Predictive Analysis of Heart Disease in Georgia: A Health and Socioeconomic Perspective using Multiple Linear Regression

By:
Jun Ming Li
Rajivini Tiruveedhula
Thasin Wahid

BDA 610
Advanced Business Statistics

Performed Under:
Dr. Rui Sun

Mercer University
Stetson Hatcher School of Business

1. **Introduction:**

Coronary heart disease (CHD) remains one of the leading causes of morbidity and mortality in Georgia, accounting for 22,000 deaths in Georgia alone(Heart). Although healthcare and medicine have advanced, heart disease mortality is still predicted to increase by 18% by 2030 (American, 2024), making it crucial to explore the various factors contributing to its prevalence. While genetic predisposition plays a role, health outcomes are often shaped by a combination of lifestyle choices, healthcare access, and socioeconomic conditions. While there is an abundance of studies exploring the link between health and heart disease, there are not sufficient sources to support the relationship between socioeconomic factors and CHD. Socioeconomic factors play an important role in one's lifestyle and access to healthcare. Analyzing CHD trends through the lens of these factors can offer valuable holistic insights into the root causes and potential areas for intervention.

To better understand how these elements interplay, this study employs a multiple linear regression model. This statistical approach allows for the examination of several health and socioeconomic variables simultaneously to determine their influence on coronary heart disease rates. Factors such as obesity, smoking, physical activity, income levels, education, and unemployment will be integrated into the model to measure their collective and individual impact on predicting CHD incidence. By combining the two categories of factors, we aim to achieve a better fitted model in predicting the prevalence of heart disease in individuals.

By focusing on Georgia, where both urban and rural disparities exist, this analysis aims to uncover significant correlations between socioeconomic status and cardiovascular health outcomes. We employed county level data to help us evaluate Georgia through an overall lens and further explore how the differences in urban and rural environments factors into one's general well being. The findings can inform public health strategies aimed at reducing CHD through targeted policies that address not only medical treatment but also the broader social determinants of health. The results might also benefit physicians and hospitals that evaluate their patients in a holistic approach and help mitigate the risks of developing CHD.

2. **Literature Review:**

Sun et. al developed an analysis taking a look at social-economic, financial, and health indicators; the key variables were: age, sex, physical activity, BMI, race, education, smoking, income, gender, percent in college by population, parks per capita, gyms per capita, race % in population, air quality. Linear regression was used to find correlations between the CHD groups and the individual variables. A Geospatial analysis was used to see the prevalence of CHD. Both the sociodemographic and lifestyle factors were important correlates. Age, gender, race, education, income, lifestyle, and environment were especially important factors for prevalence of CHD. Asian and Hispanics have better health outcomes for CHDs than the reference group of Whites. Findings also mirrored similar findings regarding the typical lifestyle and health indicators for CHDs. Of particular note for the contextual level data: Income and education had strong effects to reduce CHD prevalence. County level data selected also followed closely with other analysis into county level prevalence of CHDs. They mentioned that adding in more data on air pollution, mortality data, and insurance data would further improve this research.

In an exploratory analysis by Adepu et. al, main variables of interest are median household income, air quality, lack of health insurance and urban-rural status. The study also looks into descriptive statistics like physical inactivity to analyze the behavioral risk factors associated with the prevalence of CHD. They aim to understand the impact of socioeconomic and behavioral factors that lead to an increased mortality rate caused by CHD, specifically in the state of Georgia. The study looked into 159 counties in the state of Georgia for the period 2014-2016 and assessed the relationship between CHD and socioeconomic determinants of health by performing a bivariate and multivariate regression. In the bivariate regression, CHD and physical inactivity were used as the main variables whereas in the multivariate regression, CHD mortality was used as the dependent variable and the other variables were used as the predictors. Upon analysis, the study found that the Median household income is a widely influenced variable and it affects the variables physical inactivity and access to health insurance. Another important variable that had a significant impact of increased CHD mortality rate was air quality. This study utilizes county level data, which is similar to our study by analyzing factors that lead to CHD. The main focus of this study is on the behavioral factors and air quality indicators which can be useful in our study to better analyze the causes of CHD.

