ASN11 – Final Research Project Report

Team Name	Team Gainz
Individual Student Names	Adam Kaushik, Richard Liang, Darwin Ma, Nikolay Pomytkin

Introduction

The project that we are working on is part of the field of public health. To be more specific, our group's goal is to research the relation between public health, socio-economic status, and the distance away sample populations are away from cities. Of these three factors, the main one that we want to focus on is public health and healthcare availability in the United States.

This topic is interesting because along with allowing us to see the relationship between income and public health, public health is a very broad field. Public health not only encompasses the physical health of individuals, but also their mental health. In addition, a person's current physical and mental health can cause them to make better or worse decisions, possibly skewing the public health of lower and higher income individuals left and right. Furthermore, a person's quality of health is directly related to their quality of life and many other factors. Relating public health to politics, we can see how recent events regarding federal laws about healthcare, like the redaction of the controversial Affordable Care Act, may alter this study's results in comparison with previous similar experiments. Another factor in public health is urban sprawl, which is defined as the expansion of populations into suburban areas from urban cities. It is important to analyze the difference between the suburbs and urban environments in terms of public health availability, especially considering the vastly different lifestyles associated with each of these developed human settlements. Public health is something which can be tied into overall quality of life, and there has always been groups of people throughout history who have not enjoyed the steady improvements in quality of life others have. The information obtained could be used to start determining a way to increase the public health and thus the quality of life of certain groups, depending on our research. However, it is also very hard to find a solution to a socioeconomic gap. The socioeconomic gap in our current society is very hard to overcome, as societal balances have been deeply rooted in our society since Our topic will further branch out and relate to great "problems of our time," and demonstrate the relationship between public health and urban sprawl, income, etc…, and will also prove useful for finding optimal residence locations for us and other college students after completing our degrees. Overall, the unique relationship between urban sprawl, socioeconomic status, and public health, and the practicality of our project is what compelled us to make this study.

This topic certainly has societal and scientific relevance to some of the great "problems of our time." This topic is applicable to the issues of social inequality, poverty, lack of education, and access to healthcare. In our topic, we discussed how the socioeconomic gap, income, and distance away from cities affect public health. What we found out is that people with a lower income, and thus in a lower socioeconomic class have significantly worse health and are more likely to engage in risky behaviors. This relates to social inequality, poverty, and lack of education. The gap in social inequality has only been getting wider over the years. The wealthy are becoming wealthier and the poor are becoming poorer. As the income gap in society continues to grow, access to healthcare is also becoming more limited for those with a lower income. As stated before, we have found in our research that people with lower income and in a lower SES have worse health and partake in risky behaviors. This further ties into the poverty and education problems. Since the poor have a greater number of health problems and aren't as aware of how to prolong their lifespan (which may be a reason they have risky behaviors), people in a lower SES

have a lower social standing and are generally looked down upon by others, sometimes because of a lack of education. When the poor are facing all these health problems and money issues, they find it hard to prioritize the future, and have little incentive to study hard and obtain a good education. This leads to limited options when they need to make money in the future, and with a lowered education status comes an unequal social standing with others who have diplomas, have degrees, etc…. Furthermore, apart from being related to great "problems of our time," our topic can be further related to other things. For example, for someone in a lower SES with social inequality and income inequality, they may be unable to have self-respect or be able to afford health insurance, regular check ups, etc… This leads to an increase in stress, depression, and tendency to take harmful substances, and a decrease in confidence to reach out and try to succeed. Thus, our topic directly ties not only into some of the great "problems of our time", but also to other important issues. The socioeconomic gap, income, and public health all only serve to widen the social inequality gap, and leave those in a low SES with more poverty and education problems.

There are many parties that could be interested in novel analytical insights in this field. Since our topic affects the lives of every American, it can be a topic of interest to anyone living in the United States. Poor Americans may be interested to see how their health compares to that of the wealthy. It may also serve as a baseline and help poor Americans become more aware as to how much they are lacking in terms of public health and longevity. Likewise, those living in urban areas may be curious to see how their health compares to that of those living in more rural places. People looking to move from a more rural area to an urban environment, and vice-versa, may also be interested in our analysis. Those who would be especially intrigued are those interested in seeing the effects of the wage gap or investigating the differences in quality of life for people living in areas more or less populous than others. Additionally, the wealthy may also be interested in these studies, as it will enforce their beliefs that what they are doing is correct, and that they are taking good steps in maintaining their public health and longevity. Those in business could potentially be interested in our data. For example they could market a product which increases general health, and make it targeted to those with a poor income. Those selling homes could also use our data to make homes in a denser area more attractive, as it will lead to a higher health. People in medicine will definitely find the results of our data worthwhile, as it shows a clear disparity in health. There is much room for improvement in those with a lower income, and much less improvement in those with a higher income, which may affect how people in medicine try to increase the public health of Americans. More specifically, politicians may be interested in our analysis as it may help them decide what health-related reforms to push for based on where they work. Furthermore, this may also be a topic of interest for other countries. They may use analytical insights from studies in this field to compare their overall public health with ours, and perhaps look to add funding or decrease funding into research for public health.

Analytics is a relatively new part of public health research, however in recent years it has become a more important aspect of it. As the world transitions to a digital era, much more data is available, especially regarding population health. Analyzing all of this data will lead to new insights with regards to how to keep the general population healthier and how to increase health care accessibility. As more and more data is being collected in a variety of different forms, there will definitely be a need for analytics in this field, as it will help give our society a better understanding of what steps need to be taken in order to

improve public health and get more people access to higher quality healthcare. These insights will be incredibly important to organizations both in the private and public sector, especially government agencies/officials that seek to improve public health and healthcare accessibility, which recently have been very important issues that the both local, and federal governments have tried to tackle (Philadelphia's sugar tax and the Obamacare at the federal level). Overall, analytics already is, and will continue to be, a very important part of public health research, and it will be crucial in providing the many insights that our society needs in order to improve public health.

