Policy Analysis of California Clean Air Act on the Reduction of PM2.5 in Bay Area Counties

Web address for GitHub repository

Wanchen Xiong

Abstract

Experimental overview. This section should be no longer than 250 words.

Contents

1 Research Question and Rationale

Air pollution is an important aspects of environmental quality management. There have been numerous studies showing the adverse health effects of air pollutants, such as ozone, particulate matter, nitrogen oxide, lead and so on. In the United States, there have been progress in establishing legislations to control these air pollutants since the Clean Air Act of 1970, which established the federal framework to reduce air pollutants from stationary and mobile sources. This project chose PM2.5, ambient particulate matters with a diameter of 2.5 micrometers that have been studied extensively to be proved to have devastating effects on human lungs and cardiovascular systems. And the area of interest is bay area in northern California, as California has been progressive in implementing state air quality management programs and the bay area is both the political, industrial, and urban hub of the state. The state of California established a new annual standard for daily maximum PM2.5 of 12 ug/m3 in June, 2002, and this standard became effective in 2003. Therefore, this environmental data analysis project is investigating whether this policy intervention led to significant reduction of ambient PM2.5 in bay area over the years since it's effective. Meanwhile, California has frequent natural disaster such as wildfires annually, so this project will explore whether air monitoring captures such temporal variability of air quality and provide meaningful data support for management.

This analysis is developed into 4 research questions:

- 1. Overall reduction across the bay area: Is there a signficant reduction in PM2.5 from 1999 to 2016 in the bay area(9 counties)? Is the ambient PM2.5 concentration after the policy significantly less comparing to that before the policy was implemented?
- 2. Is each county having significant reduction in its ambient PM2.5? How does this comparison look like geographically?
- 3. Looking at one bay area county, Santa Clara County, is there a point of change since the policy change? A time series analysis was performed to look at when the data reflects a change in the trend of ambient PM2.5.
- 4. Is the air monitoring data capturing temporary events such as wildfires? Solano County had three wildfires over the year 2018 and was explored to investigate the effectiveness of the air monitoring data.

2 Dataset Information

The air monitoring data for PM2.5 were collected from here. PM2.5 data for California are available from 1999 to 2018. These data were archived by year. Each file included the daily maximum PM2.5 concentration and the date recorded were continuously from January 1 to December 31 for each year(though some days were missing in some years). All the available PM2.5 data were downloaded as individual csv files and will be combined and wrangled for the analysis of this projet.

The spatial data included the California county boundaries were from California governmental open data here and will be used for spatial analysis in this project.

Data Name	Information
EPA Air Quality PM2.5	The dataset contains data from air quality monitoring of PM2.5 in California from 1999 to 2018. All the data are in csv formats and named as pm_year_ca.
California County Boundary	The dataset contains the shapefile for all the county boundaries in California.

3 Exploratory Data Analysis and Wrangling

After pulling in all the PM2.5 monitoring data from 1999 to 2018, a data wrangling is performed to only include 9 bay area counties, and make sure date is in the right format and recognized by R. Save the processed data to the folder.

```
# wrangle this new dataframe to only include 9 bay area counties
# Make sure date is in the right format and recognized by R. Save the processed data t

bay_pm_99_18 <-
  ca_pm_99_18 %>%
  filter(COUNTY == "Alameda" |
                   COUNTY == "Contra Costa" |
                   COUNTY == "Marin" |
                   COUNTY == "Napa" |
                   COUNTY == "San Francisco" |
                   COUNTY == "San Mateo" |
                   COUNTY == "Santa Clara" |
                   COUNTY == "Solano" |
                   COUNTY == "Sonoma") %>%
  select(Date, Site.ID, Daily.Mean.PM2.5.Concentration, UNITS, DAILY_AQI_VALUE,
         COUNTY, SITE_LATITUDE, SITE_LONGITUDE)
bay_pm_99_18 $Date <- as.Date(bay_pm_99_18 $Date, format = "%m/%d/%Y")
bay_pm_99_18  <- mutate(bay_pm_99_18 ,
                        month = month(as.Date(bay_pm_99_18$Date, "%Y-%m-%d")))
bay_pm_99_18  <- mutate(bay_pm_99_18 ,
                        year = year(as.Date(bay_pm_99_18 $Date, "%Y-%m-%d")))
bay_pm_99_18 $year <- as.factor(bay_pm_99_18 $year)

# Save the processed data into the "processed" folder.
write.csv(bay_pm_99_18, file = "./Processed/bay_pm25_99to18.csv")
```

```
# Display the data's structure, dimension, columns and have a look at first/last sever
str(bay_pm_99_18)

dim(bay_pm_99_18)

colnames(bay_pm_99_18)

head(bay_pm_99_18)

tail(bay_pm_99_18)
```

