

# **A Study of the Correlation Between Wealth and Health Based on Geographic Location**

Analysis Performed By: Michael Bass, Ki Yeob Lee, Shirish Pandagare, and Ali Shafi

## Introduction

Researchers have demonstrated that there is a strong correlation between the wealth and health of communities. To put it to the test, our team decided to investigate this relationship. In our investigation, we used income to assess the socioeconomic status and likewise used diabetes and obesity as indicators of overall health. We were able to analyze data at the zip code and country levels to understand how geographic areas are impacted by wealth. We analyzed not only the direct correlation between health and wealth, but also tried to analyze the underlying factors such as food habits and exercise. Our ultimate breakthrough came in analyzing the income breakdown between each zip code. Ultimately we concluded that not only is there a strong correlation between wealth and health, but wealth directly impacts the underlying factors of food and physical activity.

## Data Set

For our analysis, we extracted data from the tax statistics for 2016 which is publicly available Internal Revenue Service's (IRS) website [1]. The dataset contains many tax related statistics from 2016, including income, but also deductions, dividends, tax credits, etc. The data is aggregated by zip codes, and additionally each zip code is broken into 6 income slabs. We collected health data, in the form of diabetes and obesity rates, from the United States Department of Agriculture (USDA) [2]. This data contains information related to health, physical activities, different forms of grocery store availability, fast food and restaurant statistics, etc. The data from the USDA is aggregated at the county level.

In order to begin our investigation, we had to first determine how to combine the IRS and USDA datasets, which had varying levels of geographic granularity. Each county is uniquely identified by a five-digit Federal Information Processing Standards (FIPS) code. Through additional searching we found processes and data linking the county FIPS with the zip codes contained in each county. In order to connect

the county and zip code data, we analyzed county data as percentages, assumed that the percentages were evenly distributed through the county's zip codes, and applied those percentages to each zip code. Unfortunately, the IRS data was not present for all zip codes, and hence those zip codes were not included in our study.

A major part of our investigation process was to clean and segregate data. We cleaned data as needed using python and joined and visualized data using a software called Tableau.

## Diabetes and Obesity

The very first step of our investigation was to establish the link between diabetes and obesity. While this is a well known fact, we had to look at it for ourselves. Looking at Figure 1, this correlation is extremely evident.

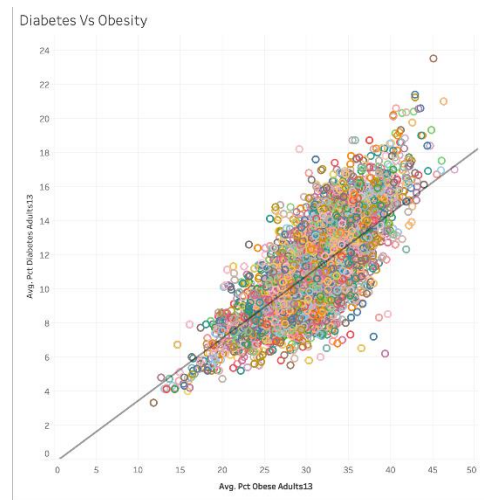


Figure 1: Obesity vs Diabetes

## Income, Diabetes, and Obesity

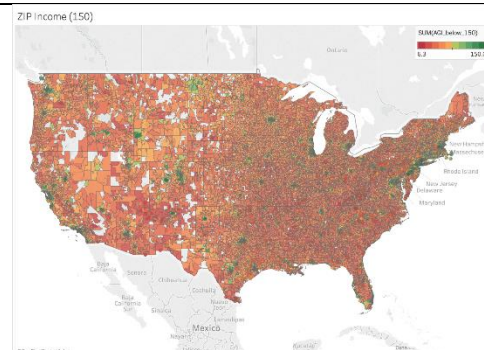


Figure 2: US Income Distribution

For the next step of our investigation, we wanted to see the geographical distribution of income, diabetes, and obesity. Looking at Figure 2, we can start to understand the distribution of wealth, in the form of adjusted gross income (AGI), in the United States (US). Because of the few outliers, we have capped the AGI at 150 thousand dollars, to better visualize the data. From Figure 2 it is most evident that wealth is concentrated in the large US urban centers.

To continue, let's see how diabetes and obesity are geographically distributed. Looking at Figures 3a and 3b we see some similarities. Clearly, the US Deep South has a strong concentration of diabetes and obesity, but it is hard to visually see a connection elsewhere.