Data Sources

NAME	DESCRIPTION	URL
Community Health Status Indicators (CHSI) to Combat Obesity, Heart Disease and Cancer	This dataset provides key health indicators for local communities and encourages dialogue about actions that can be taken to improve community health (e.g., obesity, heart disease, cancer). The CHSI report contains over 200 measures for each of the 3,141 United States counties collected in 2012.	https://healthdata.gov/dataset/community-health-status-indicators-chsi-combat-obesity-heart-disease-and-cancer?fbclid=IwAR3iXPW1-bAi-o69VrOPQm0C0DgG5xMqO80i-ckCBJ6-Az6_w6EDThLpCKs
US Census Demographic Data	This dataset was compiled from the 2015 government census data for all census tracts. It includes population and income data for over 3,000 United States counties.	https://www.kaggle.com/muonneutrino/us-census-demographic-data
Chronic Disease Indicators	This dataset is a compiled from the Center for Disease control and contains data that tracks chronic disease indicators in all 50 states of the U.S.	https://www.kaggle.com/cdc/chronic-disease
Health, Nutrition, and Population Statistics	This dataset contains data on different health, nutrition, and population indicators throughout the world from 1960-2016. This shows the progression of how health indicators and population sizes have changed over the last 50 years.	https://www.kaggle.com/theworldbank/health-nutrition-and-population-statistics#data.csv

Results Achieved

The health indicators in all of the graphs are compiled from the CHSI dataset, which contains a variety of data on public health in thousands of U.S. counties. A higher indicator value indicates that the individuals in a county are on average **more** unhealthy that their counterparts in a county with a lower indicator value.

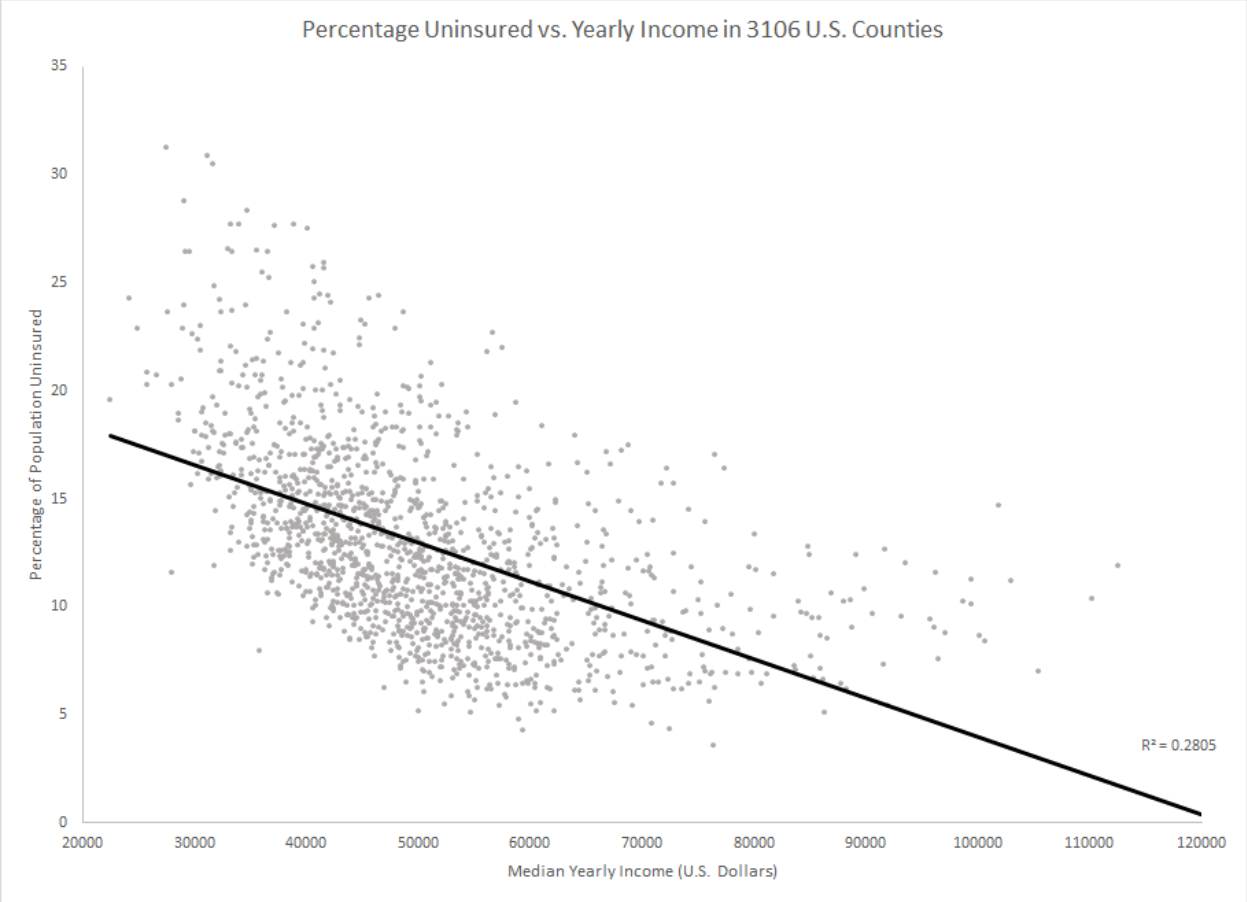
The **Obesity indicator** is the calculated percentage of adults at risk for health problems related to being overweight, based on body mass index (BMI) in each county. A BMI of 27.8, for men, and 27.3, for women, or more is considered obese.

The **High Blood Pressure indicator** is the percentage of adults who responded yes to the question, "Have you ever been told by a doctor, nurse, or other health professional that you have high blood pressure?" in each county.

The **Physical Inactivity indicator** is the calculated percentage of adults at risk for health problems related to lack of exercise in each county.

The **Uninsured Percentage** is the estimated number of uninsured individuals in the county from the 2013 Current Population Survey (CPS) divided by the total population in each county multiplied by 100.

