

Accelerating California's Transition to Zero Emission Vehicles

Final Report

Primary Client: Union of Concerned Scientists

Advisor: Pablo Saide Peralta

Team Members:

Vivian Kha

Anna Schneider

Weihao Shao

Aria Soeprono

Michelle Yep

University of California, Los Angeles

Table of Contents

1. Abstract	3
2. Introduction	4
3. Research Questions	4
4. Methodology	5
5. Results and Discussion	18
6. Outreach	58
7. Conclusion and Recommendations	60
8. Literature Cited	63
9. Appendix	65

Abstract

The purpose of this study was to determine where older vehicles are located in California, what the demographics of these communities are, and what impacts older vehicles have on air quality and public health. The main focus was on pre-2003 light-duty passenger vehicles, before the implementation of the Low Emission Vehicle II standards in 2004. We used a database of vehicles across the state of California to locate where the older vehicles are, and based off of those locations, we estimated the resulting NO_x tailpipe emissions. We then found the resulting air quality due to combined primary and secondary PM_{2.5} pollutants, and estimated the health impacts in terms of morbidity and mortality. All of these components had an environmental justice aspect that helped identify the demographics in most need of immediate relief. Our analysis uncovered great inequity in terms of race/ethnicity and income. This study may give policy-makers insight on geographic areas to prioritize for public health and environmental justice concerns when designing programs to accelerate the transition to zero-emission vehicles. While it is clear that older vehicles contribute disproportionately to air pollution in certain communities in California, the benefits of removing older polluting vehicles is unevenly distributed across the state, and the amount of older vehicles as well as their resulting emissions and secondary impacts are concentrated primarily in parts of Southern California, Central Valley, and the Bay Area. Further research into sensitivity testing for various model year cutoffs may be necessary to analyze if there is a different priority cutoff date than 2003 to help California reach its air quality goals.

Introduction

Transportation is the primary source of greenhouse gas pollution in the United States. According to the U.S. Environmental Protection Agency, the transportation sector generates the largest share of greenhouse gas emissions by burning fossil fuel for cars, trucks, ships, trains, and planes (U.S. Environmental Protection Agency, 2020-b). Those that are fueled by gasoline and diesel release harmful air pollutants such as sulfur dioxide, nitrogen oxides, and particulate matter (PM) through tailpipe emissions. These passenger vehicles typically have a lifetime average of 10-15 years. However, they need to be replaced by zero-emission vehicles (ZEVs) at a faster rate in order to meet California's long-term goal of carbon neutrality by 2045.

As defined by Caltrans, ZEVs are vehicles that produce zero exhaust emissions of any criteria pollutant regardless of its operational mode (Caltrans, 2021). This can include battery electric vehicles (BEVs), which run exclusively on batteries, as well as hydrogen fuel cell electric vehicles (FCEVs), which are powered by electricity stored in hydrogen fuel. As of 2020, the California New Car Dealers Association reports that true ZEVs currently represent approximately 6% of new vehicle sales in California (California New Car Dealers Association, 2021). In September of 2020, California Governor Gavin Newsom had announced a new Executive Order declaring that by 2035, all new cars and passenger trucks sold in California will be ZEVs (Office of Governor Gavin Newsom, 2020). This would help achieve more than 35% reduction in statewide greenhouse gas emissions from current levels. However, even if all newly bought vehicles are mandated to be ZEVs, there will still be negative air quality impacts from older vehicles that are still being driven.

Unfortunately, these negative impacts of tailpipe emissions and its effects on local and regional air quality are disproportionately distributed across the population, with the burden weighing heaviest on low-income and minority communities. Our goal is to identify the areas that are most in need of switching to zero-emission vehicles to provide immediate relief from air pollution and the resulting health impacts.

Research Questions

1. Where are older combustion engine vehicles located throughout the state of California?
2. What are the demographics of locations with high numbers of older combustion engine vehicles?
3. What are the resulting air quality impacts of older vehicles based on emissions modeling?
4. What are the cumulative health impacts of older (pre-2003) combustion engine passenger vehicles?
5. How can we accelerate California's transition to zero emission vehicles?
6. What are some key takeaways for policymakers?

Methodology

1.1 Vehicle Fleet Database

Raw data processing

To analyze the pattern and distribution of older combustion vehicles throughout the state of California, we used RStudio data analysis. We prepared the vehicle fleet data from the California Air Resources Board database, EMISSION FACTOR (EMFAC) website by filtering out the variables and only keeping what we needed (California Air Resources Board, 2021-b). The data for this light fleet vehicle task required raw data processing of the vehicle category, fuel type, census block group code, model year, and vehicle population - all variables of the EMFAC data. We modified the vehicle category types and filtered them to only include passenger vehicles (LDA), light-duty trucks (LDT1), and light-duty trucks (LDT2). For the fuel type, we focused only on the diesel and gasoline categories. We used 2003 as the boundary year to differentiate the older and newer vehicle since there was a significant decrease in NO_x emissions after 2003 (Figure 2.2.1) due to the 2004 Low Emission Vehicle II standards. Furthermore, we removed all the “unknowns” and “scrubbed” rows from the EMFAC under variable census block group code. Approximately 7% of the data were “unknown” and “scrubbed” values, totaling 7966093 rows of data removed. This removal might have slightly shifted the data's standard deviation, but since it was a minor influence on the total of about ten million rows, we did not look further into it. The data outputs were in the format of comma-separated files (CSV files), shapefiles, and image files.

Additionally, we determined the percentage of unknown census codes and the trend of vehicle development from 1973 to 2019 (separated by decade). Following that, we filtered data input with census block group code and joined it with two spatial variables with a GEOID code: income and race. We also determined the percentage of the households with only one vehicle registered to the address (hhv=1) and compared it to the emissions rate. The dataset was then used for spatial analysis at three different resolutions: county, census tract, and block group.

Mapping distribution

We found that using RStudio to plot tract and block group resolutions resulted in some polygons clustered in certain locations. To avoid this issue, we used ArcMap and QGIS to do analysis for the whole state and instead used RStudio to focus only on certain counties. By comparing the different resolutions, we decided to use the finest resolution for mapping the distribution of the old vehicles. A logarithmic scale was used to analyze the county-level data due to the order of magnitude differences among counties. We used the census tract and block group levels to compute fractions of the vehicle population and further filtered out the rows with vehicle population over 20,000 since these are possibly vehicle renting agencies or airport parking.

1.2 Emissions

For the emissions task, the goal was to combine the two datasets of vehicles and their corresponding emissions. We obtained the total tailpipe emissions data for California (at the statewide level) for several different air pollutants — $PM_{2.5}$, NO_x , Reactive Organic Gases (ROG), SO_x , and NH_3 —from the Emissions Inventory from EMFAC (California Air Resources Board, 2021-a) for the base year 2019. Because the vehicle database included model years from 1973 to 2019, 2019 was chosen as the base year to ensure that the "end years" matched up and wouldn't create an imbalance of vehicles. The purpose of using those particular pollutants was due to their role in atmospheric chemistry and their impact on air pollution. Using RStudio, we filtered the data to passenger and light-duty tiers (LDT1, LDT2) of vehicles, as these are the focus of the overall project. We focused on these particular vehicles because they are among those commonly found in communities that we are aiming to assist. We then differentiated the data by the vehicle manufacture date, considering a year-mark of 2003 that we identified as being representative of the boundary between production of "older" versus "newer" vehicles. Then we made some conversions with the data to make it compatible with the vehicle data that resulted from the first task, such as matching up the vehicle categories and the model years. We calculated the per vehicle emissions rates by dividing the total emissions for each pollutant, year, and vehicle category by the corresponding vehicle population given by EMFAC. We then joined the two datasets together and multiplied the per vehicle emission rates by the vehicle population from the vehicle database in the first section, in order to give the total emissions with respect to the census data's registered population. The file was then joined with TIGER/Line block group and tract level shapefiles for California provided by the US Census Bureau ("TIGER/Line Shapefiles, n.d."), so that the end file could be used to map the results in InMAP in the Air Quality section. We looked at correlations between NO_x and $PM_{2.5}$ emission rates in 2019 and the characteristics of the vehicles, like the model year, fuel type, and vehicle category, and the emissions per mile travelled.

There were some assumptions that were made in this process. First off, we had to decide whether we would use total tailpipe emissions data from EMFAC at the county level or state level. In order to decide, we obtained EMFAC data for the statewide level as well as several counties. We then plotted the base 2019 emissions for gasoline LDA vehicles in tons/yr/vehicle and g/mile per model year for four air pollutants (NO_x , $PM_{2.5}$, ROGs, and SO_x) using the data from all of these different places. Figure 1.2.1 shows results for NO_x and $PM_{2.5}$. We came to the conclusion that the statewide data would be able to approximate the emissions for most California counties, so we made the assumption that the emissions for all vehicles in the vehicle database would be approximately equal to the values given in the EMFAC state emissions data.

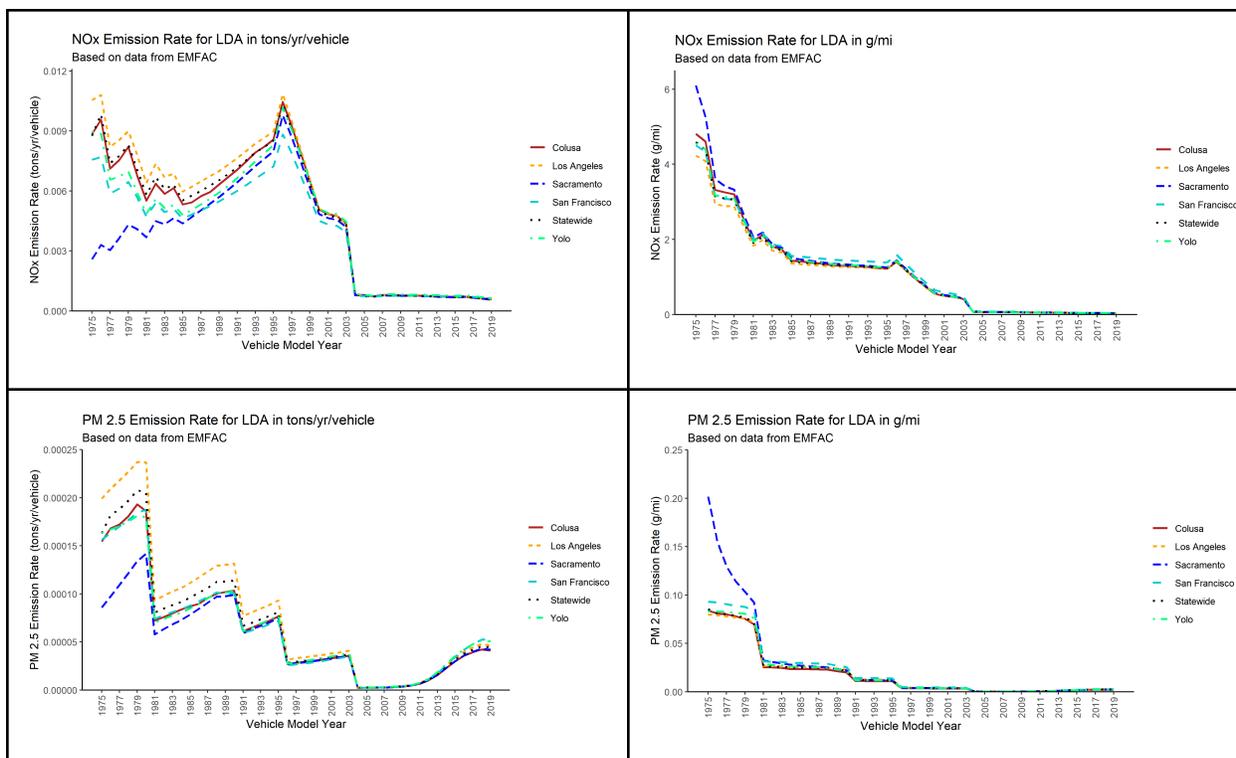


Figure 1.2.1 Four graphs indicate the emissions of NO_x and $\text{PM}_{2.5}$ in both tons/yr/vehicle and g/mi per model year. Several curves on the graphs indicate the statewide data as well as data from the Colusa, Los Angeles, Sacramento, San Francisco, and Yuba Counties. These counties as well as Glenn and Yuba counties (not shown here) in order to get a variety of locations.

In addition, the vehicle database included cars with model years back to 1973, while the EMFAC data included cars with model years back to 1975. In order to account for the difference, we assumed vehicles of model years 1973 or 1974 had the same emissions data as vehicles from 1975, for which there was usually data from EMFAC. However, some observations were missing from EMFAC so sometimes it was necessary to make similar extrapolations and fill in the missing data with data from the nearest available model year present.

Finally, the 1973 vehicles in the vehicle database included both vehicles made in 1973 and any vehicles with model years before 1973. Because of this, we assumed that all of the vehicles in this category had the same emissions, the data extrapolated from 1975 or the closest model year with data available.

1.3 Air Quality Modeling

Troubleshooting the model

The air quality group tested the InMAP modeling tool prior to running official simulations in order to adjust several parameters and troubleshoot for potential errors.

We adjusted the extent of the model so that the model covered California and minimized coverage of neighboring states (Figure 1.3.1(a) and Figure 1.3.1(b)). This helped shorten the time it took to model air quality since there were less areas to be analyzed.

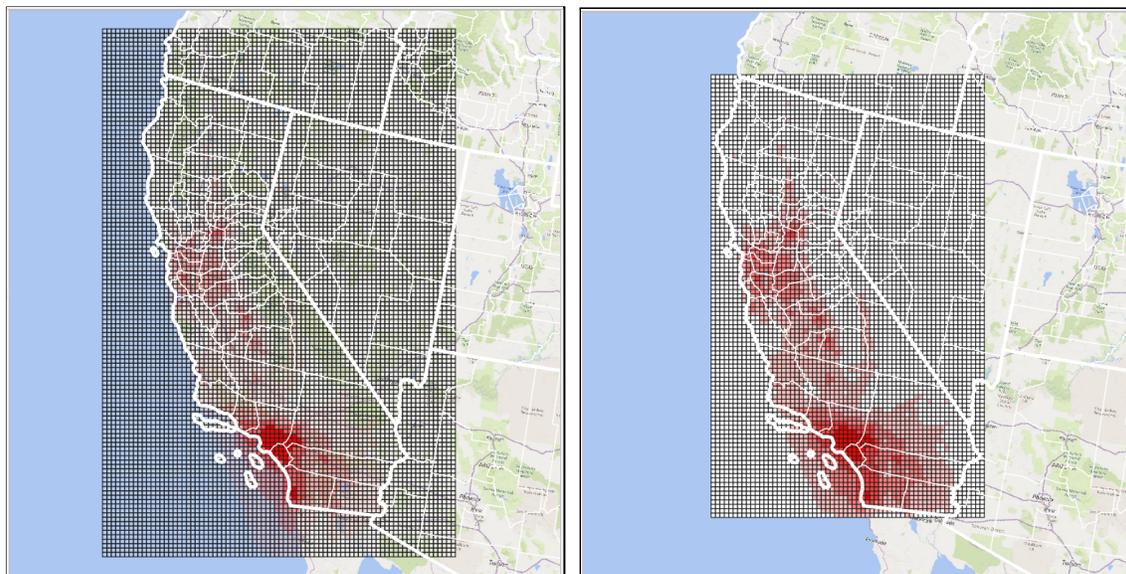


Figure 1.3.1 Display of InMAP outputs before (a) and after (b) the extent of the model was adjusted to be as close as possible to the boundary of California.

Then, we checked the various files InMAP needed to model air quality, including a baseline meteorological and pollutant dataset and a population dataset. The baseline meteorological and pollutant dataset was used to simplify the model by making assumptions of atmospheric chemistry. The population dataset was required to relate “human exposures by employing higher resolution in urban areas and lower spatial resolution in rural and remote locations and at high altitude” (Tessum et al. 2017). In other words, areas with high population density were considered to be potential for high air pollution activity and vice versa.

Finally, we tested how long each simulation would take, on average, at different resolutions in order to budget sufficient time for modeling. Several emission datasets were obtained from the Union of Concerned Scientists as test files to achieve this. Based on the run times of more than 50 simulations, it took approximately 20 to 30 minutes to run a 12-km resolution, 30 to 40 minutes for a 4-km resolution, and 2.5 hours for a 1-km resolution. Variations in run time are mainly due to the resolution of the model. The higher the resolution, the longer the simulation took to run. After checking for errors and making adjustments, we ran InMAP using the official emissions obtained from the emission group.

Understanding the simulation process

InMAP outputs annual average exposure to PM_{2.5} coming from the emissions used as input through a series of grid cells. Each of these grid cells is 12-km in length and width. When

InMAP detects high population density and high emissions in one of the grid cells, it conducts further analysis to predict the resulting $PM_{2.5}$ concentration in ug/m^3 . If the resolution of the model is set at 12-km, InMAP analyzes the grid and predicts the concentration of the particular grid cell immediately. However, if the resolution of the model is set at 4-km or 1-km, respectively, and if any of the grid cells has high population density and high emissions, then InMAP divides the grid cells down to 4-km or 1-km, respectively, for further analysis before predicting the concentrations of each of these grid cells. As such, the model output shows a variable grid—where some grids are larger and some smaller (Figure 1.3.2).

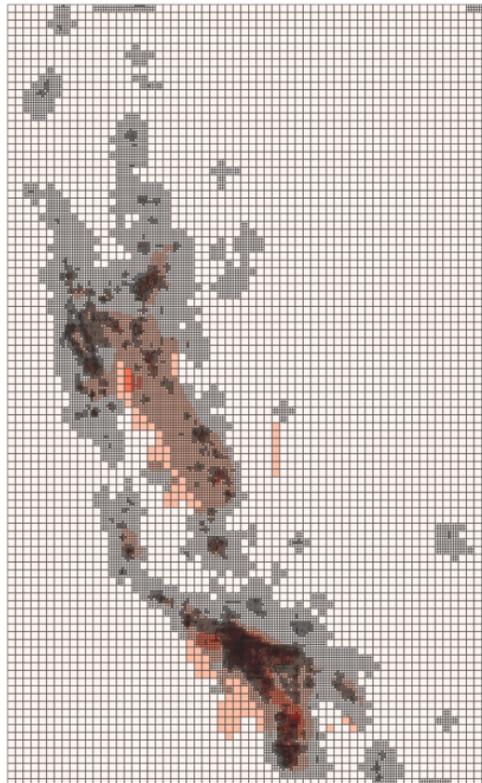


Figure 1.3.2 *InMAP modeling output.*

Formatting the output for the next modeling phase

Before utilizing this model output to predict health effects through BenMAP, we regridged the output from a variable to a fixed grid, as a fixed grid was required for BenMAP. To do this, we first used QGIS to create a reference grid of 1-km. We then overlaid the 1-km grid over the variable-grid InMAP output and used the vector geoprocessing tool “intersection” to combine the two grid layers. Intersecting the two grid layers produced a final grid where each grid cell is 1-km in length and width everywhere in the grid, with air quality concentrations data in each of the grid cells. Figure 1.3.3 shows the result of regridging the variable grid InMAP output.

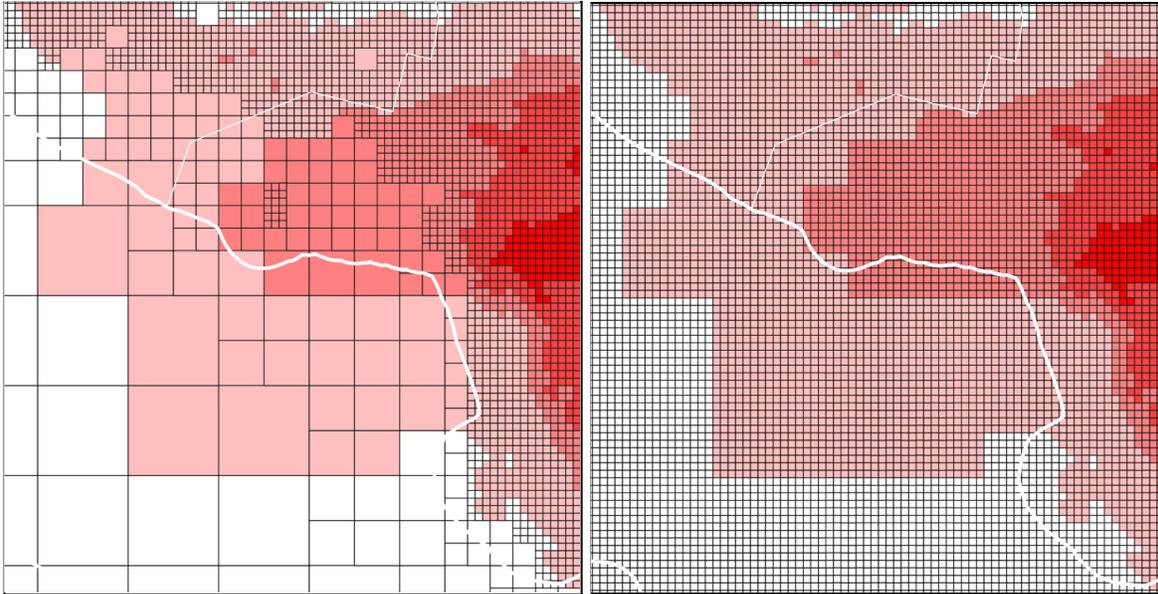


Figure 1.3.3(a) Variable grid. **Figure 1.3.3(b)** Fixed grid.

1.4 Health Effects Modeling

Overview

Next, we focused on modeling potential health effects of air pollutants from the previously-computed air quality model, as well as an assigned cost estimate to be used for further analysis. The aim of this step in the study was to find out the health impact of older emitting vehicles by testing a hypothetical scenario in which all vehicles older than 2003 are removed. We used an EPA standard health effects equation to map the health effect difference between a baseline (business as usual) scenario and a control (with policy applied) scenario, to show mortalities that could be avoided if action were taken to reduce older vehicle emissions. To do so, we utilized the Environmental Benefits Mapping and Analysis Program-Community Edition (BenMAP-CE), which was developed by the Environmental Protection Agency, but can be accessed as an open-source software. BenMAP-CE is a useful tool for this because it has a graphical and user-friendly interface with pre-existing databases to enhance the user-inputs (U.S. Environmental Protection Agency, n.d.).

We used a health function for all-cause mortality by Krewski, et al. (2009), which is one of the EPA Standard Functions for BenMAP-CE. In addition, we ran health impact results for heart disease and lung cancer with equations from the same study, as well as incidence of hospital admissions taken from Ostro, et al. (2009), and childhood asthma incidence from Tétreault, et al. (2016). We used these health impact functions to estimate morbidity and mortality, given the existing BenMAP-CE database of health endpoints and existing census population data by county for the year of 2019. By computing the net difference between a

baseline and control scenario, we were able to identify areas with the largest health-related benefits from reducing older vehicles.

The modeling process takes air quality scenarios as inputs, which were obtained through InMAP for the state of California at a 4-km resolution and 1-km resolution, and then clipped to specific regions of interest at the 1-km level for the air quality regions of Los Angeles, South Coast, San Joaquin, and the Bay Area. In each modeling iteration, the baseline scenario represented a “business as usual” approach including all vehicles from 1997 to 2019, whereas the control scenario represents the vehicles newer than 2003 only. BenMAP then takes the difference to produce incidence results caused by passenger vehicles that are older than 2003. The population dataset was obtained from the 2019 census data (for consistency with the vehicle data) and was pre-loaded into Benmap.

To get the InMAP output data ready for BenMAP input, the shapefiles were brought into QGIS and functions were used to add column and row data to the attribute table, since this is the primary identifying system that BenMAP uses to link each air quality value to a specific spatial area. A reference grid had to be created with the same numbering system for rows and columns, which was then brought into BenMAP and cross-walked with the existing built-in reference grids that Ben-MAP has, so that it can attribute geographic location to the air quality grids when they are uploaded to BenMAP. BenMap takes inputs for air quality data in a specific format for baseline and control scenarios that utilize the same reference grid to relate the air quality data to spatial information. The results were downloaded as csv files and shapefiles, and the “Point Estimate” column, which was the estimated quantification of the negative health outcome for each endpoint of interest, was used for mapping processes.

The statewide 4-km resolution was aggregated by county level to get mortality counts per county. We then used these county-level mortality counts to estimate a cost associated with increased mortality from older emitting vehicles based on the Environmental Protection Agency's estimated value of a statistical life (Environmental Protection Agency, 2020-a).

Understanding BenMAP Health Functions

BenMAP-CE Version 1.5.8 comes pre-loaded with sets of health incidence functions, which may be associated with $PM_{2.5}$ or Ozone, and were obtained by the EPA from various health studies. These existing health functions are divided into datasets which include the following: “EPA Standard Health Functions (2021),” “Expert PM25 Functions,” “GEMM Functions,” and “Additional Health Functions.” For this study, we only used functions related to $PM_{2.5}$, because our air quality impacts were only computed for $PM_{2.5}$. Each function has a health endpoint that correlates to either mortality or a specific condition of morbidity. Within both the mortality and morbidity metrics, there are subcategories of endpoints, which specify the condition (e.g. Lung Cancer could be a subcategory of Mortality). We performed health effects analysis for six health functions, based on particular endpoints that were chosen in accordance with the interest of the Union of Concerned Scientists.

From the “EPA Standard Health Functions (2021),” dataset, we utilized Tetrault et al. (2016) for childhood asthma incidence from ages 0-4 and ages 5-17 to find the total childhood asthma incidence from PM_{2.5} concentrations, as well as Ostro, et al. (2009) for “Hospital Admissions, All Respiratory” for our analysis of hospitalizations due to respiratory impacts of PM_{2.5} emissions. Lastly, all of our mortality analysis came from Krewski, et al. (2009), a study which produced three health functions related to mortality, including all-cause mortality, Ischemic heart disease mortality, and lung cancer mortality. Since the prevalence of Ischemic heart disease mortality and lung cancer mortality are reflected within all-cause mortality, the majority of this health impacts research analyzed in this study came from the mortality (All-Cause) health function. However, the figures corresponding to the other health functions are included in the Appendices F-L.

Preparing Reference Grids

To test out the methods of preparing reference grids, we started with InMAP air quality outputs which were obtained as shapefiles, each with an ID column signifying the location within the assigned geographic coordinate system. However, to properly connect this coordinate system to BenMAP's grid system, column and row information is necessary. This was added using the field calculator in QGIS. First, a new grid was created using the “Create Grid” function in QGIS, and was set to the extent and resolution of the original shapefile. This function outputs a grid in the same coordinate system as the first, but with added x and y columns, in units of meters. To convert the “X” and “Y” columns to “Row” and “Column” columns, a standardization function was used as outlined in Appendix A. This updated grid was saved as the reference grid shapefile (Note: reference grids did not contain air quality data), uploaded to the BenMAP system, and cross-walked with all of the existing grids so that the air quality values could later be assigned to a specific grid cell, allowing BenMAP to attribute health effects equations to the area. Once the grid definitions were assigned, BenMAP was ready for inputting the air quality information.

Preparing Input Files

Although BenMAP inputs are in the comma separated value (CSV) format, they still need to have column and row data to properly be assigned to the grid in BenMAP. Since InMAP outputs shapefiles with no X and Y coordinate information, the ID column was used to generate both column and row information for all of the air quality grid cells, using the QGIS field calculator and functions explained in Appendix B. These functions require the pre-existing knowledge of the number of columns and rows, which were obtained through the methods described in the “Preparing Reference Grids” section, and therefore results in an identical grid system to the reference grid. This is essential for proper validation of the air quality data to BenMAP.

Once the column and row information was added, the data was downloaded as a CSV, and was the origin of both the control and the baseline files. There are only two data inputs to a BenMAP analysis: the baseline or business as usual air quality scenario, and the control scenario with a change implemented (in this case a hypothetical removal of pre-2003 vehicles). To obtain the baseline, the “BaselinePM25” column was used from the pre-2003 vehicle simulation conducted in InMAP. The control was produced by subtracting the “TotalPM25” column from the “BaselinePM25” column, which essentially results in all of the air quality effects except those from vehicles older than 2003. This represents a scenario in which all of these older vehicles are removed. When these files are uploaded to BenMAP for analysis, it essentially subtracts out the control values from the baseline values to produce the total effects from pre-2003 vehicles alone. The last step that was necessary to prepare the CSV files for input was to add the variable columns for annual, seasonal, and daily metrics. Appendix C contains sample baseline and control CSV formats for reference.

Testing an alternative method of obtaining control scenario data

While in the process of conducting analysis, we realized that there was a second option for obtaining control scenario data by utilizing results from two separate air quality InMAP analyses (one for baseline and one for control). This alternative was pursued to test whether it resulted in different health effects estimates than the first method, but it was found that they only differed slightly. The second method was carried out as follows: the baseline was represented by the “TotalPM25” column of InMAP output from all passenger vehicles from 1997-2019, while the control was also obtained by the “TotalPM25” column from an InMAP output of 2003-2019 (excluding the pre-2003 vehicles). Therefore, it represents the same information as before in the first method, and then Benmap subtracts the control of post-2003 vehicles to get the effects of pre-2003 vehicles isolated. For the estimates of Los Angeles County at a 1-km level, 95 deaths were estimated for the first method whereas 99 deaths were estimated for the second method. Although it is difficult to deduce why these differences occurred, it is likely due to the non-linearities in air quality modeling from the two simulations. For the rest of the analysis, we continued with Method 1.

Simulations

This analysis started off at the 12km resolution for the testing and troubleshooting phase, and then progressed to 4-km and later 1-km to test how high of a resolution the computer could process. Early in the process, the simulations were limited by the capacity of the computer (8 gigabytes of RAM), and was only capable of producing 1-km grids for very small areas of interest. Later, the simulations were able to be run on a computer with significant processing power (32 gigabytes of RAM). Even so, there was a limit when it came to crosswalking a 1-km grid for the state of California, which was attempted several times, and on the most successful attempt, crashed after about 8 hours of running.

Therefore, this study was limited to analyzing smaller areas like counties and air quality districts at the 1-km level. The first attempts to do so utilized a grid with an extent clipped to the exact border of the region of interest, which resulted in unreasonably high health effects estimates around the edges of the clipped boundaries. To fix this occurrence, new rectangular reference grids had to be made that encompassed a buffer beyond the edges of the county. This method was successfully found to prevent this anomaly from occurring, and was used to model health effects for all of the areas of interest for the 1-km level. 1-km results were obtained for all of the previously described health functions for the following areas of interest: San Joaquin Valley Air District, South Coast Air District, San Francisco Bay Air District, Los Angeles County, and Orange County. The air district and county area shapefiles were downloaded from CARB's online GIS data library. Statewide analysis was conducted at the 12km level and the 4-km level, but was not able to proceed for a 1-km resolution.

