

# **Impact of access to parks on the health of the community**

**Team 18**

**Ana Monay, Angel Keele, Angelica Villanueva, Breana  
Whittington & Latiff Evans**

**Final Report Requirements:**

[https://docs.google.com/document/d/1o8jckwra-Fwvj2ChLzhDuS\\_SRiJ4P9BkXukyOnhsppQ/edit?usp=sharing](https://docs.google.com/document/d/1o8jckwra-Fwvj2ChLzhDuS_SRiJ4P9BkXukyOnhsppQ/edit?usp=sharing)

**Tables of Contents**

<b>Introduction</b>	<b>2</b>
<b>Data</b>	<b>3</b>
<b>Data Analysis &amp; Computation</b>	<b>6</b>
- Correlation Matrix	
- Scatter Plots & Pair Plots	
- Statistical Model	
- OLS	
- Analysis of Variance	
<b>Results</b>	<b>8</b>
<b>Conclusion</b>	<b>10</b>
<b>References</b>	<b>12</b>

# Introduction

---

With heart disease being the leading cause of death in the United States, it's crucial to explore ways to improve public health and prevent the onset of this and other health ailments. Physical activity, as recommended by the CDC, is a proven way to improve health and lower the risk of heart disease and other health problems. The suggestion of physical activity is somewhat vague. There are those who want to be more physically active, however, there may be a barrier of finding accessible and affordable ways to engage in physical activity can be a challenge, particularly for those living in urban areas.

Parks provide a unique solution to this problem, offering a wide range of opportunities for physical activity such as walking paths, trails, basketball courts, and tennis courts - all for free or low cost. This presents an opportunity to investigate the potential impact of increasing the number of parks in a city on the health of its citizens. By exploring this relationship, we hope to inform policy decisions and contribute to a healthier, more active community.

Our hypothesis is that cities with higher population-to-park ratios will have higher rates of certain health ailments. We believe that limited access to green spaces and recreational opportunities may contribute to poor health outcomes.

To test this hypothesis, we will examine the relationship between the ratio of population to parks in cities and the incidence of health problems such as heart disease, obesity, high blood pressure, and other health conditions.

We believe that if a city has a lower population-to-park ratio, it means that its citizens have greater access to parks and recreational opportunities, leading to increased access to physical activity and better health outcomes. It's important to note that while the raw number of parks in a city is important, the population-to-park ratio provides a more meaningful indication of accessibility and the potential impact on public health.

# Data

---

For this study, we used data from 4 counties in Texas:  
Denton, Williamson, Montgomery, and Harris. Our  
sample locations were chosen based on the availability  
of data and represent a diverse range of urban  
environments.

