

Covid on the Border

Raymond H. Levine*
Westview High School
San Diego, CA 92129, USA

March 30, 2024

Abstract

The goal of this research was to study the effect of demographic and economic data on Covid case rates in San Diego County. Data were obtained from the San Diego Association of Governments (SANDAG) and the statistical software R was used for the analysis. The results found that Covid case rates varied greatly by zip codes. In particular, regions with more college educated people had lower Covid case rates, while regions with higher poverty rate had higher Covid case rates. In addition, zip codes with higher Covid case rates tended to have more Hispanics and Blacks.

Keywords: Covid pandemic impact, statistical visualizations, education level, poverty, ethnicity, regression smoothing, bar chart, choropleth map

1 Introduction

The Covid pandemic started in the Spring of 2020 in the United States. It altered so much of our daily life and had a profound impact on our society. Have you ever wondered whether the toll of Covid was even across different communities?

This research focuses on Covid case rates by zip code in San Diego County. San Diego County is located in the Southwestern corner of the United States. As San Diego is bordered by the Pacific Ocean to the West and Mexico to the South, it encompasses a wide range of socio-economic levels and ethnic groups.

The questions we aimed to answer were whether Covid affected regions in San Diego County by demographic and economic variables. In particular, do zip codes have different Covid case rates because of different levels and compositions in education, poverty, and ethnicity?

*raymlb8@gmail.com

2 Data

The data used in this paper were from San Diego Association of Governments (SANDAG) [2]. The first data set contains daily Covid case counts, between 3/30/2020 and 12/28/2021, for each of the 113 zip codes in San Diego County. The second data set contains demographic data, also organized by zip codes. We merge the two data sets together before performing statistical analysis. In addition, we also standardize case counts to the number of cases per 100,000 people so that the comparisons across zip codes are fair.

3 Methods

The following statistical visualizations were performed to answer the questions of interest:

- Choropleth maps of San Diego County zip codes were created using case rates and demographic variables, education and poverty.
- Time series graphs were created to display daily case rate over time, while demographic data were split into high and low groups where high is the upper quartile and low is the lower quartile among zip codes.
- Scatterplots were created with regression lines and scatterplot smooths to study the relationship between case rates per zip code and demographic variables.
- Stacked bar charts were created for the categorical variable, ethnicity. The bar charts were grouped into four zip code categories ordered by case rate quartiles.

All the data summaries and statistical analyses were done in R [3]. The R packages used include tidyverse [5], ggplot [4], choroplethr, and choroplethrZip [1]. A description of choropleth map may be found at en.wikipedia.org/wiki/Choropleth_map.

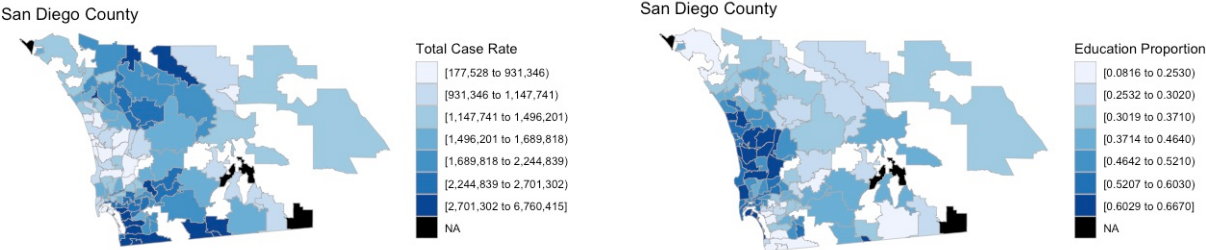


Figure 1: Left: Map of total Covid case rate per 100,000 people between 3/30/2020 and 12/28/2021 in San Diego County by zip codes. Right: Map of the proportion of population with a college degree in San Diego County by zip codes.

4 Results

The left panel of Figure 1 shows a map of the total Covid case rate, between 3/30/20 and 12/28/21, in San Diego County. The rates have been standardized to the number of total cases per 100,000 people. The lighter colors on the map show zip codes with lower total case rates, while darker colors on the map show zip codes with higher total case rates. The map indicates that higher case rates are found close to the Mexican border (south or zip codes to the bottom) and more inland (away from ocean or zip codes to the right). Note that the border and inland areas are poorer and have higher rates of minorities. The right panel of Figure 1 presents a map of education levels across zip codes in San Diego County. Education levels are defined as the proportion of the population that are college educated in each zip code. The darker colors indicate higher proportions of the population with a college education. We see that the coastal areas (west or left side of the map) have the highest education levels.

