Analyzing Health Disparities: The Relationship Between Obesity, Diabetes, and Environmental Factors in U.S. Counties

Itzel Vázquez, Emiliano Zarza, Carlos Sánchez

November 19th, 2024.

Content

Introduction

1. **Obesity and Diabetes in Metropolitan vs. Population Loss Counties**
   1.1. Descriptive Analysis of Variables
   1.2. Regional Patterns of Obesity and Diabetes Across the United States
   1.3. Linear Regression Analysis
   - Metropolitan Counties
   - Population Loss Counties
   1.4. Why the Null Hypothesis is Rejected

2. **Impact of Fast Food on Different Age Groups**
   2.1. Data Aggregation
   2.2. Analysis of Relationships Between Diseases and Fast Food Restaurant Density by Age Groups
   2.3. Linear Regression Analysis
   - Population Over 65 Years Old
   - Population Under 18 Years Old
   2.4. Main Findings

3. **Relationship Between the Number of Physical Activity Facilities and Obesity/Diabetes Rates**
   3.1. Descriptive Analysis of Variables
   3.2. Analysis of Relationships Between Physical Activity Facilities and Health Outcomes
   - Linear Regression Analysis
   - T-test Analysis
   3.3. Main Findings

Source

Introduction

In this report we use data from the Food Environment Atlas of the U.S. Department of Agriculture to analyze trends in Health. The main goal of this analysis is to find disparities in counties of the US in the relationship between obesity and diabetes rates and factors like rural/urban areas, age, and number of fast food restaurants and physical activity facilities in the county.

- **Data cleaning:**

We worked on a Jupyter Notebook and used a Python Kernel. The Atlas was in an xls file. We selected three Sheets: HEALTH, SOCIOECONOMIC, and RESTAURANTS. The key variable we used was the FIPS identificator, which was a unique number for each county.

We drop the observations (counties) that had NA values. Initially, the Atlas had 3143 observations and, after the cleaning, the dataframe was left with 2213. The data was saved in a csv file. The three analyses contained in this report used that same csv file (n=2213).

1 . Obesity and diabetes in Metropolitan vs Population Loss Counties

This section of the report delves into the relationship between obesity and diabetes rates across metropolitan and population loss counties in the United States. By analyzing key health indicators, we aim to uncover meaningful insights that highlight potential disparities between urban and rural contexts.

The analysis begins with a descriptive examination of the data, addressing the question:

- **"How does the rate of obesity and diabetes compare in population loss counties versus metropolitan counties?"**

To set the stage, we present **regional patterns** of obesity and diabetes across the United States, offering a broad perspective on the geographical distribution of these health issues.

Finally, we perform a **linear regression analysis** to explore the relationship between obesity and diabetes rates in both county types, comparing the strength of the relationship and identifying key trends that could guide future health interventions.

1.1 Descriptive analysis of the variables

This analysis focuses on two critical health indicators: diabetes rates in 2013 and obesity rates in 2017, comparing their distribution and variability between two county categories: metropolitan counties and population loss counties. These statistics provide a foundation for understanding the differences in health outcomes across urban and rural contexts.

Comparison Between Metropolitan and Population Loss Counties

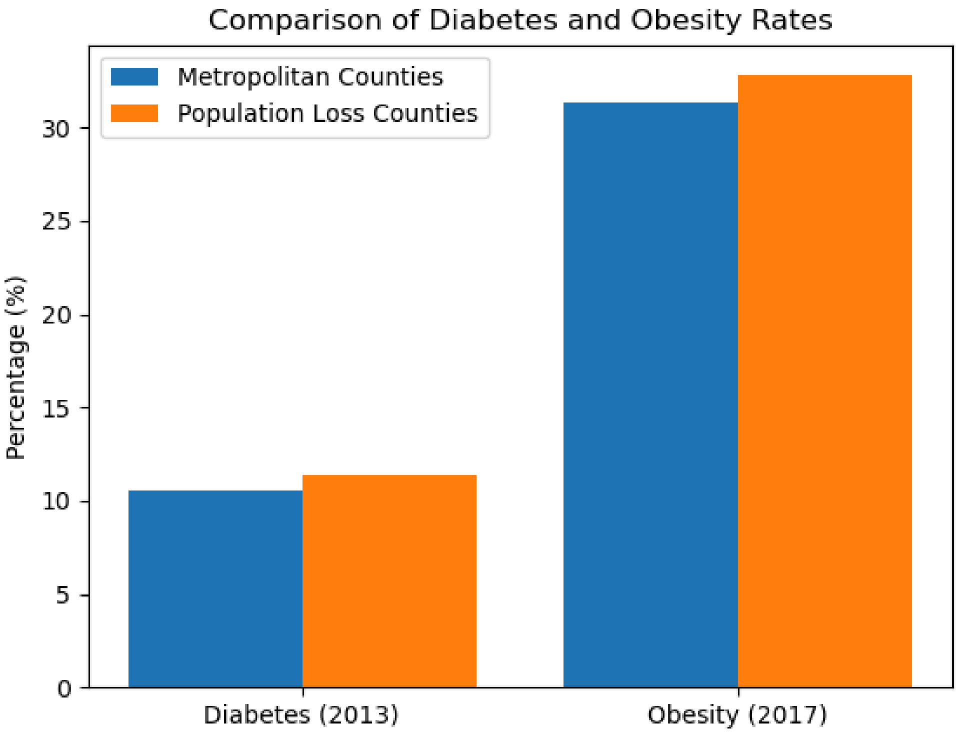
Metric	Metropolitan Counties	Population Loss Counties
Mean Diabetes Rate	10.57%	11.38%
Mean Obesity Rate	31.28%	32.80%
Diabetes Variability (Std)	2.25%	1.91%
Obesity Variability (Std)	3.41%	3.07%
Minimum Diabetes Rate	4.10%	6.50%
Maximum Diabetes Rate	18.60%	19.20%
Minimum Obesity Rate	22.60%	22.60%
Maximum Obesity Rate	38.10%	38.10%

Source: Food Environment Atlas, U.S. Department of Agriculture.

Population loss counties have higher average rates of diabetes (11.38%) and obesity (32.80%) compared to metropolitan counties, which have averages of 10.57% and 31.28%, respectively. These differences suggest that rural areas face slightly worse overall health outcomes. However, diabetes rates in metropolitan counties show greater variability, with a standard deviation of 2.25%, compared to 1.91% in population loss counties. This indicates that urban areas have a wider range of health outcomes, potentially influenced by diverse socioeconomic conditions and healthcare access.

The ranges of diabetes and obesity rates further highlight these differences. Metropolitan counties exhibit diabetes rates spanning from 4.10% to 18.60%, while population loss counties range from 6.50% to 19.20%. For obesity, both county types show a similar range, from 22.60% to 38.10%, suggesting comparable variability in obesity regardless of the county type. These findings imply that while population loss counties face worse averages, the variability in metropolitan counties could signal disparities within urban populations that deserve further investigation.

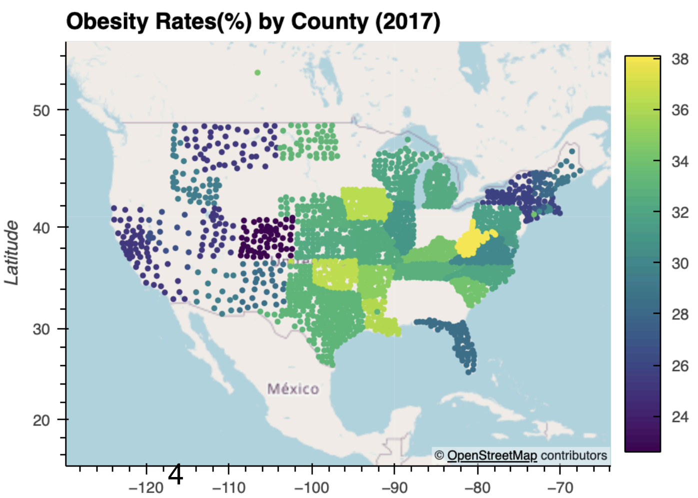
How does the rate of obesity and diabetes compare in population loss counties versus metropolitan counties?

