# Impacts of Tree Canopy Cover Percentage on the Rate of COVID Deaths per Positive Case in California

Clarissa Boyajian

12/1/2021

## Contents

## What impact does the percentage of tree canopy cover have on the rate of COVID deaths per confirmed positive case in California?

This blog post will conduct analysis at a county level within California to look at the impact of tree canopy cover percentages on the likelihood of dying after COVID-19 is contracted.

### Importance

This is an important and interesting environmental justice question to consider because tree canopy cover percentages and COVID-19 disproportionately impacted historically marginalized populations throughout the United States. Additionally, both are linked to respiratory and cardiovascular diseases. There is proven evidence that areas with more urban tree canopy have increased public health indicators (6). This includes lower rates of disease such as asthma, strokes, and cardiac disease. There is also evidence that individuals with existing respiratory and/or cardiovascular disease are not only more likely to contract COVID-19, but also more likely to be sicker or even die (8).

## Data

**Tree Canopy Data**  The tree canopy data used for this analysis is publicly available from the Public Health Alliance of Southern California (2), which reports California Healthy Places Indexes. The data is available in CSV format, with a canopy cover percentage for each census tract within California. The website was last updated in April of 2021, but the tree canopy data is from 2011.

**COVID-19 Data**  The COVID-19 data used in this analysis was publicly available on the LA Times DataDesk GitHub repository (1). This data was collected using scrapers written in Python and Jupyter notebooks, scheduled and run via GitHub Actions, and archived using git. The scrapers collection data from the California Department of Public Health and other government agencies. This data is at a county-level spatial resolution and includes a daily number for both confirmed cases and deaths from February 1st, 2020 to today. The data used in this analysis included daily numbers from February 1st, 2020 through November 22, 2021. Potential bias in this data, is that confirmed cases are calculated based on positive test results. This means that any individuals who contracted COVID-19, but did not get tested, are not included in this data. Due to the high rate of asymptomatic cases (9), there is likely a large quantity of missing data.

**Geographic Data**  The geographic data used in this analysis includes California county borders and U.S. Census regions, which subsets the state into 10 different regions. The county geographies were downloaded as a ShapeFile from the LA Times DataDesk GitHub repository (3). The U.S. Census regions were manually entered into R based on a map publicly available on the U.S. Census website (5).

## Basic Data Analysis

For my analysis I planned to conduct a simple OLS linear regression, but first conducted some basic analysis to explore the data. To begin, I needed to transform the tree canopy and COVID-19 data to be at the same spatial and temporal resolution. First, I calculated the tree canopy cover percentage for each county using the `group_by()` and `summarize()` functions to create an average from the census tract data. Next, I calculated each county's average population, average daily number of confirmed positive cases, and average daily reported deaths also using `group_by()` and `summarize()` functions. Finally, I calculated a rate of deaths per confirmed case and per capita for each county. Once this was completed, I combined all datasets based on the county Federal Information Processing System (FIPS) codes to create a dataframe including tree canopy, COVID-19, and income data as well as geometries for each county.

## Basic Data Visualization

Before deciding to use a simple OLS linear regression, I wanted to conduct some basic data visualization to explore the correlation between tree canopy and the rate of COVID-19 deaths per positive case.

First, I aggregated the county data further into the 10 county regions as defined by the U.S. Census and plotted the canopy cover percentages and COVID-19 deaths per capita for each region (Fig 1). This exploration showed a potential relationship between lower canopied regions within California and COVID-19 deaths per capita.