Due to slight calculation differences within modeling resolutions, the 4-km level estimations turned out to be different from 1-km level by a factor of 1.1 times larger. Therefore, the 4-km mortality counts were assumed to be overestimated in comparison to the 1-km level by about 10%. Los Angeles County was used as the testing location for resolution comparisons, as it was computed at a 1-km with clipped reference extent, 1-km with rectangular reference extent, 4-km with clipped extent, 4-km with rectangular extent, 4-km aggregated by county in BenMAP, and 12km from the preliminary tests. We tested various resolutions and extents to gain an understanding of the relation between the resolution results, and these results are shown for Los Angeles County in Appendix D. All of the data from the simulations were downloaded as both a shapefile and a CSV and then QGIS was used to conduct further visual refinements.

Endpoint Estimates

The point estimate information from the resulting 4-km statewide CSV files were compiled into a single Excel file representing the outcome from all of the health effects equations. We computed aggregate values for all of the studied endpoints for the entire state at the 4-km level, as well as by county. Further, aggregated estimates were obtained for the three air districts of interest, in addition to Los Angeles County and Orange County. For the creation of the maps that are shown in section four of the results and discussion section, the output was displayed in graduated values by point estimate, and the color distribution was scaled by Jenks Natural Breaks for all of the maps, which is why each map has its own scale.

Aggregating by County

For the data that showed county aggregates, this was primarily conducted through BenMAP. After running the simulation for the 4-km level statewide, each health function endpoint point estimate was projected as an aggregation by county, and then the resulting data was downloaded as both a shapefile and CSV for further visualization in QGIS.

Cost Estimates

Conveniently, BenMAP is capable of conducting cost analysis through aggregation and valuation of health effects, but this process requires an extensive understanding of both health effects equations and valuation functions to choose to make an accurate estimate. The BenMAP guidelines provided by the EPA suggested researchers consult an epidemiologist with questions for health effects valuations, but this presented a limitation of this study due to lack of access to experts in the field of both medicine and economics.

Instead, another less precise estimation of health-related costs was produced using one health effects measure rather than an aggregation of multiple. Specifically, costs were estimated based on the total mortality related to all-causes from older, pre-2003 vehicles. This data column was taken from the results at the 4-km resolution statewide extent. Then, the calculation was computed in an Excel file using the a function of the described mortality column, multiplied by a constant factor of 9.398 million 2019 US dollars per person, leaving units millions of 2019 dollars. The estimated value of a statistical life as outlined by the EPA is 7.4 million 2006 dollars, and to adjust to 2019 dollars it needs to be multiplied by a factor of 1.27. This leaves the resulting dollar amount adjusted to the year of analysis, which is based on the year that we pulled census data from to conduct the health effects analysis.

1.5 Environmental Justice

Overview

The last study was built upon all of the previous studies' results in order to investigate the environmental justice aspect of vehicle tailpipe emissions pollution and identify the areas that policymakers should prioritize when it comes to accelerating the transition of California's vehicle fleet to ZEVs. We used Excel and ArcMap to produce calculations and maps that contribute to our analysis. Although there already exists a California Communities Environmental Health Screening Tool (CalEnviroScreen) produced by the California Environmental Protection Agency, the existing analysis omits race when evaluating vulnerable communities disproportionately burdened by pollution. However, CalEnviroScreen served as a useful tool to guide our analysis. We conducted analysis at the state level, as well as specific air districts, including the Bay Area, San Joaquin Valley, and South Coast.

Acquiring Socioeconomic Data

In order to conduct environmental justice analysis, data from the previous studies were compiled and integrated with socioeconomic factors, such as population size, race, and income. Socioeconomic data were pulled from the IPUMS National Historical Geographic Information System (NHGIS) (Manson, et al., 2020), the United States Census Bureau's American Census Survey (U.S. Census Bureau, 2020), California's Department of Housing and Community Development (State of California Department of Housing and Community Development, 2020),

and the United States Department of Housing and Urban Development (U.S. Department of Housing and Urban Development, 2021), at either the county level, census tract level, or both. We used the most recent data available, which at the time was 2015-2019 from American Census Survey dataset for racial data, 2014-2018 from American Census Survey for median income data, and 2020 for average household size and poverty level data. County level and census tract level boundaries were also found from NHGIS and were provided as GIS files. Additionally, we downloaded EPA's CalEnviroScreen GIS files to use as a reference in our analysis (California Environmental Protection Agency, 2018).

Producing Distribution Maps

To map the distribution of various races, we first used Excel to calculate the percentage of each census tract each race made up. For each race, we divided the number of people of a certain race by the total population within that census tract for all the census tracts within California. We then imported the CSV file to ArcMap and performed a join with the census level boundaries shapefile. To produce a choropleth map that visually displays the distribution for each race, we changed the symbology to be graduated colors and classified the values into 5 classification groups using natural breaks (jenks).

To map income at the census tract level, we decided to do the ratio of median income for each census tract to the poverty level, based on average household size. This decision was based on our goal to identify the areas that need financial assistance in switching to cleaner vehicles, so we wanted to use a similar method to the one used to determine Covered California and MediCal eligibility. We first downloaded data for average household size from the United States Census Bureau and data for median income from California's Department of Housing and Community development. Next, we used federal poverty data from the United States Department of Housing and Urban Development and matched the federal poverty level based on average household size for each census tract. Then we took the ratio of median income to federal poverty level for each census tract. After importing the csv file to ArcMap, we employed the same methods used to map racial distribution. However, instead of using natural jenks to classify the samples, we manually set the ranges to be ≤ 1.5 , 1.5-2, 2-2.5, 2.5-3, 3-4, and >4 . These ranges were based on those used to determine eligibility with Covered California and MediCal (Covered California, 2021).

Comparison Calculations for Analysis

Using the data prepared for the vehicle fleet database, we conducted further analysis. We looked at vehicles per household population and distribution at the block group level. We used the column "*numbers of vehicles registered at the same address (hhv)*" from the EMFAC data and split them into two groups based on the number of household vehicles registered at the address: 1 and more than 1. This allowed us to visualize the correlations between older

vehicles and local demographics. We then used these two groups (mainly focused on hhv=1) of data to identify trends in the fraction of older vehicles and racial makeup of the population.

Next, we wanted to compare emissions, air quality, and health effects for each race and median income to federal poverty level ranges. To do so, we used emissions data and air quality data provided from the previous steps. Using Excel, we merged the racial and income data with emissions and air quality data for each census tract.

As for analyzing emissions, we decided to use NO_x for our analysis because they represent tailpipe emissions more accurately by removing the effects of secondary pollutants. To calculate how much NO_x emissions each race contributed, we first divided the NO_x value for each census tract by the area of the corresponding census tract to obtain an emissions flux. We then computed population weighted emissions by multiplying the emissions flux for each census tract by the total number of people of a certain race in each certain census tract, and then divided by the total number of people of the certain race in California. This produced NO_x emissions flux for each census tract, which was then summed for each race to produce emissions values. The same formula was used to compute population weighted $\text{PM}_{2.5}$ and mortality.

To compare emissions, air quality, and health effects by the ratio of median income to federal poverty level, we took a slightly different approach. Using the filter option, we filtered out the ratio of median income to federal poverty level by the ranges previously mentioned. Then, instead of summing the values for each census tract, we took the average of emission flux, concentration of $\text{PM}_{2.5}$ exposure and mortality for each range.

Analysis of Specific Air Districts

For spatial comparison, we put together maps of emissions, air quality, mortality, CalEnviroScreen's disadvantaged communities, median income to federal poverty level, Hispanic/Latino distribution, African American Distribution, and Asian Distribution side by side. These comparison maps were produced at a finer scale, one for each specific air district region, including Bay Area Air District, San Joaquin Valley Air District, and South Coast Air District. Additionally, we conducted further analysis of the aforementioned air districts. To make calculations at this finer scale, we imported the CSV file to ArcMap and joined it to the census tract boundaries shapefile. We then clipped this layer to the appropriate air district boundaries and exported the table to Excel. We used the same emissions, air quality, and health effects methods employed for the state level analysis to complete calculations for each air district.

Results and Discussion

2.1 Vehicle Fleet Analysis

There was more than nine million vehicle information in the EMFAC dataset. The raw data was processed into more finite scales at different resolutions. We found that there were approximately 6 million older vehicles, accounting for about 25% of the total vehicle population. The older vehicles were mostly located in the South Coast region, especially Los Angeles County and Orange County, as well as the Bay Area region. This may be in correlation to the high population density in these areas. Another pattern possibly influenced by population density is the unexpected older vehicle distribution located in lower population density areas. Since the population density is significantly low in these areas and most people have older vehicles, this can result in a higher fraction. Further, we analyzed households by number of vehicles to find that the majority of the emissions are coming from households with multiple vehicles (Appendix E).

The maps produced aim to inform policymakers where the most older vehicles are located and their impacts on air pollution and health problems. By targeting specific areas where there is light-duty vehicle distribution, we can effectively accelerate the transition to ZEVs. After-2003 newer light-duty vehicles had around 17 million population, we were also concerned about this trend of increasing population. More vehicles will be considered as older each new year. We did not focus on this shifting pattern in our analysis, but it could be considered as a variable in a future analysis (Appendix R).

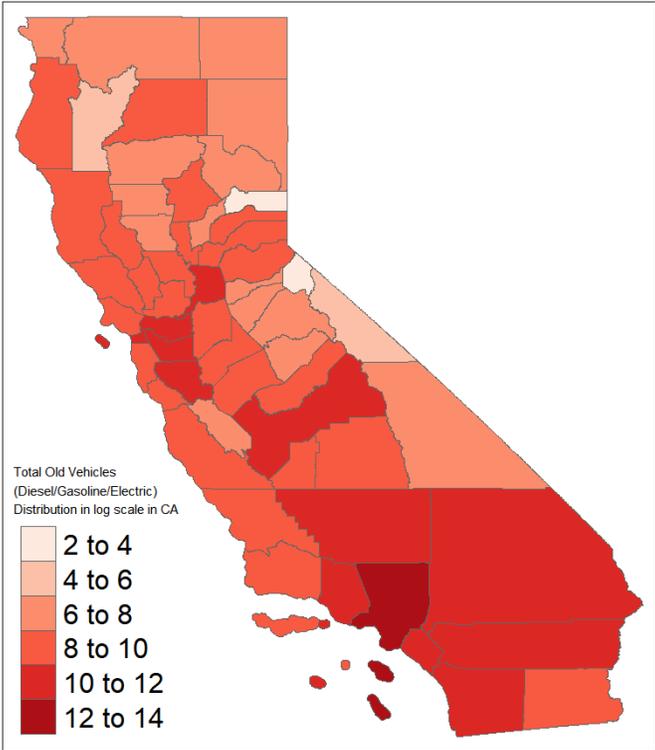


Figure 2.1.1 Map of older vehicle distribution in California at the county level

Figure 2.1.1 shows the older vehicle distribution pattern statewide at a county level. We determined that the South Coast region, Bay Area, and San Joaquin Valley have heavier older vehicle distribution. Los Angeles County shows the highest older vehicle concentration and this may be due to the factor of population density. The entire state has a total 6 million older light-duty vehicles out of 24 million light-duty vehicle population.

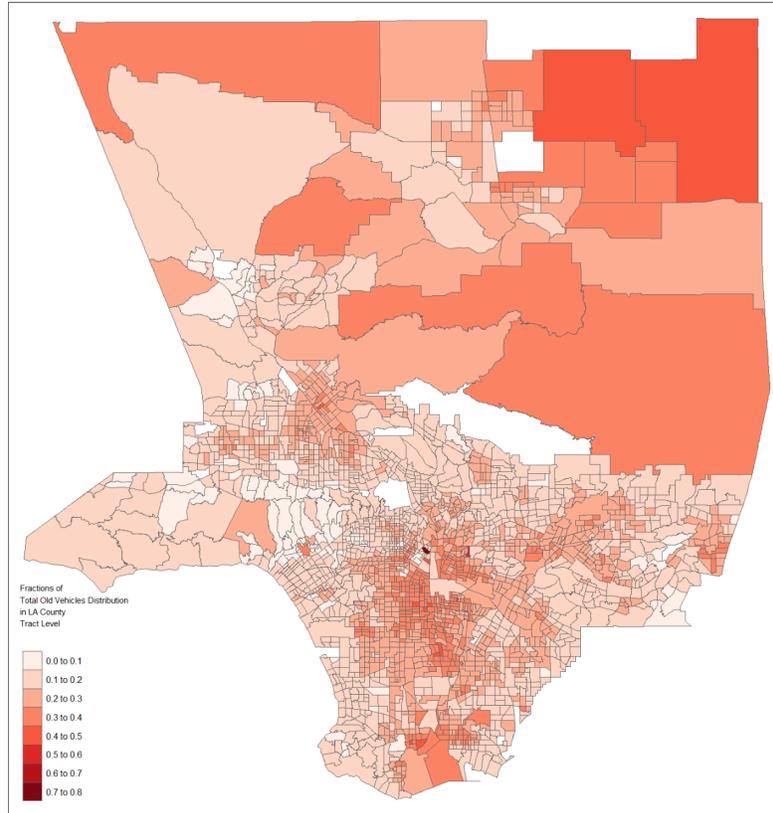


Figure 2.1.2 Map of older vehicle distribution in Los Angeles County at the census tract level

Figure 2.1.2 shows the older vehicle distribution in Los Angeles County at the census tract level. The map shows an older vehicle distribution mainly concentrated in South Central LA, and an unexpected pattern can also be seen in Northeast LA. The unexpected patterns in Northeast Los Angeles County are due to the population density factor. As mentioned previously, the Northeast Los Angeles County area has a low population density due to its geography, and this can lead to a higher fraction of the older vehicles since the analysis was not based on the vehicle per capita. Los Angeles County has around 300,000 older light-duty vehicles out of a total 1.6 million light-duty vehicle population.

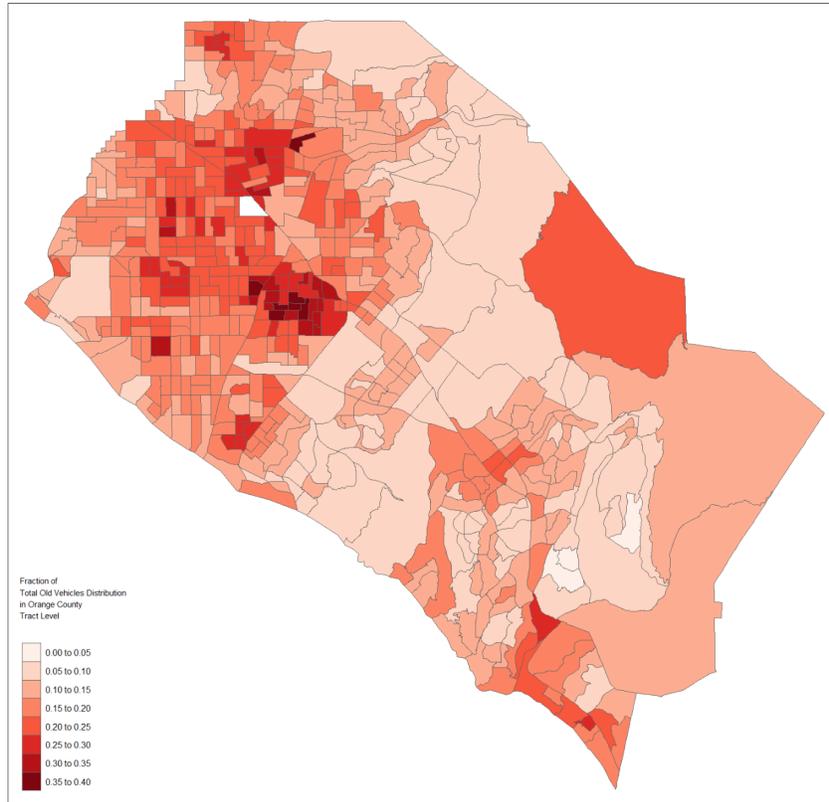


Figure 2.1.3 Map of older vehicle distribution in Orange County at the census tract level

Figure 2.1.3 spatially matches the mortality rate analyzed in the health effect subgroup, and it has a negative influence on $PM_{2.5}$ concentrated in the area where it has a higher older vehicle population. Orange County has around 80,000 older light-duty vehicles out of a total 470,000 light-duty vehicle population.

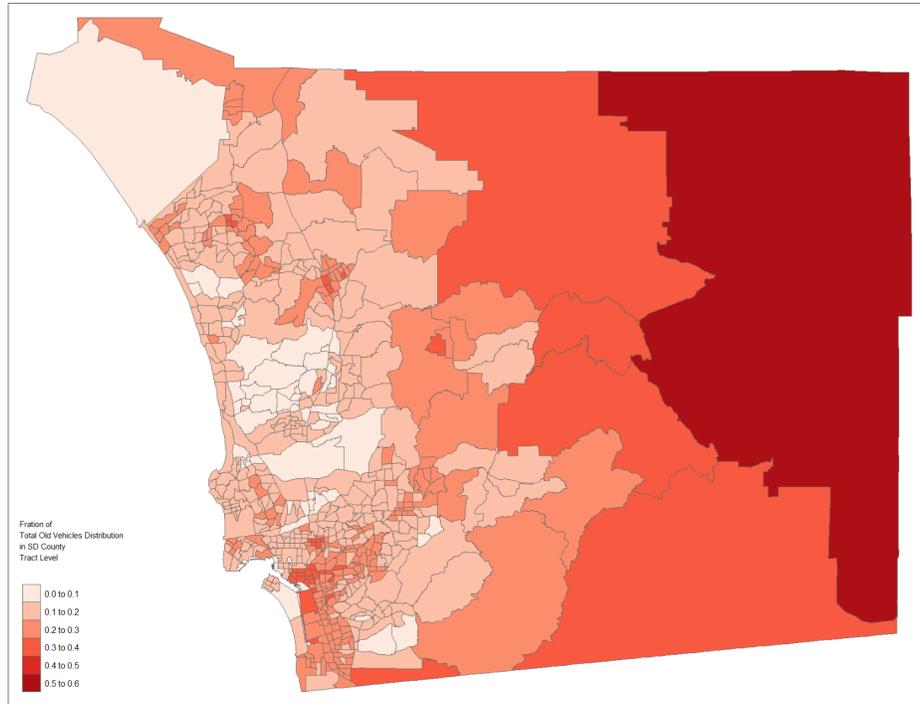


Figure 2.1.4 Map of older vehicle distribution in San Diego County at the census tract level

We also performed analysis on San Diego County (Figure 2.1.4), and noticed that in San Diego, older vehicles are concentrated in the east where there is low population density. Since we did not take the older vehicle per capita into account in our analysis, this may result in an unexpectedly high fraction of old vehicles in certain areas. On the other hand, a high older vehicle fraction could also be due to the higher population density. San Diego County has around 90,000 older light-duty vehicles out of a total 500,000 light-duty vehicle population.

2.2 Emissions Analysis

The graphs below were generated using the file from the joined vehicle and emissions data.

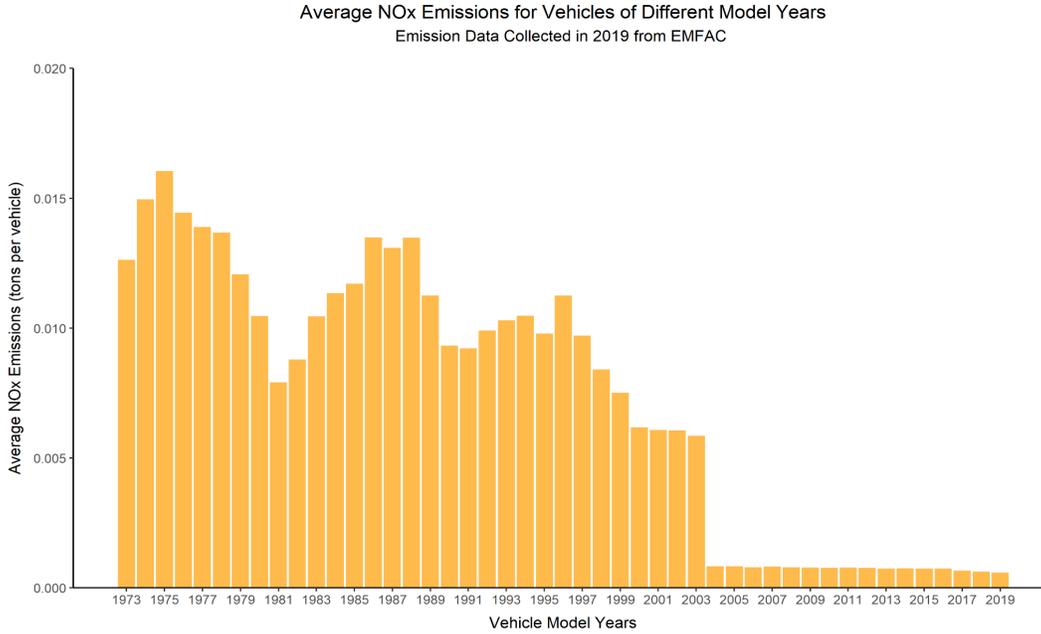


Figure 2.2.1 Graph depicting the mean per-vehicle NO_x emissions for cars from each model year from 1973-2019. The mean emissions are based on the vehicle database and emissions data collected by EMFAC for base year 2019.

Figure 2.2.1 shows the mean emissions of NO_x from the emissions data for 2019, differentiated by the model years of the cars. In calculating the mean emissions, any N/A values in the data were not considered. We see an overall decreasing trend of NO_x emissions as cars get newer, although there are some spikes in the data as seen on the graph. Model years after 2003 show the emissions remaining steady at lower values than any of those from earlier models. 2003 was decided as the split between the "older" and "newer" cars due to vehicle regulations in 2003, so the rapid change in emissions between 2003 and 2004 is not unexpected. This supports the hypothesis that older cars are emitting more and suggests that it would be beneficial to replace them.

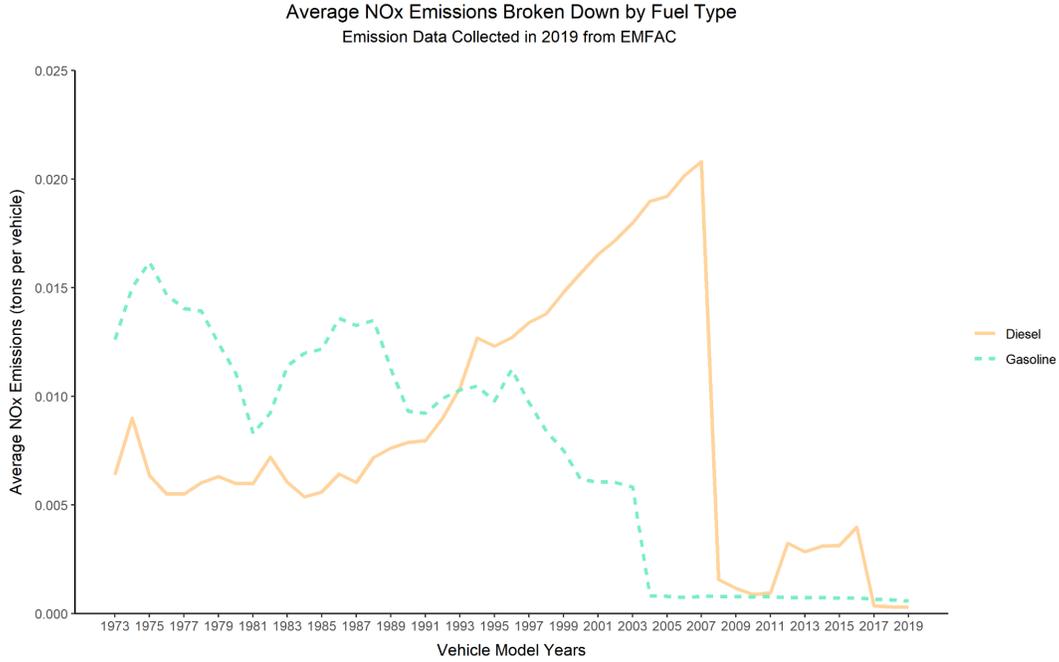


Figure 2.2.2 Graph showing the average per model year of NO_x emissions per mile travelled. There are two lines designating gasoline and diesel vehicles, and this graph includes model years from 2003 to 2019. The data was collected by EMFAC in 2019 and CARB.

Figure 2.2.2 shows the average NO_x emissions per vehicle for each model year. The two lines represent gasoline and diesel vehicles. Gasoline and diesel vehicles experienced sudden drops in NO_x emissions in 2004 and 2008 respectively, prior to which gasoline had an overall decreasing trend and diesel had an overall increasing trend. The decrease in emissions from gasoline vehicles appears to occur in conjunction with the regulations put in place in 2003. The decrease in emissions for diesel appears in vehicles made some time later, but such a large drop in 2008 vehicles might have been caused by different regulations, aimed specifically at diesel vehicles.

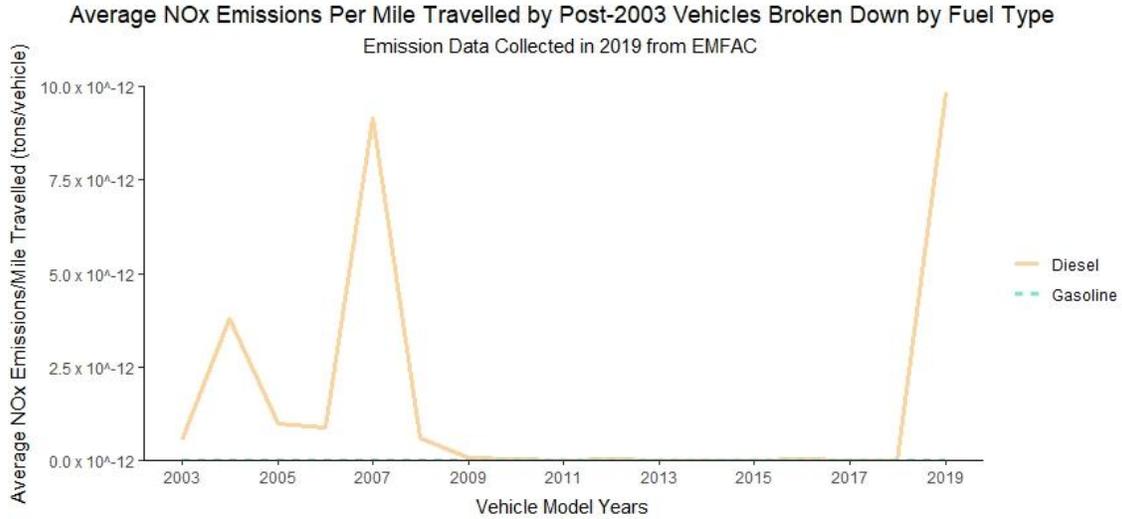


Figure 2.2.3 Graph showing the average per model year of NO_x emissions per mile travelled. There are two lines designating gasoline and diesel vehicles, and this graph includes model years from 2003 to 2019. The data was collected by EMFAC in 2019 and CARB.

Figure 2.2.3 shows the average NO_x emissions per mile travelled by vehicles for each model year after 2002. The two lines represent gasoline and diesel vehicles. While diesel vehicles experience several fluctuations, gasoline vehicles remain near zero. This shows that the diesel vehicles on average emit more NO_x than gasoline vehicles for newer vehicles.

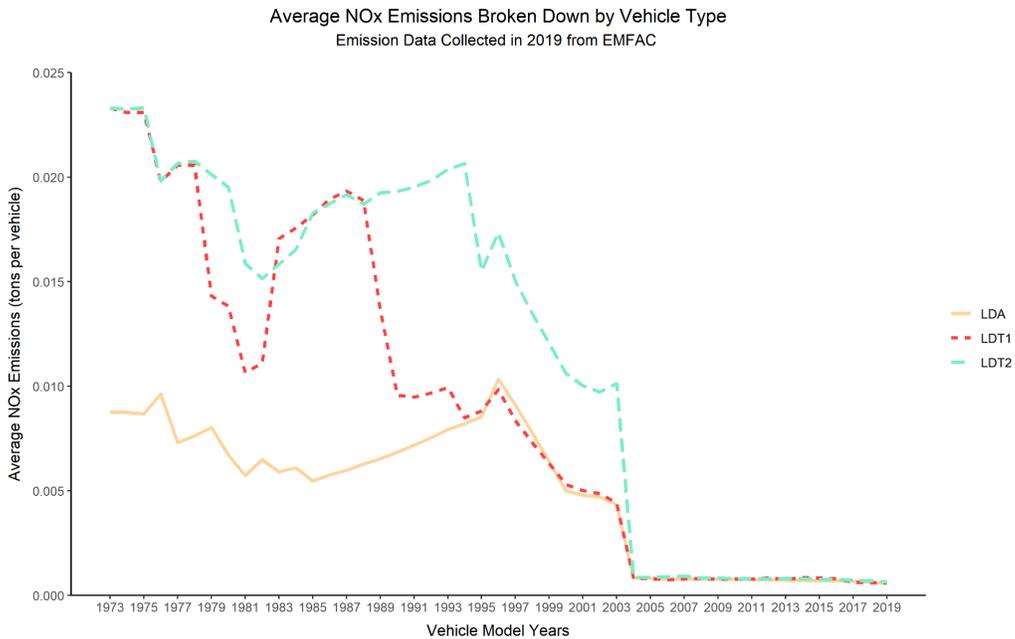


Figure 2.2.4 Graph showing the average amount of NO_x emitted per vehicle, differentiated by vehicle type and model year. The data was collected by EMFAC in 2019 and CARB.

Figure 2.2.4 shows the average NO_x emissions for each model year, categorized by vehicle type. All vehicle types fluctuate at first but experience a sudden drop after 2003 and remain steady relatively close to zero. The graph shows a similar trend to gasoline vehicles in Figure 2.3, as the emissions for all three vehicle types plummet after 2003, which is when regulations were put into place. This result is supportive of our hypothesis that vehicles made in or before 2003 emit more than those made after 2003, as the graph shows that this is the case for this data sample. In addition, it indicates that emissions for all three vehicle types studied were reduced in models after 2003.

