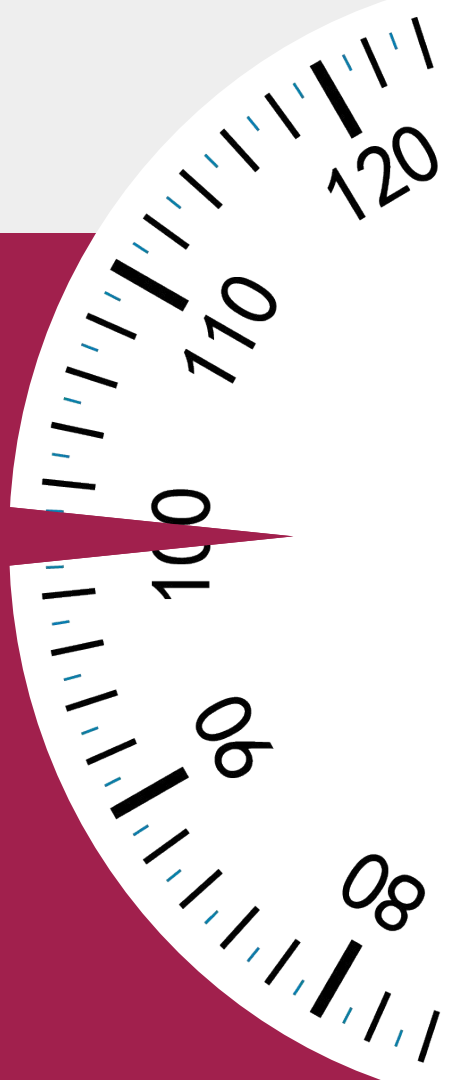
OBESITY, INCOME AND FOOD ACCESS

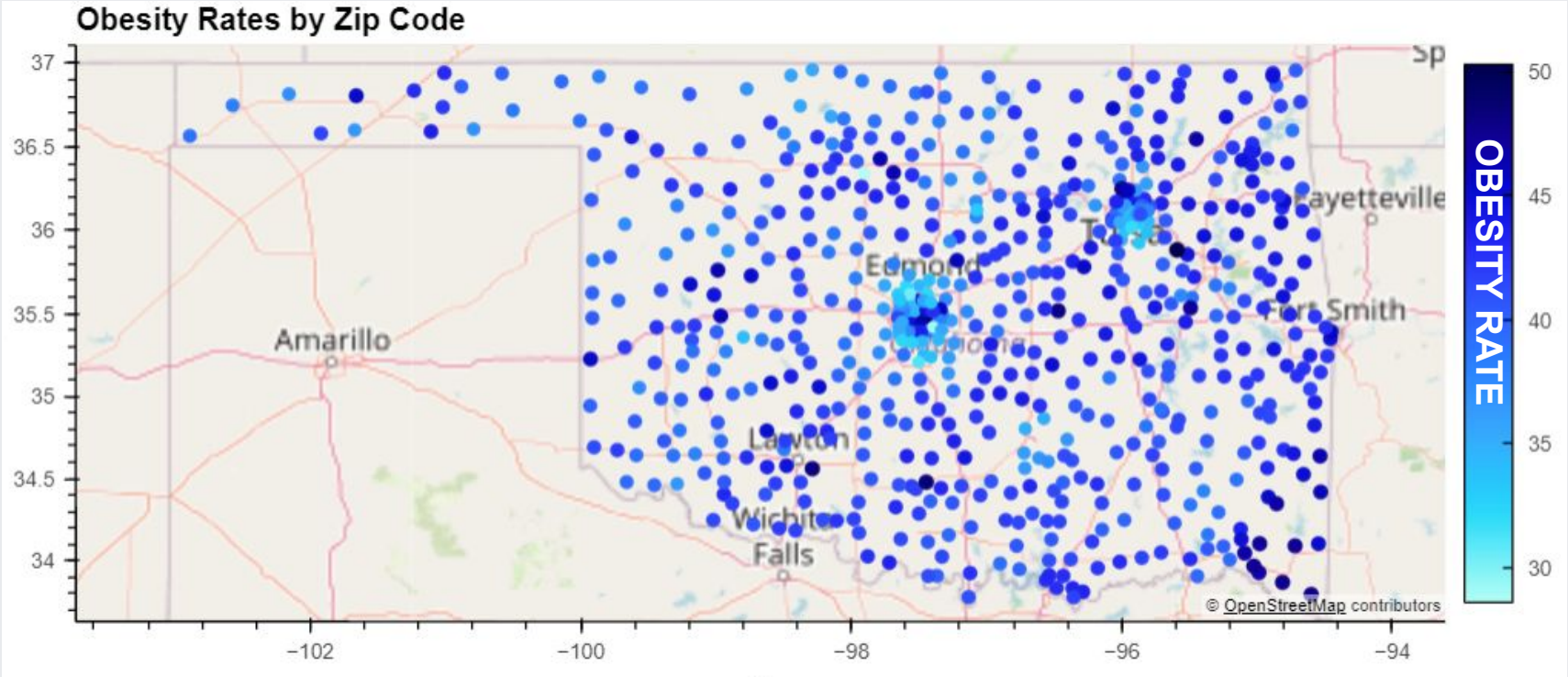
Angie Foust, Cassidy Simmons, Stephanie Rathod, Katie Starnes

OVERVIEW

In this project, we aimed to investigate possible factors that impact the obesity rate in Oklahoma. We were interested in income, access to grocery stores (hoping to see a relationship to “food deserts”), access to fast food, and available transportation. Through a comprehensive exploration of these possible relationships, we aspire to contribute valuable insights that can inform policies and interventions aimed at addressing public health challenges associated with socio-economic disparities and nutritional access.



BASIC OBESITY STATISTICS



2021 Obesity among adults (aged 18+): Centers for Disease Control
The average rate of obesity in Oklahoma is 40.89% per zip code.

