



Miami Dade County: Exploring healthy living by income level

Douglas Moura
Jeff Bednar
Valeria Claros

November 2, 2020

Motivation & Summary

Hypothesis: Individuals in higher income (or in low poverty rate) brackets will be healthier and have access to more fresh and nutritional food

Questions we asked:

1. Do people in higher income brackets have more access to supermarkets, gyms and liquor stores?
2. Do people in higher income brackets have healthier lifestyles (focusing on obesity and unhealthy behaviors as metrics)?

As part of the community, we want to better understand the distribution of healthy living in all neighborhoods in Miami Dade by analyzing the income level of people and health data.

We were not able to answer these two questions to our satisfaction. We were limited by how much health data was available and how it is reported.

We were able to answer one out of two questions.

Questions & Data

- We were curious to understand the correlation between healthy living and income. We assumed that if people have more access and are in proximity to healthy choices then they would be more inclined to live a better lifestyle.
- We looked at household income and poverty rate data from the Census API (2013) to come up with a estimate by neighborhood
- We also used Google Maps API to come up with the number of supermarkets, liquor stores and gyms by neighborhood. The goal of this step was to understand the options available to people in the community. We assumed that the more choices people have nearby would influence their type of lifestyle in regards to health.
- Lastly, we explored the CDC databases (2017) and used obesity as a metric to identify areas where people were healthier and show less health problems.

Data Cleanup & Exploration

Describe the exploration and cleanup process

- We used the Census API website to pull data by zip code. The data shows the income level per household, poverty rate and per capita income. We used this data to understand the distribution of income among all neighborhood and zip codes in Miami Dade.
- The Census data reported multiple zip codes per city. The cities also use zip codes for specific PO boxes, not associated with any area in the city.
- We found various outliers in the Census data. One particular zip code in Miami (Fisher Island) is one of the wealthiest zip codes in the country and it stood out as an outlier.
- Supermarket, liquor store and gym data. We used Google Maps API to pull locations for Supermarket, liquor store and gyms in each neighborhood in Miami Dade County. We then aggregated each type to count how many of them were located in a zip code.
- Originally, we planned on including fast food restaurants in our analysis. However, Google Maps did not allow us to use “fast food” as a variable to search for. Google Maps only allowed us look for restaurants in general. We decided to exclude restaurants in general because it didn’t give us an accurate representation of the analysis we are looking for.
- CDC Data: CDC does not report obesity data for every zip code in Miami Dade. The data availability was limited to a few neighborhoods in the county. This data was also reported by latitude and longitude, not by zip code.
A major challenge for us was deciding if we were going to merge all 3 data frames together using zip code. However, the CDC data reported multiple data points for a single zip code using geolocations.
We decided to use the CDC Data in a map as a layer to visualize the population with highest obesity and unhealthy behavior rates .

Data Cleanup & Exploration

Describe the exploration and cleanup process

- We were able to combine the Census data and Google Maps data by using the zip code as the link between these tables. However, we ran into trouble when the CDC data didn't report any zip codes and only geolocations. CDC data also reported multiple data points per zip code, making it very difficult to accurately pinpoint the exact value of obesity in each neighborhood. Additionally, the CDC does not report data for every location in Miami Dade county. Most health websites report data by state or by city, it was difficult to find very detailed data. Health data excludes anyone below 18 years of age.
- The Census data reported income less than 0, we had to exclude the zip codes reporting this number as it was obviously incorrect.
- We used one data frame that combined Census data and Google Maps data to calculate how many locations (Super Markets, Gyms and Liquor Stores) were available in each zip code and neighborhood in Miami Dade. We found several outliers in the data due to income levels that are reported in the county.

