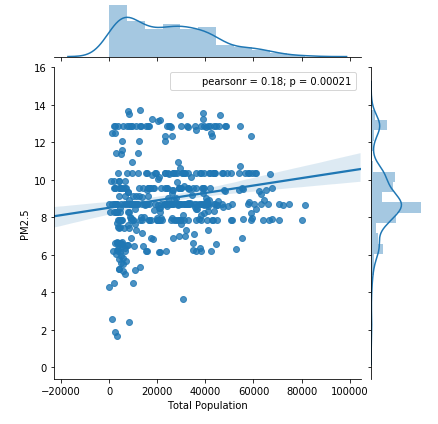
Amy Stewart April 18, 2018

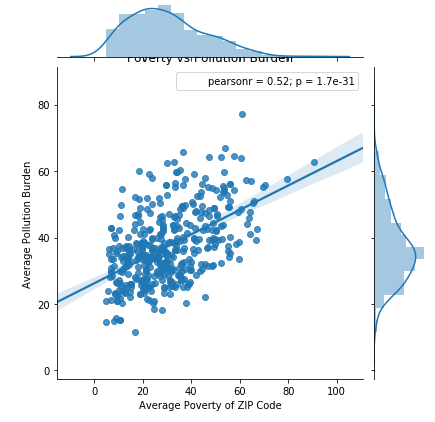
Stefani Robnett

Steven Zulim

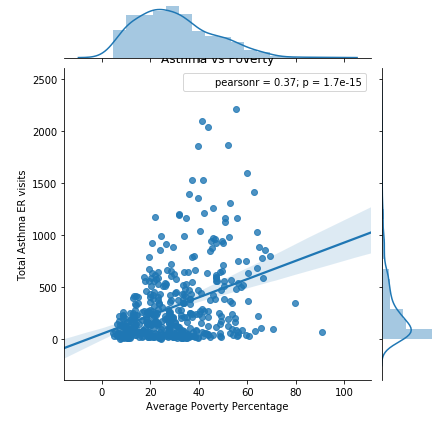
CalEnviroScreen and CALARB Analysis

**1.Is there a relationship between population characteristics and air pollution?**

* 
* In the above visualization, each point represents a zip code within our selected 22 counties. We chose zipcodes as our unit because zipcodes lend to a more granular analysis of population vs. pollution metrics. We found that though PM 2.5 concentration does tend to increase with population, there isn’t a clear linear relationship between the two. This ended up being useful further on in our analysis as it primed us to look for sources of pollution outside of urban areas.



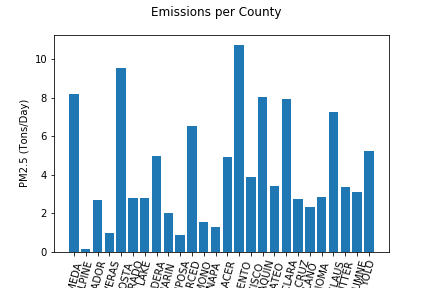
Our next visualization explored if there was an between the average poverty of a zip code (% of people living at 2x federal poverty line or less) and their pollution burden score. We found that indeed, more impoverished areas tended to suffer more from pollution in aggregate.



