mentioned CSV file holding mean daily pollutant concentrations. Once all data had undergone initial treatment, a significance correlation plot was created to visualize r-values for each weather factor and each pollutant. A similar process was carried out for VMT data, however initial treatment to calculate mean daily averages was unnecessary as the provided data had already been treated by previous researchers. Once the data was added to the CSV file, a significance correlation plot similar to that of the weather correlation plot was created.

## 3. Results

In order to assess differences in pollutant concentrations between geographic locations, $PM_{2.5}$, $NO_2$, and $O_3$ concentrations were compared between seven metropolitan areas using an ANOVA test. Furthermore, pollutant concentrations were compared annually by region using ANOVA tests to examine sources of outliers in the overall pollutant comparison and additional ANOVA tests compared seasonality for each metropolitan area to understand how natural variations impact pollutant concentrations. Spearman Rank-Order Correlation tests were conducted between pollutant concentrations and weather factors to examine natural climate impacts and anthropogenic impacts were examined using Spearman Rank-Order Correlation tests between pollutant concentrations and VMT data.

### *3.1 Pollutant Concentration Comparisons*

Overall, mean concentrations for all pollutants fell below the annual standards set by the EPA, but all regions had outliers that did not meet the standards with the exception of Houston for $NO_2$. Post hoc comparisons using the Tukey HSD tests indicated that there was a significantly higher concentration of $PM_{2.5}$ in LA with a median concentration of 9.7 $\mu g/m^3$ followed by Atlanta and Houston both with a median concentration of 8.5 $\mu g/m^3$ (Table S1). The lowest overall concentrations were exhibited by NY with a median concentration of 6.0 $\mu g/m^3$. Outliers were especially prevalent for $PM_{2.5}$ observations overall (Fig. 2A), and SF exhibited the most outliers above the upper quartile for this pollutant. The lowest $O_3$ concentrations were present in SF, Houston, and NY with median concentrations of 0.034 ppm, 0.034ppm, and 0.033ppm, respectively. Furthermore, SF exhibited the most outliers below the lower quartile. Phoenix exhibited significantly larger concentrations of $NO_2$ and $O_3$ than the other

studied regions (Fig. 2B, 2C) with a median $NO_2$ concentration of 31.7 ppb and a median $O_3$ concentration of 0.048 ppm.

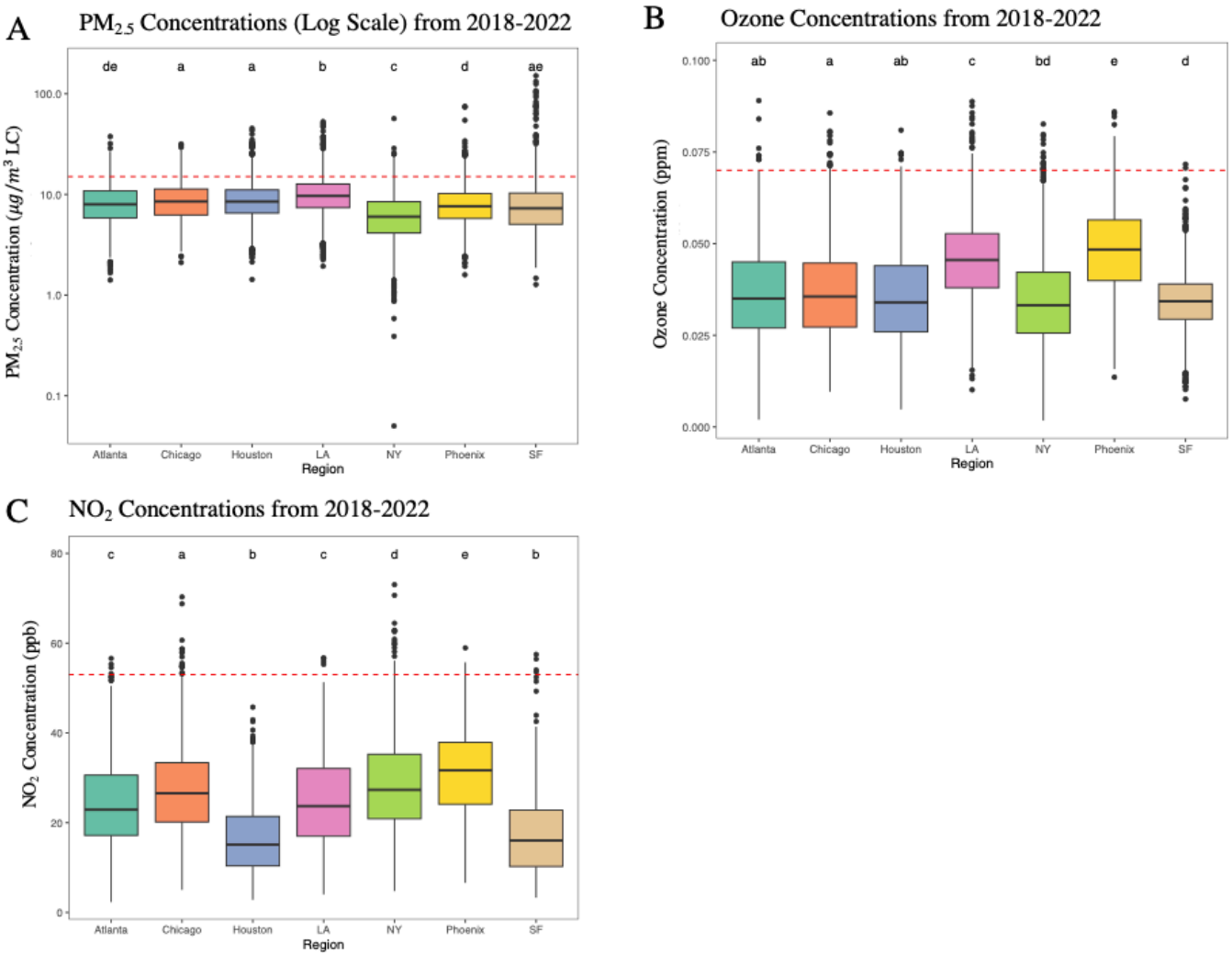


**Figure 2. Fine particulate ($PM_{2.5}$ ), $NO_2$, and $O_3$ concentrations for 7 selected metropolitan regions between 2018 and 2022.** Horizontal lines within the boxes represent median values while vertical lines indicate the overall spread of the dataset. Dots indicate outliers. Letters indicate significance between years for each region where two boxes that share the same letter are not statistically significantly different from each other. Red dotted lines mark the EPA annual standard for each pollutant.

### *3.1.1 $PM_{2.5}$*

Significant annual $PM_{2.5}$ differences were exhibited in all regions (Fig. 3). A general positive trend in pollutant concentration occurred over the course of 5 years in Houston where the lowest median concentration of 7.6 μg/m$^3$ occurred in 2018 and the highest concentration occurred in 2022 with a median of 9.0 μg/m$^3$. The highest concentrations in LA occurred in 2018 with a median of 10 μg/m$^3$ while the lowest concentration occurred a year later in 2019 with a median of 8.9 μg/m$^3$. A similar trend occurred in SF where the highest median concentration in 2018 was 8.2 μg/m$^3$ and the lowest median concentration in 2019 was 6.3 μg/m3. The lowest concentrations for Atlanta and Phoenix occurred in 2020 with median concentrations of 7.5 μg/m$^3$ and 5.1 μg/m$^3$ respectively and the highest concentrations occurred in 2019 with 8.4 μg/m$^3$ for both metropolitan areas. Chicago and NY exhibited the lowest median concentrations in 2020 with values of 8.0 μg/m$^3$ and 5.3 μg/m$^3$ respectively. Both metropolitan areas also exhibited the highest median concentrations in 2021 where Chicago had a median concentration of 9.1 μg/m$^3$ and NY had a median concentration of 6.3 μg/m$^3$. Outliers were present for every year in all regions, but extensive outliers were observed in 2020 for SF and LA.

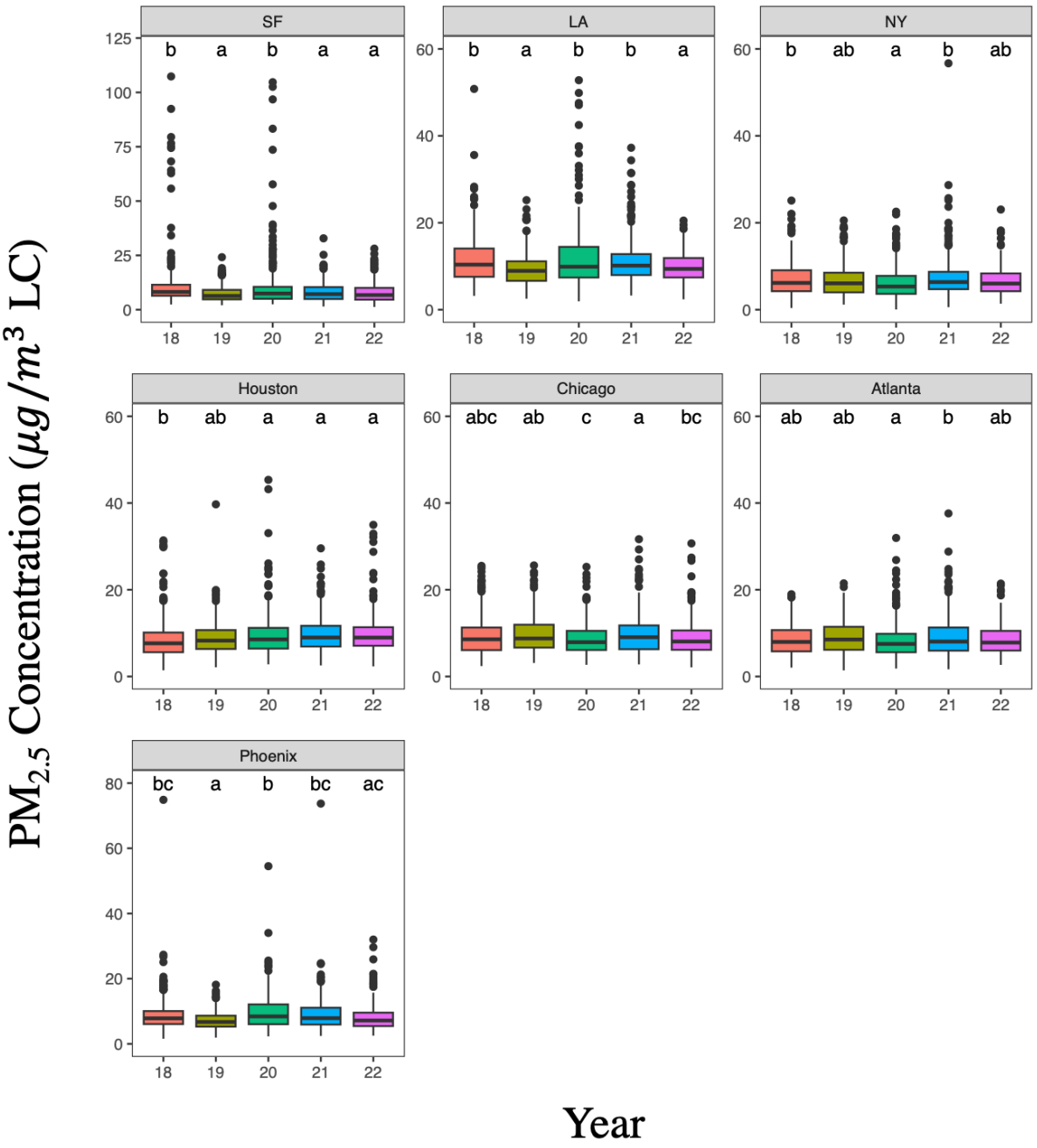


**Figure 3. Annual fine particulate ($PM_{2.5}$) concentrations for 7 selected metropolitan regions between 2018 and 2022.** Red boxes represent 2018, yellow boxes represent 2019, green boxes represent 2020, blue boxes represent 2021, and purple boxes represent 2022. Horizontal lines within the boxes represent median values while vertical lines indicate the overall spread of the dataset. Dots indicate outliers. Letters indicate significance between years for each region where two boxes that share the same letter are not statistically significantly different.

### *3.1.2 $NO_2$*

ANOVA tests indicated overall significant annual differences between $NO_2$ concentrations for the seven regions (Fig. 4). Concentrations in Phoenix were significantly higher in 2018 when compared to all other years with a median concentration of 25.3 ppb as indicated by a Tukey HSD post hoc comparison. The highest concentrations in LA, SF, and Chicago occurred in 2018 with median values of 24.2 ppb, 17.4 ppb, and 27.6 ppb, respectively. The highest concentrations for NY and Atlanta occurred in 2019 with median concentrations of 29.2 ppb and 25.25 ppb, respectively. Houston's highest concentration occurred in 2022 with a median concentration of 16.0 ppb. The lowest concentration occurred in 2021 for LA and SF with median values of 23.4 ppb and 14.7 ppb, respectively. All other regions experienced their lowest median concentrations in 2020: NY's concentration was 25.7 ppb, Houston's concentration was 14.1 ppb, Chicago's concentration was 26.0 ppb, Atlanta's concentration was 21.6 ppb, and Phoenix's concentration was 21.6 ppb. Outliers were present for all regions except for Phoenix with SF exhibiting extensive outliers in 2018.