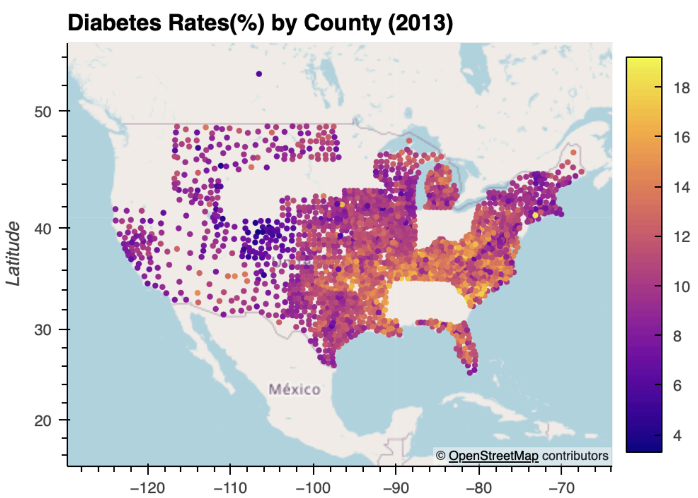


Diabetes and obesity rates are higher in population loss counties compared to metropolitan counties. This shows that rural areas face unique health challenges that are different from those in urban regions. The gap between the mean and median diabetes rates in population loss counties suggests a slightly wider range of diabetes levels, meaning some rural counties might have very high or very low rates.

The higher diabetes and obesity rates in population loss counties may be caused by systemic differences. Rural areas often have less access to healthcare facilities, which can delay diabetes diagnosis and treatment. Lifestyle factors, such as higher poverty rates, food deserts, and fewer opportunities for physical activity, also contribute to these problems. Additionally, lower levels of education and awareness about healthy habits and early disease detection in rural areas make these health issues worse. These factors show the need for special programs to reduce health risks in population loss counties.

1.2 Regional Patterns of Obesity and Diabetes Across the United States





The regional comparison of obesity and diabetes rates reveals distinct patterns across the United States. The Southeastern U.S. consistently shows the highest rates for both health indicators. In 2017, counties in states like Mississippi, Alabama, Arkansas, and Louisiana exhibited obesity rates exceeding 35%, forming clusters of high prevalence. Similarly, diabetes rates in 2013 were most concentrated in the Southeast, with states like Alabama, Mississippi, and West Virginia displaying significant overlaps between high obesity and diabetes rates. Parts of the Midwest, such as Missouri and Kentucky, also demonstrated elevated rates for both conditions, reinforcing regional disparities. In contrast, Western states like Colorado, Utah, and California, as well as parts of the Northeast, including Vermont and Massachusetts, consistently showed lower rates of obesity and diabetes, often below 26% for obesity and with comparatively low diabetes prevalence.

There is a strong geographic overlap between high obesity and diabetes rates, particularly in the Southeast and parts of the Midwest, emphasizing obesity as a key risk factor for diabetes. However, some outliers exist, where counties have moderate obesity rates but disproportionately high diabetes rates, potentially reflecting other factors such as healthcare access, genetic predisposition, or economic challenges. The broader geographic spread of diabetes rates compared to obesity may also highlight the influence of systemic issues like poverty and healthcare availability. Overall, the Western and Northeastern U.S. benefit from better access to healthy food, higher socioeconomic conditions, and infrastructure for physical activity, while the Southeastern U.S. faces significant public health challenges due to its clustering of high obesity and diabetes rates.

1.3 Linear Regression: Is there a significant correlation between obesity and diabetes rates in population loss counties compared to metropolitan counties?

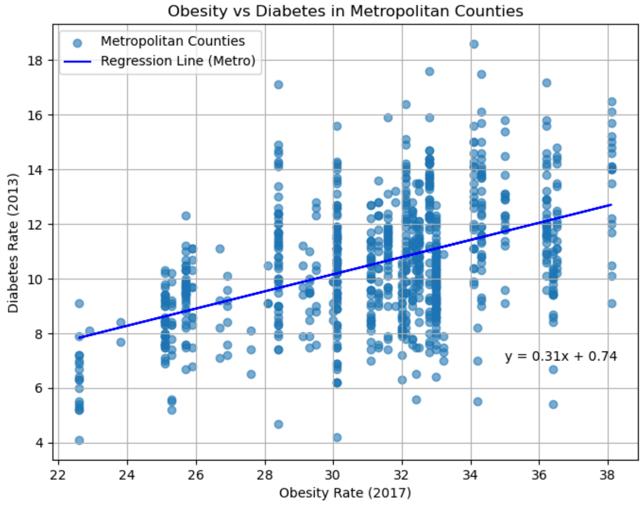
Hypotheses for Metropolitan Counties

- **Null Hypothesis (H_0):** There is no statistically significant relationship between obesity rates (2017) and diabetes rates (2013) in metropolitan counties.
- **Alternative Hypothesis (H_1):** There is a statistically significant positive relationship between obesity rates (2017) and diabetes rates (2013) in metropolitan counties.
- Decision Rule:
    - Reject the null hypothesis (H_0) if the p-value is **< 0.05**, indicating a statistically significant relationship.
    - Fail to reject the null hypothesis (H_0) if the p-value is **≥ 0.05**, indicating no statistically significant relationship.

Hypotheses for Population Loss Counties

- **Null Hypothesis (H_0):** There is no statistically significant relationship between obesity rates (2017) and diabetes rates (2013) in population loss counties.
- **Alternative Hypothesis (H_1):** There is a statistically significant positive relationship between obesity rates (2017) and diabetes rates (2013) in population loss counties.
- **Decision Rule:**
    - Reject the null hypothesis (H_0) if the p-value is **< 0.05**, indicating a statistically significant relationship.
    - Fail to reject the null hypothesis (H_0) if the p-value is **≥ 0.05**, indicating no statistically significant relationship.

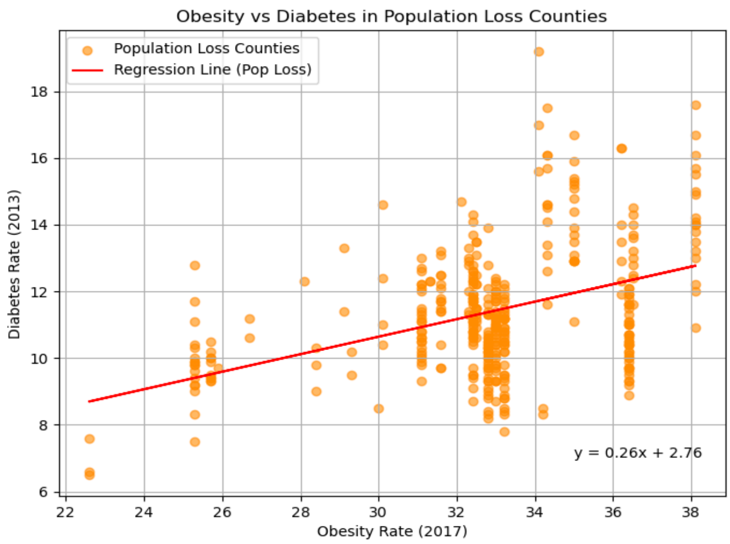
Linear Regression: Metropolitan Counties



Slope	0.3141
Intercept	0.7394
R^2 Score	0.2255
P-value	0.0000

The graph illustrates the relationship between obesity rates (2017) and diabetes rates (2013) in metropolitan counties. A positive correlation is observed, as shown by the upward slope of the regression line. The regression analysis reveals a slope of 0.3141, indicating that for every 1% increase in obesity rate, the diabetes rate increases by approximately 0.31%. The intercept is 0.7394, and the R^2 score of 0.2255 suggests that 22.55% of the variation in diabetes rates can be explained by obesity rates. The results are statistically significant (p-value = 0.0000), with a standard error of 0.0200, indicating reliable parameter estimates.

