Public Transportation Accessibility in Correlation to Income and Vulnerability to COVID-19

Lucius Wu, Jodie Lu, Maya Facciolo, Erin Bhan

CYPLAN88

Prof. Marta Gonzalez

May 10, 2021

Abstract

Based on census data and transit networks, we analyzed vulnerability within the city of Santa Monica. Our hypothesis was that a lack of access to public transit would correlate to higher vulnerability. In our analysis, we were able to determine which tracts in Santa Monica were most vulnerable based on its average income and its access to health insurance, public transportation, and vehicles. Each of these factors contribute to our vulnerability index. To run the analysis, we looked at each of the nineteen census tracts individually, organized them based on their attributes, and created a map of the cluster groups. We then overlay the bus network on top of the cluster map to see the correlation between concentration of public transit stops and population vulnerability.

In our analysis, nodes are bus stops of the Santa Monica Big Blue Bus network. We included only the BigBlueBus stops that were within the city limits of Santa Monica, 275 individual stops. Each node has coordinates which correspond to its geographical location in Santa Monica. The links are the connection between bus stops based on the bus route network data. We detail our entire process in the rest of the paper, beginning with a background, then data, methods for analysis, results, and conclusion, and end with ideas for further research. We conclude that our hypothesis is incorrect because the most vulnerable tracts are the downtown areas, which have the highest degree centrality and closeness centrality of public transit stops.

Introduction

Background

Our project is focused on the Santa Monica Big Blue Bus network, a municipal bus system in the city of Santa Monica, California. With an initial interest in transportation networks and accessibility, we decided to study the Santa Monica Big Blue Bus network, especially because we all moved down to Southern California when COVID-19 hit. It was impossible not to notice inequalities highlighted by the pandemic throughout the cities we live in, not to mention around the entire world. There are reasons the pandemic hit worse in certain areas, including a lack of access to healthcare, and these communities have been shown to be majority people of color. In our curiosity to understand how varying demographics and levels of access might show patterns of inequality, this project was born, and we decided to begin close to home in Santa Monica.

We were initially inspired by a master's project that analyzed transit accessibility and its implications on food deserts in the Bay Area (Huang and Lee, 2020). Witnessing a change of demographics in the area, Huang and Lee wanted to visualize and analyze ways for policymakers to increase accessibility to retailers that accepted SNAP. Our project also looked at vulnerabilities, but instead we focused on analyzing Santa Monica to understand which census tracts were most vulnerable, and thus may be more susceptible to the impacts of COVID-19. We achieved this by clustering census tracts into groups of k=5, then we analyzed which tracts were more vulnerable given household income, health insurance access, access to public transportation, and whether they have access to a vehicle. We will then overlay the network map onto the census tracts to further delve into transit accessibility since our hypothesis is: greater transit accessibility will have a negative impact on whether a neighborhood is more vulnerable.

Moreover, we found another paper from the Southern California Association of Governments (SCAG) by Hiroshi Ishikawa and multiple co-authors called “Snapshot of COVID-19 Transportation Impacts in the SCAG Region” that was published in August 2020. The paper was related to our topic in terms of the relationship between health vulnerability and transportation after the pandemic. The paper portrayed a significant decline in bus ridership, falling 71 percent overall in April 2020, compared to April 2019 due to the stay-at-home orders and social distancing.

Table 1: Year-Over-Year Monthly Bus Ridership Change (2019 vs 2020)

TABLE 1 Year-Over-Year Monthly Bus Ridership Change (2019 vs. 2020) — Continued

	JAN	FEB	MAR	APR	MAY
Santa Clarita Transit	1.6%	-4.4%	-45.9%	-81.4%	-81.5%
Santa Monica's Big Blue Bus	6.6%	9.4%	-29.5%	-73.1%	-73.6%
SunLine Transit Agency	4.5%	6.6%	-35.5%	-62.9%	-59.6%
Torrance Transit System	12.9%	11.9%	-25.6%	-69.0%	-62.2%
Ventura Intercity Service Transit Authority	-3.3%	5.7%	-37.6%	-82.5%	-75.8%
Victor Valley Transit Authority	11.9%	20.6%	-33.4%	-70.4%	-67.5%
TOTAL	3.1%	3.8%	-36.0%	-71.2%	-66.5%

Source: National Transit Database monthly module adjusted database (https://www.transit.dot.gov/ntd/data-product/monthly-module-adjusted-data-release)

Source: National Transit Database monthly module adjusted database

Data and Methods

To run our analysis, we used Big Blue Bus GTFS bus route data and U.S. Census API 2017 data for Santa Monica. The network used was the Santa Monica Big Blue Bus network, as mentioned above. We overlaid the network on top of the census tracts of Santa Monica, and analyzed geographical information, to visually demonstrate whether transit accessibility is correlated with vulnerability to COVID-19. Characteristics of the undirected network itself include 275 bus stop nodes, and edges, which are the existence of a connection between the nodes. The edges are weighted, where higher weight means more edges pass between certain nodes, often located in more heavily populated areas.

The census tract data used in the notebook is downloaded through Cenpy. Calling out the nineteen census tracts specifically to Santa Monica, we then used the census tracts and clustered the data into five different groups to determine whether a tract is more 'vulnerable'. Vulnerability as mentioned earlier is defined by aggregating the tract's average income, and access to health insurance, public transportation, and vehicles.

To analyze the data, we used degree centrality, closeness centrality, and K-Means clustering. Degree centrality helps us understand how many bus stops are connected to each other. This can give us an idea of how busy each station is in order to draw conclusions about transportation accessibility of each census tract. Closeness centrality helps us understand how close a bus stop is to other bus stops in the network. This can show us where the bottlenecks of the bus route system are and also help visualize which stations we should expect to be most utilized. Considering the variables (total rented occupied, total population with Bachelor degree, Gini Index for income, total population taking public transportation to work, total population of those driving alone to work, and median household income, we used K-Means clustering to determine the most "vulnerable" tracts in Santa Monica. It helps us visualize the statistics from the census data as well as show the distributions of each census variable for every cluster.

Results

The Network

Our first step is to visualize the network that we will be analyzing. Using the shapefile of the city of Santa Monica, and the Big Blue Bus GTFS data, we plotted the bus routes and stops that are located within Santa Monica's city limits. As shown in the graph below, our network contains 275 bus stops spread out across 19 census tracts.