Woodruff et. al conducted a study to analyze the increase in mortality rate caused by CHD to understand the increase during the COVID-19 pandemic and understand the changes in trends among adults aged 35 and above. The study also compares the CHD-related deaths from 2020-2022 in comparison to the trends of 2010-2019. The key variables used in this study are age, race and ethnicity groups and sex for 50 US states and the District of Columbia. The study also focuses mainly on state level data, which is similar to the scope of our study, but over a wider and broader range of data. The study looks into the mortality rate between 2010-2022 for 50 US states and the District of Columbia based on age-adjusted mortality rates. It also examines secondary outcomes which are heart disease and stroke and uses 95% CI by sampling normal distributions along with log-linear regression models to identify trends in time segments across the two time periods During the period 2010-2022, there were 10,951,403 deaths caused due to CHD. It declined by 8.9% from 2010-19 however, there was a significant increase of 9.3% from 2019-22.

In a study done by Dr. Fiscella and Dr. Tancredi, the authors utilize socioeconomic risk factors to predict cardiovascular disease in individuals. The main variables used to predict the risk of CHD is income at a national level in the United States, employment status which measures employed vs. unemployed individuals in the US, and living conditions. It also uses physical risk factors for prediction, such as overweight and obesity, age, alcohol consumption, blood pressure, and cholesterol level. The data was collected on individuals and were measured in each category, such as: income, employment, education, and physical factors like blood pressure. Individual data such as employment and income were collected directly from the individuals. Data collection was derived using CHD risk tool using collected data from all individuals associated with the study and statistics from health databases of the postal codes associated with each individual. In summary, there is a relationship between socioeconomic and physical factors with risk of CHD. Studies suggest that lower income and education are linked to not only high rates of CHD, but also higher rates of CHD mortality. This applies to our study because of the systematic approach to analyzing the relationship between SES and prevalence of heart disease within counties of Georgia. The article is a much broader study which encompasses a higher population but can give us insight of how a general population of the United States was used to conduct the study.

In a systematic review by Pawel Borkowski et. al, the key variables used are well known risk factors associated with CHD, such as smoking, hypertension, diabetes, and obesity. The study also illustrates the impact of socioeconomic status (SES) factors that include education, income levels, and rates of unemployment. These were all collected at a national level but demonstrate multiple populations. The study focuses on the health factors normally associated with heart disease but also researches the impact on socioeconomic status on the risk as well by providing a country level approach to how individuals with different health risks and economic living conditions are able to predict their risk for CHD. Data was also pulled from the Centers for Disease Control and Prevention's Wide-Ranging Online Data for Epidemiologic Research (CDC WONDER) database, which encompassed 17,357,312 subjects, to examine trends in CHD mortality. In summary, the study found that economic status can be used to predict the risk of heart diseases, just as well as normal health factors can. The biggest SES factors used to predict heart disease are education and employment. This applies to our study because of the systematic approach to analyzing the relationship between SES and prevalence of heart disease within counties of Georgia.

3. Data and Data Characteristics:

This project used a custom dataset that contained county level data on Georgia counties in the year 2021. The dataset was created by pooling health and socioeconomic variables from the CDC, County Health Rankings, and FRED. The CDC database was used to find the prevalence of coronary heart disease (CHD), the focus of this study. County Health rankings were used to find many of the socioeconomic and health variables such as diabetes prevalence, median household income, and unemployment rate. Moreover, County Health Rankings also recorded uncommon variables that were of interest to our study: Water Violations, Air pollution, and Insufficient Sleep. Financial and economic variables were cross-checked with FRED to verify validity.

The County Health Rankings database was used as the structural and format foundation to create this project's database. The raw data was first filtered by year, by the county and by alphabetical order. Counties that did not have complete observations were removed, resulting in 1 county being removed from our study. In total, 12 variables and 158 counties were chosen to be used to predict the CHD prevalence among adults in counties in Georgia. **Table 1** shows the summary statistics of the data collected. A more detailed list of the variables and their descriptions are provided in the index.