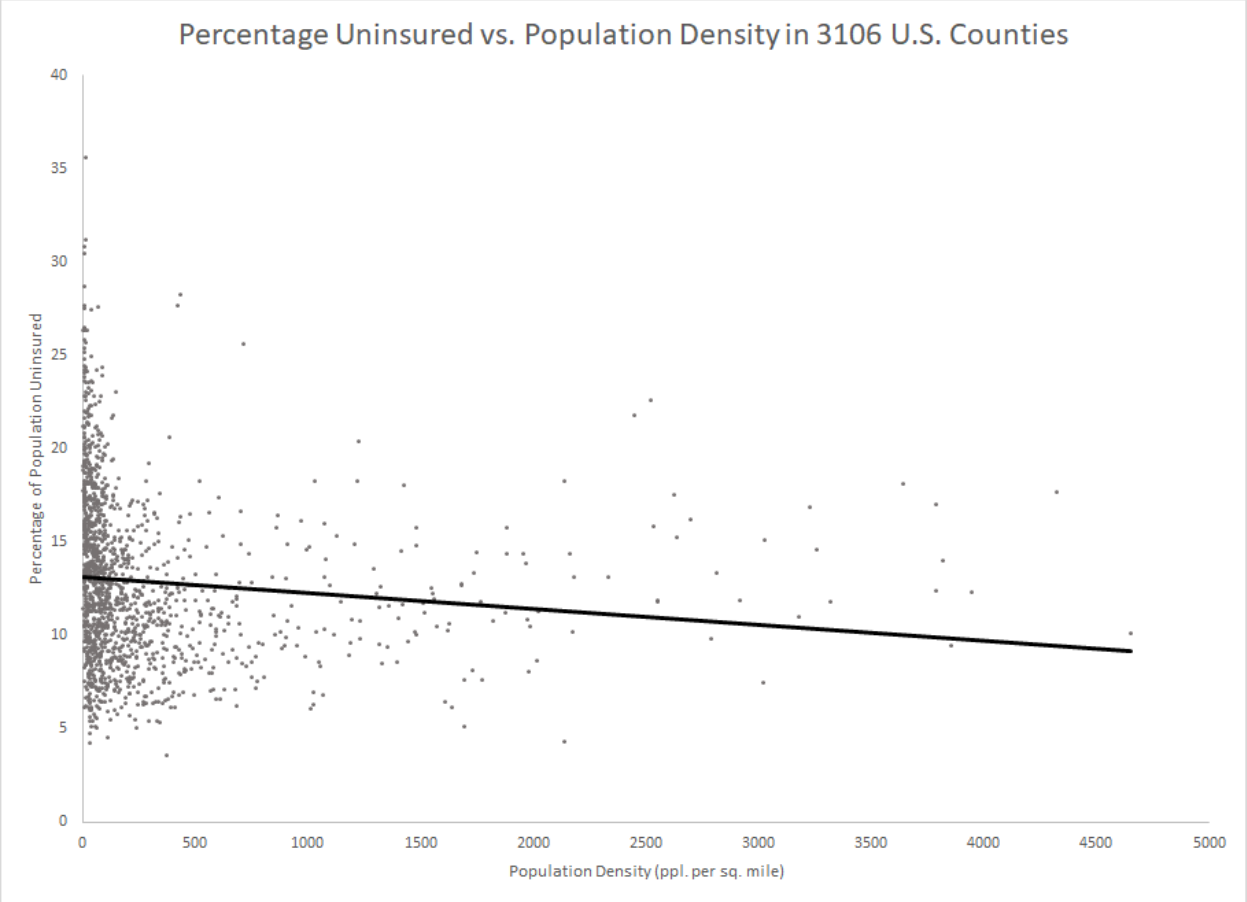
The graphs also contain income data (Median Yearly Income in each county) which was taken from the 2015 U.S. Census dataset. The population density per county data was taken from the CHSI dataset.



One of the main characteristics we were interested in studying was healthcare availability and how it varies with SES. The percentage of population uninsured is a good measurement of healthcare accessibility as it shows how many people have affordable access to quality healthcare. In counties with a high percentage of people uninsured, it is likely that there is lower access to quality healthcare, and it is also much more expensive to use that healthcare without proper insurance. This graph, which shows the relationship between the percentage of people without healthcare insurance and the median yearly income in over 3,000 U.S. counties, has a clear trend line that shows how healthcare access increases as median income increases.

Of all of our data, this graph had the best correlation with an R-squared value of 28%, which shows that there is definitely variance in the dataset, however the data is definitely pretty close to the trend line.

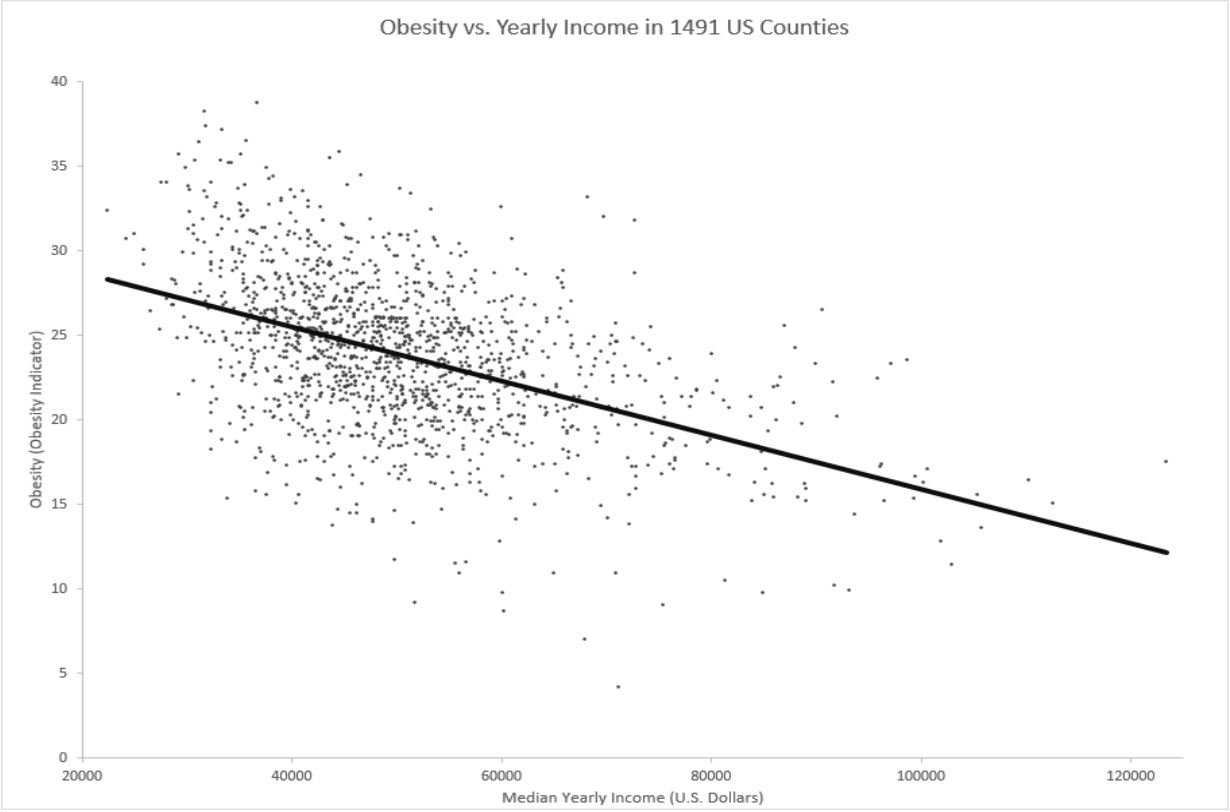
In counties where the median income is between $20,000 and $60,000, the data points trend downwards quite steeply, showing that there is a huge discrepancy in healthcare access between those at the lower and higher ends of that range. After the median yearly income passes $80,000, the percentage uninsured tends to level out, at around 10%, which shows that after a certain income level, healthcare accessibility becomes relatively steady.