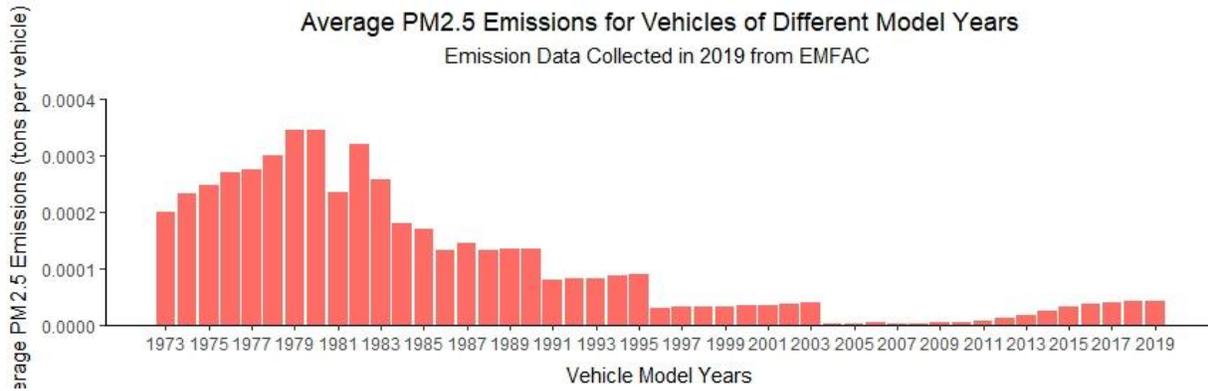


Figure 2.2.5 Graph depicting the mean PM_{2.5} emissions for cars from each model year from 1973-2019. The mean emissions are based on data collected by CARB and emissions data collected by EMFAC for 2019.

Figure 2.2.5 shows the mean emissions for PM_{2.5} from the emissions data for 2019, differentiated by the model year of the cars. While calculating the mean emissions, N/A values were not considered. The values from 1973 until about 1995 are visibly higher than those after 1995, and there is an interesting value in 1981 which is lower compared to the other emissions in model years surrounding 1981. This graph shows that older vehicles emitted more than newer vehicles, particularly vehicles with model years from 1973 to 1990.

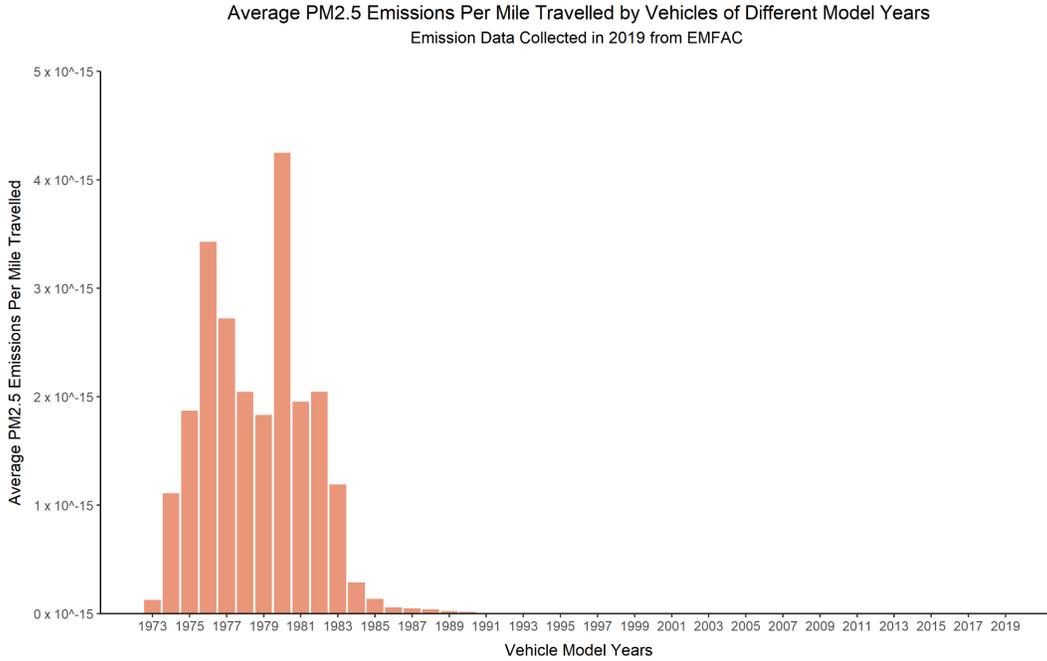


Figure 2.2.6 Graph showing the average PM_{2.5} emissions per mile travelled by vehicles, differentiated by model year. All model years are included in this graph. The data was collected by EMFAC in 2019 and CARB.

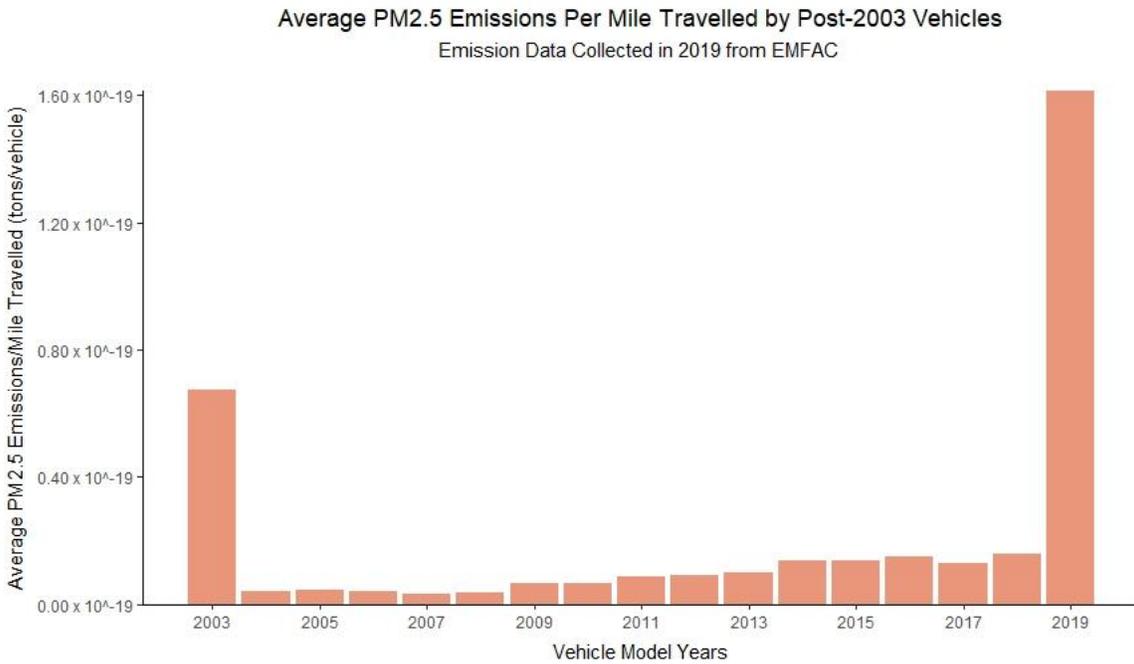


Figure 2.2.7 Graph showing the average PM_{2.5} emissions per mile travelled by vehicles, differentiated by model year. Cars with model years from 2003-2019 are included in this graph. The data was collected by EMFAC in 2019 and CARB.

Figure 2.2.6 shows the average $PM_{2.5}$ emissions per mile travelled by vehicles, for each model year. There appears to be some fluctuation but in about 1986 the emissions appear to decrease to near-zero values. Figure 2.2.7 shows the average $PM_{2.5}$ emissions per mile travelled by vehicles for each model year as well, but the graph focuses on model years after 2002. Besides the increased value in 2003, levels remain about the same until 2009, when they begin to increase to a much higher value in 2019. Despite the increase in 2019, it is still a lower value than those for model years 1973 to 1985, which is likely why it is not visible in Figure 2.2.6. The high values from model years 1973 to 1985 (Figure 2.2.6) could contribute to the higher overall emissions from these model years, while the higher emissions from years between 1985 and 2003 might be due to increased vehicle population or people driving more (these might compensate for the low "Emissions Per Mile Travelled" values seen here). Figure 2.2.7 is interesting because of the sudden and dramatic increase in emissions in cars from 2019. One hypothesis is the use of gasoline direct injection (GDI) engines in vehicles, which reportedly increased after 2008 (Pontecorvo, 2020). Pontecorvo (2020) says that these engines result in emission of PM, which might explain why $PM_{2.5}$ emissions increase after 2008 as shown in Figure 2.2.7. The higher emissions from 1973 to 1985 (Figure 2.2.6) support our idea that replacements of older vehicles might be in order.

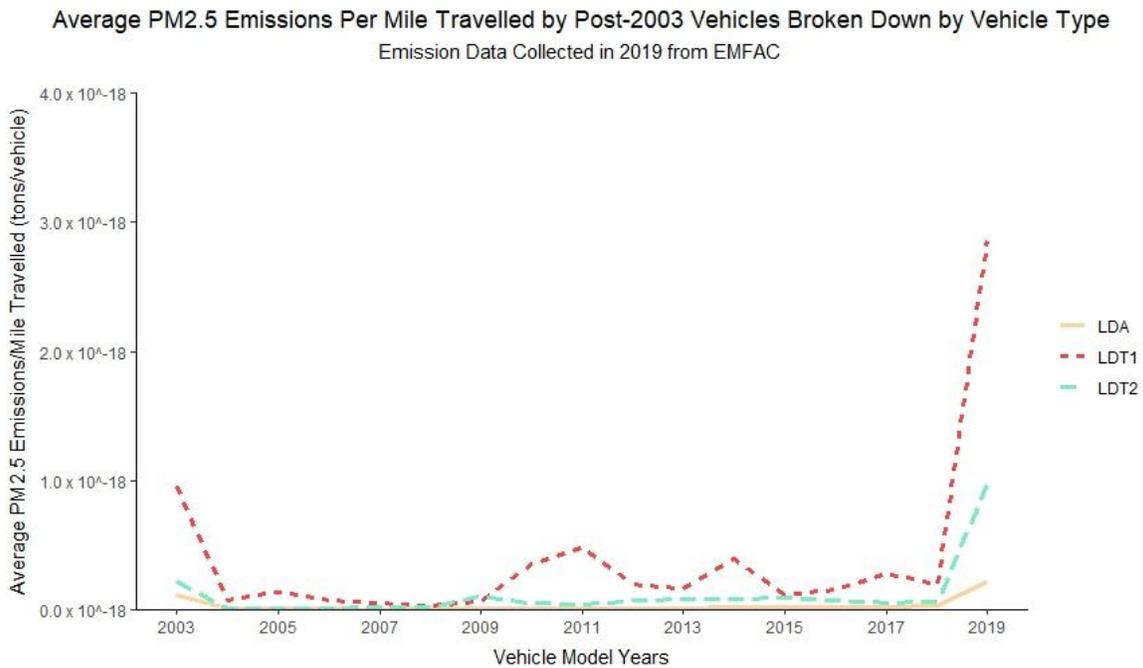


Figure 2.2.8 Graph showing the average $PM_{2.5}$ emissions per mile travelled by vehicles, differentiated by vehicle type and model year. The data was collected by EMFAC in 2019 and CARB.

Figure 2.2.8 is a more detailed analysis of Figure 2.7. It shows the average $PM_{2.5}$ emissions per mile travelled for each model year after 2002, but split into the three vehicle

types. All three categories experience a drop from 2003 to 2004, then LDT1 vehicles experience some fluctuation until 2018, after which all three types increase, LDT1 in particular increasing a large amount. The LDT1 curve is interesting because of the varying values that are not shown as much by either LDA or LDT2 vehicles. Figure 2.2.8 gives more detail about the high emissions from 2019 cars (Figure 2.2.7), and it looks like LDT1 vehicles are mostly responsible. If 2019 vehicles merit further investigation, LDT1 vehicles look to be those most in need of change.

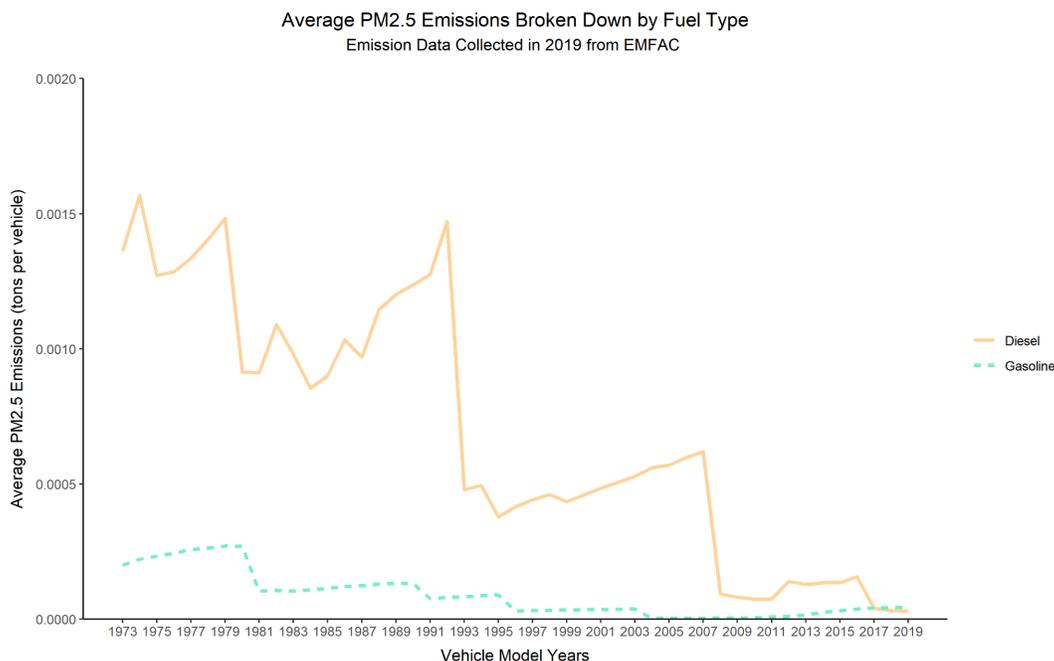


Figure 2.2.9 Graph showing the average PM_{2.5} emissions, differentiated by fuel type and model year. The data was collected by EMFAC in 2019 and CARB.

Figure 2.2.9 shows the average PM_{2.5} emissions per vehicle for each model year, with two lines representing diesel and gasoline. Gasoline vehicles start at relatively low values and undergo several drops while otherwise remaining fairly consistent, while diesel vehicles also experience several drops with fluctuating values in between. By about 2017, diesel and gasoline vehicles have about the same average PM_{2.5} emissions. The sudden drops in PM_{2.5} emissions in gasoline vehicles may be due to changes in regulation and this may apply to diesel vehicles as well, as there are several large drops in the diesel vehicle trend. These drops correspond with different years compared to gasoline vehicles, so different regulations that apply to diesel vehicles specifically may be responsible. The graph suggests that vehicles made even several years after the 2003 split emit less, but diesel vehicles even some years after still have room for improvement. This suggests that there are vehicles other than those made in or before 2003 that should be considered. The general trends do indicate that older vehicles emit more and replacing them would probably be helpful for reducing emissions.

2.3 Air Quality Analysis

Air quality results using emissions at census block group vs. tract

We used the emissions outputs to map air quality at both the census block group or the tract level. In order to determine whether the extent will have an effect on air quality, we ran simulations using both census levels. Figures 2.3.1(a)-(b) show the air quality results of these simulations zoomed into Los Angeles County. Even though there were no large differences between the two when compared, we decided to use emissions at the block group level.

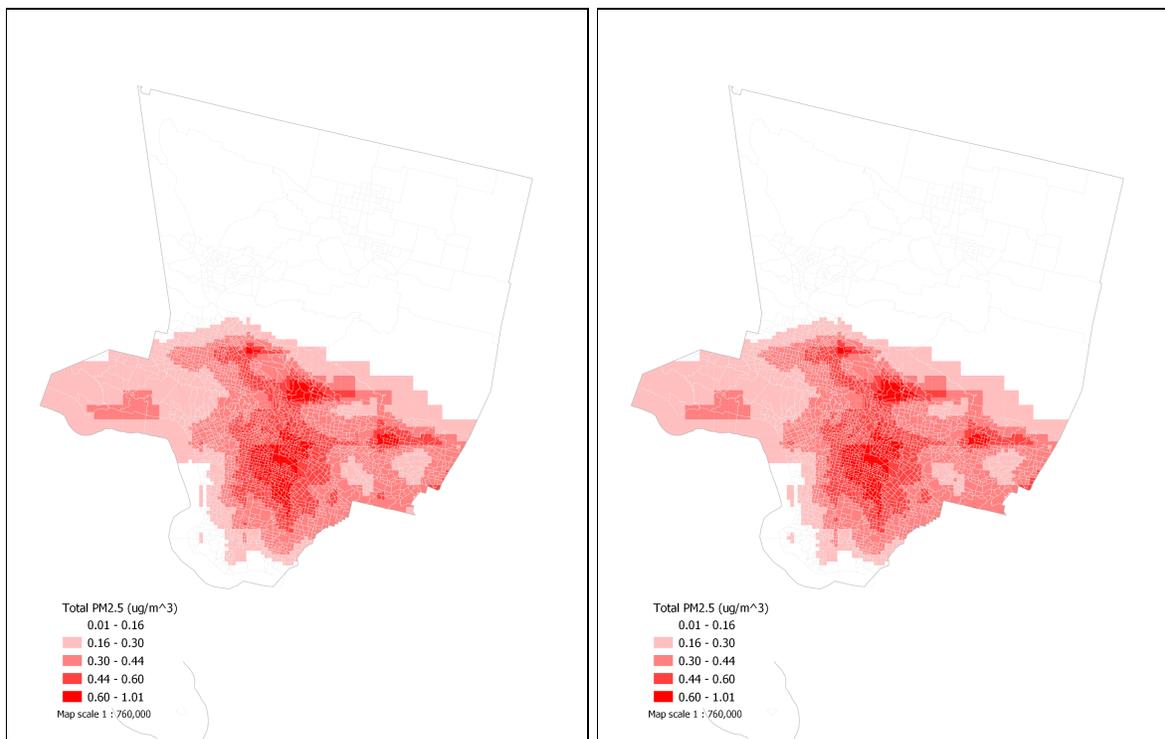


Figure 2.3.1 InMAP output at 1-km extent with emissions at **(a)** the census block group level and **(b)** at the tract level.

Air quality results with and without NH_3 emissions

In addition, we wanted to confirm how big of a role NH_3 *running* emissions played in affecting air quality. This is because all pollutant (NO_x , SO_x , $\text{PM}_{2.5}$, ROG/VOC) emissions retrieved from EMFAC for this research were total emissions released from vehicles, which aggregated emissions released when vehicles were running, idling, and starting. InMAP also required NH_3 emissions to model air quality. However, EMFAC only had NH_3 emissions released from vehicles when they were running, known as NH_3 running emissions. Even though we knew that NH_3 would have an effect on air quality as it listed one of the four main precursors to $\text{PM}_{2.5}$ listed by the Environmental Protection Agency (EPA), we were not sure whether NH_3 *running* emissions would have a noticeable effect since running emissions only make up one of the three parts of

total emissions. As such, two simulations were run, one included NH_3 running emissions and the other did not. The results were shown in figures 2.3.2(a)-(b), respectively. Comparing the two figures, it was clear NH_3 emissions played a significant role in affecting resulting $\text{PM}_{2.5}$ concentrations, even if only running emissions and not the total were available, because the areas with high $\text{PM}_{2.5}$ expanded substantially especially in central Los Angeles County. Therefore, all further analyses used emissions that included NH_3 running emissions.

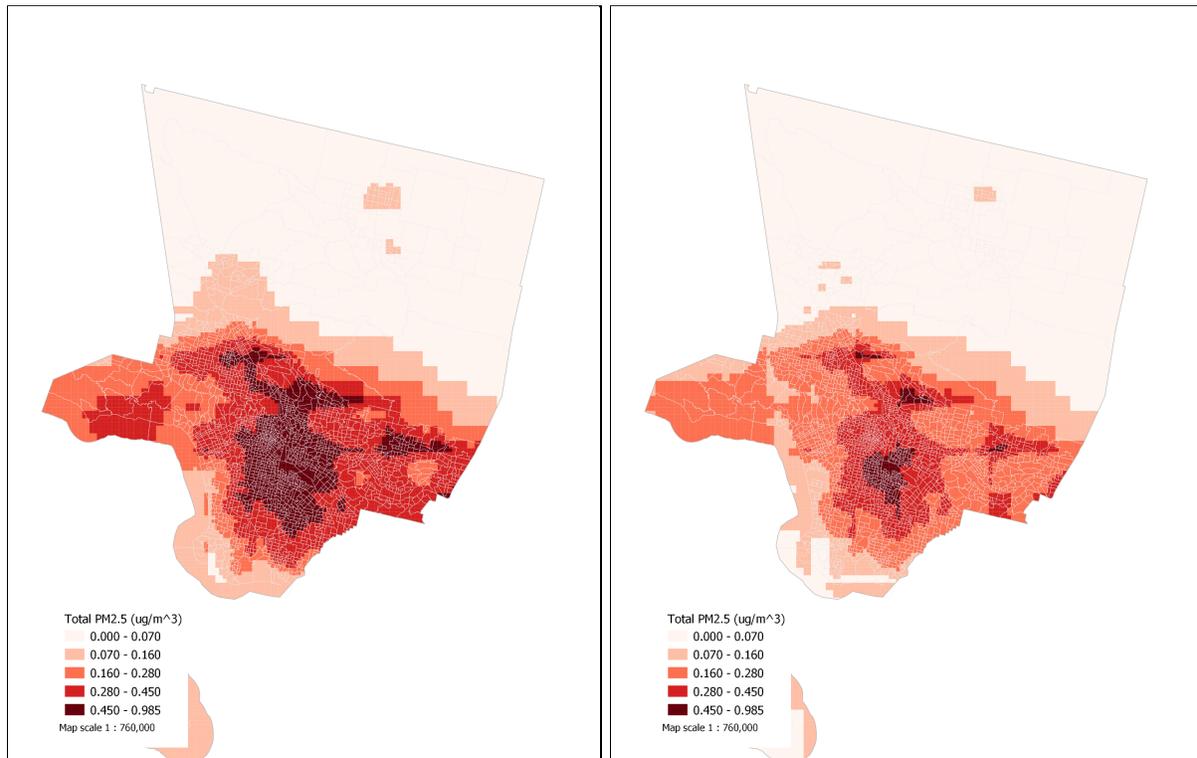


Figure 2.3.2 InMAP output at 1-km extent (a) with NH_3 running emissions and (b) without NH_3 running emissions.

Air quality results from pre-2003 vehicles (1-km vs. 4-km resolution)

InMAP can model air quality down as far as 1-km, which is a very fine resolution for the scope of our analysis. Even though our analysis strived to be as accurate as possible, because of computer memory limitations, the next portion of our research, which was health effects analysis, could only analyze health effects across California as a whole at a maximum of 4-km resolution. Therefore, we also wanted to analyze the difference in air quality at both the 1-km and 4-km resolution levels at the extent of California. Figures 2.3.3(a)-(b) show the results of such analysis, respectively.

Based on figures 2.3.3(a)-(b), InMAP predicted high $\text{PM}_{2.5}$ concentrations in many regions of Southern California, the Bay Area, as well as Central California. However, the areas with high $\text{PM}_{2.5}$ concentration were much larger for the 4-km resolution simulation than for the 1-km, suggesting that 4-km simulation was overestimating. Though we were not certain why

this was the case, one possible explanation was InMAP was able to make a more detailed analysis at finer scale and did not have to overestimate for uncertainties.

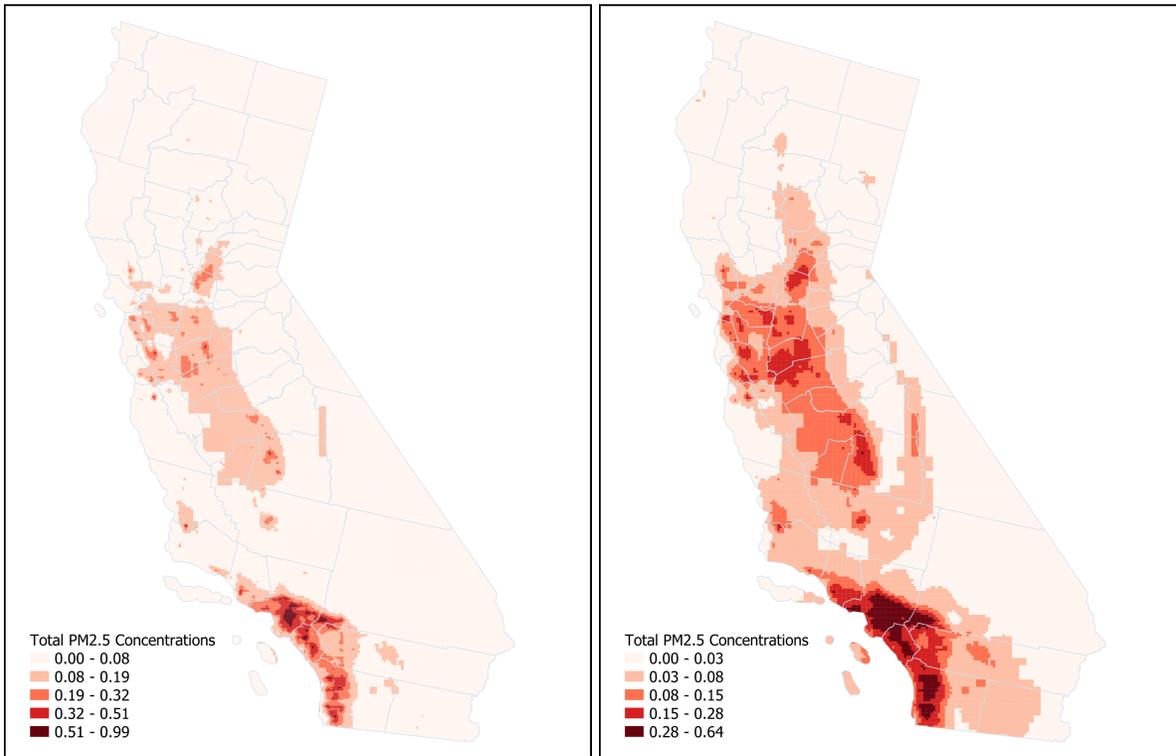


Figure 2.3.3 InMAP output at (a) 1-km resolution clipped to the extent of California and at (b) 4-km resolution clipped to the extent of California.

Air quality results from pre-2003 vehicles across various air districts

Three air quality district regions were selected for finer analysis based on how severe PM_{2.5} concentrations were in the regions. These were the Bay Area Air Quality Management District (BAAQMD), San Joaquin Valley Unified Air Pollution Control District (SJAPCD), and South Coast Air Quality Management District (SCAQMD). Figures 2.3.4-2.3.6 show the air quality in these regions respectively. PM_{2.5} concentrations were particularly high around the bay in BAAQMD, the northern and southeastern part of the SJAPCD, and the western part of SCAQMD. Comparing all three, however, SCAQMD on average had the highest PM_{2.5} concentrations.

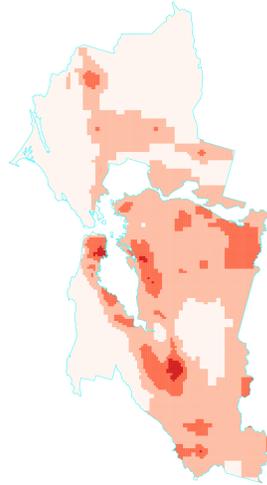


Figure 2.3.4 InMAP output at 1-km clipped to Bay Area Air Quality Management District (BAAQMD)

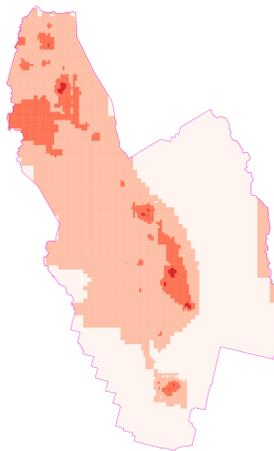


Figure 2.3.5 InMAP output at 1-km clipped to San Joaquin Valley Unified Air Pollution Control District (SJAPCD)

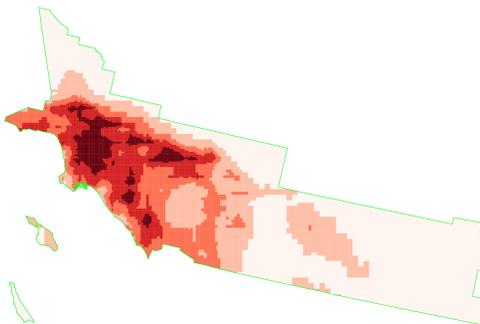


Figure 2.3.6 InMAP output at 1-km clipped to South Coast Air Quality Management District (SCAQMD)

2.4 Health Effects Analysis

All of the health effects modeling results show increased incidence of certain mortality and morbidity endpoints as a direct result of $PM_{2.5}$ from pre-2003 passenger vehicles. Since Los Angeles County and Orange County had the highest mortality rates associated with older vehicles, they were studied at a higher resolution, in addition to the air districts of the San Francisco Bay, San Joaquin Valley, and South Coast (which includes Orange County and parts of South LA). These geographic representations of mortality can be thought of as either the number of lives that could have been saved in 2019 if older vehicles were removed, or the amount of annual death and morbidity caused by not removing pre-2003 vehicles from the roads. Either way, there were direct health consequences of older vehicles. In addition, the analysis shows that for all metrics studied, 95% of the health impacts were located in just 20 out of California's 58 counties (Figure 2.4.8), which were primarily located in the air districts of San Francisco Bay, San Joaquin Valley, and South Coast. Therefore, it is likely that prioritizing these regions would have the greatest positive effect on health, even without a statewide program.

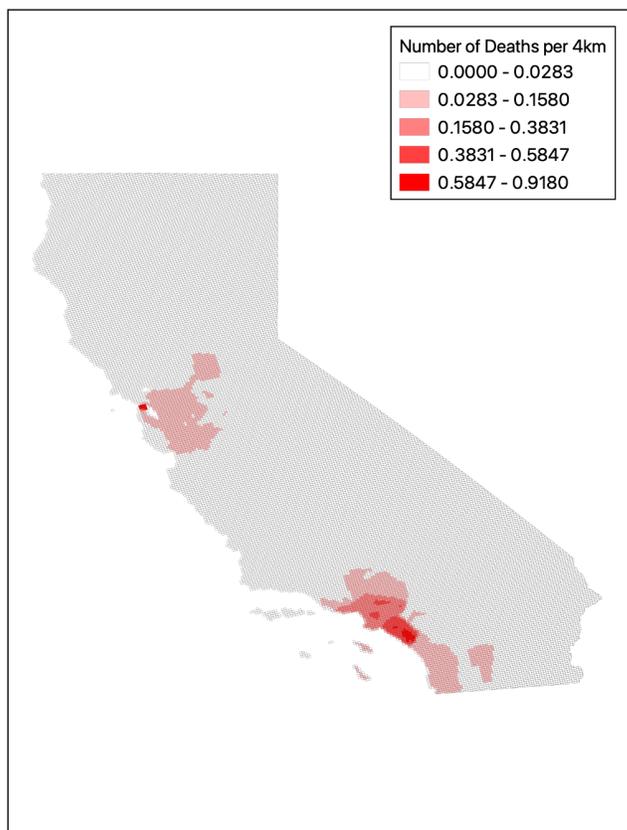


Figure 2.4.1 Number of deaths per 4-km due to the air quality effects of pre-2003 vehicle emissions in 2019, for an aggregated estimated total of 307.23 deaths statewide.