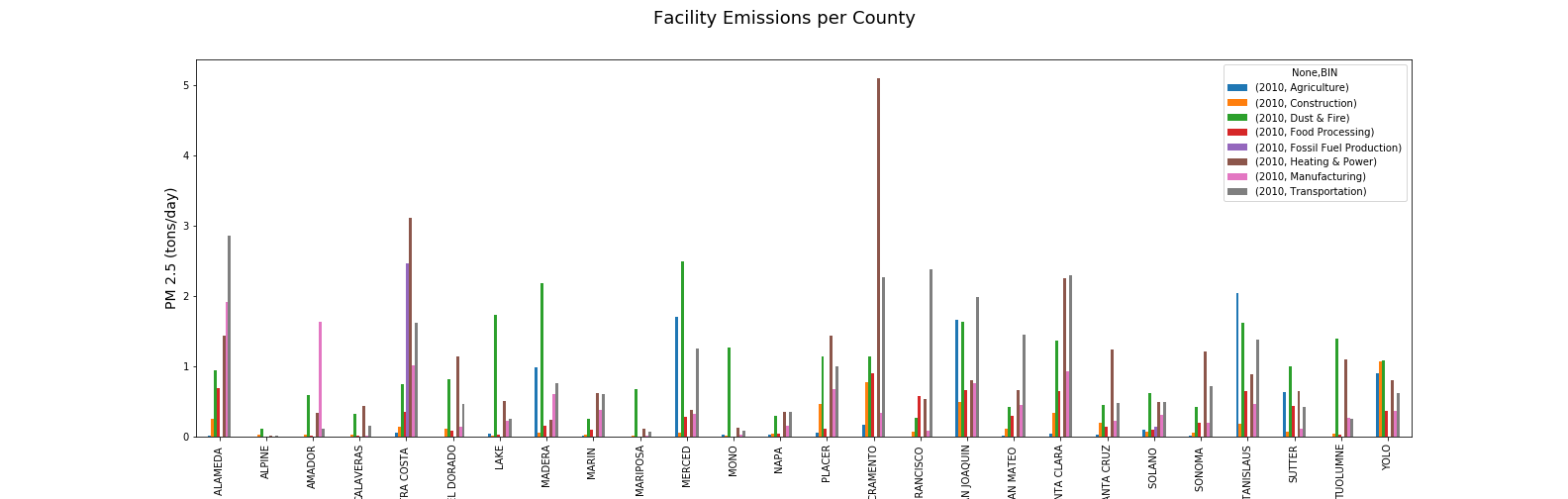
Our final visualization using CalEnviroScreen explored whether Asthma related ER visits correlated with Average Poverty percentage. (We chose to target asthma because our pollutant of interest for our next data set (PM 2.5) has particularly egregious effects on the respiratory system.) While zip codes with lower poverty had lower numbers of asthma-related ER visits, greater poverty resulted in a greater range of asthma-related ER visits. Caveat: CalEnviroScreen did not have an adequate way to control for average community distance to ER. While not a direct link to PM 2.5, higher poverty with its higher pollution burden does appear to weakly correlate with ER visits.

**2. How does air pollution vary geographically between the selected regions?**

The majority of data to answer this question came from the California Air Resources Board’s Standard Emissions Tool. We were able to download a PM 2.5 emissions report from our counties of focus (22 counties in total) and aggregate the reports to get one large csv file. From there we grouped the data by county and got the sum daily emissions for each county. We found that Sacramento had the worst emissions by far, followed by Contra Costa, Alameda, San Joaquin, and Santa Clara. See chart below:



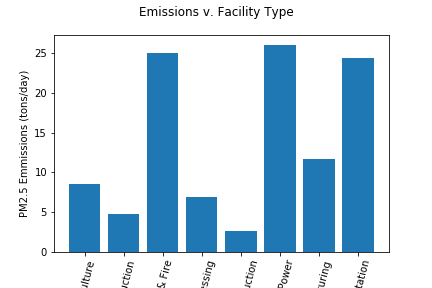
Next, in order to try and understand how emissions varied from county to county, we grouped by county and our pre-binned categories of facility type.



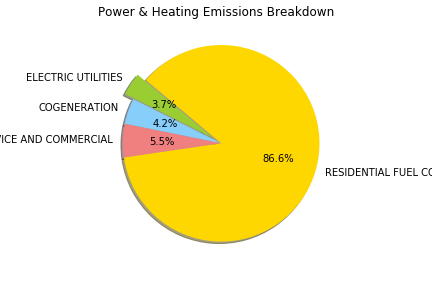
The outcome illuminated the difference between Sacramento and the rest of the counties: Power & Heating. Other notable categories included Dust & Fires, Transportation, Agriculture, Food Processing.

**3. What are the worst sources of air pollution, that could potentially have the most effective interventions?**

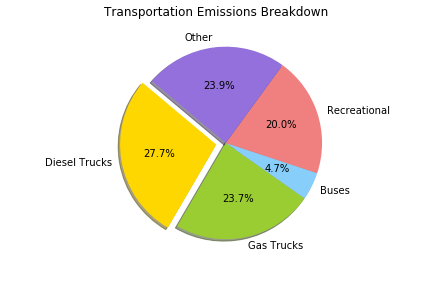
In order to get a baseline idea of how pollution varys from category to category, we grouped the data by these categories and got the sum daily output.