QUESTION 1

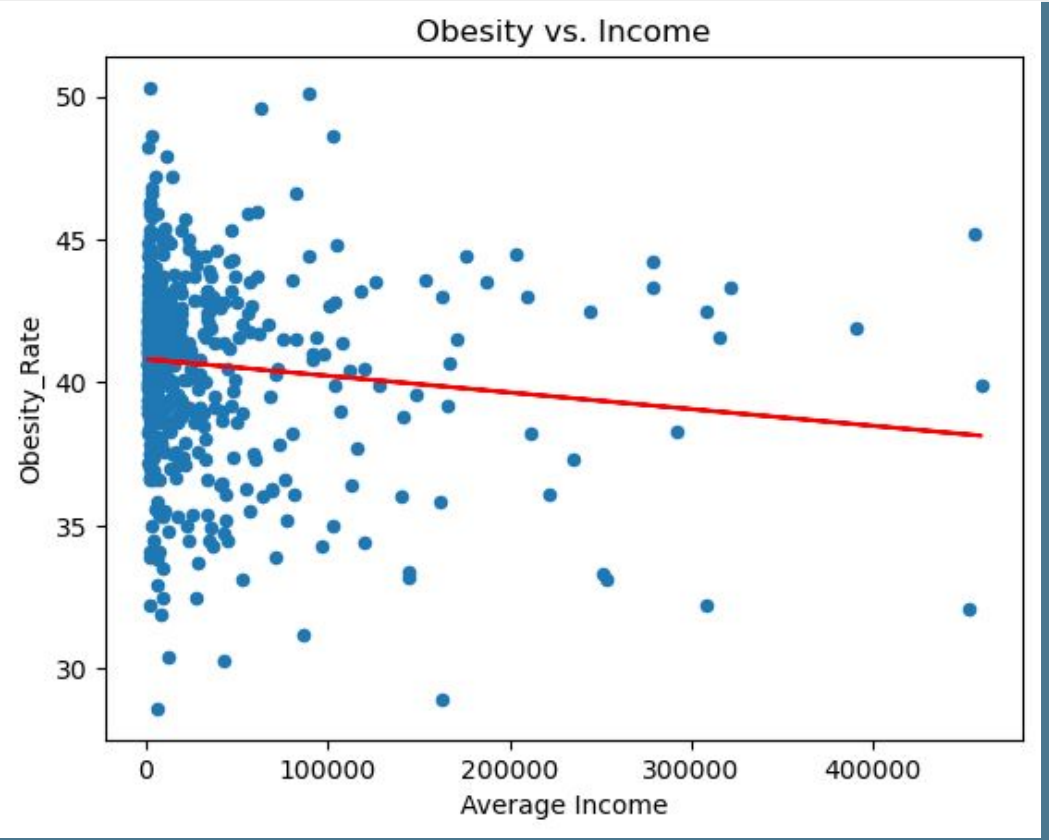
How does income affect the chances of developing obesity?

Hypotheses:

- **The lower the income the higher the rate of obesity**
- **Null: no relationship exists between income level and obesity**

Income by Zip code vs. Obesity Rate

Scatter plot of income vs. obesity rate. Most of the points are skewed right and are below an average income of $50,000, and above an obesity rate of 35%.



Summary Statistics

	Average Income	Obesity_Rate
count	547.000000	547.000000
mean	32785.412699	40.621024
std	62770.088106	3.221905
min	675.666667	28.600000
25%	2927.416667	39.000000
50%	8245.166667	41.000000
75%	32996.916665	42.800000
max	459684.666700	50.300000

QUESTION 2

How does proximity to grocery stores impact obesity?

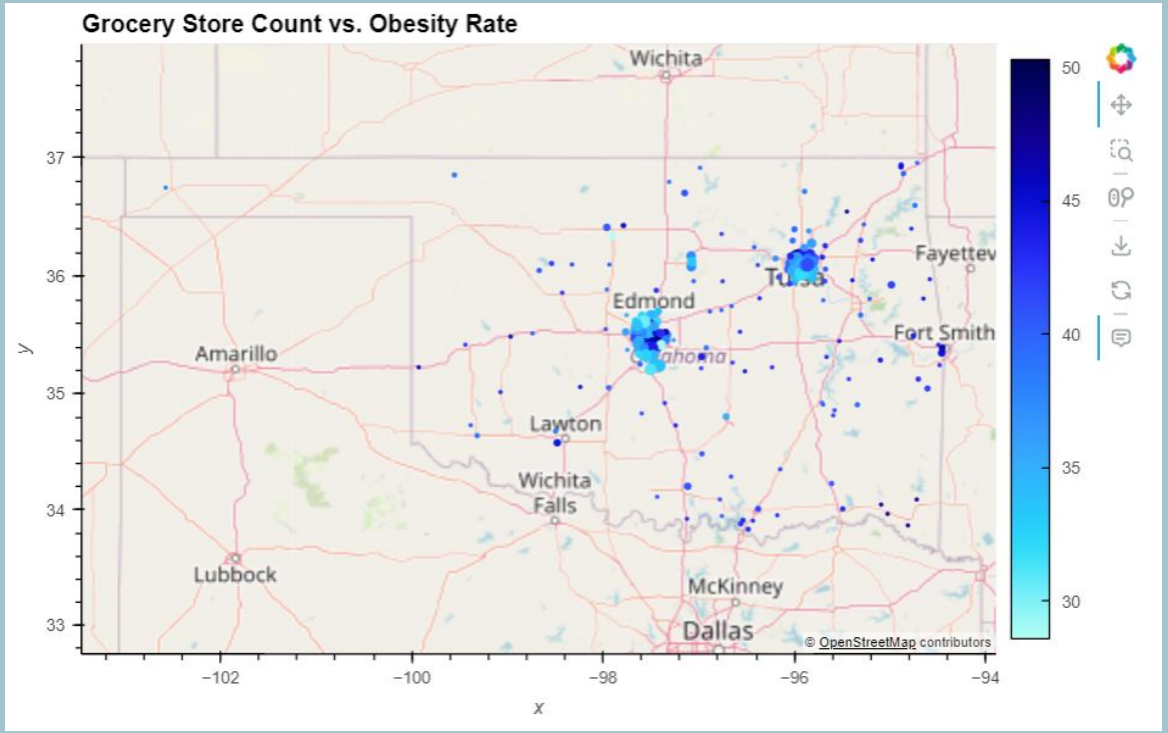
Hypotheses:

