Capstone Project: Food Access in Los Angeles

Coursera: Applied Data Science Capstone Course

Riley Franks
June 9, 2019

Introduction / Business Problem

In the last decade, nutrition in the United States has come into focus as a national crisis. Millions of Americans lack access to an adequate supply of fresh, healthy food. And, according to the CDC, there appears to be a link between access to affordable nutritious foods and the incorporation of healthy foods into the populations' diets. There is growing interest in the idea of "food desert" communities. There's no standard definition, but generally these are neighborhoods that lack grocery stores and farmers' markets. The U.S. Department of Agriculture summarizes the problem well: "Some people and places, especially those with low income, may face greater barriers in accessing healthy and affordable food retailers, which may negatively affect diet and food security."

This project will focus on Los Angeles County in California. We will analyze geographic variation in food access and population demographics, and overlay that onto a map of grocery stores, convenience stores, and restaurants in the area. Community planners and local governments may find this more useful than data reports and tables in getting an intuitive understanding of the food landscape in their cities.

Data

The U.S. Department of Agriculture's Economic Research Service publishes a Food Access Research Atlas[1] with census-tract level data on food access in the United States. There are tons of indicators, but we will be exploring a few specific variables. We get counts of the total population in each census tract, the population living more than a 1/2 mile from a supermarket, and the low-income population living more than a 1/2 mile from a supermarket.

The University of Southern California provides a crosswalk[2] to get latitude and longitude coordinates for a census tract. We will use this to connect the food access data to geospatial coordinates.

The Los Angeles Times publishes geographic boundary data[3] that we will use to draw the census tract divisions on our map.

Finally, we will access the Foursquare API[4] to look up the venues in each neighborhood. We will request information for each census tract using their latitude/longitude coordinates. The Foursquare data gives us the name, location, and category for each venue. We'll use the category variable to subset our results to food-related venues.

1. https://www.ers.usda.gov/data-products/food-access-research-atlas/
2. https://usc.data.socrata.com/widgets/atat-mmad
3. https://boundaries.latimes.com/sets/
4. https://developer.foursquare.com/

Methodology

First, I loaded food access data from the USDA Economic Research Service food access research atlas. I wrangled the data to format the population shares as percentages and extract the 6 digits census tract number that will link to other data sources. Here is a peek at the first few rows of the data:

	CensusTract	State	County	Urban	POP2010	lapophalf	lapophalfshare	lalowihalf	lalowihalfshare	atlas_tract
0	6037101110	California	Los Angeles	1	4731	3461.636143	73.169227	1443.337800	30.508091	101110
1	6037101122	California	Los Angeles	1	3664	3660.694039	99.909772	317.312226	8.660268	101122
2	6037101210	California	Los Angeles	1	5990	5329.504750	88.973368	3035.663558	50.678857	101210
3	6037101220	California	Los Angeles	1	3363	1874.230204	55.730901	681.286018	20.258282	101220
4	6037101300	California	Los Angeles	1	4199	739.348889	17.607737	76.682345	1.826205	101300