Because insurance is the very definition of healthcare availability, our group thought it a good idea to correlate that to population density, a measure of how far or close counties are from urban environments.

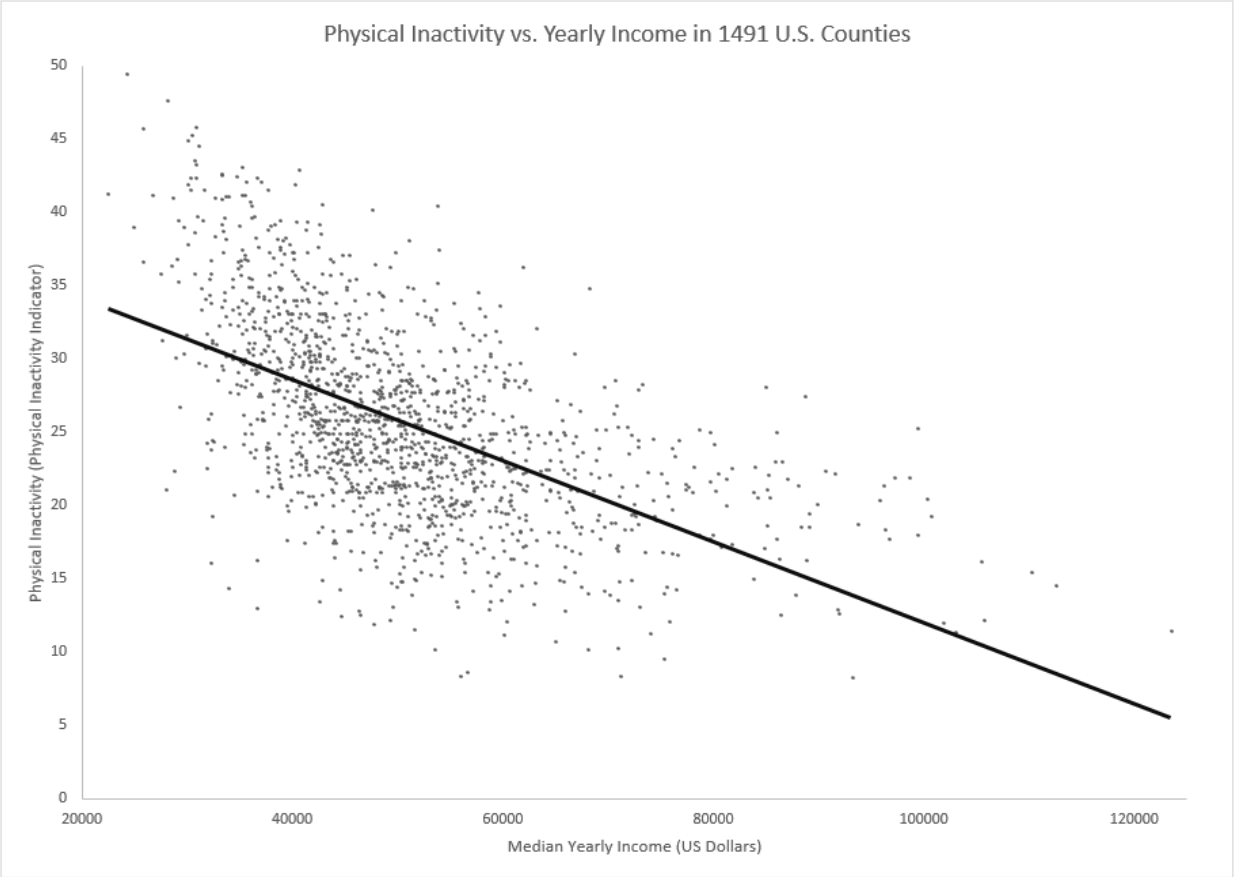
There is a lot of variety for the data representing counties with population densities of 500 or below, which skewed our trend line a little; however, we can still see that the percentage of the population that is uninsured goes down as population density increases. We noticed that in this graph, the highest percentage of population that is uninsured went down from around 30% to 20% as population density went up.

We suspect that this is the case since people living in urban areas (i.e high population density) not only have higher accessibility to basic commodities like healthcare, but they are in general wealthier, allowing them to purchase better healthcare. On the other hand, those in the suburbs or rural areas (i.e low population density) are not able to access commodities as easily as urban residents. This, along with the fact that suburban and rural residents are of average or low wealth, meaning that they inherently will not be able to get the healthcare that urban residents can get.