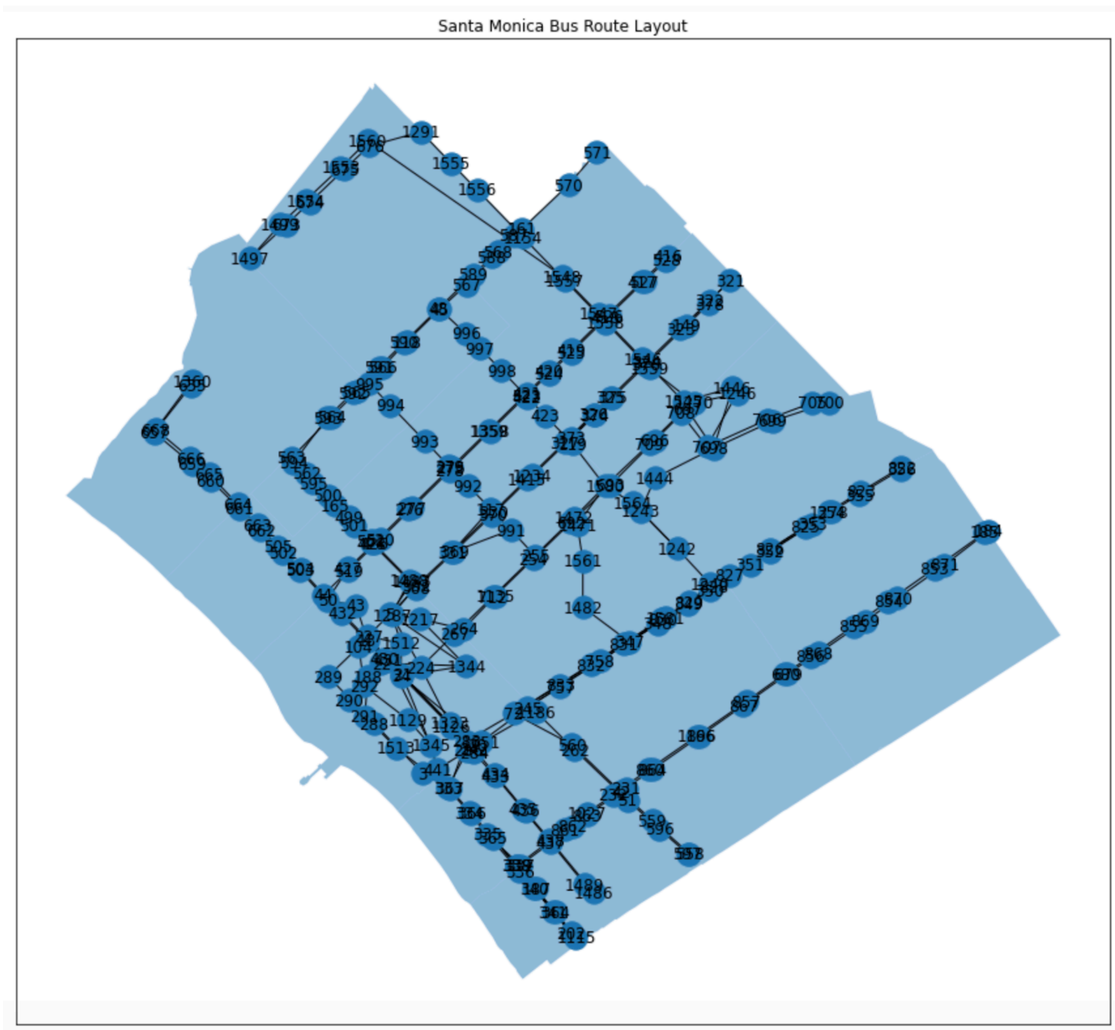

Figure 1: Visual Representation of Santa Monica Bus Route

From the US Census API, we determined that the cluster variables we would use for analysis were: total rented occupied, total population with a bachelor's degree, Gini index, total population taking public transportation to work, total population driving alone to work, and median household income. We plotted colormap graphs with legends to help visualize the distributions of data.