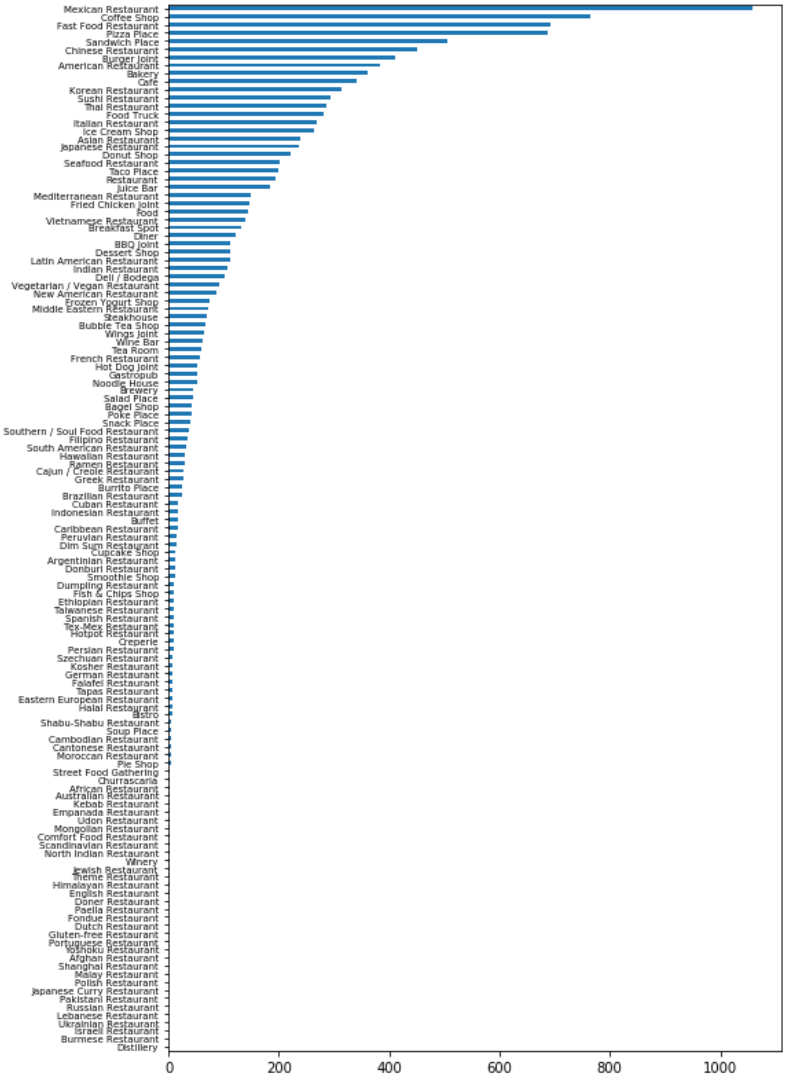
I combined that with data from the University of Southern California's Neighborhood Data for Social Change (NDSC) platform. This gives us specific latitude and longitude coordinates that we can use to query Foursquare data for each census tract. As an additional bonus, we also get neighborhood names!

	POP2010	lapophalf	lapophalfshare	lalowihalf	lalowihalfshare	Latitude	Longitude	Neighborhood	Tract Number
0	4731	3461.636143	73.169227	1443.337800	30.508091	34.259555	-118.293602	Tujunga	101110
1	3664	3660.694039	99.909772	317.312226	8.660268	34.267357	-118.290240	Tujunga	101122
2	5990	5329.504750	88.973368	3035.663558	50.678857	34.251998	-118.292687	Tujunga	101210
3	3363	1874.230204	55.730901	681.286018	20.258282	34.251190	-118.281014	Tujunga	101220
4	4199	739.348889	17.607737	76.682345	1.826205	34.245595	-118.271731	Tujunga	101300

Then I connected to the Foursquare API to query information about food-related venues in the area. I looped through all 2,342 census tracts and extracted the venue name, ID, geographic coordinates, and venue category. I wanted to make sure each venue was only counted once, so I de-duplicated the data and kept the result with the smallest 'distance' value. We ended up with 11,949 results. Here's a sample:

	Neighborhood	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Northridge - Census Tract 111205	Starbucks	34.256671	-118.537146	Coffee Shop
1	Mission Hills - Census Tract 109604	Matt & Tony's	34.264956	-118.466957	Sandwich Place
2	Northridge - Census Tract 113301	Cold Stone Creamery	34.242330	-118.559864	Ice Cream Shop
3	Northridge - Census Tract 113301	Romano's Macaroni Grill	34.242092	-118.556188	Italian Restaurant
4	Northridge - Census Tract 111205	Shogun Sushi	34.255875	-118.535482	Sushi Restaurant

The venue categories are too granular to be useful. There are 129 different categories including labels as specific as Japanese Curry Restaurant, and Pie Shop.