It appears as though the health effects roughly follow county boundaries even though it wasn't aggregated by county for these results. This may be because the population data that BenMAP uses is on the county level, so it is not necessarily designed to zoom into very small areas, and is better equipped for large aggregations over huge regions. In retrospect, that is the opposite of the purpose of this portion of the study, which was intended to zoom into as fine of a resolution and scale as possible. Still, we were able to compute results to the 1km scale within counties to find that there is still variation despite having a uniformly designated population count for the county. Figure 2.4.1 shows the relative distribution of deaths due to older vehicles throughout the state in 2019. Although the unit is in a scale that is difficult to conceptualize (the number of deaths within a 4-km area) it is clear to see that the regions with high population density like the San Francisco Bay and the southern coastal region had the most deaths.

This observation makes sense because the health function equations use the population to compute the estimated number of deaths, and this result has not been refined per-capita. However, from the perspective of saving lives and avoiding negative health consequences, Figure 2.4.1 provides a good overview of where deaths could be avoided by removing older vehicles.

To analyze further, we aggregated the deaths by county, resulting in a more understandable scale (Figure 2.4.2). Figure 2.4.2 highlights Los Angeles County with the highest mortality of 100 deaths, followed closely by Orange County with 66 deaths. Notably, the 4km resolution overestimates deaths in comparison to the 1km resolution, which resulted in an estimated total mortality of 64 deaths for Los Angeles County (Figure 2.4.3). This estimated difference is due to inherent air quality modeling functions at different resolutions that alter the input data used for BenMAP. Interestingly, the Central Valley region has a notable amount of deaths that doesn't show up on the 4-km resolution alone, prompting further analysis in the San Joaquin Valley Air Pollution Control District (Figure 2.4.6). However, the most significant health impacts were located in Los Angeles, Orange, and San Diego Counties.

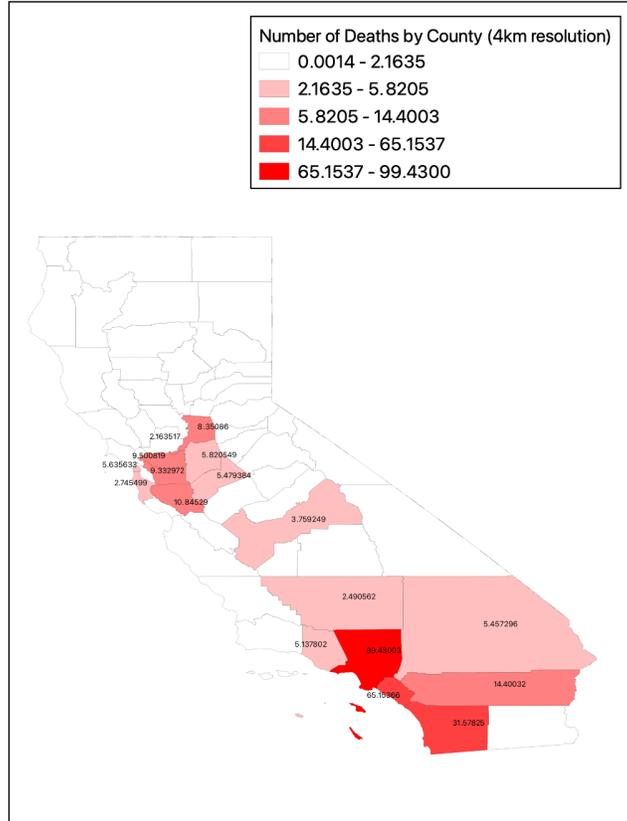


Figure 2.4.2 Number of deaths due to the air quality effects of pre-2003 vehicle emissions at a 4-km resolution, aggregated to total number of deaths by county in 2019, and labeled with the number of deaths in the highest-mortality counties.

Our first 1km resolution test looked at Los Angeles County. As described briefly previously, because BenMAP uses population data by county, when we ran health effects analysis of a single county only, there was only one population number that affected all of the resulting health effect outputs. This essentially acts as a control for the population, because any uneven distribution of health effects visible at a smaller scale than the county level is attributable to PM 2.5 concentrations from older vehicles alone, rather than directly influenced by differences in population density from city to city. Although population is inherently still a factor at the county scale due to the fact that more older vehicles may be located in heavily populated areas, population density is not as much of an influence in the health effects data at the county level as it is at larger scales.

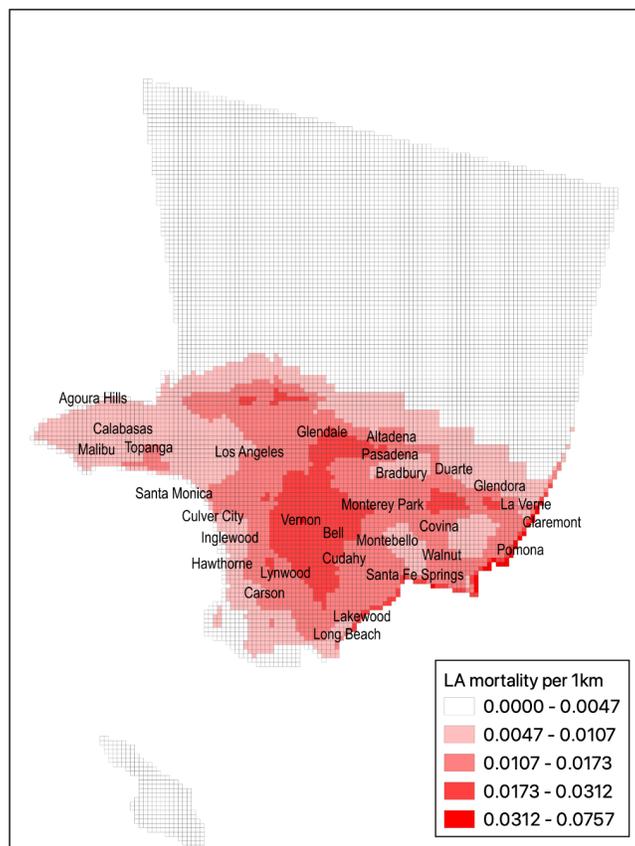


Figure 2.4.3 Number of deaths per kilometer due to air quality effects from pre-2003 vehicle emissions in Los Angeles County in 2019, for a total estimate of 63.10 deaths county-wide.

When we analyzed health effects for Los Angeles County at the 1 kilometer level, South Central was of particular concern (Figure 2.4.3). One consideration is that health effects could potentially be exacerbated in communities that have poor access to healthcare, so to test for this we analyzed the overlap between CalEnviroScreen's list of disadvantaged communities and the areas of highest health concern due to older vehicles. The highest natural break grouping based on deaths from pre-2003 vehicles was compared to CalEnviroScreen's designation of disadvantaged communities. Interestingly, eight cities (Pasadena, Monrovia, Azusa, Montebello, Huntington Park, Downey, Bellflower, and Compton), of the highest-mortality cities, were also listed by CalEnviroScreen.

While the majority of the deaths were located in the southern half of the county, the central region of South Los Angeles County is particularly negatively affected. The mortality estimate for Los Angeles County depends on the resolution and aggregation. Mortality at the 4-km level was estimated to be 105 deaths, which was reduced to an estimate of 100 deaths when aggregated to county within Benmap. The 1-km resolution estimate for Los Angeles County was about 100 deaths when it was clipped to the exact county extent, but only 63.10 deaths when it used an extended rectangular reference grid. Perhaps when it is aggregated

within Benmap there is some method that helps reduce overestimation effects, but it is unclear why there is such a drastic difference due to the reference grid extent. However, the deaths in this county were the most significant in comparison to all other counties in California.

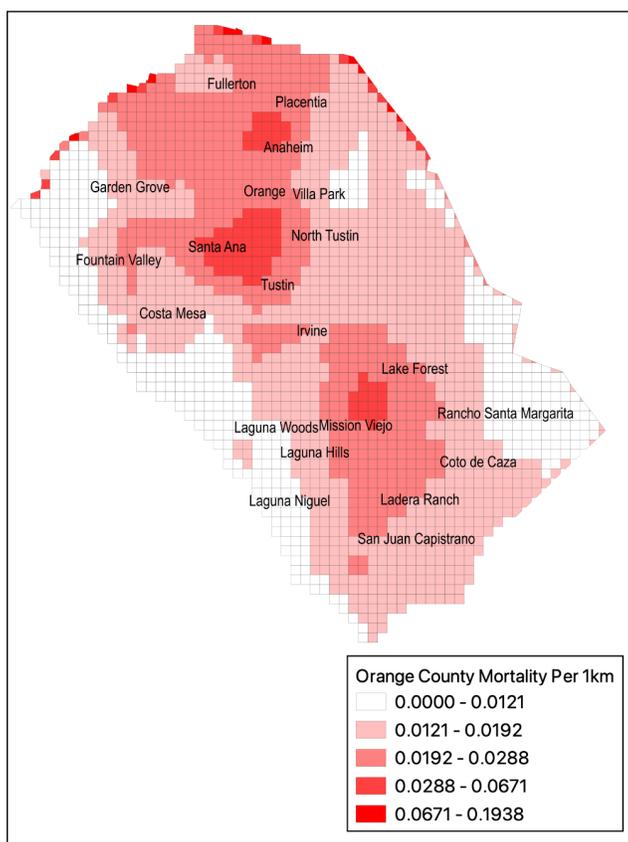


Figure 2.4.4 Number of deaths per kilometer due to air quality effects from pre-2003 vehicle emissions in Orange County in 2019, for a total of 36.88 deaths (Note: edge effects were removed for this estimate, the estimate was 49.98 when cropped to a different extent).

As a fairly large and populous county with a well-known history of air quality concerns, it makes sense why Los Angeles County would have the majority of the air quality health impacts of the state. However, the much smaller Orange County had an even higher per 4-km mortality than LA, despite being less populated (Figure 2.4.1). When we zoomed in, we found that health impacts were prevalent throughout, with the majority occurring in the central regions of the county (Figure 2.4.4). However, when we analyzed this at the 4-km level, more health effects were notable throughout the entire county (Appendix F). The estimated mortality at a 1-km level for Orange County when using the extended analysis grid was 50 deaths, but when we cropped it to reduce the overexaggerated edge effects, the estimate dropped to 37 deaths.

Since the distribution of health effects closely aligns with where the older vehicles were (shown in Figure 2.1.3), these patterns of high mortality in Santa Ana and Mission Viejo were likely due to high amounts of older vehicles. Notably, of the top 10 cities within Orange County

with the highest mortality, only Santa Ana showed up on CalEnviroScreen's list of disadvantaged communities, showing that there is an opportunity to alleviate a disproportionate health burden if policies were to focus on Santa Ana.

The results presented so far give evidence to support that the replacement of older light-duty vehicles with ZEVs in the South Coast Air Quality Management District would have the highest health benefits, because it includes all of Orange County and South LA, the two counties with the highest health impacts. So, next we looked at the South Coast at a 1-km resolution to find that within this area, mortality was highest in the Southern half of Los Angeles County and the entirety of Orange County, with relatively minuscule mortality in the rest of the air district (Figure 2.4.5).

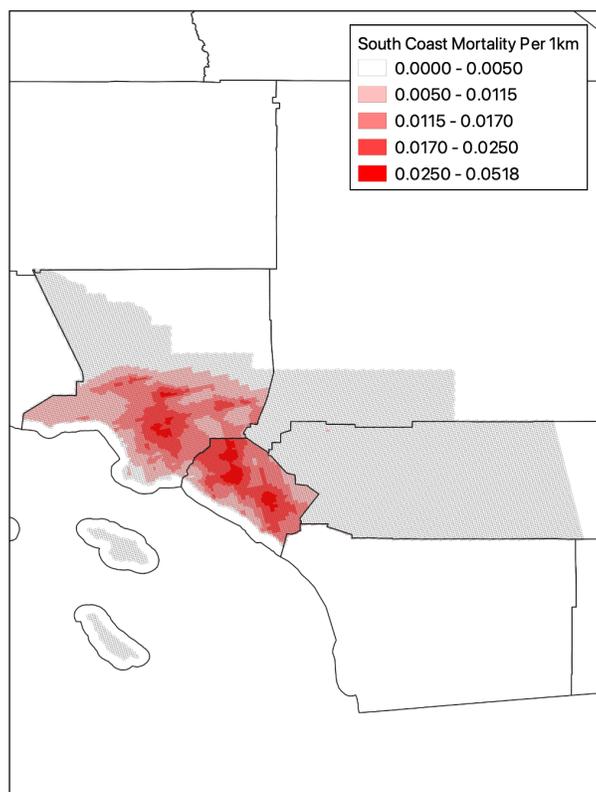


Figure 2.4.5 Number of deaths per kilometer due to air quality effects from pre-2003 vehicle emissions in the South Coast Air Quality Management District in 2019, for a total of about 30 deaths.

When the total mortality for the South Coast air district was computed at a 1-km level, an estimated 105 deaths occurred due to older vehicle emissions in 2019. While this is less than the aggregate of the county-wide health impact estimate of Los Angeles County and Orange County combined (166), this is likely due to the fact that the 4-km resolution overestimates health impacts in comparison to the 1km resolution. The combination of Los Angeles County

and Orange County at the 1km clipped level was 100 deaths, so these two counties represent nearly the entirety of South Coast air quality-related health impacts due to older vehicles.

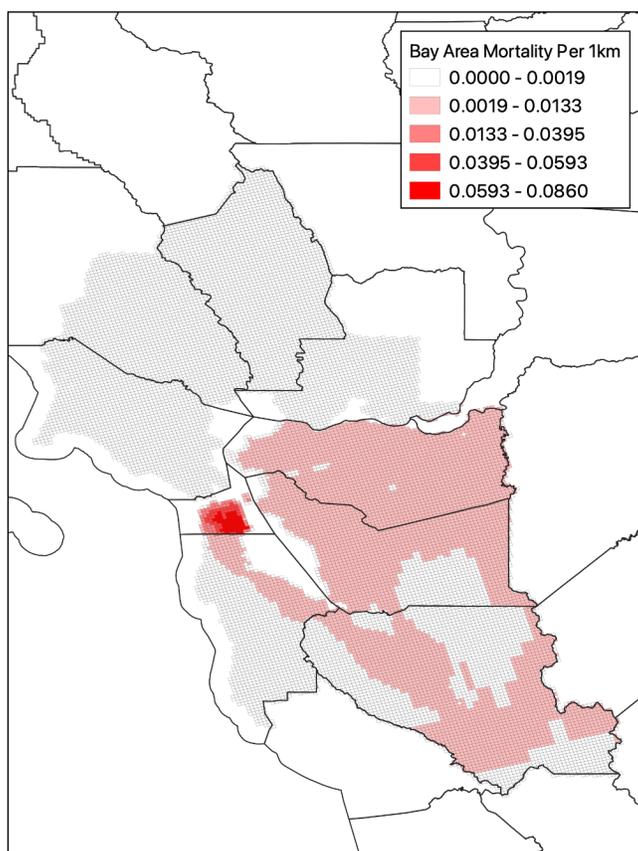


Figure 2.4.6 Number of deaths per kilometer due to air quality effects from pre-2003 vehicle emissions in the Bay Area Air Quality Management District (AQMD) in 2019, with an estimated total of 30 deaths.

To gain a more refined understanding of the health effects of air pollution from older vehicles in other parts of the state, we zoomed into the San Francisco Bay air district to find that the majority of the health effects were concentrated in San Francisco County (Figure 2.4.6). The total all-cause mortality due to older vehicles came out to 30 deaths for 2019. Interestingly, while the air quality results show significant PM 2.5 concentrations in San Francisco, Oakland, and San Jose (Figure 2.3.3), only San Francisco stands out in the health effects map at the Bay Area level. For the same reason that health effects seem to align on county boundaries, the population data at a county level could also contribute to underrepresentation of areas smaller than the county level. While this is true for the entire state of California, it is especially noticeable in this representation of the Bay Area (Figure 2.3.3). Since San Francisco is a county but Oakland and San Jose are cities within counties, San Francisco is highlighted more on the health effects map than either of the two, and the other two cities' high population densities

were diluted throughout their respective counties, and therefore do not significantly add mortality to the health effects data at this level.

Finally, we analyzed the San Joaquin area because when the resulting statewide health effects data was aggregated by county, Fresno County became visible with high health effects (Figure 2.4.2). Primarily affecting the counties of Merced, Fresno, Stanislaus, and San Joaquin, it is clear to see that the health impacts in the Central Valley were far from negligible (Figure 2.4.7). In fact, there were an estimated 16 deaths from this region that also did not visually appear on the county map (Figure 2.4.2). While these health impacts may not be nearly as concentrated as they were in the other areas of study such as Los Angeles, Orange, and San Francisco, Central Valley should still be considered as a major contributor to overall mortality due to older vehicles across the state.

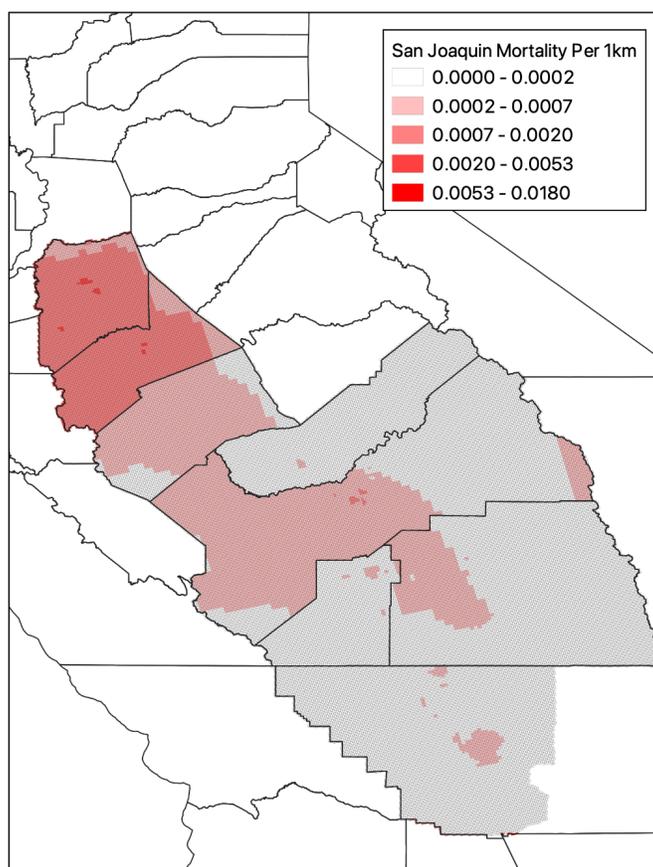


Figure 2.4.7 Number of deaths per kilometer due to air quality effects from pre-2003 vehicle emissions in the San Joaquin Valley Air Pollution Control District in 2019, with an estimated total of 16 deaths.

So far, we discussed health effects in terms of total mortality from all-causes as a result of pre-2003 passenger vehicles because it is most representative of the negative health impact, but we also obtained similar mortality results related specifically to lung cancer and Ischemic heart disease, as well as other endpoints of hospitalizations from respiratory causes, and the

incidence of childhood asthma. The majority of these results visually resemble the maps of mortality all-cause, and for reference all of these health effects results were included in Appendices G-M.

To bring it all together, Figure 2.4.8 was compiled to compare these numerical results in the top 20 counties with the greatest mortality due to older vehicle emissions, as compared to the state (Figure 4.8). The percentage of total row indicates that about 95% of the health impacts from the various endpoints were located in these 20 counties, suggesting a highly uneven distribution of health impacts throughout the state of California. Additionally, Los Angeles and Orange Counties combined account for about a third of all estimated mortality, suggesting that more resources should be put into helping the South Coast Air Quality Management District resolve the air quality issues that were leading to these negative health impacts.

Top 20 Counties with Highest Health Impacts due to Older Vehicles

County	Mortality (All-Cause)	Mortality (Ischemic Heart Disease)	Mortality (Lung Cancer)	Hospital Admissions (All-Respiratory)	Asthma Incidence (Ages 0-17)	Associated Cost of Total Mortality (Millions USD)
Los Angeles	99.43	74.83	10.76	11.28	324.28	933.55
Orange	65.15	41.44	7.74	5.82	198.57	611.73
San Diego	31.58	18.01	3.91	2.44	97.60	296.49
Riverside	14.40	9.85	1.86	1.00	47.10	135.20
Santa Clara	10.85	5.61	1.29	1.18	40.78	101.83
Contra Costa	9.50	4.22	1.26	0.36	26.53	89.20
Alameda	9.33	4.34	1.14	1.35	29.70	87.63
Sacramento	8.35	4.74	1.10	0.70	24.53	78.41
San Joaquin	5.82	3.12	0.74	0.48	19.47	54.65
San Francisco	5.64	2.59	0.71	0.45	11.85	52.91
Stanislaus	5.48	3.87	0.67	0.49	17.79	51.45
San Bernadino	5.46	3.17	0.64	0.67	19.66	51.24
Ventura	5.14	2.85	0.55	0.27	15.92	48.24
Fresno	3.76	2.26	0.41	0.10	14.44	35.30
San Mateo	2.75	1.37	0.32	0.05	7.94	25.78
Kern	2.49	1.60	0.30	0.26	10.24	23.38
Solano	2.16	0.91	0.33	0.02	5.78	20.31
Merced	2.00	1.24	0.25	0.10	8.29	18.80
Santa Cruz	1.60	0.79	0.17	0.12	4.52	15.00
Tulare	1.50	0.95	0.17	0.13	6.88	14.13
Top 20 Total	292.39	187.78	34.33	27.27	931.87	2745.22
Statewide Total	307.23	195.86	36.22	28.67	970.81	2884.57
Percentage of Total	95.17	95.88	94.78	95.10	95.99	95.17

Figure 2.4.8 The top 20 counties with the highest health impact estimates due to pre-2003 vehicles, in comparison to the statewide total health impacts at a 4-km level.

2.5 Environmental Justice Analysis

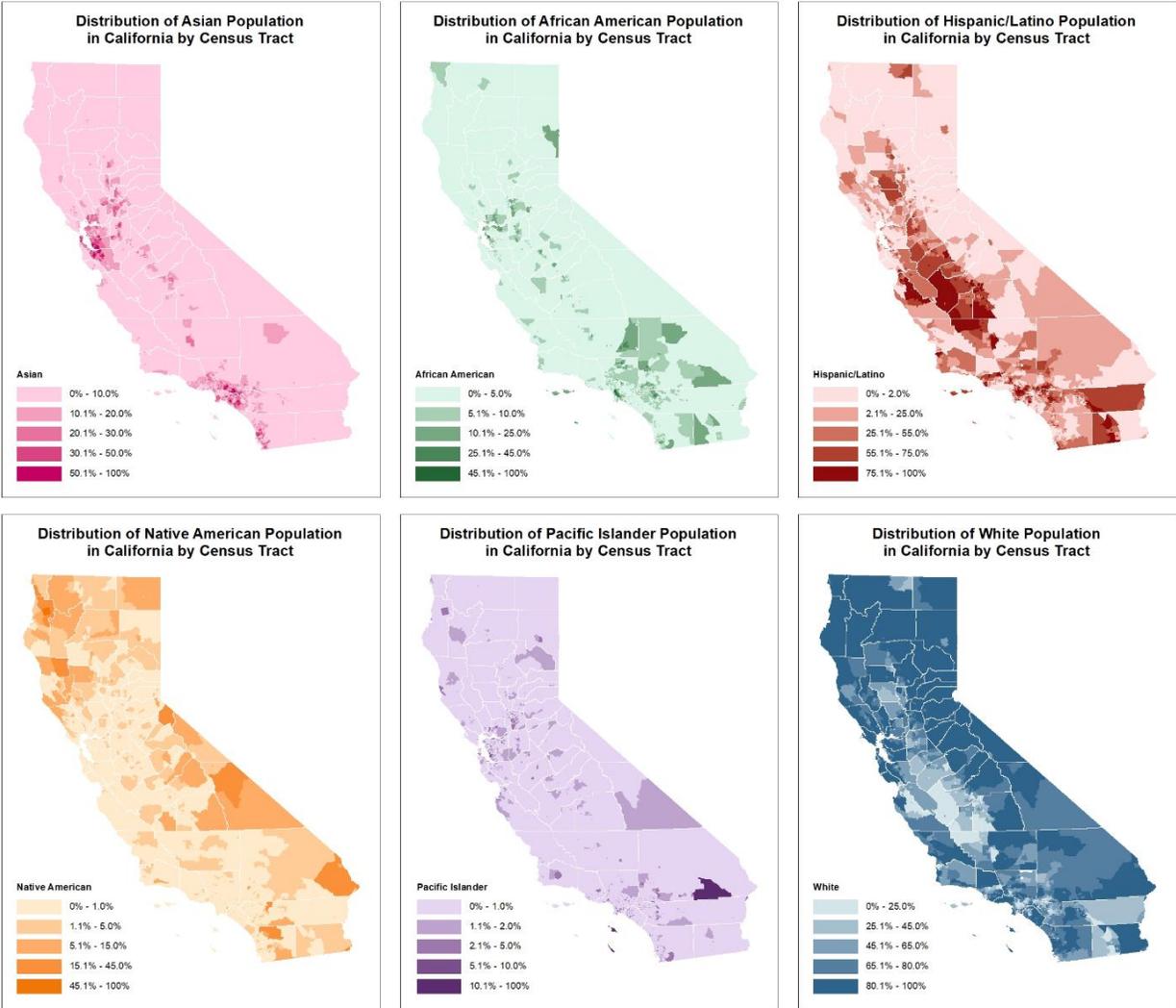


Figure 2.5.1 Racial Distributions of Asian, African American, Hispanic/Latino, Native American, Pacific Islander, and White populations as a percentage of the census tract population

Figure 2.5.1 shows the racial distribution of Asian, African American, Hispanic/Latino, Native American, Pacific Islander, and White populations. The ranges for the maps vary, as each one was broken up by natural breaks (jenks).

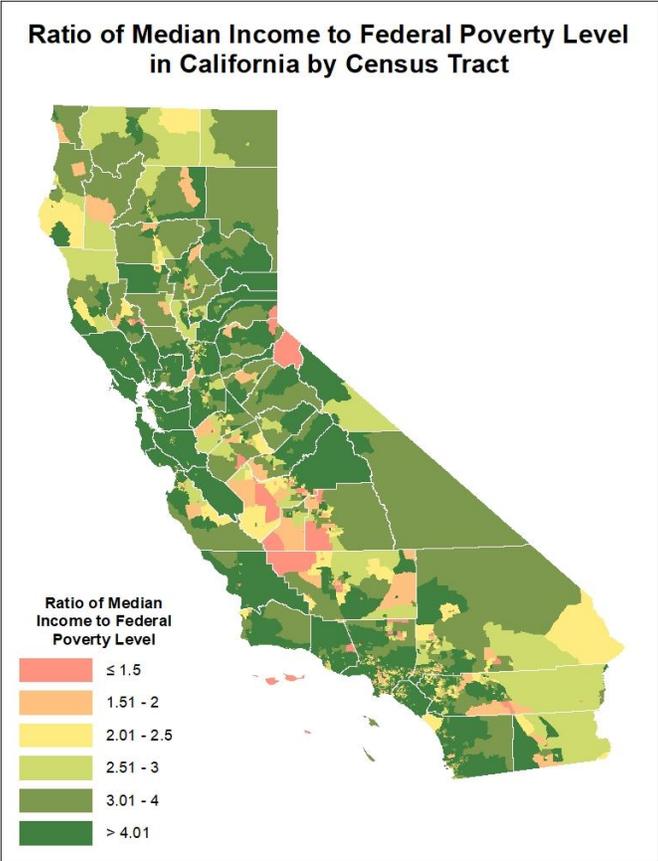


Figure 2.5.2 *Distribution of ratio of median income to federal poverty level in California by census tract.*

Figure 2.5.2 shows the ratio of median income to federal poverty level broken up into ranges used by Covered California and MediCal: ≤ 1.5, 1.5 - 2, 2 - 2.5, 2.5 - 3, 3 - 4, and > 4.

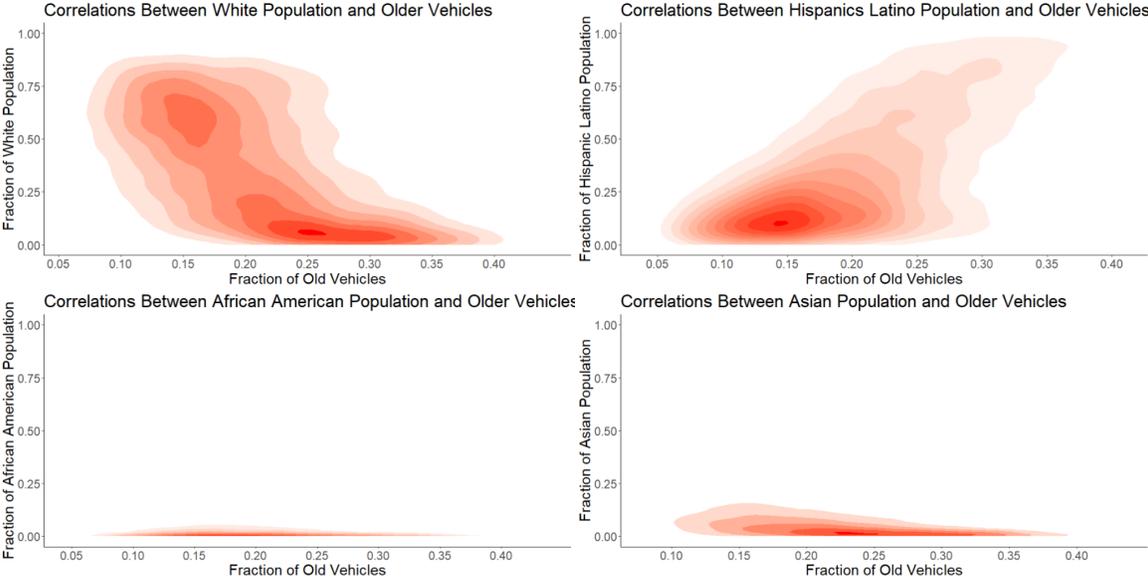


Figure 2.5.3 Correlation between the fraction of older vehicles per household and White, Hispanic Latino, African American, and Asian population makeup at the block group level, under the condition of one vehicle per household

Figure 2.5.3 shows the correlation between different races and older vehicles. The fraction of older vehicles is on the x-axis and the racial makeup of the population is on the y-axis. Looking at the top right Hispanic Latino correlation plot, we can see that there is the strongest positive correlation compared to other races. The denser area is concentrated in the bottom left of the plot and has around a 0.15 fraction of old vehicles and a fraction of 0 ~ 0.25 for the Hispanic population.