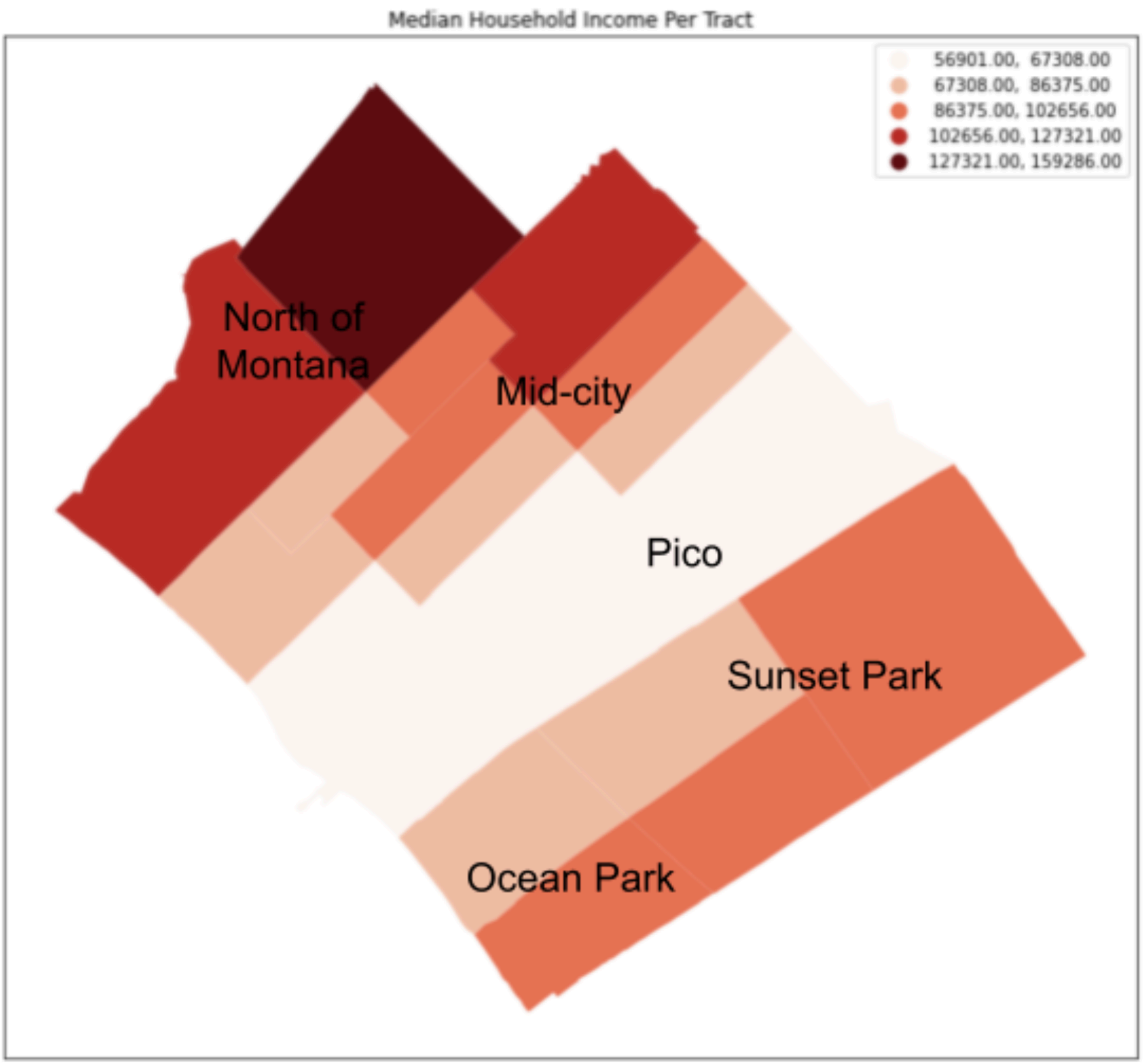

Figure 2: Distribution of Median Household Income Per Tract

The graph above is a larger example of the distributions that we considered. Figure 3 also shows the distributions for each of the census variables.

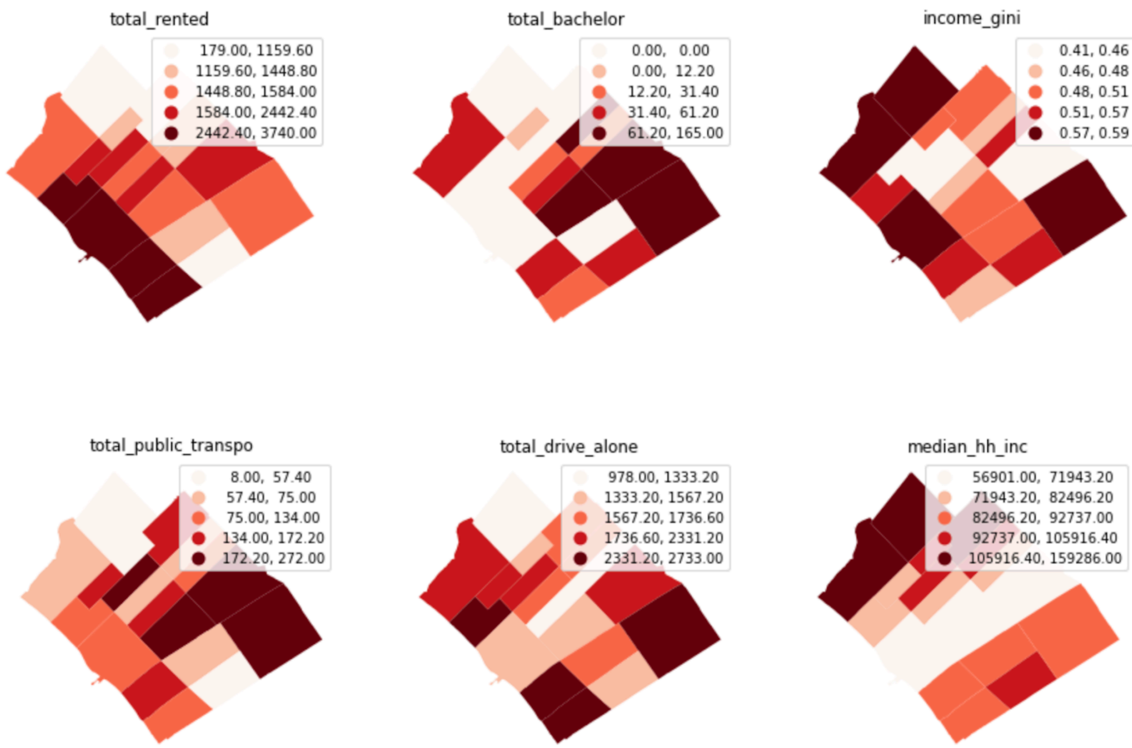

Figure 3: Distribution for Census Variables: Total Rented Occupied, Total Bachelor Degree, Gini Index For Income, Total population of those taking Public Transportation to work, Total Population of Those Driving Alone to Work, & Median Household Income

K-Means and Centrality Visualizations

As mentioned in the Data and Methods section, we used K-Means clustering on the tracts of Santa Monica in order to determine the most vulnerable neighborhoods. Following examples from class, we ran K-Means and plotted the five geodemographic clusters, shown next in Figure 4.