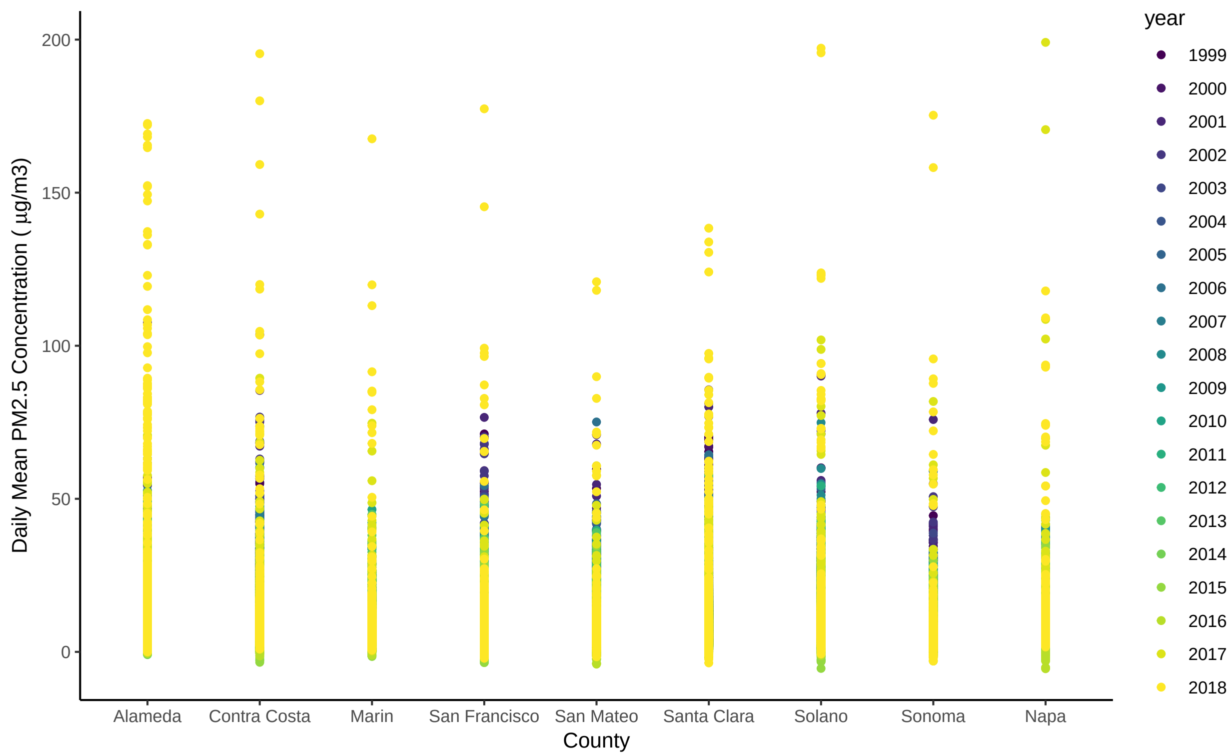

Figure 1: Data Exploration: 9 Bay Area Counties and PM2.5

After data is wrangled, a clean dataframe that includes the necessary information for the purpose of this analysis is produced. However, in order to proceed to data analysis and create meaningful synthesis of the data, initial data exploration is necessary.

Therefore, several plots were produced to look at the patterns of the concentration of PM2.5 by county and by year to identify potential interesting factors that are worth exploring in addition to the research goals.

First, figure 1 is produced to look at each of the 9 bay area counties and their respective PM2.5 concentration level over 1999-2018.

Then, figure 2 is generated to identify if there is seasonal effects that influence ambient PM2.5.

After looking at the first and second exploration plot, it looks like 2017 and 2018 have some extreme data out of the normal range. Therefore, for the purpose of this project, data from 2017 and 2018 will be excluded for the most part except when California wild fire events were considered for the last research question.

Figure 3 is generated to explore which county has the best and worst air pollution based on PM2.5 concentration.

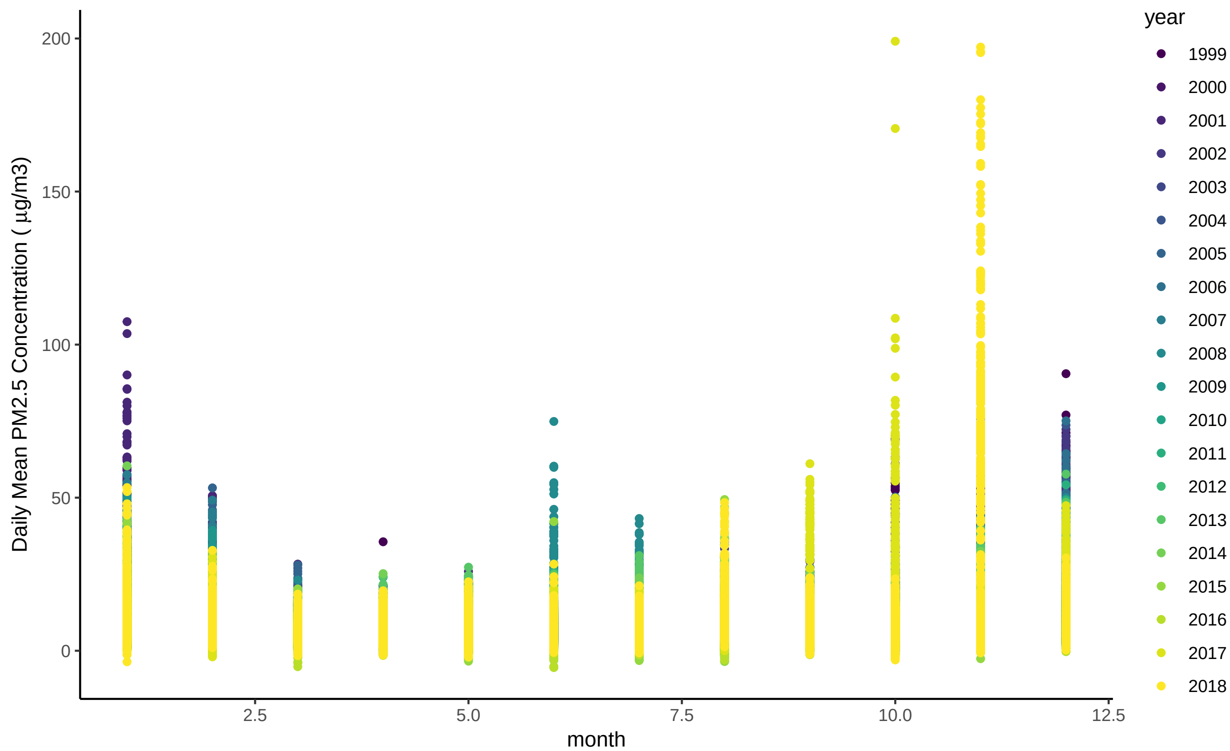

Figure 2: Data Exploration: Is There Seasonal Variability of PM2.5?