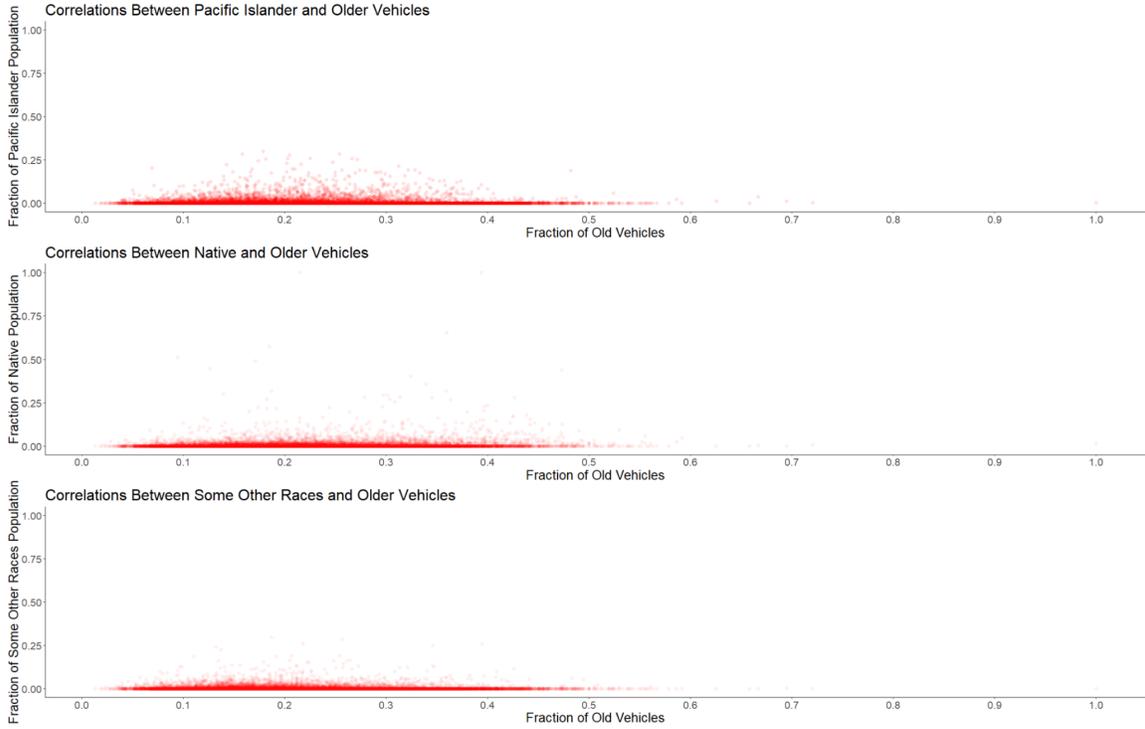


Figure 2.5.4 Correlation plots between older vehicles and different races with a smaller population size at the block group level, under the condition of one vehicle per household.

During the analysis, we found there were three other race types that were not able to be graphed in density plots (Figure 2.5.4). These correlation plots are similar to the ones in Figure 2.5.2, but the fraction of the race population had over 18,000 zero values out of the total 22,868, indicating that the process of plotting the density plots had failed. Since the size of the race population was too small, we neglected it for our analysis.

For the following spatial analysis conducted on the South Coast, Bay Area, and San Joaquin Valley air districts, spatial mismatch of areas with heavy PM_{2.5} concentrations and mortality could be due to population density or other factors such as base concentration and all-cause mortality (BenMAP functions). Another consideration is that mortality could be exacerbated in communities with poor healthcare access.

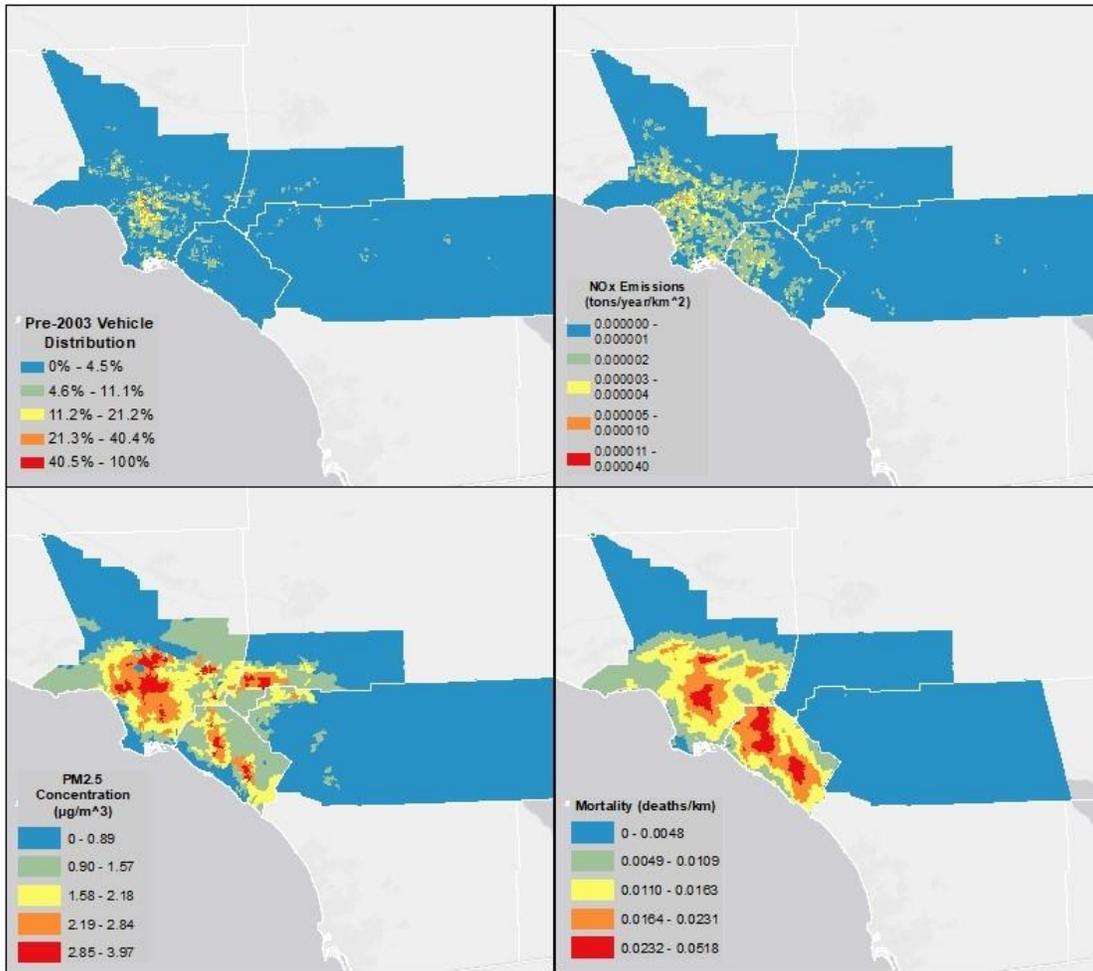


Figure 2.5.5 South Coast Air District maps of pre-2003 vehicle distribution, NO_x emissions, PM_{2.5} concentrations, and mortality.

Pre-2003 vehicles are mainly located in South Central Los Angeles County, and are scattered throughout the eastern areas of the South Coast Air District (Figure 2.5.5). NO_x emissions are heavier in Southern Los Angeles County and Northwestern Orange County. PM_{2.5} concentrations are concentrated in the same areas, as well as Southwest San Bernardino County. Mortality reflects air quality, as there are higher deaths per kilometer in South Central Los Angeles County and Orange County.

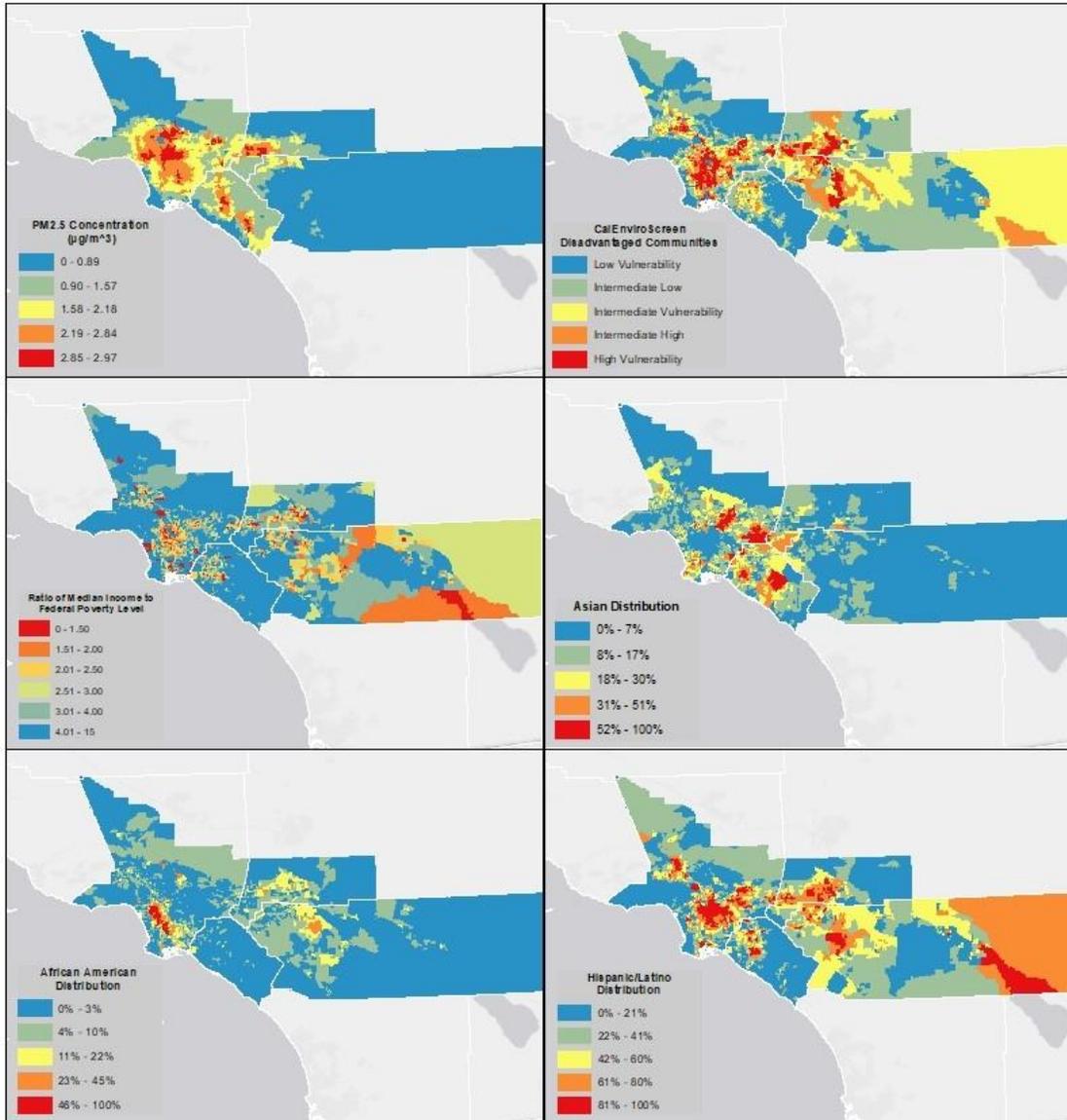


Figure 2.5.6 South Coast Air District maps of PM_{2.5} concentrations, CalEnviroScreen disadvantaged communities, ratio of median income to federal poverty level, and distributions of Asian, African American, and Hispanic/Latino populations.

Represented in Figure 2.5.6, CalEnviroScreen identifies disadvantaged communities in the South Coast Air District to be located in South Central Los Angeles County, Northwestern Orange County, Southwestern San Bernardino County, and Northwestern Riverside County. Similarly these areas are also where there is a lower ratio of median income to federal poverty level. Where the affected regions of Los Angeles County are is also where there is a greater distribution of the Hispanic/Latino population as well as the African American population. In the affected regions of Orange County, we can find a higher distribution of the Hispanic/Latino and

Asian populations. As for San Bernardino County and Riverside County, affected areas are mostly made up of the Hispanic/Latino population.

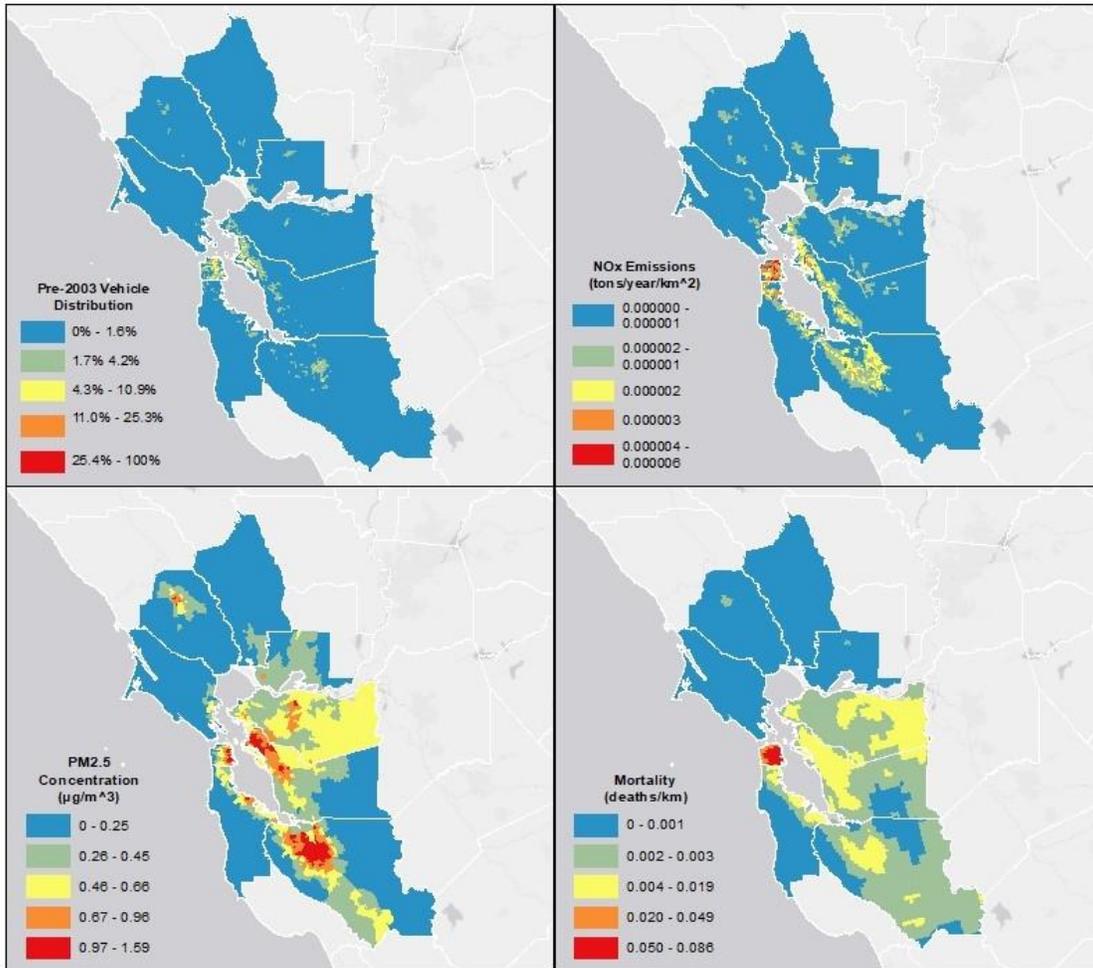


Figure 2.5.7 Bay Area Air District maps of pre-2003 vehicle distribution, NO_x emissions, PM_{2.5} concentration, and mortality.

In the Bay Area Air District (Figure 2.5.7), the distribution of pre-2003 vehicles is mainly concentrated along the coasts of the bay. Furthermore, most NO_x emissions due to pre-2003 vehicles seem to emit from the coast of the bay, which translates to higher PM_{2.5} concentration and poor air quality as well. Additionally, mortality is higher along the coast of the bay, with the highest deaths per 1 km concentrated in the San Francisco Area, where NO_x emissions and PM_{2.5} concentrations are higher as well.

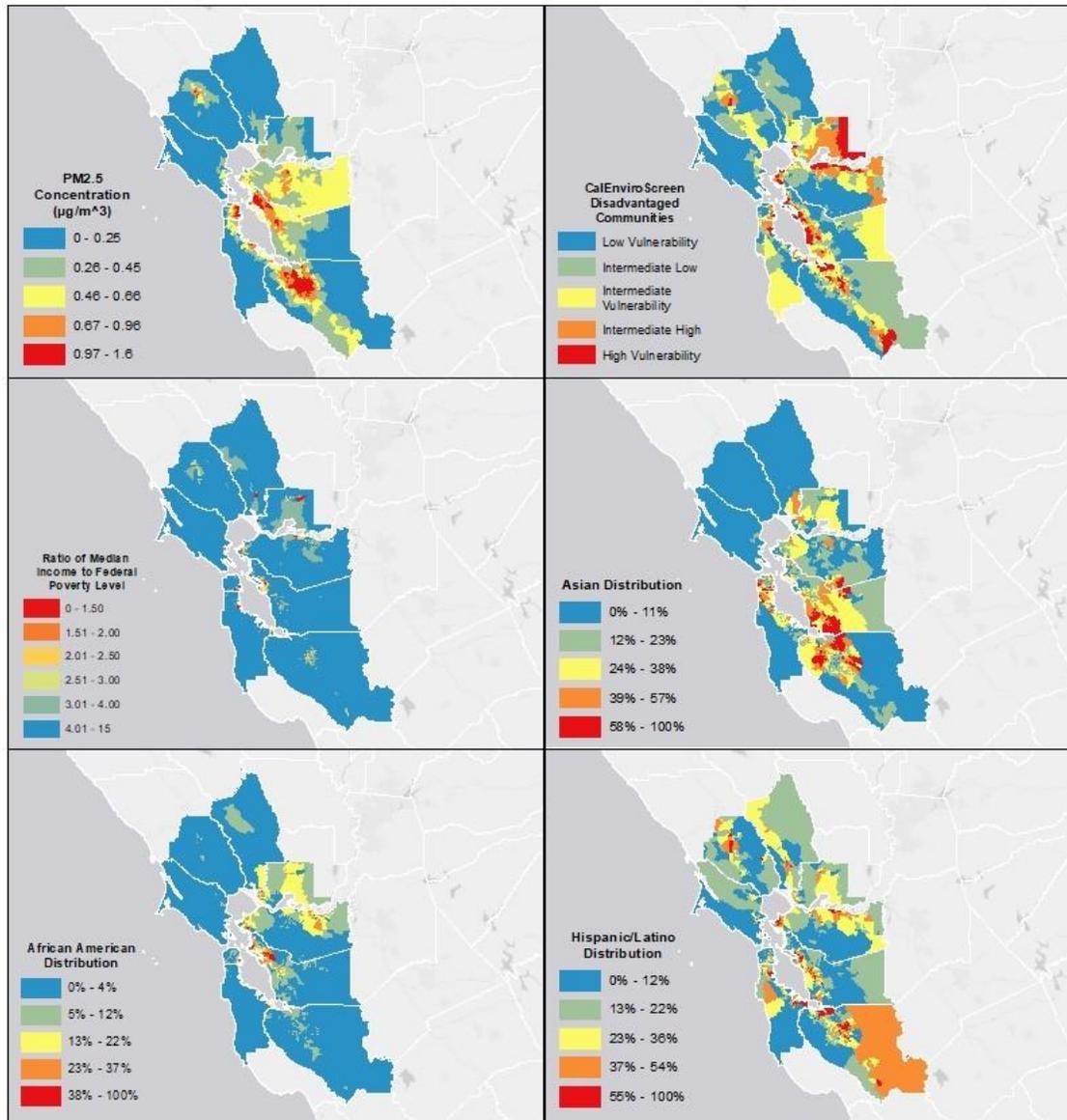


Figure 2.5.8 Bay Area Air District maps of $\text{PM}_{2.5}$ concentration, CalEnviroScreen disadvantaged communities, ratio of median income to federal poverty level, and distributions of Asian, African American, and Hispanic/Latino populations.

As seen in Figure 2.5.8, CalEnviroScreen defines disadvantaged communities to lay along the coast of the bay, as well as certain East and South Central communities in the region. In line, is the ratio of median income to federal poverty level, where lower ranges are also along the coast. The Asian and Hispanic/Latino populations are heavily distributed along the coast of the bay and South, while African American populations mainly reside on the East coast of the bay and further inland. In looking at the maps of Figure 2.5.3, we can deduce that where the concentrations of $\text{PM}_{2.5}$ are higher, is also where CalEnviroScreen defines disadvantaged

communities, as well as where there is a higher distribution of Asian and Hispanic/Latino populations.

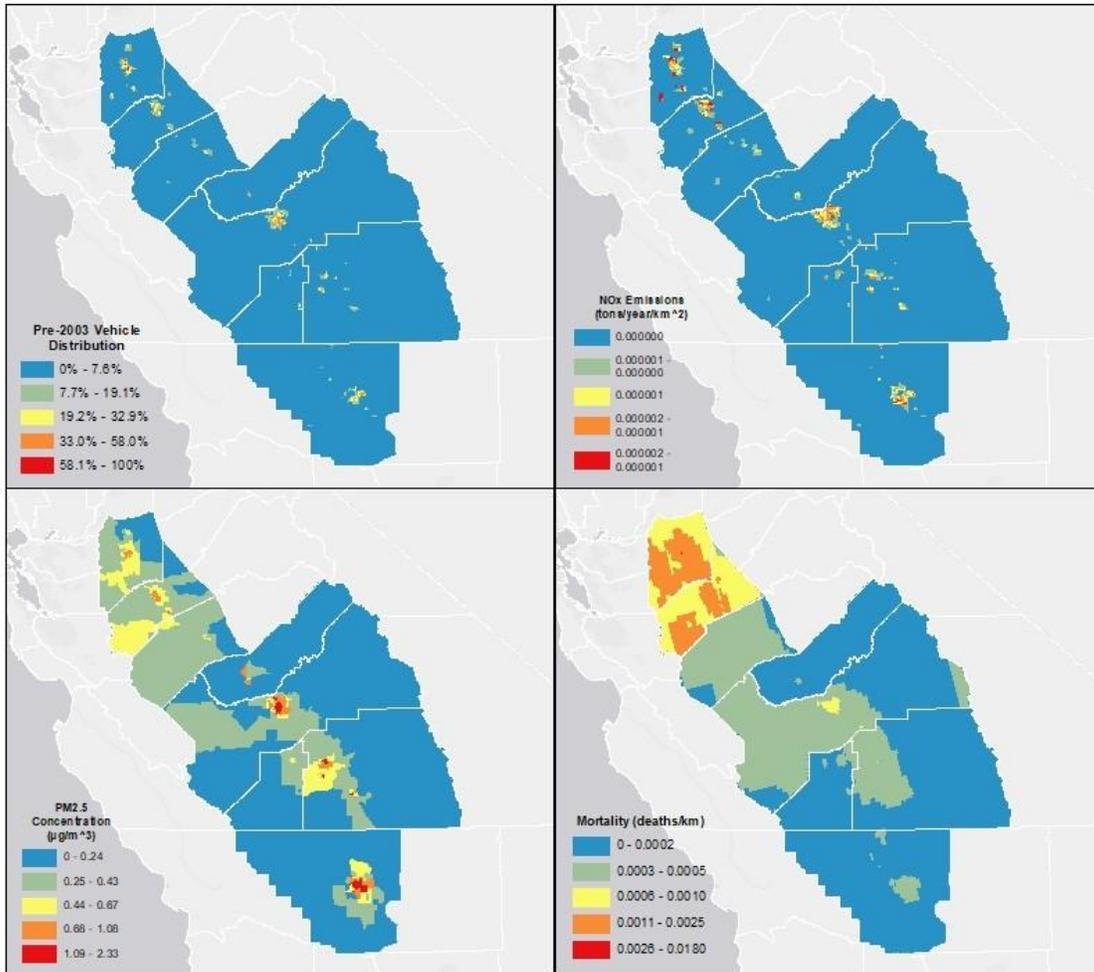


Figure 2.5.9 San Joaquin Valley Air District maps of pre-2003 vehicle distribution, NO_x emissions, PM_{2.5} concentrations, and mortality.

As for the San Joaquin Valley Air District (Figure 2.5.9), pre-2003 vehicle distribution and NO_x emissions are aligned and have a high concentration throughout the middle of the air district. PM_{2.5} concentrations are high along the same areas of older vehicle distributions but are more dispersed in the North and Central regions of the air district. Similarly, mortality is also concentrated in the Central area of the air district, but more so in the North where Stockton and Modesto are.

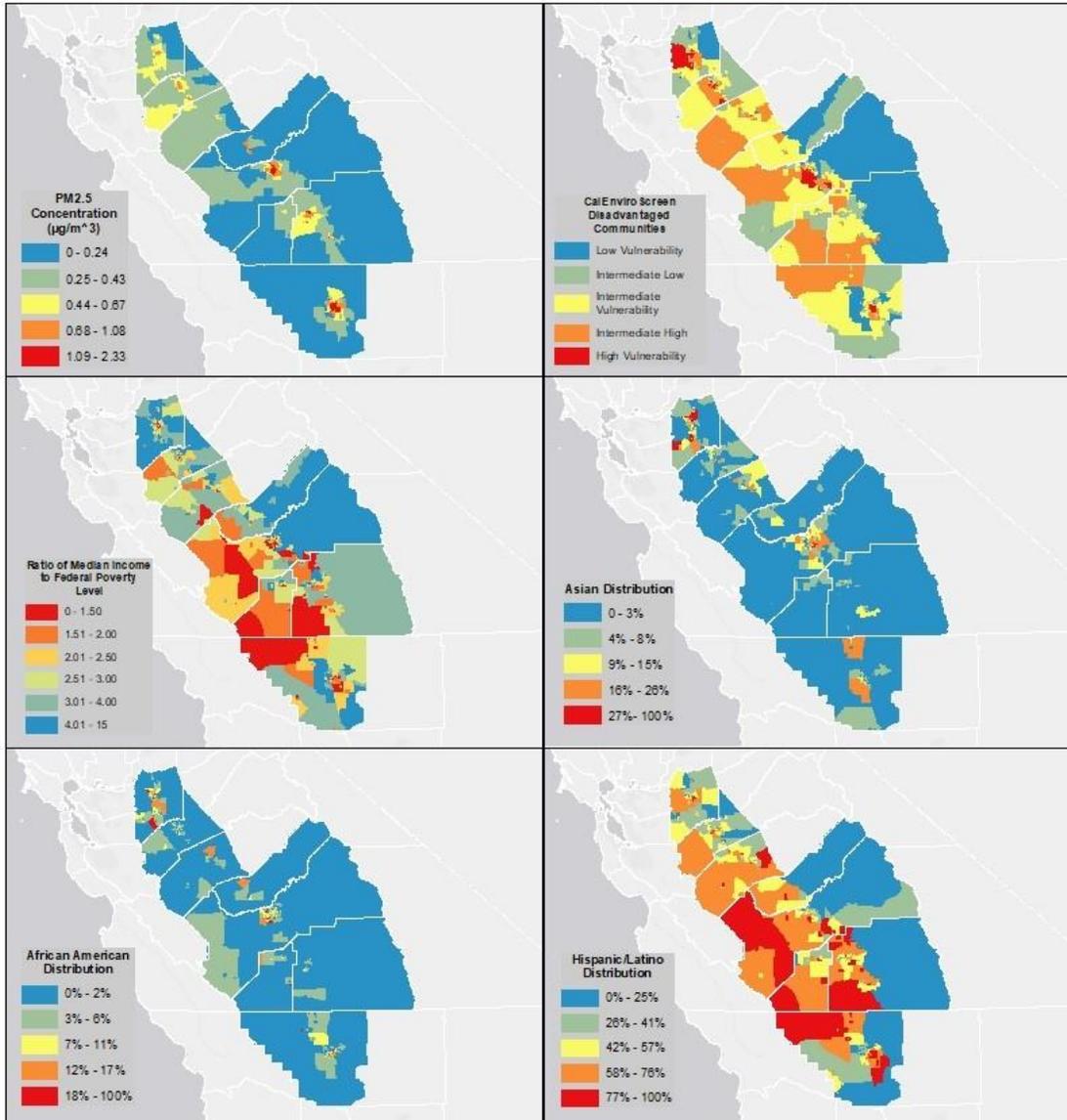


Figure 2.5.10 San Joaquin Valley Air District maps of $\text{PM}_{2.5}$ concentrations, CalEnviroScreen disadvantaged communities, ratio of median income to federal poverty level, and distributions of Asian, African American, and Hispanic/Latino populations.

Figure 2.5.10 demonstrates the CalEnviroScreen defined disadvantaged communities in this area to be along the West half of the San Joaquin Valley Air District. It can be seen that the most disadvantaged communities suffer from poorer air quality, fall into a lower range of ratio of median income to federal poverty level, and consist of a higher distribution of the Hispanic/Latino population. However, it is Central Fresno that is consistent in its vulnerability to poor air quality, classification as CalEnviroScreen's disadvantaged community, lower ratio of median income to federal poverty level, and high distribution of Asian, African American, and Hispanic/Latino populations.

The following analysis was done by comparing each category to the state and air district totals so the values are in percentages. The figures with raw numbers can be found in the appendix (Appendices N-Q).