3.1 Health Variables

In this study, smoking status, physical inactivity, Insufficient sleep, diabetes prevalence, excessive alcohol consumption, obesity prevalence, and air pollution were used as our health variables. Smokers was used as adults who are currently smoking. Physical inactivity was adults who reported having no leisure-time for physical activities. Insufficient sleep was adults who reported fewer than 7 hours of sleep per night on average. Diabetes prevalence was adults who were diagnosed as diabetic. Excessive alcohol consumption was defined as those drinking 5 or more alcoholic beverages a day or 15 a week. Obesity prevalence was adults that reported a BMI of 30 or greater. Lastly, air pollution was a measure of the average daily density of fine particulate matter in micrograms per cubic meter (PM2.5) (County Health Rankings & Roadmaps). Typically, heart health studies focus on most of these variables as their main variables of influence because of their high correlation to heart diseases and other related heart problems (Sun et al. & Borkowski et al.). In our study, these health variables served as only half of the research model and were analyzed alongside socioeconomic variables.

3.2 Socioeconomic variables

In this study, median household income, education, housing problems, water violations, and unemployment rate were our socioeconomic variables. The median household income was the average median household income for all houses within a county. Education was adults aged 25 and over with a high school diploma or equivalent. Housing problems were households with at least 1 of 4 housing problems: overcrowding, high housing costs, lack of kitchen facilities, or lack of plumbing facilities. The unemployment rate was determined by applying the typical unemployment rate formula to the county level population. Lastly, Water Violations was the presence of health-related drinking water violations. Water violation was coded as a binary variable with 0 as the reference level and denoting ‘No’, and 1 for ‘Yes’. ‘Yes’ indicates the presence of a violation, ‘No’ indicates no violation (County Health Rankings & Roadmaps).

	Mean	SD
MedIncome	50067.14	14135.16
Education	82.35	5.9
Unemploy	3.91	0.94
Diabetes	14.61	4.36
Obesity	34.9	6.25
Alcohol	16.28	1.72
Physical	30.37	5.63
Insleep	40.36	2.71
Smoker	21.93	3.14
AirPollution	9.59	0.7
HousingProb	15.6	3
WaterVio	0.2	0.4

Table 1: Summary Statistics of all collected variables

4. Methodology and Results:

Multiple linear regression was chosen as the best avenue to look at the joint effects of health and socioeconomic variables in predicting CHD. **Table 2** below summarizes the adjusted R-squared and the joint P-value for each model. The initial model of just pitting the predictors against the outcome variable had problems with a low adjusted R-squared of .1188. The residual plots were reviewed for Model 1 and outliers and high leverage points were removed from the data set. Model 2 had an increased adjusted R-squared of .1625 but had issues with multicollinearity. Scatter plots and correlation matrices were used to determine which variables had the most correlated effects and a further VIF analysis was used to verify these results. MedIncome, Insleep, Alcohol, and Smoker all had VIF scores over 5. After testing multiple combinations of variables to remove, it was determined that alcohol could be left within the model without causing significant collinearity issues with the other variables. Ultimately, for Model 3, MedIncome, Insleep, and Smoker were removed from the analysis, but the adjusted

R-squared dropped to .1197.

Next, both forward and backward stepwise regression was used to determine the most relevant variables to use to increase the fit of our model to the data. It was determined that water violations, physical inactivity, and diabetes prevalence were not relevant for increasing the fit of our model, thus Model 4 removed them. Lastly, interaction effects were considered based on the studies in our literature review and it was determined that interactions terms for physical inactivity and obesity, and smoker and education increased the fit of our model. The model (5) that was ultimately chosen uses unemployment rate, smoker, education, air pollution, housing problems, obesity rate, physical inactivity, and the pertinent interaction terms to explain CHD prevalence in the counties of Georgia. The adjusted R-squared and joint significance of the models are summarized in **Table 2** and model 5 is summarized in **Table 3.**

Model	Adj. R Squared	Fstat: P-Value
1	0.1188	2.05E-03
2	0.1625	1.79E-04
3	0.1197	1.11E-03
4	0.1344	1.27E-04
5	0.1819	1.52E-05

Table 2: Final Output Model - R^2 and Joint Significance

	Coefficient	P-values
(Intercept)	-10.11	0.35
Unemploy	-0.33	0.0019
Smoker	0.87	0.03
Education	0.18	0.11
AirPollution	-0.11	0.34
HousingProb	-0.02	0.59
Obesity	0.00	0.95
Physical	-0.04	0.65
Smoker:Education	-0.01	0.08
Obesity:Physical	0.0010	0.69

Table 3: Final Output Model - Coefficients and P-value

4.1 The Model

Model 5 had had an adjusted R-squared of .1819 and a joint significance at the .1% level, nearly doubling the ability to account for the variability in CHD prevalence in Georgia counties and providing a greater understanding of the relationship between the predictor variables and our outcome variable when compared to Model 1.