Data Analysis

Super Markets, Gyms and Liquor Stores by Zip Code

1. Do people in higher income brackets have more access to super markets, gyms and liquor stores?

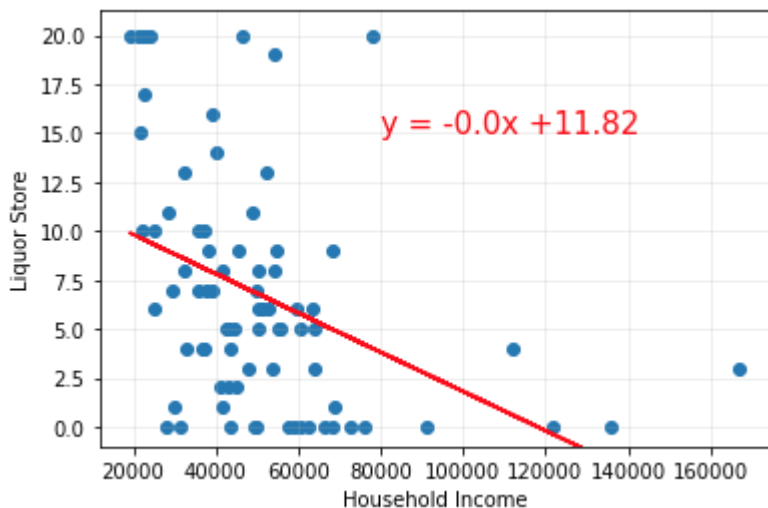
One interesting point about the income level vs. supermarket, liquor store and gym is that we saw a huge difference in the number of locations that showed up for income above \$100K. There are very few locations in the income brackets above \$100K when we thought these locations would show the most super markets and gyms nearby. This results were completely different than what we assumed we would see.

We can see that the strongest correlation is between super markets and income level with a r-squared value of 0.29, followed by liquor stores.

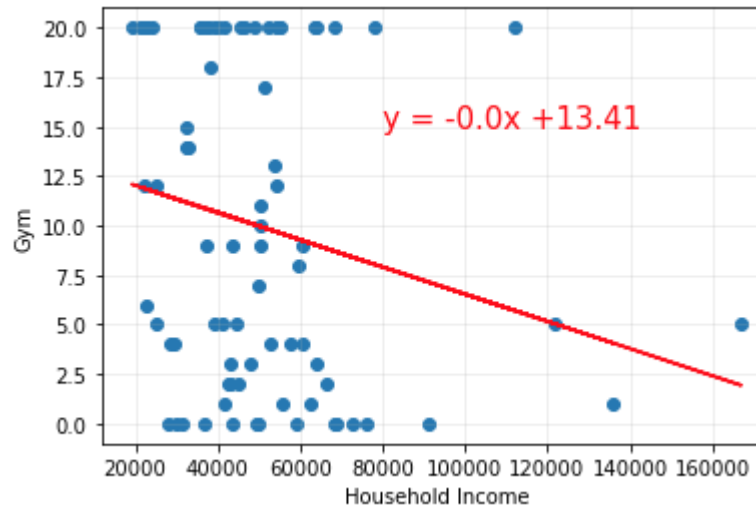
Gyms showed the smallest r-squared with a value 0.04, showing that income and number of gyms in each neighborhood is not correlated.

Miami in general might not be a good location to make an accurate assessment of this answer. In the county, we have Fisher Island, one of the wealthiest zip codes in the country, but also have Liberty City that is extremely poor. Both neighborhood are close by and are drastic representations of incomes.