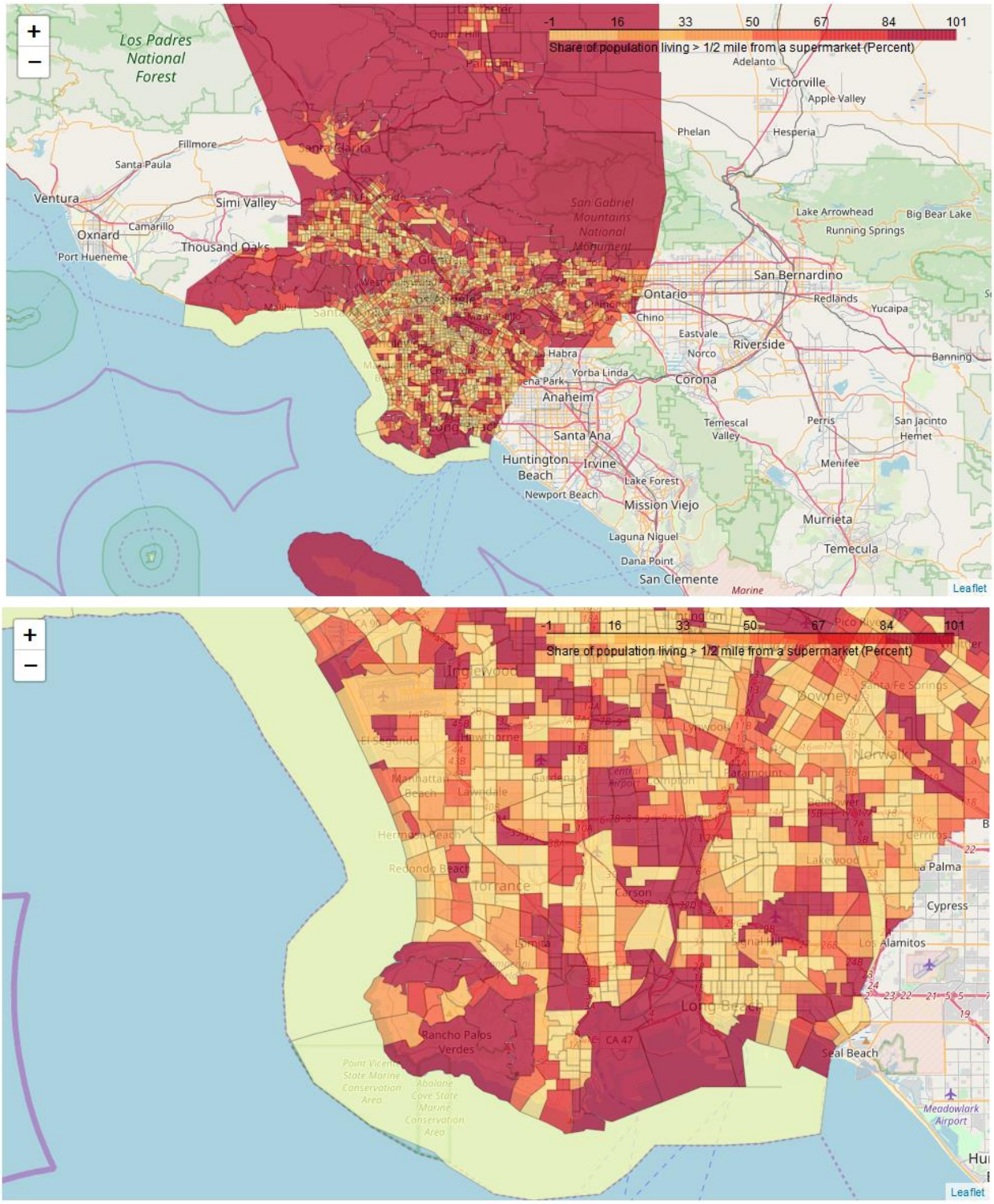
I grouped the categories into broader categories: Restaurant, Fast Food, Café, Bakery/Dessert, Street Food, Bar, and two “Other” categories:

Restaurant	5737
Fast Food	3040
Cafe	1430
Bakery / Dessert	1101
Street Food	284
Other - Food	144
Bar	110
Other - Deli / Bodega	103

Finally, I put all of this together into a map visualization using folium.