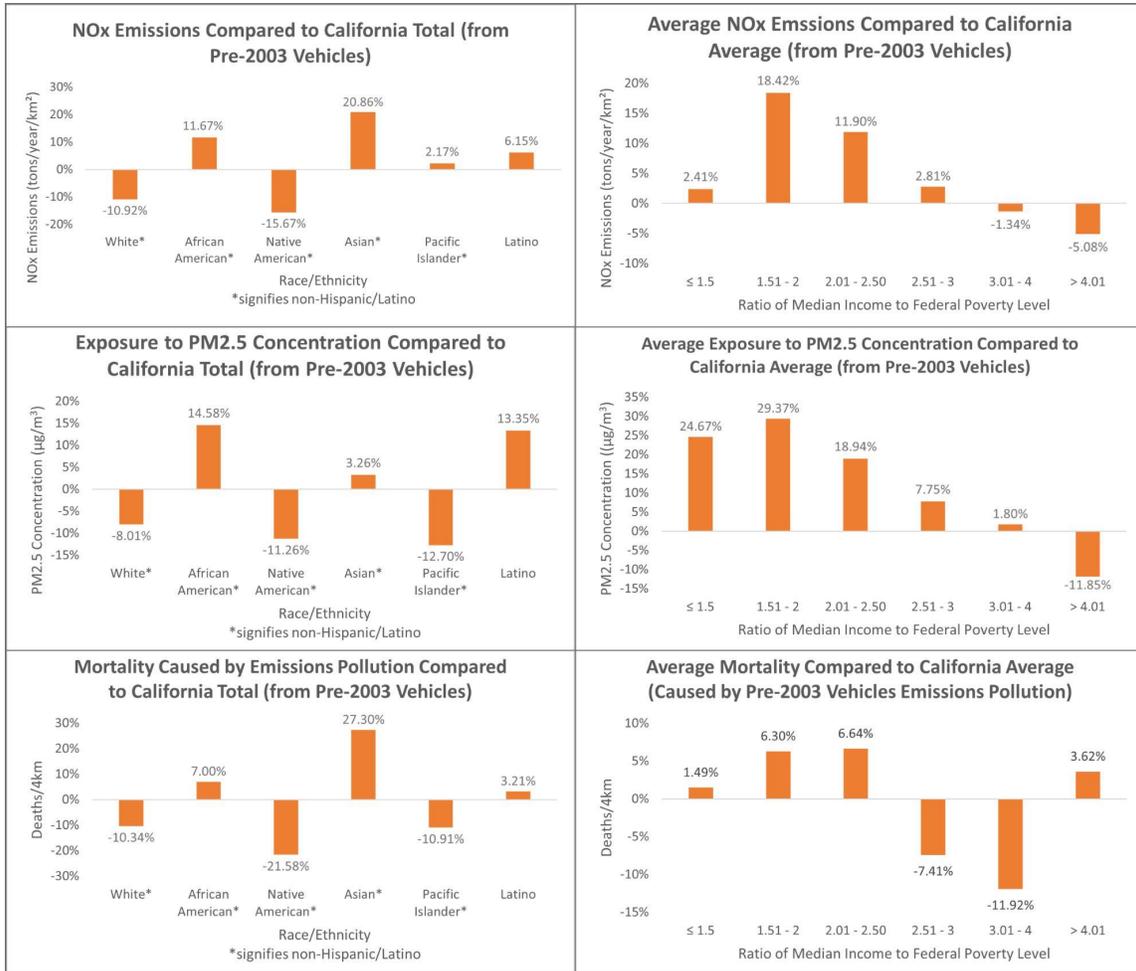


Figure 2.5.11 California race and income comparisons to state totals/averages of NO_x emissions, PM_{2.5} concentrations, and mortality.

In looking at the whole state of California (Figure 2.5.11), we can see that the Asian (20.86%), African American (11.67%) and Latino (6.1%) populations live in areas with above average (compared to total) NO_x emissions from pre-2003 vehicles. Meanwhile, those who fall in the ratio of median income to federal poverty level range of 1.51 - 2 (18.42%) and 2.01 - 2.5 (11.90%) inhabit areas that contribute the most to NO_x emissions. As for air quality, African American (14.58%) and Latino (13.35%) populations experience the poorest air quality, as they had greater exposure to PM_{2.5} concentrations compared to the state total. Those whose ratio of median income to federal poverty level is less than 1.5 (24.67%) experience poor air quality, but more so for those in the 1.51 - 2 range (29.37%). When looking at mortality, the Asian (27.30%) population experiences more deaths compared to the state total. Those with a

ratio of median income to federal poverty level in the ranges of 1.51 - 2 (6.30%) and 2 - 2.5 (6.64%) experience the highest mortality.

In California, African American, Asian, and Latino populations reside in areas that contribute more to NO_x emissions, experience poorer air quality, and suffer from more deaths compared to the state totals. Those with a ratio of median income to federal poverty level equal to or less than 2.5 consistently, on average, live in areas that contribute more NO_x emissions, experience poorer air quality, and higher mortality than the state average. Those in the 1.51 - 2 range produce the most emissions and also experience the poorest air quality. Although they do not suffer from the greatest mortality, they are a close second. An interesting note when looking at the whole state shows that those with a ratio of median income to federal poverty level greater than 4.01 reside in areas that contribute the least emissions, experience the best air quality, yet suffer from more deaths.

For the following analysis conducted on the South Coast, Bay Area, and San Joaquin Valley air districts, some of the changes in trends from emissions to PM and from PM to mortality can be explained by the shifts seen in the spatial maps (Figures 5.2.5 - 5.2.10).

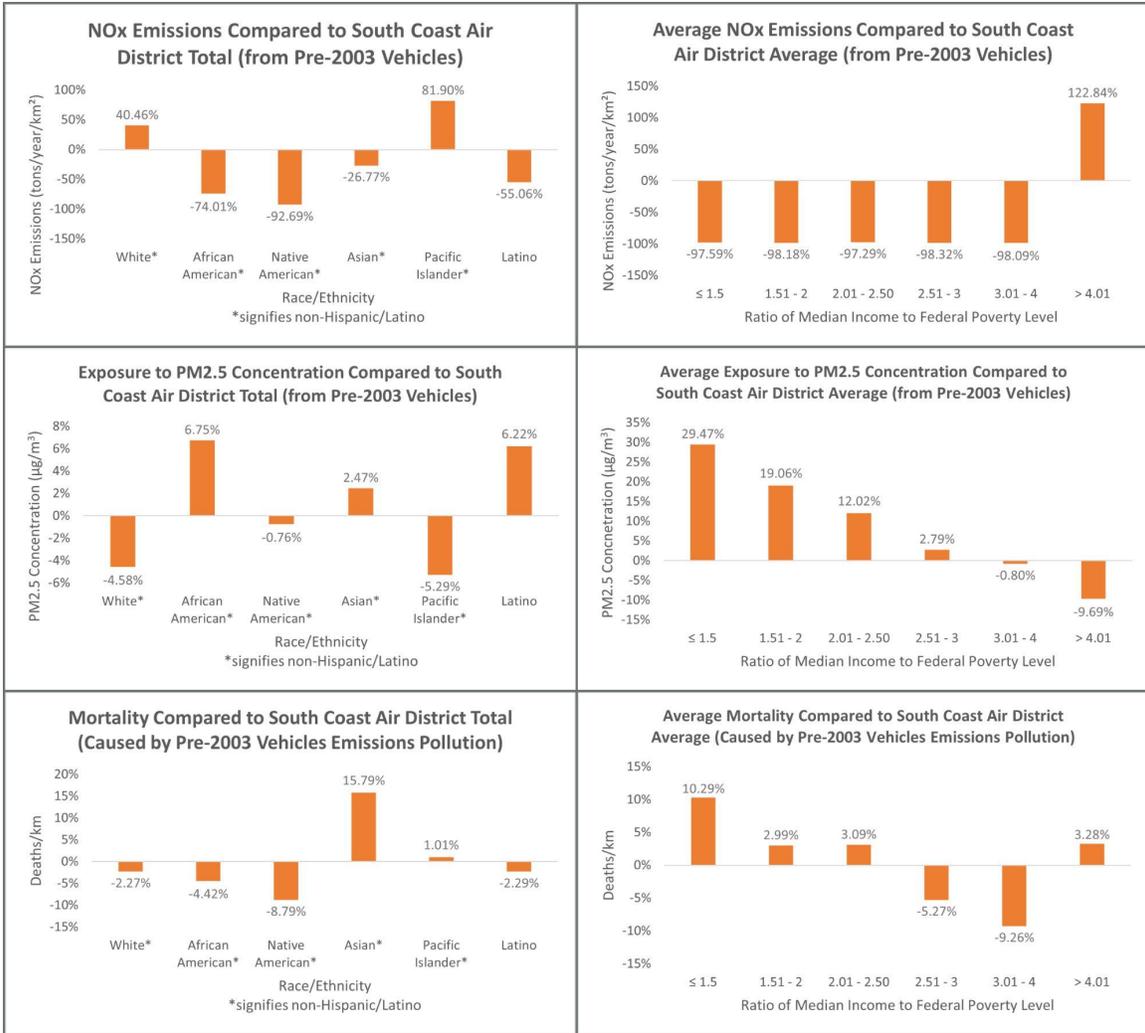


Figure 2.5.12 South Coast Air District race and income comparisons to air district totals/averages of NO_x emissions, PM_{2.5} concentrations, and mortality.

Focusing on the South Coast Air District (Figure 2.5.12), we can see that Pacific Islander (81.90%) and White (40.46%) populations reside in areas that contribute the most to NO_x emissions compared to the whole air district, and so do those with a ratio of median income to federal poverty level greater than 4.01 (122.84%). African American (6.75%) and Latino populations (6.22%) experience a higher concentration of PM_{2.5}. Those with a ratio of median income to federal poverty level equal to or less than 1.5 (29.47%) experience the greatest PM_{2.5} concentration and each subsequent range experiences better air quality. In the South Coast Air District, the Asian (15.79%) population experiences the greatest mortality compared to the whole air district. Those with a ratio of median income to federal poverty level equal to or less than 1.5 (10.29%) experience the most deaths per km.

In the South Coast Air District, the White and Pacific Islander populations live in areas that contribute the most NO_x emissions, yet the African American and Latino populations

experience the poorest air quality and the Asian population suffers from the highest mortality. Similar to the trend seen in the San Joaquin Valley Air District, those with a ratio of median income to federal poverty level greater than 4.01 inhabit areas that contribute the most emissions and air quality improves as the range of ratio of median income to federal poverty level increases. However, in the South Coast Air District, it is those with a ratio of median income to federal poverty level less than or equal to 1.5 who suffer from the highest mortality.

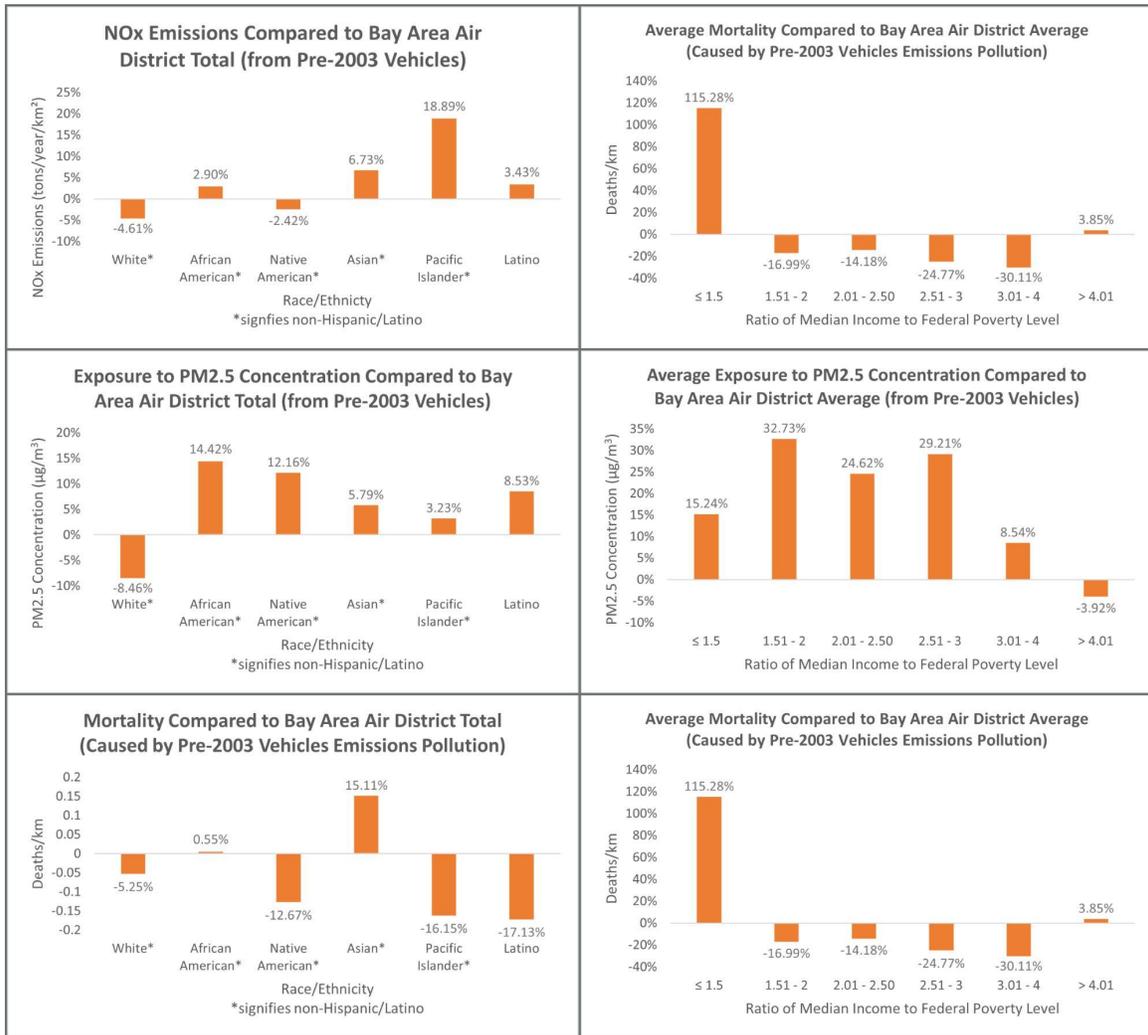


Figure 2.5.13 Bay Area Air District race and income comparisons to air district totals/averages of NO_x emissions, PM_{2.5} concentrations, and mortality.

In the Bay Area Air District (Figure 2.5.13), the Pacific Islanders (18.89%) population is distributed in areas that contribute the most to NO_x emissions compared to the air district total. Those whose ratio of median income to federal poverty level fall between the range of 1.51 - 2 (22.18%) also reside in areas that contribute greatly to NO_x emissions. African American (14.42%) and Native American (12.16%) populations experience poorest air quality in the Bay Area. Those with a ratio of median income to federal poverty level between the ranges of 1.51 -

2 (32.73%), 2.01 - 2.5 (24.62%), and 2.51 - 3 (29.21%) experience a higher concentration of PM_{2.5} compared to the entire air district. As for health effects, the Asian (15.11%) population experiences the highest deaths in the Bay Area. Those with a ratio of median income to federal poverty level equal to or less than 1.5 experience the greatest mortality.

In the Bay Area Air District, although African American and Asian populations don't live in areas that contribute much to NO_x emissions, they still experience poorer air quality and more deaths compared to the air district total. Additionally, those in the ratio of median income to federal poverty level range of 1.51 - 2 inhabit areas that produce the most emissions and suffer from the poorest air quality, but those in the range of less than or equal to 1.5 experience the highest mortality.

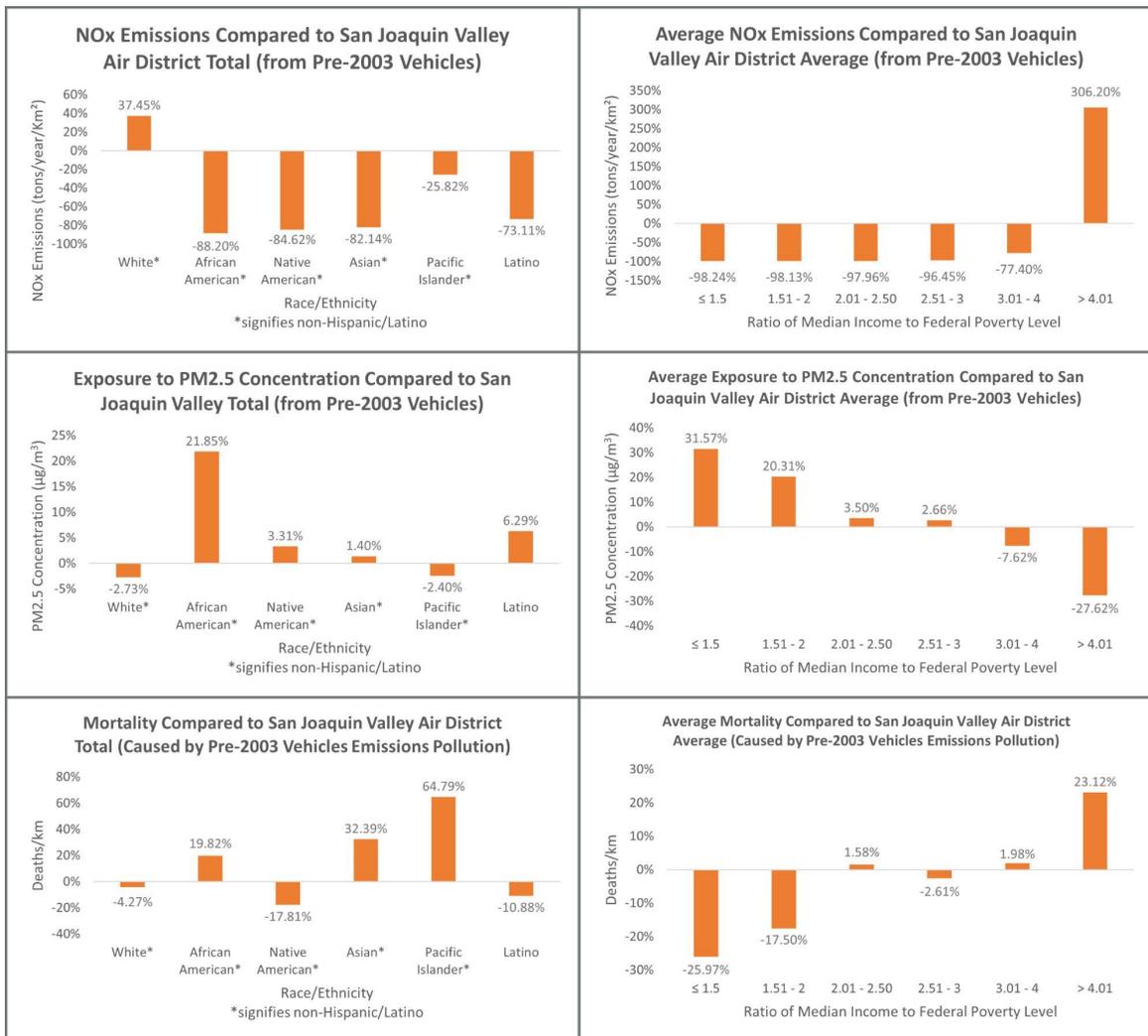


Figure 2.5.14 San Joaquin Valley Air District race and income comparisons to air district totals/averages of NO_x emissions, PM_{2.5} concentrations, and mortality.

In the San Joaquin Valley Air District (Figure 2.5.14), the White (27.45%) population is distributed in areas that contribute the most to NO_x emissions compared to the air district total. Those whose ratio of median income to federal poverty level are greater than 4.01 (306.20%) live in areas that contribute the most to NO_x emissions. The African American population in the San Joaquin Valley Air District experience the poorest air quality and so do those with a ratio of median income to federal poverty level less than or equal to 1.5 (31.57%) and in the range of 1.51 - 2 (20.31%). As for mortality, the Pacific Islander (64.79%) population experiences the greatest number of deaths. Those with a ratio of median income to federal poverty level greater than 4.01 (23.12%) experience the greatest mortality.

Although the White population is heavy in areas that produce the most NO_x emissions in San Joaquin Valley Air District, it is the African American population that experiences the poorest air quality and the Pacific Islander populations that suffer from the most deaths. Those with a ratio of median income to federal poverty level greater than 4.01 inhabit areas that contribute the most emissions but also suffer from the highest mortality. Meanwhile, air quality improves as the range of ratio of median income to federal poverty level increases.

Outreach

Towards the end of our research progress, we contacted several regulators and environmental justice organizations to share preliminary findings from our study as well as gain insights for overall policy recommendations in light of our analysis. We hope that this report can motivate collaboration among decision-makers and environmental justice advocates to support a healthy and safe future for all Californians.

Coalition for Clean Air

We spoke to Chris Chavez, Deputy Policy Director of Coalition for Clean Air. Chavez mentioned that this research is particularly good timing due to the state's surplus of over \$70 billion in funding, and upcoming budget proposal hearings. Currently, transportation programs are trying to get more funding for equity programs to help low-income and disadvantaged communities make the switch to ZEVs. Traditionally, rebate programs and the Clean Cars for All program have been funded through California's cap and trade system, but according to Chavez this funding is often not enough revenue to actually meet the demand of these programs, especially when it comes to environmental justice concerns.

One of Coalition for Clean Air's main priorities has been to get more funding for these programs, and in particular to enhance equity programs to help low-income and disadvantaged people purchase cleaner vehicles. One main suggestion that came out of this meeting was for our research team to overlay a CalEnviroScreen map with the distribution of the older vehicles. While we didn't include this in the report, we did side-by-side comparisons of CalEnviroScreen

to older vehicle distribution in our research presentations which are uploaded to Youtube, and available at this link: <https://youtu.be/ld814Nsou7w>. Some other considerations that we did not include in this analysis were the availability of charging stations and a multiple air toxics study that looks at cancer risk in comparison to the distribution of older vehicles, both important research questions that would be good to address in future studies.

South Coast Air Quality Management District

We initially intended to meet with Sarah Rees, Assistant Executive Officer in the Planning, Rule Development and Area Sources Division (PRDS) at the South Coast Air Quality Management District, but Rees also invited more of her co-workers in the planning division so we were able to get insights from multiple perspectives. Rees suggested that we look particularly into even older vehicles, or limit our definition of an older vehicle to even earlier than 2003, rather than treat all of the vehicles as the same. By analyzing at a different cutoff point, we would be able to test whether our assumption that pre-2003 vehicles are a good cutoff for the most polluting cars. This analysis would be good to explore further in a future project, especially regarding the addition of heavy-duty vehicles, if possible.

Greenlining Institute

We met with Leslie Aguayo, Environmental Equity Program Manager for the Greenlining Institute. Aguayo works specifically with light-duty and heavy-duty vehicles, as well as charging infrastructure. She was also particularly concerned with urban planning and the connection between housing and transportation in a transition to a clean energy grid. One main point from this meeting was regarding the Charge Ahead Coalition, established by AB8, which allowed the DMV to use registration fees towards the Clean Cars for All program. Most notably, there were no equity provisions within AB8, prompting efforts from the Greenlining Institute and other organizations to rectify that. Further, Aguayo mentioned that charging infrastructure is of particular importance, especially in where charging stations are located to be the most accessible and equitable.

Breathe Southern California

We conducted the last outreach meeting with Raj Dhillon, Senior Manager of Advocacy and Public Policy at Breathe Southern California. Breathe Southern California mostly focuses on air quality and lung health impacts, particularly concerning medium and heavy-duty vehicles. Breathe Southern California is advocating a transition away from diesel due to health impacts, but they are having trouble reaching the most affected areas. One reason behind this is that the zip codes that are the most affected don't show up to the outreach events as much as the less affected areas. Likewise, the intended benefits of outreach don't often reach the desired

populations. The biggest takeaway from this meeting in terms of our research report is that even though we only focused on light-duty vehicles, there is also a huge concern on the polluting effects of medium and heavy-duty vehicles. Further, the availability of charging stations may be of even larger concern for heavy-duty vehicles than for light-duty vehicles because they are unable to charge from home charging stations. Still, the availability of charging stations is a concern for light-duty vehicles, and Dhillon mentioned that most charging stations are probably not in the locations that they are most needed, but this is another aspect of study that requires future research.

Conclusions and Recommendations

In our study, we found that older combustion engine vehicles produced before 2003 were unevenly distributed throughout the state of California. Older vehicles are highly concentrated in Southern California, Fresno County, and the San Francisco Bay region. Overall, we found the highest correlation between fraction of older vehicles and the Hispanic/Latino population, although the correlations between race/ethnicity and older vehicles varies regionally. From the emissions graphs, we found that older vehicles seemed to have higher emissions rates when compared to newer vehicles in 2019 overall. The resulting air quality effects shows that removing older vehicles in areas of Southern California, Central Valley, and Bay Area would provide immediate air quality improvements for those regions. The cumulative health impacts of older, pre-2003, combustion engine passenger vehicles show that older combustion engine passenger vehicles contributed an estimated total of 308 deaths across the state of California in 2019, with about 95% of the associated negative health impacts occurring in just 20 counties. This gives reason to believe that a ZEV incentive program that is focused on these specific areas with the highest health impacts could save lives, as well as the burden on the healthcare system, even if it isn't implemented throughout the entire state of California.

In considering the switch to ZEVs in the state of California, we recommend that policymakers focus specifically on communities with a high proportion of African Americans, Asians, and Latinos, since their populations are more heavily distributed in areas with high NO_x and other tailpipe emissions, poor air quality due to secondary $\text{PM}_{2.5}$ emissions, and greater mortality. Similarly, those with a ratio of median income to federal poverty level at or below 2.5 could particularly benefit from financial assistance in switching, especially in single-vehicle households. Additionally, this report in itself is intended as a guiding document to help California make the switch to electric vehicles, and we hope that by distributing the results to the environmental justice organizations that helped inform our research, the communities that are in most need of assistance to purchase an electric vehicle will be able to do so.

We also recommend improving funding for vehicle rebate programs such as Clean Cars for All and other similar ones like the EV rebate for low income residents. However, even if they were granted more funding and worked more effectively, they may not sufficiently account for

environmental justice concerns. While our report aims to address this by correlating environmental justice in relation to older vehicles, there are many aspects of this technological transition process that are left unaddressed by the results of this study. One factor in particular that hinders the adoption of electric vehicles is the availability of vehicle charging stations, in both public spaces and privately owned buildings. Without the ability to conveniently and affordably charge an electric vehicle, single-car households are unlikely to make the switch. This is just one of many factors that serve as a barrier to disadvantaged communities when it comes to purchasing electric vehicles. There are many potential solutions to help remove these barriers, such as subsidizing the price of using charging stations to be even more cost-effective, and increasing Spanish language outreach to make existing scrap and replace programs more accessible.

Additionally, even if programs were able to sufficiently address environmental justice considerations, the remaining question of power grid would remain. While this may be less of a concern in terms of air quality due to the immediate reduction in tailpipe emissions, the long-term increase in electricity demand and consequences of a dirty energy grid contradict the goal of creating a clean vehicle fleet, and contribute to the global climate crisis. Even electric vehicles will not be truly clean until the power grid that is manufacturing and powering the electric vehicles is running on renewable and sustainable energy sources. Additionally, the energy grid itself could be limited in its ability to provide for a significant increase in electric vehicle users, so consideration needs to be taken in expanding the energy grid to meet even higher electricity demands in response to ZEV incentivization programs and increasing gasoline prices. Improving infrastructure goes beyond electrifying the grid, however. One way of reducing the need for older vehicles is to promote denser land use development and improved public transportation to reduce car dependency and the culture of car reliance in California.

Furthermore, this study represents just a fraction of the vehicle emissions, as it is representative of only the light-duty passenger vehicle fleet. The fact that this study excluded medium and heavy duty vehicles is not in any way indicative of their importance, only that there is not enough information presently available to conduct such research. The air quality impact of medium and heavy duty vehicles is concerning, especially to those who live near ports, freight corridors, and/or highways. Unfortunately, research opportunities are further limited by the fact that the vehicle registration locations for medium and heavy duty vehicles are less likely to be correlated with the location of their emissions than those of light-duty vehicles due to the prevalence of large commercial fleets. Thus, other methods to distribute these emissions would need to be implemented. While this limitation also exists to a smaller extent for light-duty passenger vehicles, it is a safer assumption that personal light-duty vehicles will be driven in the area of their registration, which is how we were able to conduct this study.

While the division of this research into five sections served as an advantage for our time constraints and allowed us to conduct many levels of analysis, it also brought several

disadvantages, and a potential source of error or misrepresentation. Having distinct teams assigned to particular tasks made it more difficult to maintain communication between the results of the groups and have a broader understanding of the interconnectedness of these factors. While we have attempted as best we could to weave together the analysis, there are likely other interconnected aspects that we have failed to identify. Moreover, the process of sharing files itself adds complication if certain aspects of the data are not fully communicated. However, this risk is small given our frequent communications, and it was well-worth the tradeoff of being able to combine so many aspects of analysis under one report.

In conclusion, further studies will need to be done to assess the sensitivity of removing older vehicles to find out if setting a more realistic vehicle removal threshold would still yield significant effects on air quality and health effects, or alternatively, if a strategy to remove vehicles starting at an even older cutoff date. Additionally, an evaluation of how different incentive levels would accelerate the pace of older vehicle removal would prove useful. While our analysis is not advocating in particular for the removal of all pre-2003 vehicles indiscriminately, these results suggest that focusing on older vehicle removal to reduce emissions could have tangible air quality, health effects, and environmental justice benefits. While 2003 seems to be a suitable year to distinguish the cars with the most negative impacts, it is clearly 6 millions of light-duty older vehicles that would have to be taken off of the roads, and is unlikely to occur rapidly. More realistically, programs designed to phase out older polluting vehicles should be designed to target areas with the most significant impacts, the most older polluting vehicles, and the areas with households that potentially could benefit the most, which we identified in this study. Furthermore, focusing on particular regions of interest could help reduce the environmental injustices throughout the state that contribute unequal harms to vulnerable communities. Due to the correlations between the location of older vehicles, income, and race, there may be evidence to suggest that certain demographics may need more assistance by means of policies or incentives to make the switch to ZEVs. This may be especially true in cases where the only household vehicle is a high-emitting older vehicle. Overall, given the distribution of older vehicles, these results call for a consideration of income and race as deciding factors for how to conduct programs intended to accelerate the transition to ZEVs in California.