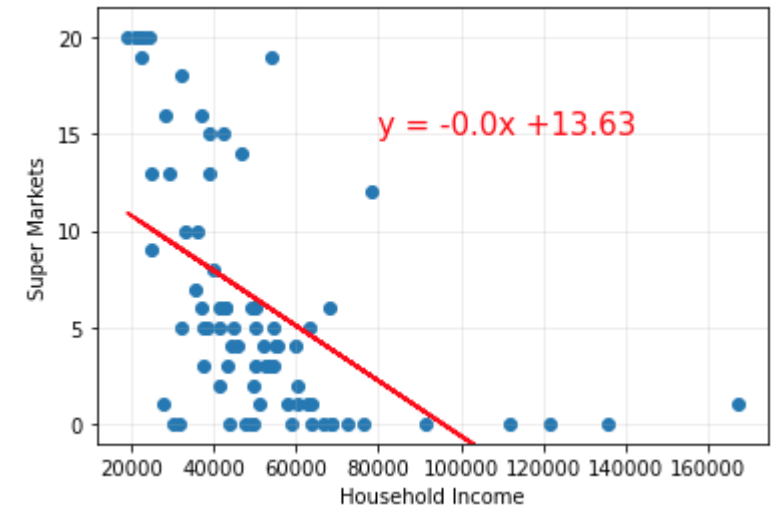
The r-squared is: 0.16975614410969947



The r-squared is: 0.04546798045365874



The r-squared is: 0.29100930450674134



Data Analysis

Super Markets, Gyms and Liquor Stores by Zip Code

1. Do people in higher income brackets have more access to supermarkets, gyms and liquor stores?

After seeing such a big difference in the number of locations (supermarket, liquor store and gyms), we broke up the income levels into bins reflecting the tax brackets. In this analysis, we can definitely see the difference in number of locations by bracket. Incomes above \$20K and below \$139K show the most number of locations.

We would also recommend going a step deeper and breaking up the 2nd and 3rd bracket in to smaller bins to get a more accurate representation of the data.