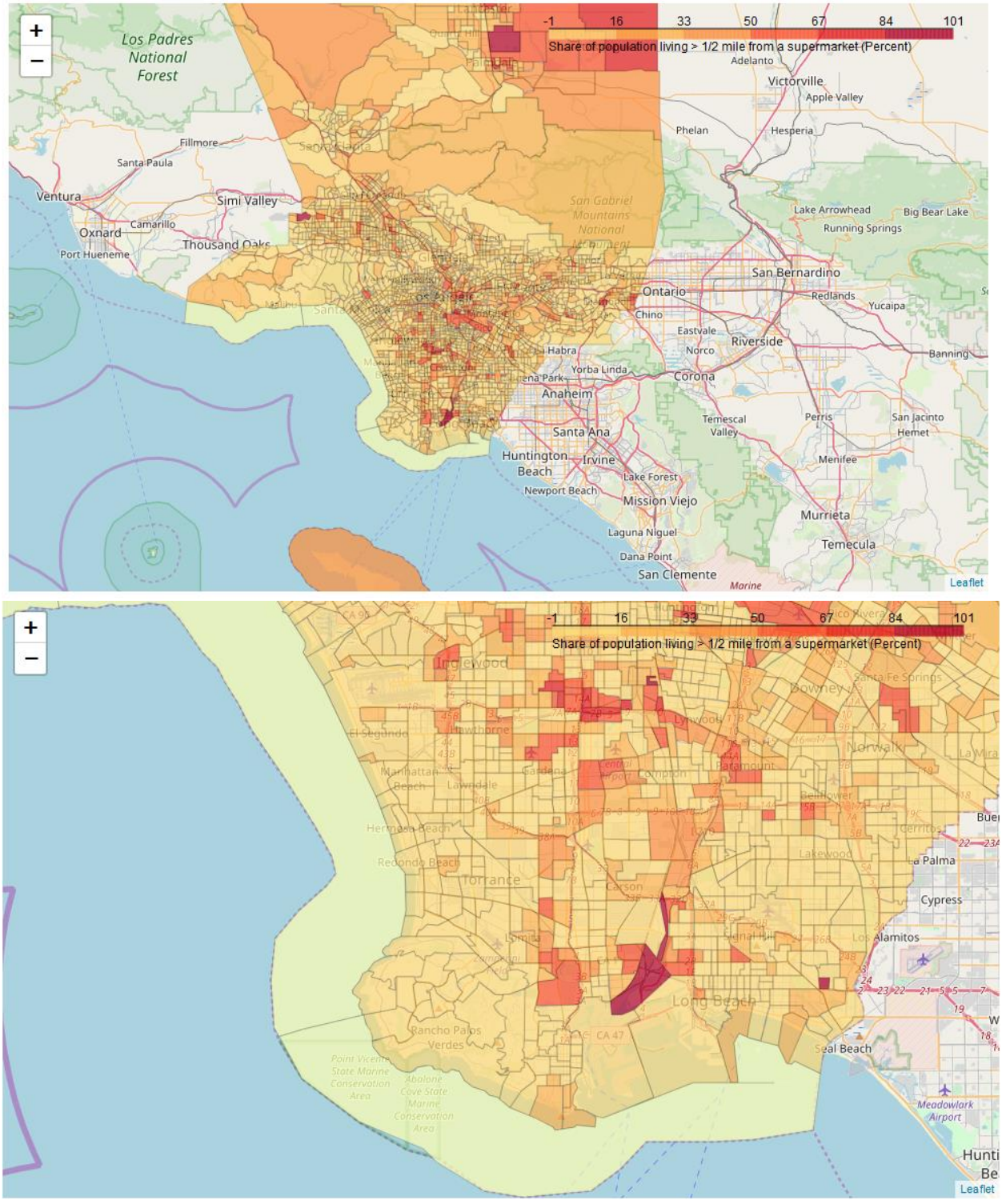
Results

My first choropleth visualized the share of the population of each census tract that live more than half a mile from a supermarket.



This is interesting. There's a wide range of values. Some of the neighborhoods that seem to have low food access are in areas generally known to be quite wealthy (e.g. Rancho Palos Verdes). I decided to

focus on a different metric in the food access data: the share of the population in a census tract that is ***low-income*** and living more than ½ a mile from a supermarket.