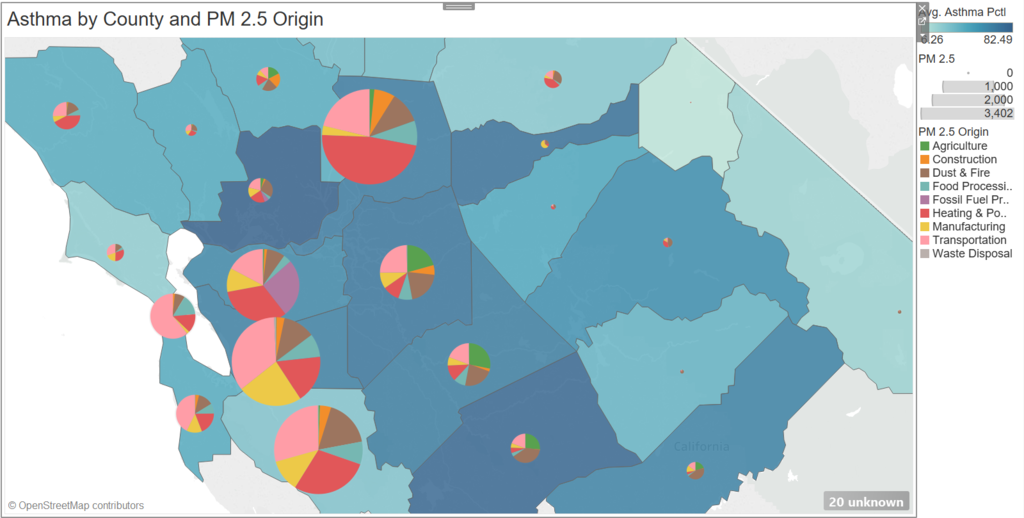
The results highlighted Power & Heating, Dust & Fires, and Transportation as the top 3 categories for daily output. Wanting to get a more granular insight into the different facilities included in these top categories, we created separate data frames with only the facilities in 2 categories: Power & Heating, and Transportation. (Dust & Fires was not chosen as a snapshot because of mostly natural causes - windblown dust, wildfires, etc. However, because dust is more likely to be prevalent in rural areas, this breakdown lends insight into why rural areas had just as much, if not more pollution than urban areas.)



We found that Residential Fuel Combustion comprises of the vast majority of Power & Heating emissions. Residential and Fuel Combustion means woodfire stoves and heaters. While these methods of energy production are generally rare in urban areas (as urban residents have access to utilities), rural residents who may be undocumented or just outside of the scope of public utilities are forced to use these much less clean methods. Therefore, without expanding utilities reach and solving basic socio-economic inequalities such as poverty, it would be difficult to target this area for effective intervention.



Our Transportations breakdown was a bit less depressing. We found that Trucks (both diesel and gas) make up for over half of the emissions in the entire transportation category. Even though other forms of transportation have a greater mean daily output (such as Urban Buses, Aircrafts, ships), it is clear that the quantity of trucks accounts for the disproportionate output. If Elon Musk succeeds in building is hyperloop, maybe transportation emissions from trucks will go by the wayside. Either way, innovation has been targeting increasing fuel efficiency, so we can reasonably expect trucking emissions to go down in the near future.



Finally, we were able to create a map of the 22 counties we selected using Tableau ([see interactive version](https://public.tableau.com/profile/steven5876#!/vizhome/Air_Quality/Dashboard1?publish=yes)). This required merging our datasets in a way that left the unique format of the facilities data from CalARB intact. We wanted to map the potential effects of pollution directly against the potential cause. For this map, shading of counties correlates with asthma cases and overall size of the pie charts correlates to overall PM 2.5 concentration.

We were curious if a specific PM 2.5 source had a stronger correlation with health effects like asthma, so we added pie charts of the different facilities categories over each county. We found no specific link between any one PM 2.5 source and asthma or other health problems, leading us to believe the causes of these health issues are far more nuanced.