Figure 2 shows the relationship between education and Covid case rates. The left panel shows the cumulative case rates from 3/3/2020 to 12/28/2021, for zip codes with the highest and lowest quartiles in % of the population with college degrees. We see that the Covid case rates are much lower in the most educated zip codes (blue line) than in the least educated zip codes (black line). The right panel of Figure 2 shows the scatterplot of total case rate versus the proportion of the population with a college degree. The pink line is the linear regression (ordinary least squares) fit, and the blue line is a locally weighted scatterplot smoother (LOWESS), with 95% confidence bands (shaded areas). We see that, as the proportion of the population with college degrees increases, the total case rate decreases ($p = 0.00038$).

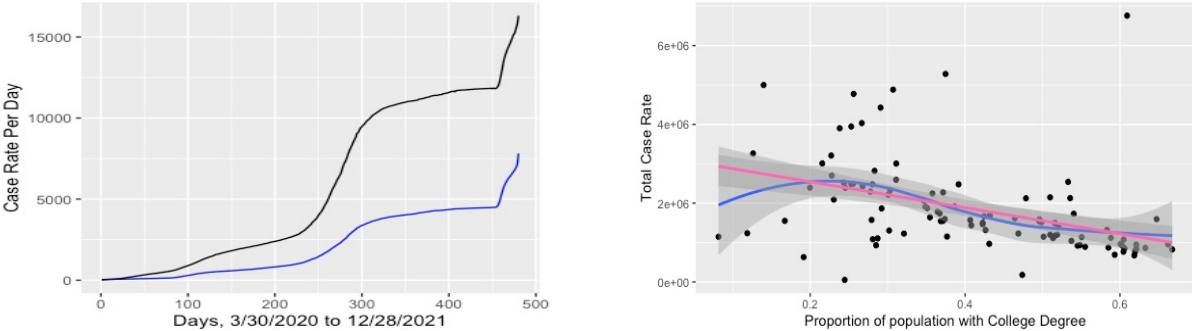


Figure 2: Left: Comparison of cumulative Covid case rates by education levels, grouping zip codes into two categories by percentage of the population with a college degree (black the higher percentage and blue the lower percentage). Right: Plot of total case rate against proportion of the population with a college degree overlay by a linear regression fit (pink line), scatterplot smooth (blue curve), and 95% confidence bands over each (grey shaded regions).

Next we study the relationship between poverty and Covid case rates. Figure 3 shows a map of poverty levels, defined as the proportion of the population below the poverty level, across zip codes in San Diego County. The darker colors indicate poorer areas. The color

scheme in this map is almost the opposite of the right panel of Figure 1, with border regions (south/bottom) and Eastern regions of the county (right side of the map) having higher proportions of the population below the poverty level.

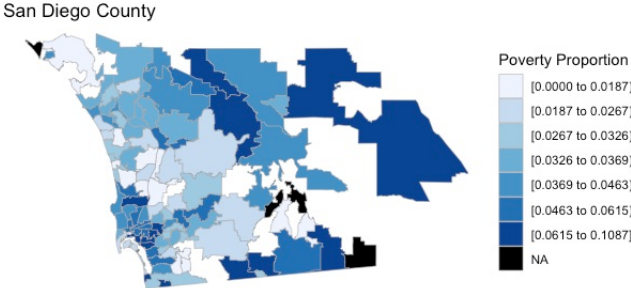


Figure 3: Map of the proportion of the population below the poverty level in San Diego County by zip codes.

Figure 4 is similar to Figure 2, but based on poverty levels. We see that the Covid case rates are higher in poorer zip codes ($p = 0.064$).