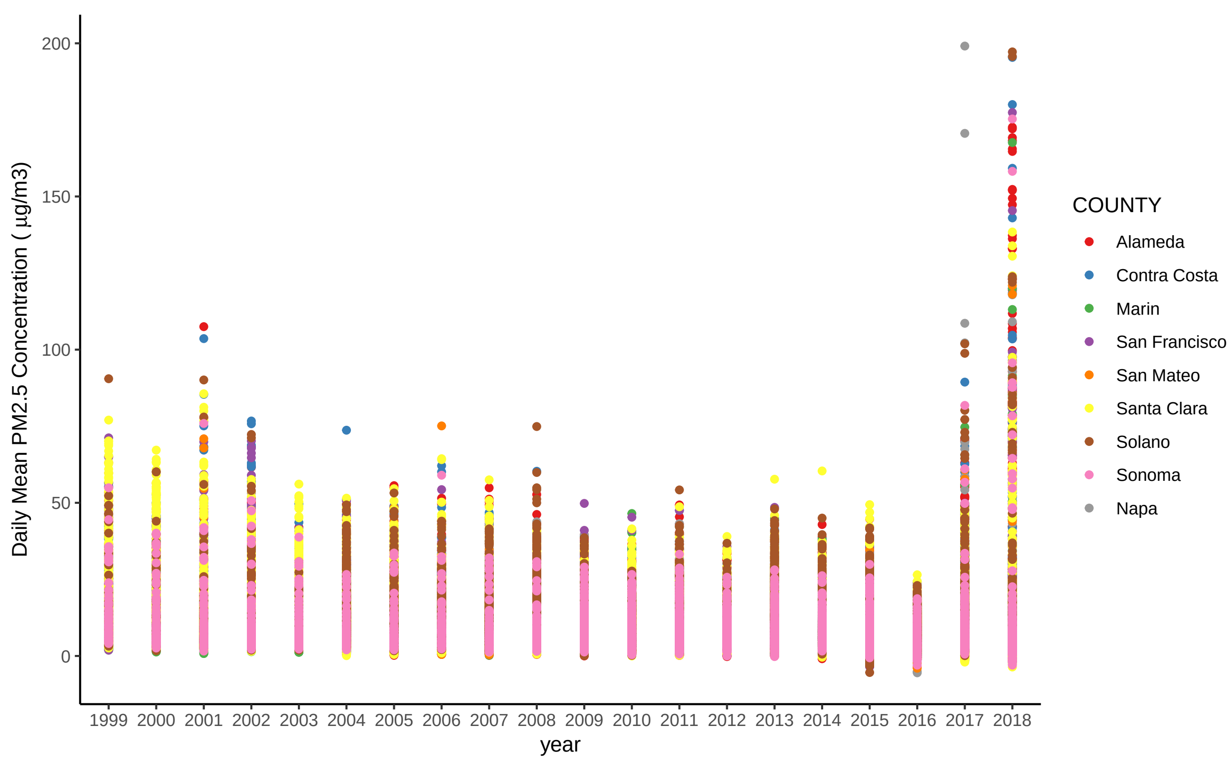

Figure 3: Data Exploration: Comparison among Counties

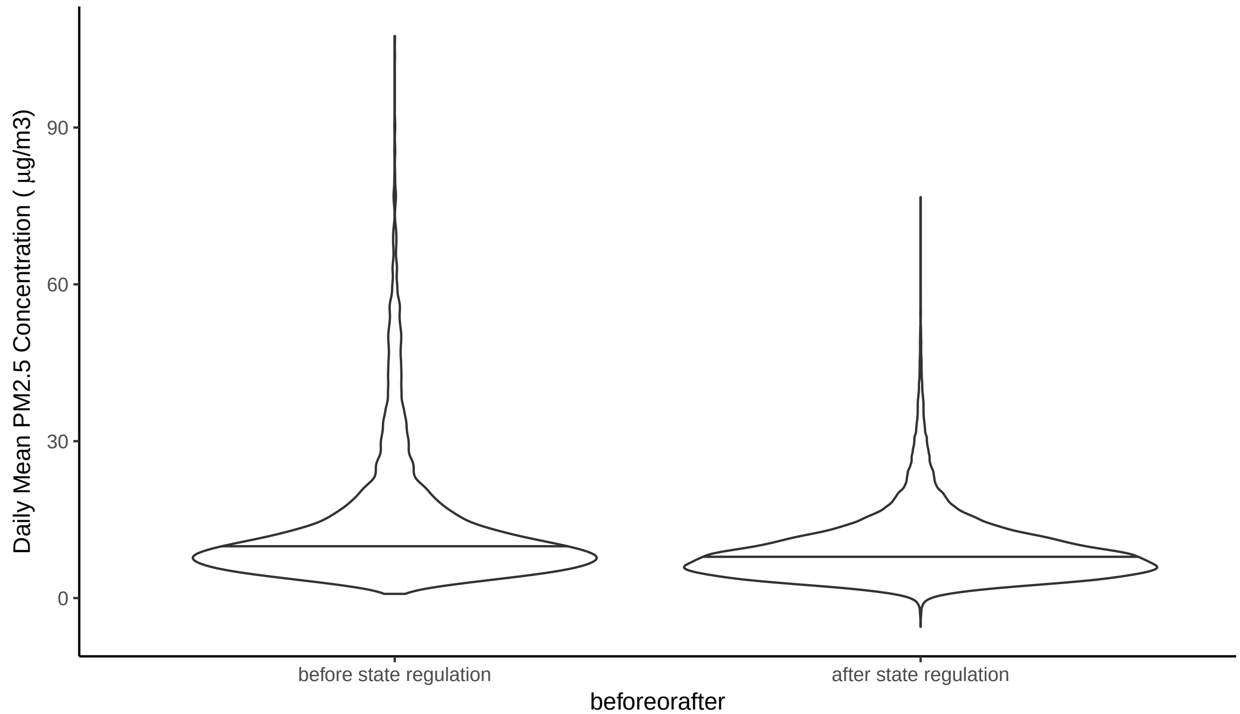

Figure 4: Difference in Mean PM2.5 before- and after-regulation

4 Analysis

4.1 Research Question 1: Is PM2.5 Reducing in Bay Area ?

Since California implemented the PM2.5 standard in 2003. The data before 2003 will be referred as "before the state regulation", whereas data after 2003 will be referred as "after the state regulation".

