Measuring Health in Food Deserts - 2020 Research Paper | Hannah Perlstein

Introduction

Seeing as COVID-19 has caused massive stockpiling of essential resources, like food items, I was curious to whether the most vulnerable populations were susceptible to failing to get the necessary items due to the exploitation of resources by the wealthy. In narrowing the topic, I was struck to investigate the prominence of healthy grocery stores in neighborhoods and the possible implication of this on various health factors. More specifically, I wanted to investigate whether access to healthy grocery stores coincides with lower poverty and obesity rates and additionally if there is a correlation between these findings and racial demographics.

Data Collection & Exploratory Data Analysis

For my research, I used the AskCHIS Neighborhood Edition, the UCLA Center for Health Policy Research's extensive database with data sourced by the California Health Interview Survey (CHIS). The dataset goes as far back as 2001 and adds new data each year with surveys from 20,000 Californians. Since the dataset is for all of California, and my project is more narrow, I selected the data to use by the definition of a constituent core based Metropolitan Statistical Area, specifically the Los Angeles-Long Beach-Anaheim, CA Metropolitan Statistical Area. This area was chosen due to its availability within the database, personal familiarity, in addition to a largely heterogeneous population. Within this area, I broke it down further into 322 zip code regions in Los Angeles County and Orange County, CA. Using AskCHIS I selected health related variables for each zip code: Total Population, proportion of the population with diabetes, heart disease, poverty, food insecurity, and obesity (defined as a Body Mass Index greater than or equal to 30) in addition to racial demographics for each Zip Code.

I then compiled a list of Trader Joe's Grocery Store locations falling into any of the Zip Codes of the Metropolitan Statistical Area (MSA). From here I calculated the longitude and latitude of each Zip Code and each Trader Joe's location. With this information, I was able to calculate the minimum distance to a Trader Joes from each subset of the MSA.

Data Cleaning

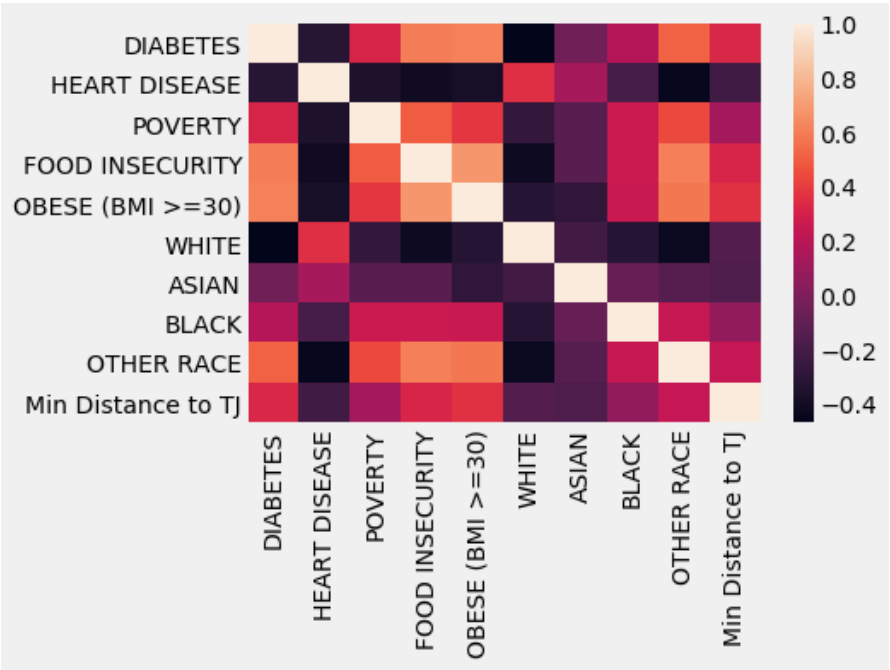
If a Zip Code had a Trader Joe's within it, I normalized its distance to 0 as the coordinates of the Zip Code were arbitrarily placed in the center of the area (based upon Google's map data). Additionally, if any Zip Code had a null value in any of the 10 criteria I was looking for, it was removed from the data set.