Next, I honed my exploration more closely in on my research questions: what impact does tree canopy coverage have on the likelihood that someone who contracts COVID-19 will die? To do this I created two maps. One map shows the average tree canopy cover percentage for each county within California (Fig 2). The other map displays the rate of average daily COVID-19 deaths per average daily confirmed positive cases for each county within California (Fig 2). As with my previous exploration this visualization does not show anything conclusively, but does indicate that there is a relationship between California counties with lower tree canopy cover and higher likelihood of death for individuals who contract COVID-19.
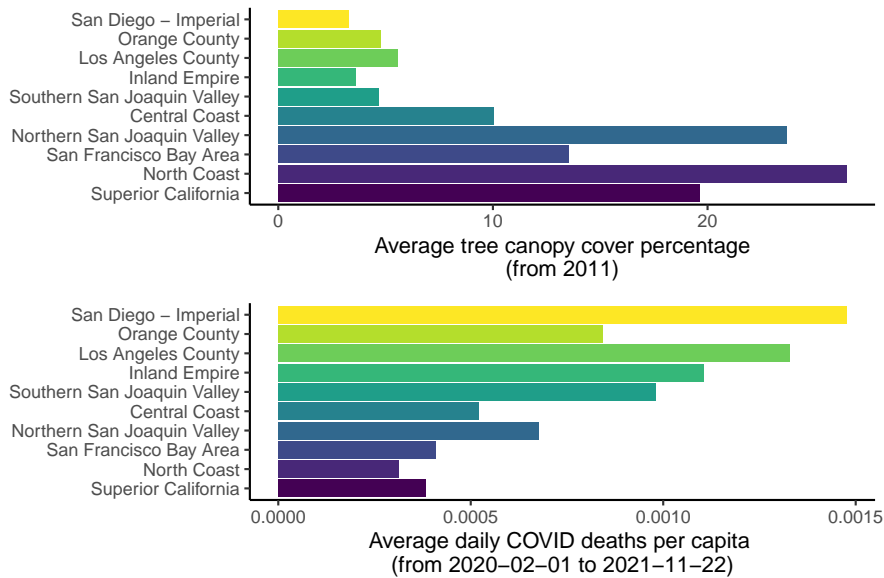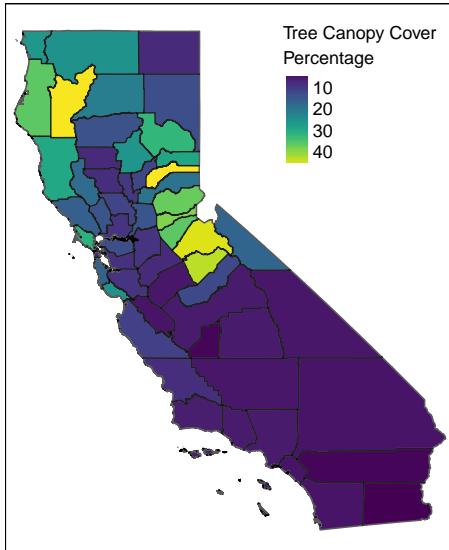
Figure 1: Plot showing the average tree canopy cover percentage and average daily COVID-19 death per capita for each of California's 10 Census Regions.
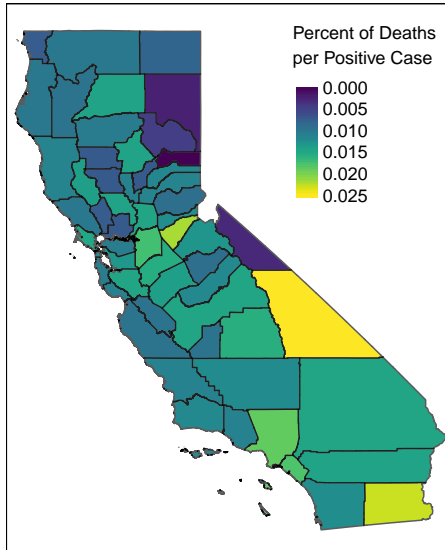


Figure 2: Maps showing the average tree canopy cover percentage and average daily COVID-19 death per capita for each of the 57 California Counties. (Note: the tree canopy cover data used did not include Alpine County, so it was excluded from all analysis.)

## Simple Linear Regression

I conducted a simple OLS linear regression to look at the impacts of tree canopy cover percentages on the rate of COVID-19 deaths per positive case. First, I used `ggplot` to create a scatter plot comparing the tree canopy cover percentage to the rate of COVID-19 deaths per positive case for each of the 57 California counties. Then I used `geom_smooth()` to plot a simple OLS linear regression of the data. Visually, it appears there is a negative correlation between the two rates (Fig 3).
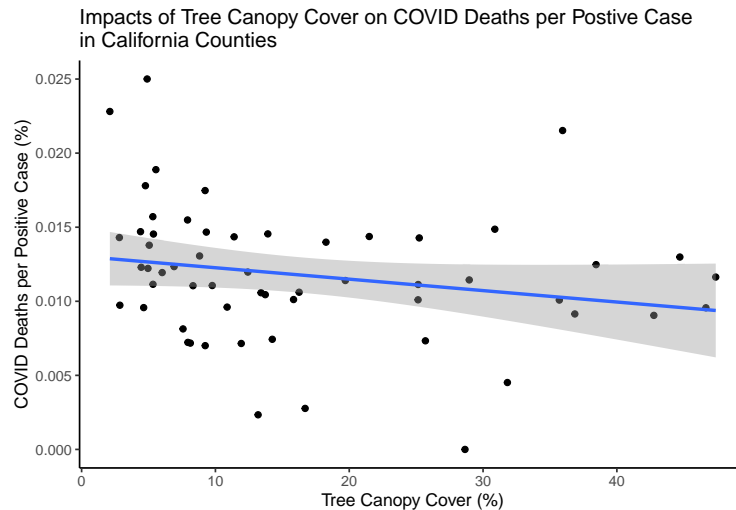


Figure 3: Correlation plot of tree canopy percentage compared vs. rate of COVID-19 death per positive case for 57 California Counties, including simple linear regression line.