Denton	26
Harris	14
Montgomery	6
Williamson	8

To calculate the population-to-park ratio, we used data on the population and park count  
for each of the 4 counties. This data was collected from a variety of sources, including  
county data resources and the online platform koordinates.com. We focused on data  
from the year 2019 to minimize any biases due to changes in regulations or  
pandemic-related policies.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 54 entries, 0 to 53
Data columns (total 17 columns):
 #   Column           Non-Null Count  Dtype  
---  --  
 0   LocationName     54 non-null    object  
 1   COPD             54 non-null    float64 
 2   Cholesterol Screening  54 non-null  float64 
 3   Chronic Kidney Disease 54 non-null  float64 
 4   Coronary Heart Disease 54 non-null  float64 
 5   Diabetes          54 non-null    float64 
 6   High Blood Pressure 54 non-null    float64 
 7   High Cholesterol   54 non-null    float64 
 8   Obesity            54 non-null    float64 
 9   Physical Health    54 non-null    float64 
 10  Physical Inactivity 54 non-null    float64 
 11  Stroke             54 non-null    float64 
 12  Taking BP Medication 53 non-null    float64 
 13  TotalPopulation    54 non-null    int64  
 14  ParkCount          54 non-null    int64  
 15  people_park_count_ratio 54 non-null  float64 
 16  County Name        54 non-null    object  
dtypes: float64(13), int64(2), object(2)
memory usage: 7.3+ KB
```

The health data used in the study was obtained from the Centers for Disease Control  
and Prevention (CDC). We used the CDC's annual Places dataset and the Behavioral

Risk Factor Surveillance System (BRFSS) survey data to gather information on the prevalence of stroke, COPD, coronary heart disease, diabetes, obesity, high blood pressure, and high cholesterol in those counties. There were additional columns kept to refer to from the CDC: chronic kidney disease, cholesterol screening, physical health, physical inactivity, and taking BP medications. The age-adjusted percentages of the population affected by each condition were calculated based on the sample data collected from the survey. Our focus was on physical health conditions that have the potential to be prevented through physical activity. The population data for each city was also obtained from the CDC. We ended up with 54 cities between the 4 counties.

Cleaning the data included removing numerous unrelated columns, creating park counts for cities, merging city and park counts with CDC data, removing extra letters, words, or splitting strings (\*ex: "City of", pulling the city only from the address, keeping consistent capitalization), and calculating the population-to-park ratio. The health ailments in the CDC dataset are in one column. The dataset had to be pivoted to create individual columns for each health ailment.