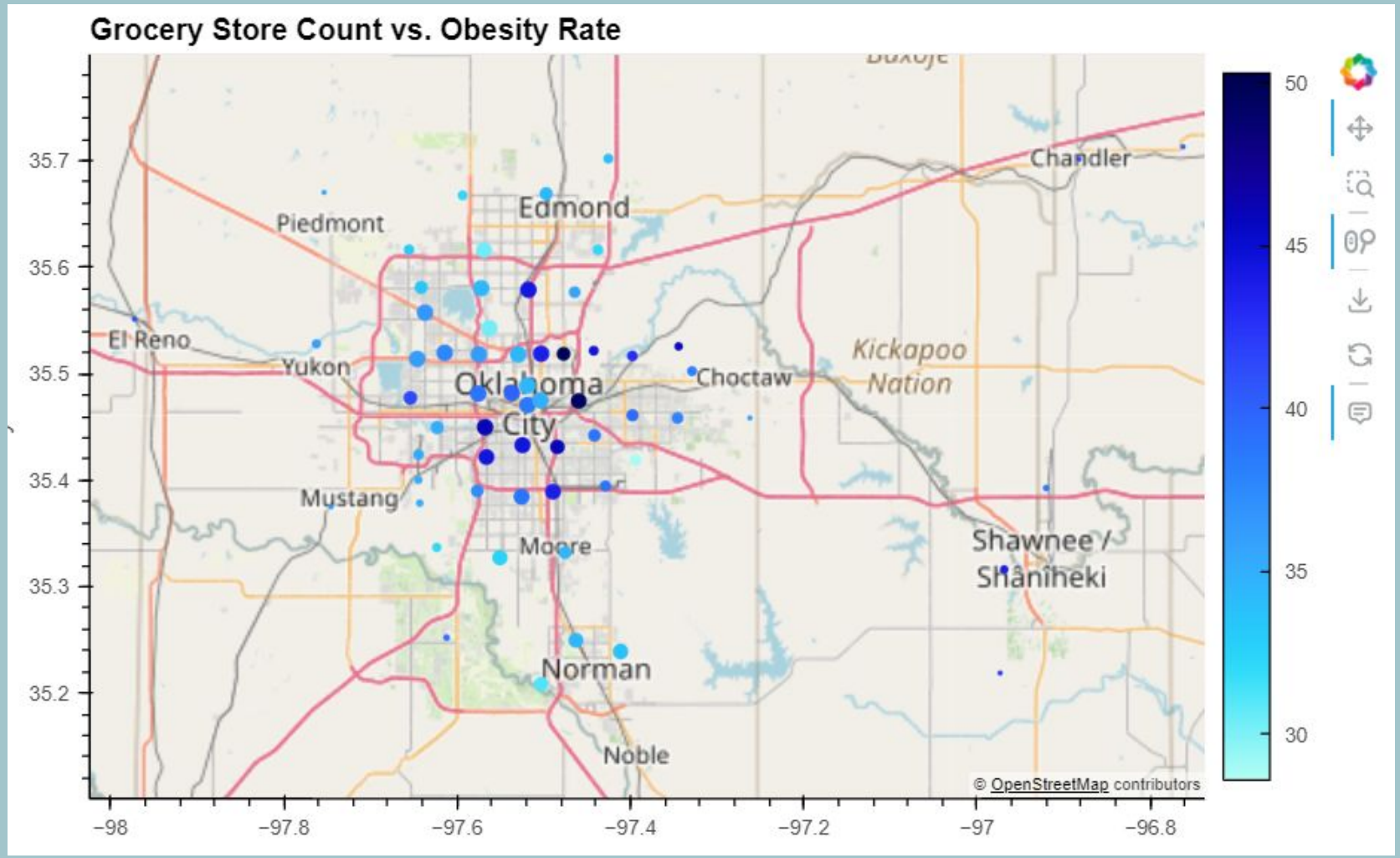
- **Null: No relationship exists between the number of stores available within 5 miles and the obesity rate.**
- **Hypothesis: The number of accessible stores has an impact on the obesity rate.**

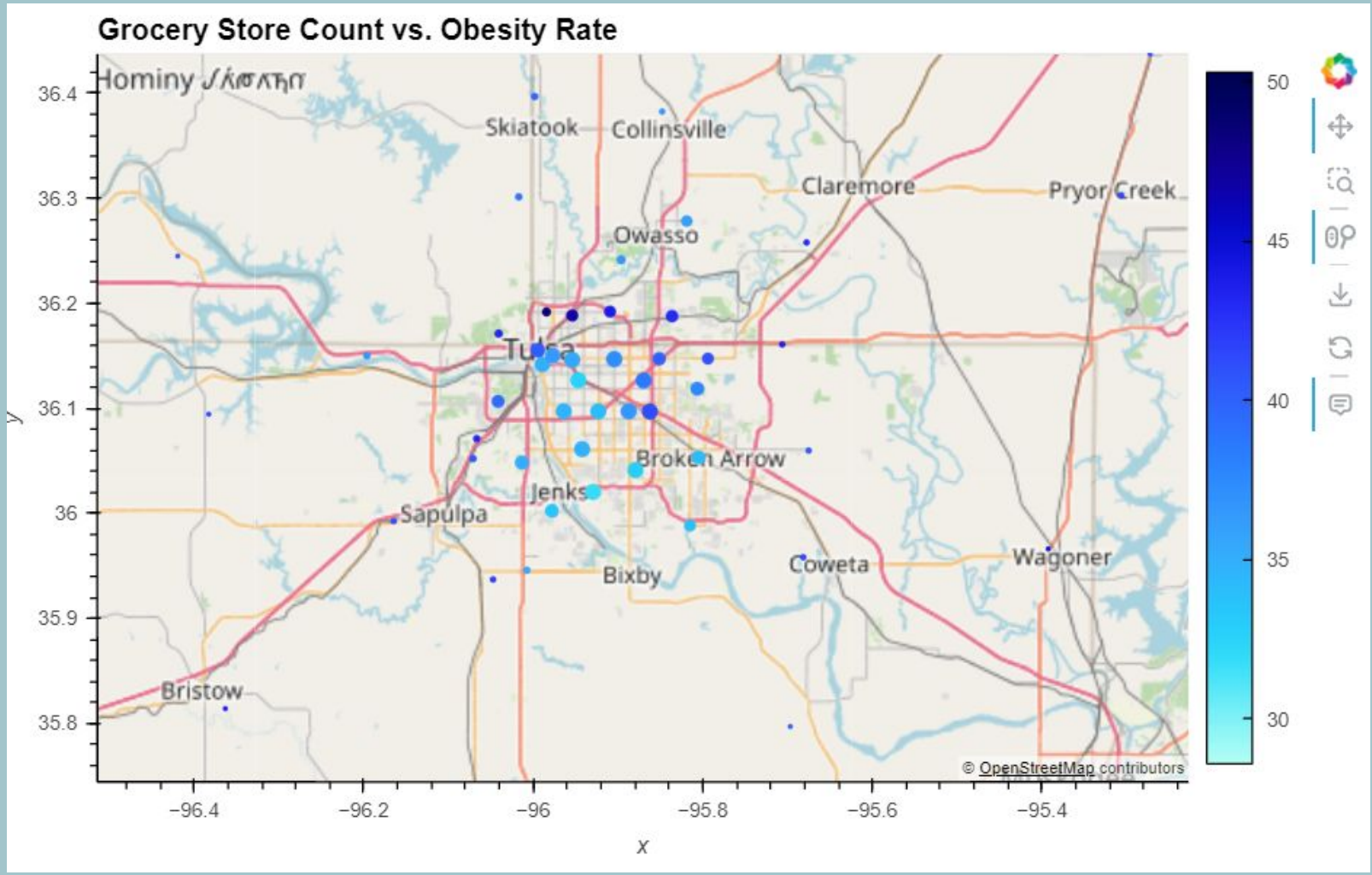
Store Count by Zip Code vs. Obesity Rate

Average: 2 (per 5 mile radius)
Range: 20

Method: Because the data from the CDC included zip codes with a long/lat reference point, we were able to use Geoapify to pull the number of grocery stores within 5 miles of each point.