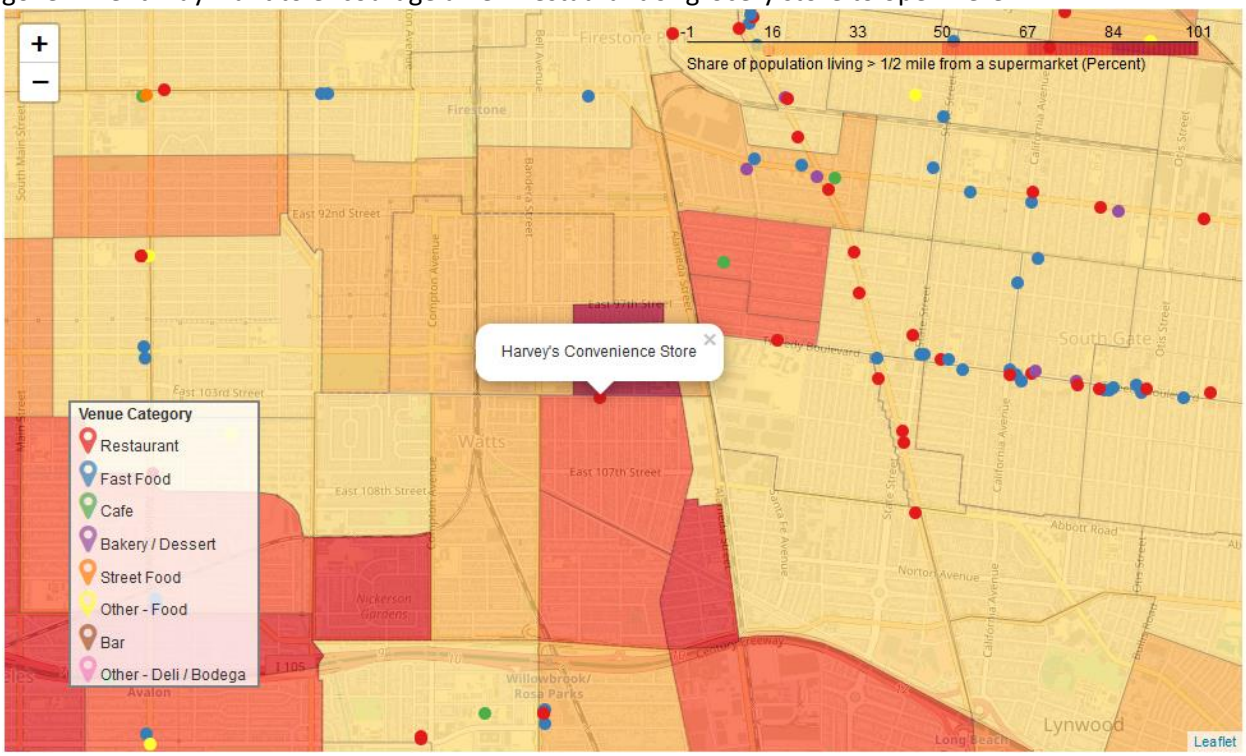


This version makes more sense to me and seems more useful to a community planner interested in food deserts. After adding markers for the food venues identified via Foursquare, we have an interactive map

that overlays a visualization of the level of food access in a neighborhood with the actual restaurants and fast food places in the area.

Discussion

A community planner or local government official can use this map to get a better understanding of the food access issues in their communities. For example, zooming in on the Watts neighborhood, you see a few census tracts with very high proportions of low-income, low food access populations. In one neighborhood, there is only one venue – Harvey's Convenience Store on East 103rd Street. Local government may want to encourage a new restaurant or grocery store to open here.



Conclusion

This project focused on Los Angeles County in California, but this analysis could be extended to any community of interest. There is a lot of valuable information available through public data resources. Data visualization can help turn those raw numbers into meaningful insights. Community planners and local governments may find this more useful than data reports and tables in getting an intuitive understanding of the food landscape in their cities.