Literature Cited

- California Air Resources Board. (2021-a). *Emissions Inventory*. EMFAC.
<https://arb.ca.gov/emfac/emissions-inventory>.
- California Air Resources Board. (2021-b). *Fleet Database*. EMFAC.
<https://arb.ca.gov/emfac/fleet-db/6315133214b0fa903134ec2112f763a2f231109b>.
- California Environmental Protection Agency. (2018, June). *CalEnviroScreen 3.0*. California Office of Environmental Health Hazard Assessment.
<https://oehha.ca.gov/calenviroscreen/report/calenviroscreen-30>.
- California New Car Dealers Association. (2021, February). California Auto Outlook.
<https://www.cnca.org/wp-content/uploads/Cal-Covering-4Q-20.pdf>
- Caltrans. (2021). Zero-Emission Vehicles. Zero-Emission Vehicles | Caltrans.
<https://dot.ca.gov/programs/sustainability/zero-emission-vehicles>.
- Covered California. (2021). *What is the federal poverty level?: Covered California™*. What is the federal poverty level?
<https://www.coveredca.com/support/financial-help/federal-poverty-level/>.
- Krewski D, Jerrett M, Burnett RT, Ma R, Hughes E, Shi Y, Turner MC, Pope CA III, Thurston G, Calle EE, Thun MJ. 2009. Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality. HEI Research Report 140. Health Effects Institute, Boston, MA.
- Office of Governor Gavin Newsom. (2020, September 23). Governor Newsom Announces California Will Phase Out Gasoline-Powered Cars & Drastically Reduce Demand for Fossil Fuel in California's Fight Against Climate Change. California Governor.
<https://www.gov.ca.gov/2020/09/23/governor-newsom-announces-california-will-phase-out-gasoline-powered-cars-drastically-reduce-demand-for-fossil-fuel-in-californias-fight-against-climate-change/>.
- Ostro, B., Tobias, A., Querol, X., Alastuey, A., Amato, F., Pey, J., Pérez, N., & Sunyer, J. (2011). The Effects of Particulate Matter Sources on Daily Mortality: A Case-Crossover Study of Barcelona, Spain. *Environmental Health Perspectives*, 119(12), 1781–1787.
<https://doi.org/10.1289/ehp.1103618>
- State of California Department of Housing and Community Development. (2020, April 30). 2020 Income Limits.
<https://www.hcd.ca.gov/grants-funding/income-limits/state-and-federal-income-limits/docs/income-limits-2020.pdf>.
- Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 15.0 [dataset]. Minneapolis, MN: IPUMS. 2020. <http://doi.org/10.18128/D050.V15.0>

- Tessum, C., Hill, J., & Marshall, J. (2017-a). InMAP: An Intervention Model for Air Pollution. InMAP. <http://spatialmodel.com/inmap/>
- Tessum, C. W., Hill, J. D., & Marshall, J. D. (2017-b). InMAP: A model for air pollution interventions. PLOS ONE, 12(4), e0176131. <https://doi.org/10.1371/journal.pone.0176131>
- Tétreault, L.-F., Doucet, M., Gamache, P., Fournier, M., Brand, A., Kosatsky, T., & Smargiassi, A. (2016). Severe and Moderate Asthma Exacerbations in Asthmatic Children and Exposure to Ambient Air Pollutants. International Journal of Environmental Research and Public Health, 13(8), 771. <https://doi.org/10.3390/ijerph13080771>
- TIGER/Line Shapefiles. (n.d.). United States Census Bureau. <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html>
- U.S. Census Bureau. (2020, April 7). *Average Household Size and Population Density - County*. Census COVID-19 Data Hub. https://covid19.census.gov/datasets/21843f238cbb46b08615fc53e19e0daf_1/explore?location=16.793879%2C0.315550%2C1.98&showTable=true
- U.S. Department of Housing and Urban Development. (2021, April 1). *Income Limits*. Income Limits | HUD USER. <https://www.huduser.gov/portal/datasets/il.html#2021>
- U.S. Environmental Protection Agency. (N.d.). Environmental Benefits Mapping and Analysis Program - Community Edition (BenMAP-CE). <https://www.epa.gov/benmap>
- U.S. Environmental Protection Agency. (2020-a, November 20). *Mortality Risk Valuation*. EPA. <https://www.epa.gov/environmental-economics/mortality-risk-valuation>.
- U.S. Environmental Protection Agency. (2020-b, December 4). Sources of Greenhouse Gas Emissions. U.S. Environmental Protection Agency. <https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions#:~:text=The%20largest%20source%20of%20greenhouse,electricity%2C%20heat%2C%20and%20transportation.&text=Transportation%20>.

Appendix

A. Row and Column Conversions from X and Y

Row and Columns were assigned using QGIS. First, a reference grid was created using an existing "Create Grid" function in QGIS, that automatically has X and Y columns assigned based on a numerical number system in meters. As you move down rows from top to bottom on the grid, the Y coordinate gets smaller, so the Y column was used to compute the rows. Similarly, when you move across columns from right to left on the grid, the X coordinate gets larger, so the X column was used to compute columns. In other words, anytime the X value is the same for two grid cells, the column number is the same, and anytime the Y value is the same for two grid cells, the row value is the same. A singular reference origin point was chosen at the top left point of the grid, but rather than representing an origin of (0,0), it needed to represent (1,1) due to the numbering system that BenMAP uses for column and row data points. Each step from one grid cell to the next is based on the resolution of the grid (for example, the step size for a 12 kilometer grid is 12,000 meters, so the X value would increase by 12,000 as the column value increases by 1). Likewise, each step on the grid also needs to correspond to a change of 1 according to the row and column numbering system. This conversion method depends on knowing the value of the origin point at the top left corner of the reference grid.

To assign columns, we used the Field Calculator of the Attribute Table in GIS to input the

following function: $\frac{(X - X_0)}{(X_1 - X_0)} + 1$, where X is the variable that the Field calculator

automatically fills in for each of the X variables, X_0 is the X value of the origin point, and X_1 is the second X value, such that $(X_1 - X_0)$ is equivalent to the step size for the resolution, and 1 is a constant that was added to make sure that the first computation for the origin point outputs 1 rather than zero, and the following points are computed accurately as well.

To assign rows, we used the Field Calculator of the Attribute Table in GIS to input the

following function: $\frac{(-Y + Y_0) + 1}{(Y_1 - Y_0)}$, where Y is the variable that the Field calculator

automatically fills in for each of the Y variables, Y_0 is the Y value of the origin point, and Y_1 is the second X value, such that $(Y_0 - Y_1)$ is equivalent to the step size for the resolution, and 1 is a constant that was added to make sure that the first computation for the origin point outputs 1 rather than zero, and the following points are computed accurately as well. The two equations for Column and Row differ in the sign of the variable and whether it is added or subtracted in

relation to the origin value. The reasoning behind this is because X values started out negative and increased from left to right, whereas Y values started out positive and decreased from top to bottom, so it was to account for these differences and create a column-row grid such that the largest values for Column and Row would be in the bottom right corner.

B. Row and Column Conversions from ID Column

The air quality data needed to be attached to corresponding Column and Row attributes, but the output shapefile only contained data for an ID column that needed to be split into columns and rows. Because we used the same underlying grid as the reference grid, we already knew how many rows there were, so we used a function that took the total number of rows in the reference grid as an input in the Field Calculator in QGIS, and returned a unique Column and Row attribute for each grid cell of the air quality data.

A modulus is a function represented by “%” that returns the remainder of a division operation, and this type of function was used for computing each row. Since we knew for the 12 kilometer grid there should be 101 distinct rows based on the number of rows in the reference grid, the remaining values the modulus function needed to output for this resolution ranged from 0 to 100 (and similar assessments were made for other resolutions). To make this the case, the following function was used:

```
CASE
WHEN (ID % N) = 0 THEN 101
ELSE (ID % N)
END
```

In this function, ID is the variable for the ID column, which QGIS automatically replaces with each unique ID value when run. N represents the number of rows in the reference grid. The “WHEN” statement represents the fact that the remainder value of “0” actually resembles row 101, so rather than outputting 0 as the row number, this function will correctly identify the cell with ID 101 as in row 101. The rest of the modulus results output a value that aligns with the row number.

To assign column values, the following function was used:

```
Ceil(ID/N)
```

Ceiling functions round up to the nearest whole number. So for all of the cell ID numbers that are in column 1, will return a fraction that may be much smaller than 1, but will round up to the nearest whole number due to the ceiling function. For example, cell ID 1 on a 12 kilometer grid

would be assigned to Column 1, because even though 1/101 is a miniscule number, the nearest whole number is still 1. As soon as the ID number becomes larger than the row number, the nearest whole number becomes 2, and so forth. While the 12 kilometer resolution had 101 rows, other finer resolutions had a greater number of rows, but we simply replaced the total number of rows in the equations.

C. CSV Input Formats

Link to Health Effects section data files for Inputs:

https://drive.google.com/drive/folders/1HIJz5xj8ljbX9W2U2Pj_yL2IaTFwbkTw?usp=sharing

To find a particular input, pick a resolution folder and a location folder, and then select the folder titled "Inputs" to see the CSV files uploaded to BenMAP for Baseline and Control scenarios.

D. All Resolutions for LA County Comparison

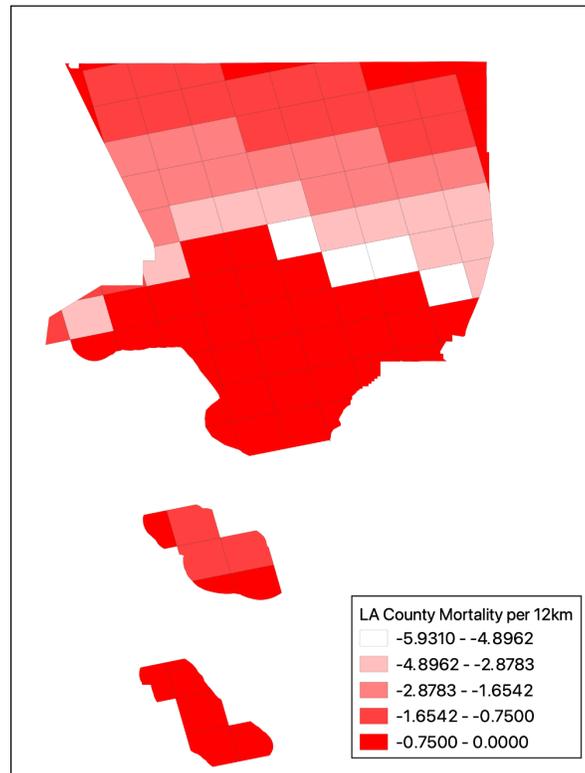


Figure D1 Number of deaths per kilometer due to air quality effects from pre-2003 vehicle emissions in Los Angeles County in 2019. Total estimated deaths: 121.98.

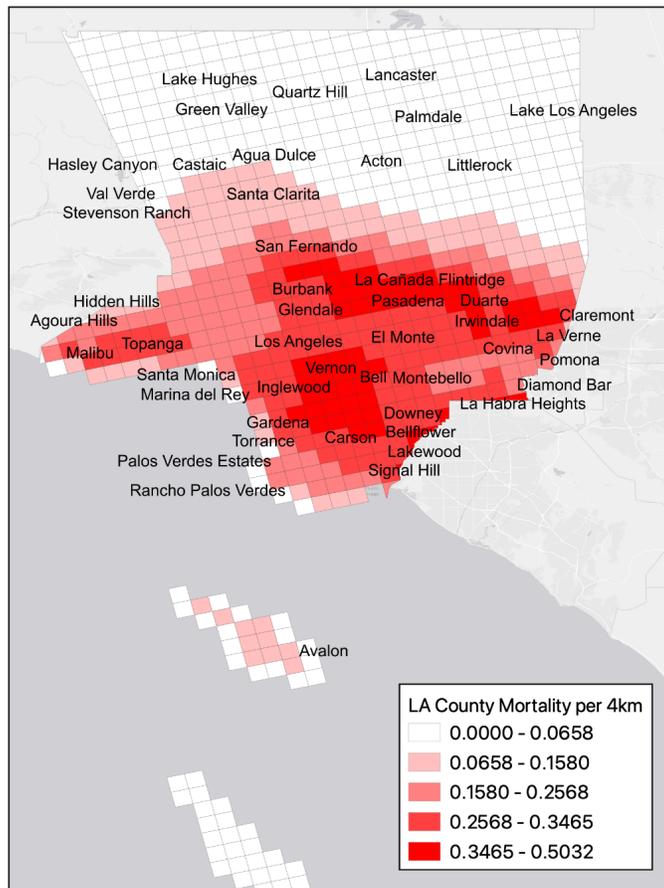


Figure D2 *Number of deaths per 4 kilometers due to air quality effects from pre-2003 vehicle emissions in Los Angeles County in 2019. Total estimated deaths: 104.77.*

E. Number of vehicles per household shows that households with multiple vehicles contribute to 75% of the emissions

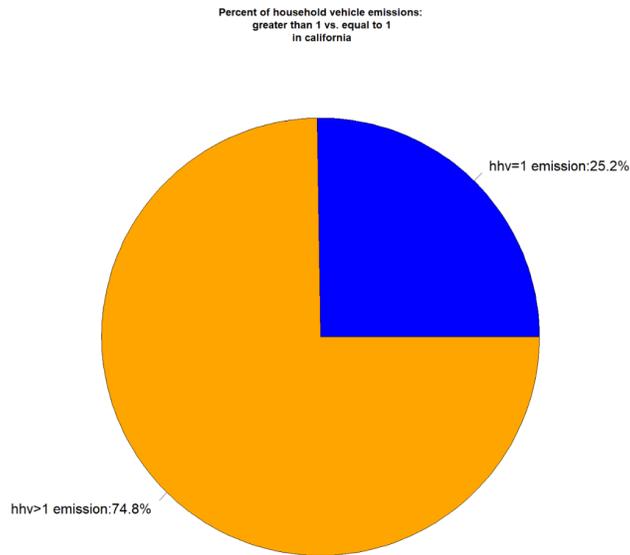


Figure E *Pie chart of the percentage of emission rate for household vehicles equal to one and greater than one.*

F. 4km level for Orange County

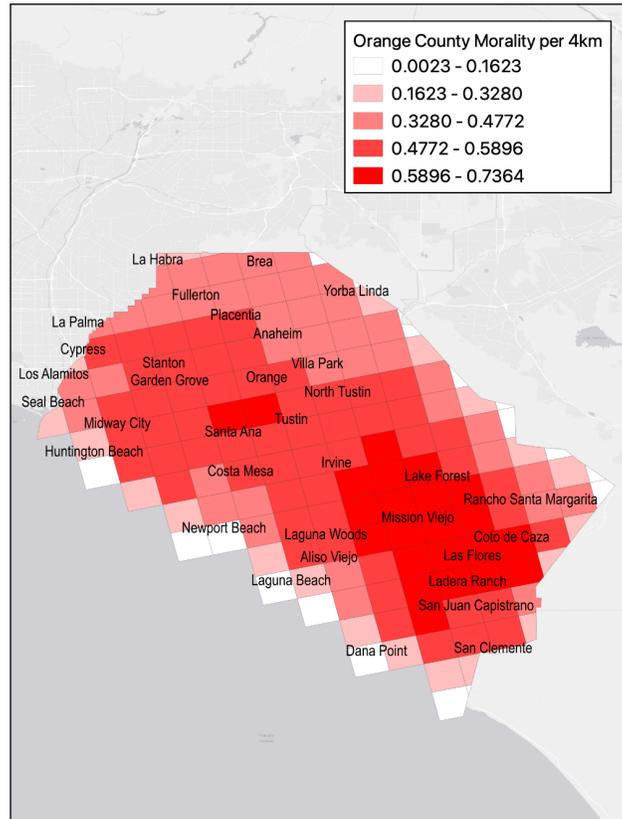


Figure F *Number of deaths per 4 kilometers due to air quality effects from pre-2003 vehicle emissions in Orange County in 2019.*

G. Other Health Functions for 4km California

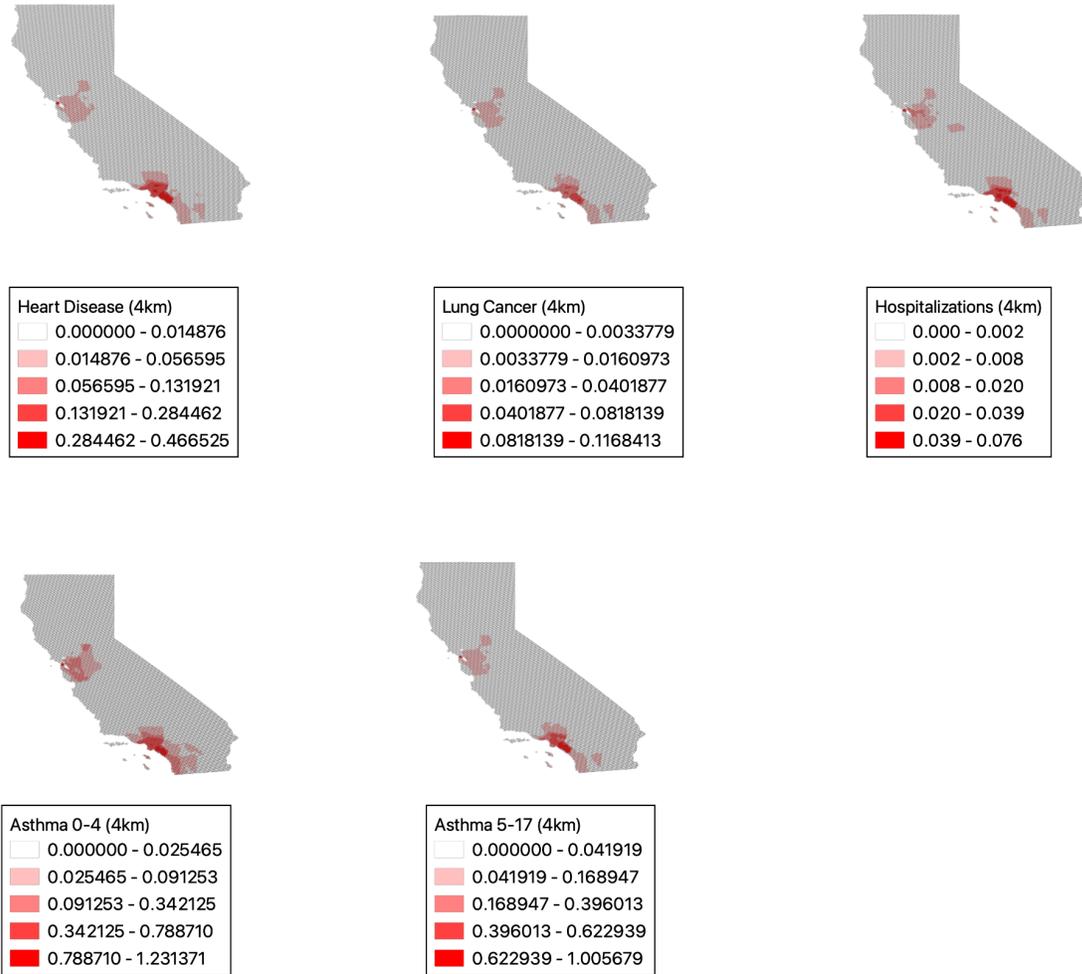


Figure G Each of these five maps represents one of the health equations that was tested at the 4km resolution statewide. Heart disease and Mortality estimates represent mortality endpoints, whereas hospitalizations refers to all respiratory related hospital admissions, and asthma estimates are based on incidence rates.

H. Other Health Functions by County (4km Resolution Aggregated by County)

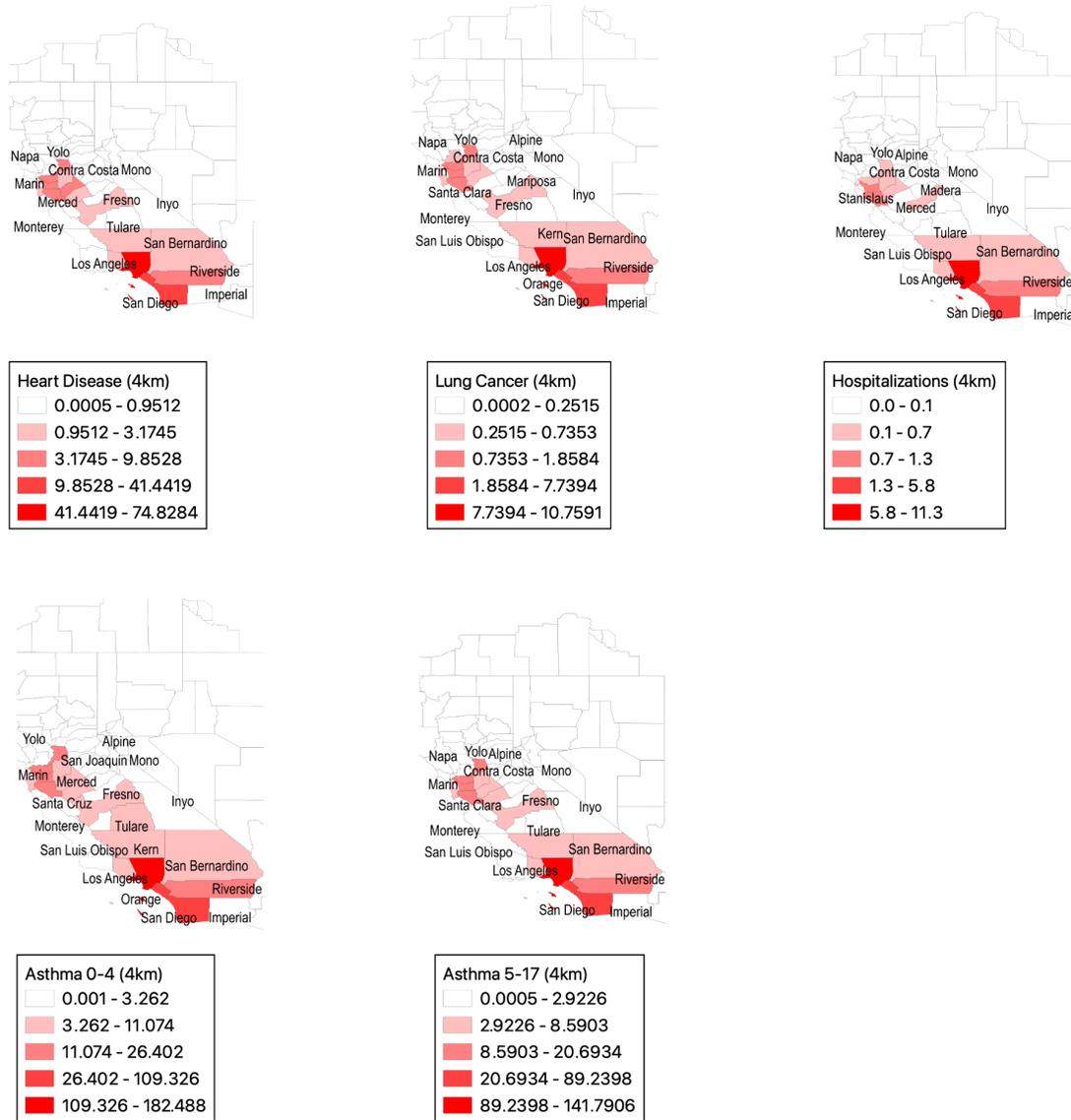


Figure H Each of these five maps represents one of the health equations that was tested at the 4km resolution statewide, and then aggregated by county. Heart disease and Mortality estimates represent mortality endpoints, whereas hospitalizations refers to all respiratory related hospital admissions, and asthma estimates are based on incidence rates.

I. Other Health Functions for San Joaquin Valley Air District

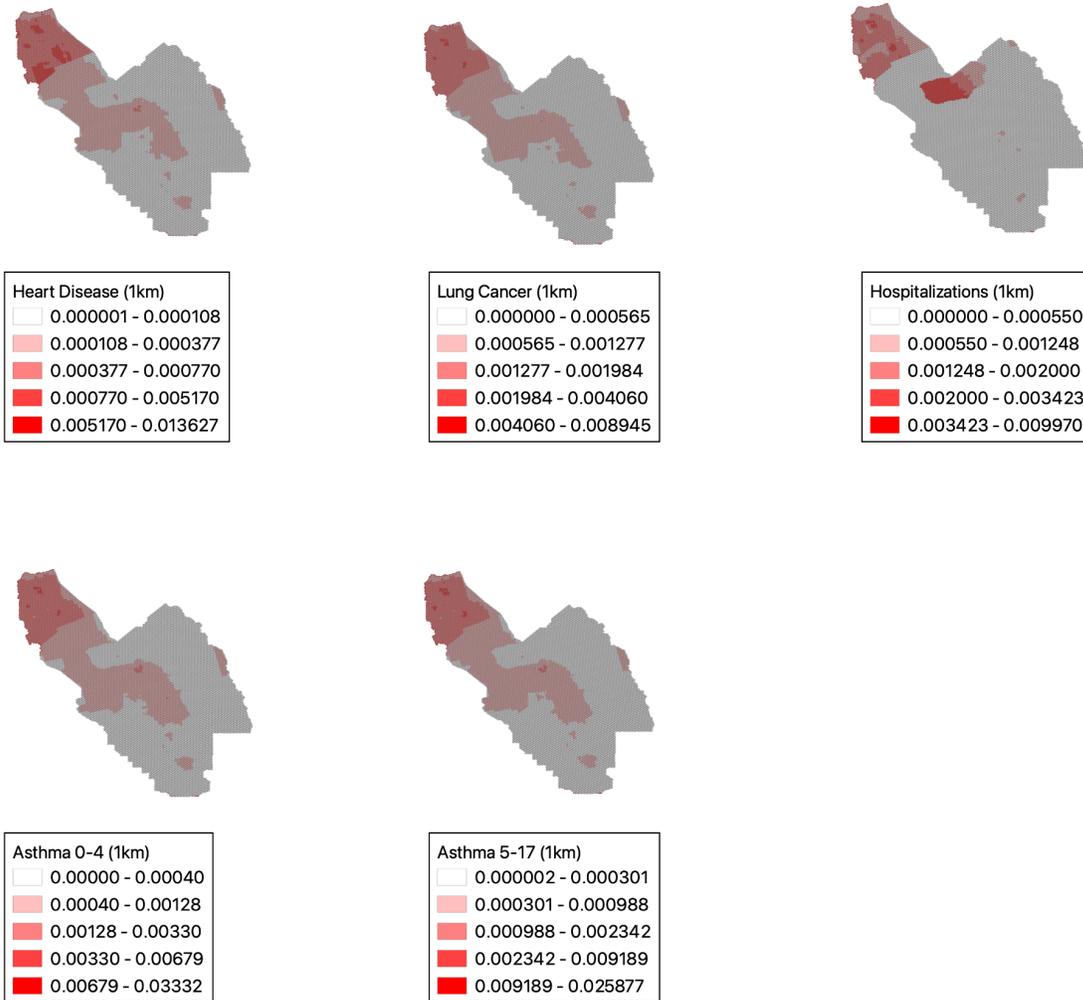


Figure I Each of these five maps represents one of the health equations that was tested at the 1km resolution for San Joaquin Valley Air Quality District. Heart disease and Mortality estimates represent mortality endpoints, whereas hospitalizations refers to all respiratory related hospital admissions, and asthma estimates are based on incidence rates.

J. Other Health Functions for Bay Area Air District

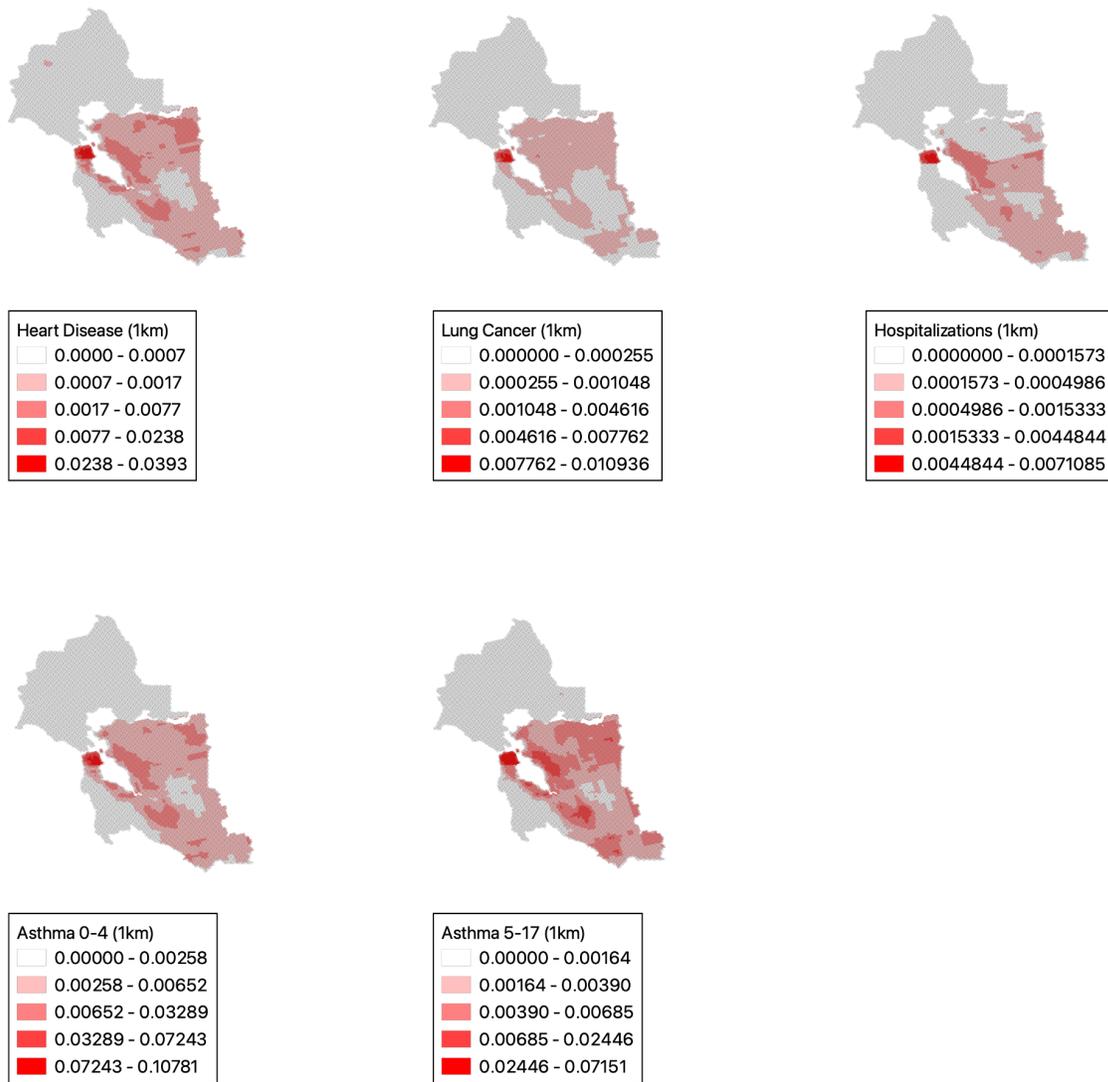


Figure J Each of these five maps represents one of the health equations that was tested at the 1km resolution for the Bay Area Air Quality Management District. Heart disease and Mortality estimates represent mortality endpoints, whereas hospitalizations refers to all respiratory related hospital admissions, and asthma estimates are based on incidence rates.