The results of the OLS regression are shown in figure 4. The results show that there is in fact a negative correlation between tree canopy cover percentage and the rate of COVID-19 deaths per positive cases within California counties. The slope results indicates that for each 1% increase in tree canopy there is a 0.008% decrease in the COVID-19 death rate per positive case. However, the p-value is 0.11 meaning that there is no statistical significance that can be taken from this analysis. Additionally, the R-squared value is 0.05, meaning that only 5% of the variation in the rate of COVID-19 deaths per positive case are explained by average tree canopy cover percentages.

Table 1: Simple Linear Regression Results: Impact of tree canopy cover on rate of COVID-19 deaths per positive case in California

| Intercept | Slope | P-value | R-Squared |
|---|---|---|---|
| 0.0130365 | -0.0000772 | 0.107893 | 0.0463171 |

**Hypothesis test**  Because of the lack of statistical significance discussed above, we are unable to reject the null hypothesis with the current analysis.

*Null hypothesis:* In California counties, the tree canopy cover percentage has no impact on the rate of COVID-19 deaths per positive reported case.
*Alternative hypothesis:* In California counties, the tree canopy cover percentage has an impact on the rate of COVID-19 deaths per positive reported case.

**Confidence Interval**   I calculated a confidence interval and found that I was 95% confident that the true change in COVID-19 deaths per positive case for each 1% increase in tree canopy cover percentages was withing the range of (-0.000172, 0.000017).

## Conclusions and Future Analysis

As discussed above, there is no statistically significant conclusion to be taken from this analysis. However, I believe that future analysis is warranted.

**Full United States**   The analysis that I conducted only included 57 observations, one for each California county included in the original tree data. This is a low number of observations, so there may be a different result if analysis were conducted looking at all 3,006 counties within the United States.

**Median Income**   Other future analysis would be to include median income in the regression model. Because those with higher median incomes generally live in communities with higher tree canopy cover percentages (7), it is possible that the negative correlation we saw was more not due to tree canopy. For future analysis I would use median income data from the United States Department of Agriculture (USDA) Economic Research Service website (4), which has 2019 median income data is aggregated to a county level.

**More Recent Tree Data**   Lastly, the data used in this analysis was from 2011. While tree canopy cover tends to change slowly, there have been technological advances in the last decade (such as LiDAR) that allow for more accurate tree canopy cover percentage estimates.

## References

**Data**

1. LA Times DataDesk, California Coronavirus Scrapers GitHub repository: https://github.com/datadesk/california-coronavirus-scrapers
2. Public Health Alliance of Southern California's California Healthy Places Index Report: https://healthyplacesindex.org/data-reports/
3. LA Times DataDesk, Geographic Boundaries GitHub repository: https://github.com/datadesk/boundaries.latimes.com/blob/master/shapefiles/counties/2012/counties.prj
4. United States Department of Agriculture Economic Research Service County-level Data Sets: https://www.ers.usda.gov/data-products/county-level-data-sets/
5. United States Census, California Complete Count Office: https://census.ca.gov/regions/

**Literature**

6. Astell-Burt, Thomas, and Xiaoqi Feng. "Urban Green Space, Tree Canopy, and Prevention of Heart Disease, Hypertension, and Diabetes: A Longitudinal Study." The Lancet Planetary Health 3 (September 1, 2019): S16. https://doi.org/10.1016/S2542-5196(19)30159-7.
7. Schwarz, Kirsten, Michail Fragkias, Christopher G. Boone, Weiqi Zhou, Melissa McHale, J. Morgan Grove, Jarlath O'Neil-Dunne, et al. "Trees Grow on Money: Urban Tree Canopy Cover and Environmental Justice." PLOS ONE 10, no. 4 (April 1, 2015): e0122051. https://doi.org/10.1371/journal.pone.0122051.
8. Yang, Jing, Ya Zheng, Xi Gou, Ke Pu, Zhaofeng Chen, Qinghong Guo, Rui Ji, Haojia Wang, Yuping Wang, and Yongning Zhou. "Prevalence of Comorbidities and Its Effects in Patients Infected with SARS-CoV-2: A Systematic Review and Meta-Analysis." International Journal of Infectious Diseases 94 (May 1, 2020): 91–95. https://doi.org/10.1016/j.ijid.2020.03.017.

9. Zhao, Hongjun, Xiaoxiao Lu, Yibin Deng, Yujin Tang, and Jiachun Lu. "COVID-19: Asymptomatic Carrier Transmission Is an Underestimated Problem." Epidemiology & Infection 148 (ed 2020). https://doi.org/10.1017/S0950268820001235.