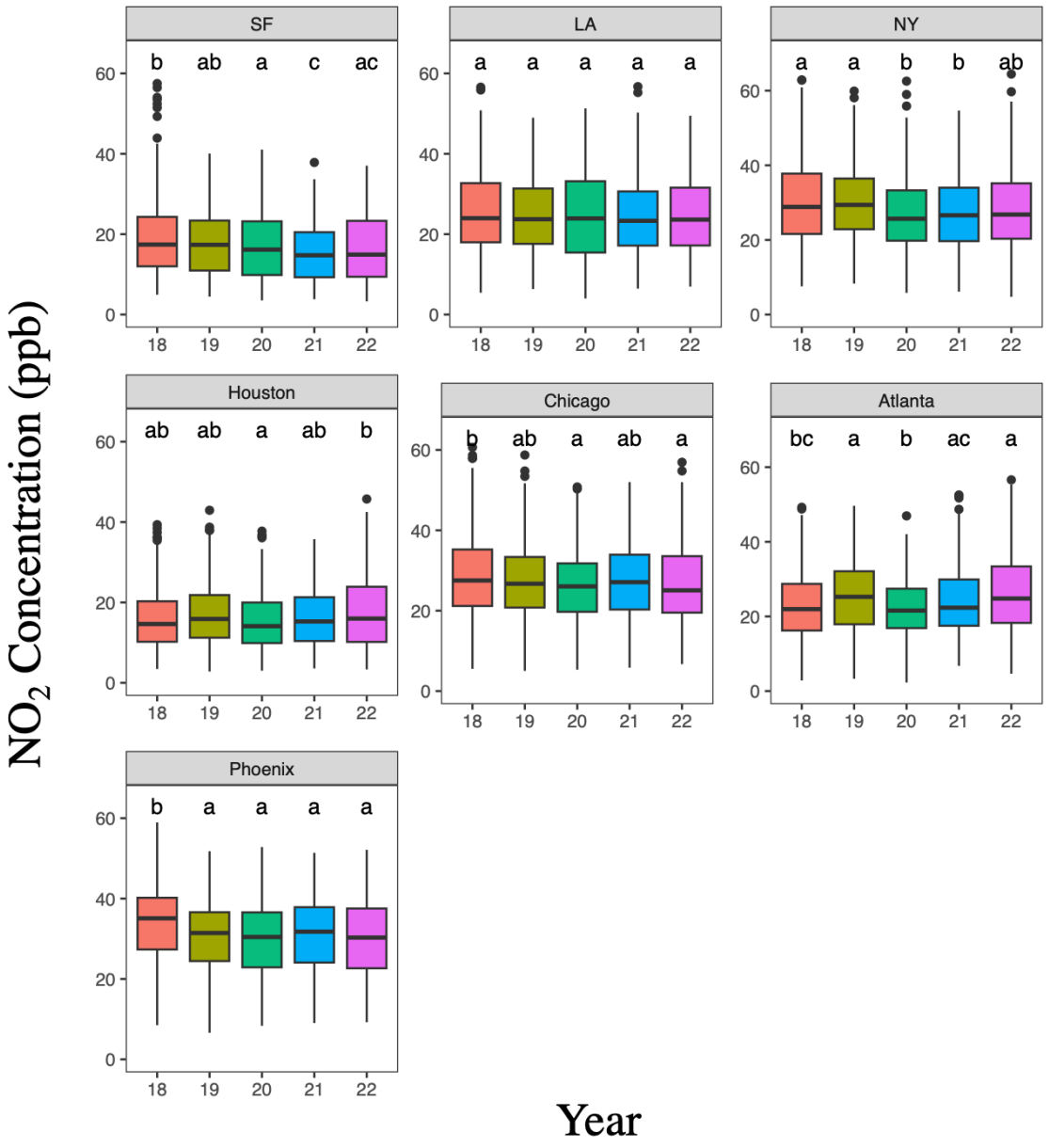


**Figure 4. Annual $NO_2$ concentrations for 7 selected metropolitan regions between 2018 and 2022.** Red boxes represent 2018, yellow boxes represent 2019, green boxes represent 2020, blue boxes represent 2021, and purple boxes represent 2022. Horizontal lines within the boxes represent median values while vertical lines indicate the overall spread of the dataset. Dots indicate outliers. Letters indicate significance between years for each region where two boxes that share the same letter are not statistically significantly different.

*3.1.3 Ozone*

Post hoc comparisons using Tukey HSD tests indicated a significantly higher $O_3$ concentration in 2019 for Atlanta with a median concentration of 0.037 ppm as well as a significantly higher concentration in 2022 than 2018 in Houston. However, there were insignificant differences between years for all other regions (Fig. 5). LA's highest median concentration of 0.046 ppm occurred in 2018 and 2022 while the lowest median concentration of 0.045 ppm occurred in 2020. In SF, the highest concentration occurred in 2021 with a median value of 0.035 ppm while the lowest concentrations occurred in 2018, 2019, and 2022 with a median concentration of 0.034 ppm. Ozone concentrations in NY were lowest in 2018 with a median concentration of 0.031 ppm. The highest concentrations in this region were exhibited in 2019, 2021, and 2022 with a median concentration of 0.034 ppm. The highest median concentrations for Houston and Chicago occurred 2022 with values of 0.035 ppm and 0.037 ppm, respectively. Houston's lowest median concentration was 0.031 ppm in 2018 while the lowest median concentrations of 0.035 ppm occurred in 2018, 2019, and 2020 in Chicago. The highest median concentration in Phoenix occurred in 2019 with a median concentration of 0.037 ppm and the lowest median concentration occurred in 2020 with a median concentration of 0.032 ppm. Outliers were present for all regions but were more prevalent in SF, LA, NY, and Chicago.

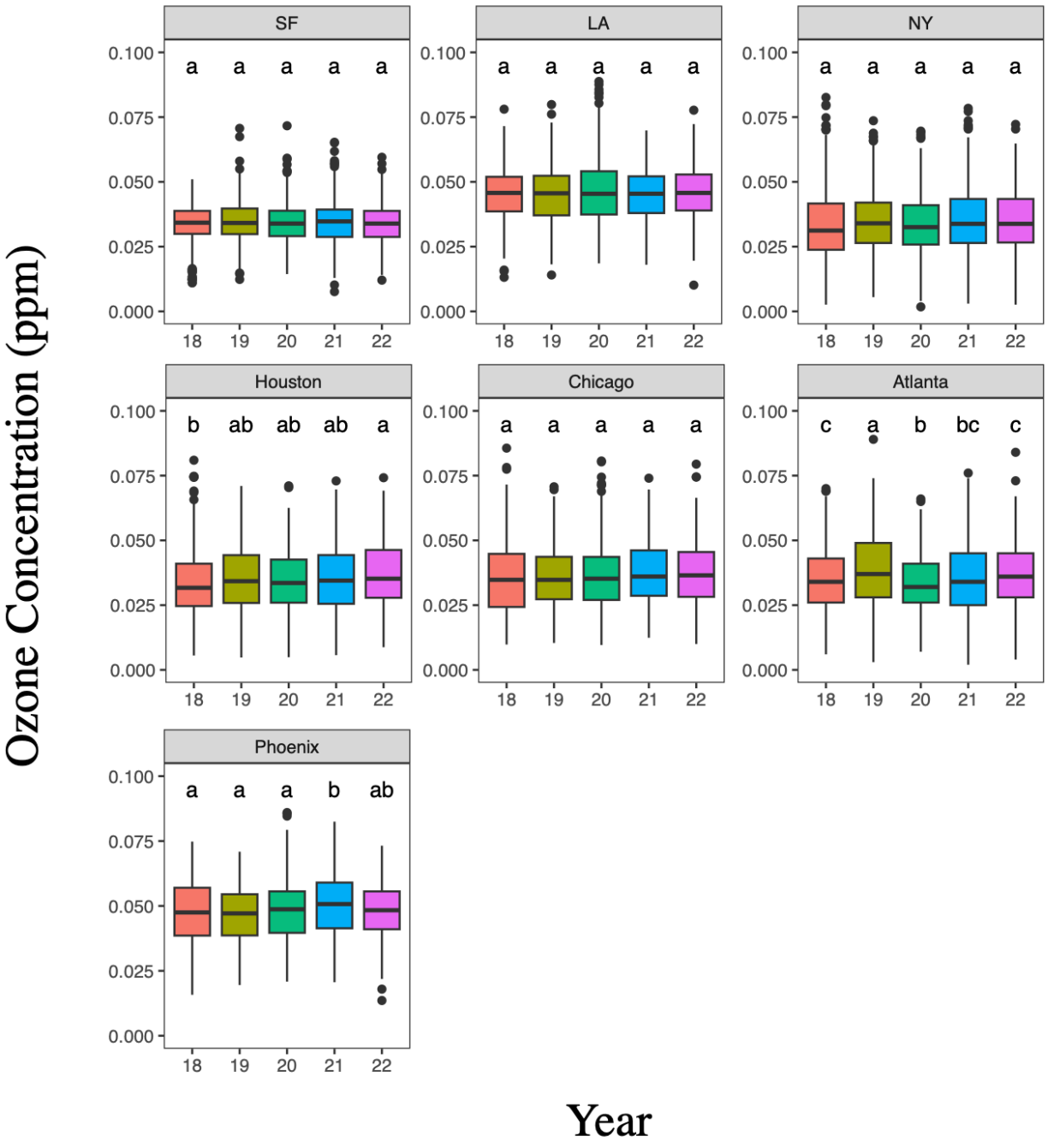


**Figure 5. Annual $O_3$ concentrations for 7 selected metropolitan regions between 2018 and 2022.** Red boxes represent 2018, yellow boxes represent 2019, green boxes represent 2020, blue boxes represent 2021, and purple boxes represent 2022. Horizontal lines within the boxes represent median values while vertical lines indicate the overall spread of the dataset. Dots indicate outliers. Letters indicate significance between years for each region where two boxes that share the same letter are not statistically significantly different.

*3.1.4 Seasons*

Tukey HSD post hoc comparisons indicated $PM_{2.5}$ concentrations were significantly highest in the fall for SF and LA with many outliers above the upper quartile for SF (Fig. 6). Regions on the West Coast exhibited similar trends regarding low concentrations with SF, LA, and Phoenix exhibiting significantly lower concentrations during the fall season. All other regions experienced lowest overall concentrations in either the fall or winter seasons.

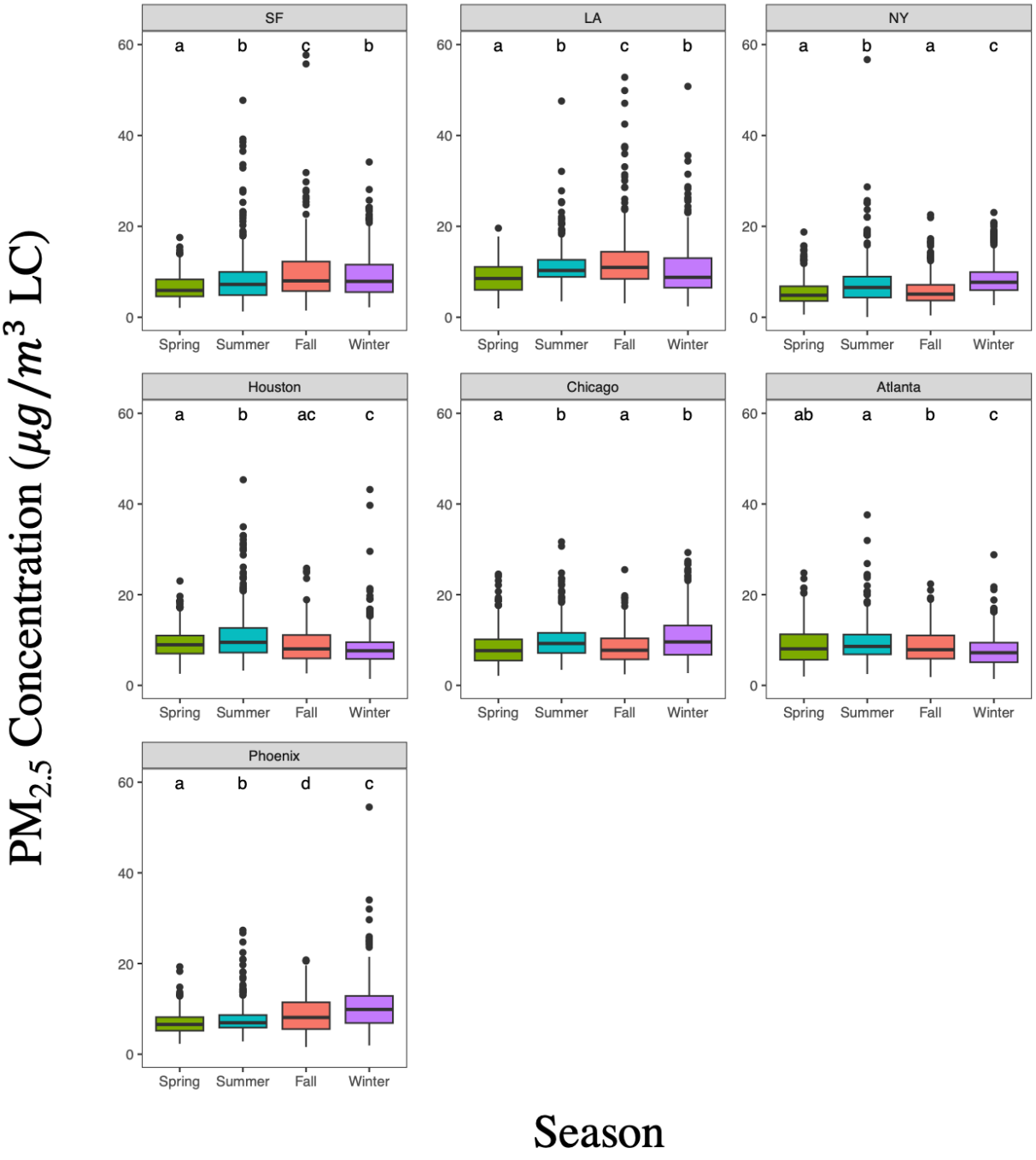


**Figure 6. Seasonal fine particulate ($PM_{2.5}$) concentrations for 7 selected metropolitan regions between 2018 and 2022**. Green boxes indicate spring months, blue boxes indicate summer months, red boxes indicate fall months, and purple boxes indicate winter months. Horizontal lines within the boxes represent median values while vertical lines indicate the overall spread of the dataset. Dots indicate outliers. Letters indicate significance between years for each region where two boxes that share the same letter are not statistically significantly different from each other.