Linear Regression: Population Loss Counties



Slope	0.2626
Intercept	2.7629
R^2 Score	0.1772
P-value	0.0000

The graph illustrates the relationship between obesity rates (2017) and diabetes rates (2013) in population loss counties. A positive correlation is evident, as represented by the upward slope of the regression line. The regression analysis shows a slope of 0.2626, meaning that for every 1% increase in obesity rate, the diabetes rate rises by approximately 0.26%. The intercept is 2.7629, and the R^2 score of 0.1772 indicates that 17.72% of the variation in diabetes rates is explained by obesity rates. The results are statistically significant (p-value = 0.0000), with a standard error of 0.0292, signifying robust and reliable coefficient estimates.

Why the Null Hypothesis is Rejected

The null hypothesis, which posits that there is no significant relationship between obesity rates from 2017 and diabetes rates from 2013, is rejected based on the following findings from the analysis of metropolitan and population loss counties:

Statistical Significance:
The p-value for both metropolitan and population loss counties is 0.0000, far below the conventional threshold of 0.05. This indicates that the relationship between obesity and diabetes is highly statistically significant and unlikely to be due to chance.

Positive Slope:
The slopes of the regression lines indicate a clear positive relationship between obesity and diabetes:

- In metropolitan counties, a 1% increase in obesity correlates with a 0.31 percentage point increase in diabetes rates.
- In population loss counties, a 1% increase in obesity correlates with a 0.26 percentage point increase in diabetes rates.
  These results confirm that higher obesity rates are consistently associated with higher diabetes rates across both county types.

Relationship Strength:
The R^2 values highlight the proportion of variation in diabetes rates explained by obesity:

- In metropolitan counties, obesity accounts for 22.55% of the variation in diabetes rates.
- In population loss counties, obesity explains 17.72% of the variation in diabetes rates.
  Although additional factors influence diabetes, these R^2 values underscore obesity as a significant contributor.

Precision of Estimates:
The low standard error of the slopes—0.0200 for metropolitan counties and 0.0292 for population loss counties—indicates that the estimates are precise and reliable, enhancing confidence in the results.

Conclusion:
The null hypothesis is rejected because the data demonstrates a real, positive, and statistically significant relationship between obesity and diabetes. This evidence underscores the importance of addressing obesity as a key strategy in reducing diabetes rates, particularly in both metropolitan and population loss counties.

2. Are populations over 65 years old more protected or more vulnerable to the effects of fast food on health compared to those under 18?

The objective of the analysis is to assess whether there is a significant difference in how the **population over 65 years old** and the **population under 18 years old** are affected by the presence of **fast food restaurants** in their respective areas. Specifically, the analysis seeks to determine if one population group is more vulnerable or more protected from the health impacts (such as **obesity** and **diabetes**) associated with fast food consumption.

Key Questions:

- Are older populations (over 65) more protected from the effects of fast food on health, or do they experience more severe health outcomes (obesity, diabetes) in areas with more fast food outlets?
- Do younger populations (under 18) show different patterns in health outcomes (obesity, diabetes) related to fast food restaurant density?

Analysis Focus:

Obesity and Diabetes Rates: These are the two key health outcomes being studied, as they are strongly linked to the consumption of fast food and poor dietary habits.

- **Fast Food Restaurant Density**: This variable represents the number of fast food outlets available to the populations in different regions, which could influence dietary choices and health outcomes.

Approach:

The analysis uses correlation and t-tests to explore the relationships between the number of **fast food restaurants** and the **obesity** and **diabetes rates** for both population groups (those over 65 and those under 18). This helps determine if the impact of fast food on these health outcomes differs significantly between the two groups.

2.1. Data aggregation

To conduct this analysis, first, it was calculated the mean obesity rates, diabetes rates, the proportion of populations over 65 and under 18 years old, and the number of fast food restaurants across states. These aggregated measures were used to observe trends and relationships between these variables.

```
The total number of states is : 51
```

[130…

State	Mean Obesity	Mean Diabetes	Mean over 65 population	Mean under 18 population	Mean Fast Food Rests
AK	25.7	7.017241	8.573574	25.487328	13.275862
AL	33.0	15.489552	15.110755	23.477278	50.089552
AR	34.5	13.308000	16.686740	23.542034	25.000000
AZ	26.0	11.093333	16.523411	25.306010	273.800000
CA	25.0	8.774138	13.721435	23.656666	449.086207

Source: Food Environment Atlas, U.S. Department of Agriculture.

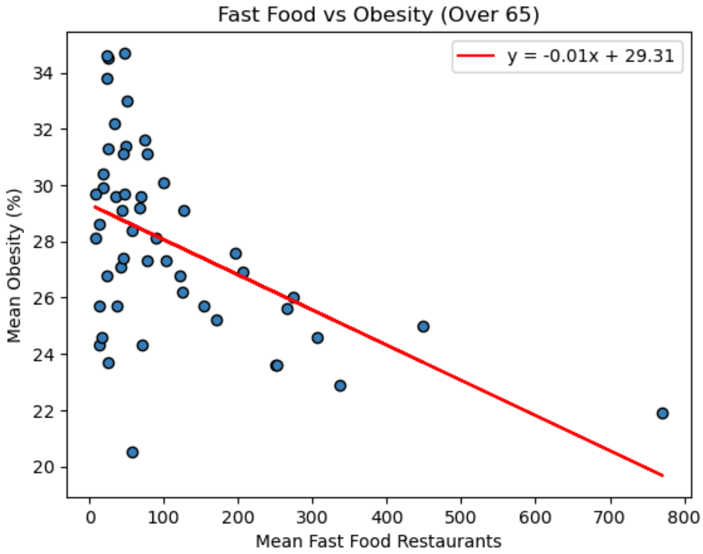
Groping the data allowed us to account for regional differences in health outcomes and fast food availability. Since the effect of fast food on obesity and diabetes may vary significantly from state to state due to factors like population demographics, healthcare policies, and socioeconomic conditions, grouping the data by state provides a more accurate and localized understanding of these relationships.