All\_cities\_cdc dataset created and used for analysis.

Column Name	Data Type	Description
LocationName	object	City Name
COPD	float	Chronic obstructive pulmonary disease Data Value - Age-Adjusted Percentage
Cholesterol Screening	float	Cholesterol Screening Data Value - Age-Adjusted Percentage
Chronic Kidney disease	float	Chronic Kidney disease Data Value - Age-Adjusted Percentage
Coronary Heart Disease	float	Coronary Heart Disease Data Value - Age-Adjusted Percentage
Diabetes	float	Diabetes Data Value - Age-Adjusted Percentage
High Blood Pressure	float	High Blood Pressure Data Value - Age-Adjusted Percentage
High Cholesterol	float	High Cholesterol Data Value - Age-Adjusted Percentage
Obesity	float	Obesity Data Value - Age-Adjusted Percentage

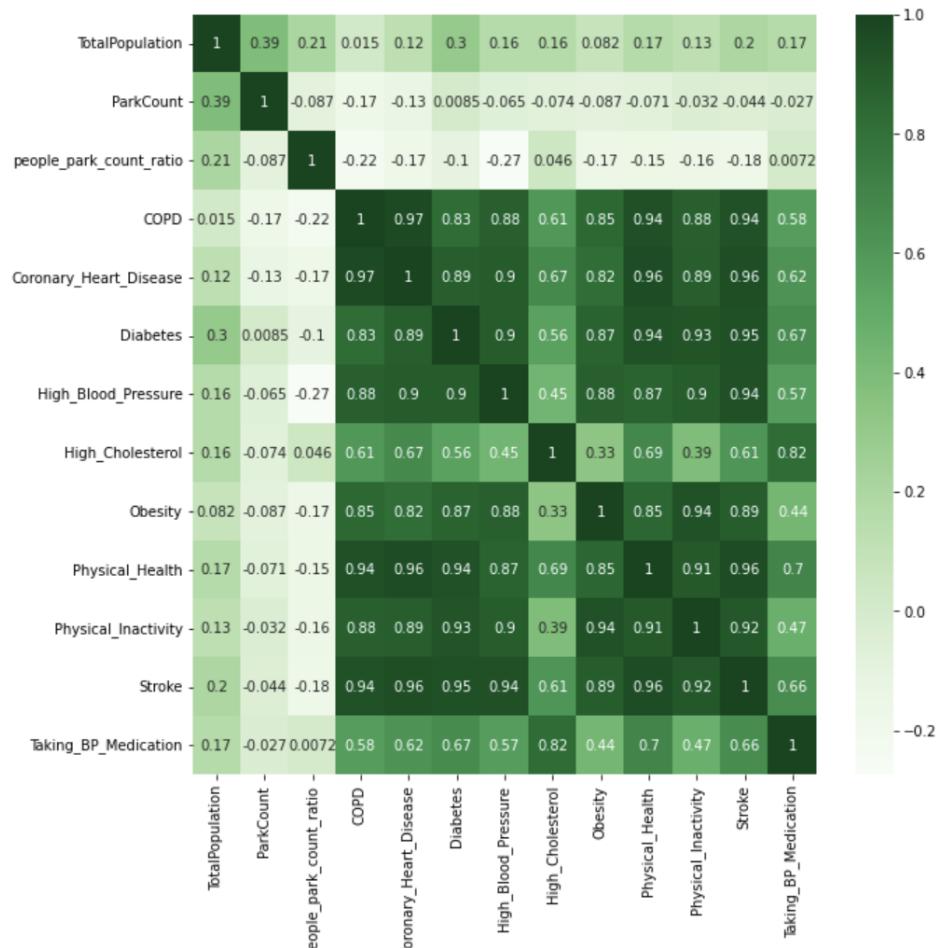
Physical Health	float	Experienced negative physical health in the last 14 days. Data Value - Age-Adjusted Percentage
Physical Inactivity	float	Been physically inactive for the last 14 days. Data Value - Age-Adjusted Percentage
Stroke	float	Stroke Data Value - Age-Adjusted Percentage
TotalPopulation	int	City Total Population
ParkCount	int	Number of parks per city city
people_park_count_ratio	float	<i>TotalPopulation</i> divided by <i>ParkCount</i>
County name	object	County Name

# Data Analysis & Computation

Since we had a smaller sample size of 54 cities, we wanted to look at a few methods. Initially, we started by looking at the data and comparing the health ailments and park counts for each city and county. A correlation matrix was created and visualized to see if there were any obvious relationships to look into further.

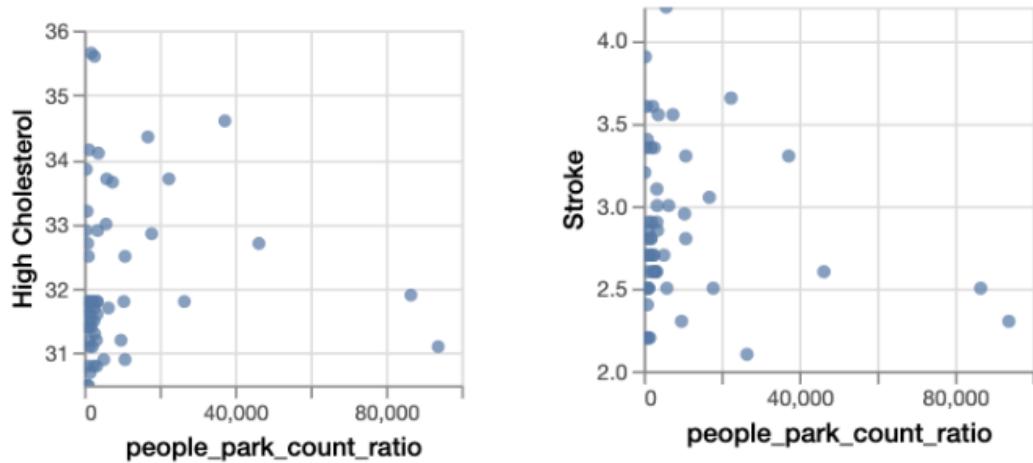
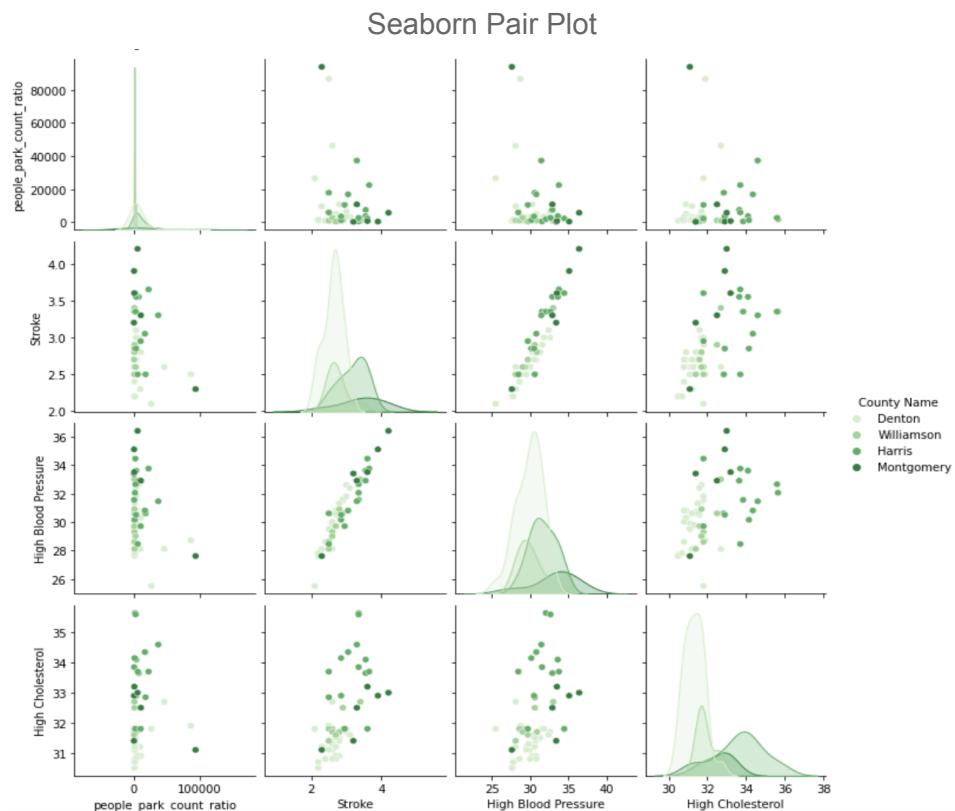
## Correlation Matrix

The initial overview gives some weak correlations between some of the health conditions and people to park ratio.



## Pair Plots & Scatter Plots

To get a better visualization of the models there was some visualization done. Seaborn pair plots, various scatter plots, and regression-specific plots were created.

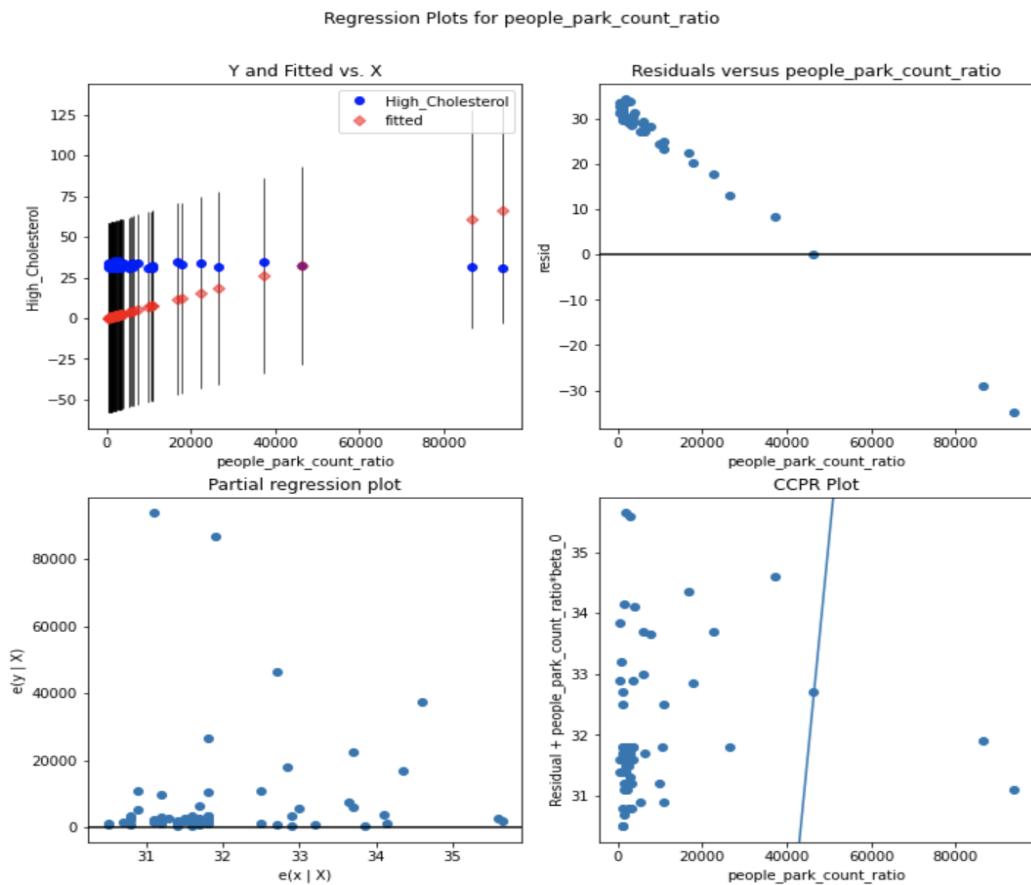


# Statistical Analysis

The analysis used two statistical methods to analyze the data on the relationship between the population-to-park ratio and health outcomes in four counties in Texas. The methods used were ANOVA (Analysis of Variance) and OLS (Ordinary Least Squares) regression.