Grocery Store Count vs. Obesity Rate
35.7
35.6
35.5
35.4
35.3
35.2
−98
−97.8
−97.6
−97.4
−97.2
−97
−96.8
Edmond
Piedmont
El Reno
Yukon
Oklahoma City
Choctaw
Kickapoo Nation
Chandler
Mustang
Moore
Shawnee / Shâniheki
Norman
Noble
© OpenStreetMap contributors
50
45
40
35
30


Grocery Store Count vs. Obesity Rate
Hominy
Skiatook
Collinsville
Claremore
Pryor Creek
Owasso
Tulsa
Broken Arrow
Jenks
Sapulpa
Bixby
Coweta
Wagoner
Bristow
© OpenStreetMap contributors
36.4
36.3
36.2
36.1
36
35.9
35.8
y
−96.4
−96.2
−96
−95.8
−95.6
−95.4
x
50
45
40
35
30

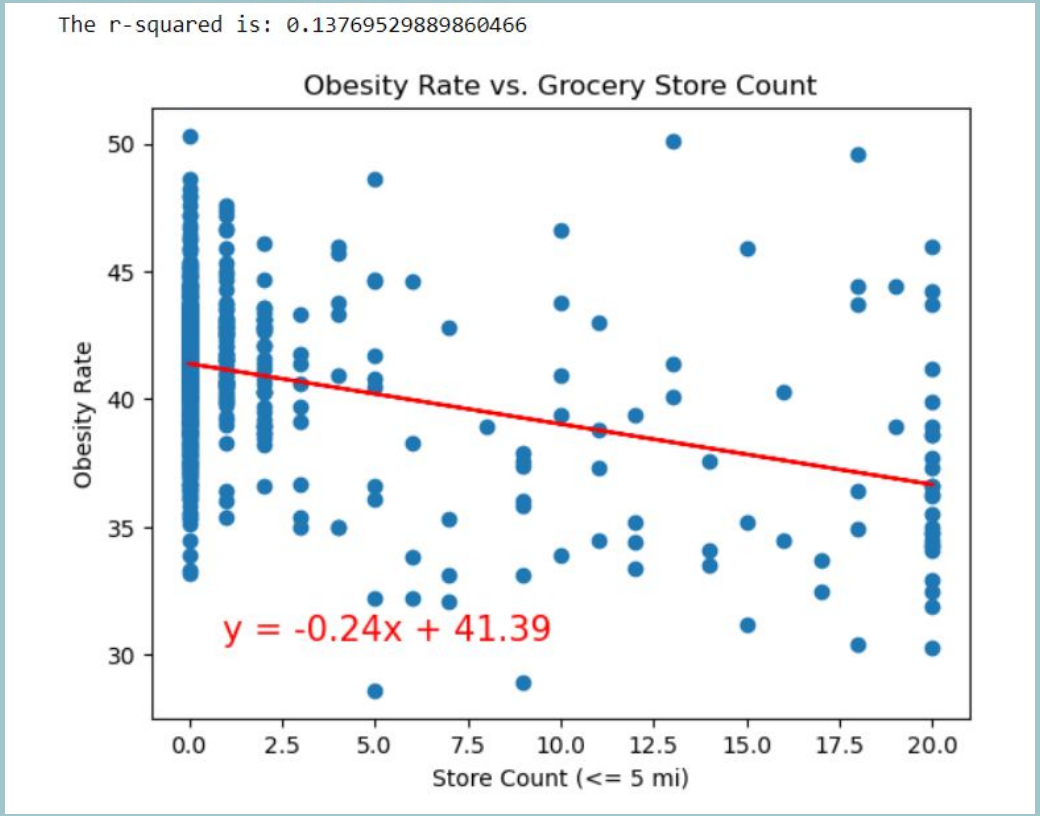
DATA ANALYSIS

Three types of data analysis and visualization were used to determine if there is a relationship between the number of accessible grocery stores and the obesity rate.

- Linear Regression (for all data points)
- Summary Statistics (with grouped data)
- Box Plot by Store Count Groups with ANOVA test

LINEAR REGRESSION

Ungrouped data was plotted as a scatter plot with a line of regression. The line visually indicates the lower store count areas have higher rates of obesity, however, the r value is less than 0.3 which indicates a very weak correlation. Because there are so many areas with the same count of stores, looking at the data grouped by store count will allow hypothesis testing.