Perform a t test to see whether there is:

1) a significant change in PM2.5 concentration across the bay area before and after the state started regulating PM2.5;
2) a decrease in the PM2.5 concentration since the state regulation was effective.

The results of the T-test shows that there is a significant difference in PM2.5 concentration for all 9 counties as a whole between before the regulation and after the regulation; the test shows that there is a reduction of 4.35 micrograms/cubic meter for the bay area since the regulation ($p < 0.0001$, mean difference = 4.35 with 95 CI).

Figure 4 is a graphic result of the T-test.

Figure 5 displays the an overall change in PM2.5 concentration before the regulation and after the regulation.

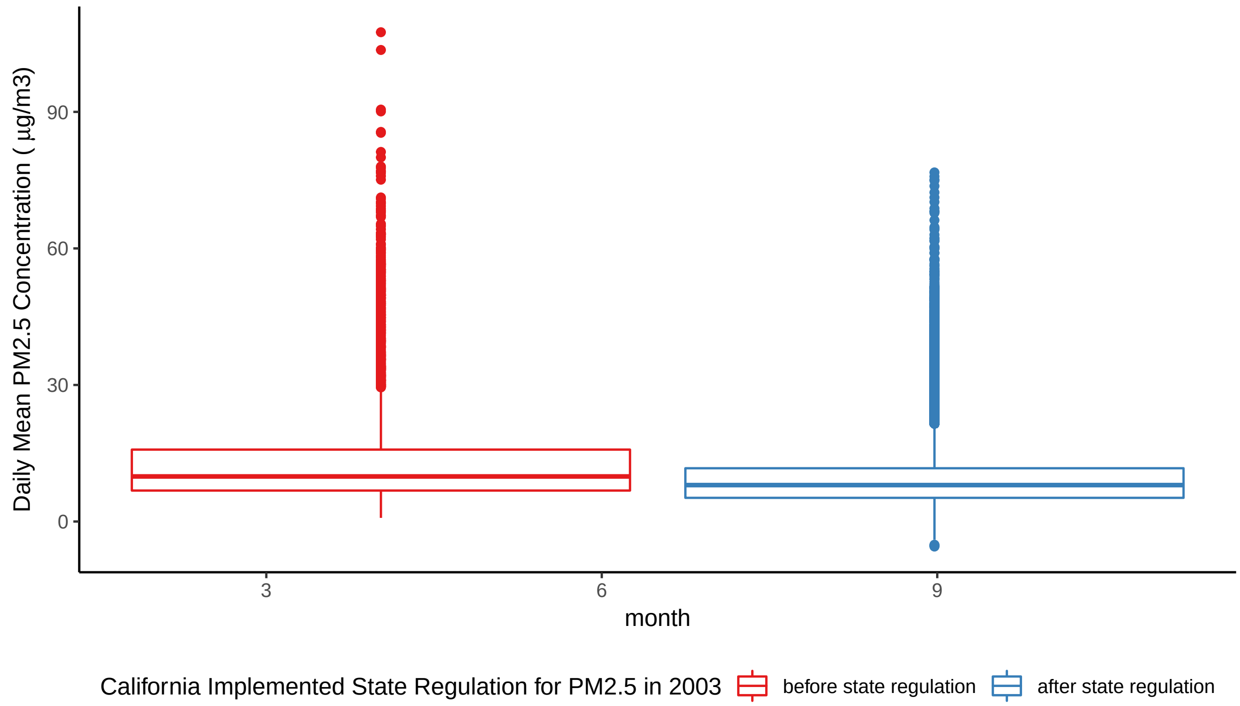

Figure 5: Mean PM2.5 Across 9 Bay Area Counties

4.2 Research Question 2: How Does Each County Experience Change in PM2.5 ?

After running the statistical test on all of the bay area counties as a whole, a signficant reduction in PM2.5 is observed. To further explore this trend, a T-test is performed to see whether every county in bay area is reducing its PM2.5. Since Napa county does not have data before the regulation, it is excluded from this analysis. The following code is a T-test performed on Alameda County.

```
# First, wrangle the data to divide Alameda PM2.5 data into before- and after-regulati
Alameda.be <- before_reg %>%
  filter(COUNTY == "Alameda")
Alameda.af <- after_reg %>%
  filter(COUNTY == "Alameda")

# Run a T-test Alameda County to see if it has reduced in PM2.5 since 2002.
t.test(Alameda.be$Daily.Mean.PM2.5.Concentration,
       Alameda.af$Daily.Mean.PM2.5.Concentration, alternative = "greater")
```

Run the same test for each of the 8 counties. A result is summarized in this table:

County Name	Results
Alameda	Significant reduction, mean differnce = -3.89 (p<0.0001)

County Name	Results
Contra Costa	Significant reduction, mean differnce = -4.30 ($p<0.0001$)
Marin	No significant change, mean difference = +1.38 ($p = 1$)
San Francisco	Singificant reduction, mean difference = -4.16 ($p<0.0001$)
San Mateo	Significant reduction, mean difference = -2.49 ($p<0.0001$)
Santa Clara	significant reduction, mean difference = -5.59 ($p<0.0001$)
Solano	Significant reduction, mean difference = -3.60 ($p<0.0001$)
Sonoma	Significant reduction, mean difference = -3.37 ($p<0.0001$)

As seen in figure 6, every county except Marin county shows a signficant reduction in PM2.5 concentration since the regulation. Even though Marin County seems to show a non-significant increase in PM2.5, Marin County is already at the lower end of PM2.5 level to begin with, comparing to the rest of the counties.