## Linear Regression

OLS was used to conduct a linear regression analysis to investigate the relationship between the population-to-park ratio and the prevalence of different health outcomes. It found the line of best fit that minimizes the sum of the squared differences between the predicted and actual values and determines the strength and direction of the relationship between the variables. OLS is a robust method that can handle missing data and outliers, making it suitable for this analysis.



## Analysis of Variance

ANOVA was used to compare the means of health outcomes such as stroke, diabetes, obesity, and others, between the different cities in Texas. It allowed for the determination of significant differences in the means of these health outcomes between the cities and the use of a continuous variable, population-to-park ratio, as an independent variable. However, this may not be the best choice for the data since it does not do as well with the continuous relationship between park ratios and health conditions. When we did run it, there was some significance for high blood pressure. If the park data were grouped more in categories (ex: low, med, high for park ratios) or more variables that could affect health to add (environment, air quality, stress levels) this might vie a better analysis. [reword →] There were also huge differences in results when the number of health outcomes changed for variables in the analysis.

Obesity					
	sum_sq	df	F	PR(>F)	
people_park_count_ratio	3.271807	1.0	2.541411	0.116956	
Residual	66.944675	52.0	NaN	NaN	
High_Blood_Pressure					
	sum_sq	df	F	PR(>F)	
people_park_count_ratio	18.360478	1.0	4.229954	0.04475	
Residual	225.710448	52.0	NaN	NaN	
High_Cholesterol					
	sum_sq	df	F	PR(>F)	
people_park_count_ratio	1.401867	1.0	0.539945	0.465757	
Residual	135.008364	52.0	NaN	NaN	
Stroke					
	sum_sq	df	F	PR(>F)	
people_park_count_ratio	1.401867	1.0	0.539945	0.465757	
Residual	135.008364	52.0	NaN	NaN	

# Results

---

The study found that there is a significant relationship between the ratio of population to parks and three health ailments: High Blood Pressure, High Cholesterol, and Stroke. The results showed that as the ratio of population to parks increases, so does the prevalence of these health issues. This suggests that a lack of access to parks may be associated with these health problems.

OLS Regression Results						
	coef	std err	t	P> t	[ 0.025	0.975 ]
Dep. Variable:	High_Cholesterol	R-squared (uncentered):	0.199			
Model:	OLS	Adj. R-squared (uncentered):	0.184			
Method:	Least Squares	F-statistic:	13.17			
Date:	Mon, 23 Jan 2023	Prob (F-statistic):	0.000641			
Time:	03:37:55	Log-Likelihood:	-258.07			
No. Observations:	54	AIC:	518.1			
Df Residuals:	53	BIC:	520.1			