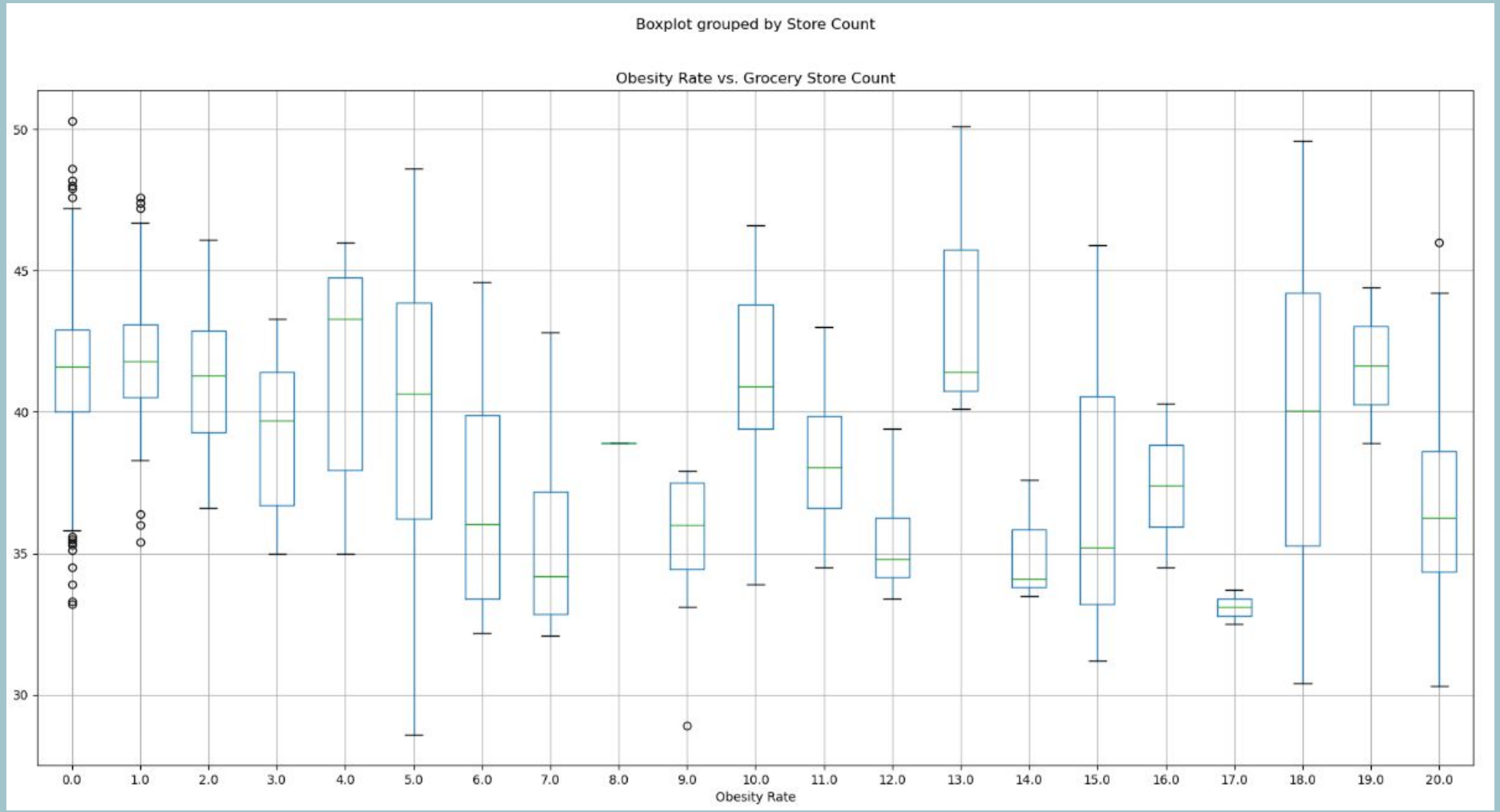
SUMMARY STATISTICS

Note: Only one geographical location had 8 stores.

The data was grouped by store count for further analysis.

The mean for these groups appear to be similar, so further testing is needed.

Store Count	mean	median	var	std	sem
0.0	41.345909	41.60	6.354653	2.520844	0.120177
1.0	41.987879	41.80	6.240159	2.498031	0.307486
2.0	41.211765	41.30	4.704100	2.168894	0.371962
3.0	39.222222	39.70	8.619444	2.935889	0.978630
4.0	41.385714	43.30	21.864762	4.675977	1.767353
5.0	39.440000	40.65	37.269333	6.104861	1.930527
6.0	37.225000	36.05	30.842500	5.553602	2.776801
7.0	35.825000	34.20	23.409167	4.838302	2.419151
8.0	38.900000	38.90	NaN	NaN	NaN
9.0	35.242857	36.00	10.496190	3.239782	1.224523
10.0	40.920000	40.90	23.037000	4.799687	2.146485
11.0	38.400000	38.05	12.580000	3.546830	1.773415
12.0	35.600000	34.80	6.960000	2.638181	1.319091
13.0	43.866667	41.40	29.563333	5.437217	3.139179
14.0	35.066667	34.10	4.903333	2.214347	1.278454
15.0	37.433333	35.20	57.763333	7.600219	4.387989
16.0	37.400000	37.40	16.820000	4.101219	2.900000
17.0	33.100000	33.10	0.720000	0.848528	0.600000
18.0	39.900000	40.05	51.256000	7.159330	2.922784
19.0	41.650000	41.65	15.125000	3.889087	2.750000
20.0	36.792308	36.25	14.679938	3.831441	0.751407


Boxplot grouped by Store Count
Obesity Rate vs. Grocery Store Count
50
45
40
35
30
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
17.0
18.0
19.0
20.0
Obesity Rate

Boxplot with ANOVA

Running a one-way ANOVA test resulted in a pvalue of 3.2 e -24

Because the pvalue is under 0.05, we can reject the null hypothesis.

Conclusion: A relationship does exist between the number of stores available within 5 miles and the obesity rate.

QUESTION 2: CONCLUSION

This exploration of data shows there is likely a relationship between the number of grocery stores accessible (within 5 miles) and the obesity rate. However, this aspect of the project has several limitations including:

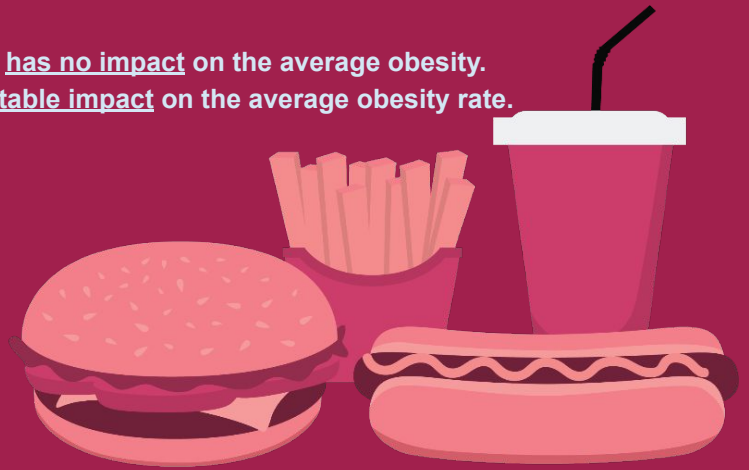
- **Rural areas may likely be farming/ranching communities who do not rely heavily on store bought goods.**
- **GeoApify may not have included smaller mom & pop stores more common in smaller towns, or other stores that carry grocery items.**
- **The variety of items at the grocery stores was not looked at and could have an impact on obesity rates. (i.e. Sprouts vs. Dollar General)**

QUESTION 3

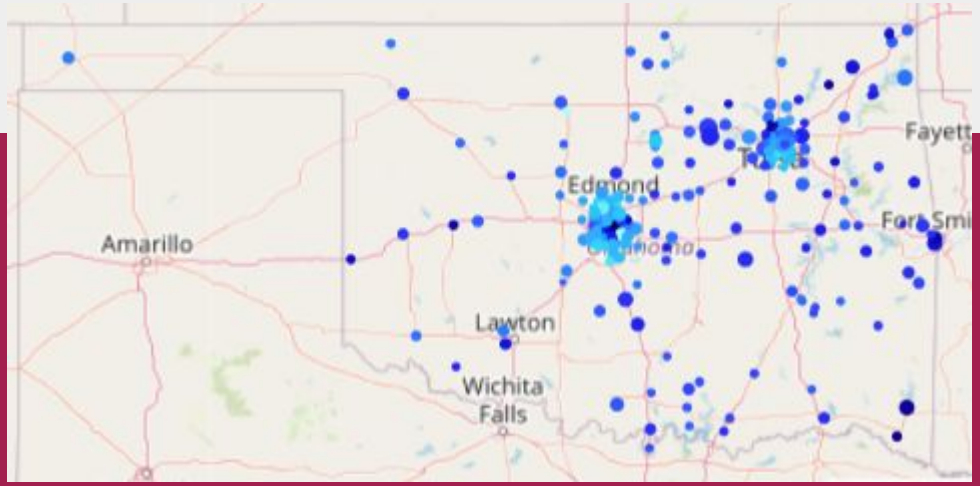
How does the number of fast food restaurants in a given area impact obesity?