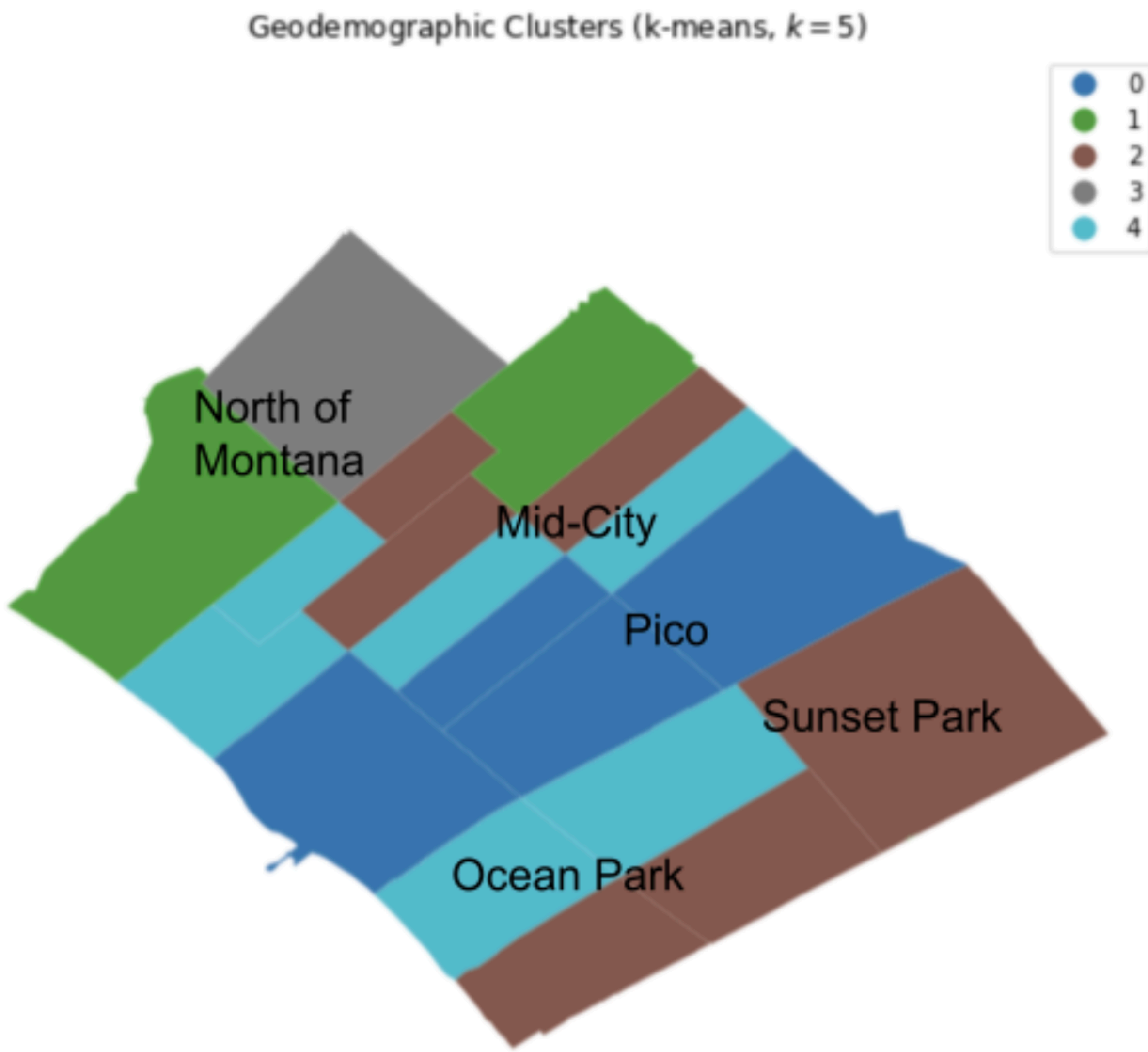

Figure 4: Geodemographic Clusters for k = 5

Then, we used the seaborn library to plot the KDE distributions of each variable by cluster. In the diagrams below in Figure 5, note that there is no distribution shown for Cluster 3 because it only contains one census tract. However, the data points for cluster 3 are plotted as the red lines on each of the subgraphs. As shown by the data, Cluster 3 has the highest median household income, as well as the smallest population of people taking public transportation or driving alone to work, so we will not consider this cluster as the most vulnerable.

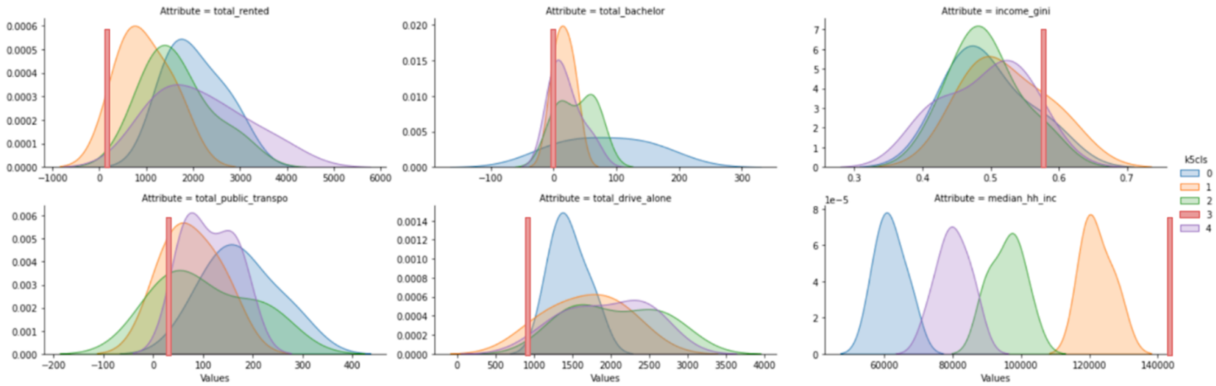

Figure 5: KDE Distributions of Each Variable by Cluster

From the KDE plots, we can determine that Cluster 0 is the most vulnerable, followed by Cluster 4, given that they have the lowest median household income, a high number of rented occupied, a low population with a bachelor's degree, and a high number of people taking public transportation to work, which are all attributes that we deemed as contributing to the vulnerability of an area when comparing tracts against one another.

After we determined that Cluster 0 and Cluster 4 were the most vulnerable, we wanted to see how that would compare to the current layout of the Big Blue Bus stops and routes currently available to the populations in that cluster. As shown in the graph below, we do see that the bus stops with the highest degree centrality are located in Cluster 0 and the bus stops with the highest closeness centrality are located within Cluster 0 and on the outskirts of Cluster 4.