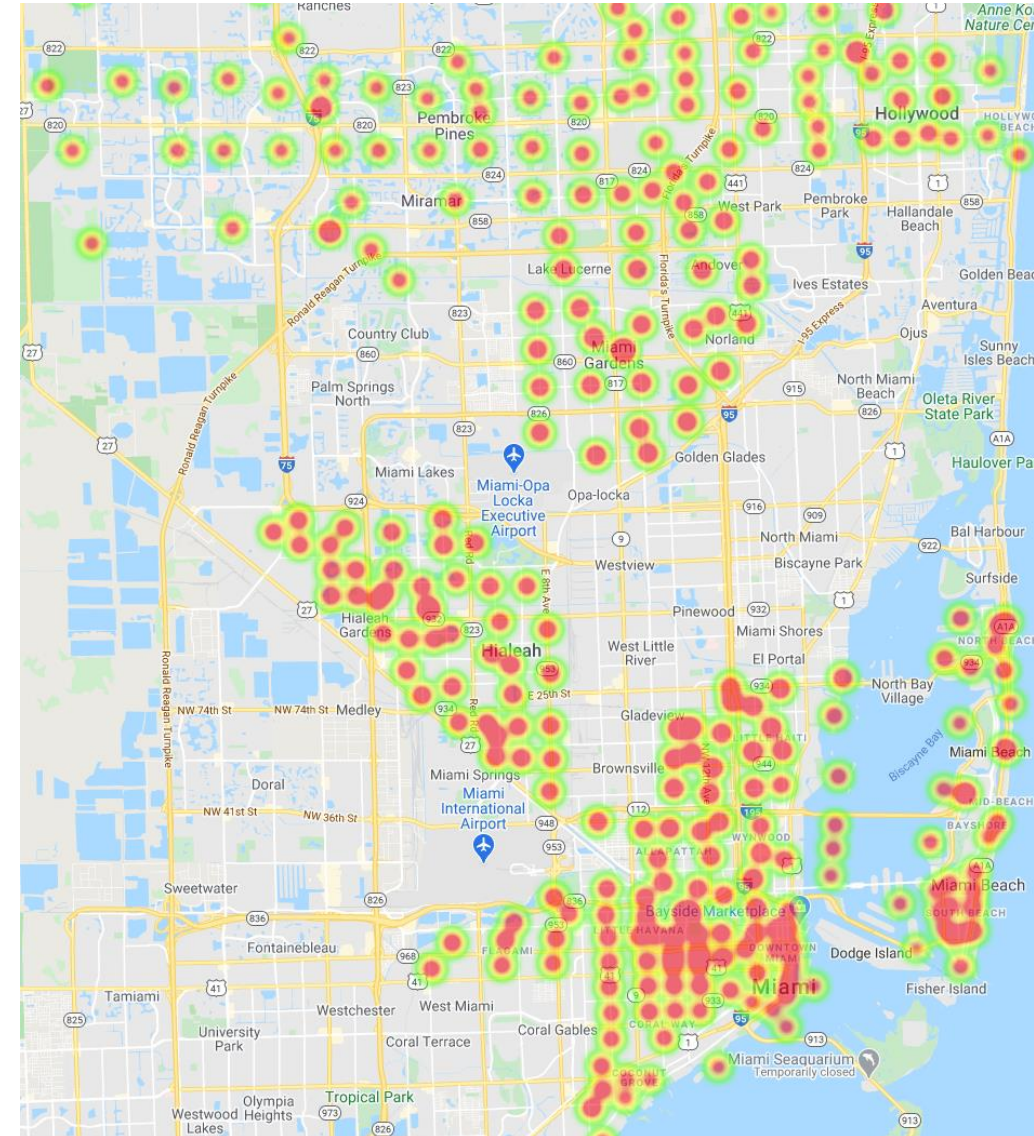
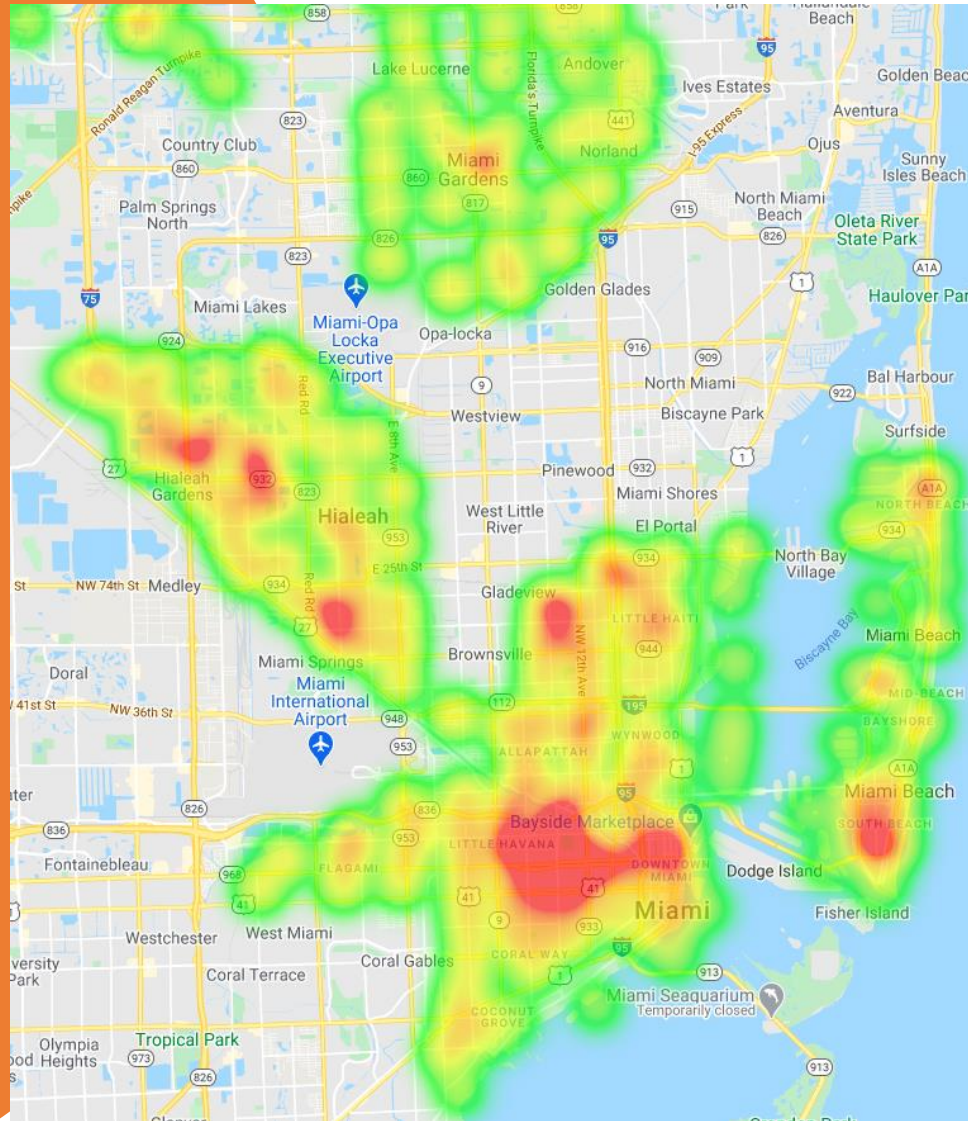
We would also only focus on income levels below \$100K.