2.2. Analysis of Relationship between diseases and fast food restaurants density by age groups.

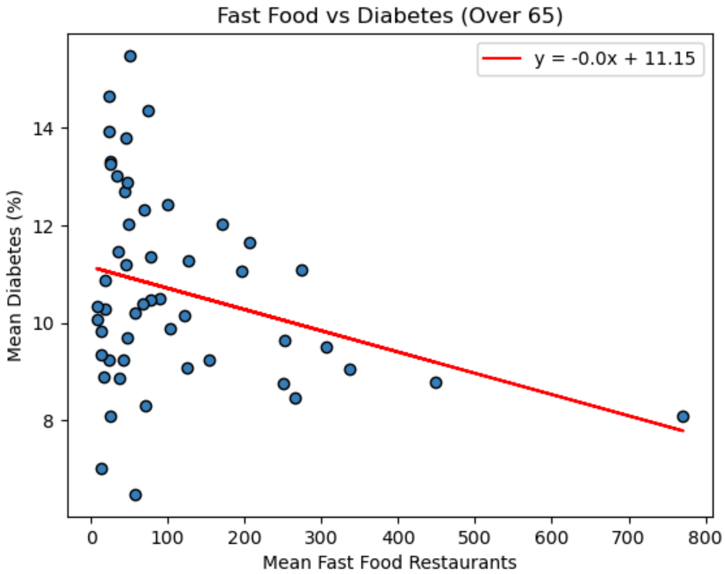
The linear regression analysis was conducted to explore the relationship between the density of fast food restaurants and the prevalence of obesity and diabetes for both the population over 65 years old and under 18 years old. This approach was chosen to quantify and visualize any potential correlation between fast food availability and health outcomes in these two age groups. By employing **linear regression,** we aimed to determine whether the impact of fast food on obesity and diabetes varied between these populations, thus providing insights into how age might influence susceptibility to the health effects associated with fast food consumption

Linear relationship for the population over 65 years old

The r^2 value is: 0.26



The r^2 value is: 0.09



<u>Obesity relationship</u>

The slope of the regression line, -0.01, suggests a very slight negative relationship between the number of fast food restaurants and obesity rate. For every additional fast food restaurant, the obesity rate decreases by 0.01%, according to the model.

The intercept value of 29.31 indicates that if there were zero fast food restaurants, the predicted obesity rate would be 29.31%.

The R^2 value of 0.26 implies that only 26% of the variability in the obesity rate can be explained by the number of fast food restaurants. This indicates a weak explanatory power and suggests that other factors not included in the model significantly influence obesity rates.

The regression model demonstrates a minimal negative association between the number of fast food restaurants and obesity rates among individuals over 65 years old. However the relationship is weak, as evidenced by the low R^2 value.

The analysis does not provide strong evidence to support a substantial impact of the number of fast restaurants on the obesity rate in the over-65 population.

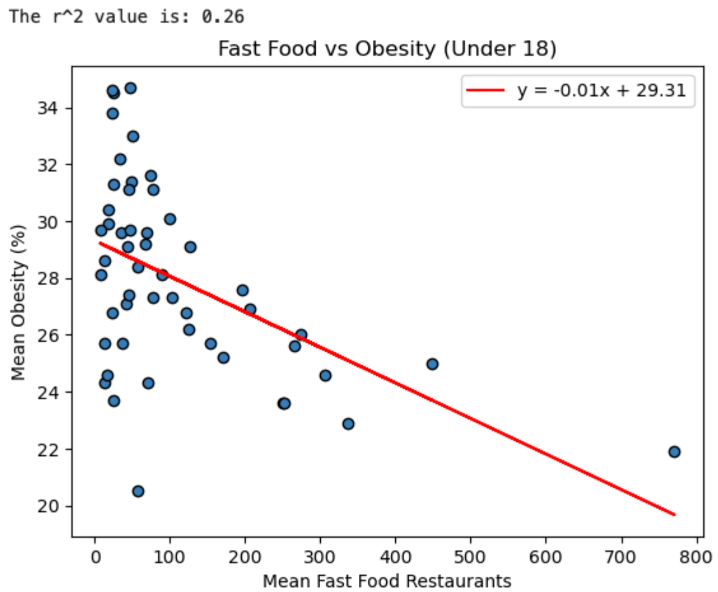
<u>**Diabetes relationship**</u>

The slope of the regression line, −0.0 indicates no meaningful change in diabetes rates as the number of fast food restaurants increases. This suggests a negligible relationship.

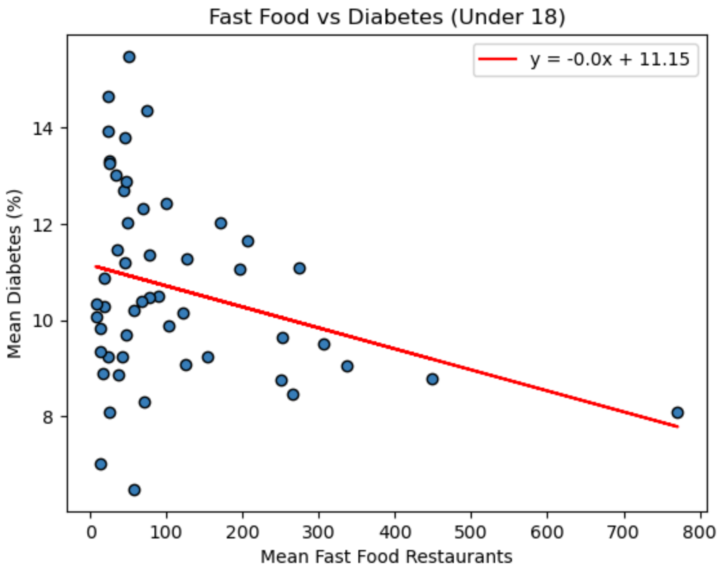
The intercept value of 11.15 represents the predicted diabetes rate when there are zero fast food restaurants. This baseline value implies that other factors likely contribute to the diabetes rate.

The R^2 value of 0.09 suggests that only 9% of the variation in the diabetes rate can be explained by the number of fast food

Linear relationship for the population under 18 years old



```
The r^2 value is: 0.09
```



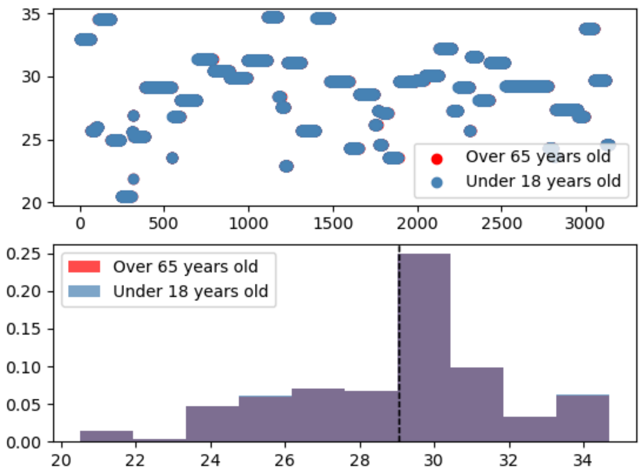
In the analysis, the linear relationship between the prevalence of obesity and diabetes and the density of fast food restaurants for both the population over 65 years old and the population under 18 years old resulted in similar findings. Despite the fact that these two groups, older adults and younger individuals, are generally assumed to have different health risks and responses to fast food, the regression analysis revealed that the trend between fast food density and the prevalence of obesity and diabetes was consistent for both groups.

2.3 Linear Regression: Is there a significant correlation between obesity and diabetes rates in the population over 65 years old and with those under 18 years old?

- **Null Hypothesis (H_0):** There is no significant difference in the impact of fast food on obesity or diabetes between the population over 64 years old and the population under 18 years old
- **Alternative Hypothesis (H_1):** There is a significant difference in the impact of fast food on obesity or diabetes between the population over 65 years old and the population under 18 years old

In order to test the null hypothesis, a two-sample t-test to compare the impact of fast food on obesity and diabetes between the population over 65 years old and the population under 18 years old. The null hypothesis stated that there would be no significant difference in the impact of fast food on these health outcomes across the two age groups.