In a comparison of the residual vs. fitted values plot of Model 1 (**Figure 1**) to Model 5 (**Figure 2**), model 5 is a better fit of the data. In model 5's residual plot, the 0-line is more linear as the right-side end of the 0-line is less sharply curved upward. Moreover, in model 5, the heteroscedasticity has improved more on the right side as well.

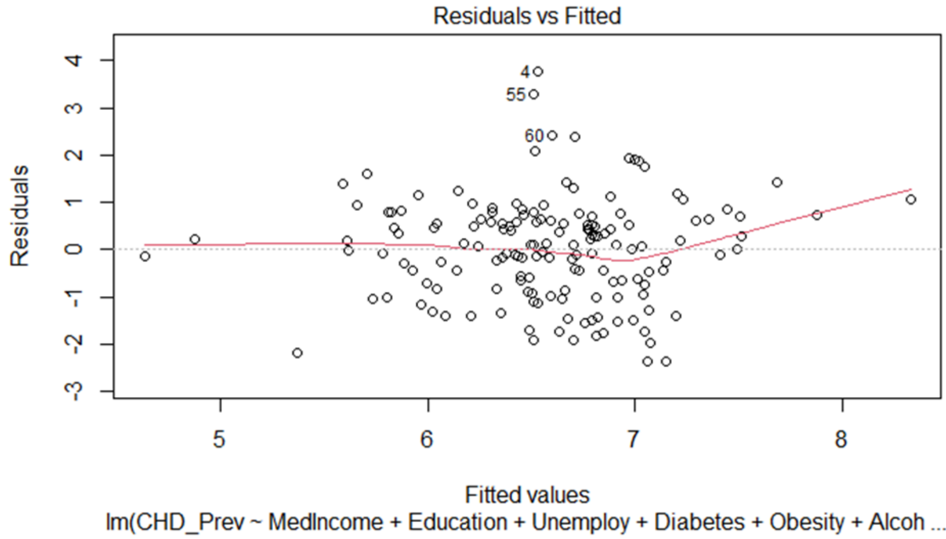

Figure 1: Model 1

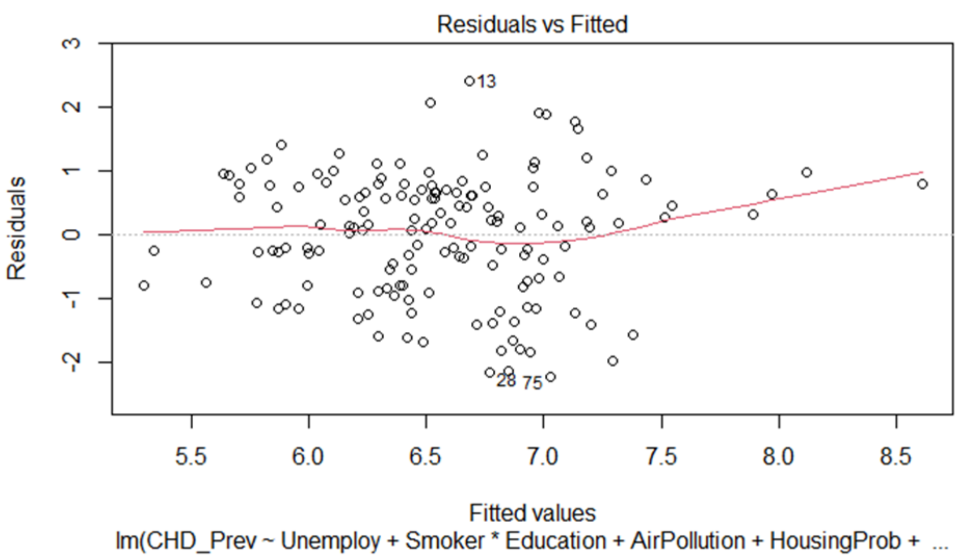

Figure 2: Model 5

5. Discussion:

The results of the model analysis provide a comprehensive understanding of the different variables influencing heart disease prevalence within the 158 counties in Georgia. The overall model, formulated to assess the impact of unemployment, smoking, education, air pollution, housing problems, obesity, physical activity, and their interactions, resulted in an adjusted R-squared value of 0.1819. The adjusted R-squared is a modified version of the R-squared that accounts for the amount of variables in a regression and for predictors that are not significant to the model. In other words, it helps us evaluate how well the model did in predicting the dependent variable of choice, and the higher the adjusted R-squared value, the more accurately the dependent variable can be predicted. However, an adjusted R-squared of 0.1819 indicates that a significant portion of the variability remains unexplained, warranting further investigation into additional variables that may be at play.

Furthermore, we can take a deeper look at **Table 2** to reference the overall accuracy. The significance of the F-statistic ($F = 4.755$, $p = 1.52e-05$) confirms that model 5, as a whole, provides a better fit than a model with no predictors, affirming the relevance of the chosen variables. However, a closer inspection reveals that only a few predictors demonstrate statistically significant relationships with CHD prevalence. After performing the regression and analyzing the output, it can be seen in **Table 3** that many of our variables came out to be insignificant; this was determined by solving for the coefficient p-value for each predictor. Variables, such as education, obesity, and physical activity all reported back with extremely high p-values. This contradicts the current literature as well as general medical knowledge; it is well known that obesity is highly correlated with increased risk of heart disease and would be a great predictor. The contradiction in the regression could be the result of the limitations in the model. By identifying the insignificant variables, we don't simply disregard their influence; instead, we leverage the model to understand how each variable contributes to assessing the risk of heart disease across various factors.