I will be testing whether or not the number of fast food chains within a 5 mile* (8,000 meter) radius has an impact on average obesity in the area

Null hypothesis: The number of fast food restaurants within a 5 mile radius of a given area has no impact on the average obesity.
(alternate) hypothesis: The number of fast food restaurants within a 5 mile radius has a notable impact on the average obesity rate.



Map !



Average (5 mile radius): 3.14
(more than the grocery store average)

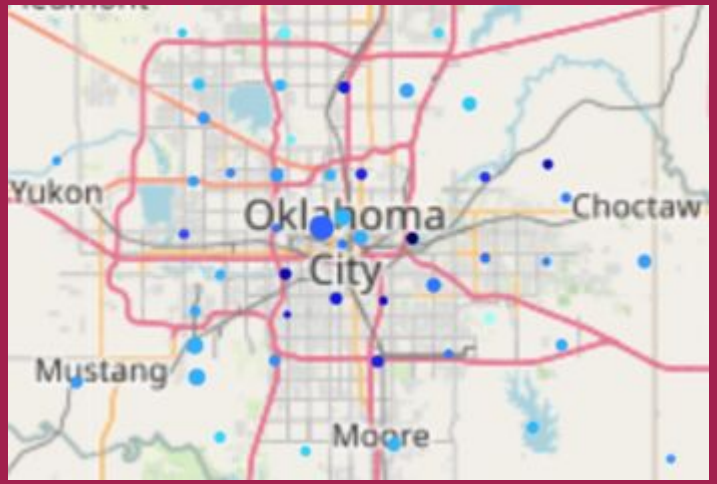
Method:
The CDC data we used grouped average obesity rate by zip code + latitude and longitude coordinates. Using those coordinates I was able to find out many fast food restaurants there are in a five mile radius using Geoapify.

A closer look at the metro areas

Size = # of fast food chains

Color = obesity rate

(that big dot in OKC has 51 fast food chains!)

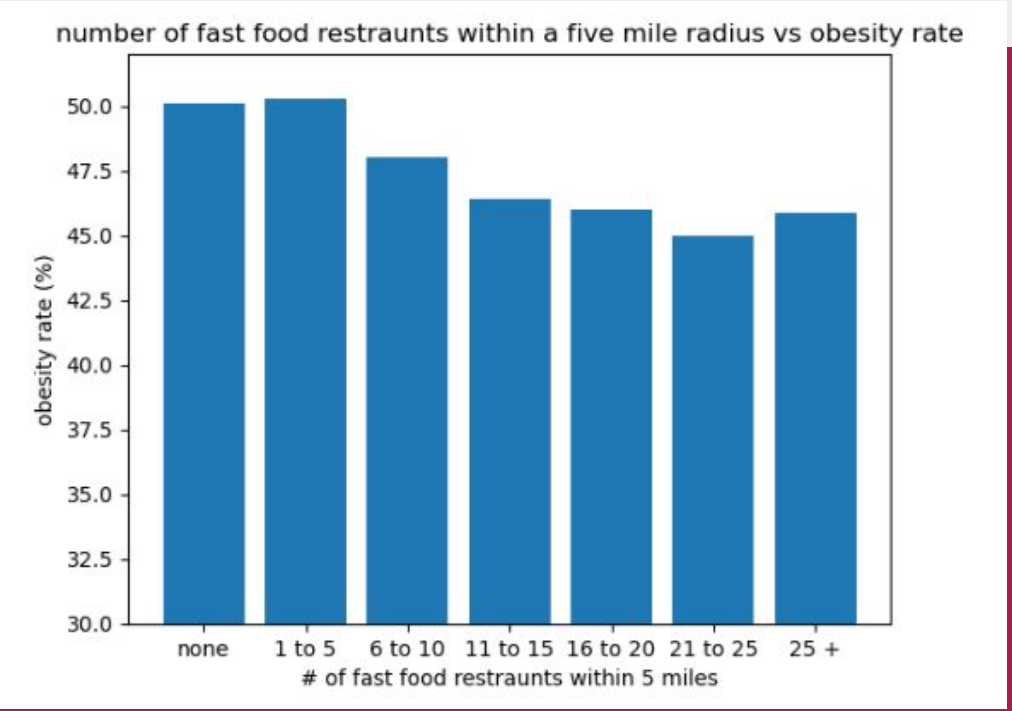


Bar chart !

Method:

After finding the number of fast food chains in each zip code I binned those values into seven different groups and graphed those values vs the average obesity rate in each group

(did not return the values i was expecting!)



ANOVA test !

The image below shows a snippet of code where the groups are defined by the number of fast food restaurants and we are comparing the obesity rate (‘data_value’)

Is there a statistical difference between the groups ?

Since the p-value is super tiny (less than 0.05), there is a statistical difference and the null hypothesis is REJECTED !

```
# ANOVA test

group1 = fast_food_df[fast_food_df["bins"] == "none"]["Data_Value"]
group2 = fast_food_df[fast_food_df["bins"] == "1 to 5"]["Data_Value"]
group3 = fast_food_df[fast_food_df["bins"] == "6 to 10"]["Data_Value"]
group4 = fast_food_df[fast_food_df["bins"] == "11 to 15"]["Data_Value"]
group5 = fast_food_df[fast_food_df["bins"] == "16 to 20"]["Data_Value"]
group6 = fast_food_df[fast_food_df["bins"] == "21 to 25"]["Data_Value"]
group7 = fast_food_df[fast_food_df["bins"] == "25 +"]["Data_Value"]

st.f_oneway(group1, group2, group3, group4, group5, group6, group7)
```

```
F_onewayResult(statistic=10.331597928295137, pvalue=6.1512105797246e-11)
```

QUESTION 3: CONCLUSION

Through the analysis I have conducted it is clear that the number of fast food restaurants and obesity rate is connected in some way, perhaps not the way I had initially assumed though.

This project has brought up several other questions that we could explore and dig deeper with to get a clearer idea on other potential causes:

- What types of fast food are people consuming the most in certain areas?
- Would people be willing to drive further for their favorite fast food rather than the grocery store?
- How often are people eating out rather than at home?
- How do more rural areas impact the dataset?

QUESTION 4

Does access to public transportation affect accessibility obesity?