$NO_2$ concentrations were significantly lowest during the summer season for all regions with the exception of Chicago where the difference between fall and winter concentrations were not statistically significant (Fig. 7). Even so, $NO_2$ concentrations in the spring and winter seasons are significantly higher than the fall and summer seasons for this region (Fig. 7). While all regions exhibited highest concentrations during the winter season, the second highest $NO_2$ concentrations were split between spring and fall amongst regions.

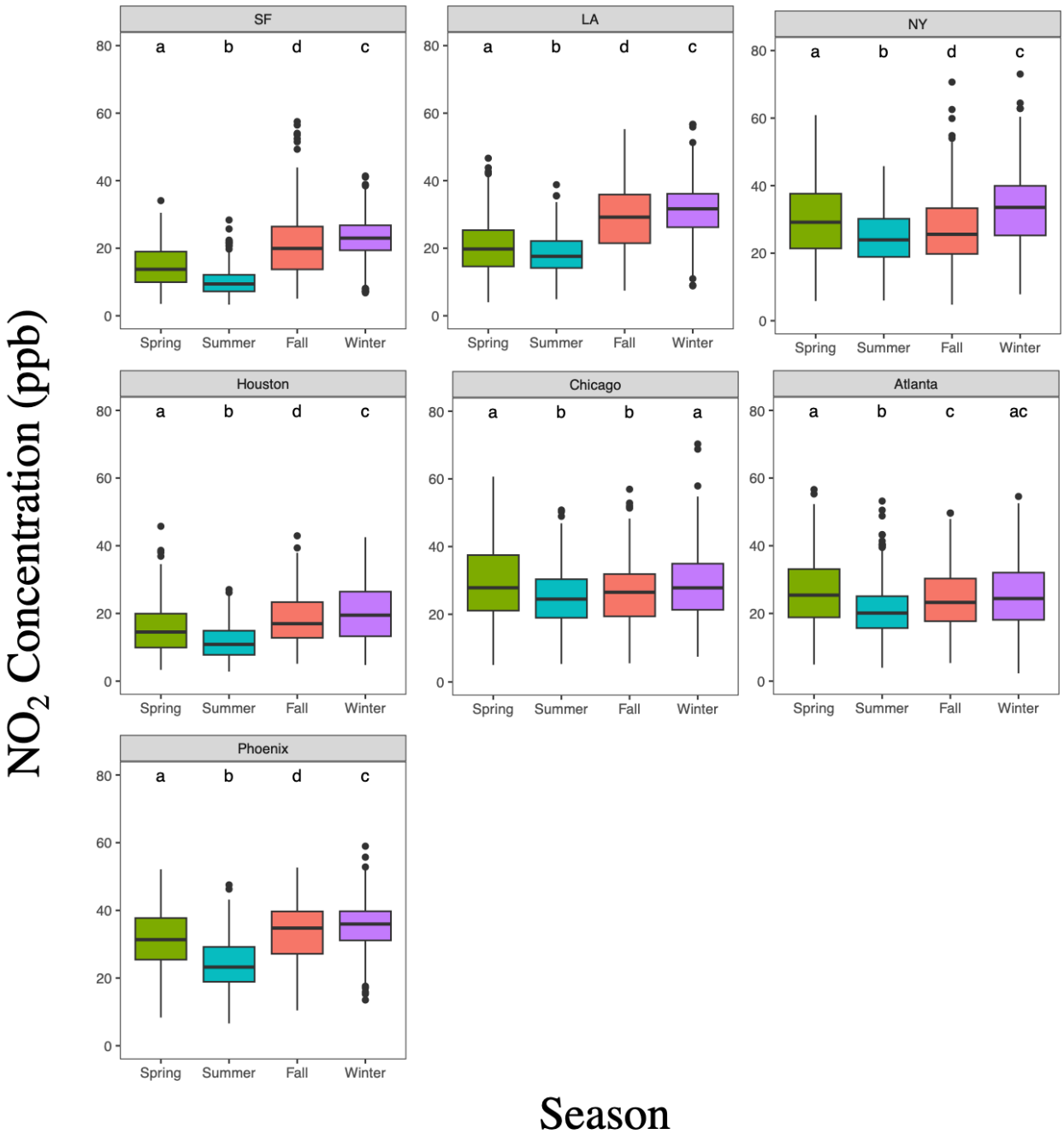


**Figure 7. Seasonal $NO_2$ concentrations for 7 selected metropolitan regions between 2018 and 2022**. Green boxes indicate spring months, blue boxes indicate summer months, red boxes indicate fall months, and purple boxes indicate winter months. Horizontal lines within the boxes represent median values while vertical lines indicate the overall spread of the dataset. Dots indicate outliers. Letters indicate significance between years for each region where two boxes that share the same letter are not statistically significantly different from each other.

The majority of regions had high $O_3$ concentrations in the spring, however LA, SF, and Phoenix exhibited this trend in the fall. Tukey HSD post hoc comparisons revealed that winter exhibited significantly lower $O_3$ concentrations for all regions while summer exhibited the highest concentrations for most regions (Fig. 8). $O_3$ concentrations were highest in the spring for SF and Houston. There was not a statistically significant difference between the two highest concentrations in the spring and summer for Atlanta.

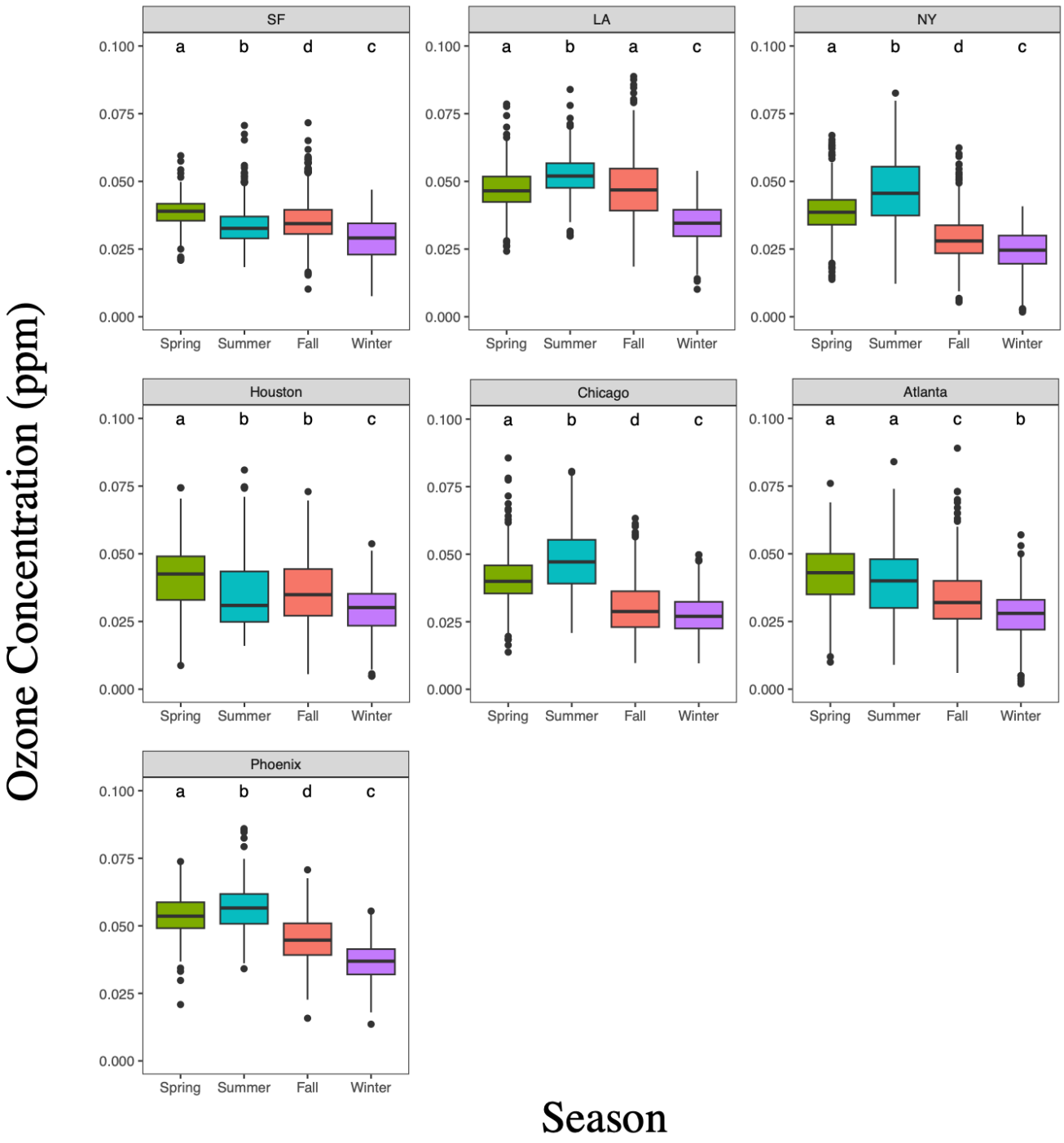


**Figure 8. Seasonal $O_3$ concentrations for 7 selected metropolitan regions between 2018 and 2022.** Green boxes indicate spring months, blue boxes indicate summer months, red boxes indicate fall months, and purple boxes indicate winter months. Horizontal lines within the boxes represent median values while vertical lines indicate the overall spread of the dataset. Dots indicate outliers. Letters indicate significance between years for each region where two boxes that share the same letter are not statistically significantly different from each other.

*3.2 Pollutant Correlations*

*3.2.1 Weather*

Overall, significant strong positive correlations between temperature and ozone were observed for all regions (Fig. 9). The strongest correlation was found for Phoenix (r = 0.69) followed closely by LA (r = 0.68). The weakest of these correlations was exhibited in Houston where a correlation coefficient of 0.13 was found. Precipitation exhibited significant negative correlations with all pollutants for all regions with the exception of SF where no significant correlation was found. While significant negative correlations between wind speed and all pollutants were generally observed, $O_3$ concentrations in SF (r = 0.15) and Phoenix (r = 0.23) were positively correlated with this weather factor. $NO_2$ concentrations in these two metropolitan areas, however had significantly strong negative correlations between $NO_2$ and wind speed. SF (r = -0.72) and Phoenix (r = -0.57) exhibited correlation coefficients that were generally stronger than other regions who experienced similar significant negative correlations, as the next largest coefficient was exhibited by LA with a correlation coefficient value of -0.51. While correlation tests indicated significant correlations for most factors in all regions, low correlation coefficients were calculated for most results.

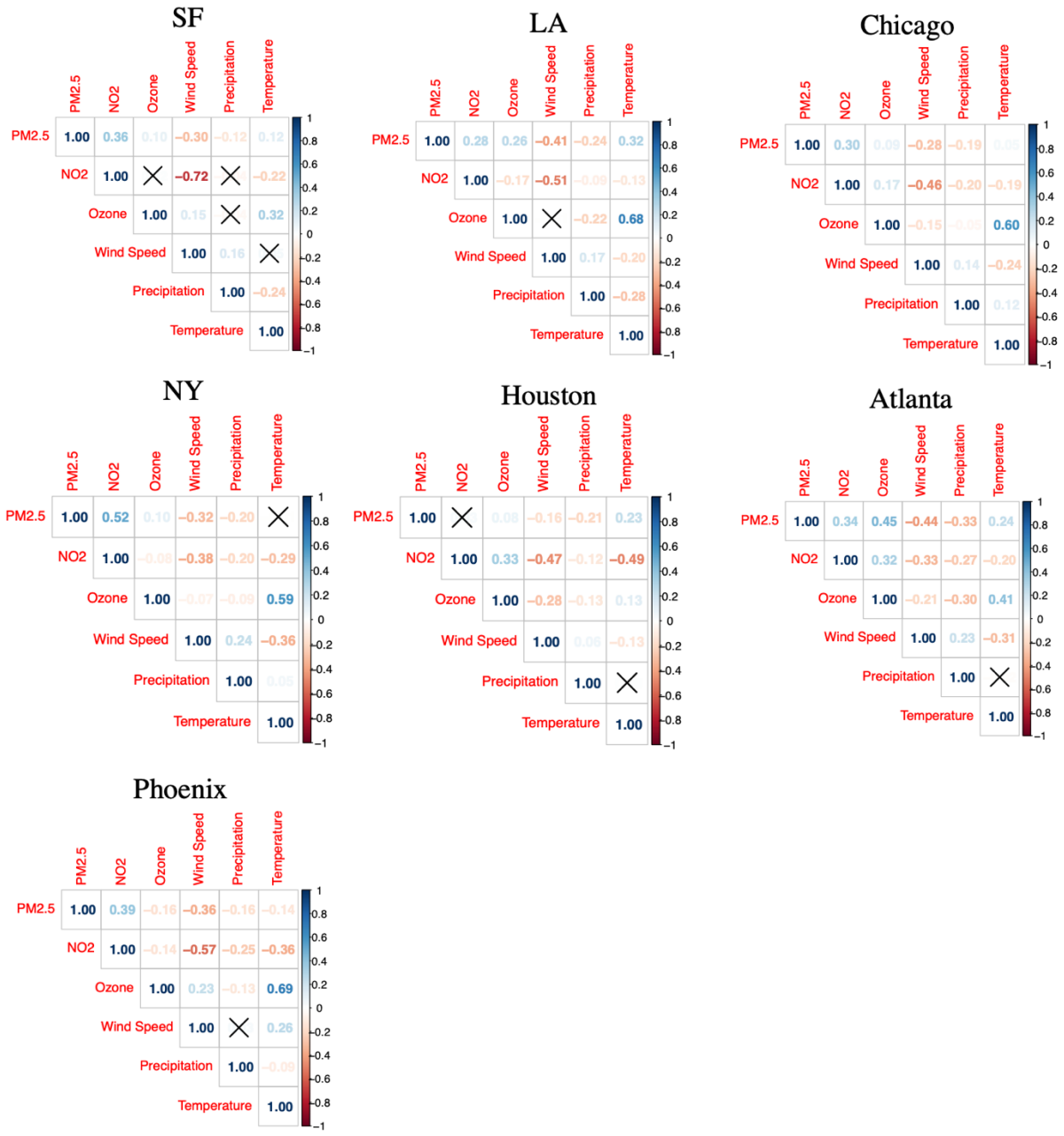


**Figure 9. Significance correlation plots between weather data (wind speed, precipitation, and temperature) and pollutant concentrations for 7 selected metropolitan regions from 2018 to 2022.** Values in boxes indicate r-values for correlation where darker text indicates a more statistically significant correlation. An "X" indicates a lack of significant correlation between the variables.