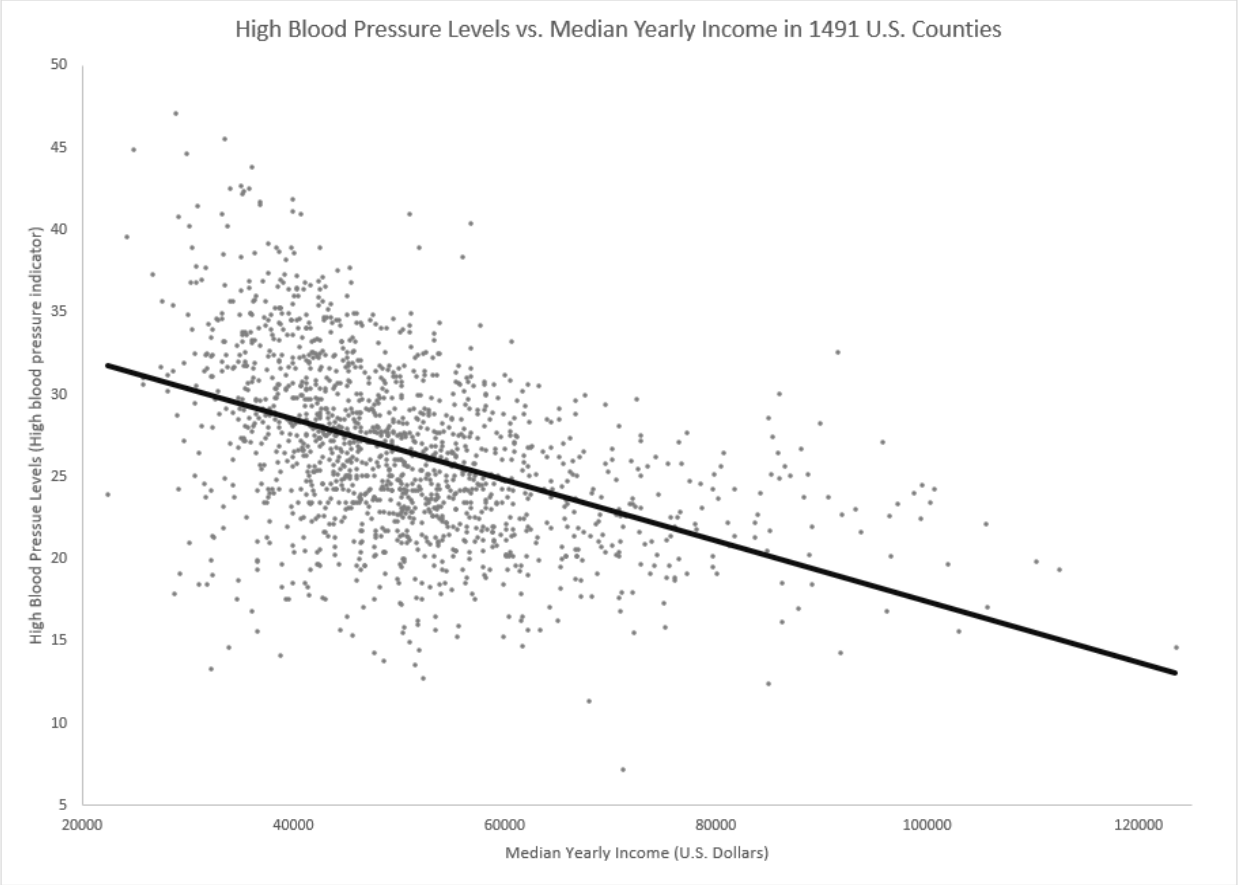
The figure above depicts a scatter plot with a trend line delineating the correlation between median salaries and obesity in 1491 counties spread evenly across all 51 U.S states. Our group analyzed this graph primarily by using the trend line. Without it, all the plot points representing counties with median yearly incomes of approximately $50,000 would look like a disarray. However, with the trend line, we saw that there was a clear linear relationship between salaries and obesity. Even for counties with median incomes between about $40,000 to $50,000, one can see that as you get closer and closer to the trend line, the denser the plot points. Overall, we see that as salaries increase, obesity levels decrease.

The pattern we noticed with this graph was that overall, the wealthier a county was, the lower that county's obesity indicator would be. Specifically, that meant that the county would have a lower percentage of adults at risk of health problems related to obesity. For counties with incomes around $30,000 to $40,000, the highest obesity indicator was at almost 40%, and for counties with incomes over $60,000, the highest obesity indicator was at about 25%.