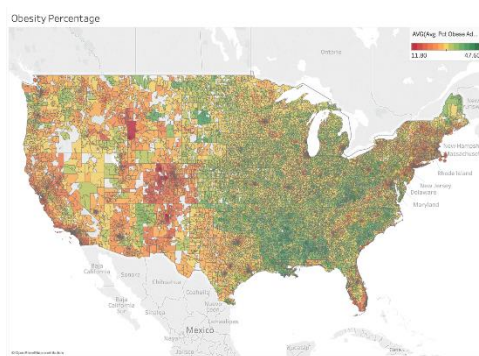


Figure 3a: US Obesity Distribution

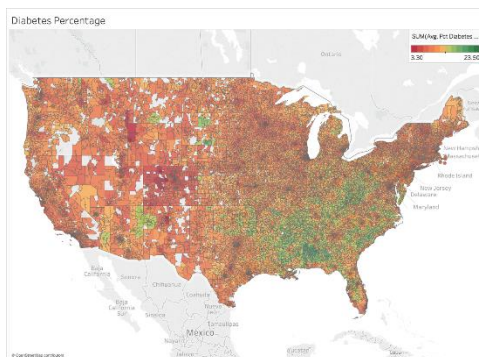


Figure 3b: US Diabetes Distribution

Comparing the geographical visualizations of wealth and health it is challenging to draw a conclusion. This is primarily due to wealth being so concentrated in urban areas. Although these visualizations have helped us gain some intuition, we must continue to dig into the data to determine a relationship.

## Wealth VS. Income

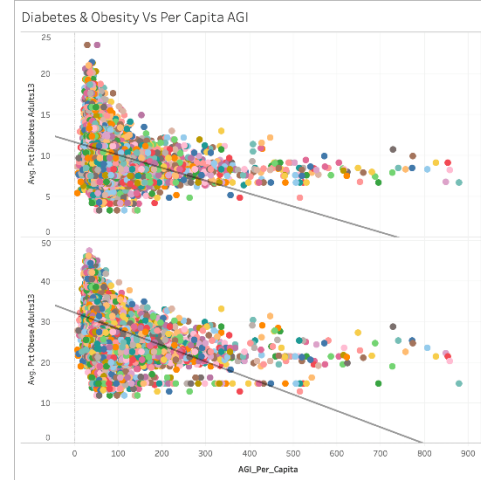


Figure 4: AGI vs Health

To determine if there is a correlation between wealth and health, we first investigated the relationship between income, in the form of adjusted gross income (AGI) and health. To visualize the relationship, we graphed the data, as well as a linear regression in Figure 4. For many of the relationships we investigated we visualized the data as scatter plots with each zip code or county being represented as one dot on the graph. The regression line shows a clear relationship between the AGI per capita and health. However, some aspects of the data make it appear as though the line may not be telling the true story. The points on the right side of the graph appear to not be decreasing, but only going further to the right. Additionally, the bottom left of the data seems unrelated to the linear fit. One thing we must keep in mind is that each plot contains almost 30,000 points, so the data density is challenging to understand. Based on this result alone, we don't think anything can yet be concluded, and we therefore must dig deeper.

## Poverty Rate VS. Health

We decided to next analyze the relationship between poverty rates for zip codes and the health indicators, visualized in Figures 5 and 6. Visually, this appears to have a much stronger relationship, and the linear regression agrees. Clearly as the poverty rate increase, so does the

prevalence of obesity and diabetes. While this shows that poorer locations are correlated with lower health, does being wealthier truly correlate with better health?