*3.2.2 VMT*

Significant positive correlations were observed between VMT and $PM_{2.5}$ (r = 0.28) as well as $NO_2$ (r = 0.38) concentrations in LA (Fig. 10). A similar trend was observed in SF for VMT and $PM_{2.5}$ (r = 0.36) and $NO_2$ (r = 0.56) in addition to a weak negative correlation between VMT and $O_3$ (r = -0.29). SF had larger correlation coefficients than the LA correlations. Furthermore, there was no significant correlation between VMT and $O_3$ concentrations in LA (Fig. 10)

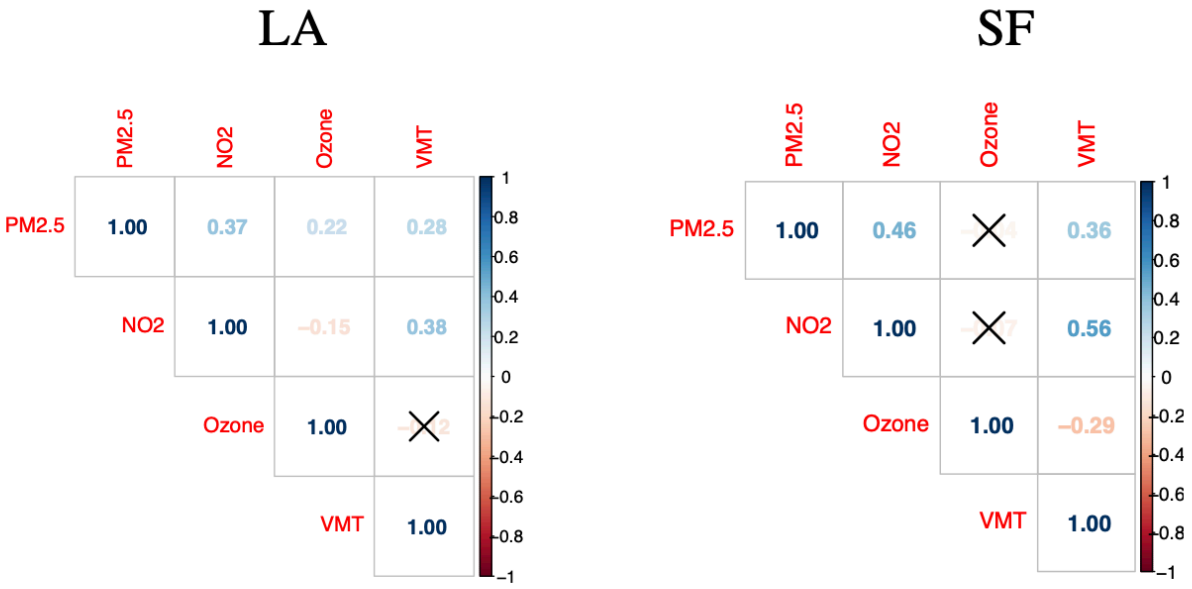


**Figure 10. Significance correlation plot between vehicle miles travelled (VMT) data and pollutant concentrations in LA and SF from January 2020 through June 2020.** Values in boxes indicate r-values for correlation where darker text indicates a more statistically significant correlation. An "X" indicates a lack of significant correlation between the variables.

## 4. Discussion

### *4.1 Temperature*

Overall, the lowest ozone concentrations occurred in the winter for all regions and most regions exhibited the highest concentrations in the summer with the exception of SF, Houston, and Atlanta (Fig. 8). Previous studies have found that temperature has a strong positive correlation with ozone concentrations and increasing trends in temperature sensitivity in ozone production have been stronger in more recent years (Lee et al. 2014). Low temperatures, as winter months generally exhibit the coldest annual temperatures across the US, may contribute to significantly lower ozone concentrations due to reduced rates of ozone formation at this time. Furthermore, ozone formation is dependent on oxygen radicals that occur through the photolysis of oxygen molecules (Chieh C. 2021). Generally, solar UV radiation in winter months is weaker than radiation exhibited during summer months in North America (Sliney & Wengraitis, 2006). Here, oxygen radicals become the limiting factor in ozone formation as the necessary conditions for photolysis to occur are not present. As a result, it is likely that this combination of lower temperatures and UV radiation restrict ozone production in the winter across all regions.

While winter months exhibit the coldest annual temperatures, summer months generally exhibit the warmest annual temperatures. Meng et. al (2023) found that ozone formation reactions increase with temperature, likely contributing to the overall strong positive correlation values between ozone and temperature exhibited across all studied regions. The aforementioned study also identified that under warmer conditions, an increase in ozone production rates was greater than the rate of ozone loss, contributing to fast net ozone accumulation in heat wave conditions. It is possible that these net positive relationships between temperature and ozone formation contributed to generally higher summer concentrations for most metropolitan regions.

The only three regions that did not exhibit highest concentrations in the summer season – SF, Atlanta, and Houston – exhibited highest concentrations in the spring season or did not have a significant difference between the spring and summer seasons. A study conducted in 2020 identified a significant shift in warm weather type frequency in eastern North America with a mean frequency of 56 days (Ellis & Marston, 2020). This approximately two month shift in warm conditions may explain the deviation from the majority of regions in ozone concentrations exhibited by SF, Atlanta, and Houston. Continued exacerbation of these shifts due to anthropogenic climate change may contribute to similar higher spring concentrations in more metropolitan areas in the future.

*4.2 Precipitation*

While there were insignificant correlations between precipitation and $NO_2$ in addition to ozone in SF, all other regions exhibited a negative correlation between precipitation and all pollutants (Fig. 9). The formation of acid rain may contribute to an understanding of these results, as previous atmospheric models have pointed to a dependent relationship between pollution's removal from the atmosphere and the rate of precipitation formation where faster rates have contributed to less remaining pollution (Naresh et al., 2007). Under a four-phase interaction composed of the formation of raindrops, the introduction of primary pollutants, the introduction of secondary pollutants, and the interactions of all pollutants, acid rain formation depends on the rate of pollutant introduction to precipitation (Naresh et al., 2006). Pollutant particles are extracted from the atmosphere as they form acid rain through these processes, thus likely resulting in a lower atmospheric concentration as measured at EPA monitoring stations.

These negative correlations, however, are generally weak as exhibited by correlation coefficients greater than -0.30 for all calculated values (Fig. 9). Previous studies have found the incorporation of air pollutant particles into orographic clouds, formed in response to local topography and uplifted winds, leads to a reduction in drop coalescence speed, thus resulting in a delay in the conversion of clouds to precipitation (Givati & Rosenfeld, 2004). Consideration of this additional relationship between precipitation and pollutants may contribute to lower confidence within these correlations as the amount of precipitation may be limited by the pre-existing concentration of pollutants in the area. Further investigation into precipitation suppression through examination of precipitation on a monthly basis in relation to pollutant concentrations would allow for a better understanding of the impact of this weather factor on the studied metropolitan areas.

*4.3 Wind*

$NO_2$ and wind were significantly negatively correlated across all metropolitan regions with an especially strong correlation in SF ($r = -0.72$) and Phoenix ($r = -0.57$) (Fig. 9). Previous studies have shown that $NO_2$ columns are strongest in the area of origin, but they do exhibit high influence on surrounding areas through wind transport (Rivera et al., 2015). While these two metropolitan areas are located in the western part of the US, both have different topographies. SF is located along the coast, thus exposing the area to winds from the Pacific Ocean. It is possible that these winds may blow the $NO_2$ pollutants inland, contributing to reducted pollutant concentrations for the metropolitan area as a whole. On the other hand, Phoenix is landlocked and located within a basin that expands into Mexico. The expansive nature of these geologic

features may contribute to the transport of pollutants to far edges of the basin, thus reducing the concentrations directly over the studied metropolitan area.

While there were strong negative correlations exhibited between wind and $NO_2$ in these regions, significantly high $NO_2$ and ozone concentrations were present in Phoenix while SF concentrations were generally lower (Fig. 2). Here, considerations of the different topographies may lead diverging processes of pollutant buildup between regions. Previous studies have found that the spread of $NO_2$ due to wind were especially strong for regions in the same basin while low wind speeds and positive vertical temperature gradients were associated with weather classification schemes that saw the highest concentrations of $NO_2$ (Rivera et al., 2015; Grundstrom et al., 2015). Additional studies have shown that a multi-day buildup of locally produced ozone did not have a significant increase during the day, but the buildup of precursors at night did impact ozone concentrations for the next day (Fast et al. 2000). It is possible that generally low wind speeds exhibited in Phoenix may combine with basin topography to build up ozone and ozone precursors, including $NO_2$, over time.

In addition to the local buildup of pollutants, Gaffey et al. (2022) found that CO and NO, two emissions that can contribute to the formation of $NO_2$, are mostly mobile and indicate long-range transport of emissions through an anticorrelation between $NO_2$ formation and $O_3$ concentrations at night. Wind activity around Phoenix and surrounding regions, therefore, may contribute to the higher concentrations of observed $NO_2$ and $O_3$ concentrations. Further spatial analysis may allow for insight into the impact of long-range transport on these regions of study.

*4.4 Vehicle Emissions*

While positive correlations between $PM_{2.5}$ and $NO_2$ with VMT were observed, there was a negative correlation between ozone and VMT (Fig. 10). Previous studies have identified a general trend in vehicle contributions to roadway pollution with especially strong contributions from diesel vehicles (Song et al., 2018). With this in mind, it is possible that these positive correlations are due to a larger fleet of diesel vehicle traffic during this period.

The negative correlations between ozone and VMT, however, may indicate high concentrations of VOCs and additional pollutants in these regions (Fig. 10). Liu et al. (2018) found that $NO_x$ emissions from vehicles have high contributions to ozone formation with the elimination of $NO_x$ and or VOC emissions from these vehicles estimated to result in a 5-2.5% decline in ozone formation. While $NO_x$ is a contributor to ozone formation, these pollutants can participate in additional atmospheric reactions to form nitric acid ($HNO_3$). The compound relies on the photolysis of ozone to form OH radicals that react directly with $NO_x$. It is possible that a net negative trend in ozone particles can be explained by $NO_x$'s participation in other reactions considering the ways in which ozone and nitric acid reactions occur simultaneously. Under this framework, the limiting factor in these reactions would become UV radiation and the rate of photolysis.

Even so, overall low correlation coefficients indicate that additional factors impacted pollutant concentrations outside of the observed VMT data (Fig. 10). These values, however, could also be in-part due to the years where data was accessible. In the year 2020, average total vehicle traffic was reduced by 12.7% and $NO_2$ concentrations dropped significantly in counties where COVID-19 lockdowns did occur in comparison to counties where lockdowns were not instated (Wegman & Katrakazas, 2021; Bar et al., 2021). These trends in reduced traffic,

stemming from complex societal adjustments to the global COVID-19 pandemic, may introduce confounding factors into the correlations. Access to additional data from years before and after the pandemic can contribute to a more robust study of vehicle impact on pollutant concentrations in these regions.

*4.5 California Wildfires and Outliers*

Out of the seven studied metropolitan areas, the highest overall $PM_{2.5}$ concentrations from 2018-2022 were observed in LA. Similar to Phoenix, the region is located in a basin where previous studies have pointed to meteorological feedbacks that contribute to a buildup of pollutant concentrations in the central basin (Wang et al. 2020). The topographic characterization of basins, composed of taller features that surround a central dip, may contribute to the confinement of these particles within the system and the resulting positive feedback loops that further exacerbate the $PM_{2.5}$ buildup.

While the observed $PM_{2.5}$ concentrations in LA may be due to the inability for wind to easily blow particles out of basins, there was a weak negative regional correlation between $PM_{2.5}$ concentrations and wind (Fig. 9). Wang and Guo (2009) found a similar pattern – the spatial patterns of the LA region combined with the Santa Ana Winds (SAWs) generally had a negative correlation where lower concentrations were exhibited in areas where the wind was blowing. This behavior, however, operated inversely when wildfires were introduced into the system. During time periods when wildfires were presently burning, SAW activity contributed to higher pollution concentrations in the region (Wang and Guo 2009; Aguilera et al. 2020). With this information in mind, California's annual wildfires may contribute to observed high concentrations even while a negative correlation between wind and pollutant concentration was

found. Furthermore, studies have identified that the amount of $PM_{2.5}$ in the basin system impacted the extent of the positive feedback loop where higher concentrations led to more effects within the system (Wang et al. 2020). An initial buildup of $PM_{2.5}$ particles within the LA basin, such as during a wildfire event, may have led to further buildup due to faster rates of the aforementioned positive feedback loops.

The presence of wildfires may also contribute to extensive $PM_{2.5}$ outliers observed for SF and LA in 2020 (Fig. 3). In 2020, Cal Fire reported that the state of California experienced 8,648 wildfires. Among these fires, the August Complex Fire in Northern California was recorded as the largest fires in California's history ("10 Dead in California as Wildfires Spread on West Coast," 2020) and the state experienced 4,304,379 of burned acreage by the end of the year, double the amount burned in 2018 (Safford et al. 2022). Large-scale biomass burning events, such as these annual wildfires, can be important point sources of $PM_{2.5}$ particles. Burning periods, however, vary on an annual basis. This unpredictable variation can make wildfires a key pollutant source which can lead to extremely high concentrations one year in comparison to the baseline of the next. This supports fewer outliers that are exhibited for years outside of 2020 when less extensive burning occurred during these time periods. With this in mind, the extent and intensity of the burning in 2020 thus likely contributed to the many observed outliers in both SF and LA for this year.