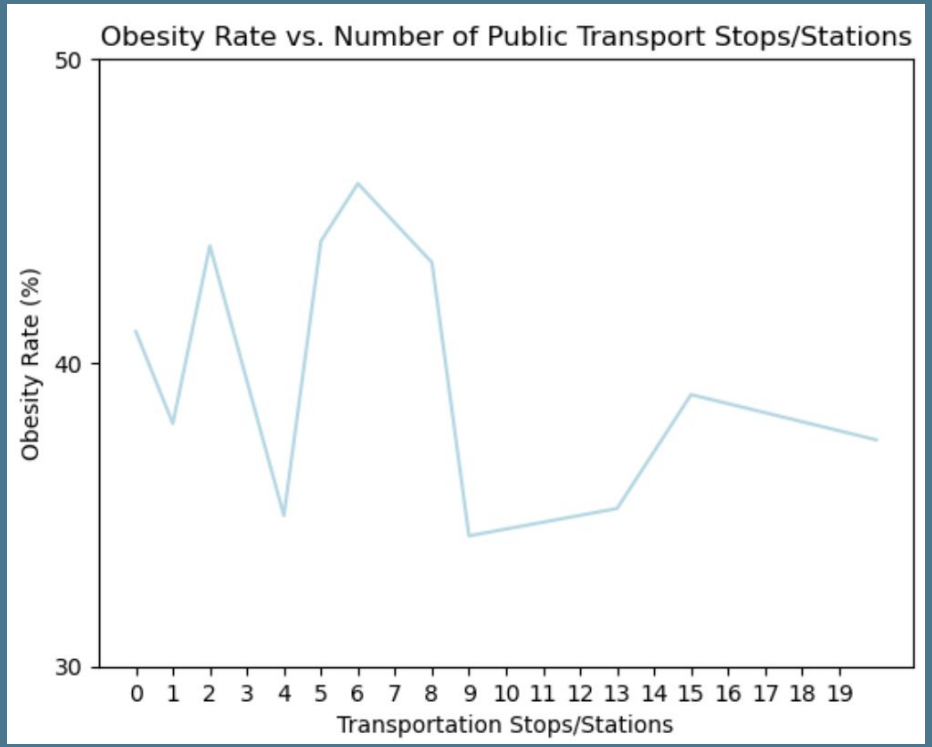
I will be testing how many public transportation stops and stations are within a 1,600 meter (1 mile) radius of each zip code in Oklahoma.

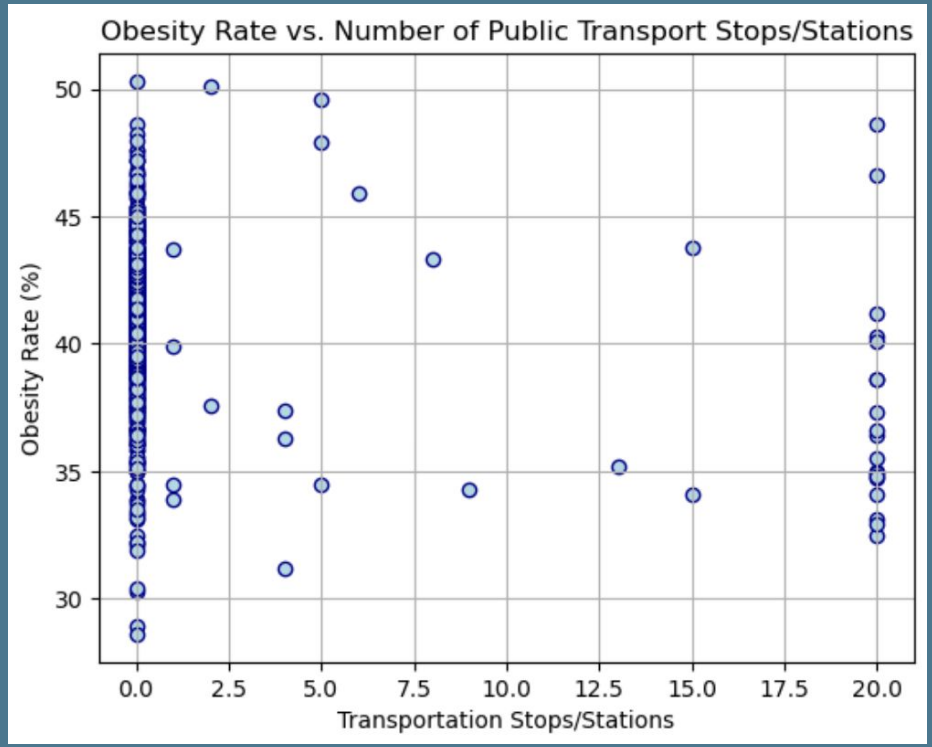
Hypotheses:

- **The amount of transportation stops/stations will have an affect on the rate of obesity in a zip code.**
- **Null: The amount of transportation stops/stations has no effect on the rate of obesity in a zip code.**