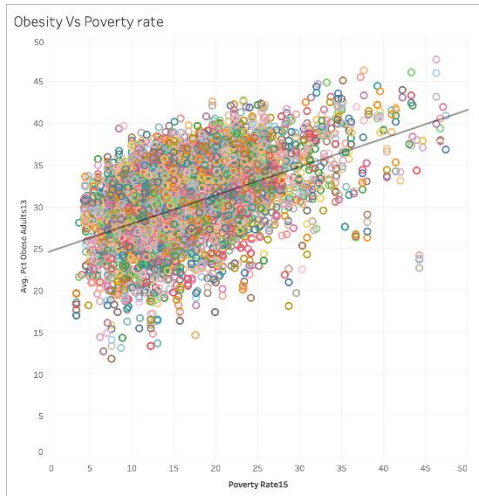


Figure 5: Poverty Rate vs Obesity

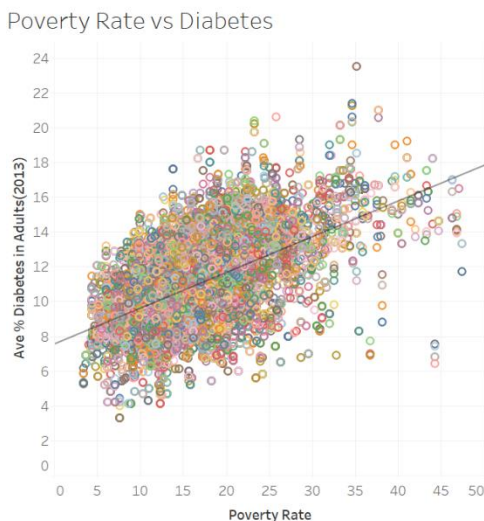


Figure 6: Poverty Rate vs Diabetes

### Median Income VS. Health

To answer this question, we tried to compare median income with health. After all, if the median income of an area is higher, the area as a whole must be wealthier. We can see that as the median income increases health risks decrease. The graph presents a pretty clear representation of this. However, the decline in

health risks as median income increases is not extremely rapid. And we must ask the question, does median income really indicate a wealthy area? With a measure such as the median, there are several different ways in which a location might have a given median. This was a step in the right direction, but maybe it doesn't completely answer the question.

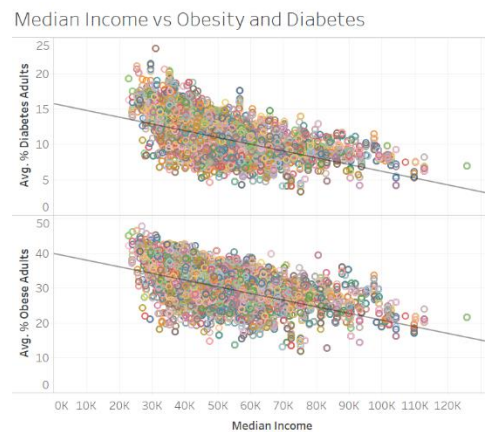


Figure 7: Poverty Rate vs Diabetes

### Slab Income VS. Health

Fortunately, the data from the IRS contained data divided into “slabs” for each zip code. Each slab represents a different income range. The different ranges are:

- Slab 1 = \$1 - \$25,000
- Slab 2 = \$25,000 - \$50,000
- Slab 3 = \$50,000 - \$75,000
- Slab 4 = \$75,000 - \$100,000
- Slab 5 = \$100,000 - \$200,000
- Slab 6 = \$200,000+

We decided to calculate the percentage of each zip code that falls into each slab. Then with this data we could analyze how diabetes and obesity relate to zip codes on a per-slab basis. This should give us a better idea of how income and health are truly related. Figure 8 shows our results of these efforts.