The 2020 California wildfires have initiated extensive studies into the state's burning period, especially in the ways it is responsive to anthropogenic climate change. Southern California's chapparal shrub communities are prone to burning and rely on postfire conditions for growth and persistence, however frequent fires on shorter intervals have inhibited these communities' abilities to grow in more recent decades (Storey et al., 2021). Previous studies

have pointed to longer and more intense burning periods as a result of expanded dry seasons, drought conditions, and recessions in snowpack (Safford et al. 2022). Additional changes in weather patterns have shifted patterns in vegetation growth and drying, thus producing more annual fuel for burning in addition to varying amounts of fuel moisture in the fall and summer (Hernández Ayala et al., 2021; Wang et al. 2023). SF and LA, the two metropolitan regions in the state of California, exhibited more overall outliers over the course of 2018 to 2022 in comparison to the other areas of study (Fig. 2). In addition to contributions from the record-breaking 2020 wildfires, these changes in ecosystem conditions may also contribute to more regular biomass burning events and a consequential increase in $PM_{2.5}$ sources. Furthermore, previous studies have identified how California's dry hot air with low moisture content is conducive to the spread of wildfires (Hernández Ayala et al., 2021). The contribution of these conditions to the spread, and therefore the length of burning, increase fire intensity which may explain the present outliers.

The majority of SF outliers were observed in the fall season when $PM_{2.5}$ concentrations were generally high for both SF and LA (Fig. 6). These trends may be explained by changes in fuel availability and seasonal conditions necessary for burning to occur. Gross et al. (2020) found a weather and climate driven trend towards higher-risk conditions of wildfire burning occurring in August. The study identified that this trend makes it more likely for northern and southern California to undergo wildfire burning simultaneously, an event that was previously rarer. This shift towards extensive burning in August introduces an additional source of $PM_{2.5}$ into the environment that is not present during the other seasons, thus increasing the overall concentrations during the end of the summer and early fall. Previous studies have also identified how strong winds can contribute to the dispersal of $PM_{2.5}$ particles across hundreds to thousands

of kilometers and can keep particles airborne for months (Aguilera et al. 2020). While August borders the end of the summer and beginning of the fall, it is possible that the introduction of SAWs for LA and additional wind patterns in SF may contribute to longer-lasting particle suspension across the entire fall period.

While high concentrations of $PM_{2.5}$ were exhibited in both California metropolitan areas, low ozone concentrations were observed in SF while high concentrations were observed in LA. Smoke events, as produced by wildfires, have been shown to reduce UVB radiation necessary for ozone production due to the absorption of radiation by soot particles (Gaffney et al. 2002). Extensive wildfire burning during 2020 in addition to smaller biomass burning events throughout 2018-2022 may contribute to more $PM_{2.5}$ particles with soot components in the SF metropolitan area and a resulting reduced rate of photolysis. Under these conditions, oxygen radicals may serve as a limiting factor for ozone production and lead to an overall smaller concentration of this pollutant.

In contrast, the high ozone concentrations observed in LA may be explained by variations in the amount of locally produced $NO_x$ and VOC emissions between these two metropolitan areas. A 2021 study by Langford et al. identified an overall higher ozone concentration in urban areas adjacent to wildfire burning due to wind transfer of ozone with wildfire smoke. They also identified that there was a possibly significant relationship between ozone production and local pyrogenic VOCs and $NO_x$ in the surrounding region. LA's history of vehicle transport may contribute to higher concentrations of these primary pollutants in comparison to SF, and buildup of these pollutants due to the LA basin's topography may provide more local emissions for the wildfire smoke to react with in comparison to SF. This topography may also provide the

necessary conditions for positive feedback loops in terms of ozone production to form, thereby causing the disparate ozone concentrations between SF and LA.

*4.6 Air Pollution Policy*

The COVID-19 pandemic led to a string of policy decisions across the US in the form of lockdowns. The first case of COVID-19 in the US was reported in January of 2020, and within five months there were 18,600 confirmed deaths alongside over 500,000 confirmed cases (CDC, 2023). The state governments began implementing shutdowns across the country as a mitigation strategy to fight against the virus in March of 2020, and these conditions were instated through 2021 (CDC, 2023). During the beginning of the shutdown period in 2020, all regions except for SF and LA experienced their lowest overall $NO_2$ concentrations (Fig. 4). A study conducted by Liu et al. (2021) drew connections between minimized transportation during the pandemic shutdowns and changing trends in $NO_2$ concentrations: $NO_2$ concentrations near major power plants decreased following the shutdown period, while $NO_2$ concentrations in heavily residential areas with close interactions to highways increased. It is possible that similar trends were exhibited in the metropolitan regions of study which encompass both industrial and residential areas. With this in mind, the lower overall concentrations may be attributed to a larger decrease in pollutant concentrations in industrial areas than increases in residential spaces for the studied regions.

While all of the median concentrations of the studied metropolitan areas fell below EPA annual standards, most regions had outliers above the standards except for the observed $NO_2$ in Houston. These standards, as set by the Clean Air Act, were established in 1970 with the goal of having every state meet the regulations five years later (US EPA, 2013). The legislation

prioritized major pollutant sources, specified as singular or group sources that have the potential to emit 10 or more tons of pollutants on an annual basis (US EPA, 2013). While the implementation of this policy likely supported pollutant concentrations meeting an overall standard, it is possible that outliers arose from events not directly tied to the major pollutant sources that the legislation was targeting.

Legislation that built onto the Clean Air Act may contribute to further differences between metropolitan regions. Ozone concentrations were highest in the summer season for all regions with the exception of SF, Houston, and Atlanta (Fig. 8). It is probable that concentrations were higher in summer months due to higher UV radiation and consequential photolysis, however the three regions that did not follow this trend may be explained by responses to policy changes. Previous studies have identified that ozone sensitivity regimes are VOC limited in conditions of extreme heat (Meng et al., 2023), and it is possible that past policies aimed at VOC reductions have thus contributed to these unexpected results in SF, Houston, and Atlanta. Alongside the Clean Air Act, the US signed the Montreal Protocol, which aimed to reduce the production of ozone-depleting substances, in 2012 (US EPA, 2015). Since then, the EPA reports a 78% decrease in the combined concentrations of many principal pollutants, including VOCs (US EPA, 2015). Considering ozone production is VOC limited under warmer conditions, policies targeting these primary pollutants may have further limited the amount of ozone formation that could be undergone in more recent years.

When considering all of the studied metropolitan areas, NY exhibited the lowest overall $PM_{2.5}$ concentrations. Pitiranggon et al. (2021) found that major policy implementation and economic incentives led to a decrease in the use of fuel oil for ships and residual fuel oils for cars and other forms of vehicular traffic. They identified that changes in emission standards, rather

than decreased traffic, had led to an overall decrease in $PM_{2.5}$ concentrations in the area from 2002-2018. These changes, stemming from state government-level policy changes, may have contributed in-large to these low concentrations. It is also possible that the state's focus on vehicle emissions played a key role in reducing the observed concentrations.

Policy, however, spanned further than just a vehicle focus. Studies looking at NY emissions from 2005-2016 identified that policy changes had lowered electricity generation industry emissions and saw a decline in the use of No. 2 and No. 4 oils for space heating within the region (Masiol et al. 2019). The policies implemented by NY oversaw changes in all sectors, and the widespread reach of changes within the urban system may have played a large role in lowering $PM_{2.5}$ concentrations. Even so, the only source that exhibited a positive trend within NY was the use of spark-ignition vehicles, however a simultaneous introduction of ultralow sulfur fuels for most forms of transportation and the implementation of particle traps and $NO_x$ controls for heavy-duty diesel trucks was in operation (Masiol et al. 2019). It is possible that these concurrent efforts to address rising trends in pollutant sources led to the overall lower concentrations in the metropolitan area, even with less robust results exhibited. Yet, the strong focus on this pollutant source also points to vehicle traffic as an area that may require more nuanced policy changes for further pollutant reduction in the future. Differences in correlation coefficients for VMT and pollutants (Fig. 10) also suggest that the necessity to prioritize vehicular emissions in relation to other pollutant sources may vary among regions.

## Conclusion

### *5.1 Overview: Trends in Metropolitan Pollutants*

The strongest correlations were found between ozone and temperature for all regions, likely attributed to the necessity of photolysis in the formation of ozone. These findings are supported by seasonal variations in ozone concentrations where summer concentrations were highest for all areas of study. The overall highest ozone and $NO_2$ concentrations were observed in Phoenix during the period of study and both were significantly higher than the other metropolitan areas in the analysis. Negative correlations were generally found between $NO_2$ and wind while precipitation and all pollutants exhibited a weak negative correlation. Complex atmospheric reactions in tandem with different topographies may contribute to these findings, as the formation of acid rain and basin wind patterns introduce dynamic systems that may influence observed concentrations.

The introduction of wildfires into these systems complicates the generally observed trends. In addition to extensive $PM_{2.5}$ outliers in LA and SF during 2020, significantly higher overall $PM_{2.5}$ concentrations were observed in LA when compared to all studied regions. These trends may be the result of wildfire burning during these time periods. Intensified burning as a result of environmental shifts in the face of climate change paired with strong wind patterns may have further contributed to pollutant transport and buildup for these metropolitan areas.

Generally, VMT correlations from this study align with previous literature (Song et al., 2018). Positive correlations were observed between VMT and $PM_{2.5}$ as well as $NO_2$, indicating the use of diesel vehicles in the two regions of California being studied. Surprisingly, weak or insignificant anticorrelations were observed between VMT and ozone despite the positive relationship between VMT and $NO_2$.

Air pollution policy in the US, such as the Clean Air, may also contribute to lower observed concentrations in several metropolitan areas. NY exhibited significantly lower $PM_{2.5}$ concentrations than all other regions of study. Well documented policies aimed at reducing pollutants have been implemented in this region, likely contributing to these low concentrations. Even so, median concentrations for all regions fell below the EPA annual standards as set by the Clean Air Act. Outliers may be attributed to sources that are difficult to predict, such as wildfires, or from additional sources not identified as major sources by the Clean Air Act.

### *5.2 Limitations and Next Steps*

Further extensions of this research would benefit from examination of the relationships between wind transport between metropolitan regions, exploration of the role of precipitation suppression, and more extensive analysis using VMT data for all regions over a longer span of time.

While relatively strong correlations between wind and $NO_2$ were drawn from this study, a lack of spatial analysis limits the application of long-range transport of pollutants to these findings. A study by Gaffey et al. (2022) identifies this transport of pollutants to be especially prevalent in basins. With this in mind, this extension of the study would provide important insight into the observed findings for metropolitan areas with significant results, such as LA and Phoenix. This analysis can be completed using HYSPLIT, providing important implications for the relationship between wind and the studied pollutants.

Weak negative correlations between precipitation and pollutant concentrations may be further expanded on through an analysis of the relationship between pollutant concentrations and exhibited precipitation for all regions. Understanding this inverse relationship will provide

insight into the weak r-values exhibited in correlation calculations while also opening additional avenues for investigation regarding other factors influencing the observed correlations. The relationship between acid rain and pollutant concentrations may also be expanded on by analyzing the relationships between precipitation pH and observed $NO_2$ concentrations for each region.

While the available VMT data allowed for correlations between pollutants and vehicle activities in the state of California, it is possible that the results drawn from these analyses are limited in scope as they are from the same general region. A broader understanding of VMT's relationship with observed concentrations can be drawn from analysis for regions from other states, as would be the case for the other five metropolitan areas of study. Furthermore, analysis of VMT data over the course of several years would reduce the likelihood of biased results.

### *5.3 Further Implications*

The findings from this study emphasize the diverse factors that influence air pollutants within a dynamic atmospheric system. While policies have already been put in place to address existent atmospheric pollution, further policy action will be necessary as urban centers and their surrounding metropolitan areas continue to expand. Resulting anthropogenic activity will continue to impact these systems at a local and global scale and if unattended to, buildup of these concentrations can negatively impact the health of exposed populations through a decline in physical fitness within these densely populated regions (Zhao & Ma, 2020). Significant differences between overall pollutant concentrations for $PM_{2.5}$, $NO_2$, and ozone between metropolitan areas indicate that region-specific conditions need to be considered when implementing further policies aimed at reducing harmful aerosols. Topographic variations and

temperature differences should be accounted for, as they may influence wind patterns and pollutant formation rates.

The focus of prior policies on major sources of toxic particles can be expanded on to address less consistent sources. Observations from this study show the expansive influence of wildfire activity on $PM_{2.5}$ and ozone concentrations, leading to extensive outliers that surpass the EPA's annual standards. Looking forward, changing climate conditions will likely lead to more extensive and intense burning events (Hernández Ayala et al., 2021; Wang, et al. 2023; Storey et al., 2021; Safford et al. 2022). Integration of actions aimed at addressing the reduction of these events through on-site ecosystem maintenance is thus as important as policies aimed at reducing the greenhouse gas effect. The resulting emphasis on pairing regional policies with broader aims may be essential in accounting for unpredictable pollutant sources.

Shifting trends in the observed seasonal concentrations further indicate the role of climate change within observed pollutant concentrations. With this in mind, regional policies must also account for changing weather conditions that may result in atmospheric interactions that differ from historically observed trends. A thorough understanding of each region's baseline relationships between weather and pollutant concentrations is thus essential, and further examination into these trends is of utmost importance moving forward.

**Acknowledgements**

I would like to thank all the contributors to this project. A special thank you to my EA thesis advisors Professor Katie Purvis-Roberts and Professor Franck Fu for their consistent support and assistance throughout the entire process. Professor Purvis-Roberts, I am eternally grateful for the care you showed me in my introductory chemistry courses as I was adjusting to college and my time at Scripps. You have pushed me to be a better scientist and person. Professor Fu, thank you for your patience and helping hand during a busy time as I navigated RStudio. Your instrumental feedback has been essential in the comprehensive nature of the project.

Thank you to my EA major advisor, Professor Robins, for his constant support throughout my college career. I am grateful for the ways you have encouraged me to expand my horizons beyond the classroom while preparing for the future. I would also like to thank Khylah Pugh, Natalie Burton, Daniel Bonilla, Johan Martinez, and Claudia Quintero for consistently lending an ear and discussing this project with me.