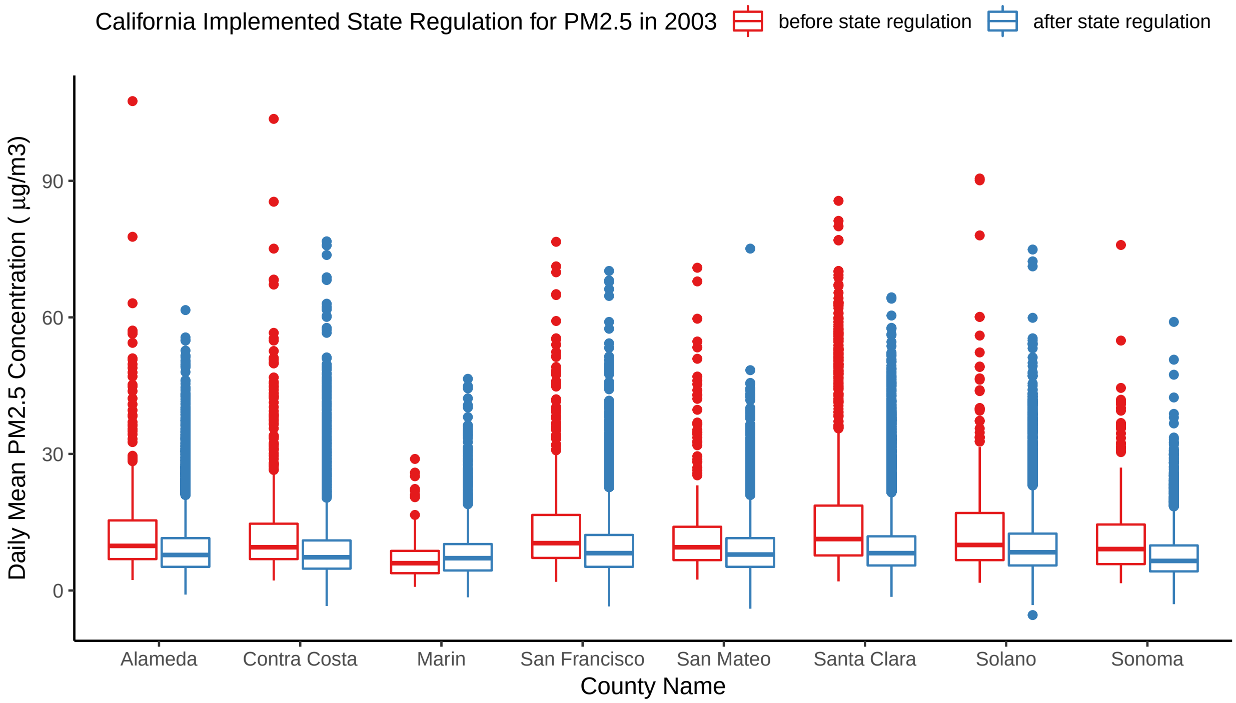

Figure 6: PM2.5 Comparison Before and After Regulation for 8 Bay Area Counties

4.3 The Spatial Display of the Results of the Regulation

Figure 7 and figure 8 shows the level of air pollution regarding PM2.5 concentration in a spatial display. The color is calibrated at the regulation standard, PM2.5 = 12 micrograms/cubic meter. The more orange the county is, the higher the PM2.5 concentration is. The more blue the county is, the lower the PM2.5 concentration is. The majority of PM2.5 high level seems to concentrate in the eastern part of the Bay Area, where it is mostly inland and urban environment.