Switching our focus to the significant variables, we find there are 3 variables that provide significant insight and plausible prediction for CHD. Unemployment, Smoker, and Smoker:Education (interaction term) were all found to be statistically significant based on their individual low p-values in the joint test. Keeping all other factors constant, we can see that based on the regression that unemployment has a negative impact on the prevalence of coronary heart disease. This can be interpreted as when unemployment increases within a county, meaning if there is a higher percentage of people within a county that are unemployed, they are less likely to have coronary heart disease. In our regression, for every one percent change of unemployment, the prevalence of CHD decreases by 0.33 approximately. This can be explained by individuals who are unemployed are able to file for unemployment; therefore, they are able to receive money from the government without having to do physical work. Having to do less physical work and receiving money is correlated with less stress which reduces health factors such as high blood pressure, heart attacks, and heart disease. Also due to the fact they are unemployed, they may receive government aid in the form of insurance or medicaid and may have to pay less for certain health screenings procedures which will allow for better prevention compared to those who have commercial insurance but have a high deductible and cannot afford the health care treatments. In this example, our regression can be justified with how more unemployed individuals are less likely to have CHD due to reduced levels of stress (Jeong, 2023).

Our regression also suggests that individuals who smoke are more likely to increase their risk for heart disease. While keeping all other variables constant, the regression results confirm that as we increase the percent of smokers by 1% in a county, the risk for the individuals will increase by 0.87, which seems to be a high influence. Our data matches up with multiple other studies that provide evidence that smoking is a major risk factor of health disease by damaging blood vessels and leading to plaque build up causing blood clots and decreasing blood flow to the heart (Health, 2024).

Our other significant variable was the interaction term between Smoker and Education. The p-value for this was 0.08 and can be rejected at a 0.1 level. Based on the estimated coefficient, the differential effect of smoking on CHD prevalence due to education is -0.01, holding all other variables constant. For every 1% increase in smokers who hold higher education, risk decreases by 0.01. This can be explained because even though smokers have a positive relationship with CHD prevalence, smokers who have a higher education on a percentage level are more likely to experience lower rates of CHD compared to smokers who do not have a high school degree or higher. We can see this to be true in literature too because those who have higher education may smoke but smoke less overall which would give evidence to believe they are less likely to be diagnosed with CHD. Another reason can be explained through individuals who are higher educated might use nicotine and alcohol but in a safer method, such as through nicotine gum or vaping or in lower quantities. There is evidence that while these methods still cause harm, they might be a safer alternative to smoking in traditional methods. Through this, we can accept this regression and understand why those who smoke but are of higher education are less likely to have prevalence of CHD compared to those who smoke but have lower education.