	Total Gyms	Total Super Markets	Total Liquor Stores
Income Range			
<20,000	20.0	20.0	20.0
20,000 - 44,999	387.0	364.0	304.0
45,000 - 139,000	364.0	119.0	203.0
140,000 - 149999	0.0	0.0	0.0
150,000 - 199999	5.0	1.0	3.0
200,000+	0.0	0.0	0.0

Unhealthy Behavior and Obesity by Zip Code

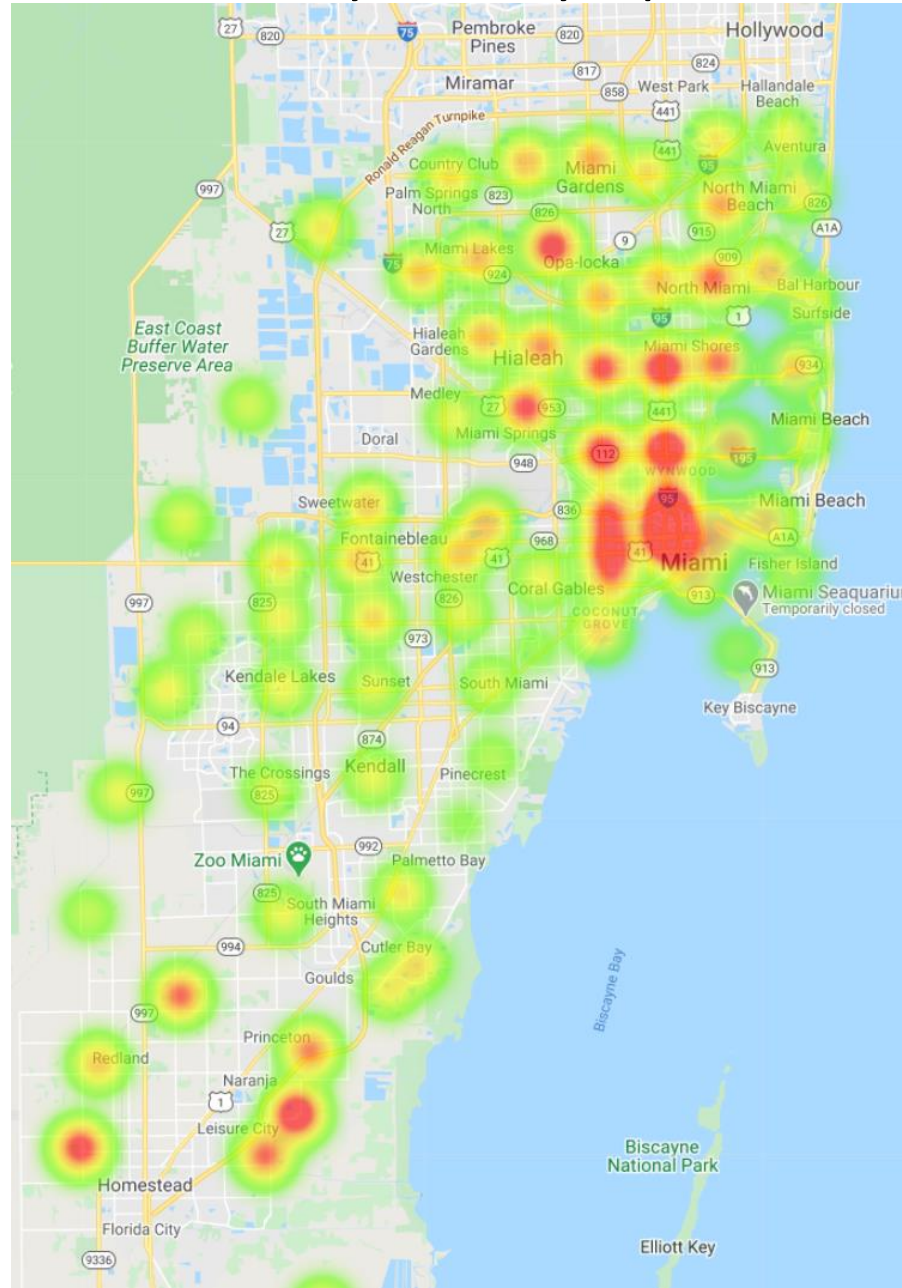
CDC health data reported by geolocation. Not every zip code/neighborhood in Miami Dade reported.