K. Other Health Functions for South Coast Air District

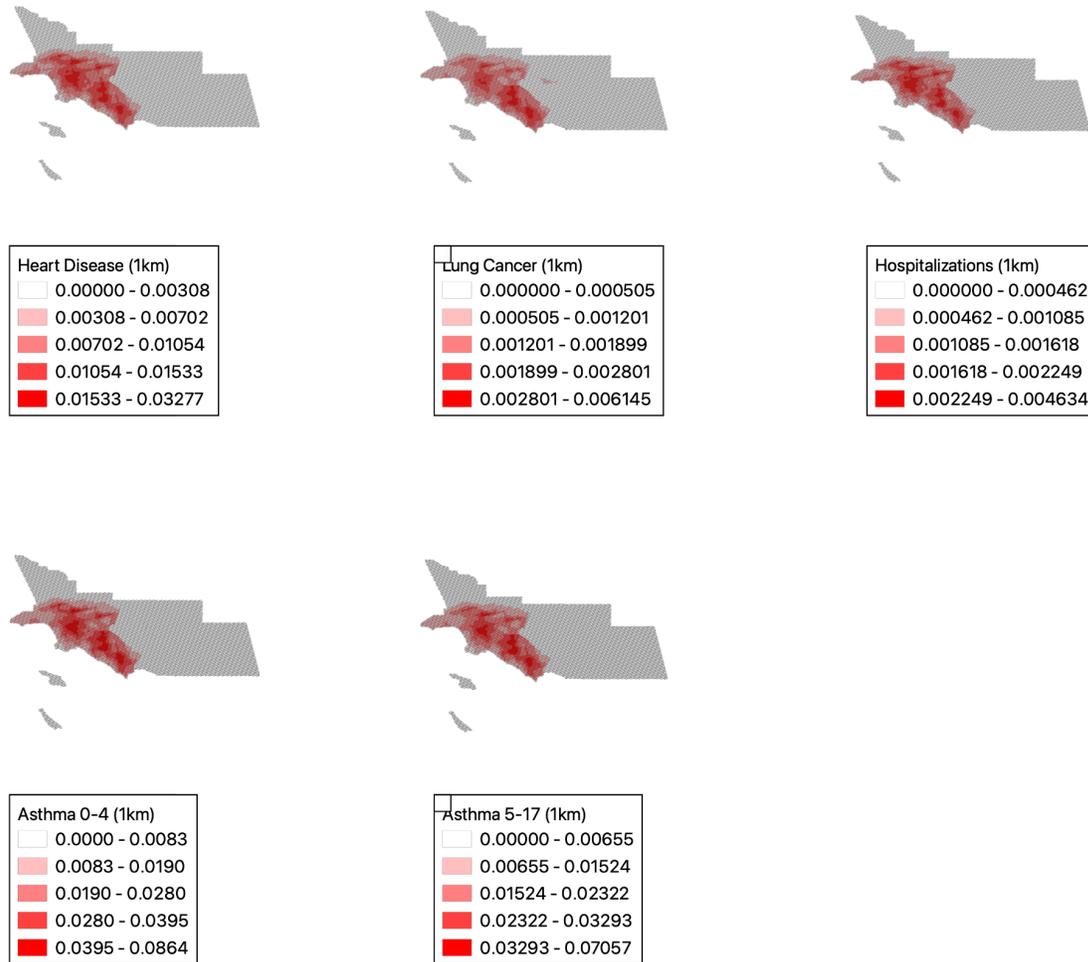


Figure K Each of these five maps represents one of the health equations that was tested at the 1km resolution for the Southcoast Air Quality Management District. Heart disease and Mortality estimates represent mortality endpoints, whereas hospitalizations refers to all respiratory related hospital admissions, and asthma estimates are based on incidence rates.

L. Other Health Functions for Los Angeles County

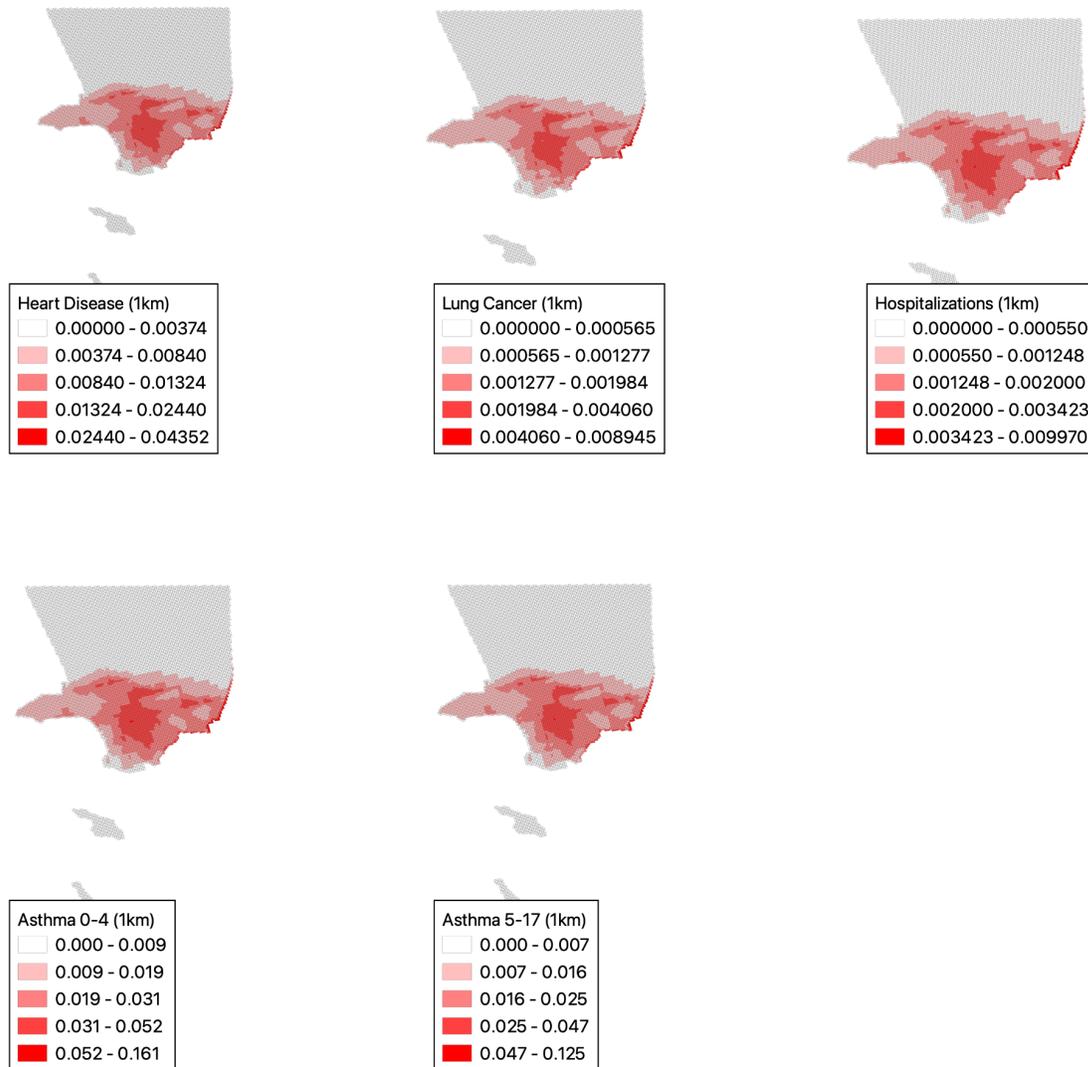


Figure L Each of these five maps represents one of the health equations that was tested at the 1km resolution for Los Angeles County. Heart disease and Mortality estimates represent mortality endpoints, whereas hospitalizations refers to all respiratory related hospital admissions, and asthma estimates are based on incidence rates.

M. Other Health Functions for Orange County

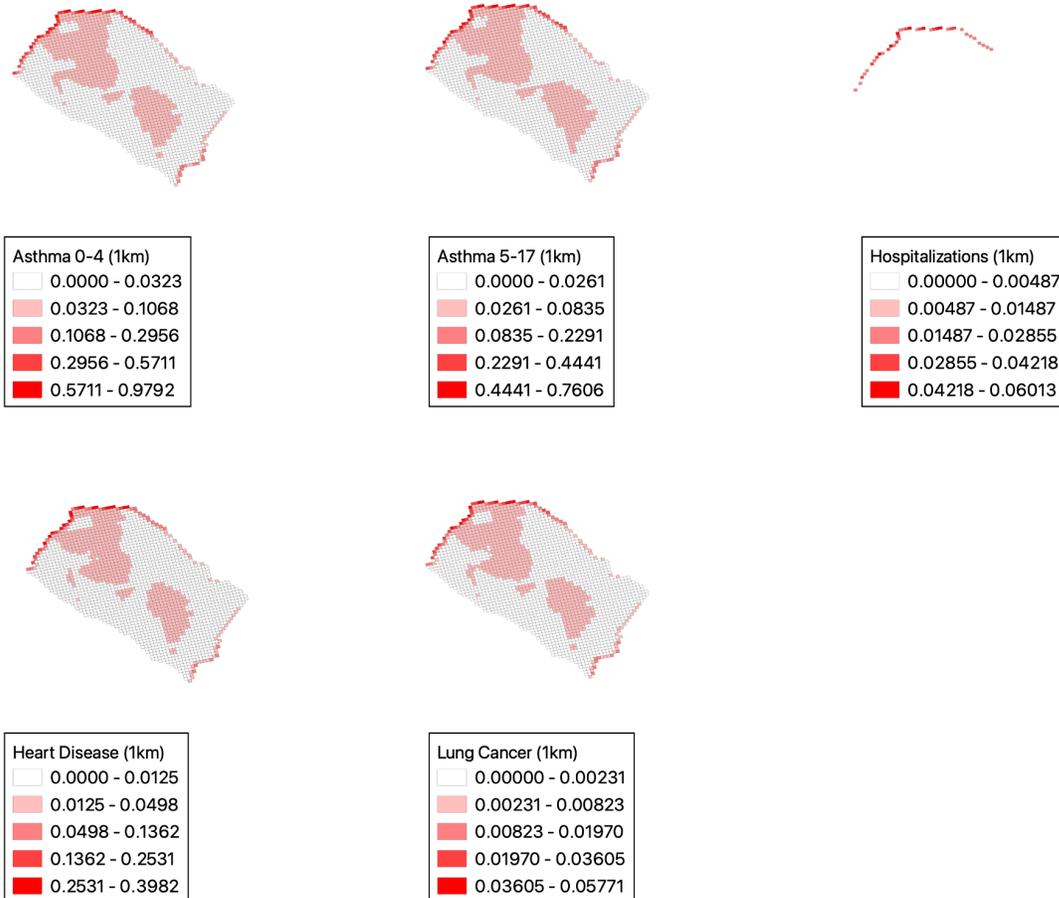


Figure M Each of these five maps represents one of the health equations that was tested at the 1km resolution for Orange County. Heart disease and Mortality estimates represent mortality endpoints, whereas hospitalizations refers to all respiratory related hospital admissions, and asthma estimates are based on incidence rates.

N. Raw number calculations of NOx Emissions, PM2.5 Concentration, and Mortality for California by race/ethnicity and ratio of median income to federal poverty level

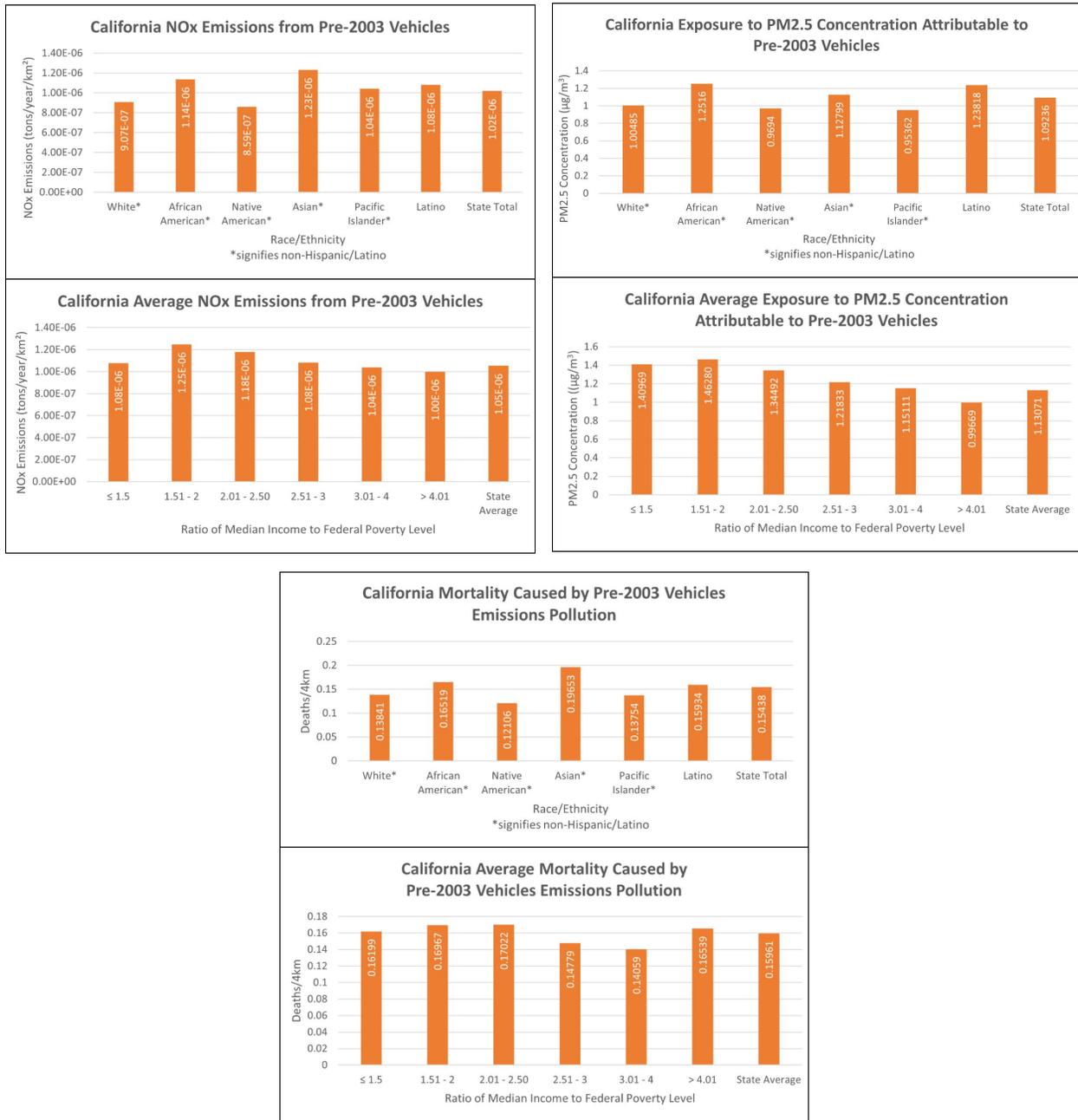


Figure N These bar plots compare the raw values of NOx emissions, PM2.5 concentrations, and mortality by race/ethnicity and ranges of ratio of median income to federal poverty level in California at the census tract level.

O. Raw number calculations of NOx Emissions, PM2.5 Concentration, and Mortality for South Coast Air District by race/ethnicity and ratio of median income to federal poverty level

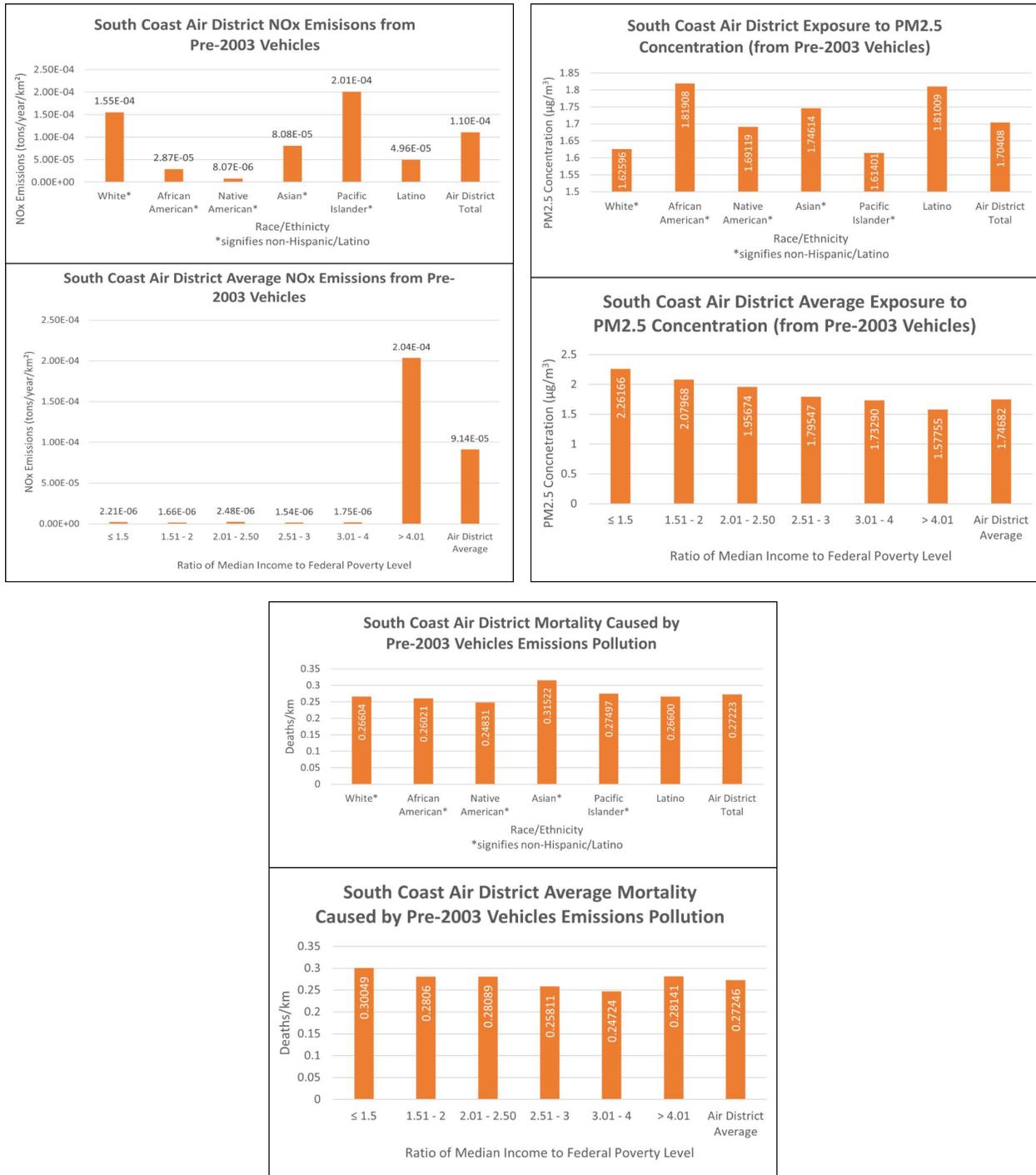


Figure O These bar plots compare the raw values of NOx emissions, PM2.5 concentrations, and mortality by race/ethnicity and ranges of ratio of median income to federal poverty level in the South Coast Air District at the census tract level.

P. Raw number calculations of NOx Emissions, PM2.5 Concentration, and Mortality for Bay Area Air District by race/ethnicity and ratio of median income to federal poverty level

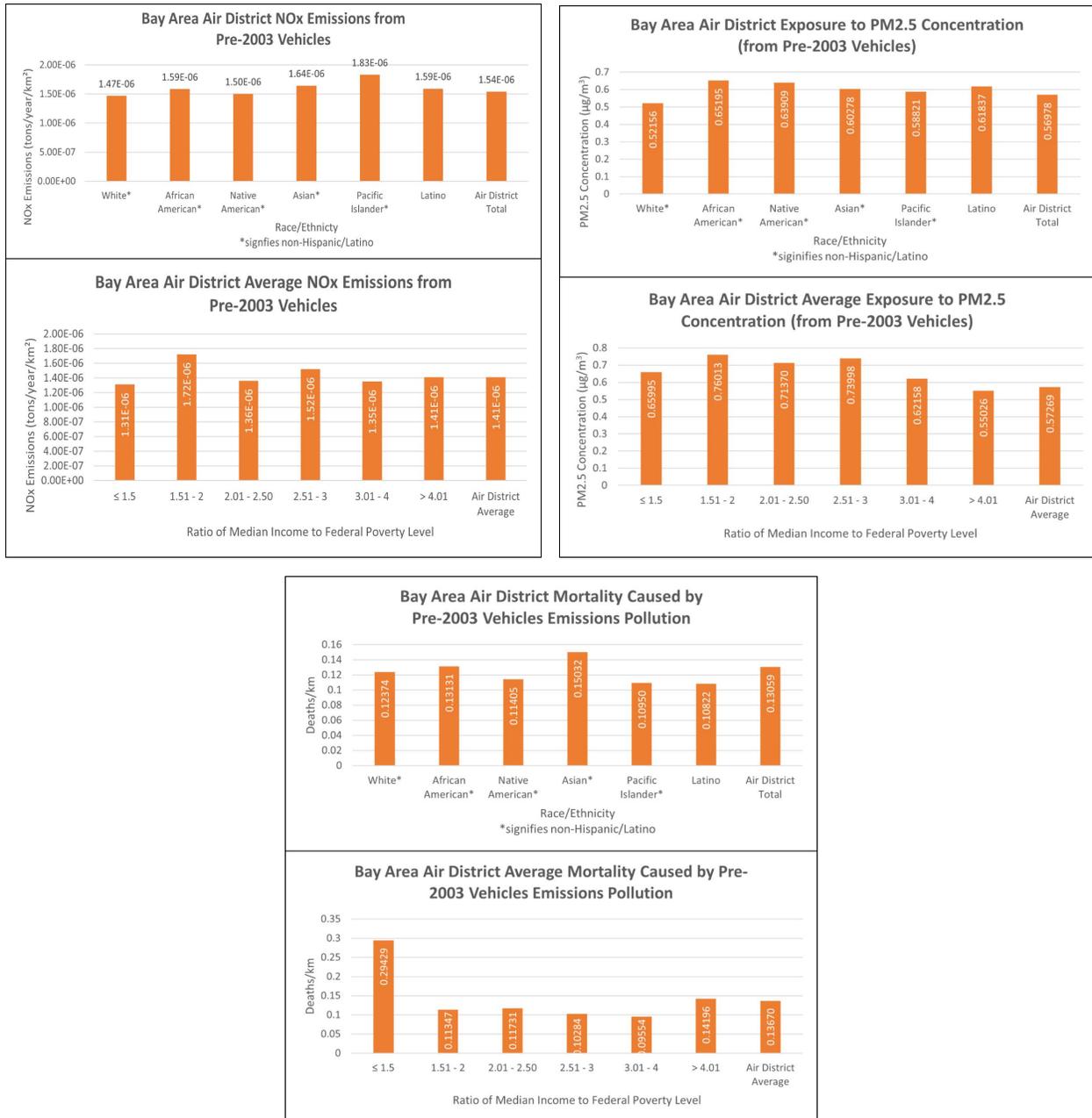


Figure P These bar plots compare the raw values of NOx emissions, PM2.5 concentrations, and mortality by race/ethnicity and ranges of ratio of median income to federal poverty level in the Bay Area Air District at the census tract level.

Q. Raw number calculations of NOx Emissions, PM2.5 Concentration, and Mortality for San Joaquin Valley Air District by race/ethnicity and ratio of median income to federal poverty level

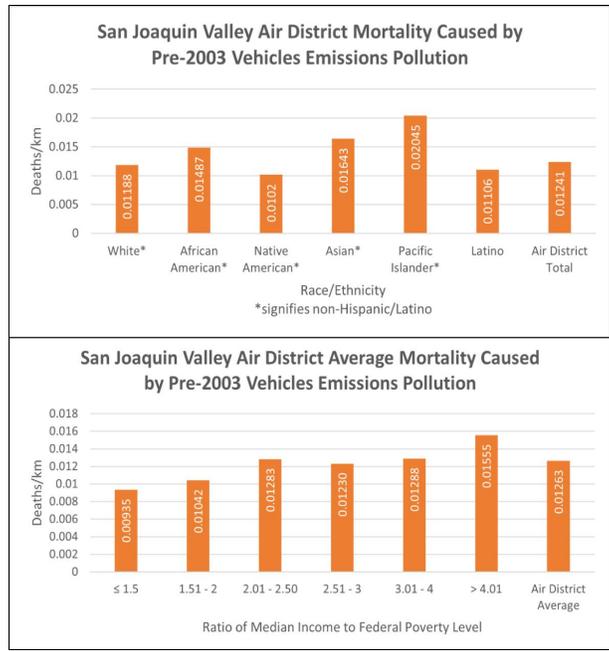
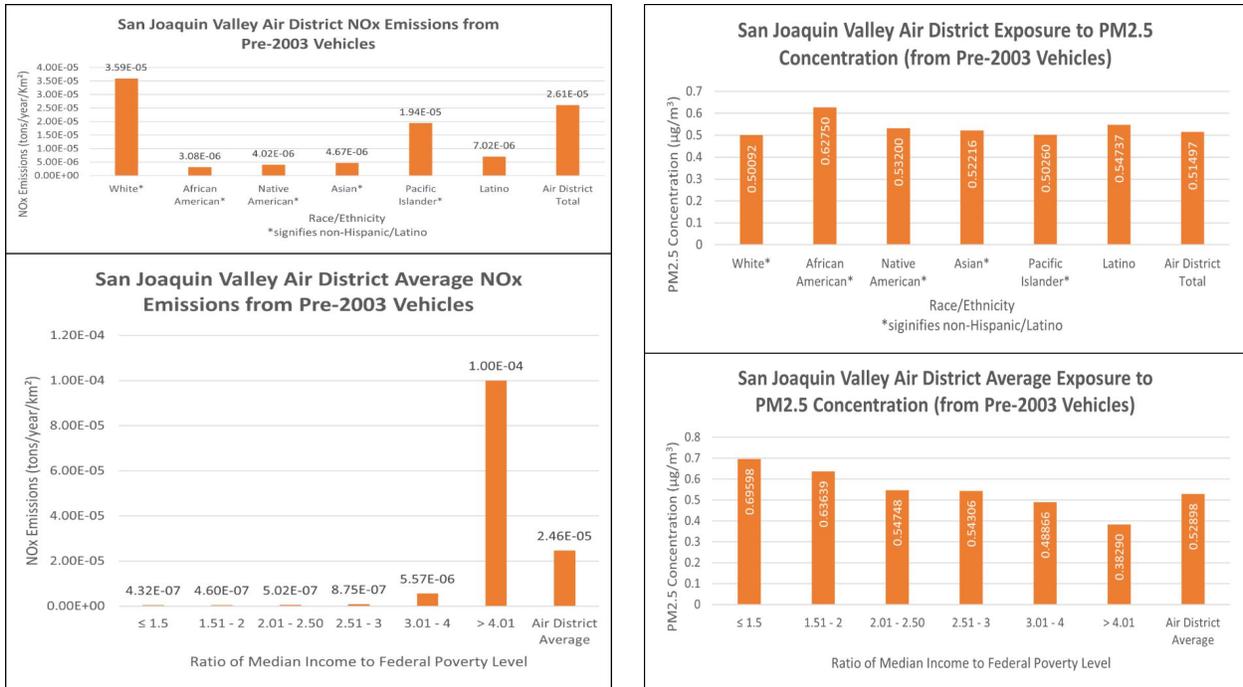


Figure Q These bar plots compare the raw values of NOx emissions, PM2.5 concentrations, and mortality by race/ethnicity and ranges of ratio of median income to federal poverty level in the San Joaquin Valley Air District at the census tract level.

R. Trend of Vehicle Population by Fuel Type and Year

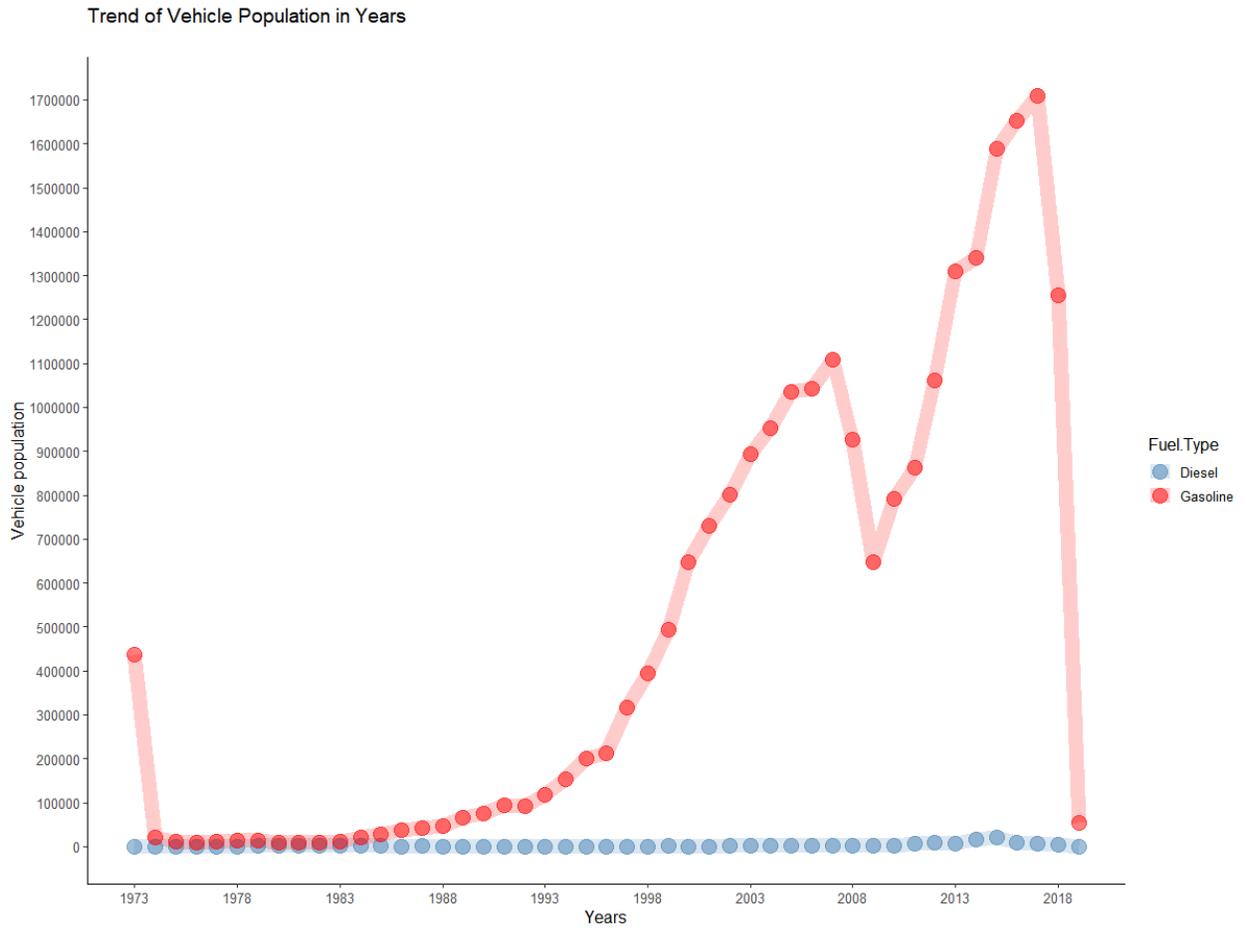


Figure R *Trend of vehicle population by fuel type and year*