5.1 Socio-economic vs Health-related factors and their respective impact on the Prevalence on CHD

	Adj R Squared	Fstat: P-values
Our Model	0.18	1.52E-05
Health Only	0.13	2.76E-04
Social Only	0.11	5.31E-04

Table 4: Main Model vs Health and Social Isolated Models

In the Health-only factors model, we focused exclusively on variables related to physical health which are diabetes prevalence, obesity prevalence, excessive alcohol consumption, physical inactivity, air pollution and sleep insufficiency. The adjusted R-squared was observed to be 0.13 indicating that approximately 13% of the variance in CHD prevalence can be explained by health factors. The f-statistic p-value for the health-only model was 2.76e-4. While we do get insights into the relationship between health factors and CHD prevalence, the low adjusted R-squared and p-value suggest that this model does not give a full enough picture on heart health.

The Social-only factors model included variables: Median household income, Unemployment rate, Education, Water violation and Housing problems. The R-squared was 0.11 with a f-statistic p-value of 5.31e-4. This suggests that these variables only explain about 11% variance in the CHD prevalence and these factors do not give us a clear idea on their effect on CHD prevalence either.

Based on these findings, we can see that the adjusted R-squared for health factors alone is 0.13 and for socioeconomic factors alone is 0.11; these findings are both lower than that of our main model that combined the two and had an adjusted R-squared of 0.1819. This further emphasizes the necessity to take a holistic approach.

6. **Conclusions and Recommendations:**

To summarize, the three variables: Unemployment, Smoker, and education: smoker (interaction term) were the most statistically significant in our regression in predicting the risk of CHD in counties of Georgia. Furthermore, we discovered that running a model with both health factors and socioeconomic factors provides a better insight than running the two types of factors isolated. While there are significant findings in our research, it is not without limitations. Upon running the final model, we noticed that there were a few weaknesses that needed to be addressed.

The weaknesses are as follows:

1. Low R-squared: One of the major issues encountered in our analysis was the low R-squared values. The R-squared values indicate how well the independent variables explain the variability of the dependent variable, which in this case is CHD prevalence. This suggests that the models do not explain the variance in CHD prevalence in Georgia's counties well enough. Due to this, the model may not be the best fit for analyzing and evaluating the effects of the independent variables on CHD prevalence.
2. Insignificant variables: When running the model, we noticed that there were a few variables that were not significantly related to the outcome variable, CHD prevalence. This suggests that a few of the predictors that were chosen may not have a meaningful impact on analyzing their effects on CHD prevalence. This raises the question whether different or more specific variables can be chosen which might yield better results and capture a better relationship.
3. Limited Scope: Another weakness that contributed largely to why the model does not predict CHD as well as we had hoped is that the scope was limited to only the counties of Georgia for 1 year, 2021, meaning our model was hyper specific to time and region. The narrow scope of the study also makes it difficult to draw broader conclusions about CHD as a whole.

Improvements:

Given the weaknesses of our study, several improvements need to be made to the existing model to improve the robustness of the findings. One of our main goals would be to expand the data set which will likely enhance the significance of the model and provide greater insights into understanding the factors that lead to CHD prevalence.

The recommended improvements are discussed below:

1. Include more states: Incorporating county level data for more states would help generalize the findings and allow the model to reflect a wider range of demographic data. This will help better assess the predictors variables and prevalence of CHD across various contexts over a wide range of observations.
2. Include a wider range of time: By extending the study and including multiple years of data, we can better understand the long-term trends in CHD prevalence and its predictors. This longitudinal approach would help track changes over time and trends.
3. Include more variables: Another key recommendation is to include different variables and a more diverse set of predictors that are less collinear with each other and help improve the overall significance of the model. Including variables like Body Mass index(BMI), age, race, height, weight and other lifestyle factors will allow for a more comprehensive analysis of different categories that contribute to CHD prevalence.