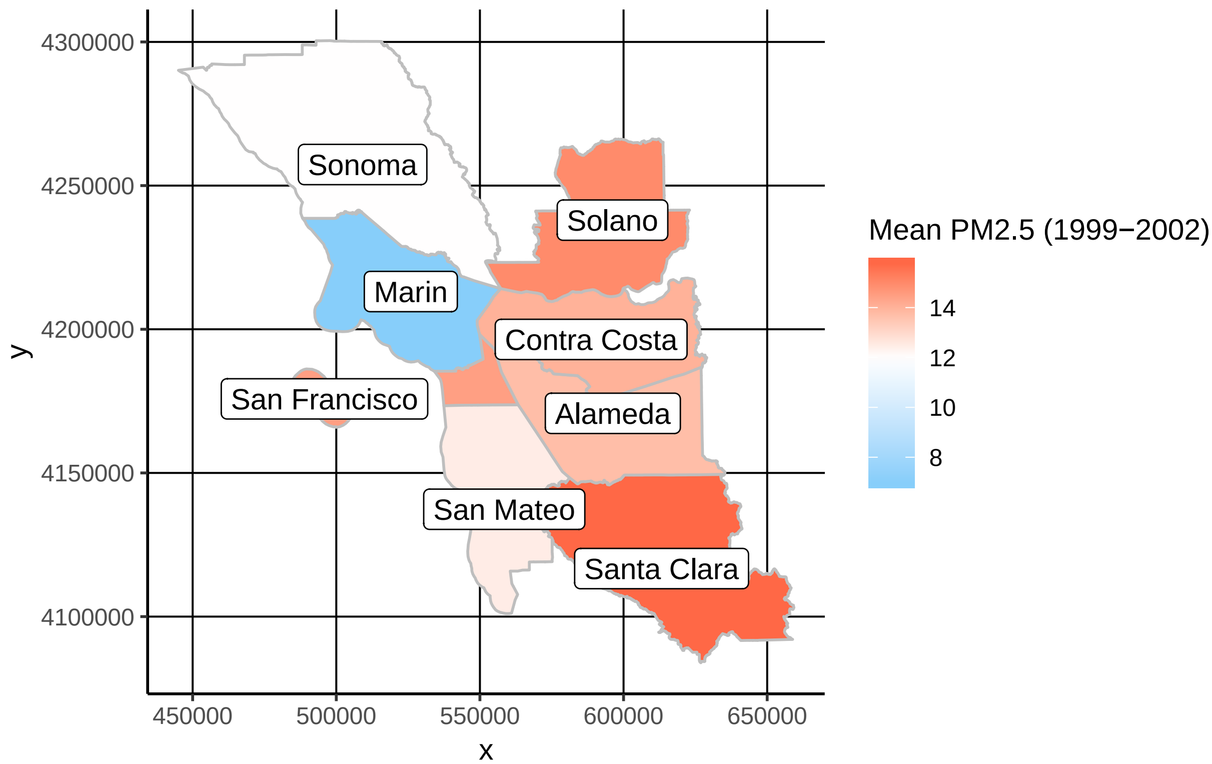

Figure 7: Mean PM2.5 in Each Bay Area County Before 2003

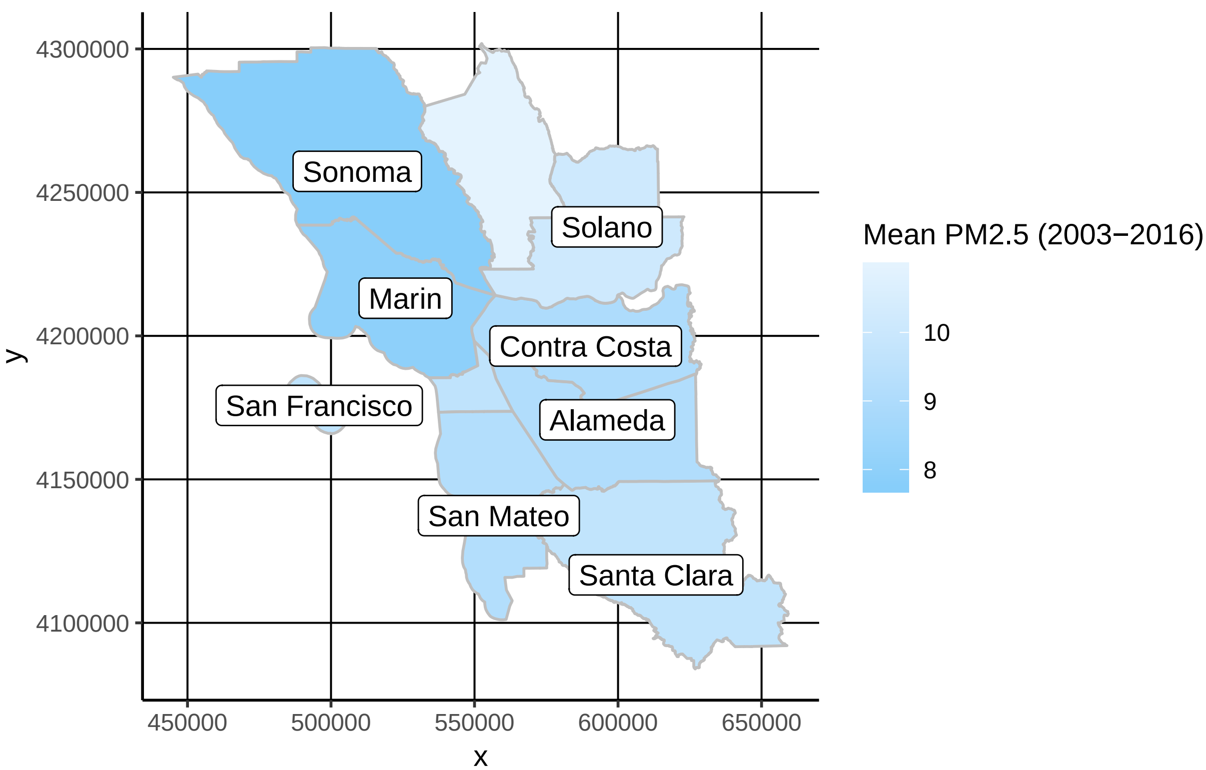

Figure 8: Mean PM2.5 in Each Bay Area County After 2003

4.4 Research Question 3: Is There a Point of Change in the Data to Reflect the Effectiveness of the Regulation? A Close Look at Santa Clara County.

First, ingle out Santa Clara County and wrangle the data to just include data for Santa Clara County and from year 1999 to 2016.