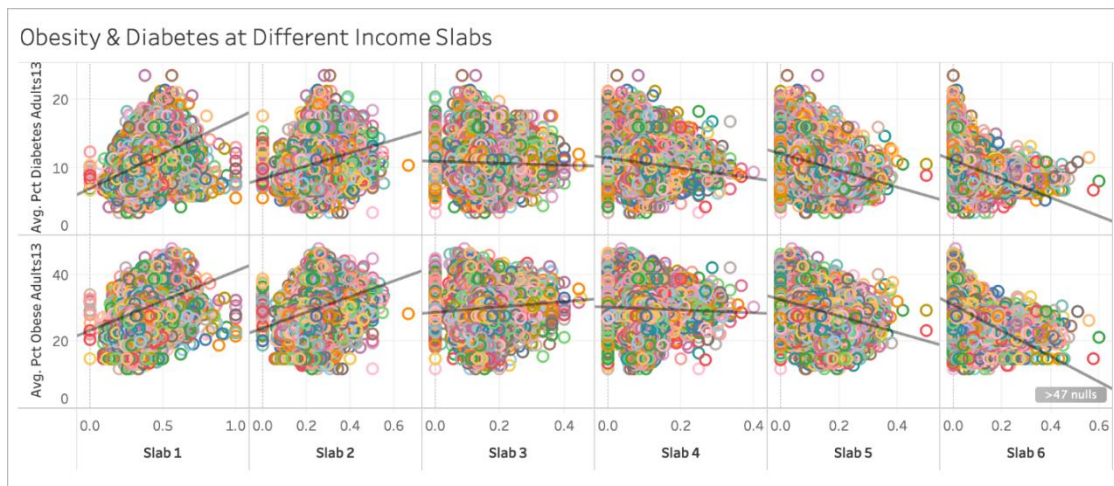


Figure 8: Income Per Slab vs Health

This result shows how health relates to what percentage of the zip belongs to a given slab. To better understand this, let's look at the data for slab 1. On the bottom axis we have graphed the percent of the zip that belongs to slab 1, the poorest slab. As the percent of the zip in slab 1 increases, the health risks also increase. Therefore, as more of a location is very poor, we see the area being less healthy. Each slab tells a story, and as we progress through the slabs we begin to understand the effect of wealth. As areas have greater populations in slabs 1 and 2, the areas have worse health. But, when locations have a greater percentage belonging to slabs 5 and 6, the wealthiest slabs, the health risks decrease. And in the middle, slabs 3 and 4, health is almost uncorrelated with these populations. This tells us that the poorer an area is, there is a good chance that it will be less healthy, and the wealthier an area is, it is much more likely to be healthier. In our opinion, this piece of evidence solidifies the connection between wealth and health. But certainly just having money doesn't make someone healthier, that simply is illogical. We need to look at more of the factors behind being wealthy that link wealth to better health.

### Lifestyle VS. Health

As we already discussed, we found a very strong correlation between obesity and diabetes. Additionally, countless studies have indicated

that obesity is the major cause of diabetes. Obesity does not only lead to diabetes but also to other health problems such as high blood pressure, heart diseases, joint problems, sleep apnea, cancer, metabolic syndrome and psychosocial effects [1]. Hence, obesity is a major factor while determining the health index of any population.

As we stated, we found a very strong correlation between obesity and wealth in the US. But, what is the cause of this correlation? To gain insights about what can be the root cause of obesity rates we decided to investigate the food habits and physical activity.

### Food Habits

Over the past few decades there have been a significant rise in fast food chains across the country, even in the most remote locations. Our hypotheses, before plotting the data, were that 1) there are more fast food restaurants in slab 1 and 2 than full service restaurants since the latter is pricier than the former, and 2) people in slab 1 and 2 will buy more fast food than full serviced food.

In Figure 9, the first row shows number of fast food restaurants/thousand people, and the second row shows sales in fast food restaurants per capita. Similarly, Figure 10 represents full service restaurants. Although the trends in slab 1 and 6 are slightly different, trends for number



of fast-food and full-service restaurants show that people in all slab have almost equal access to both types of restaurants.

Trends for the sale in fast-food and full-service restaurant clearly tell that people in slab 1 and 2 spend more money on fast food restaurants as compared to people in slab 5 and 6 who spend more money on full service restaurants. Although it is not easy to argue that what you eat is the leading contribution to the obesity and

diabetes from the data, it can be thought of a discriminating factor from our experience. Plus, this difference could be simply different taste buds that people in slab 5 and 6 have money to try variety of different menus whereas people in slab 1 and 2 do not have enough money to try different diet. Nevertheless, people in slab 5 and 6 clearly spend more money on full service restaurants than groups in slab 1 and 2 which can potentially results to healthier lifestyle.