Appendix:

Variables

- CHD_Prev: Coronary Heart Disease prevalence as a percentage of county population that has been diagnosed with coronary heart disease.
- MedianIncome: Median Household Income in a county.
- Umemploy: Unemployment Rate as a percentage of county population.
- Smoker: Adults who are current smokers as a percentage of county population.
- Education: Adults aged 25 and over with a high school diploma or equivalent as a percentage of county population.
- Diabetes: Diabetic individuals as a percentage of county population.
- Alcohol: Adults within a county that drink excessively as defined by the health data website as drinking 5 or more alcoholic beverages a day or 15 a week, as a percentage of county population.
- Obesity: Adults that report a BMI of 30 or greater as a percentage of county population.
- AirPollution: Average daily density of fine particulate matter in micrograms per cubic meter (PM2.5).
- Physical: Adults age 18 and over reporting no leisure-time physical activity, as a percentage of county population.
- Insleep: Adults who report fewer than 7 hours of sleep on average, as a percentage of county population.
- WaterVio: Indicator of the presence of health-related drinking water violations. ‘Yes’ indicates the presence of a violation, ‘No’ indicates no violation.
- HousingProb: Households with at least 1 of 4 housing problems: overcrowding, high housing costs, lack of kitchen facilities, or lack of plumbing facilities as a percentage of county population.

References:

American Heart Association (2024) Forecasting the Burden of Cardiovascular Disease and Stroke in the United States Through 2050—Prevalence of Risk Factors and Disease: A Presidential Advisory From the American Heart Association 2024 Circulation e65-e88 150 4 doi:10.1161/CIR.0000000000001256 *https://www.ahajournals.org/doi/abs/10.1161/CIR.0000000000001256*

Borkowski, P., Borkowska, N., Mangeshkar, S., Adal, B. H., & Singh, N. (2024). Racial and Socioeconomic Determinants of Cardiovascular Health: *A Comprehensive Review. Cureus, 16(5), e59497. https://doi.org/10.7759/cureus.59497*

Centers for Disease Control and Prevention. (2019, June 13). Socioeconomic differences in cigarette smoking among sociodemographic groups. *Centers for Disease Control and Prevention. https://www.cdc.gov/pcd/issues/2019/18_0553.htm*

Fiscella, K., & Tancredi, D. (2008). Socioeconomic status and coronary heart disease risk prediction. *JAMA, 300(22), 2666–2668. https://doi.org/10.1001/jama.2008.792*

Grafova, I. B., & Stafford, F. P. (2009). The Wage Effects of Personal Smoking History. Industrial & labor relations review, 62(3), 381. https://doi.org/10.1177/001979390906200307

Health effects of cigarettes: cardiovascular disease. (2024, September 17). Smoking and Tobacco Use. *https://www.cdc.gov/tobacco/about/cigarettes-and-cardiovascular-disease.html#:~:text=Smoking%20is%20a%20major%20cause,from%20cardiovascular%20disease%20and%20death*

Heart disease. Georgia Department of Public Health. (n.d.). https://dph.georgia.gov/chronic-disease-prevention/heart-disease

Jeong, S., & Fox, A. M. (2023). Enhanced unemployment benefits, mental health, and substance use among low-income households during the COVID-19 pandemic. Social science & medicine (1982), 328, 115973. https://doi.org/10.1016/j.socscimed.2023.115973

Search. (n.d.). County Health Rankings & Roadmaps. https://www.countyhealthrankings.org/search?keywords=georgia&f[0]=type:states&f[1]=type:counties

Sun, W., Gong, F., & Xu, J. (2019, July 25). *Individual and contextual correlates of cardiovascular diseases among adults in the United States: A geospatial and Multilevel Analysis - Geojournal*. SpringerLink. https://link.springer.com/article/10.1007/s10708-019-10049-7

Sanjana Adepu, Adam E. Berman, Mark A. Thompson, (2020, September) Socioeconomic determinants of health and county-level variation in cardiovascular disease mortality: An exploratory analysis of Georgia during 2014–2016, *Preventive Medicine Reports, https://www.sciencedirect.com/science/article/pii/S2211335520301200*

Woodruff, R. C., Tong, X., Khan, S. S., Shah, N. S., Jackson, S. L., Loustalot, F., & Vaughan, A. S. (2023). Trends in Cardiovascular disease Mortality Rates and excess Deaths, 2010–2022. American Journal of Preventive Medicine, 66(4), 582–589. https://doi.org/10.1016/j.amepre.2023.11.009