Perform a time series analysis using Mann-Kendall trend test to see whether there is a trend in the PM2.5 concentration over time.

```r
# Perform the time series analysis on Santa Clara County
mk.test(sc_pm_99_16$Daily.Mean.PM2.5.Concentration)
```

The result shows that there is a significant negative trend from 1999 to 2016 ($z = -27.29$, $p < 0.0001$).

```r
# Perform a pettitt test to detect where is the changing point in the data
pettitt.test(sc_pm_99_16$Daily.Mean.PM2.5.Concentration)
```

Pettitt's test shows that there is a point of change occurred at location 5463($P<0.0001$), which refers to Februry 4, 2009. This seems a reasonable time for the regulation to take effecst. However, more investigation of the data is needed before any conclusion can be made. Perform another Mann-Kendall trend test and pettitt test to see whether there is a trend in between the starting date to Februry 4, 2009.

```r
# Perform another pettitt test to see whether there is a trend.
mk.test(sc_pm_99_16$Daily.Mean.PM2.5.Concentration[1:5463])
pettitt.test(sc_pm_99_16$Daily.Mean.PM2.5.Concentration[1:5463])
```

The Mann-Kendall trend test shows that there is a significant negative trend leading up to Februry 4, 2009($z = -10.33$, $P<0.0001$). And the result of pettitt test shows that the point of change occurs at position 2108, which refers to Feb 1, 2003. This looks like a reasonable estimation of the time that it takes for the policy to take effects. To check that this is the first turning point, perform another round of Mann-Kendall trend test and pettitt test to validate this conclusion.

```r
# Perform another Mann-Kendall trend test and pettitt test
mk.test(sc_pm_99_16$Daily.Mean.PM2.5.Concentration[1:2108])
pettitt.test(sc_pm_99_16$Daily.Mean.PM2.5.Concentration[1:2108])
```

The result shows that the trend of ambient PM2.5 is a negative one as it approaches Feb 1, 2003. The pettitt test again to see if there is any significant point of change before Feb 1, 2003 and after 2003. The result shows that at location 350, which equals to Jan 12, 2000, is a possible point of change; however, this date falls out of range of our interest because we are only interested in the time after the regulation was implemented. Therefore, since the state regulation was effective in 2003, the first significant change of point occurs on Feb 1, 2003.

Figure 9 presents the change in the ambient PM2.5 concentration, separated by months, since 1999 to 2016.

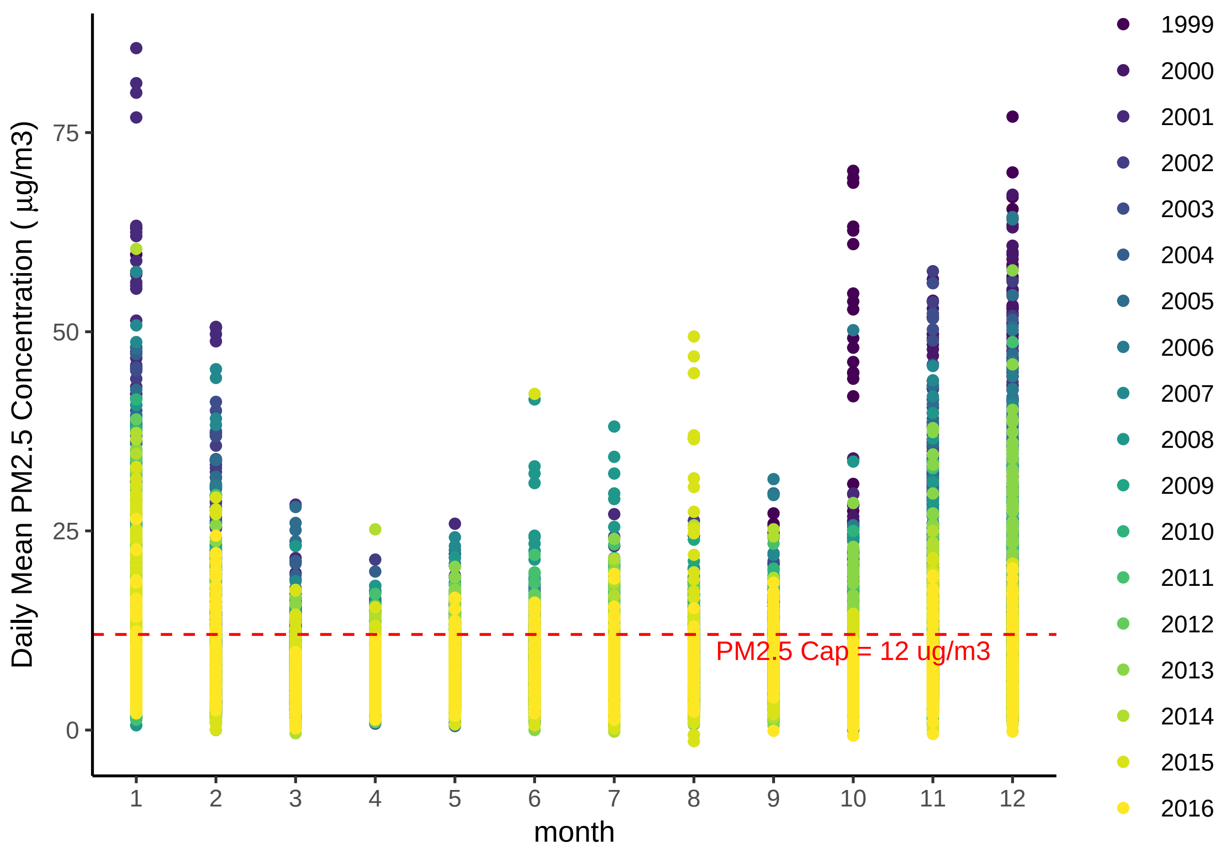

Figure 9: PM2.5 Concentration Trend from 1999 to 2016 in Santa Clara County

4.5 Research Question 4: Does EPA Air Quality Monitoring Data Capture Air Pollution Disturbance like Wildfire? A Close Look at Solano County in 2018.

EPA air quality data are used by policymakers, researchers, and environmental management professionals frequently. Therefore, it is important that these data accurately reflect what's happening in the environment. PM2.5 is a large component of the pollutants emitted from fire. This section explores the PM2.5 data for Solano County in 2018 to see if it correctly captures the wildfires events happened in this county on August 10, Oct 7, and November 8, 2018.

```r
# Perform an anova test to see if there is a significant difference between
# August, October, and November comparing to reference month, January, which
# no fire event happened.
PM2.5.anova <- lm(solano_pm_18$Daily.Mean.PM2.5.Concentration
                  ~ solano_pm_18$month)
summary(PM2.5.anova)
```

The result shows that November's mean PM2.5 concentration is significantly different from the reference month ($P < 0.0001$). Therefore, the difference between month that has wildfire and the month that does not is only reflective of November, as shown in Figure 10.

