

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

Clarissa Boyajian

12/1/2021

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 analyze the impact of tree canopy cover percentages on the likelihood of dying if COVID-19 is contracted at a county level throughout California.

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 (CITE) that areas with more urban tree canopy have increased public health indicators. This includes lower rates of disease such as asthma, strokes, and cardiac disease (CITE nature conservancy?). There is also evidence (CITE) 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.

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.

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).

Income Data The median income data used in this analysis was downloaded from the United States Department of Agriculture (USDA) Economic Research Service website (4). The median income data is aggregated to a county level and is from 2019, with the website last updated in June 2021.

Basic Analysis

For my analysis I planned to conduct a simple linear regression, but first conducted some basic analysis to explore the data. To begin, I did some basic data analysis to create tree canopy and COVID-19 data 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. Then 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 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 correlation between lower canopied regions within California and COVID-19 deaths per capita.

California County Regions

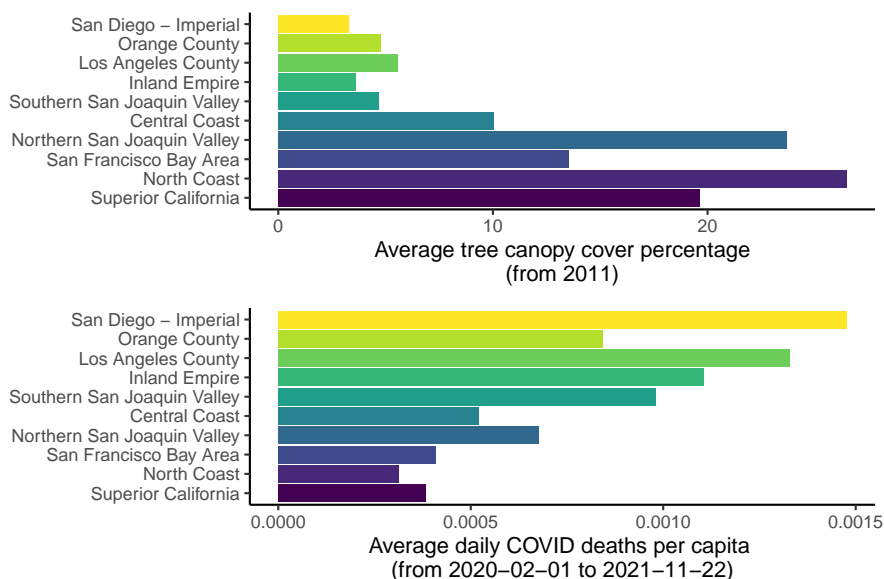


Figure 1: Plot showing the average tree canopy cover percentage and average daily COVID-19 death per capita for each of California's 10 Census Regions.

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).

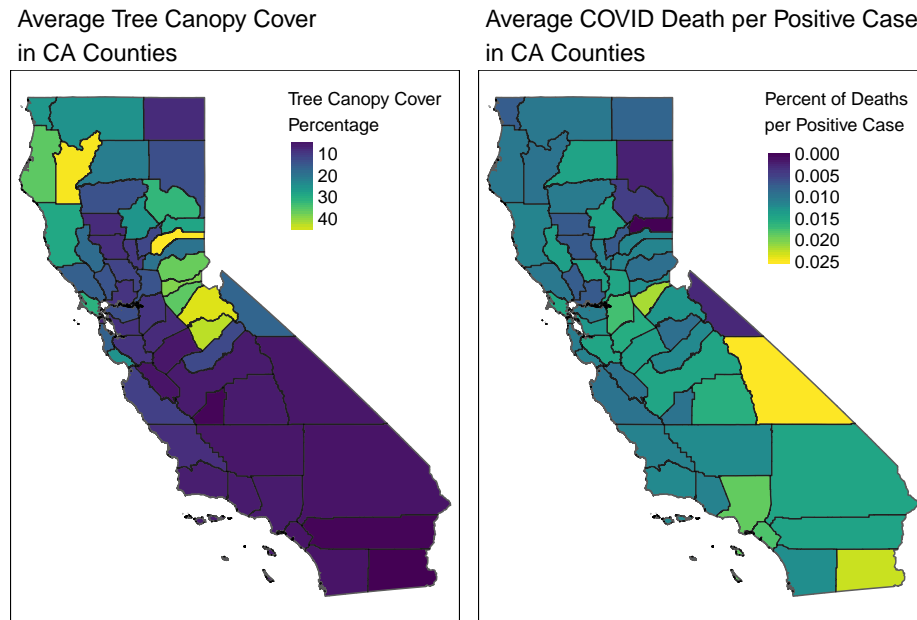


Figure 2: Maps showing the average tree canopy cover percentage and average daily COVID-19 death per capita for each of the 57 California Counties. (Note: the tree canopy cover data used did not include Alpine County, so it was excluded from all analysis.)

Simple Linear Regression

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

Alternative hypothesis: In California counties, the tree canopy cover percentage has an impact on the rate of COVID deaths per positive reported case.

Conclusions and Future Analysis

There is no

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>

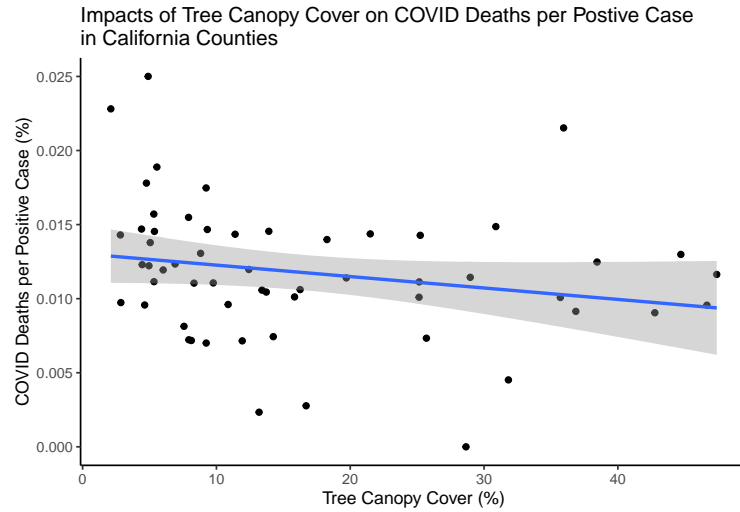


Figure 3: Correlation plot of tree canopy percentage compared vs. rate of COVID-19 death per positive case for 57 California Counties, including simple linear regression line.

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.