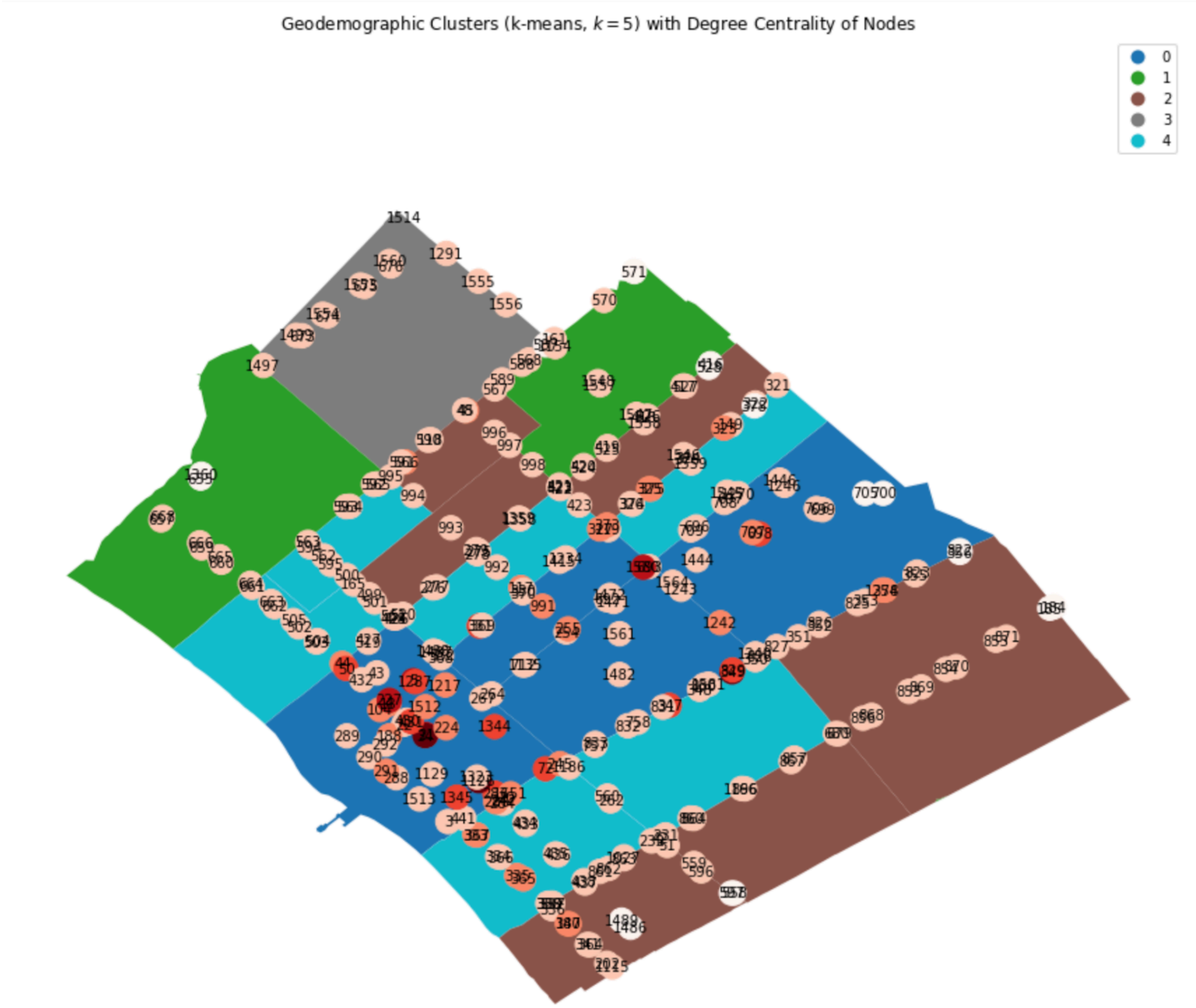

Figure 6: Geodemographic Clusters with Degree Centrality of Nodes (k = 5)

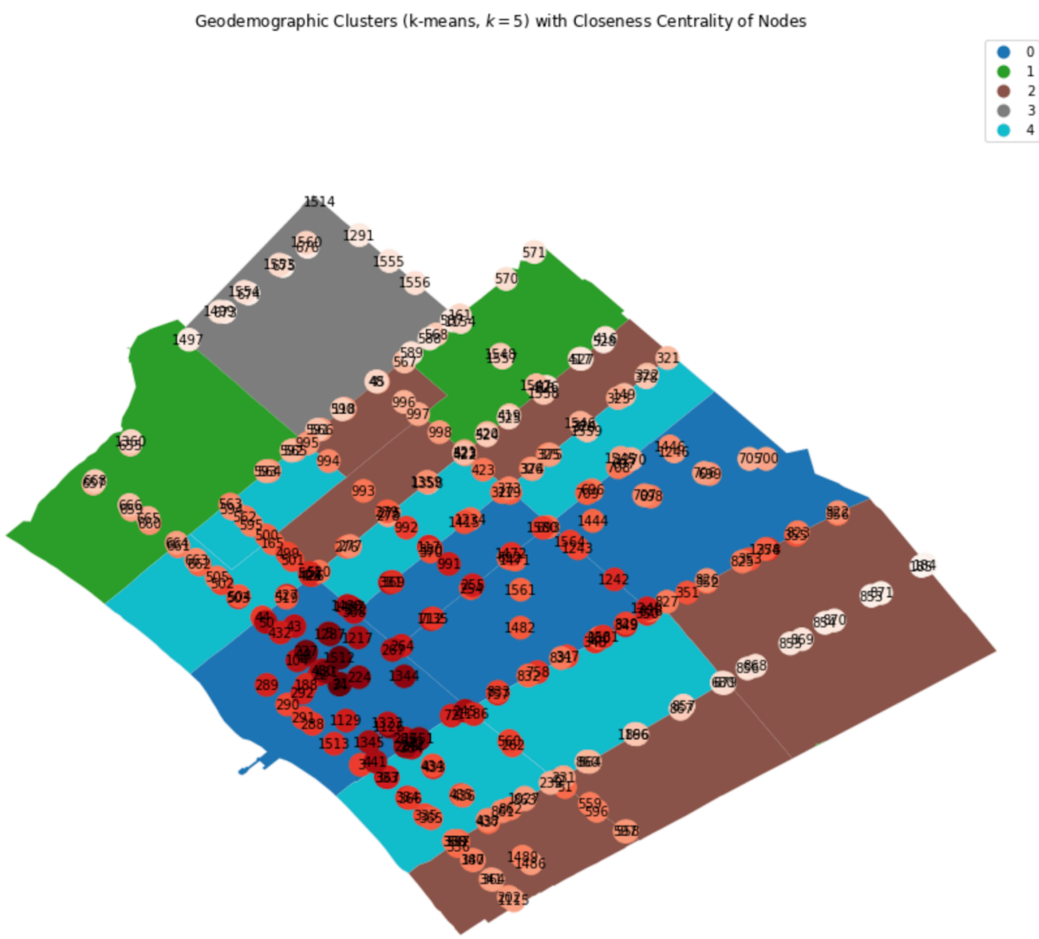

Figure 7: Geodemographic Clusters with Closeness Centrality of Nodes (k = 5)

It is interesting to note that the clusters that we deemed to be the most vulnerable seem to have the highest accessibility to public transportation. As denoted by the darker red, the bus stops that lie within Cluster 0 have the highest closeness centrality. When comparing with other census tract variables, we wanted to visualize the relationship between transit accessibility and median household income.