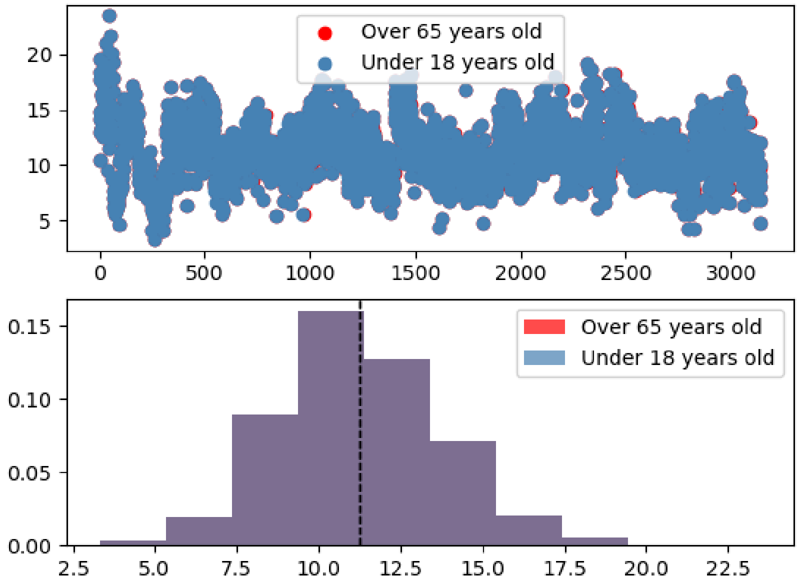
For the obesity analysis:



Measure	Value
t-statistic	-0.0231
p-value	0.9816
degrees of freedom	6283

Since the p-value is much larger than the conventional significance level of 0.05, the null hypothesis is rejected, indicating that there is no statistically significant difference in the impact of fast food on obesity between the two groups.

Similarly, for the diabetes analysis:



Measure	Value
t-statistic	**0.0003**
p-value	**0.9997**
degrees of freedom	**6281**

As the p-value is also much larger than 0.05, the null hypothesis is rejected, suggesting that there is no significant difference in the impact of fast food on diabetes between the two age groups.

In conclusion, the results of the statistical analysis indicate that there is no significant difference in the impact of fast food on obesity and diabetes between the population over 65 years old and the population under 18 years old. Both linear regression and t-test analyses suggest that the relationship between fast food consumption and these health outcomes, obesity and diabetes, remains consistent across the two age groups. Despite the differences in demographic characteristics, such as age, the data show no statistically significant variation in how fast food affects these conditions in either group.

3. Relationship between Number of Physical Activity Facilities and Obesity Rates in Counties of the US

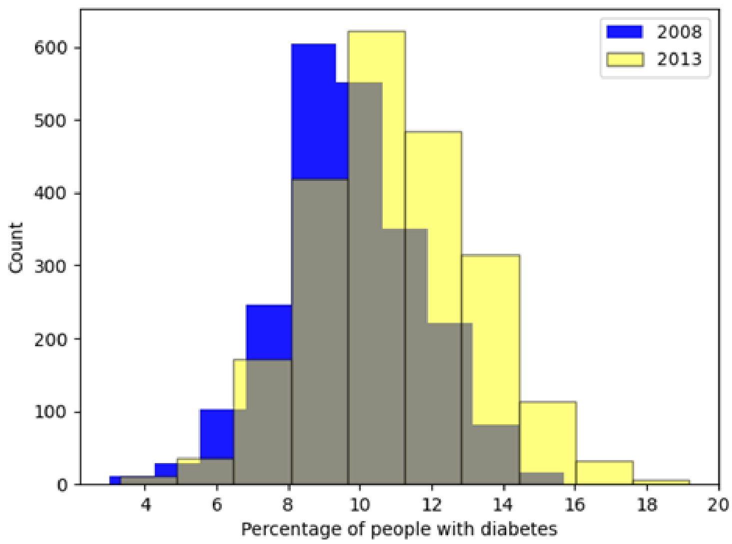
As mentioned before, it is important to identify whether there is a relationship between fast food restaurants and obesity and diabetes rates. Nevertheless, it is also relevant to analyze if the fact of having a certain quantity of gyms and physical activity facilities has an effect over obesity and diabetes rates in counties of the US.

In this section, the research question was “Do areas with a higher number of gyms and physical activity options have lower obesity and diabetes rates, even with high fast food restaurant density?”. The main goal is to explore whether access to physical activity infrastructure can counteract the effects of fast food.

3.1. Descriptive Analysis of Variables

To identify the structure of data distributions of the necessary variables, some descriptive statistics were analyzed.

Diabetes Rates

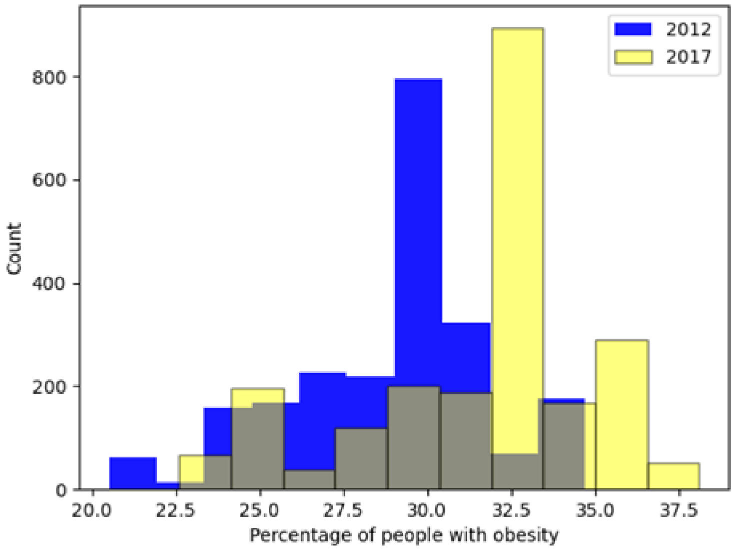


Variable	n	Mean	Median	Standard Deviation
Diabetes Rates, 2008	2213	9.7	9.6	1.9
Diabetes Rates,2013	2213	10.9	10.9	2.3

Source: Food Environment Atlas, U.S. Department of Agriculture.

The distributions of diabetes rates for 2008 and 2013 appear to be normal. Specifically, the diabetes rate for 2013 is normally distributed, as indicated by the close alignment between the mean and median. Additionally, the standard deviation suggests that there is minimal dispersion, meaning that the values are clustered closely around the mean.

Obesity Rates

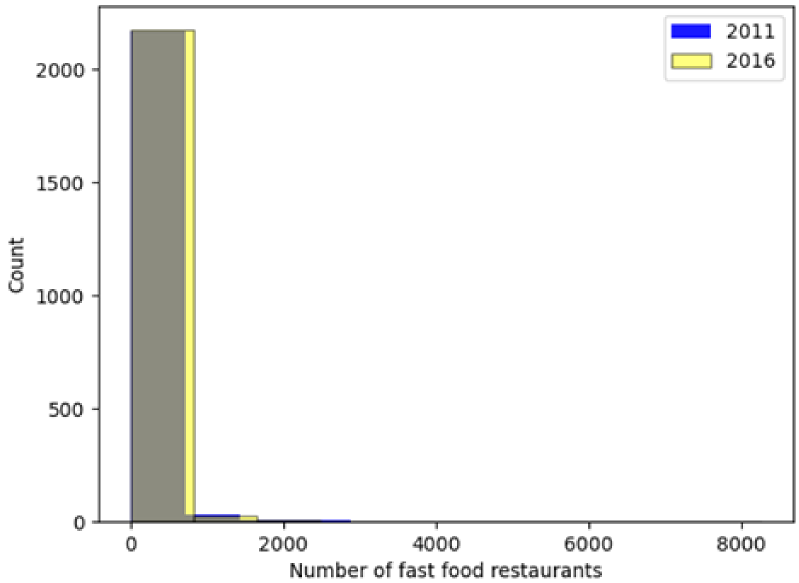


Variable	n	Mean	Median	Standard Deviation
Obesity Rates, 2012	2213	28.8	29.2	3.0
Obesity Rates, 2017	2213	31.6	32.4	3.4

Source: Food Environment Atlas, U.S. Department of Agriculture.