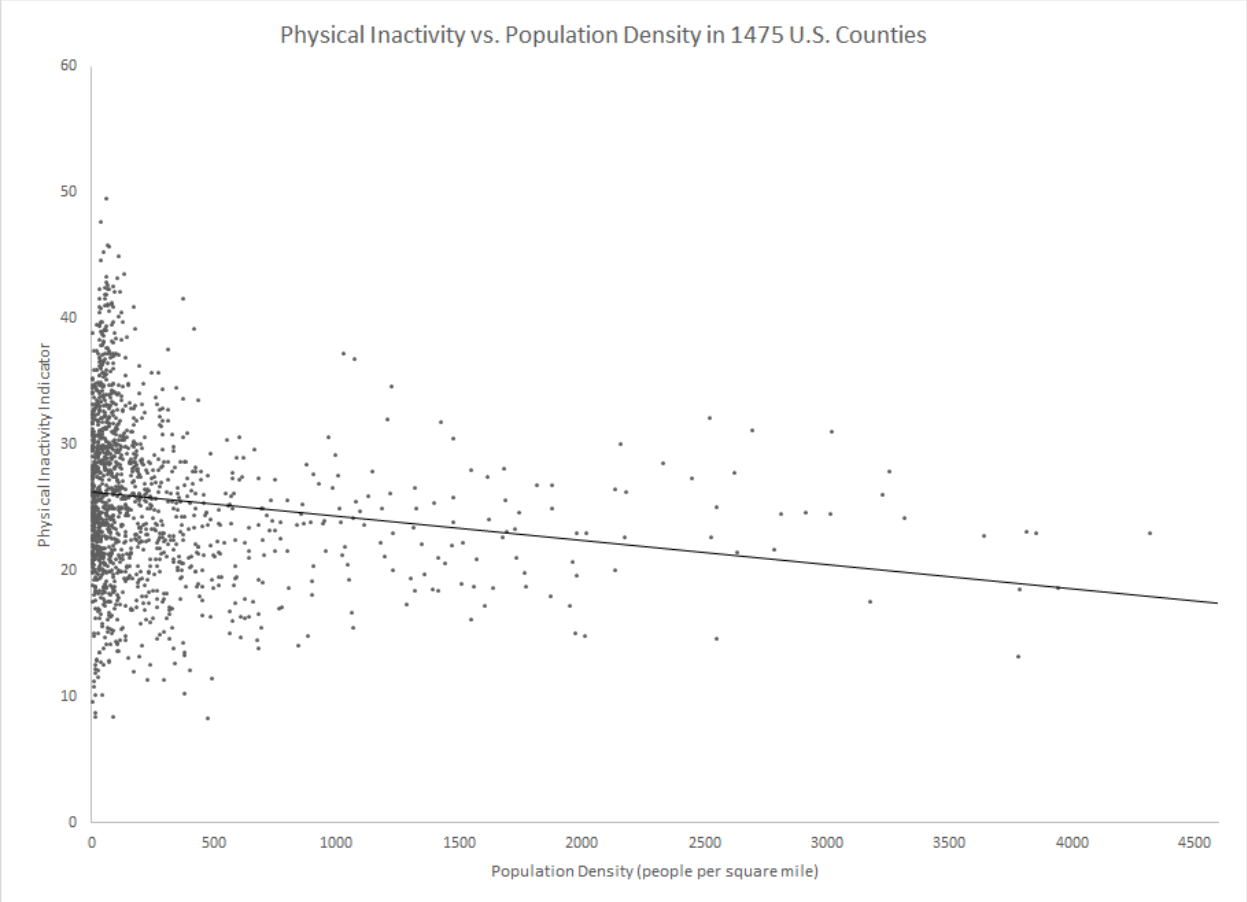
Similar to obesity, another variable that we thought was relevant to our study was physical inactivity. Previous research noted that due to urban sprawl, commodities that would have been within a few blocks of one another in urban areas would be much farther apart in suburban regions. As a result, those who reside in the suburbs need to drive to access commodities such as the grocery store; meanwhile, urban residents simply need to walk a few minutes to get what they need. Given the disparity between median household incomes and physical activity levels of different developed settlements, relating income and physical inactivity levels could be critical to our research.

The figure above is a scatter plot that depicts the effect income has on physical activity. Our datasets only had results for physical inactivity, so the negative correlation means that on the grand scale, people become more active as their incomes increase. However, there was a wide range of physical activity levels in counties with median incomes between $40,000 and $60,000. This discrepancy may be a result of the different environments affiliated with different U.S counties.



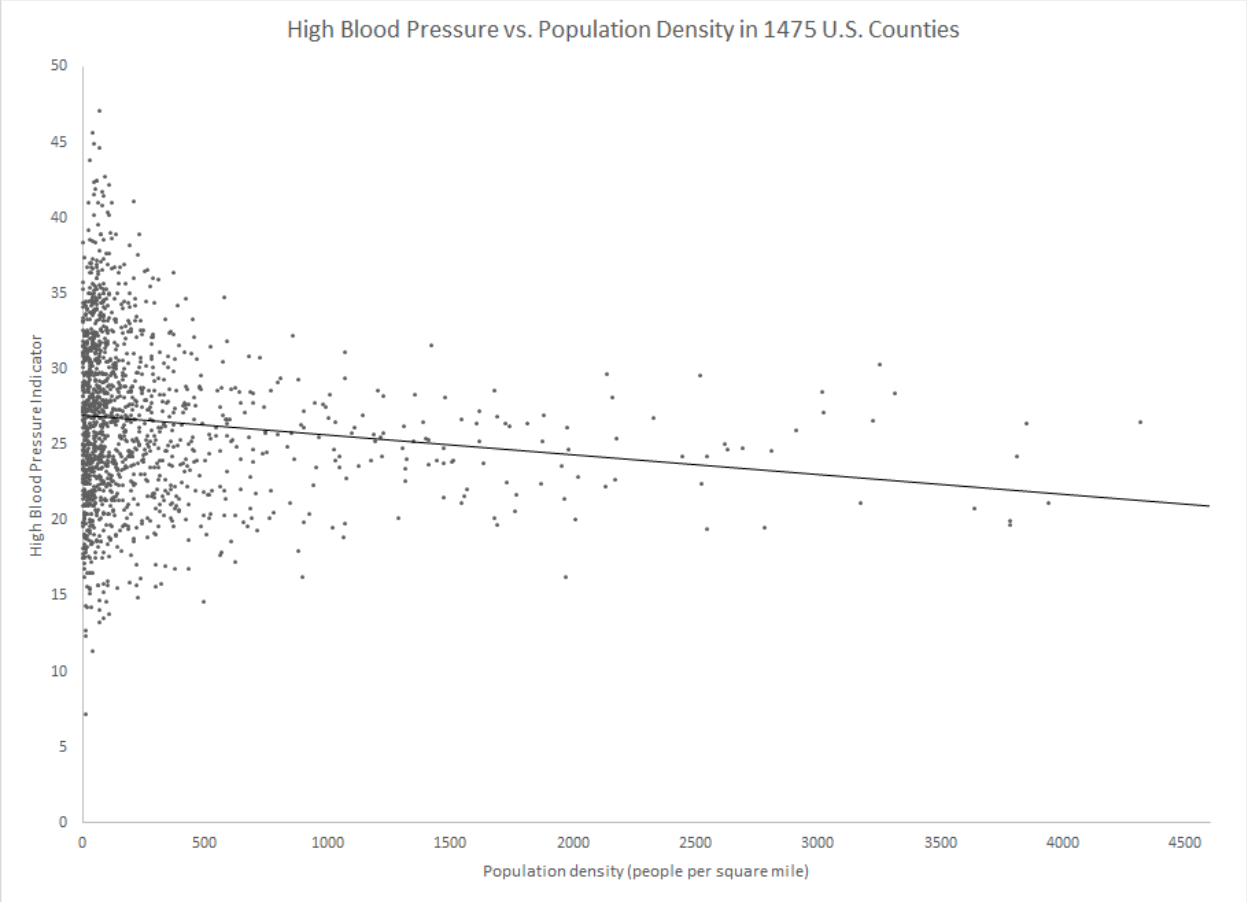
High blood pressure is a problem typically caused by poor diets and lack of exercise, which are also problems instigated or mitigated by one's socioeconomic status. We concluded that high blood pressure would be useful to us when compared to income. Furthermore, high blood pressure is another ailment that could lead to a variety of serious health issues.