Data Analysis



Data Analysis

Poverty Rate by Zip Code



2. Do people in higher income brackets have healthier lifestyles (focusing on obesity and unhealthy behaviors as metrics)?

After looking at all three maps, we can definitely see that there is overlap between unhealthy behavior/obesity and poverty rate. We can see this overlap especially in areas such as Downtown/Over town/ Hialeah areas.

We did not find a direct or indirect correlation in the data. As previously mentioned, we were not able to merge the data frames together due to issues with geocoding and zip codes.

Discussion

Final conclusion:

- We can see the information displayed in heatmaps and assume that income level has a direct affect on people lifestyle. For example, in Hialeah, Downtown and Over town, we see higher index of unhealthy lifestyle choices where poverty rate is high. However, we do not have a direct correlation.
- CDC data was not accurate and detailed enough. In general, we did not have enough data to make a good conclusion. Additionally, our data sources were from different years (2013 vs. 2017)
- Limiting factors included:
 - Health data is from highest population areas in Miami, but did not include every zip code.
 - Miami is particular city as we see extremely high and low incomes in the same county. There is a big difference in people's lifestyle and income levels just a few miles from one another. We also see a lot of private properties that include gyms and are not account for in Google Maps.
- Our data and analysis did provide insight into the current situation of Miami Dade County, however, it didn't specifically answer the questions we asked.
- We saw a moderate correlation between super markets and income level (R-squared of 0.29). We saw a low to weak correlation between income and liquor store/gym (R-squared of 0.16 and 0.04 respectively).
- We would recommend further analysis and data mining to answer our question more accurately.

Post Mortem

Difficulties:

- The biggest difficulty we experienced was finding health data that was available for the entire county. Additionally, the CDC reports a variety of information related to health, therefore finding the specific metrics that we wanted to explore was difficult.
- We focused on the heatmaps to show areas with the biggest issues. We would have liked to find specific positive or negative correlation numbers that could answer our questions more accurately.

If we had more time?

- We would have worked on the reverse geocoding – finding the zip code based on geolocations. This way we could have done more analysis along with the income data frames.
- Additional resources and information available for health
- Possibly include Broward county to get a more accurate representation of trends in South Florida.



Questions