Income slabs vs Fast-foods and expenditure/capita

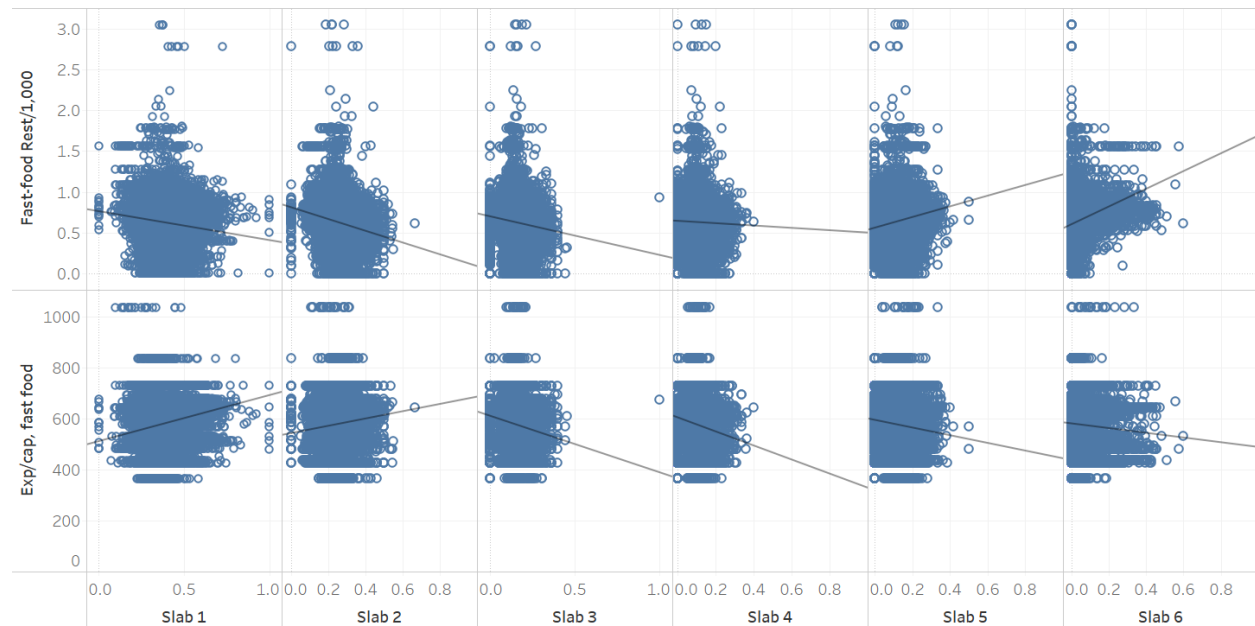


Figure 9: Fast Food Prevalence and Amounts Spent on Fast Food

Income slabs vs Full service restaurants and expenditure/capita

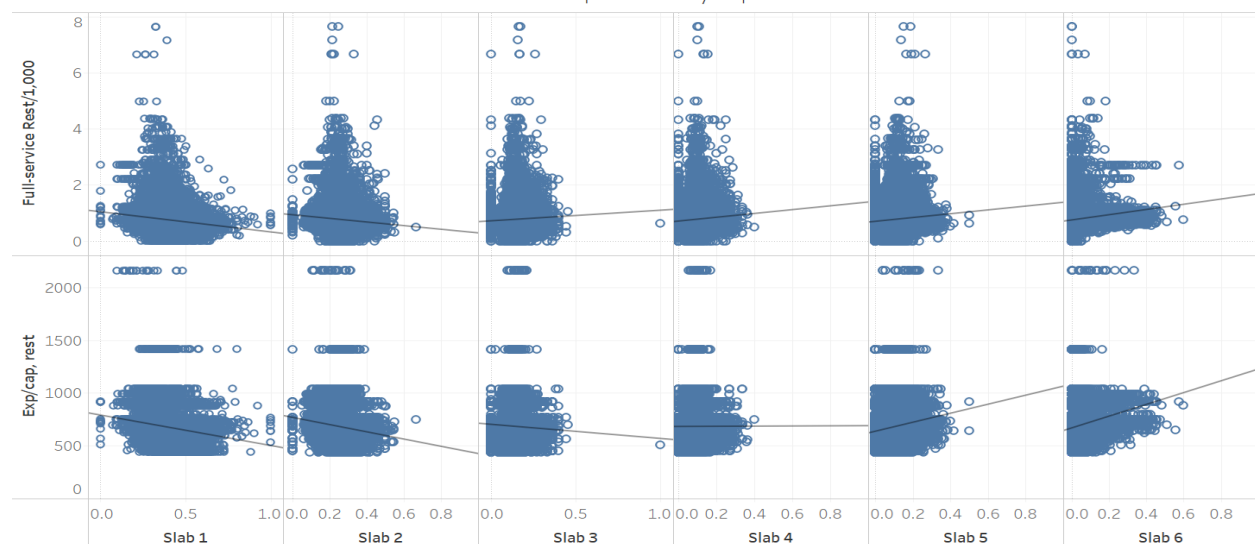


Figure 10: Full Service Restaurant Prevalence and Amounts Spent on Full Service Restaurants

Another hypothesis was that people in low income groups would purchase groceries/food in limited quantity. On the contrary, people with high income can stock up groceries in huge quantities. Thus, establishing a greater number of grocery stores in regions with high slabs makes more sense. In Figure 11 we see that as the number of grocery stores increases the diabetes and obesity rate decreases since people have more options to buy healthy food. On the

contrary to this as the population in lower income groups increases, convenience stores are present in higher numbers as compared to locations with higher concentrations of rich people, as seen in Figure 12. Convenience stores often sale food for instant gratification, such as junk food, soda, candies, etc., which adds to an already existing unhealthy diet for lower slab counties and thus fueling diabetes and obesity rates.