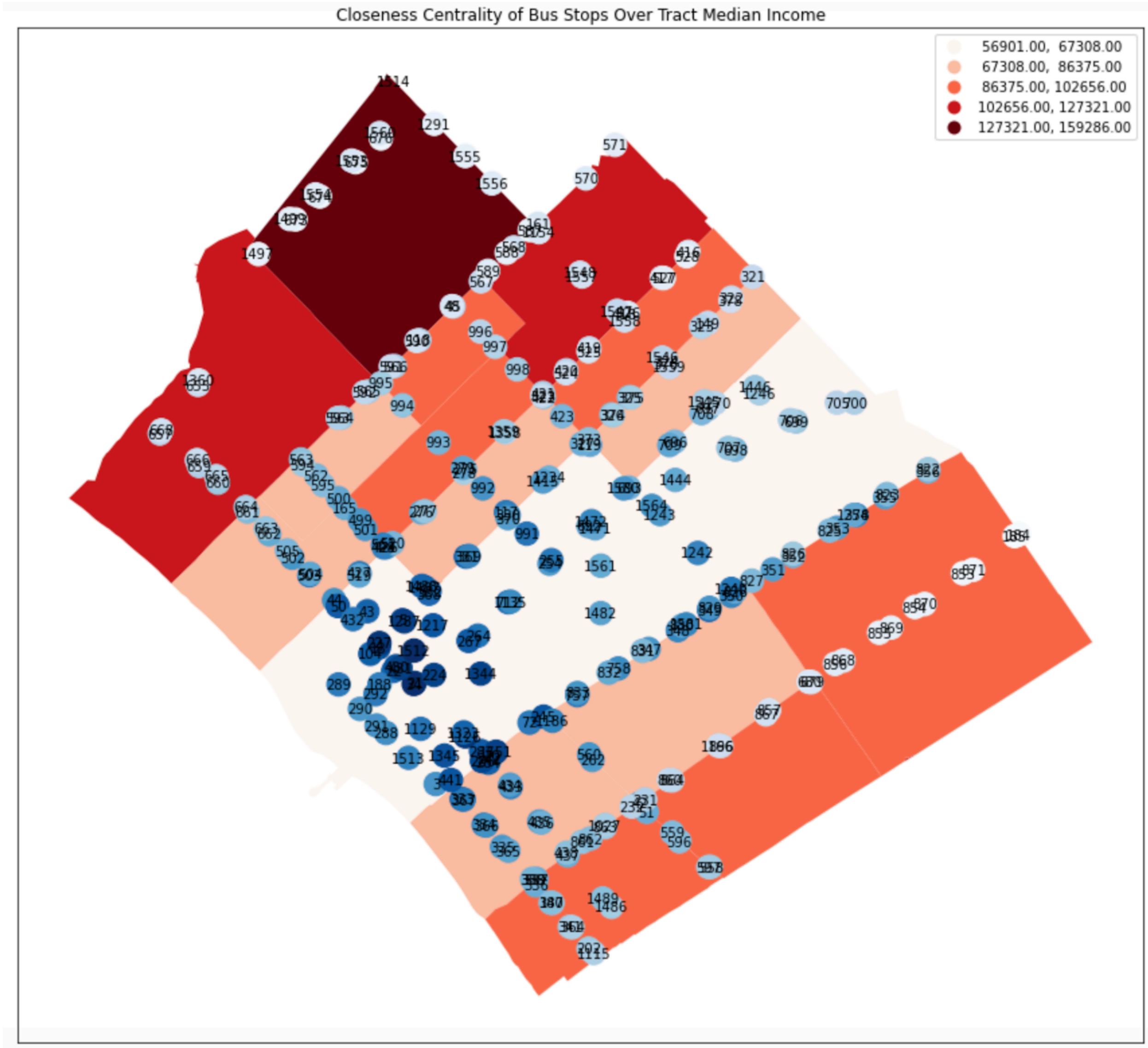

Figure 8: Closeness Centrality over Median Household Income per Tract

In the graph above, we can see again that the Cluster 0 has the lowest average household income (cream color in center), where the highest number of high-centrality nodes (darker blue circles in center) are simultaneously located. This indicates that the bus stops located in the heart of Cluster 0 are most frequently used, high-traffic areas, and make up what is considered downtown Santa Monica. Low income paired with high access to public transportation.

Furthermore, with consideration of the impacts of COVID-19 on the neighborhoods of Santa Monica, we also wanted to visualize the accessibility to health insurance of the census tracts, and see if the results corresponded with our analysis using K-Means and node centralities. Below we show a graph of the populations of the census tracts with access to health insurance, as well as the percentage of each census tract with access to health insurance.