The distribution of obesity rates does not follow a normal distribution. The mean and median values for each year differ by 0.4 and 0.8, respectively, with standard deviations of 3.0 for 2012 and 3.4 for 2017. Additionally, the histogram above reveals two prominent bars—one for 2012 and another for 2017—that are notably higher than the others. This pattern deviates from a typical bell-shaped curve, further indicating that the data is not normally distributed.

Number of Fast-Food Restaurants

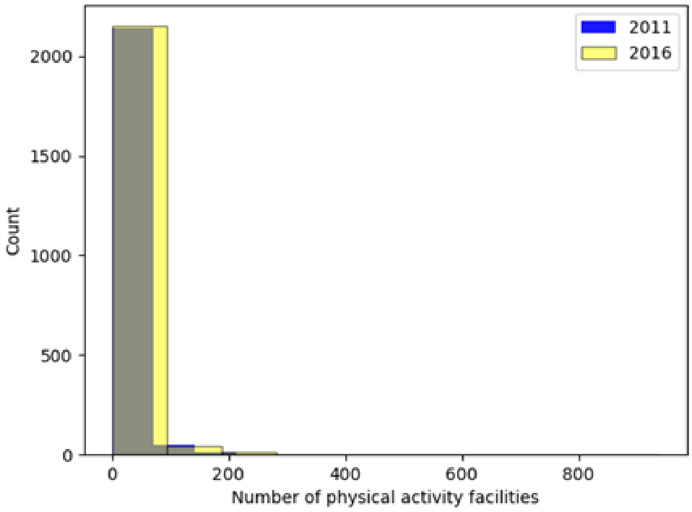


Variable	n	Mean	Median	Standard Deviation
Number of fast-food restaurants, 2011	2213	77.5	16.0	270.3
Number of fast-food restaurants, 2016	2213	85.6	16.0	307.9

Source: Food Environment Atlas, U.S. Department of Agriculture.

The distribution of the number of fast-food restaurants is highly skewed. While the means are 77.5 for 2011 and 85.6 for 2016, the median remains at 16 for both years, indicating an imbalance in the data. The high standard deviations, 270.3 for 2011 and 307.9 for 2016, suggest significant dispersion, reinforcing the skewness of the distribution.

Number of Physical Activity Facilities



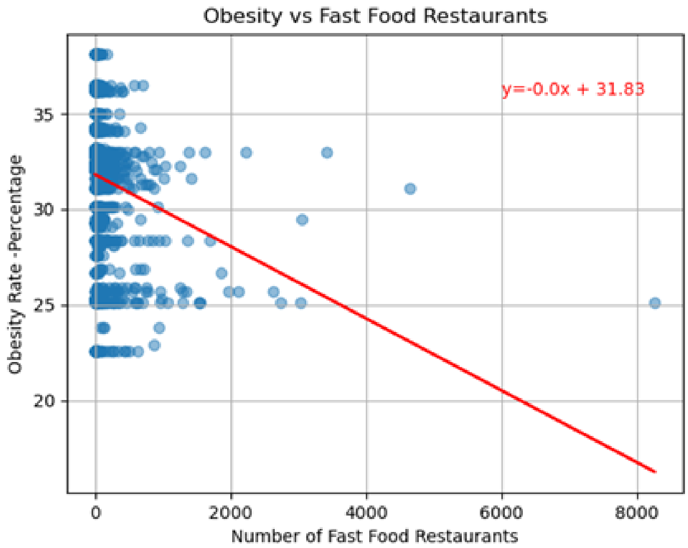
Variable	n	Mean	Median	Standard Deviation
Number of physical activity facilities, 2011	2213	10.5	2.0	32.8
Number of physical activity facilities, 2016	2213	12.1	2.0	41.4

Source: Food Environment Atlas, U.S. Department of Agriculture.

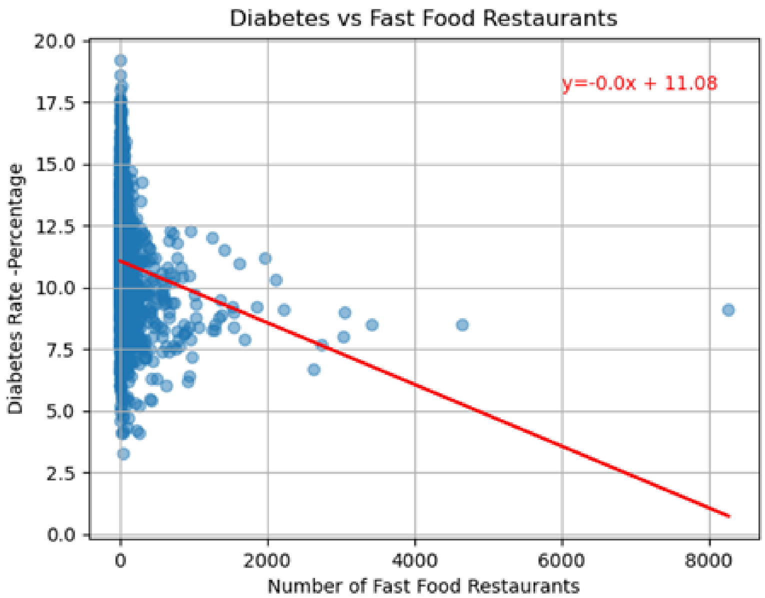
The distribution of the number of physical activity facilities is also highly skewed. While the means are 10.5 for 2011 and 12.1 for 2016, the median is at 2 for both years, also indicating an imbalance in the data. The standard deviations, 32.8 for 2011 and 41.4 for 2016, suggest significant dispersion.

3.2. Analysis of Relationship between Number of Fast-Food Restaurants and Obesity/Diabetes Rates

To investigate whether there is a relationship between the number of fast-food restaurants and rates of obesity and diabetes, a **simple linear regression** analysis was first conducted. The selected variables are the most recent to the present year.



Linear Regression between Fast-Food Restaurants and Obesity Rates: The results confirm that the coefficient of the independent variable (number of fast-food restaurants) is close to zero, indicating no significant relationship between the two variables. The R^2 value is 0.02, suggesting that variations in the number of fast-food restaurants have no meaningful effect on obesity rates.



Linear Regression between Fast-Food Restaurants and Diabetes Rates: Similarly, the linear regression analysis between the number of fast-food restaurants and diabetes rates indicates no relationship. As with obesity rates, the R^2 value is 0.02, showing that the independent variable has no substantial influence on diabetes rates.