Many thanks to my family for their love and care for my wellbeing since day one. I am forever grateful for my parents' support and encouragements as I continue to learn more about myself and pursue my interests. Anne, thank you for being just a call away when I felt overwhelmed and for always cheering me on. 公公, I am truly fortunate to engage with you in important conversations that expand my views and am indebted to you for the opportunity to attend Scripps.

Lastly, thank you to all the friends who have served as important confidants and support systems for and beyond this project. This would not be where it is without you.

**References:**

10 Dead in California as Wildfires Spread on West Coast. (2020, September 10). *The New York Times*. https://www.nytimes.com/2020/09/10/us/fires-oregon-california-wa-state.html
*2020 Fire Season Incident Archive | CAL FIRE*. (n.d.). Retrieved July 25, 2023, from https://www.fire.ca.gov/incidents/2020

Aguilera, R., Gershunov, A., Ilango, S. D., Guzman-Morales, J., & Benmarhnia, T. (2020). Santa Ana Winds of Southern California Impact PM2.5 With and Without Smoke From Wildfires. *GeoHealth*, *4*(1), e2019GH000225. https://doi.org/10.1029/2019GH000225

Ailshire, J. A., & Crimmins, E. M. (2014). Fine Particulate Matter Air Pollution and Cognitive Function Among Older US Adults. *AMERICAN JOURNAL OF EPIDEMIOLOGY*, *180*(4), 359–366. https://doi.org/10.1093/aje/kwu155

Alig, R., Kline, J., & Lichtenstein, M. (2004). Urbanization on the US Landscape: Looking Ahead in the 21st Century. *Landscape and Urban Planning*, *69*, 219–234. https://doi.org/10.1016/j.landurbplan.2003.07.004

Auch, R., Taylor, J., & Acevedo, W. (2004). Urban growth in American cities: Glimpses of U.S. urbanization. In *Circular* (1252). U.S. Geological Survey. https://doi.org/10.3133/cir1252

Axelrad, Daniel, Adams, Kristen, Chowdhury, Farah, D'Amico, Louis, and Douglass, Erika. 2013. "America's Children and the Environment, Third Edition." *Environmental Protection Agency*, January, 504. https://www.epa.gov/sites/default/files/2015-10/documents/ace3_criteria_air_pollutants.pdf

Bar, S., Parida, B. R., Mandal, S. P., Pandey, A. C., Kumar, N., & Mishra, B. (2021). Impacts of partial to complete COVID-19 lockdown on $NO_2$ and $PM_{2.5}$ levels in major urban cities of Europe and USA. *CITIES*, *117*, 103308. https://doi.org/10.1016/j.cities.2021.103308

Bell, J. N. B., Power, S. A., Jarraud, N., Agrawal, M., & Davies, C. (2011). The effects of air pollution on urban ecosystems and agriculture. *INTERNATIONAL JOURNAL OF SUSTAINABLE DEVELOPMENT AND WORLD ECOLOGY*, *18*(3), 226–235. https://doi.org/10.1080/13504509.2011.570803

CDC. (2023, March 15). *CDC Museum COVID-19 Timeline*. Centers for Disease Control and Prevention. https://www.cdc.gov/museum/timeline/covid19.html

Ellis, A. W., & Marston, M. L. (2020). Late 1990s' cool season climate shift in eastern North America. *CLIMATIC CHANGE*, *162*(3), 1385–1398. https://doi.org/10.1007/s10584-020-02798-z

Fast, J. D., Doran, J. C., Shaw, W. J., Coulter, R. L., & Martin, T. J. (2000). The evolution of the boundary layer and its effect on air chemistry in the Phoenix area. *Journal of Geophysical Research: Atmospheres*, *105*(D18), 22833–22848. https://doi.org/10.1029/2000JD900289

Fry, R. (2020, July 29). Prior to COVID-19, Urban Core Counties in the U.S. Were Gaining Vitality on Key Measures. *Pew Research Center's Social & Demographic Trends Project*. https://www.pewresearch.org/social-trends/2020/07/29/prior-to-covid-19-urban-core-counties-in-the-u-s-were-gaining-vitality-on-key-measures/

Gaffney, J. S., Marley, N. A., Drayton, P. J., Doskey, P. V., Kotamarthi, V. R., Cunningham, M. M., Baird, J. C., Dintaman, J., & Hart, H. L. (2002). Field observations of regional and urban impacts on NO2, ozone, UVB, and nitrate radical production rates in the Phoenix air basin. *ATMOSPHERIC ENVIRONMENT*, *36*(5), 825–833. https://doi.org/10.1016/S1352-2310(01)00528-3

Givati, A., & Rosenfeld, D. (2004). Quantifying precipitation suppression due to air pollution. *JOURNAL OF APPLIED METEOROLOGY*, *43*(7), 1038–1056. https://doi.org/10.1175/1520-0450(2004)043<1038:QPSDTA>2.0.CO;2

Goss, M., Swain, D. L., Abatzoglou, J. T., Sarhadi, A., Kolden, C. A., Williams, A. P., & Diffenbaugh, N. S. (2020). Climate change is increasing the likelihood of extreme autumn wildfire conditions across California. *ENVIRONMENTAL RESEARCH LETTERS*, *15*(9), 094016. https://doi.org/10.1088/1748-9326/ab83a7

Grundstrom, M., Hak, C., Chen, D., Hallquist, M., & Pleijel, H. (2015). Variation and co-variation of $PM_{10}$, particle number concentration, $NO_x$ and $NO_2$ in the urban air—Relationships with wind speed, vertical temperature gradient and weather type. *ATMOSPHERIC ENVIRONMENT*, *120*, 317–327. https://doi.org/10.1016/j.atmosenv.2015.08.057

Hernández Ayala, J. J., Mann, J., & Grosvenor, E. (2021). Antecedent Rainfall, Excessive Vegetation Growth and Its Relation to Wildfire Burned Areas in California. *Earth and Space Science*, *8*(9), e2020EA001624. https://doi.org/10.1029/2020EA001624

*Household air pollution*. (n.d.). Retrieved September 24, 2023, from https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health

Kheyroddin, R., & Ghaderi, M. (2023). Railways and urban expansion: How does rail transport affect urban expansion in metropolitan areas? (Warsaw and Copenhagen case). *INTERNATIONAL PLANNING STUDIES*, *28*(2), 124–141. https://doi.org/10.1080/13563475.2022.2137476

Langford, A. O., Senff, C. J., Alvarez II, R. J., Aikin, K. C., Ahmadov, R., Angevine, W. M., Baidar, S., Brewer, W. A., Brown, S. S., James, E. P., McCarty, B. J., Sandberg, S. P., & Zucker, M. L. (2023). Were Wildfires Responsible for the Unusually High Surface Ozone in Colorado During 2021? *Journal of Geophysical Research: Atmospheres*, *128*(12), e2022JD037700. https://doi.org/10.1029/2022JD037700

Lee, Y. C., Shindell, D. T., Faluvegi, G., Wenig, M., Lam, Y. F., Ning, Z., Hao, S., & Lai, C. S. (2014). Increase of ozone concentrations, its temperature sensitivity and the precursor factor in South China. *TELLUS SERIES B-CHEMICAL AND PHYSICAL METEOROLOGY*, *66*, 23455. https://doi.org/10.3402/tellusb.v66.23455

Li, J., Georgescu, M., Hyde, P., Mahalov, A., & Moustaoui, M. (2015). Regional-scale transport of air pollutants: Impacts of Southern California emissions on Phoenix ground-level ozone concentrations. *Atmospheric Chemistry and Physics*, *15*. https://doi.org/10.5194/acp-15-9345-2015

Ling, H., Ting, K. H., Shaharuddin, A., Aiyub, K., & Yaakob, M. (2011). *Air quality and human health in urban settlement: Case study of Kuala Lumpur city*. 510–515. https://doi.org/10.1109/CSSR.2010.5773831

Liu, F., Zhu, Y., & Zhao, Y. (2008). Contribution of motor vehicle emissions to surface ozone in urban areas: A case study in Beijing. *INTERNATIONAL JOURNAL OF SUSTAINABLE DEVELOPMENT AND WORLD ECOLOGY*, *15*(4), 345–349. https://doi.org/10.3843/SusDev.15.4:9

Liu, Q., Harris, J. T., Chiu, L. S., Sun, D., Houser, P. R., Yu, M., Duffy, D. Q., Little, M. M., & Yang, C. (2021). Spatiotemporal impacts of COVID-19 on air pollution in California, USA. *SCIENCE OF THE TOTAL ENVIRONMENT*, *750*, 141592. https://doi.org/10.1016/j.scitotenv.2020.141592

Lv, B., Zhang, B., & Bai, Y. (2016). A systematic analysis of PM2.5 in Beijing and its sources from 2000 to 2012. *ATMOSPHERIC ENVIRONMENT*, *124*, 98–108. https://doi.org/10.1016/j.atmosenv.2015.09.031

Masiol, M., Squizzato, S., Rich, D. Q., & Hopke, P. K. (2019). Long-term trends (2005-2016) of source apportioned PM2.5 across New York State. *ATMOSPHERIC ENVIRONMENT*, *201*, 110–120. https://doi.org/10.1016/j.atmosenv.2018.12.038

Meng, X., Jiang, J., Chen, T., Zhang, Z., Lu, B., Liu, C., Xue, L., Chen, J., Herrmann, H., & Li, X. (2023). Chemical drivers of ozone change in extreme temperatures in eastern China. *SCIENCE OF THE TOTAL ENVIRONMENT*, *874*, 162424. https://doi.org/10.1016/j.scitotenv.2023.162424

*Most Polluted Cities | State of the Air*. (n.d.). Retrieved November 20, 2023, from https://www.lung.org/research/sota/city-rankings/most-polluted-cities

Naresh, R., Sundar, S., & Shukla, J. (2007). Modelling the Removal of Gaseous Pollutants and Particulate Matters from the Atmosphere of a City. *Nonlinear Analysis: Real World Applications*, *8*, 337–344. https://doi.org/10.1016/j.nonrwa.2005.08.005

Naresh, R., Sundar, S., & Upadhyay, R. K. (2006). Modelling the removal of primary and secondary air pollutants by precipitation. *INTERNATIONAL JOURNAL OF NONLINEAR SCIENCES AND NUMERICAL SIMULATION*, *7*(3), 285–294.
Nations, U. (n.d.). *Population*. United Nations; United Nations. Retrieved September 24, 2023, from https://www.un.org/en/global-issues/population

Pitiranggon, M., Johnson, S., Haney, J., Eisl, H., & Ito, K. (2021). Long-term trends in local and transported PM2.5 pollution in New York City. *ATMOSPHERIC ENVIRONMENT*, *248*, 118238. https://doi.org/10.1016/j.atmosenv.2021.118238

Pleijel, H., Grundstrom, M., Karlsson, G. P., Karlsson, P. E., & Chen, D. (2016). A method to assess the inter-annual weather-dependent variability in air pollution concentration and deposition based on weather typing. *ATMOSPHERIC ENVIRONMENT*, *126*, 200–210. https://doi.org/10.1016/j.atmosenv.2015.11.053

Quijas, S., Schmid, B., & Balvanera, P. (2010). Plant diversity enhances provision of ecosystem services: A new synthesis. *BASIC AND APPLIED ECOLOGY*, *11*(7), 582–593. https://doi.org/10.1016/j.baae.2010.06.009

Requia, W. J., Coull, B. A., & Koutrakis, P. (2019). Regional air pollution mixtures across the continental US. *ATMOSPHERIC ENVIRONMENT*, *213*, 258–272. https://doi.org/10.1016/j.atmosenv.2019.06.006

Rivera, C., Stremme, W., Barrera, H., Friedrich, M. M., Grutter, M., Garcia-Yee, J., Torres-Jardon, R., & Gerardo Ruiz-Suarez, L. (2015). Spatial distribution and transport patterns of $NO_2$

in the Tijuana—San Diego area. *ATMOSPHERIC POLLUTION RESEARCH*, *6*(2), 230–238. https://doi.org/10.5094/APR.2015.027

Safford, H. D., Paulson, A. K., Steel, Z. L., Young, D. J. N., & Wayman, R. B. (2022). The 2020 California fire season: A year like no other, a return to the past or a harbinger of the future? *Global Ecology and Biogeography*, *31*(10), 2005–2025. https://doi.org/10.1111/geb.13498

Shaw, S., & Van Heyst, B. (2022). An Evaluation of Risk Ratios on Physical and Mental Health Correlations due to Increases in Ambient Nitrogen Oxide (NOx) Concentrations. *ATMOSPHERE*, *13*(6), 967. https://doi.org/10.3390/atmos13060967

Sliney, D. H., & Wengraitis, S. (2006). Is a differentiated advice by season and region necessary? *Progress in Biophysics and Molecular Biology*, *92*(1), 150–160. https://doi.org/10.1016/j.pbiomolbio.2006.02.007

Song, C., He, J., Wu, L., Jin, T., Chen, X., Li, R., Ren, P., Zhang, L., & Mao, H. (2017). Health burden attributable to ambient PM2.5 in China. *ENVIRONMENTAL POLLUTION*, *223*, 575–586. https://doi.org/10.1016/j.envpol.2017.01.060

Song, C., Ma, C., Zhang, Y., Wang, T., Wu, L., Wang, P., Liu, Y., Li, Q., Zhang, J., Dai, Q., Zou, C., Sun, L., & Mao, H. (2018). Heavy-duty diesel vehicles dominate vehicle emissions in a tunnel study in northern China. *SCIENCE OF THE TOTAL ENVIRONMENT*, *637*, 431–442. https://doi.org/10.1016/j.scitotenv.2018.04.387

Storey, E. A., Stow, D. A., O'Leary, J. F., Davis, F. W., & Roberts, D. A. (2021). Does short-interval fire inhibit postfire recovery of chaparral across southern California? *SCIENCE OF THE TOTAL ENVIRONMENT*, *751*, 142271. https://doi.org/10.1016/j.scitotenv.2020.142271