Visualizations

	Min	Max	Mean	St. Dev	Median
Total	1100	20500	28500.310559	19172.937344	2550

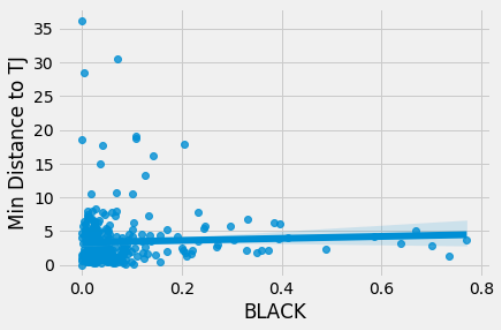
Population					
Diabetes	0.0125	0.135	0.089449	0.024335	0.088500
Heart Disease	0.034	0.15	0.059009	0.012373	0.058000
Poverty	0	0.598	0.134069	0.078888	0.119500
Food Insecurity	0.001	0.7	0.071807	0.066658	0.057000
Obesity (BMI >= 30)	0.085	0.442	0.266202	0.082905	0.257500
White	0	0.97	0.576591	0.200934	0.600500
Asian	0	0.725	0.159674	0.145772	0.117000
Black	0	0.771	0.07365	0.115907	0.035000
Other	0	0.555	0.182311	0.126126	0.150500
Min Distance to TJ (miles)	0	36.21095	3.431438	4.094115	2.256336

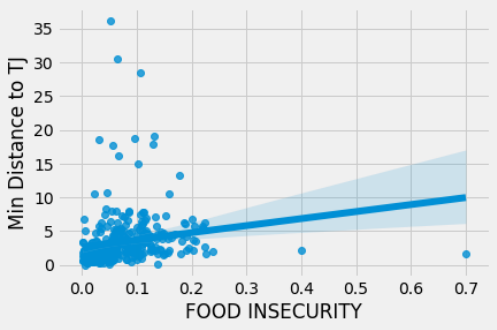
Seeing as my investigative question relies on the relationship between access to low cost healthy grocery stores (Trader Joes) and a myriad of health risks, I found the correlation between these variables and plotted them in a heat map.

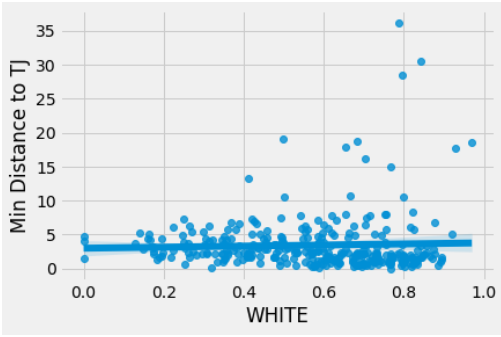


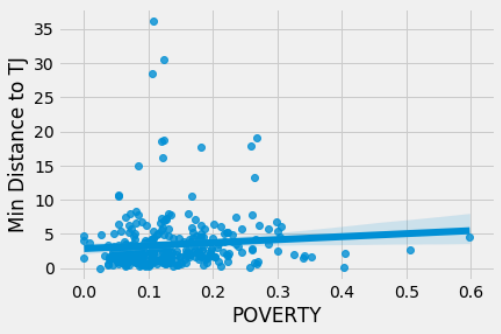
This heat map shows the highest correlation between Min Distance to TJs and another variable are Obesity (BMI >=30) (.367 correlation) followed by Diabetes (.332), Food Insecurity (.320), and Other Race (.239).