Accessibility to food vs Obesity and Diabetes Rates

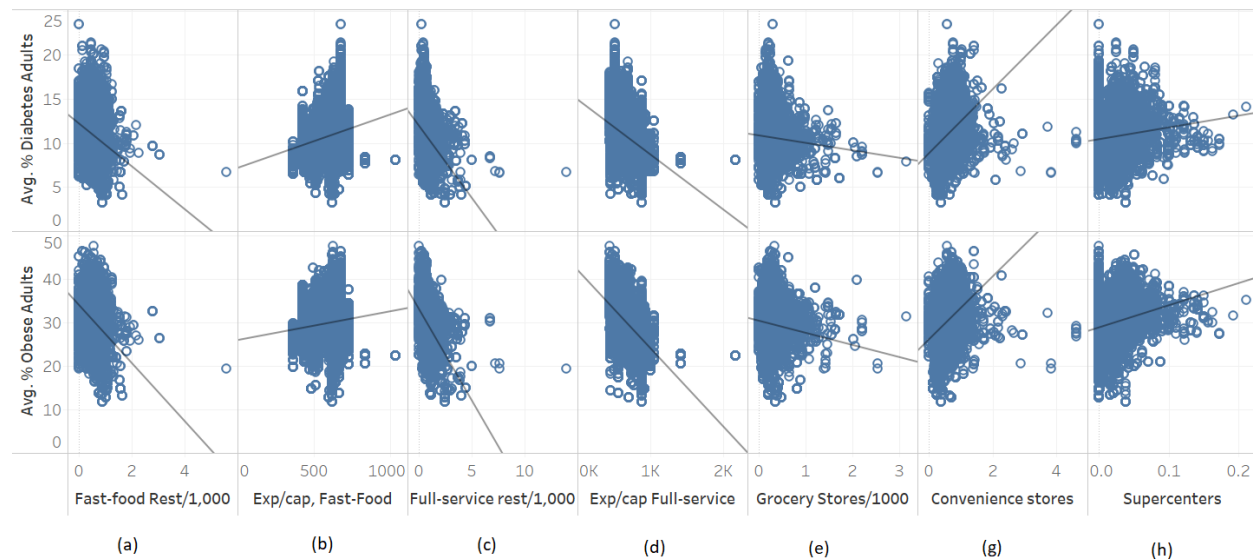


Figure 11: Different Types of Stores vs Health

Income slabs vs No. of Convenience stores

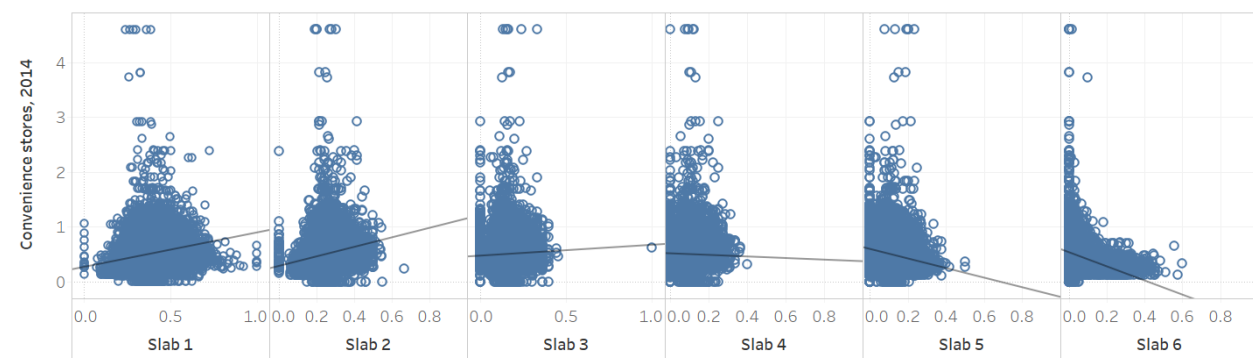


Figure 12: Convenient Store Prevalence vs Slabs

## Physical Activity

The last link we need to investigate is how physical exercise impacts health, and does physical activity relate to wealth? We believe that lower amounts of physical activity directly

lead to higher obesity rates, and this is confirmed by Figure 13. Looking at the figure, we can see that as the number of recreation facilities per 1000 people increase, the health risks are decreasing. But is this related to wealth?