Shown above is a scatter plot with each plot point representing a county's high blood pressure levels in relation to the average household income of a resident of that county. Like the previous graphs, our team perceived an overall negative correlation between high blood pressure levels and median yearly incomes. However, unlike the previous graphs, the plot points that represent counties with median yearly incomes between $40,000 and $50,000 seem to be more random. Essentially, wealthier families are more likely to be healthy than more economically challenged families, supposedly due to those families' greater accessibilities to resources. because the plot points are more spread apart than the previous graphs, there could be more factors other than income that may affect people's blood pressures.

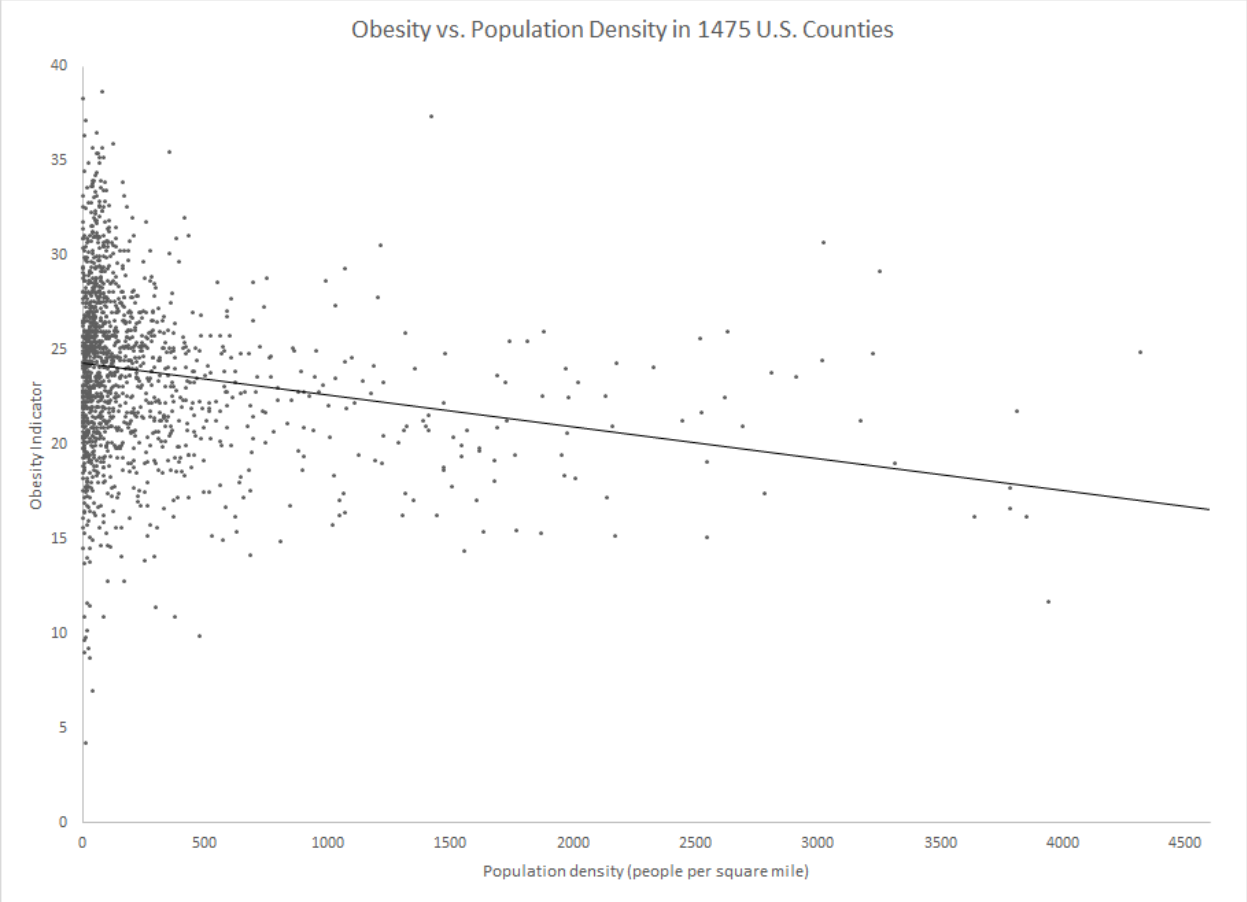


In addition to household income, our group is interested in how living in an urban, suburban, or rural environment could impact individual's health. One metric in our dataset that connects to living environment is population density. Since each dot on the graph represents a county, the dots closer to the right side of the graph are more likely to represent counties closer to cities and in a more urban environment, while those closer to the Y-axis represent counties in more rural areas.

By comparing population density to different health indicators, we can determine if living environment has an effect on public health. Although there is a clustering of counties with population densities between 0-500 ppl/sq. mile which have a very wide range of physical inactivity indicator values, this disarray decreases as population density increases, and there becomes a clear correlation between physical inactivity and population density. Overall, this graph shows that as population density increases physical inactivity decreases, which means that those living in more densely populated regions on average have more active lifestyles than those living in less densely populated regions, although there is a wide range of physical inactivity levels in the least densely populated regions.



This graph shows the relationship between high blood pressure and population density in 1475 U.S. counties. As shown by the trendline, high blood pressure cases are lower in more densely populated regions. Although there is a clustering of spread out data points in the lower population density portion of the graph, on average the blood pressure indicator is higher in counties with low population density as compared to counties with higher population density. Since high blood pressure is closely related to lifestyle habits like diet and physical activity, this graph can explain how those living in more densely populated areas tend to make healthier lifestyle choices, thus leading to a less cases of high blood pressure in these regions.