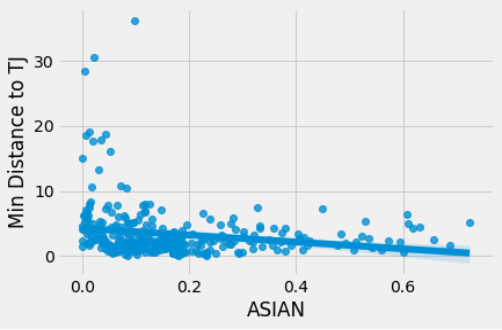
I also display a few of the correlations in the following linear regression plots. The correlation in these are weaker than the variables mentioned above. I have also attached a box plot which shows that the majority of Zip Codes in Los Angeles are very close to a Trader Joe's location (75% are <5 miles), directly opposes the idea of minimal access to healthy affordable grocery stores in los angeles

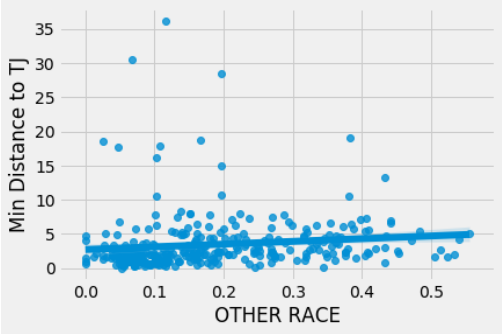


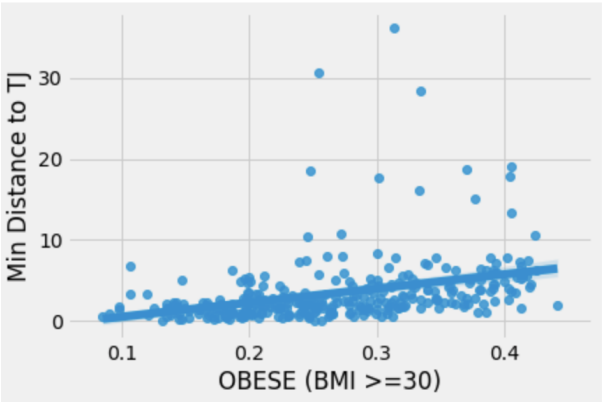


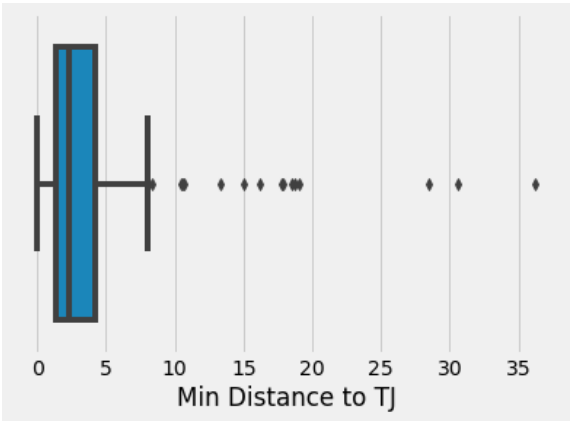


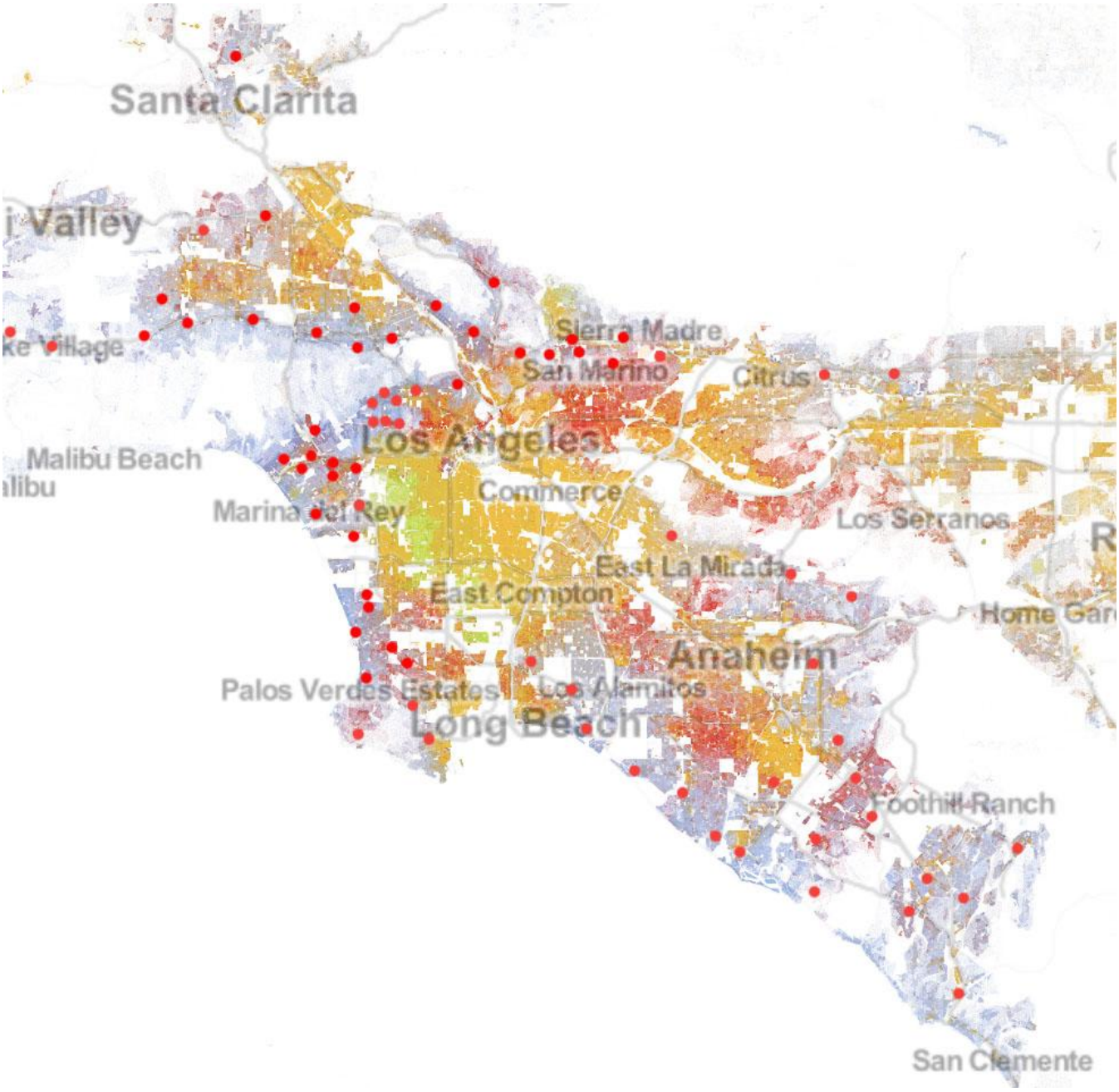












Here, I included a map visualization of the Trader Joe's locations across the two counties with each location denoted as a red dot mapped over the University of Virginia's Racial Dot Map. 1 dot in this map equals one person and the Blue dots = White, Green = Black, Red = Asian, Orange = Hispanic, Brown = Other.

Visually, however, it appears that there should be a high correlation between race and Trader Joe's locations. The red dots almost exclusively appear in areas with a high concentration of White and Asian people and are noticeably absent in Black and Other Race dominated areas.

Conclusion & Future Research

Visually, but not numerically, there appears to be a relationship between minimum distance from a Trader Joe's location and predominantly minority (excluding Asian) populations. Additionally, as mentioned in the data section, Min Distance to TJs is only weakly correlated with the health factors I have highlighted. While 5 miles does not seem much numerically, in Los Angeles without a car it could mean the difference between going to that grocery store and not. In the future, I would like to investigate this with some maximum distance considered “normal” to see whether this changes the interpretation of the correlations. Los Angeles might be too small of a metropolitan area in terms of distance, and thus while the general trend should be there it does not appear when measured in miles. In the future, I would like to expand to include all of California or also metropolitan areas that are larger in area than Los Angeles.

References

[1] https://askchisne.ucla.edu/ask/_layouts/ne/dashboard.aspx#/
[2] https://en.wikipedia.org/wiki/Combined_statistical_area
[3] http://racialdotmap.demographics.coopercenter.org
[4] https://locations.traderjoes.com/ca/