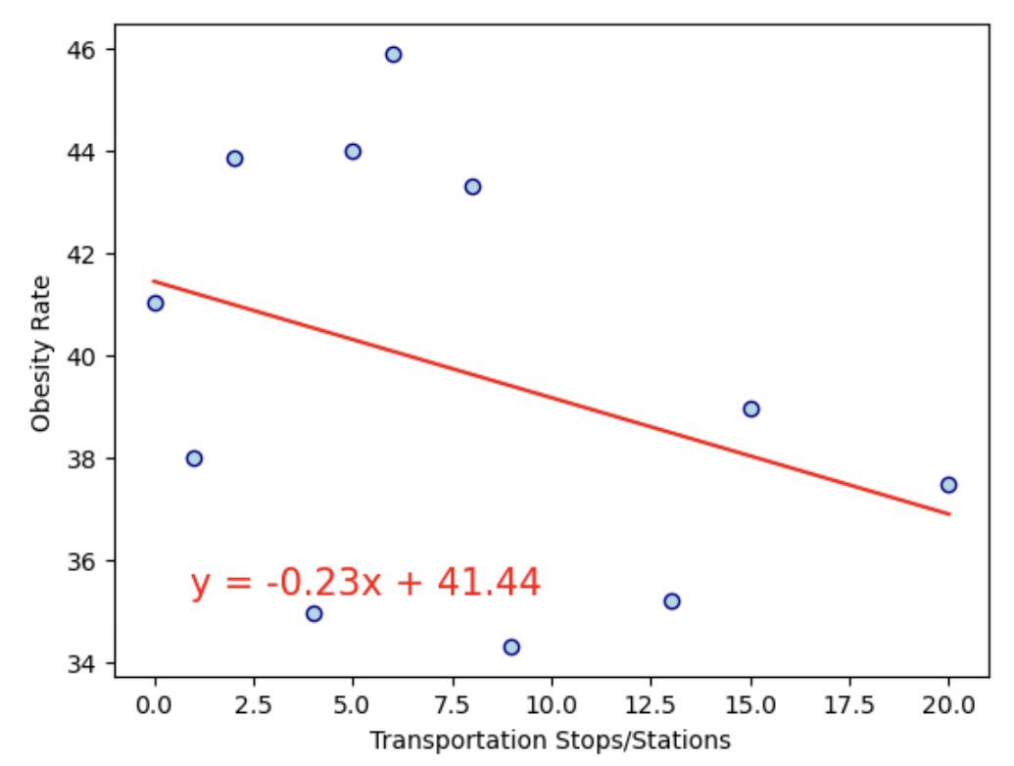
QUESTION 4: MAP

QUESTION 4: GRAPHS



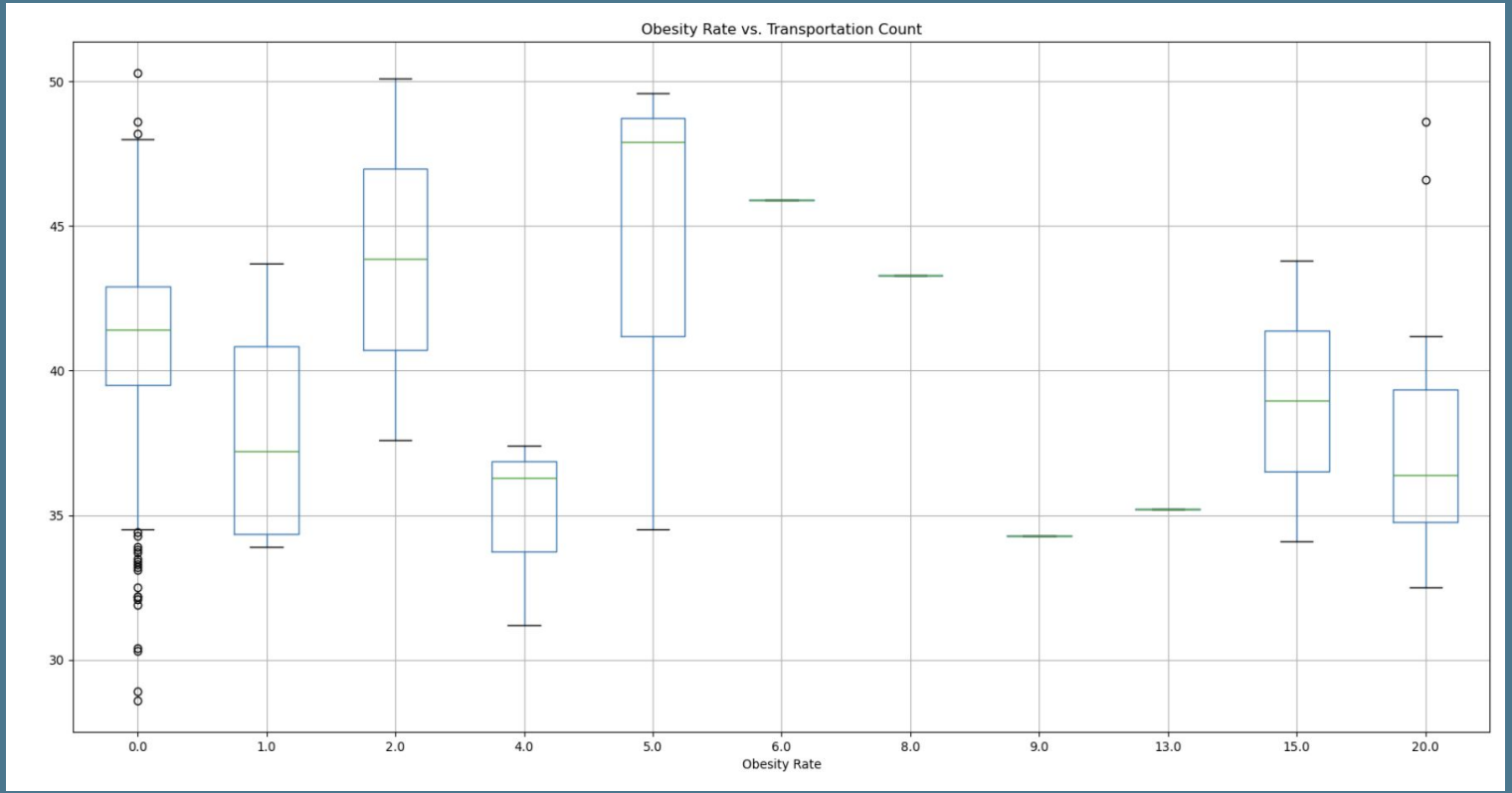


QUESTION 4: STATISTICAL ANALYSIS



The correlation between amount of transportation stops/stations and the obesity rate is very weak (<0.5), with an R^2 of 0.121038.

QUESTION 4: STATISTICAL ANALYSIS



pvalue= 5.6e-08

We can reject the null hypothesis that the amount of transportation stops/stations has no effect on the rate of obesity in a zip code.

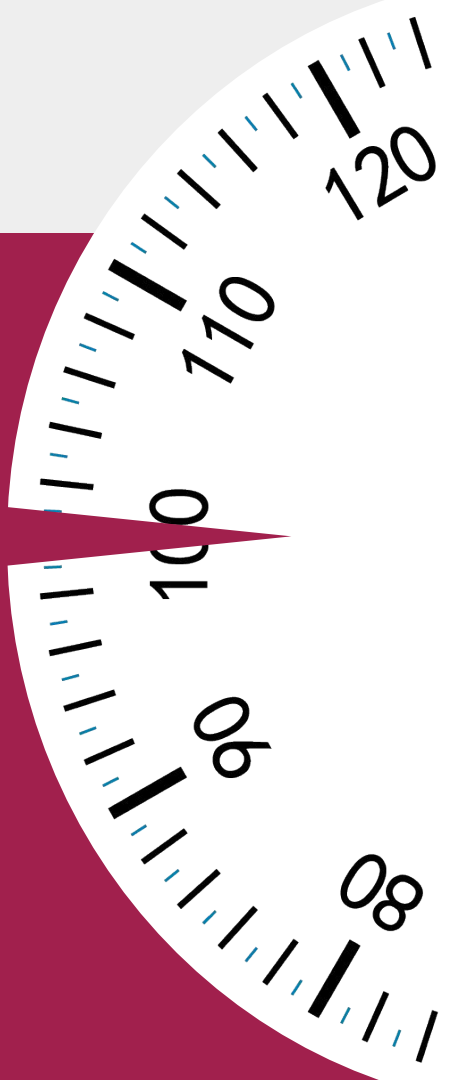
QUESTION 4: CONCLUSION

It appears that there is a minimal correlation between obesity rates and access to public transportation. The calculated R-squared value, which measures the proportion of the variance in one variable that is predictable from another, stands at a mere 0.12. This low R-squared value suggests that only a small fraction of the variability in obesity rates can be explained by access to public transportation. Thus, it seems that other factors beyond transportation accessibility play a larger role in influencing obesity rates.

Some limitations I found in the data was that many zip codes seemed to max out on the count for transportation stops/stations at 20. I believe this is from using api data from Geoapify and the limitations placed on free accounts.

RESOURCES

- Geoapify Location Platform
- SOI Tax Stats - Individual Income Tax Statistics- Oklahoma
- Local Data for Better Health, ZCTA Data



the end :)

QUESTIONS ?