However, this test does not conclude that the data are not accurate. There are days that are missing, especially during the second fire event, which half of month's data were missing. Therefore, this might lead to the result that the expected worse air pollution is not reflected in the data.

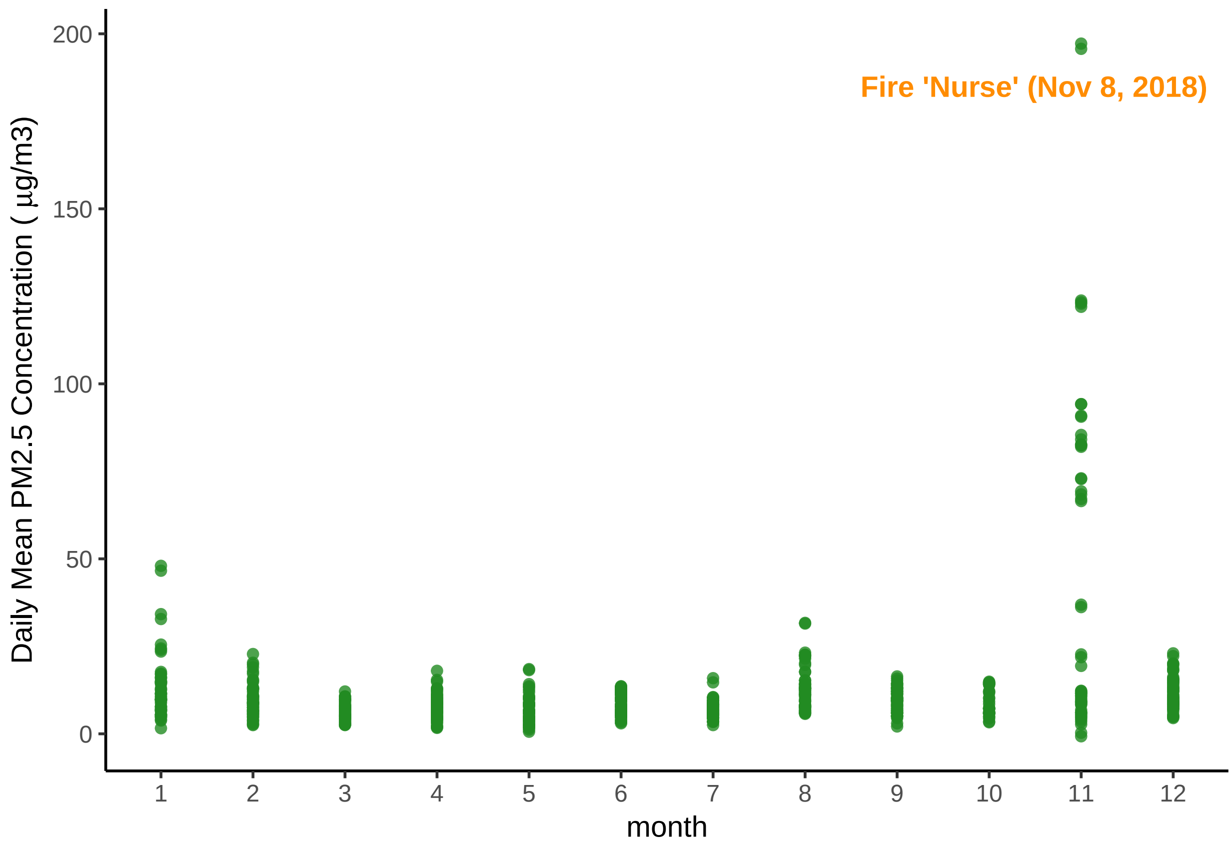

Figure 10: Overview of PM2.5 Concentration in 2018, Solano County

5 Summary and Conclusions

Policy interventions in air quality management have brought tremendous success nation-wide in the U.S. since Clean Air Act. In order to fully assess the benefits of these interventions, data analysis is necessary to make conclusions about the success of them. This research project serves to provide an in-depth policy analysis at a regional scale in Bay Area, California. This analysis covers the larger spectrum of the data by looking at 9 bay area counties as a whole. Also, individual county is compared before- and after- regulation.

As a region, Bay Area reduced its PM2.5 concentration significantly after the new regulation of daily maximum PM2.5 was implemented. Also, most counties except Marin County, by itself, has shown that PM2.5 concentration dropped significant comparing to the level before the regulation. The spatial distribution shows that eastern part of the Bay Area is where majority of PM2.5 pollution concentrates. This is significant because it can provide a spatial direction for future air quality management to focus on. When considering the time it takes for a policy to take effects, Santa Clara County shows that it takes 2 months to see a point of change; there is a negative trend in PM2.5 level following that point. Last but not least, EPA air monitoring data can capture some of the natural disturbance events such as wildfires but the data are not reflecting all of the events, partially due to lack of data during these events. This identifies an important gap in the data monitoring where it is much needed to record the air quality data during natural disturbance events.

In a nutshell, this research project concludes that since the state of California endorsed the federal regulation on PM2.5 standard, proofs from the data have shown significant reduction in PM2.5 at a regional level. However, this project acknowledge that the implementation of the state regulation standard might not be the only contribution of PM2.5 reduction. It is still worth exploring other kinds of factors that contributed to the reduction in PM2.5 across the Bay Area. Also, future air quality management effort still needs to focus on urban and inland areas as the spatial data shows that these areas tend to have higher concentration of PM2.5 and longer period time of accumulation of pollutants.