US EPA, O. (2013, February 22). *Summary of the Clean Air Act* [Overviews and Factsheets]. https://www.epa.gov/laws-regulations/summary-clean-air-act

US EPA, O. (2015, June 8). *Progress Cleaning the Air and Improving People's Health* [Reports and Assessments]. https://www.epa.gov/clean-air-act-overview/progress-cleaning-air-and-improving-peoples-health

US EPA, O. (2016a, April 19). *Particulate Matter (PM) Basics* [Overviews and Factsheets]. https://www.epa.gov/pm-pollution/particulate-matter-pm-basics

US EPA, O. (2016b, July 5). *Timeline of Nitrogen Dioxide (NO2) National Ambient Air Quality Standards (NAAQS)* [Data and Tools]. https://www.epa.gov/no2-pollution/timeline-nitrogen-dioxide-no2-national-ambient-air-quality-standards-naaqs

US EPA, O. (2020, July 10). *Ozone National Ambient Air Quality Standards (NAAQS)* [Other Policies and Guidance]. https://www.epa.gov/ground-level-ozone-pollution/ozone-national-ambient-air-quality-standards-naaqs

Wang, P., Qiao, X., & Zhang, H. (2020). Modeling PM2.5 and O-3 with aerosol feedbacks using WRF/Chem over the Sichuan Basin, southwestern China. *CHEMOSPHERE*, *254*, 126735. https://doi.org/10.1016/j.chemosphere.2020.126735

Wang, S. S.-C., Leung, L. R., & Qian, Y. (2023). Extension of Large Fire Emissions From Summer to Autumn and Its Drivers in the Western US. *Earth's Future*, *11*(7), e2022EF003086. https://doi.org/10.1029/2022EF003086

Wang, W., & Guo, Y. (2009). Air Pollution PM2.5 Data Analysis In Los Angeles Long Beach With Seasonal ARIMA Model. *2009 INTERNATIONAL CONFERENCE ON ENERGY AND ENVIRONMENT TECHNOLOGY, VOL 3, PROCEEDINGS*, 7-+. https://doi.org/10.1109/ICEET.2009.468

Wegman, F., & Katrakazas, C. (2021). Did the COVID-19 pandemic influence traffic fatalities in 2020? A presentation of first findings. *Iatss Research*, *45*(4), 469–484. https://doi.org/10.1016/j.iatssr.2021.11.005

Zhang, J., Wei, Y., & Fang, Z. (2019). Ozone Pollution: A Major Health Hazard Worldwide. *FRONTIERS IN IMMUNOLOGY*, *10*, 2518. https://doi.org/10.3389/fimmu.2019.02518

Zhao, Y., & Ma, T. (2020). Research and Countermeasures on the Influence of Air Pollution on Human Health and Fitness. *3RD INTERNATIONAL CONFERENCE ON AIR POLLUTION AND ENVIRONMENTAL ENGINEERING*, *631*, 012023. https://doi.org/10.1088/1755-1315/631/1/012023

Zhao, Y., & Ma, T. (2021). Research and Countermeasures on the Influence of Air Pollution on Human Health and Fitness. *IOP Conference Series: Earth and Environmental Science*, *631*(1), 012023. https://doi.org/10.1088/1755-1315/631/1/012023

**Supplementary Material**

**Table S1.** Summary Data for concentrations of $PM_{2.5}$, $NO_2$, and Ozone from 2018-2022 for 7 regions.

| Metropolitan Area | $PM_{2.5}$ (µg/m³ LC) | | | $NO_2$ (ppb) | | | Ozone (ppm) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Median | Mean | SD | Median | Mean | SD | Median |
| **SF** | 9.1 | 9.8 | 7.3 | 17.1 | 8.24 | 16.0 | 0.034 | 0.0079 | 0.034 |
| **LA** | 11 | 5.1 | 9.7 | 24.7 | 9.51 | 23.7 | 0.046 | 0.011 | 0.046 |
| **NY** | 6.8 | 3.8 | 6.0 | 28.6 | 10.5 | 27.3 | 0.035 | 0.013 | 0.033 |
| **Houston** | 9.3 | 4.3 | 8.5 | 16.4 | 7.67 | 15.1 | 0.036 | 0.012 | 0.034 |
| **Chicago** | 9.2 | 4.1 | 8.5 | 27.3 | 9.76 | 26.6 | 0.037 | 0.013 | 0.036 |
| **Atlanta** | 8.7 | 3.9 | 8.0 | 24.1 | 9.35 | 22.9 | 0.036 | 0.013 | 0.035 |
| **Phoenix** | 8.5 | 4.6 | 7.6 | 30.9 | 9.16 | 31.7 | 0.048 | 0.011 | 0.048 |

**Appendix**

*I. Downloading EPA Data in .CSV Format*

1. Open https://www.epa.gov/outdoor-air-quality-data/air-data-concentration-plot
2. Select pollutant
3. Select year period
4. Select Geographic Area by county
5. Select "Plot Data"
6. Download data into .csv files

*II. Downloading NOAA Data in .CSV Format*

1. Open https://www.ncei.noaa.gov/cdo-web/search
2. Select "Daily Summaries" as Weather Observation Type
3. Select date range
4. Select "Counties"
5. Under "Enter a Search Term," write the appropriate state abbreviation (e.x. California as CA)
6. Identify county of interest and select "Add to cart" and visit cart
7. Select "Custom GHCN-Daily CSV" then "Continue"
8. Select "Station Name," "Geographic Location," and "Metric" units
9. Select "Precipitation," "Air Temperature," and "Wind" then "Continue"
10. Enter email address and select "Submit Order"
11. Download .csv file from email

*III. Small Scale Air Filter Sampling and Collection*

$PM_{2.5}$ concentrations were collected in Bangkok, Thailand and the Bernard Field Station in Claremont, CA using 47 mm Whatman QMA quartz filters (Fig. S1). MesaLabs PQ200 Ambient Air Particulate Samplers (Fig. S2, S3, S4) were placed in each location and filters were removed and replaced on 24-hour intervals for each sampler. Once removed, filters were carefully placed into Petri dishes labeled with a filter number and the dates of collection before the dishes were wrapped in parafilm and placed in a freezer for later analysis.

*Figure S 1. 47 mm Whatman QMA quartz filters*


*Figure S 4. MesaLabs PQ200 Ambient Air Particulate Sampler front exterior*


*Figure S 3. MesaLabs PQ200 Ambient Air Particulate Sampler side exterior*


*Figure S 2. MesaLabs PQ200 Ambient Air Particulate Sampler interior*

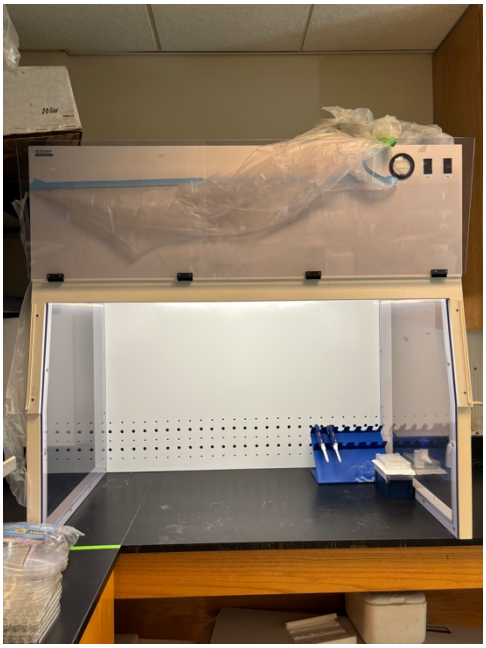

*Figure S 5. Cleatech 24" Portable Ductless Exhaust Fume Hood*

12 digestion tubes were cleaned prior to filter digestion. Each bottle underwent a scrubbing processing using a bottle scrubber and Alconox followed by three rinses with deionized (DI) water and an additional three rinses using Milli-Q water. The process was repeated for each tube and all tubes were then left to dry under a Cleatech 24" Portable Ductless Exhaust Fume Hood (Fig. S5, S6).

An empty digestion was run using a 16% HCl acid solution made from 25 mL HCl and 85 mL MilliQ water. 8 mL of the 16% acid solution was placed into each tube using a micropipette and tube stoppers were carefully placed on the tube openings so that the areas of the tube stoppers that were to be exposed to the tube contents were not touched. Lids were then screwed on. This process was repeated for all 12 tubes under a fume hood before being placed into the PerkinElmer Titan MPS Microwave Sample Preparation System (Fig. S8) to run the digestion.

*Figure S 6. Cleatech 24" Portable Ductless Exhaust Fume Hood interior ventilation*

While the empty digestion was run, filters were retrieved from a freezer and prepared for digestion under a Cleatech 24" Portable Ductless Exhaust Fume Hood. The hood surface, tweezers, and scissors were cleaned using DI water and tissue wipes before parafilm from all Petri dishes holding filters were removed. Filters were carefully cut in half so that they could be used for multiple analyses. During this process, tweezers only touched the outside section of the filters where particles had not been collected. Tweezers and scissors were cleaned and dried using DI water and tissue wipes between the cutting process for each filter. Once all filters had been cut, the whole mass and half mass of each filter was collected using an Adam Equipment PW 184 Analytical Balance (Fig. S7). To do this, am empty Petri dish lid was placed on the scale. Once tared, both halves of the filter were carefully placed on the lid in the

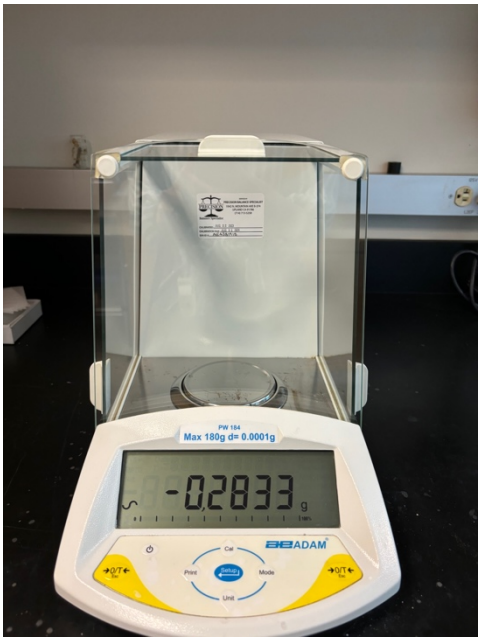


*Figure S 7. Adam Equipment PW 184 Analytical Balance*

scale using tweezers so that the area of particle collection was not touched. The full mass was weighed and recorded before one of the halves was removed so that the mass of a half filter was weighed and recorded. Once all masses had been recorded, the half filter whose mass was recorded was carefully cut into six pieces using tweezers and scissors that had been cleaned and dried with DI water and tissue wipes so that they could easily be inserted into the digestion tubes. All pieces of the filter were placed back into their initial Petri dish. Recording of the whole mass, half mass, and filter cutting was repeated for all filters. Tweezers and scissors were cleaned using DI water and tissue wipes between each filter.

Following the initial empty digestion, each tube was removed from the microwave digester and rinsed twice with 4 mL of MilliQ water. Tweezers were cleaned and dried with DI water and tissue wipes before the half filters that had previously been cut into six pieces were placed into tubes and the tube and corresponding filter were recorded. This process was repeated for all 12 tubes and filters and tweezers were cleaned between each process. Meanwhile, a 35% HCl acid solution was made under a fume hood using 55mL HCl and 55 mL MilliQ water. Once all six pieces of the half filters had been dropped into their tubes, 8 mL of the 35% acid solution was put into each tube using a micropipette. Tube stoppers were carefully placed on the tube openings so that the areas of the tube stoppers that were to be exposed to the tube contents were not touched and lids were then screwed on. One tube acted as the control group, as it had no filter and only contained the acid

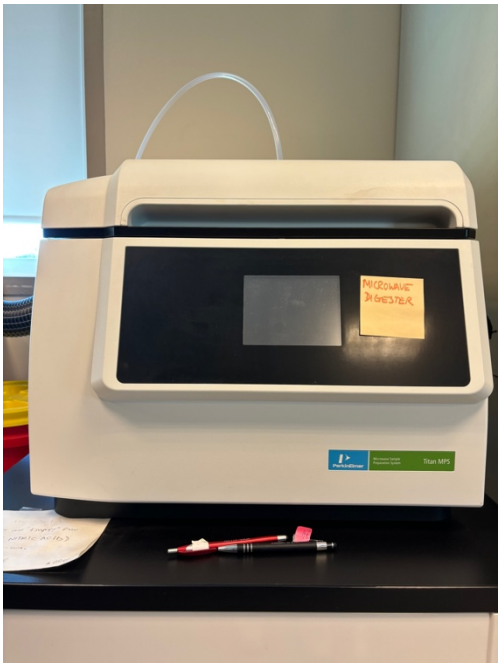


*Figure S 8. PerkinElmer Titan MPS Microwave Sample Preparation System*

solution. This process was repeated for all 12 tubes before being placed into the PerkinElmer Titan MPS Microwave Sample Preparation System to run the digestion.

Once completed, digestion tubes were removed from the digester and cooled for five minutes under the fume hood. 15 mL centrifuge tubes were labeled with filter numbers or "control." In the fume hood, cooled digestion tubes were unscrewed while holding the lid's small opening facing away from the body to avoid exposure to dangerous gaseous chemicals. A micropipette was used to add 8 mL of MilliQ water to each tube followed by return of stoppers and gentle shaking of the tube to combine the contents of the tube. The stopper was then removed, and the contents of the tube was poured into the centrifuge tube with the corresponding label for the filter number. This process was repeated for all remaining tubes.