Df Model:	1					
Covariance Type:	nonrobust					
people_park_count_ratio	0.0007	0.000	3.629	0.001	0.000	0.001
Omnibus:	67.811	Durbin-Watson:	0.439			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	477.077			
Skew:	-3.550	Prob(JB):	2.54e-104			
Kurtosis:	15.713	Cond. No.	1.00			

Notes:  
[1] R<sup>2</sup> is computed without centering (uncentered) since the model does not contain a constant.  
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

The R-value for these were all above .15, around .2. This means there was a significant amount of variance in the data.

However, it is important to note that this is just a correlation and more research is needed to determine if there is a cause-and-effect relationship and if these findings are applicable to other regions or populations. The data used in the study was collected from four counties in Texas and came from the Centers for Disease Control and Prevention (CDC).

Overall, the results suggest that increasing access to parks may have a positive impact on reducing the prevalence of High Blood Pressure, High Cholesterol, and Stroke, but more research is needed to confirm these findings and understand the underlying mechanisms.

## Conclusion

---

In conclusion, the results of this study show a significant correlation between the ratio of population to park count and the prevalence of High Blood Pressure, High Cholesterol, and Stroke in four counties in Texas. This suggests that there may be a relationship between a lack of access to parks and these health issues. However, it is important to keep in mind the limitations of this analysis, such as the presence of other variables that can affect someone's health, the different types of parks and their suitability for physical activity, and the accessibility of the parks to citizens. These limitations highlight the need for further research to better understand the link between access to parks and health outcomes.

## References

---

*PLACES: Local Data for Better Health, Place Data 2020 release*

<https://chronicdata.cdc.gov/500-Cities-Places/PLACES-Local-Data-for-Better-Health-Place-Data-202/q8xq-ygsk>

*PID Parks Points (Harris County, TX Parks)*

<https://geo-harriscounty.opendata.arcgis.com/datasets/pid-parks-points/explore?location=29.825164%2C-95.407200%2C10.44>

*Williamson County, Texas Parks and Open Spaces*

<https://koordinates.com/layer/99384-williamson-county-texas-parks-and-open-spaces/data/>

Denton County, Texas Parks

<https://koordinates.com/layer/98122-denton-county-texas-parks/data/>

*Montgomery County, Texas Public Parks*

<https://koordinates.com/layer/99222-montgomery-county-texas-public-parks/>