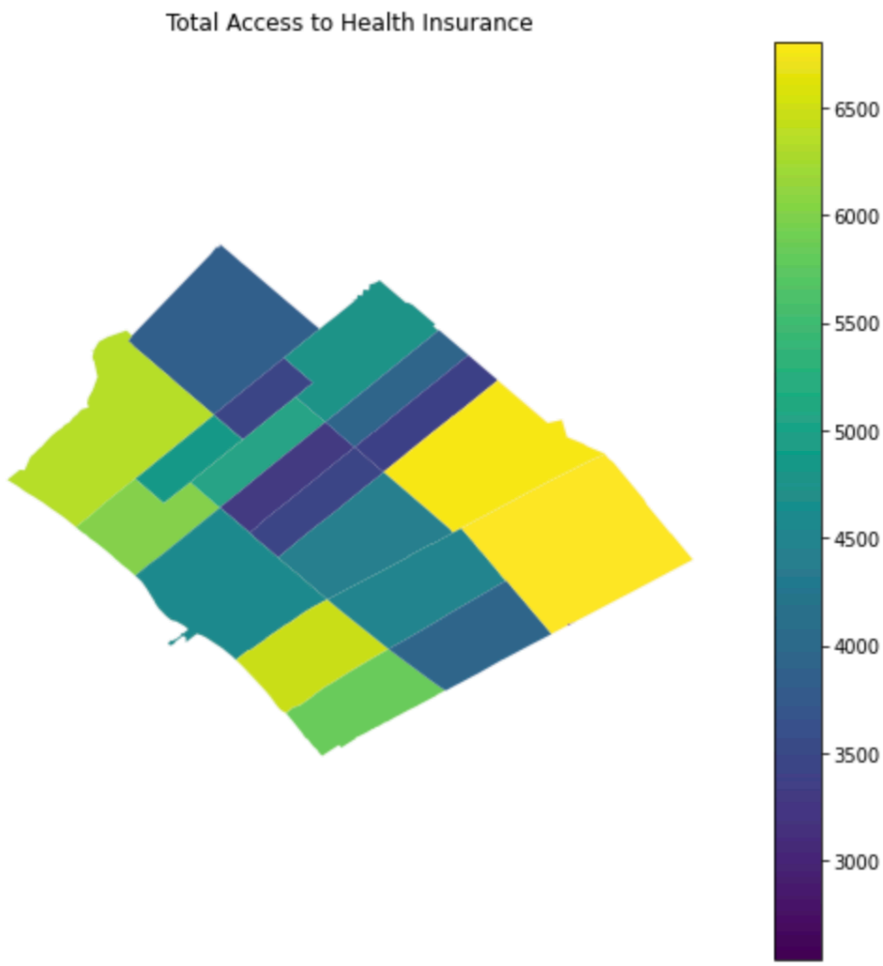

Figure 9: Population of Census Tracts with Access to Health Insurance

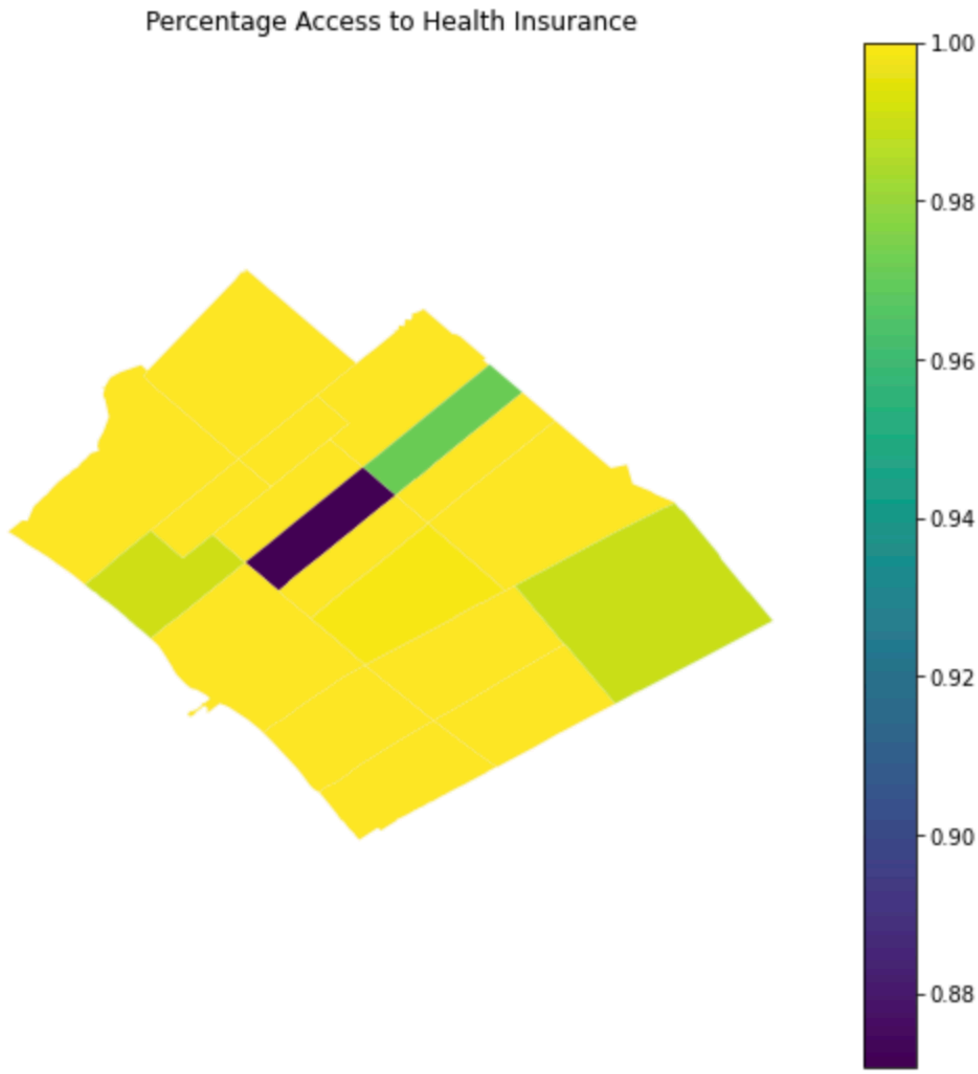

Figure 10: Percentage of Each Census Tract with Access to Health Insurance

To get the percentage of the population with access to health insurance, we divided the total health insurance by the total population of each census tract. While a majority of the census tracts have almost 100% of the population with access to either private or public health insurance, the data confirms that Clusters 0 and 4 are the most vulnerable in Santa Monica. The census tracts in these clusters are the only ones that have less than 90% accessibility to health insurance.