## *IV. R Code Used*

**Merging Pollutant Data and Mean Calculations**
*The code below is merging and calculating mean values for LA's $PM_{2.5}$ concentrations. Region and pollutant name must be appropriately changed for merging operations for different pollutants or regions.*

```
library(dplyr)
library(plyr)

#list.files combines data from the same folder
setwd("~/Downloads/Summer Research/LA/PM2.5")
data2 <- ldply(list.files(), read.csv, header=TRUE)

#selecting columns 1 & 5 (Date & PM2.5)
data1=data2 %>% select(1,5)
data1

#formatting the data to the correct "date" format and
data1$Date <- format(as.Date(data1$Date))
data1$Date <- as.Date(data1$Date,format = "%Y-%m-%d")
data1$Date <- format(as.Date(data1$Date))
data1

#Average the PM2.5 values for each day
```

```
data=aggregate(data1$PM2.5, by=list(data1$Date), mean)
data

#Produce CSV file that has 2 columns: date & the average PM2.5
write.csv(data,file="~/Desktop/data.csv",quote=F,row.names = F)
```

**$PM_{2.5}$ Overall Concentration Comparison Significance Boxplots**

```
library(dplyr)
library(plyr)
library(ggplot2)
library(hrbrthemes)
library(agricolae)
library(multcompView)

mastersheet<-read.csv("Combined Air Pollution.csv")
mastersheet

mastersheet$Region<-as.character(mastersheet$Region)

model=lm(mastersheet$PM2.5 ~ mastersheet$Region )
ANOVA=aov(model)
TUKEY <- TukeyHSD(x=ANOVA, 'mastersheet$Region', conf.level=0.95)
plot(TUKEY , las=1 , col="brown")
TUKEY

generate_label_df <- function(TUKEY, variable){

  Tukey.levels <- TUKEY[[variable]][,4]
  Tukey.labels <- data.frame(multcompLetters(Tukey.levels)['Letters'])
  Tukey.labels$Region=rownames(Tukey.labels)
  Tukey.labels=Tukey.labels[order(Tukey.labels$Region) , ]
  return(Tukey.labels)
}

LABELS <- generate_label_df(TUKEY , "mastersheet$Region")

PM2.5Significance <- ggplot(mastersheet,aes(x=Region,y=PM2.5, fill=Region)) +
  scale_y_log10()+
  geom_boxplot() + theme_bw() + theme(legend.position = "none") +
  labs(y="PM2.5 (ug/m3 LC)", x="Region")+
```

```
  ggtitle("PM2.5 Concentrations (Log Scale) from 2018-2022")+
  geom_text(data=LABELS, aes(y=190, label=LABELS[,1]), col='black', size=4) +
theme(panel.grid = element_blank()) +
  scale_fill_brewer(palette="Set2") +
  scale_x_discrete(breaks=c("Atlanta", "Chicago", "Houston", "LA","NY", "Phoenix", "SF"),
                   labels=c("Atlanta", "Chicago", "Houston", "LA","NY", "Phoenix", "SF"))

PM2.5Significance + geom_hline(yintercept=15, linetype="dashed", color = "red")
```

**$NO_2$ Overall Concentration Comparison Significance Boxplots**

```
mastersheet$Region<-as.character(mastersheet$Region)

model=lm(mastersheet$NO2 ~ mastersheet$Region )
ANOVA=aov(model)
TUKEY <- TukeyHSD(x=ANOVA, 'mastersheet$Region', conf.level=0.95)
plot(TUKEY , las=1 , col="brown")
TUKEY

generate_label_df <- function(TUKEY, variable){

  Tukey.levels <- TUKEY[[variable]][,4]
  Tukey.labels <- data.frame(multcompLetters(Tukey.levels)['Letters'])
  Tukey.labels$Region=rownames(Tukey.labels)
  Tukey.labels=Tukey.labels[order(Tukey.labels$Region) , ]
  return(Tukey.labels)
}

LABELS <- generate_label_df(TUKEY , "mastersheet$Region")

NO2Significance <- ggplot(mastersheet,aes(x=Region,y=NO2, fill=Region)) +
  geom_boxplot() + theme_bw() + theme(legend.position = "none") +
  labs(y="NO2 (ppb)", x="Region")+
  ggtitle("NO2 Concentrations from 2018-2022")+
  geom_text(data=LABELS, aes(y=80, label=LABELS[,1]), col='black', size=4) +
theme(panel.grid = element_blank()) +
  scale_fill_brewer(palette="Set2") +
  scale_x_discrete(breaks=c("Atlanta", "Chicago", "Houston", "LA","NY", "Phoenix", "SF"),
                   labels=c("Atlanta", "Chicago", "Houston", "LA","NY", "Phoenix", "SF"))
```

```
NO2Significance + geom_hline(yintercept=53, linetype="dashed", color = "red")
```

**Ozone Overall Concentration Comparison Significance Boxplots**

```
mastersheet$Region<-as.character(mastersheet$Region)

model=lm(mastersheet$Ozone ~ mastersheet$Region )
ANOVA=aov(model)
TUKEY <- TukeyHSD(x=ANOVA, 'mastersheet$Region', conf.level=0.95)
plot(TUKEY , las=1 , col="brown")
TUKEY

generate_label_df <- function(TUKEY, variable){

  Tukey.levels <- TUKEY[[variable]][,4]
  Tukey.labels <- data.frame(multcompLetters(Tukey.levels)['Letters'])
  Tukey.labels$Region=rownames(Tukey.labels)
  Tukey.labels=Tukey.labels[order(Tukey.labels$Region) , ]
  return(Tukey.labels)
}

LABELS <- generate_label_df(TUKEY , "mastersheet$Region")

OzoneSignificance <- ggplot(mastersheet,aes(x=Region,y=Ozone, fill=Region)) +
  geom_boxplot() + theme_bw() + theme(legend.position = "none") +
  labs(y="Ozone (ppm)", x="Region")+
  ggtitle("Ozone Concentrations from 2018-2022")+
  geom_text(data=LABELS, aes(y=0.10, label=LABELS[,1]), col='black', size=4) +
theme(panel.grid = element_blank()) +
  scale_fill_brewer(palette="Set2") +
  scale_x_discrete(breaks=c("Atlanta", "Chicago", "Houston", "LA","NY", "Phoenix", "SF"),
                   labels=c("Atlanta", "Chicago", "Houston", "LA","NY", "Phoenix", "SF"))


OzoneSignificance + geom_hline(yintercept=0.07, linetype="dashed", color = "red")
```

**Summary Data Table**

```
library(dplyr)

data2 <- read.csv(file = "Combined Air Pollution.csv")
head(data2)

data=data2 %>% select(1, 3)
data

data3<- na.omit(data)

data_summary2 <- data3 %>%
  group_by(Region) %>%
  dplyr::summarize_all(list(my_mean = mean,my_median = median,
                            my_sum = sum,
                            my_sd = sd)) %>%
as.data.frame()

data_summary2

write.csv(data_summary2 ,file="~/Desktop/data.csv",quote=F,row.names = F)
```

**Annual Concentration Boxplots per Region with Significance**

*The code below is for annual boxplots for SF's $PM_{2.5}$ concentrations. Region and pollutant name must be appropriately changed for plot formation of different pollutants or regions.*

```
library(dplyr)
library(plyr)
library(ggplot2)
library(hrbrthemes)
library(agricolae)
library(multcompView)

mastersheet<-read.csv("Combined Air Pollution.csv")
mastersheet
data <- filter(mastersheet, Region == "SF")
data

data$Year<-as.character(data$Year)
```

```
model=lm(data$PM2.5 ~ data$Year )
ANOVA=aov(model)
TUKEY <- TukeyHSD(x=ANOVA, 'data$Year', conf.level=0.95)
plot(TUKEY , las=1 , col="brown")
TUKEY

generate_label_df <- function(TUKEY, variable){

  Tukey.levels <- TUKEY[[variable]][,4]
  Tukey.labels <- data.frame(multcompLetters(Tukey.levels)['Letters'])
  Tukey.labels$Year=rownames(Tukey.labels)
  Tukey.labels=Tukey.labels[order(Tukey.labels$Year) , ]
  return(Tukey.labels)
}

LABELS <- generate_label_df(TUKEY , "data$Year")

ggplot(data,aes(x=Year,y=PM2.5,fill=Year)) +
  geom_boxplot()   + facet_wrap(~Region) + theme_bw() + theme(legend.position = "none")
+ ylim(0,60) +
  geom_text(data=LABELS, aes(y=58, label=LABELS[,1]), col='black', size=4) +
theme(panel.grid = element_blank()) +
  theme(plot.title = element_blank(), axis.title.x = element_blank(), axis.title.y =
element_blank()) + scale_x_discrete(breaks=c("2018","2019","2020", "2021", "2022"),
labels=c("18", "19", "20", "21", "22"))
```

**Seasonality by Region Concentration Boxplots with Significance**

*The code below is for seasonal boxplots for SF's $PM_{2.5}$ concentrations. Region and pollutant name must be appropriately changed for plot formation of different pollutants or regions.*

```
library(dplyr)
library(plyr)
library(ggplot2)
library(hrbrthemes)
library(agricolae)
library(multcompView)

mastersheet<-read.csv("Combined Air Pollution.csv")
mastersheet
data <- filter(mastersheet, Region == "SF")
data

data$Season<-as.character(data$Season)
```

```
model=lm(data$PM2.5 ~ data$Season )
ANOVA=aov(model)
TUKEY <- TukeyHSD(x=ANOVA, 'data$Season', conf.level=0.95)
plot(TUKEY , las=1 , col="brown")
TUKEY

generate_label_df <- function(TUKEY, variable){

  Tukey.levels <- TUKEY[[variable]][,4]
  Tukey.labels <- data.frame(multcompLetters(Tukey.levels)['Letters'])
  Tukey.labels$Season=rownames(Tukey.labels)
  Tukey.labels=Tukey.labels[order(Tukey.labels$Season) , ]
  return(Tukey.labels)
}

LABELS <- generate_label_df(TUKEY , "data$Season")

ggplot(data,aes(x=Season,y=PM2.5,fill=Season)) +
  geom_boxplot()   + facet_wrap(~Region) + theme_bw() + theme(legend.position = "none")
+ ylim(0,60) +
  geom_text(data=LABELS, aes(y=58, label=LABELS[,1]), col='black', size=4) +
theme(panel.grid = element_blank()) +
  theme(plot.title = element_blank(), axis.title.x = element_blank(), axis.title.y =
element_blank()) + scale_x_discrete(limits=c("Summer","Spring","Fall", "Winter"),
                labels=c("Summer","Spring","Fall", "Winter"))
```

**Merging Weather Data and Mean Calculations**

*The code below is merging and calculating mean values for SF's weather factors. Region name must be appropriately changed for merging operations for different regions.*

```
setwd("~/Downloads/Summer Research/SF/Weather Data")
data2 <- ldply(list.files(), read.csv, header=TRUE)

datawind=data2 %>% select(6,7)
data1 <- na.omit(datawind)
data1

data1$DATE <- format(as.Date(data1$DATE))
data1$DATE <- as.Date(data1$DATE,format = "%Y-%m-%d")
data1$DATE <- format(as.Date(data1$DATE))

data=aggregate(data1$AWND, by=list(data1$DATE), mean)
data

write.csv(data,file="~/Desktop/data.csv",quote=F,row.names = F)
```

```
dataRF=data2 %>% select(6,8)
data3 <- na.omit(dataRF)
data3

#Format all of the columns correctly
data3$DATE <- format(as.Date(data3$DATE))
data3$DATE <- as.Date(data3$DATE,format = "%Y-%m-%d")
data3$DATE <- format(as.Date(data3$DATE))
data3

data=aggregate(data3$PRCP, by=list(data3$DATE), mean)
data

write.csv(data,file="~/Desktop/data.csv",quote=F,row.names = F)

dataT=data2 %>% select(6,9)
data4 <- na.omit(dataT)
data4

data4$DATE <- format(as.Date(data4$DATE))
data4$DATE <- as.Date(data4$DATE,format = "%Y-%m-%d")
data4$DATE <- format(as.Date(data4$DATE))
data4

data=aggregate(data4$T, by=list(data4$DATE), mean)
write.csv(data,file="~/Desktop/data.csv",quote=F,row.names = F)
```

**Weather Factor Correlation Plots**

*The code below is for a correlation plot between pollutant concentrations and weather factors for SF. Region name must be appropriately changed for different regions.*

```
library(ggcorrplot)
library(ggplot2)
library(dplyr)
library(tidyverse)
library(corrplot)
library(Hmisc)

data2 <- read.csv(file = "Combined Air Pollution.csv")
head(data2)

colnames(data2)[6]="Wind Speed"
colnames(data2)[7]="Precipitation"
colnames(data2)[8]="Temperature"
```

```
data1 <- filter(data2, Region == "SF")
data1

data=data1 %>% select(3, 4, 5, 6, 7, 8)
data

data<- na.omit(data)

M<-cor(data, use="pairwise.complete.obs")
round(cor(data),
      digits = 2 )

corrplot(cor(data),
         method = "number",
         type = "upper"
)
cor.vals = cor(data)
cor.p = cor.mtest(data, conf.level = 0.95)$p
rownames(cor.p) = rownames(cor.vals)
colnames(cor.p) = colnames(cor.vals)
cor.p

corrplot(cor(data), type="upper",  method = "number",
         p.mat = cor.p, sig.level = 0.05)
```

**VMT Correlation Plots**

*The code below is for a correlation plot between pollutant concentrations and VMT data for SF Region name must be appropriately changed for different regions.*

```
library(ggcorrplot)
library(ggplot2)
library(dplyr)
library(tidyverse)
library(corrplot)
library(Hmisc)

data2 <- read.csv(file = "Combined Air Pollution.csv")
head(data2)

colnames(data2)[12]="VMT"
```

```
data1 <- filter(data2, Region == "SF")
data1

data=data1 %>% select(3, 4, 5, 12)
data

data<- na.omit(data)

M<-cor(data, use="pairwise.complete.obs")
round(cor(data),
      digits = 2 )

corrplot(cor(data),
         method = "number",
         type = "upper"
)
cor.vals = cor(data)
cor.p = cor.mtest(data, conf.level = 0.95)$p
rownames(cor.p) = rownames(cor.vals)
colnames(cor.p) = colnames(cor.vals)
cor.p

corrplot(cor(data), type="upper",  method = "number",
         p.mat = cor.p, sig.level = 0.05)
```