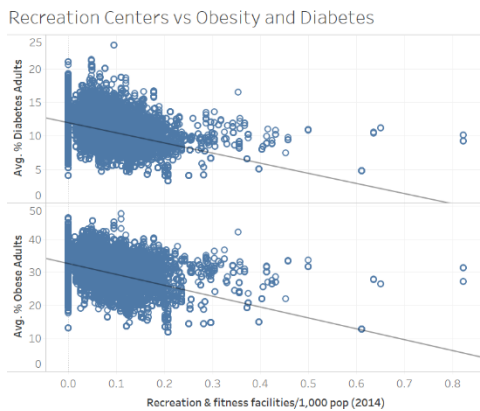


Figure 13: Physical Activity vs Health

Taking it a step further, we analyzed how recreation is related to the different income slabs. You can see the results of our comparisons in Figure 14. For the number of recreational centers, we see trends similar to full service restaurants across the income slabs. Zip codes

with a high percentage of lower income, such as slabs 1 and 2, have less recreational centers, thus restrict the poorer areas from access to physical activities. On the contrary, as a zip code becomes wealthier, as in slabs 5 and 6, the area gets more access to recreational activities. We attribute this to that gymnasiums or recreational businesses being more likely to establish themselves in wealthier areas that can afford their monthly membership, in contrast to poorer areas that would consider the memberships too expensive.

With this analysis of food and physical activity we think it clearly evident that wealthier areas have access to better food and physical activity, which ultimate lead to them having lower rated of diabetes and obesity. The evidence is clear, and the data speaks for itself.

## Recreation and Fitness facilities for each slab

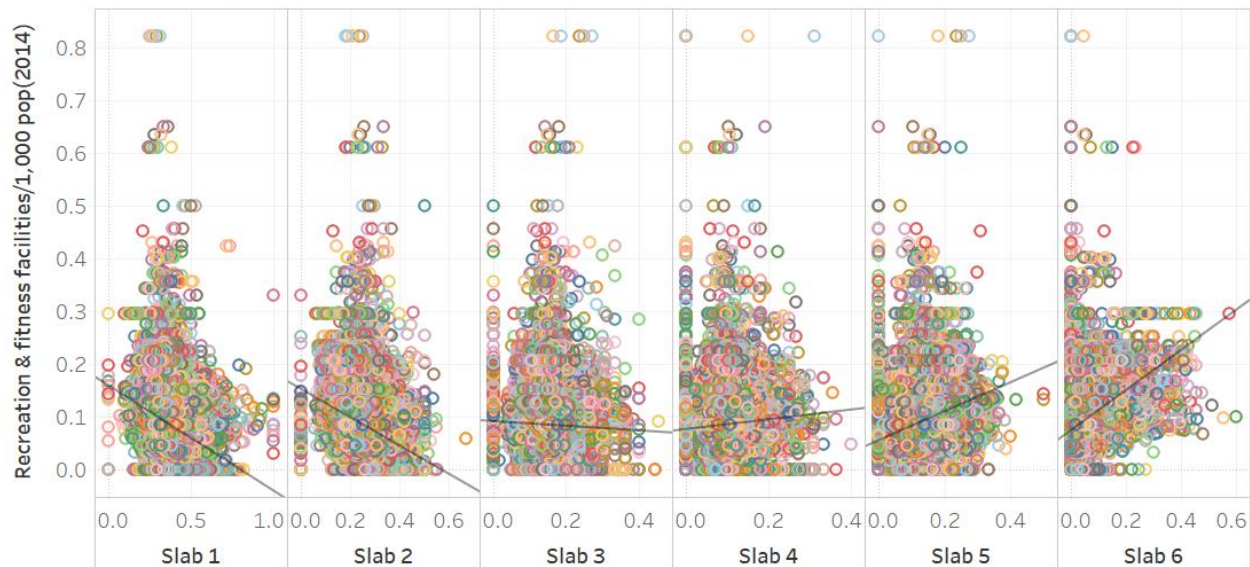


Figure 14: Slabs vs Available Recreation

## Tax Credits

One final angle we considered in our investigation was how wealth and tax credits for medical insurance are related. We know that people with low income receive government support for health insurance, but how effective are the programs?