To further examine this relationship, a t-test was performed to compare two groups of counties—those with a small number of fast-food restaurants and those with a large number of fast-food restaurants—for both obesity and diabetes rates:

The two groups were separated by the number of fast-food restaurants:

Sample	Delimitation	Variable
Group 1: Population with small number of fast-food restaurants	Number of fast-food restaurants < 16	FAST_FOOD_RESTAURANTS_16
Group 2: Population with large number of fast-food restaurants	Number of fast-food restaurants >= 16	

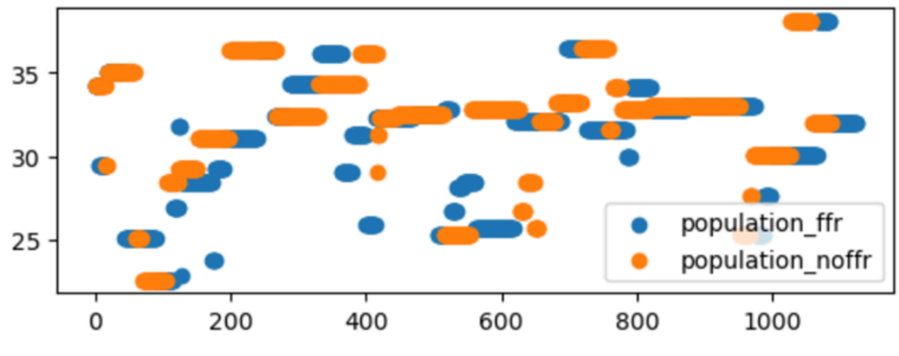
The election of the threshold is based on the median of the distribution. We took that statistic because it is the second quartile, and the data would be equally distributed below and above.

Two sample T-test Obesity Rates vs. Number of Fast-Food Restaurants:

The two hypothesis were:

> H_0: There is no difference in obesity rates between counties with more than 16 fast-food restaurants (the median) and counties with fewer than 16 fast-food restaurants.
>
> H_a: There is a statistically significant difference in obesity rates between counties with more than 16 fast-food restaurants and those with fewer than 16.

First, the two groups were graphic in a scatterplot. And the t-test was calculated for this samples with the results showed below:



Sample	Obesity Rate, 2013
Group 1	32.2
Group 2.	31.1
T-test result	**Statistic**: -7.24 **p-value**: 5.829321572109927e-13

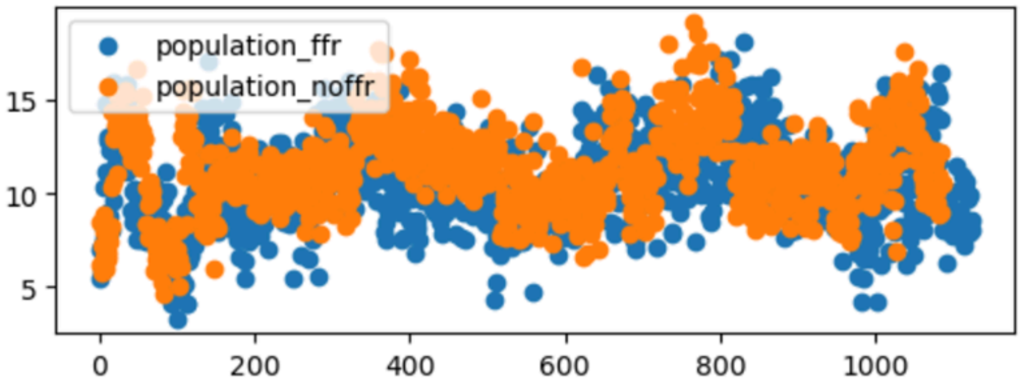
The p-value from the t-test is <0.01, indicating that we can reject the null hypothesis, which states, "There is no difference in obesity rates between counties with more than 16 fast-food restaurants (the median) and counties with fewer than 16 fast-food restaurants." Therefore, we accept the alternative hypothesis: "There is a statistically significant difference in obesity rates between counties with more than 16 fast-food restaurants and those with fewer than 16."

However, contrary to our initial expectations, the relationship is negative (t-statistic = -7.24). This means that counties with fewer than 16 fast-food restaurants actually have higher obesity rates.

Two sample T-test Diabetes Rates vs. Number of Fast-Food Restaurants:

The hypothesis were:

> H_0: There is no difference in diabetes rates between counties with more than 16 fast-food restaurants (the median) and counties with fewer than 16 fast-food restaurants.
>
> H_a: There is a statistically significant difference in diabetes rates between counties with more than 16 fast-food restaurants and those with fewer than 16.



Sample	Diabetes Rate, 2013
Group 1	11.3
Group 2.	10.5
T-test result	**Statistic**: -8.05 **p-value**: 1.2886218552001376e-15

The results are not surprising, because they show the same trend as the one observed with obesity rates. The p-value from the t-test is <0.01, indicating that we can reject the null hypothesis. Therefore, we accept the alternative hypothesis: "There is a statistically significant difference in diabetes rates between counties with more than 16 fast-food restaurants and those with fewer than 16."

The relationship is also negative (t-statistic = -8.05). This suggests that counties with fewer than 16 fast-food restaurants actually have higher diabetes rates.

These findings could potentially be explained by two socioeconomic factors:

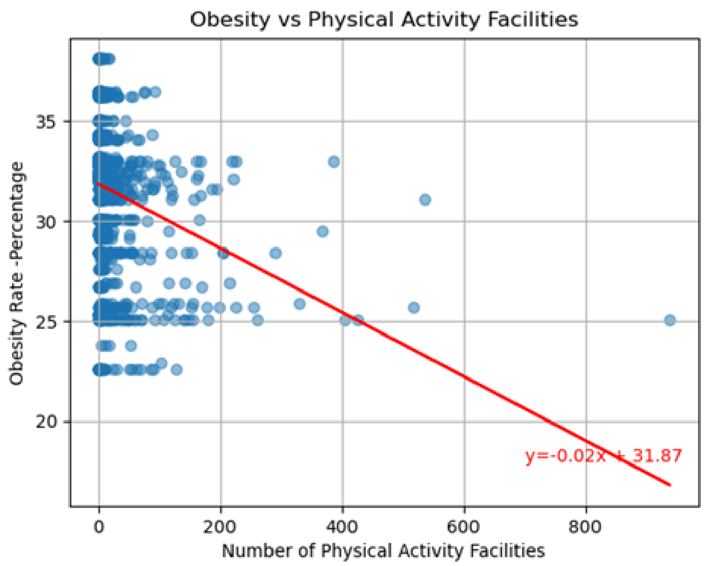
A. **Economic Development and Access to Healthier Options**: Counties with a higher number of fast-food restaurants might also have greater economic development. The population in these areas may have higher purchasing power, allowing them access to a broader range of dining options, including more expensive and healthier restaurants.

B. **Access to Health and Fitness Resources**: Areas with more fast-food restaurants might also have better access to physical activity facilities or public health programs that promote healthier lifestyles. These resources could contribute to lower obesity and diabetes rates among residents.

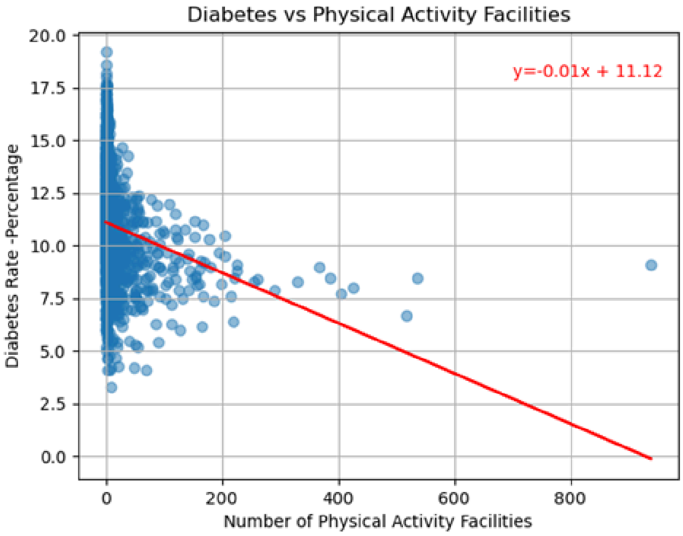
To further investigate the influence of socioeconomic factors on obesity and diabetes rates, we planned to test whether a high number of physical activity facilities has a positive effect on reducing obesity and diabetes rates.