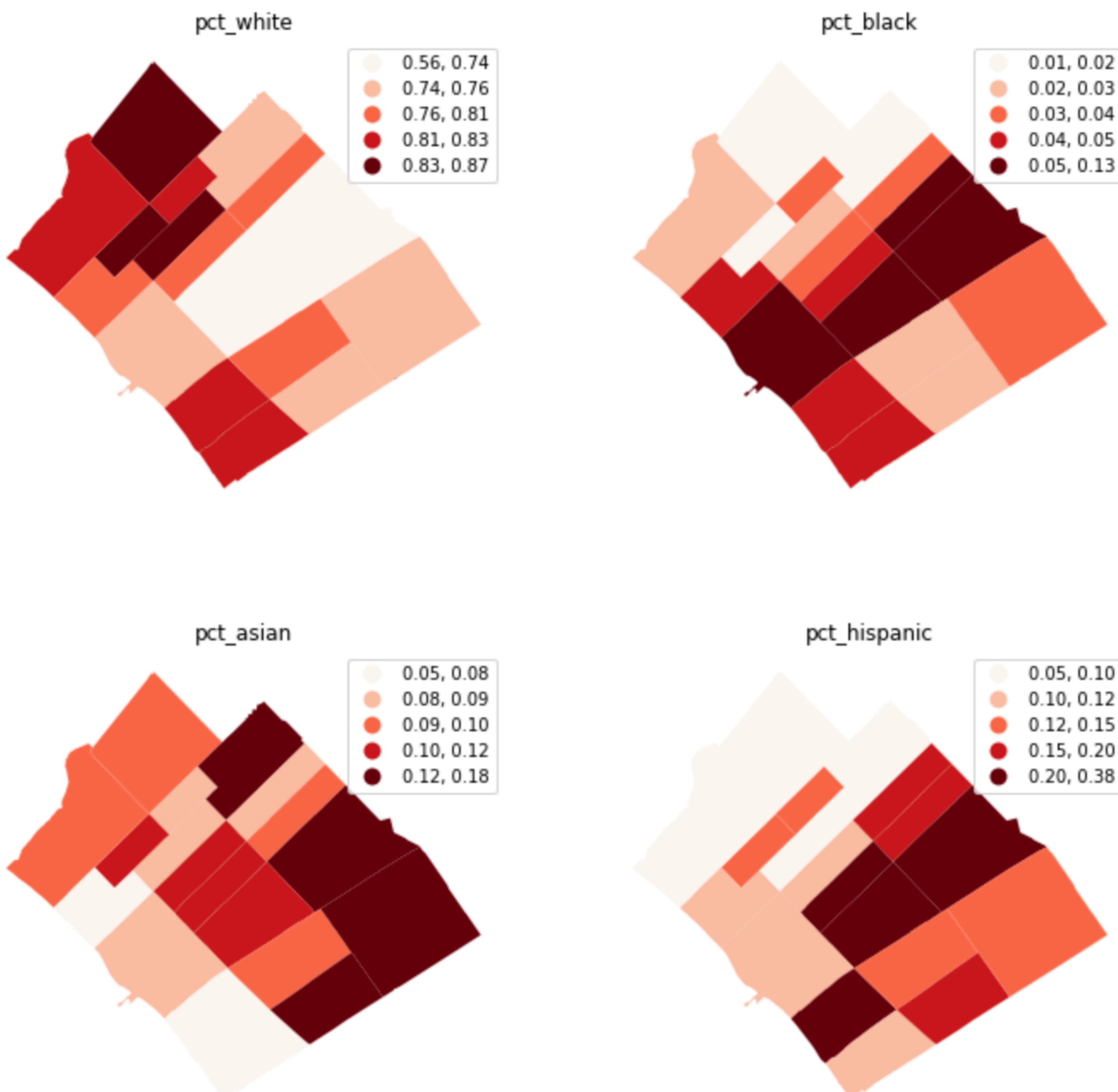

Figure 11: Percentage of Each Race Per Census Tract

In the introduction, we discussed that race and ethnicity played a role in COVID-19 vulnerability, due to the lack of healthcare access in communities of color. In the graph above, we show the distributions of race percentages in each of the census tracts. The diagrams primarily indicate that a majority of the census tract populations are white, with the lowest percentage white population being 56%. In addition, in the four census tracts where there is not a near 100% of healthcare access, there is a relatively higher percentage of people of color, which could indicate reasons for lack of healthcare access. Those census tracts are also in the middle-to-low range relative to the median household income of all tracts (Figure 8). Therefore, the data supports that Clusters 0 and 4 are the most vulnerable in terms of COVID-19 susceptibility, but not necessarily for transit accessibility.

Conclusions and Future Work

Our initial hypothesis included that transit accessibility is a key and important factor in whether a neighborhood is considered vulnerable or not. The data shows that the tracts with the highest accessibility to public transit are often the most vulnerable according to our index. For example, if looking at one of the characteristics we considered for our vulnerability index, the median household income from tracts with the highest closeness centrality of stations often had the lowest income. Therefore, our hypothesis was proven wrong as we predicted that higher transit accessibility correlated with a decrease in vulnerability. We also find that the GINI coefficient, a measure of the distribution of income across a population gauging economic inequality, widely varies between all the tracts which further complicates the issue. For a follow up to this particular project, other factors should be considered into the vulnerability index, including food accessibility and insecurity and exposure to pollution for the average resident of the city (pollution can affect health injustices) just to name a few.

The article on COVID-19 transportation impacts analyzes the public response to transportation during the pandemic in Santa Monica. This article and our paper only scratch the surface of research on this subject, there is still much need for further analysis, well beyond the scope of this class project. For future analysis, the geographical scope of census tracts may expand to surrounding cities. This project focuses only on Santa Monica, but it would be valuable to compare between many tracts to best analyze how transit accessibility and income correlate to susceptibility to COVID-19. Also, another interesting avenue to explore is where a 'poorer' city, one with lower average household income, has resilient pockets in certain census tracts. It would prove helpful to expand this scope across many cities throughout the country. In addition, any future analysis on COVID-19 vulnerability would benefit from incorporating tourism data from each city. For example, while we saw high rates of COVID-19 in some areas, there happen to be a plethora of tourists visiting from other cities in those areas, and those visitors may play a big role in the spread of COVID-19.

In the masters' paper on food deserts by Huang & Lee, they used SNAP retailers and accessibility to those retailers by public transportation. Although our project does not measure food accessibility, food is one of the most important human resources. Food insecurity, according to the USDA, is one of the main barriers individuals face to live a healthy, active lifestyle. It would be beneficial to include this aspect in future projects when analyzing a city's bus networks and vulnerability. High rates of food insecurity makes a group of people more vulnerable, especially when they live in a food desert, and transit accessibility is key to getting access to any healthy consistent source of food. There is much left to explore in terms of the correlation between average income, access to public transportation, and vulnerability.

We also suggest using the same methods of analysis used in this project and using it to measure vulnerability across other cities across the country and world. Census data is available across the country and GTFS data is readily available for many transit networks across the world. Other countries have their own versions of the census so we see that it is applicable and easy to replicate this project to other cities across the world. Lastly, we suggest also looking at 'resilience' and building a resilience index for different cities. We focused on vulnerability in this paper but analyzing where pockets of ‘resilience’ occur will be just as valuable as it can lead to conclusions whether the cities analyzed might have a higher income inequality (where the rich live in a specific tract while the poor live in others hence the resilient pockets in a city) or whether local policymakers are improving the livelihoods of their citizens. We hope that this project will help policymakers improve their cities and to scrutinize every part of their city, and to make each of their neighborhoods have the same quality of life regardless of race, ethnicity, and income.

Works Cited

Big Blue Bus GTFS Service, gtfs.bigbluebus.com/.

"Census Tracts - 2010: Open Data - City of Santa Monica." Socrata, data.smgov.net/dataset/Census-Tracts-2010/86sk-d5gw.

DeCaprio, Dave, et al. "Building a COVID-19 Vulnerability Index." ArXiv.org, 18 July 2020, arxiv.org/abs/2003.07347.

"Definitions of Food Security." USDA ERS - Definitions of Food Security, www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in the-us/definitions-of-food-security.aspx.

Huang , Amy, and Jasper Lee. Transit Accessibility and Its Implications on Food Deserts in the Bay Area. University of California Berkeley, 14 Dec. 2020.

Ishikawa, Hiroshi, et al. Snapshot of COVID-19 Transportation Impacts in the SCAG Region. Southern California Association of Government, Aug. 2020, p. 15, https://scag.ca.gov/sites/main/files/file-attachments/scag covid-19-transportation_impacts.pdf?1604269751.

SMI Adviser. "Double Jeopardy: COVID-19 and Behavioral Health Disparities for Black and Latino Communities in the U.S." SMI Adviser, Substance Abuse and Mental Health Services Administration (SAMHSA), 19 June 2020, smiadviser.org/knowledge_post/double-jeopardy-covid-19-and behavioral-health-disparities-for-black-and-latino-communities-in-the u-s.