The final indicator that we included in our analysis is obesity, and its correlation with population density. Compared to the other two indicators that we analyzed in relation to population density, obesity seems to be the health indicator that most strongly correlates with population density. Although there is definitely some data points in the lower population density counties with obesity values far above and below the trendline, which may result from these countries having lower populations and thus less individuals to get data from, the trendline shows that as population density increases, obesity definitely decreases on average.

Literature Support

Provide the **scholarly** literature references related to your proposed field of study. For **each reference**, provide a short paragraph summarizing the paper.

Fred P, Patrick K, Justin D. (2010). *Socioeconomic Disparities in Health Behaviors.* Annual Review of Sociology, *36*, 349-370.

This article talks about how people with lower socioeconomic status(or SES) tend to engage in activities which are typically detrimental to their health, such as smoking and illegal activities. The article states that this may be caused by social interactions, coupled with the fact that people with a lower SES are typically unable to afford things which can improve their general health, such as regular medical checkups and a good diet. This article depicts a correlation between socioeconomic class and health. With an increase in risky behaviors, factors such as stress and inequality, and inability to afford healthy additions to their lives, this article shows that there is generally a decrease in health associated with a lower socioeconomic class.

James, P., Troped P., Hart J., Joshu C., Colditz G., Brownson R., Ewing R., Laden F. (2013). *Urban Sprawl, Physical Activity, and Body Mass Index: HealthStudy and Nurse's Health Study II*, 103(2), 369-375.

In the article, the authors measured the association between county sprawl, defined as the spreading of cities farther from their cores, and health, which was quantified with body mass index and physical activity. The goal of the authors was ultimately to use their results to fight obesity, one of the most common health issues plaguing the us. They concluded that compared to those living in a suburban area, where commodities are more spread out, people in urban cities get more physical activity and typically have a lower BMI.

Chetty, R., Stepner, M., Abraham, S. (2016). *The Association Between Income and Life Expectancy in the United States, 2001-2014.* The Journal of the American Medical Association, 315(16), 1750-1766.

The authors examined the correlation between life expectancy and household income and how it changed between 2001 and 2014. They acknowledge that, for the most part, it is accepted that a higher income correlates to a greater life expectancy, but questioned whether there was a threshold past which income no longer had an impact and wanted to see how the socioeconomic gaps in longevity changed over time. According to the article, the difference in life expectancy due to income grew between 2001 and 2014. This is significant because, assuming the trend continued to present-day, poorer individuals' life expectancy has been steadily decreasing relative to those who are wealthier. It shows that the correlation between life expectancy isn't just present, it's also become more significant over time.

Fitzpatrick, T., Rosella, L.C., Calzavara, A., Petch, J., Pinto, A.D., Manson, H., Goel, V.,

Woodchis, W.P.(2015). *Looking Beyond Income and Education: Socioeconomic Status Gradients Among Future High-Cost Users of Health Care.* American Journal of Preventive Medicine, 49(2), pages 161-171.

The authors of this article aimed to investigate associations between a range of socioeconomic status determinants and future high cost healthcare users (HCU's). Top 5% of healthcare users in the U.S. account for over 50% of costs, and there have not been many studies that attempt to analyze the relationship between SES and those most likely to become HCU's. This article showed that having a low income, less than post-secondary education, and living in highly deprived neighborhoods greatly increased the odds of becoming an HCU, which indicates those on the lower-end of the socioeconomic spectrum are more likely to end up have poor health conditions requiring a lot of spending. This is important because it shows that SES markers like food and housing security may be important factors that could be targeted by policy makers aiming to improve health outcomes and reduce healthcare costs.

Paula, Braveman, Laura Gottlieb, (2014), *The Social Determinants of Health: It's Time to Consider the Causes of the Causes, Public Health Reports. 129(Suppl 2): 19–31.*

In this article, socioeconomic factors such as education, income, and employment are discussed. Then, evidence is used to relate these factors to average mortality rates, life expectancy at 25, etc… all of which were significantly lowered with decreasing levels of income, education, and employment. In the article, "Galea and colleagues conducted a meta-analysis, concluding that the number of U.S. deaths in 2000 attributable to low education, racial segregation, and low social support was comparable with the number of deaths attributable to myocardial infarction, cerebrovascular disease, and lung cancer, respectively." While this number may have changed over the years, the general trend and health of those in a lower socioeconomic status is significantly lowered compared to those in a higher SES.

Jae, Hyun, Kim, Eun-cheol, Park, (2015), Impact of socioeconomic status and subjective social class on overall and health-related quality of life, 13, 783.

In this article, the researchers "...performed a longitudinal analysis using data regarding 8250 individuals drawn from the Korean Longitudinal Study of Aging (KLoSA)." They then analyzed the health related quality of life(QOL) and QOL between individuals in different SES's. The results showed that those in a higher social status and with a higher income were generally better off in HRQOL and QOL, while those in a low social status with a low income had low HRQOL and QOL. They further stated that "Similar trends were seen when education was used as a proxy for socioeconomic status." The general trends observed were that health and quality of life decreased as SES decreased.

Austin, M.., Chen, E., Ross, K., McEwen, L., Maclsaac, J., Kobor, M., Miller, G. (2018). *Early-life socioeconomic disadvantage, not current, predicts accelerated epigenetic aging of monocytes.* Psychoneuroendocrinology, 97, 131-134.

This study examines the way low socioeconomic status affects risk for aging-related chronic diseases. They found that having a low socioeconomic status at a young age (0-5 years old)

mattered much more than one's current socioeconomic status. These people were more likely to face increased cellular aging and thus develop these aging-related chronic diseases, while a low socioeconomic status in late childhood and adulthood seemed to not have any impact. This is important because it shows a different impact of socioeconomic status than most other studies look at; how low socioeconomic status impacts one's children rather than themselves.

Shawky, S. (2018). *Measuring Geographic and Wealth Inequalities in Health Distribution as Tools for Identifying Priority Health Inequalities and the Underprivileged Populations*. Global Advances in Health and Medicine.

The authors of this study looked at how wealth inequality and geographic differences related to inequalities in health. They found that while wealth inequalities lined up with health inequalities in a predictable way (i.e. wealthier individuals tended to be healthier), geographic inequalities actually provided a "more comprehensive profile of health inequalities within a country".