# California County Ozone Analysis (2018-2024)

Christian Solares

10/20/2025

[GitHub Repository](https://github.com/Christian-Solares/California-County-Ozone-Analysis-2018-2024-PSTAT-Solo-Project)

## 1. Executive Summary

I began this independent study to examine trends in ozone air quality for each county in California from 2018 through 2024 to quantify environmental and public-health effects using the COVID-19 lock-down period as a natural experiment. The analysis is concerned with characterizing those counties that saw dramatic shifts in ozone pollution, examining the urban-rural inequities, and describing associations between ozone exposure and unhealthy AQI days.

Using R (tidyverse, ggplot2), I combined county-level yearly AQI datasets for EPA AirData and carried out descriptive, spatial, and inferential analysis. A Welch two-sample t-test for comparing pre-COVID (2018–2019) and lockdown (2020) phases, and linear modeling for quantifying relations between ozone days and unhealthy AQI days among counties were performed. The project contains reproducible R scripts, figures, and an organized repository that complies with best practices for openness in research.

## 2. Key Findings

* Statewide ozone patterns (2018–2024)
  + Average annual ozone-dominant days: 180.36 days.
  + Counties with highest ozone exposure: Amador, El Dorado, Lake, Tuolumne, Shasta.
  + Counties showing greatest improvement post-2020: Napa, Colusa, Mendocino, Shasta, Tuolumne.

A graph of the number of countries/regions with the most frequent ozone pollution

AI-generated content may be incorrect.

A graph of a number of states

AI-generated content may be incorrect.

A graph of different colored lines

AI-generated content may be incorrect.

A graph of different colored lines

AI-generated content may be incorrect.

* COVID-19 natural experiment
  + Pre-COVID mean = 182 days, lockdown mean = 169 days, post-COVID mean = 182 days
  + Welch t-test: t = 0.91216, p = 0.3635, df = 118.32
  + Result suggests the observed 13-day decrease (7% relative reduction) during 2020 is not statistically significant at conventional levels (p = 0.3635).

A graph with red and green bars

AI-generated content may be incorrect.

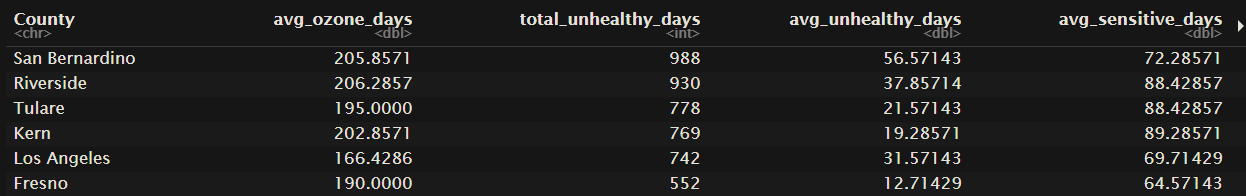
A graph of a graph showing a number of different colored bars

AI-generated content may be incorrect.

A graph with red green and blue bars

AI-generated content may be incorrect.

* Urban vs. rural differences
  + Average number of days of ozone of Urban counties (Los Angeles, San Francisco, San Diego, Santa Clara, Alameda, Orange) - Pre: 122, Lockdown: 124, Post: 135.
  + Average number of days of ozone of Rural counties (list of 18 counties used) - Pre: 207, Lockdown: 186, Post: 192.
  + Mixed group (remaining counties) - Pre: 182, Lockdown: 170, Post: 186.
  + Interpretation: rural counties show higher mean ozone-days across periods and larger absolute changes. Urban counties have much lower mean ozone-days but show modest increases post-COVID. These patterns suggest diversity in how the lockdown and subsequent years affected ozone exposure by county type.
* Health correlations
  + To find whether higher ozone exposure is correlated with poor air quality results, the correlation between average ozone-dominant days and unhealthy or sensitive-group AQI days for major California counties was examined.



* These records also initially propose that higher frequencies of ozone-dominant days do not always result in higher unhealthy air days. Riverside and San Bernardino, for example, both equally share similar ozone day averages (205 per year) but are highly distinct from each other when it comes to unhealthy day occurrences (38 versus 57 days). Likewise, Los Angeles, with fewer ozone days (166), also has a relatively higher unhealthy day occurrence count (32).
* These results suggest that ozone is a significant but not sole determinant for overall air quality health effects. Other pollutants, for example PM\_2.5, NO2, and particulates from wildfires, are responsible for large counts of unhealthy days, especially for areas that are densely inhabited or prone to fires. In addition, the relationship for ozone-dominant days is an observation for pollutant prevalence not concentration strength, so this relationship is not very robust.
* A simple linear regression of average\_ozone\_days and unhealthy\_days\_total yields an essentially flat relationship with a practically zero coefficient of determination (R² = 0.001), signifying that ozone frequency explains essentially none of the variance among unhealthy air days. That argues powerfully in support of the necessity for a stronger model involving PM\_2.5, wildfires, and weather to account for the joint impact of a number of pollutants to the public health risk.
* Frequent ozone-dominant counties can have fewer unhealthy days with moderate ozone levels, whereas other counties with fewer ozone-dominated days can have higher unhealthy-day counts due to other pollutants. This highlights the importance of multi-pollutant monitoring and comprehensive air-quality management approaches instead of using ozone as the sole public-health proxy.

A graph of air pollution

AI-generated content may be incorrect.

## 3. Technical Implementation

* Data Sources
  + EPA AirData: Annual County AQI datasets (2018–2024)
  + Variables: Median AQI, Days Ozone, Days PM\_2.5, Unhealthy Days, Primary Pollutant
  + Scope: 58 California counties
* Tools
  + R (≥ 4.3)
  + tidyverse (for data wrangling)
  + ggplot2 (for visualization)
  + Environment locked using renv for reproducibility
* Methods
  + Data cleaning: merged annual CSVs to filtered State == “California”
  + Aggregations: county-level means, medians, and totals
  + Temporal analysis: trends plotted by year
  + Hypothesis testing: two-sample t-tests (pre- vs during COVID)
  + Regression modeling: linear fit between ozone days and unhealthy days

## 4. Limitations & Future Work

### Limitations and Future Work

### Uneven monitoring of rural counties and discrepancy of mixed (rural and urban) counties.

### COVID impact affected by wildfire events.

### The t-test (p = 0.3635) may not be meaningful even with a 13-day average decline. County variance is too large to indicate clear conclusions.

## 5. Conclusion & Public-Health Implications

These statewide average ozone-days (180.4) and county deviations suggest that ozone loading in the state is highly variable, with some rural counties (e.g., Amador, El Dorado, Lake, Tuolumne) experiencing very high annual average ozone-day counts (300+ days yr.) and the urban areas experiencing lower averages with varying AQI patterns (urban Median.AQI higher at some times).

COVID-lockdown year (2020) shows a non-significant mean reduction in ozone-days (~13 days, ∼7% decrease, p = 0.3635). It suggests that, though at mean, some reduction did occur, but not strong given county variability. A stronger causal claim requires meteorological and wildfires controls.

Counties with higher avg\_ozone\_days tend to have more unhealthy-day burden (group summaries show substantially different avg\_bad\_days by ozone level). This supports prioritizing monitoring and intervention where ozone-days are persistently high.