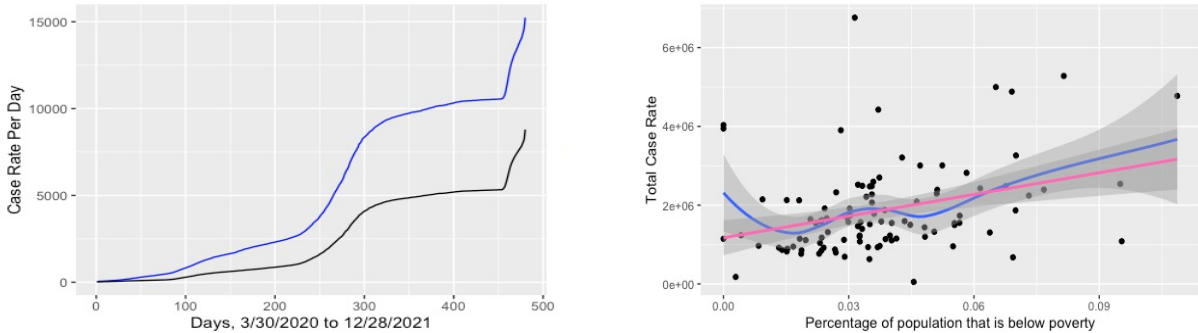


Figure 4: Left: Comparison of cumulative Covid case rates by poverty levels, grouping zip codes into two categories (black higher poverty rate and blue lower poverty rate). Right: Plot of total case rate against the percentage of the population below the poverty level overlay by a linear regression fit (pink line), scatterplot smooth (blue curve), and 95% confidence bands over each (grey shaded regions).

Figure 5 shows the relationship between ethnicity and Covid case rates. The left panel shows stacked bar charts of ethnicity, depicted separately for the four case rate quartiles, with “1” indicating the lowest case rate quartile and “4” indicating the highest case rate quartile. We see from this figure that, in the zip codes with higher Covid case rates, there are higher proportions of the population that are Hispanic and Black, and lower proportions of the population that are White. The right panel of Figure 5 shows the cumulative case rates from 3/3/2020 to 12/28/2021, among different ethnicity groups in the county. We see

that Hispanics have the highest standardized case rates, Blacks have the second highest case rates, and Whites have the lowest case rates.

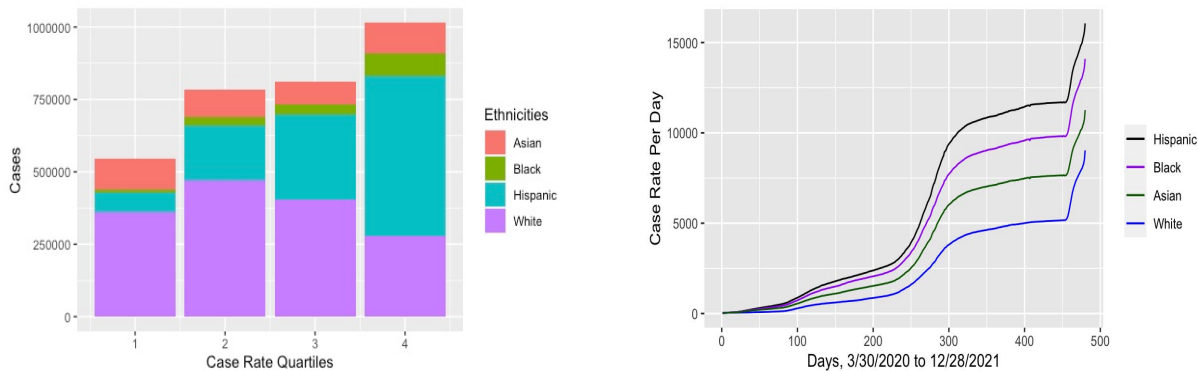


Figure 5: Left: Bar charts of ethnicity, by Covid case rate quartiles. Right: Comparison of cumulative Covid case rates by ethnicity.

5 Conclusion

The results of our analyses show that the total case rates, between 3/3/2020 and 12/28/2021, vary greatly across different zip codes in San Diego County. Regions with a higher proportion of educated citizens have lower case rates, regions with higher rates of poverty have higher case rates, and zip codes with higher Covid rates tend to have a higher proportion of the population that are Hispanic and Black. These findings may be used to improve our pandemic responses in the future. For example, more healthcare workers and educational efforts may be needed in poorer areas. Studies may be conducted on costs of such policy levels as well as return on investment as we prepare for future pandemics as a nation.

We also studied the data split before and after Covid vaccines became available and found the results to be very similar. We wanted to include death rates in this research. Unfortunately, we found death rates only at the county level but not at the zip code level. In the future, we will include death rates at the zip code level to study differential impacts of demographic and economic data on Covid deaths.

Acknowledgement

Many thanks to Professor Joey Lin at San Diego State University and Ms. Diana Loo at Westview High School who served as mentors to this research project. This paper is based on a winning poster at the 2023 American Statistical Association (ASA) Southern California Chapter Statistics Data Visualization Poster Competition and national 2023 ASA Data Visualization Poster Competition.

References

- [1] Ari Lamstein. *choroplethr: Simplify the Creation of Choropleth Maps in R*, 2023. R package version 3.7.2.
- [2] San Diego Association of Governments (SANDAG). Sangis regional gis data warehouse open data portal.
- [3] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023.
- [4] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- [5] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kokske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019.