Looking at Figure 15, we compared the medical tax credits received for each slab and compared them to the obesity and diabetes rates. The most insight can be gained from looking at slabs 1 and 2. For slabs 3 and higher, it is not surprising to see individuals in these tax divisions not receiving many, if any, of the tax credits. For slab



1, as a higher percent of a zip code is in slab 1 and receives a tax credit we see the rate of diabetes increase, while obesity remains nearly constant. However, in slab 2, we actually see the diabetes and obesity rates decrease as more people receive tax credits. This tells us that the

poorest individuals are still not able to improve their health with the tax credit, while those a little better off are definitely able to. Again, this reflects that having even some money compared to those making very little allows individuals to better care for their health.

Tax Credit for each slab vs Obesity and Diabetes

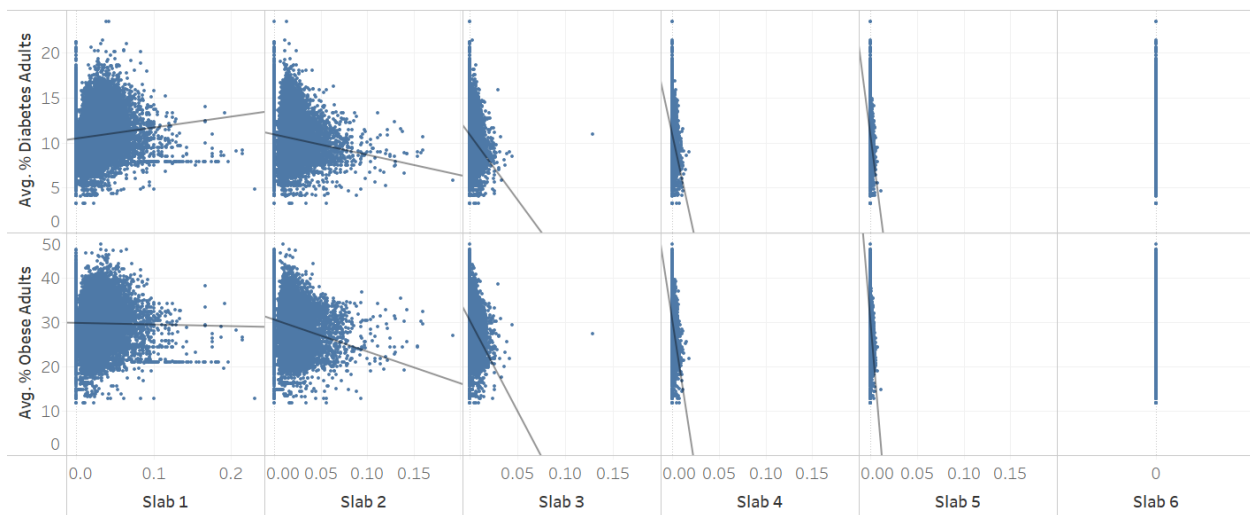


Figure 15: Medical Tax Credits Per Slab vs Health

## Conclusion

This report has studied several topics, all with the goal to see if we could find a link between wealth and health of an area. We have observed that there is a positive correlation between health, represented by diabetes and obesity, and wealth, but further questions needed to be addresses. What are the major factors that lead to obesity and diabetes, and were those factors also related to wealth? Is it the food people eat, or does exercise matter? Do government subsidies help improve health? We have tried to find answers and let the data dictate our journey. To recall, we found the greatest

answers by looking at populations based on the IRS allocated income slabs. Zip codes with a high percentage of lower income had higher rates of diabetes and obesity, and we found that they spent more on fast food and had far less access to recreational gyms. However, areas with a larger percentage of higher income individuals had lower rates of diabetes and obesity, and additionally spent more on full service restaurants and had greater access to gyms. It is clear that the access to healthier options and better work out facilities leads to a healthier lifestyle, and ultimately leads to wealthy areas having lower health risks.

## References

- [1] IRS - <https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-2016-zip-code-data-soi>
- [2] USDA - <https://www.ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads/>
- [2] ALI - <https://stanfordhealthcare.org/medical-conditions/healthy-living/obesity.html>