3.3. Analysis of Relationship Between Number of Physical Activity Facilities and Obesity/Diabetes Rates

To analyze this relationship, we followed the same steps as in previous analyses. First, we conducted linear regressions:



Linear Regression between Physical Activity Facilities and Obesity Rates: The results indicate that the coefficient for the independent variable (number of physical activity facilities) is approximately zero, suggesting there is no meaningful relationship between the number of facilities and obesity rates. The R^2 value is 0.04, which indicates that the number of physical activity facilities explains very little of the variation in obesity rates.



Linear Regression between Physical Activity Facilities and Diabetes Rates: The linear regression analysis between diabetes rates and the number of physical activity facilities shows an x-coefficient of 0.01, indicating that there is almost no relationship between the two variables. The R^2 value is 0.04, suggesting that the number of facilities explains very little of the variation in diabetes rates.

Since we did not find evidence of a significant relationship, we proceeded to compare two groups using a t-test. In this analysis, the groups were defined by:

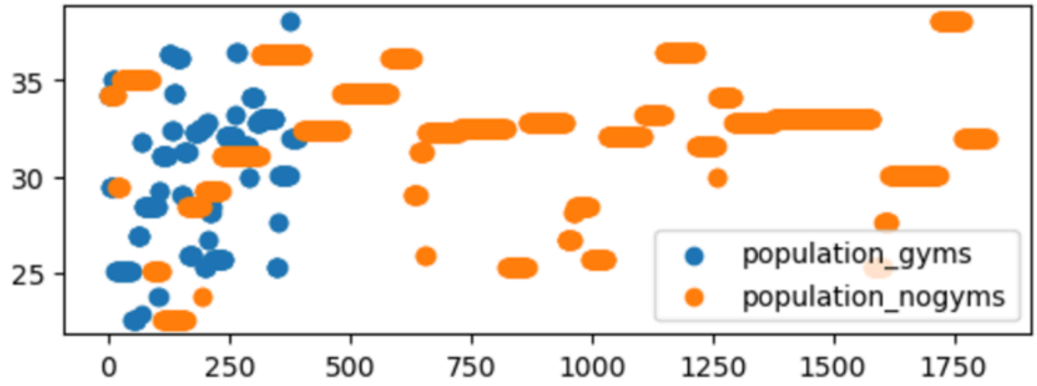
Sample	Delimitation	Variable
Group 1: Population with small number of physical activity facilities	Number of physical activity facilities < 12	GYMS_16
Group 2: Population with large number of physical activity facilities	Number of physical activity facilities >= 12	

The election of the threshold is based on the means of the distribution. We took that statistic because the median was 2 and we thought the difference between groups would show only if the large number of gyms were larger.

Two sample T-test Obesity Rates vs. Number of Physical Activity Facilities:

The two hypothesis were:

> H_0: There is no difference in obesity rates between counties with more than 12 physical activity facilities (the mean) and counties with fewer than 12 physical activity facilities.
>
> H_a: There is a statistically significant difference in obesity rates between counties with more than 12 physical activity facilities and those with fewer than 12.



Sample	Obesity Rate, 2017
Group 1	32.0
Group 2.	30.0
T-test result	**Statistic**: -10.05 **p-value**: 6.333292109738437e-22

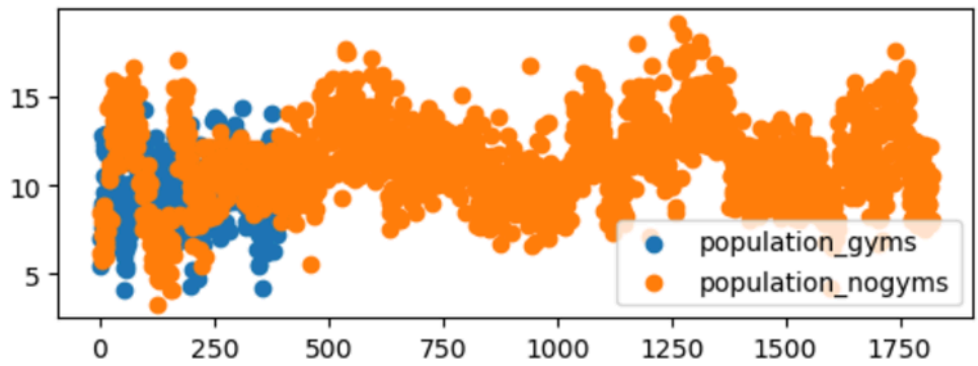
The p-value of the t-test is <0.01, indicating that we can reject the null hypothesis. Therefore, we accept the alternative hypothesis. The relationship is negative (t-statistic = -10.05), meaning that counties with fewer than 12 physical activity facilities have higher obesity rates.

In contrast to the hypothesis test concerning fast-food restaurants and obesity rates, these results suggest that socioeconomic factors within counties may have a stronger influence on health outcomes, such as obesity rates.

Two sample T-test Diabetes Rates vs. Number of Physical Activity Facilities:

The two hypothesis were:

> H_0: There is no difference in diabetes rates between counties with more than 12 physical activity facilities (the mean) and counties with fewer than 12 physical activity facilities.
>
> H_a: There is a statistically significant difference in diabetes rates between counties with more than 12 physical activity facilities and those with fewer than 12.



Sample	Diabetes Rate, 2013
Group 1.	11.2
Group 2.	9.4
T-test result	**Statistic**: -16.54 **p-value**: 9.801915611846224e-52

The p-value of the t-test is <0.01, which means that we can reject the null hypothesis. In that sense, the alternative hypothesis is accepted, so there is statistical difference in diabetes rates of the counties with more than 12 (mean) physical activity facilities in contrast to counties with less than 12 physical activity facilities.

The relationship is negative (t-statistic=-10.05), so that means that the counties with less than 12 physical activity facilities have higher diabetes rates. In contrast with the hypothesis test about fast food restaurants and diabetes rates, these results could confirm that socioeconomic factors of the counties have a stronger effect over health variables, such as diabetes rates.

3.4 Main Findings

The results of this analysis show relevant findings about the relationship between metabolic illnesses and factors that could potentially impact on the increase or decrease of the rates. In this study, we found that:

- Linear regression between obesity and diabetes rates did not show a relationship with the number of fast-food restaurants. This could be explained by:
    - The linear regression was not the best statistical approach to analyze the data, because the observations were counties. Probably, if the observations were people and we tried to measure the consumption of people in fast food restaurants, the method could have shown relationships.
- The t-test between two samples: counties with more than 16 fast-food restaurants and counties with less than 16 fast-food restaurants have a statistically significant difference in obesity and diabetes rates. The relationship was negative, which means that the group with more than 16 restaurants had higher illnesses rates.
    - This could be explained because of other socioeconomic factors such as counties with greater economic development, which means higher acquisition power for people, higher options for food intake and more health public policies.
- In order to identify the relationship between socioeconomic factors and diabetes and obesity rates, we perform a linear regression. The method suggested that there is no relationship between the number of physical activity facilities and disease rates.

- As mentioned before, this result could be the consequence of the data observations. If the observations were people, it may have shown other results.

- The t-test showed a significant difference between the mean diseases rates and the alternative hypothesis was accepted. That means that the counties with more than 12 physical activity facilities had lower rates of diabetes and obesity rates.
    - This could confirm that the socioeconomic factors are more powerful than nutrition or access to fast-food factors.

Sources

U.S. Department of Agriculture, Economic Research Service. (n.d.). *Food environment atlas: Data access and documentation downloads*. Retrieved November 7, 2024, from https://www.ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads/