Distribution Agreement

D’Anndria LaSean Kendrick April 20, 2023

The Relationship Between Energy Burden and Cardiovascular Disease Mortality in Georgia between 2018 and 2020

By

D’Anndria LaSean Kendrick
MSPH
2023

Environmental Health - Epidemiology

Stefanie Ebelt, Sc.D.
Committee Chair

The Relationship Between Energy Burden and Cardiovascular Disease Mortality in Georgia between 2018 and 2020

By

D’Anndria LaSean Kendrick

BS
University of Alabama at Birmingham
2018

Thesis Committee Chair: Stefanie Ebelt, Sc.D.

An abstract of
a thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Gangarosa Department of Environmental Health
2023

Abstract

The Relationship Between Energy Burden and Cardiovascular Disease Mortality in Georgia between 2018 and 2020

By D’Anndria LaSean Kendrick

Background: As progressively extreme temperatures become more common in our society, the use of energy to counteract the weather with either heating or cooling appliances increases, leading to an increase in energy costs annually (EPA, 2021). Energy poverty, also known as energy burden (EB), is a measure of a household’s overall energy costs, that mostly affects low-income individuals. It is estimated that the national estimate for energy burden is about 9% for low-income families, while non-low-income families average about 3%, suggesting that low-income families pay more for energy annually than middle and high-income families. With fewer financial opportunities for nutrition and medication, the energy burden leaves families more susceptible to developing chronic diseases over time. The goal of this thesis was to investigate the relationship between energy burden and cardiovascular disease mortality in the state of Georgia during 2018-2020. **Methods**: Current literature mentions covariates related to energy burden that are also associated with cardiovascular disease. These literatures were used to create Direct Acyclic Graphs to conceptualize the ideas and determine important variables on which to collect data and include in the analysis. County-level data on energy burden, cardiovascular disease mortality, and covariates were collected from the Department of Energy and the Centers for Disease Control and Prevention. Linear regression analyses were conducted to estimate the association of energy burden and cardiovascular disease mortality, adjusting for potential confounders. Analyses were also stratified by gender. **Results**: Overall, no statistically significant associations were observed between energy burden and cardiovascular disease mortality. There was indication of a larger point estimate for the association among women compared to men. **Conclusion:** Although the results of the analysis were null, the possibility of a relationship between energy burden and cardiovascular disease mortality among women is interesting. Research on this association should be further explored with more integrated variables to confirm its existence and establish more substantial, statistically significant results.

The Relationship Between Energy Burden and Cardiovascular Disease Mortality in Georgia between 2018 and 2020

By

D’Anndria LaSean Kendrick

BS
University of Alabama at Birmingham
2018

Thesis Committee Chair: Stefanie Ebelt, Sc.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Gangarosa Department of Environmental Health
2023

Acknowledgments

I would like to thank my incredible mentor, Dr. Stefanie Ebelt, for her guidance, insight, patience, and expertise in the development of this project. I would also like to thank Dr. Michael Caudle for his continued support in the beginning stages of this project. I would like to acknowledge my friends who helped influence the completion of this project and finally, a big thank you to my mother who never stopped believing in me and my future.

Table of Contents

INTRODUCTION

Background and Introducing the Problem:

Human activity, such as tropical deforestation and burning fossil fuels in industrial environments, has undoubtedly contributed to an influx of greenhouse gases (Jevtic, 2021), which causes extreme climate consequences like global warming (IPCC, 2019; Wolff, 2020). Unfortunately, because the results of climate change are not discovered immediately, many do not understand its importance. The use of energy to stay cool in scorching hot and extreme cold weather has increased alongside increasing temperatures, therefore, leading to an increase in energy-related costs for residential housing (EPA, 2021). Reports of energy usage in residential buildings and space conditioning, such as heating and cooling, across the United States account for 22% and 41% of primary energy consumption, respectively (Bradshaw, 2014).

Several related data sources have indicated that affordable household energy is an integral segment for preserving general wellness (Reames, 2021). However, unfortunately, energy poverty, also known as energy burden (EB) is a challenge that endangers a household's overall health and interferes with its ability to sufficiently maintain affordable energy services (Reames, 2021).

Families globally fight hard against energy burdens daily, especially those living in warmer, humid climates. Although this is a huge problem globally that can lead to several hundred deaths per year (Shindell, 2020) locations like the southern parts of the United States have a similar issue with slightly less lethal consequences (when compared to countries with more extreme weather conditions). Vulnerable populations in the United States, for years, have been the target communities for environmental injustice. When considering Environmental (in)Justice, the most common topics include redlining, food insecurity, Superfund sites, contaminated drinking water, and industrial plants polluting the air. However, many never consider the idea of energy burden or energy poverty.

Energy burden can be described as a measure of energy poverty that displays the percentage of monthly household income spent on utility expenditures including power, electricity, gas, heat, air conditioning, and water costs for the home. Studies illustrate an amalgamation of negative effects regarding both the physical and mental health of populations at risk that contribute to several social inequalities (Chen, 2022; Hernandez, 2015).

Residents paying large chunks of their monthly income on energy bills are energy-burdened (Brown et al., 2020). The United States Energy Information Administration (EIA) estimates that one in three American households experience a form of energy insecurity and meet necessary energy needs (EIA, 2018; Reames, 2021). It can be inferred that energy burden is an extension of energy burden as disproportionate allocations of energy burden, whether it is beneficial or harmful, is apparent in systems similar to class, education, and race/ethnicity (Reames, 2021).

Over the years, Black Americans, especially those living in lower-income communities, have been the top populations that experience an energy burden. Researchers have also established that low-income, Black American/African American, Hispanic/Latino, and renters consume less electricity than their counterparts, yet have higher energy use when normalized by housing quality and efficiency (Reames, 2016; Chen, 2022). Further, low-income, American households spend up to three times more on energy costs than higher-income households (Drehobl, 2020; Chen, 2022).

However, because of climate change, lower-income White Americans experiencing an energy burden have increased over twenty-five years (Wang et al., 2020). There are two reasons for this according to Mastropietro; one is the diminishing marginal utility of electricity supply and the other is simply that more income leads to better sustainability and more insulated appliances and households (Mastrpietro et al., 2019). Energy is extremely important for survival, especially if one resides in locations with regular occurrences of extreme weather conditions. High heat can influence the development of multiple chronic illnesses that can easily result in higher mortality rates (Sanchez-Guevara, 2019). Locations like the southeastern states, typically experience hotter summers, leading to 38% of low-income households in these states suffering an energy burden of 6% or higher, in comparison to the 29% of low-income homes in other parts of the United States (Chen, 2022).

Households who struggle with whether to utilize their monthly income on energy bills to stay cool in summer or have a well-balanced diet or medicine to treat pre-existing conditions are the most at risk for developing chronic disease, heart disease, and respiratory disease (Reames, 2021).

Efforts to Address and Improve the Energy Burden Issue:

There have been some policies and programs created to help alleviate high energy bills, but unfortunately, lower-income individuals are still spending way too much. There are many communities in Memphis, Tennessee facing up to a 27% percent energy burden (SELC, 2022), with very little assistance from local energy efficiency programs. According to America’s Council for an Energy-Efficient Economy (ACEEE,2020), the most efficient process to address energy insecurity involve designing appropriate weatherization programs to accommodate at-risk communities. Researchers conclude that Weatherization treatments and processes can improve homes' energy efficiency, which results in reduced energy bills, decreased carbon emissions, better indoor air quality, and even job opportunities (Bradshaw, 2014).

The ACEEE organization first suggests creating an energy burden goal for each city and creating strategies to improve energy insecurity and achieve the goal. For example, the organization's first goal for the city of Atlanta is to solidify a goal of less than ten percent energy burden, then continue to track progress annually. The second suggestion involves increasing funding for low-income weatherization, which can be explained as the local government's allocation of funds for energy-efficient projects, such as addressing energy burden, to assist specific households with energy costs. Finally, it is suggested by the ACEEE organization to address the overall issue of environmental injustice. High energy burdens, food deserts, poor health, unfit living conditions, and inadequate housing are linked and require similar resources.

The United States Department of Energy (DOE) Weatherization Assistance Program (WAP) was created in 1976 and has improved the lives of 7 million low-income households. WAP is a weatherization program dedicated to helping reduce energy insecurity in low-income households. They provide services such as roof repair, insulation installation, and pipe repair to eligible households. The WAP program supports weatherization services to about 35,000 households every year from DOE funds. This program also allows households to save an average of $372 or more annually. Families can apply for the program yearly and must meet certain requirements, but the DOE WAP program provides many inexpensive or free solutions to reduce energy costs through the completion of an energy audit, which explains four main measures that can address the energy burden. These measures include "Mechanical"," Health and Safety"," Building Shell", and "Electric Baseload". Examples for each measure include "installing duct and heating pipe insulation”, “evaluating mold/moisture hazards”, “installing wall, floor, ceiling, attic, and/or foundation insulation”, and “installing efficient light sources”, respectively.

These programs are great theoretically, however, the idea of energy burden is still limited knowledge among many households, leading to unknowing families suffering in silence and compromising their health simultaneously. By identifying potential risks regarding cardiovascular disease, the projected impact of this project will be to lessen the ideal aspect of one’s health and include environmental factors, spread the ideas of energy burden to the public and display the results, and further push for more programs to assist with energy burden related issues in other locations for better overall public well-being.

Purpose of Project:

Energy burden is a concept that many may not realize affects their households. It is defined as spending around 10% of monthly income on energy expenses. According to the Department of Energy’s Low-Income Energy Affordability Data tool, the national average energy burden for low-income homes is 8.6%. For middle and high-income households, the average is estimated to be about 3%. Because low-income homes spend more on energy costs, there are fewer financial opportunities for medication, nutrition, and other costs associated with everyday living, thus leading to a decline in general well-being.

The purpose of this project is to identify the relationship between chronic health outcomes, such as cardiovascular disease in any form, and high energy burden households within the years 2018 and 2020 for the state of Georgia by county. The project's main idea is to analyze the association between cardiovascular disease and energy burden and address how impactful and influential energy burden is on an individual’s general well-being. Current literature focuses on the impact of energy burden regarding Race/Ethnicity as an indicator associated with the likelihood of exposure to energy poverty and European locations with extreme weather fluctuations and higher heat-related and cold-related deaths (Thomson, 2019; Shindell, 2020). Scholars and other researchers highlighted the idea of an intersectional approach regarding energy poverty as households could include multiple indicators, such as race and gender, that place them at a higher risk of health issues and a deeper burden (Graff, 2021).

This project will focus on the gender gap in current literature and focus more on the energy burden impact in the United States.

Specifically, the project was designed to address three primary aims:

- Aim # 1: Create a Directed Acyclic Graph (DAG) to conceptualize how demographics, socioeconomic status, and environmental stressors are associated with and related to health disparities due to energy burden in vulnerable communities.
- Aim #2: Collect data on energy burden, cardiovascular disease mortality rates, and potential confounders for the state of Georgia during 2018 - 2020.
- Aim #3: Estimate the association of county-level energy burden and cardiovascular disease mortality rates, overall and by gender.

METHODS

***Figure 1.0 Directed Acyclic Graph:* Energy Burden Exposure Leading to the Development of Cardiovascular Disease**

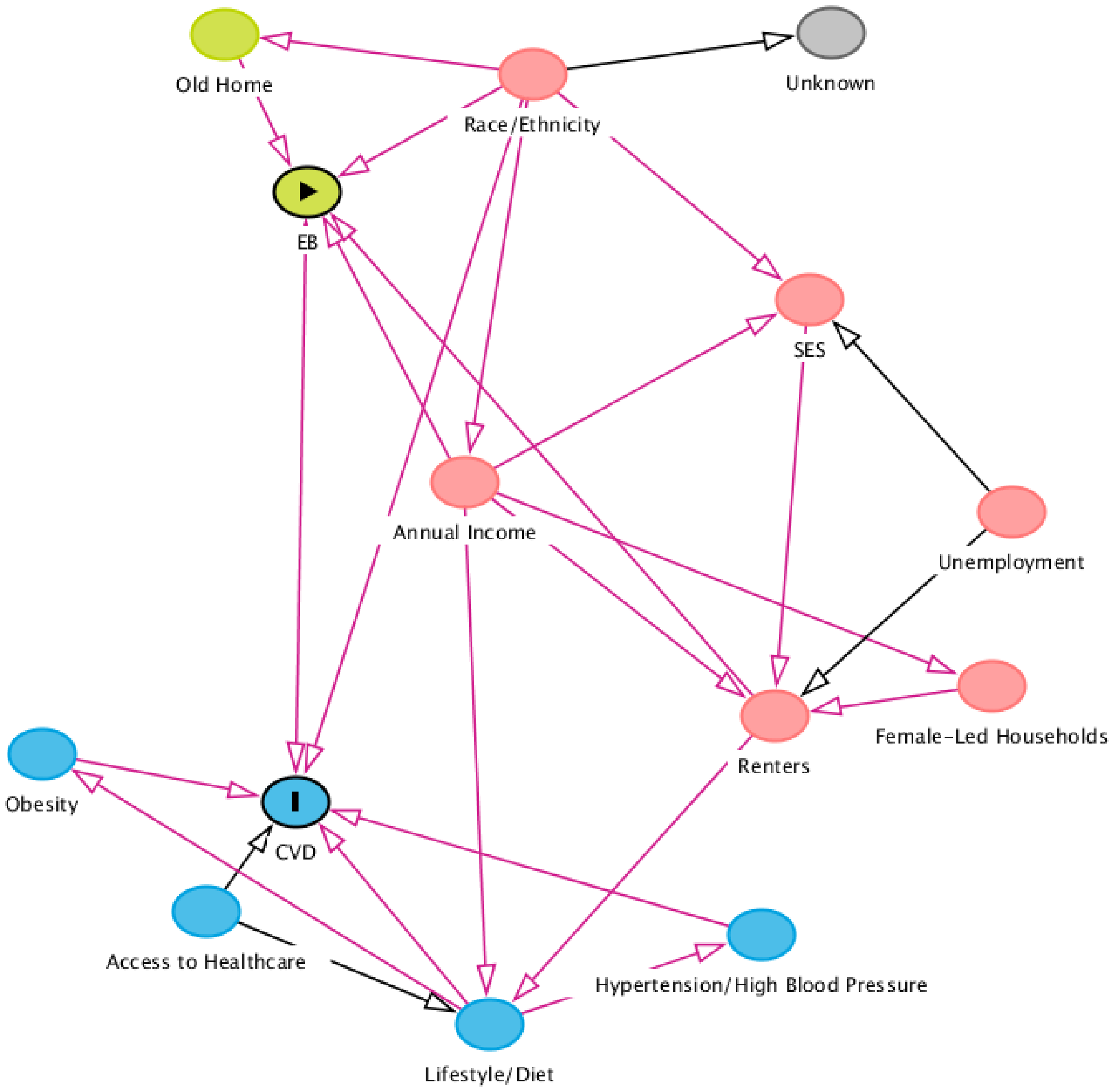

Figure 1

Figure 1 was created using DAGitty, version 3.0.

EB (Energy Burden) is the exposure

CVD (Cardiovascular Disease) is the outcome of interest

All others are covariates

Blue is an ancestor of the outcome (CVD)

Green is an ancestor of the exposure (EB)

Red is the ancestor of exposure and the outcome

Red path is a biasing path

Figure 1 is a conceptual, visual representation of the association between energy burden and cardiovascular disease by identifying and linking causal pathways. The order in which the activities are shown in this graph represents the relationship to each variable and the link to the influence and development of cardiovascular disease. This DAG was created to illustrate known sources of bias and confounders to include in the analyses and identifies potential variables to include in the linear regression analysis model. This is the original DAG created for this analysis.

Figure 1.1 Directed Acyclic Graph: **Energy Burden and Cardiovascular Disease Development Associations with Environmental Indicators**

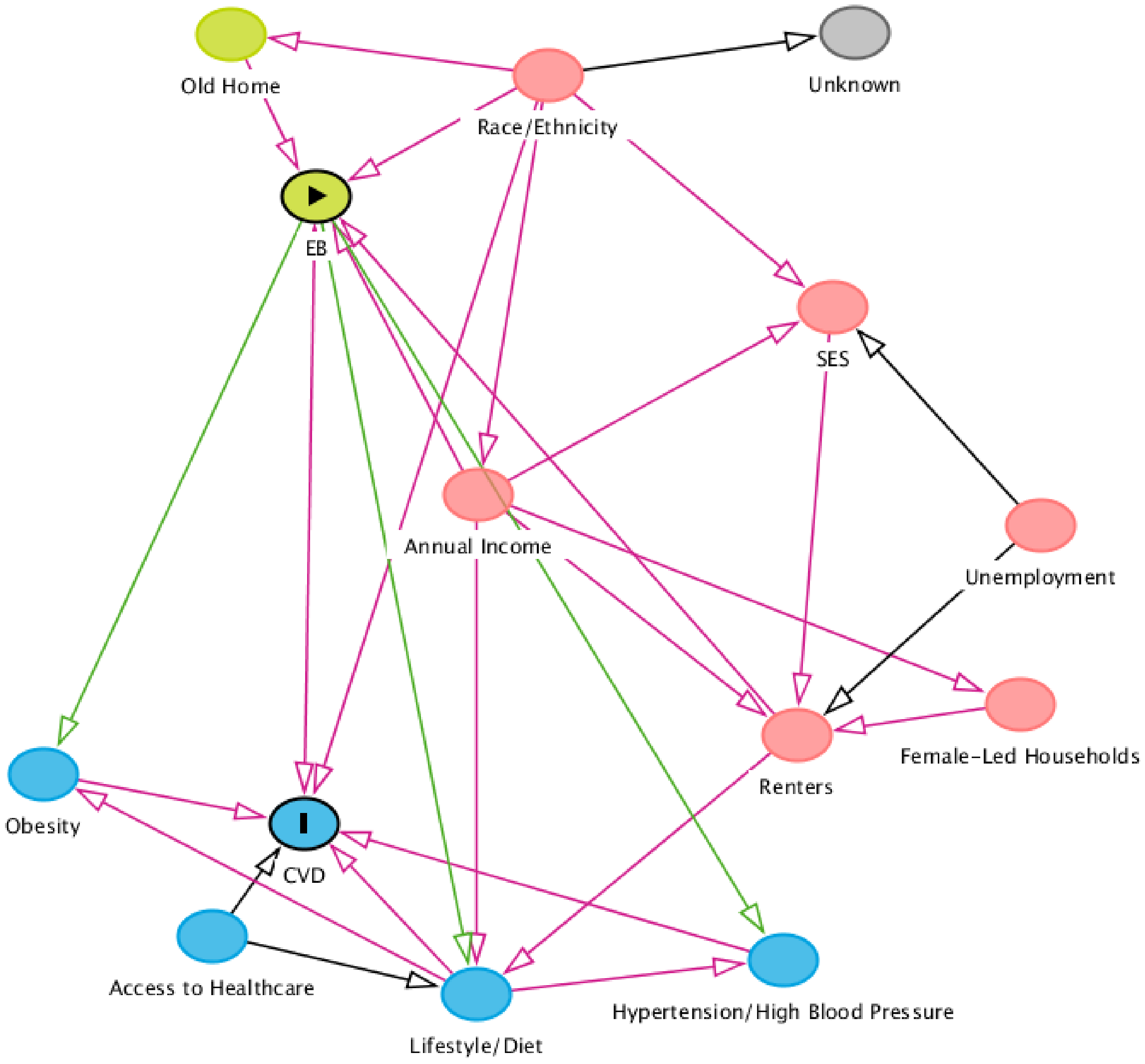

Figure 1.1

Figure 1.1 was created using DAGitty, version 3.0.

Where

EB (Energy Burden) is the exposure

CVD (Cardiovascular Disease) is the outcome of interest

All others are covariates

Blue is an ancestor of the outcome (CVD)

Green is an ancestor of the exposure(EB)

Red is the ancestor of exposure and the outcome

Red path is a biasing path

Green path is a casual path

Figure 1.1

Figure 1.1 is also a conceptual, visual representation of the association between energy burden and cardiovascular disease by identifying and linking causal pathways, similar to *Figure 1.0.* However, *Figure 1.1* connects the exposure, Energy Burden, to other variables that it influences (Lifestyle/Diet, Hypertension/High Blood Pressure, Obesity) that may not necessarily lead to the development of cardiovascular disease. Connecting these variables creates a casual path from energy burden to the listed variables. This is the final DAG created for this analysis.

Data Collection

The data for this project was obtained from the United States Census Tract, the Low-Income Energy Affordability Data (LEAD) Tool created by the United States Department of Energy and the Center for Disease Control and Prevention's Division for Heart Disease and Stroke Prevention Interactive Atlas of Heart Disease and Stroke. The housing data used in the LEAD tool originates from the United States Census Bureau's American Community Survey 2018 Public Use Microdata Samples.

The energy burden measures recorded by the United States Department of Energy LEAD tool display data in the form of interactive graphs and maps to illustrate and communicate characteristics regarding energy and housing data for low-income households by state and county. The average energy burden variable is calculated for electricity, natural, gas, and alternative fuel expenditures in the interactive LEAD tool. The LEAD tool also calculates the average percentage of income spent on energy bills for low-income households. The US DOE defines low-income homes as a household earning between 0 and 80% of the Area Median Income (Reames, 2021).

The cardiovascular disease-related data were obtained from the Center for Disease Control and Prevention's Division for Heart Disease and Stroke Prevention Interactive Atlas of Heart Disease and Stroke. The Interactive Atlas of Heart Disease and Stroke is an interactive mapping tool that allows for customized viewing of heart disease and stroke-related data that can be stratified by gender/sex, race/ethnicity, and age group. This tool also accounts for state-level and county-level data for a more specific data source. The Interactive Atlas of Heart Disease and Stroke can be used to create maps that display geographical disparities that relate to and influence the development of heart disease and stroke. For this project, the data was collected from the state of Georgia and was filtered to include “Mortality, Hospitalization (state, county)”, “Total Cardiovascular Disease”, “Date Rate”, “2018-2020”,” All Genders”,” All Races/Ethnicities”, “All Ages”, and “Smoothed”. Additional data included for the customized report for this project were are listed in *Table 1.* It was then stratified by gender, men, and women, by changing “All Genders” to “Men” and then to “Women”, for a total of three analyses.

Additional Covariates of Consideration:

Multiple covariates were considered as confounders for this project's final analyses, including Old Homes, Race/Ethnicity, Renters, and Female-led Households. However, covariates from the Interactive Atlas of Heart Disease and Stroke tool were used to provide data. Old Homes report a statistical association between households of color and energy insecurity (Graff, 2021). Older homes, built with lesser quality materials, lack insulation, and have inefficient home appliances can increase the likelihood of higher monthly energy-related costs (Hernandez, 2015). Energy efficiency upgrades can potentially reduce energy expenses and improve thermal comfort, while also addressing problematic issues in the home environment (Hernandez, 2015; DOE WAP, 2016).

Female-Led Households were another variable of consideration in the model analysis. According to research and findings, the results stated that counties with a higher percentage of families with female heads saw their energy burden increase in 2020, after controlling for COVID-19-related mortality records (Chen, 2022). These results included the impact of Covid-19, however, there is an opportunity for further research data before and after the effects of Covid-19.

Race and Ethnicity have been a variable included in many studies and research regarding EB. However, with continued research, many records have noted that White Americans are progressively becoming a part of the demographics that are affected as locations continue to grow(Wang, 2020). As the years pass and survival necessities become more expensive, race/ethnicity seems to become less involved in the concept of EB (Wang, 2020). With this in mind, Race/Ethnicity was dropped as a variable to study other indicators that are less well-known.

EB has been associated with both racial discrimination and inequality within housing policies (Chen, 2022). The risk of energy burden increases as high rates of residential discriminatory housing policies continue to saturate the housing market for Black Americans, Hispanic Americans, and Native Americans/Indigenous groups. This project recognizes that race and ethnicity and other physical demographic indicators can result in a higher risk of exposure to energy burden, however, because of Georgia's diverse population and unique housing infrastructures, “race and ethnicity” were eliminated from the final three models. For other locations with high energy burden rates, such as Alabama, race and ethnicity would be included in further studies.

Finally, the Female-Led Households variable was also considered for this analysis. This variable was considered as it considers the indicators that impact the gender wage gap, sexism in the workplace, and violence against women that may lead to temporary or transitional housing situations. There is very limited data on this topic, so this project stratified the cardiovascular mortality data by sex to address the hypothesis. With these things included, it is hypothesized that women will have a higher risk of cardiovascular disease due to exposure to energy burden than men.

Specifying the Data Sources:

The United States Low-Income Energy Affordability Data tool is an interactive tool able to break down the average energy burden percentage, the average annual energy cost, and housing costs, for each state in the United States. The data provided through the lead tool is specific data captured in 2018. Each data capture can be split into three microdata samples and can be specified using certain income models.

The first model is the area median income, which is described as the midpoint of a region's income distribution. This is described as taking half of the families in a region, earning more than the median, and half earning less than the median.

The second income model is the federal poverty level, which is described as a measure of income used by the government to determine who is eligible for programs and benefits. The third and final model is called the state median income, which is described as the midpoint of regions, and income distribution again with half of the families in the reading, earning more than the median, and earning less than the median income used by the government to determine who is eligible for programs and benefits.

The third and final model is called the state median income, which is described as the midpoint of a region, income distribution again with our families in the reading, earning more than the median, and earning less than the median.

Because the LEAD tool provided multiple energy burden percentages for each county, The Energy Burden dataset was manipulated in R studio to calculate the average EB for each of the 159 counties in Georgia, resulting in only one value per county.

The second data set originated from the Center for Disease Control and Prevention Division for Heart Disease and Stroke Prevention Interactive Atlas of Heart Disease and Stroke.

Here, I was able to collect data for the years between 2018-2020 and analyze the total cardiovascular disease-related deaths per 100,000 by men, women, and a combination of both. This specification for data also accounted for all ages and all races/ethnicities. Further along, other environmental indicators are allowed to be added to the main dataset, resulting in a custom dataset. Indicators are separated into six main groups (Prevalence, Risk Factors, Social, Economic, Environmental Data, Demographics, Health Care Delivery and Insurance, and Health Care Costs which each have several subgroups for a more detailed report. Variables included in this analysis are listed in Table 1, except ENEBUR and CVDDEATH. These steps were repeated to find the Cardiovascular Disease mortality rate for men and women, for a total of three separate datasets each merged with the manipulated EB dataset. Once the dataset was finalized and variables were renamed for more clarity, the collinearity assessment began to determine potential variables for the linear model.

***Table 1**: Description of Variables Used in EB/CVD Analysis*

Variable Description Table		
CVDDEATH	Total Cardiovascular Mortality	Total Cardiovascular Disease Death Rate per 100,000, All Races/Ethnicities, All Genders, All Ages, between 2018-2020
ENEBUR	Average Energy Burden Percentage (%)	Average annual housing energy costs divided by the average annual household income. Housing energy costs are based on household monthly expenditures for electricity, gas, and other fuels
CORHEART	Coronary Heart Disease (%)	Coronary Heart Disease Among Adults Ages 18+, 2020
HIBP	High Blood Pressure (%)	High Blood Pressure Among Adults Ages 18+, 2019
STROKE	Stroke (%)	Stroke Among Adults Ages 18+, 2020
HICHOL	High Cholesterol (%)	High Cholesterol Among Adults Screened in Past 5 Years Ages 18+, 2019
DIABET	Diabetes (%)	Diagnosed Diabetes, Age-Adjusted Percentage, 20+, 2019
OBESE	Obesity (%)	Obesity, Age-Adjusted Percentage, 20+, 2019
LIFESTYLE	Physical Inactivity (%)	Leisure-Time Physical Inactivity,

		Age Adjusted Percentage, 20+, 2019
SMOKE	Current Smoking Status (%)	Current Smoker Status Among Adults Ages 18+, 2020
HISDIP	No high school diploma (less than high school education)(%)	Percentage without High School Diploma, Ages 25+, 2016-2020 (5-year)
CODIP	No college diploma (less than college education)(%)	Percentage without 4+ Years College, Ages 25+, 2016-2020 (5-year)
INCOME	Median Household income ($)	Median Household Income, 2020
PCTPOV	Percent Poverty (%)	Percentage Living in Poverty, All Ages, 2020
UNEMPLOY	Unemployment Rate	Unemployment Rate, Ages 16+, 2021
CHOLSCRE	Cholesterol Screenin (%)	Cholesterol Screening Among Adults Ages 18+, 2019
HEALTHI	Health Insurance Status (%)	Percentage without Health Insurance, Under Age 65, 2019
HEARTDI	Heart Disease Prevalence (%)	Prevalence of Diagnosed Heart Disease Among Medicare Beneficiaries, 2020
HOME	Home Value ($)	Median Home Value, 2016-2020 (5-year)

Epidemiological Analysis:

All analyses were conducted in *RStudio*, version 1.3.1093

A crude model was initially developed from the merged data. The form of the model was as follows:

Y(death due to CVD) = α + β1(ENEBUR)

Where ‘death due to CVD’ represents the county-level CVD mortality rate and ENEBUR represents the county-level energy burden.

Income and Socioeconomic Status is impactful when analyzing EB. Instead of including both “income” and “Socioeconomic Status”, I felt that only one covariate should represent the idea of “Socioeconomic status” rather than including multiple. “Income” was tested along with the crude association to determine influence and to prove this theory. The model is shown below:

(Death due to CVD) = α + β1(ENEBUR) + γ1(INCOME)

The results concluded that energy burden and income were very similar, leading that the income variable may skew the results of the study. The variable “income” may be already included in the data for “energy burden”, which is why it was eliminated.

Correlation Matrices:

A correlation matrix was used to determine correlation coefficients for variables with potential collinearity. Several variables chosen for the model may fall under the same category, therefore leading to a repeated pattern in the analysis. For this project, there is potential collinearity within the Socioeconomic status variables (i.e., income, unemployment, educational attainment) and the general health outcome variables (i.e., high blood pressure, diabetes, obesity). For this project, there are two correlation matrices performed to help exclude redundant representations of demographic indicators. These matrices are displayed in *Figure 2* and *Figure 3*.

Figure 2: *Determining the Correlation Among SES Variables*

SES Collinearity Assessment (Correlation Matrix)

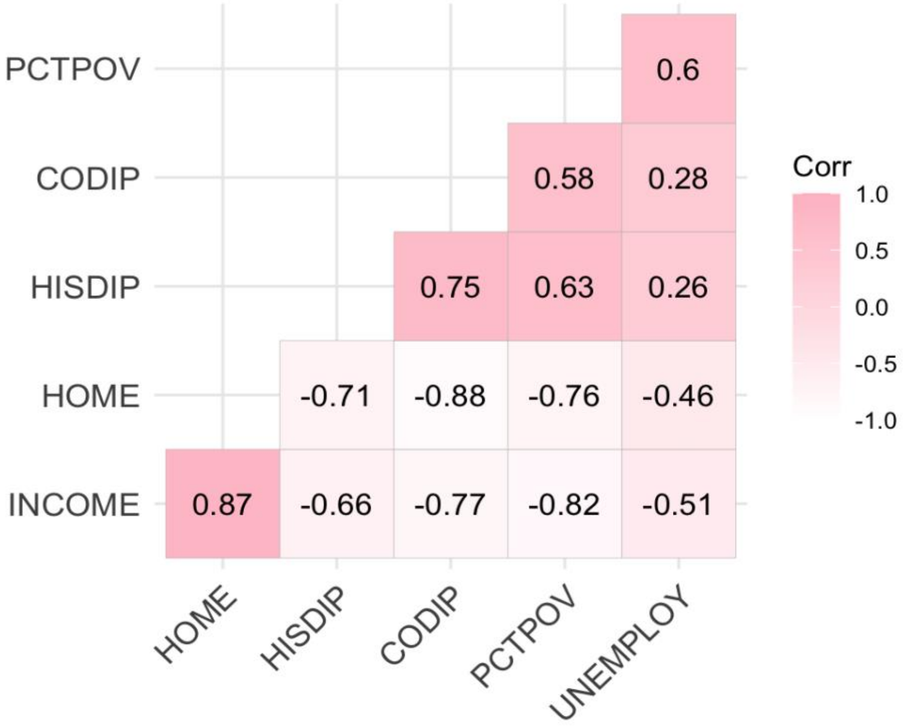

Figure 2

All variables with potential collinearity related to socioeconomic status were chosen and placed in a correlation matrix to determine collinearity and simplify the final model. The figure illustrates a high correlation between HOME and INCOME(0.87), HISDIP and CODIP(0.75), HISDIP and PCTPOV(0.63), CODIP and PCTPOV(0.58), and PCTPOV and UNEMPLOY(0.6). Because of the higher correlations, some variables were dropped from the model.

Figure 3: *Determining the Correlation Among General Health Variables*

Health Conditions Collinearity Assessment (Correlation Matrix)

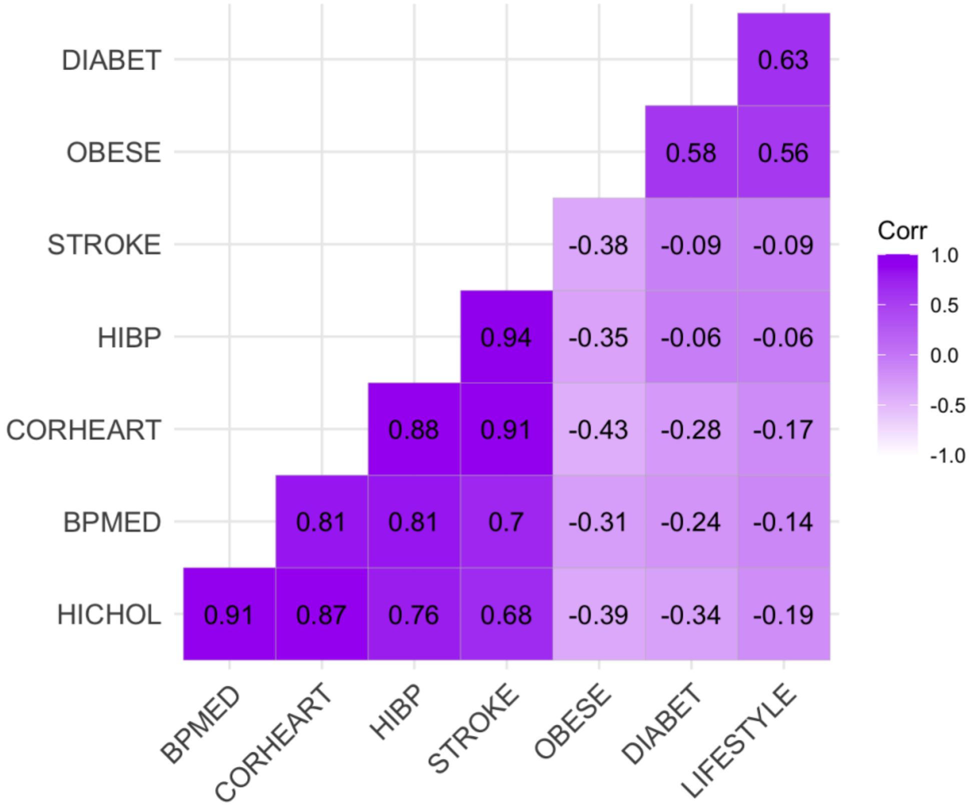

Figure 3

All variables with potential collinearity related to general health status were chosen and placed in a correlation matrix to determine collinearity and simplify the final model. Figure 3 illustrates a high correlation between HICHOL and BPMED(0.91), HICHOL and CORHEART(0.87), BPMED and CORHEART(0.81), HIBP and HICHOL(0.76), HIBP and BPMED(0.81), STROKE and HICHOL(0.68), BPMED and STROKE(0.70), HIBP and CORHEART(0.88), STROKE and CORHEART(0.88), and HIBP and STROKE(0.94). Because of the higher correlations, some variables were dropped from the model.

Epidemiologic Models Considered:

After the initial crude model testing, assessing the impact of poverty vs. income, and deterimining highly correlated covariates, I tested three epidemiologic models that differed in the included covariates. The model formulations are listed below, and in all equations:

α is the intercept

$\beta 1$ is the independent variable (ENEBUR = energy burden)

$\gamma 1,2,3,4,5,6,7,8,9,10,11$ are covariates and potential confounders

Model 1:

(Death due to CVD) = α + β1(ENEBUR) + γ1(OLDHOME) + γ2(RACE) + γ3(ANNINCOME) + γ4(FELEDHOMES) + γ5(SES) + γ6(RENT) + γ7(UNEMPLOY) + γ8(LIFESTYLE) + γ9(ACCESS) + γ10(HIBP) + γ11(OBESE)

Model 1 represents the initial model created utilizing the DAG. This model includes EB, Old Homes, Race/ethnicity, Annual Income, Female-Led Households, Socioeconomic Status, People Who Rent, Unemployment, Lifestyle, and Physical Activity, Access to Healthcare, and Obesity. Because data regarding Old Homes, Female Led Households, and Renters were not easily accessible, these variables were dropped.

Model 2:

(Death due to CVD) = α + β1(ENEBUR) + γ1(HIBP) + γ2(OBESE) + γ3(DIABET) + γ4(LIFESTYLE) + γ5(SMOKE) + γ6(PCTPOV) + γ7(HOME)

Model 2 represents the second model created to accommodate the data that was available while also incorporating the elimination of inessential variables that can be represented as other variables. *Model 2* was the final model chosen for this analysis.

Model 3:

(Death due to CVD) = α + β1 (ENEBUR) + γ1(HIBP) + γ2(OBESE) + γ3(LIFESTYLE) + γ4(SMOKE) + γ5(PCTPOV) + γ6(HOME)

Model 3 represents a simplified version of *Model 2,* while incorporation the elimination of potentially collinear variables. *Model 3* was strongly considered as the final model, but the incorporation of both Obesity and Diabetes yielded different results in the analysis, leading to the selection of *Model 2* for the final version.

RESULTS

A summary of the overall county-level dataset of energy burden and cardiovascular disease mortality and results of the linear regression analyses are presented in Figures 4-6. The corresponding parameter estimates from the linear regression analyses are also presented in Tables 2-4.

Figure 4**:** *Combined Linear Regression Plot for Men and Women, All Ages, All Ethnicities with Confidence Band and Line of Best Fit*

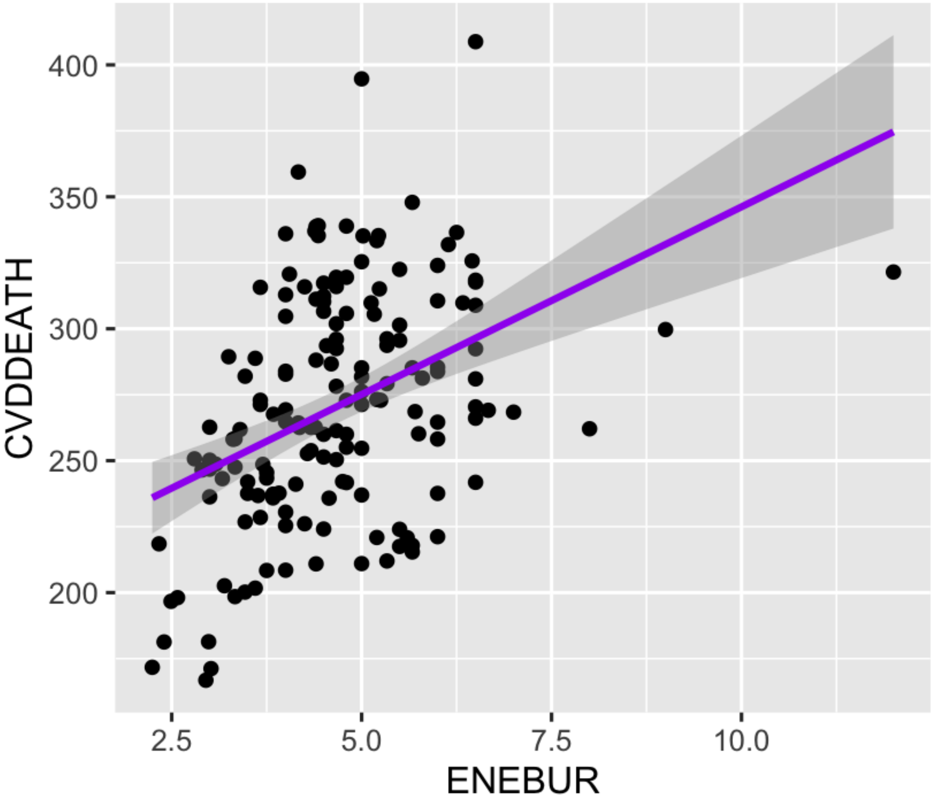

Figure 4 is a graphical output of the linear regression analysis between EB and CVD Mortality Rates in Georgia, with the dependent variable, CVDDEATH along the y- axis, and the dependent

variable, ENEBUR, along the x-axis. This output is represented as a scatterplot with a 95% Confidence Band and a Line of Best Fit included in the plot. This figure allows for a detailed visualization of a slight pattern forming towards the left side of the graph, with several outliers present. There is a variation in variance displayed in the plot, signifying that there is heteroscedasticity in the analysis, thus creates leading to the violation of the homoscedasticity assumption. This suggests that as ENEBUR (Energy Burden) increases, CVDDEATH (Cardiovascular Mortality) also increases, resulting in a positive correlation. The Line of Best Fit and Confidence Band allows for a clearer representation of the trend and relationship between the EB and CVD mortality rates.

Figure 5: *Linear Regression Plot for Men, All Ages, All Ethnicities with Confidence Band and Line of Best Fit*

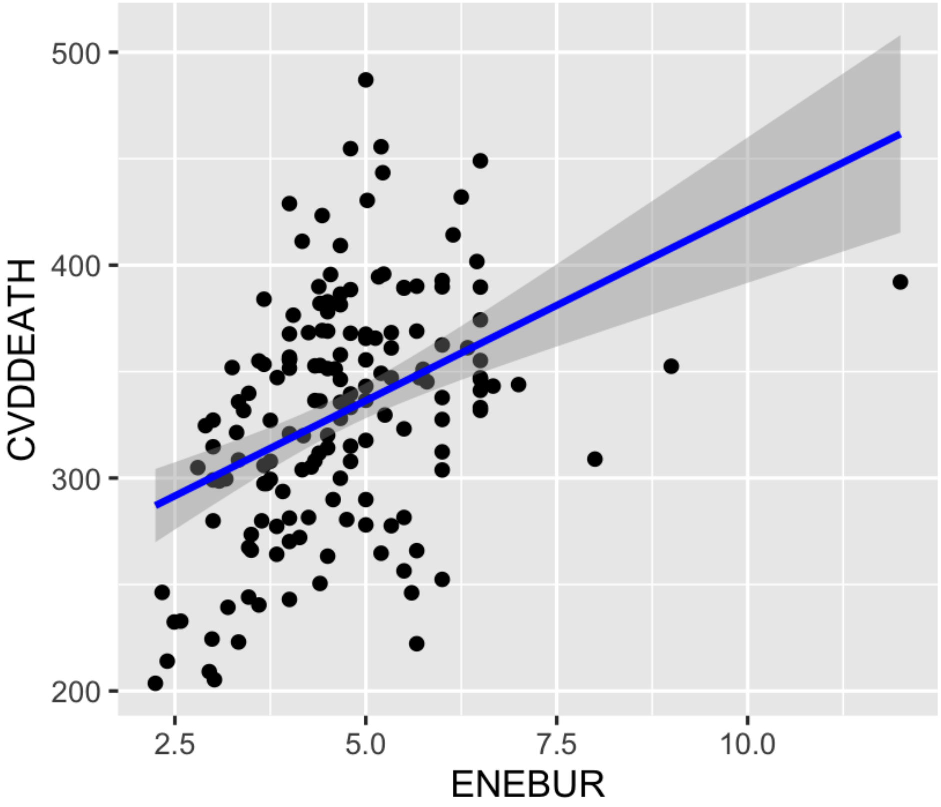

Figure 5 is an additional graphical output of the linear regression analysis between EB and CVD Mortality Rates in Georgia, with the dependent variable, CVDDEATH along the y- axis, and the dependent variable, ENEBUR, along the x-axis. This plot is one of two plots that have been stratified by sex (Figure 5 shows the results for Men). This output is represented as a more specific scatterplot with a 95% Confidence Band and a Line of Best Fit included in the plot. *Figure 5* allows for a more detailed visualization of a slight pattern forming towards the left side of the graph, with several outliers present. There is a variation in variance displayed in the plot, signifying that there is heteroscedasticity in the analysis, thus creates leading to the violation of the homoscedasticity assumption. This suggests that as ENEBUR (Energy Burden) increases, CVDDEATH (Cardiovascular Mortality) also increases, resulting in a positive correlation. The Line of Best Fit and Confidence Band allows for a clearer representation of the trend and relationship between the EB and CVD mortality rates. The results of this plot are extremely similar to Figure 4, however, the values from the analysis differ slightly.

Figure 6:*Linear Regression Plot for Women, All Ages, All Ethnicities with Confidence Band and Line of Best Fit*

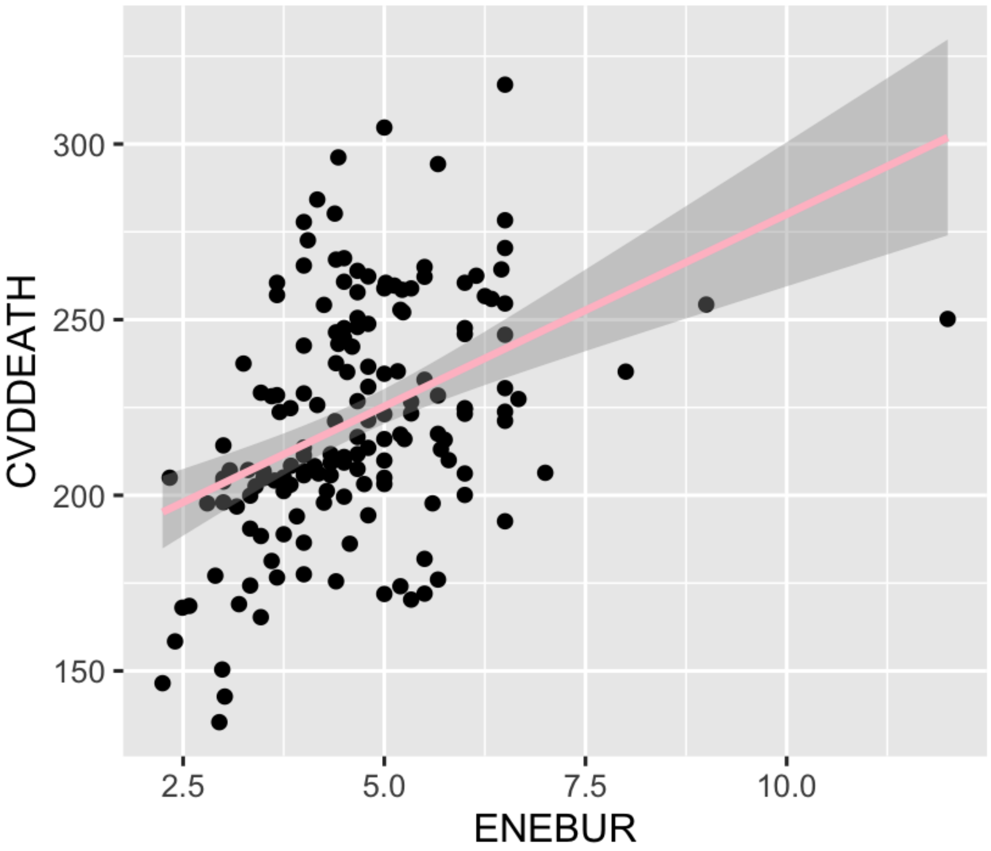

Figure 6 is the final graphical output of the linear regression analysis between EB and CVD Mortality Rates in Georgia, with the dependent variable, CVDDEATH along the y- axis, and the dependent variable, ENEBUR, along the x-axis. This plot is one of two plots that have been stratified by sex (Figure 6 shows the results for women). This output is represented as a more specific scatterplot with a 95% Confidence Band and a Line of Best Fit included in the plot. *Figure 6* allows for a more detailed visualization of a slight pattern forming towards the left side of the graph, with several outliers present. There is a variation in variance displayed in the plot, signifying that there is heteroscedasticity in the analysis, thus creates leading to the violation of

the homoscedasticity assumption. This suggests that as ENEBUR (Energy Burden) increases, CVDDEATH (Cardiovascular Mortality) also increases, resulting in a positive correlation. The Line of Best Fit and Confidence Band allows for a clearer representation of the trend and relationship between the EB and CVD Mortality rates. The results of this plot are also extremely similar to Figures 4 and 5.

Table 2: *Reported Regression Coefficients of the Linear Model Estimating the Association Between CVD Death and Energy Burden for Men and Women Combined*

Demographic Indicator	**Estimate**	**Stand. Error**	**T value**	**Pr (>|t|)**
Intercept	1.34e+02	6.45e+01	2.08	0.04*
ENEBUR	5.65e-01	3.43	0.17	0.87
HIBP	1.75	8.32e-01	2.10	0.04*
OBESE	1.07	1.03	1.04	0.29
DIABET	-1.36	2.88	-0.47	0.64
LIFESTYLE	2.48	1.21	2.05	0.04*
SMOKE	2.63	1.87	1.41	0.16
PCTPOV	-1.07	7.59e-01	-1.42	0.16
HOME	-3.26e-4	1.09e-04	-3.00	0.00

P value: 3.677e^-16

Min	**1Q**	**Median**	**3Q**	**Max**
-76.3	-23.8	-4.1	20.1	103.9

Table 2: This table represents the results of the linear regression model for the total mortality rate or deaths due to Cardiovascular disease among men and women combined. The equation for

this model is represented as: y=.0565(x)+134. Based on this information, it is suggestive that with every unit of exposure to energy burden, the rate of death due to cardiovascular disease increases by 0.565 units per year. Other health indicators that result in a higher risk of energy burden include LIFESTYLE (2.48) and SMOKE (2.63), which are controlled for in the analyses. The Standard Error is used to create the 95% confidence intervals for this analysis, which are (0.565 ± 1.96(3.43) = (-6.16,7.29). Based on this confidence interval, it can be stated that the actual slope is between -6.16 and 7.29. The t-value represents how far the standard error is away from zero. For this analysis, it can be stated that the t value is 0.17 standard errors away from zero. The P-values in the table marked with an asterisk suggest a statistically significant relationship. The results of the Total Men and Women combined analysis p-value reveal results that are not statistically significant.

Table 3: *Reported Regression Coefficients of the Linear Model Estimating the Association Between CVD Death and Energy Burden for Men*

Demographic Indicator	**Estimate**	**Stand. Error**	**T value**	**Pr (>|t|)**
Intercept	2.31e+02	8.13e+01	2.84	0.005**
ENEBUR	-1.60e-02	4.32	-0.004	1.00
HIBP	2.14	1.05	2.04	0.04*
OBESE	1.35	1.29	1.05	0.30
DIABET	-9.63e-01	3.62	-0.27	0.80
LIFESTYLE	2.89	1.53	1.89	0.06
SMOKE	7.24e-01	2.35	0.31	0.80
PCTPOV	-1.17	9.57e-01	-1.22	0.22
HOME	-5.47e-04	1.37e-04	-4.00	0.0001***

P value: 2.34e^-16

Min	**1Q**	**Median**	**3Q**	**Max**
-135.90	-25.60	-2.30	26.30	128.42

Table 3: This table represents the results of the linear regression model for the total mortality rate or deaths due to Cardiovascular disease among men. The equation for this model is represented as: y=-.0016(x)+231. Based on this information, it is suggestive that with every unit of exposure to energy burden, the rate of death due to cardiovascular disease in me decreases by 0.0016 units per year. Other health indicators that result in a higher risk of energy burden include LIFESTYLE (2.89), which is controlled for in the analyses. The Standard Error is used to create the 95% confidence intervals for this analysis, which are (-0.0016 ± 1.96(4.32) = 8.47(-8.47,8.47). Based on this confidence interval, it can be stated that the actual slope is between -8.47 and 8.47. The t-value represents how far the standard error is away from zero. For this analysis, it can be stated that the t value is -0.004 standard errors away from zero. The P-values in the table marked with an asterisk suggest a statistically significant relationship. The results of the Total Men's analysis p-value reveal results that are not statistically significant.

Table 4: *Reported Regression Coefficients of the Linear Model Estimating the Association Between CVD Death and Energy Burden for Women*

Demographic Indicator	**Estimated**	**Stand. Error**	**T value**	**Pr (>|t|)**
Intercept	1.53e+02	4.96e+01	3.08	0.002**
ENEBUR	8.90e-01	2.64	0.34	0.74
HIBP	8.11e-01	6.40e-01	1.27	0.20
OBESE	3.14e-01	7.90e-01	0.40	0.69
DIABET	-1.09	2.21	-0.49	0.62
LIFESTYLE	1.85	9.32e-01	1.99	0.05*
SMOKE	1.82	1.44	1.27	0.21
PCTPOV	-6.13e-01	5.84e-01	-1.04	0.30
HOME	-2.62e-04	8.37e-05	-3.13	0.002**

P value: 1.55e^-15

Min	**1Q**	**Median**	**3Q**	**Max**
-57.93	-19.20	-2.35	15.52	75.78

Table 4: This table represents the results of the linear regression model for the total mortality rate or deaths due to Cardiovascular disease among women. The equation for this model is

represented as: y=0.890(x)+153. Based on this information, it is suggestive that with every unit of exposure to energy burden, the rate of death due to cardiovascular disease increases in women by 0.890 units per year. Other health indicators that result in a higher risk of energy burden include HIBP (8.11e-01) and OBESE(3.14e-01), which are controlled for in the analyses. The Standard Error is used to create the 95% confidence intervals for this analysis, which are (0.890 ± 1.96(2.64) = (-4.28,6.06). Based on this confidence interval, it can be stated that the actual slope is between -4.28 and 6.06. The t-value represents how far the standard error is away from zero. For this analysis, it can be stated that the t value is 0.34 standard errors away from zero. The P-values in the table marked with an asterisk suggest a statistically significant relationship. The results of the Women's analysis p-value reveal results that are not statistically significant.

DISCUSSION

This study analyzed the relationship between exposure to energy burden (EB) and death due to cardiovascular disease within the state of Georgia, by county. Data for this analysis were collected from The Department of Energy's Low-Income Energy Affordability Data (LEAD) Tool and the Centers for Disease Control and Prevention's Division for Heart Disease and Stroke Prevention Interactive Atlas of Heart Disease and Stroke.

Overall, none of the linear regression analyses estimating the association of energy burden and cardiovascular disease mortality were statistically significant, and all point estimates had large confidence intervals. For the total combined analysis of men and women, the results suggest that there are 0.565 deaths per 100,000 for every 1 unit increase in energy burden. For the analysis of men only, the results suggest that there are 0.0016 deaths per 100,000 for every 1 unit increase in Energy Burden. For the analysis of women only, the results suggest that there are 0.890 deaths per 100,000 for every 1 unit increase in energy burden.

While non-significant, the pattern of results may suggest that women experience a higher mortality rate due to cardiovascular disease from greater energy burden exposure than men. This result could suggest that there are factors that contribute to the relationship between energy burden and cardiovascular disease that are unique to women. However, this finding is exploratory and should be investigated further in studies with more sufficient statistical power. While other works of literature have researched the impact of race/ethnicity regarding energy burden and health outcomes, there are limited research works that analyze the impact of gender on this relationship. Female-Led households have been a potential variable to include in other literature, which validates the results of the data, but there is limited data on specifics regarding women with children or single women with no children.

Strengths and Limitations:

There was finite information and little heterogeneity concerning EB by state that could account for any variables that may influence the number that is displayed within the DOE LEAD tool. The states that are highlighted within the tool that tend to have a higher energy burden also tend to have a lower population overall with less diversity to pinpoint association with certain chronic health outcomes. Because of this minimal data, it was decided to focus on one of the states encompassing one of the top cities in the US for higher percentages of EB. Georgia's, along with Michigan, Alabama, and Louisiana, main cities rank an average of about 5% or more for energy burden (NEXT CITY,2021; US Census, 2022). According to the Southern Environmental Law Center (SELC), Atlanta ranks fourth in the top five cities with the highest energy burden among low-income populations (SELC, 2019). These rankings led to a deeper analysis of Georgia, due to its higher population, which is about 10.9 million, according to the US Census, as of July 2022, and its fluctuation in weather and temperature.

Ultimately, this analysis was specifically conducted in a state with one of the highest energy burden percentages in the United States. Conducting the analysis in the south, where there is already a higher rate of cardiovascular disease, allows for more accurate reports of cardiovascular disease-related deaths. Georgia also has a higher population of people when compared to other states who also score higher in percentages of energy burden. The results of the study suggest that women experience more cardiovascular-related deaths due to energy burden exposure than men, which validates the appropriateness of the use of Female-Led Households as a variable to be considered for further investigation. According to several studies,

Women, especially Black women, lost employment at higher rates than men. The pandemic worsened employment opportunities due to pre-existing employment segregation, discrimination, unaffordable childcare options, and insufficiently paid family leave (Chen, 2022). County-level data could be a potential limitation as it is a very small representation of a bigger issue. The LEAD tool provides much more specific data based on tribal areas, but there are so many different areas to be considered. For future research behind this topic, research on the tribal areas mapped out in the LEAD tool obtaining specific reasons why an area could be at lower or higher risk for energy burden in the future would be ideal.

While using the Center for Disease Control and Prevention's Division for Heart Disease and Stroke Prevention Interactive Atlas of Heart Disease and Stroke, the data retrieved only accounted for Cardiovascular Related deaths, and did not include any Cardiovascular related Hospitalizations, which limits the data related to the at-risk population. If given the opportunity, including data on hospitalizations would be preferred and the results would be compared to the results of cardiovascular-related deaths. In the Center for Disease Control and Prevention's Division for Heart Disease and Stroke Prevention Interactive Atlas of Heart Disease and Stroke tool, Cardiovascular Disease-related Hospitalizations seemed to only include ages 65, but including all ages would be ideal and appropriate for this study.

Covid-19 influence was not accounted for in this data or in the analysis, but some additional research is needed to analyze how Covid-19 affected the costs of energy since energy use increased significantly while the majority of the population worked from home and did other activities that require energy use and internet connection. Researching the influence of Covid-19 on general health and its impacts on cardiovascular disease would be an interesting topic to consider for future research opportunities.

Although I was able to gather cardiovascular mortality data and stratified it by gender, the covariates that were chosen to be utilized in this model did not account for the differences in gender. This data only displayed the mortality rate overall between 2018 and 2020. With this in mind, the covariates were not a good representation of the population by sex. This lack of specific information is not a good representation of how the covariates, such as smoking, influence the relationship based on gender and therefore influence its specific effect on energy burden results. Simply put, the results of the indicators were the same for both men and women from the Center for Disease Control and Prevention’s Division for Heart Disease and Stroke Prevention Interactive Atlas of Heart Disease and Stroke, leading to similar results in the plots. Potential confounders such as home size, home ownership, renters, age of home, heating sources, and typical geographical weather were not included or accounted for in the project. These potential confounders influence how much energy will be used per individual household, especially the age of the home, size of the home, and localized typical weather conditions. Since more energy is typically used for these, the results may help specify a more clear representation of energy burden and would increase the likelihood of death due to Cardiovascular disease due to Energy Burden in the combined results of men and women.

REFERENCES

Bradshaw , J, et al. “Comparing the Effectiveness of Weatherization Treatments for Low-Income, American, Urban Housing Stocks in Different Climates.” *Energy and Buildings,* Elsevier, 20 Nov. 2014, https://www.sciencedirect.com/science/article/pii/S0378778813007329.

Brown, M, et al. “The Persistence of High Energy Burdens: A Bibliometric Analysis of Vulnerability, Poverty, and Exclusion in the United States.” *Energy Research & Social Science*, Elsevier, 14 Sept. 2020, https://www.sciencedirect.com/science/article/pii/S2214629620303315.

Chen , C, et al. “Localized Energy Burden, Concentrated Disadvantage, and the Feminization of Energy Poverty.” *IScience*, U.S. National Library of Medicine, 2021, https://pubmed.ncbi.nlm.nih.gov/35402875/.

EPA. “Climate Change Indicators: Residential Energy Use.” *EPA*, Environmental Protection Agency, 2021, https://www.epa.gov/climate-indicators/climate-change-indicators-residential-energy-use.

Graff , M, et al. “Which Households Are Energy Insecure? an Empirical Analysis of Race, Housing Conditions, and Energy Burdens in the United States.” *Energy Research & Social Science*, Elsevier, 18 June 2021, https://www.sciencedirect.com/science/article/pii/S2214629621002371).

Hernández, D, et al. “Benefit or Burden? Perceptions of Energy Efficiency Efforts among Low-Income Housing Residents in New York City.” *Energy Research & Social Science,* Elsevier, 23 May 2015, https://www.sciencedirect.com/science/article/pii/S2214629615000535.

IPCC — Intergovernmental Panel on Climate Change. https://www.ipcc.ch/site/assets/uploads/2019/11/SRCCL-Full-Report-Compiled-191128.pdf.

Jevtic, M. “Poverty and Energy Issues as Environmental and Health Challenges in SDGs.” *Academic.oup.com*, 2021, https://academic.oup.com/eurpub/article/31/Supplement_3/ckab164.734/6405669.

Mastropietro, P, et al. “Who Should Pay to Support Renewable Electricity? Exploring Regressive Impacts, Energy Poverty and Tariff Equity.” *Energy Research & Social Science*, Elsevier, 18 June 2019, https://www.sciencedirect.com/science/article/pii/S221462961930163X.

Reames, T, et al. “Exploring the Nexus of Energy Burden, Social Capital, and Environmental Quality in Shaping Health in US Counties.” *International Journal of Environmental*

Research and Public Health, U.S. National Library of Medicine, 2021, https://pubmed.ncbi.nlm.nih.gov/33450890/.

Reames, T, et al. "Targeting Energy Justice: Exploring Spatial, Racial/Ethnic and Socioeconomic Disparities in Urban Residential Heating Energy Efficiency." *Energy Policy*, Elsevier, 12 Aug. 2016, https://www.sciencedirect.com/science/article/pii/S0301421516304098.

Sanchez-Guevara, C, et al. "Assessing Population Vulnerability towards Summer Energy Poverty: Case Studies of Madrid and London." *UCL Discovery - UCL Discovery*, 1 May 2019, https://discovery.ucl.ac.uk/id/eprint/10070262/.

Shindell, Drew, et al. "The Effects of Heat Exposure on Human Mortality throughout the United States." *GeoHealth*, U.S. National Library of Medicine, 1 Apr. 2020, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7125937/.

Wang, Q. "Racial Disparities in Energy Poverty in the United States." *Renewable and Sustainable Energy Reviews*, Pergamon, 6 Dec. 2020, https://www.sciencedirect.com/science/article/pii/S1364032120309047.

Wolff, N, et al. *Large Scale Tropical Deforestation Drives Extreme Warming - Iopscience*. 2020, https://iopscience.iop.org/article/10.1088/1748-9326/ab96d2.

Other references and weblinks

DOE LEAD Tool Methodology **https://lead.openei.org/assets/docs/LEAD-Tool-Methodology.pdf**

DOE Interactive LEAD tool

https://www.energy.gov/scep/slsc/lead-tool

CDC Interactive Atlas of Heart Disease and Stroke Reports

https://nccd.cdc.gov/DHDSPAtlas/Reports.aspx

CDC Interactive Atlas of Heart Disease and Stroke Tool

https://nccd.cdc.gov/DHDSPAtlas/Default.aspx?state=GA

United States Energy Information Administration

https://www.eia.gov/

America's Council for an Energy-Efficient Economy

https://www.aceee.org/energy-burden

DOE WAP, 2016

https://www.energy.gov/scep/wap/weatherization-assistance-program

SELC, 2019

https://www.southernenvironment.org/news/community-and-faith-leaders-shed-light-on-georgians-energy-burden/

SELC, 2022

https://www.southernenvironment.org/news/flawed-studies-and-misleading-data-shouldnt-decide-future-of-memphis-power-supply/

NEXTCITY, 2021

https://nextcity.org/urbanist-news/new-data-shows-energy-burdens-across-50-major-cities

US Census, 2022

https://www.census.gov/quickfacts/fact/table/MI,AL,TN,GA/PST045222

APPENDICES

Appendix A : *Abbreviations*

EB - Energy Burden

CDC- Centers for Disease Control and Prevention

DOE- Department of Energy

EPA- Environmental Protection Agency

LEAD- Low-Income Energy Affordability Data

ACEEE- America’s Council for an Energy-Efficient Economy

IPCC-Intergovernmental Panel on Climate Change

CVD- Cardiovascular Disease

SES- Socioeconomic Status

WAP - Weatherization Assistance Program

Appendix B: *R Studio Code*

Code (Using R Studio):

```
library('tidyverse')
library('car')
library('haven')
library('dplyr')
library('survival')
library('survminer')
library('gtools')
library('readxl')
install.packages("plyr")                  # Install plyr package
library("plyr")
install.packages("dplyr")

AllCardio <- read_excel('/Users/danndria/downloads/AllThesisCVDdata.xlsx')
summary(AllCardio)
view(AllCardio)
MensCardio <-read_excel('/Users/danndria/downloads/MensThesisCVDdata.xlsx')
summary(MensCardio)
view(MensCardio)
WomensCardio <-read_excel('/Users/danndria/downloads/WomensThesisCVDdata.xlsx')
summary(WomensCardio)
view(WomensCardio)
energy <-read_excel('/Users/danndria/downloads/EnergyBurden.xlsx')
summary (energy)


#Calculate average Energy Burden per county
energy_county <- energy %>%                          # Specify data frame
  group_by(County) %>%                # Specify group indicator
  summarise_at(vars(PercentEnergyBurden),         # Specify column
          list(name = mean))

summary(energy_county)
view(energy_county)
```

```
##Merge Data here
total <- merge (AllCardio, energy_county, by="County")
view (total)
summary(total)

Total2 <- total

#Men
totalMen <- merge (MensCardio, energy_county, by="County")
view (totalMen)
summary(totalMen)

totalMen2 <- totalMen

#Women
totalWomen <- merge (WomensCardio, energy_county, by="County")
view (totalWomen)
summary(totalWomen)

totalWomen2 <- totalWomen

#######Rename columns##########

colnames(energy_county) <- c("County","EneBur")
print(energy_county)

colnames(AllCardio) <- c("CountNum","Display", "County", "CVDDEATH","Range",
                "CORHEART","HIBP","STROKE","HICHOL","DIABET","OBESE",
                "LIFESTYLE","SMOKE","HISDIP","CODIP","INCOME","PCTPOV",
                "UNEMPLOY","BPMED","CHOLSCRE","HEALTHI",
                "HOME","HEARTDI")

colnames(MensCardio) <- c("CountNum","Display", "County", "CVDDEATH","Range",
                "CORHEART","HIBP","STROKE","HICHOL","DIABET","OBESE",
                "LIFESTYLE","SMOKE","HISDIP","CODIP","INCOME","PCTPOV",
                "UNEMPLOY","BPMED","CHOLSCRE","HEALTHI",
                "HEARTDI","HOME")
print(MensCardio)

colnames(WomensCardio) <- c("CountNum","Display", "County", "CVDDEATH","Range",
                    "CORHEART","HIBP","STROKE","HICHOL","DIABET","OBESE",
                    "LIFESTYLE","SMOKE","HISDIP","CODIP","INCOME","PCTPOV",
                    "UNEMPLOY","BPMED","CHOLSCRE","HEALTHI",
                    "HEARTDI","HOME")
print(WomensCardio)

colnames(total) <- c("CountNum","Display", "County", "CVDDEATH","Range",
                     "CORHEART","HIBP","STROKE","HICHOL","DIABET","OBESE",
                     "LIFESTYLE","SMOKE","HISDIP","CODIP","INCOME","PCTPOV",
                     "UNEMPLOY","BPMED","CHOLSCRE","HEALTHI",
                     "HOME","HEARTDI","ENEBUR")
colnames(Total2) <- c("CountNum","Display", "County", "CVDDEATH","Range",
                     "CORHEART","HIBP","STROKE","HICHOL","DIABET","OBESE",
                     "LIFESTYLE","SMOKE","HISDIP","CODIP","INCOME","PCTPOV",
                     "UNEMPLOY","BPMED","CHOLSCRE","HEALTHI",
                     "HOME","HEARTDI","ENEBUR")
colnames(totalMen) <- c("CountNum","Display", "County", "CVDDEATH","Range",
                 "CORHEART","HIBP","STROKE","HICHOL","DIABET","OBESE",
                 "LIFESTYLE","SMOKE","HISDIP","CODIP","INCOME","PCTPOV",
                 "UNEMPLOY","BPMED","CHOLSCRE","HEALTHI",
                 "HEARTDI","HOME","ENEBUR")
colnames(totalMen2) <- c("CountNum","Display", "County", "CVDDEATH","Range",
                 "CORHEART","HIBP","STROKE","HICHOL","DIABET","OBESE",
                 "LIFESTYLE","SMOKE","HISDIP","CODIP","INCOME","PCTPOV",
                 "UNEMPLOY","BPMED","CHOLSCRE","HEALTHI",
                 "HEARTDI","HOME","ENEBUR")
colnames(totalWomen) <- c("CountNum","Display", "County", "CVDDEATH","Range",
                 "CORHEART","HIBP","STROKE","HICHOL","DIABET","OBESE",
                 "LIFESTYLE","SMOKE","HISDIP","CODIP","INCOME","PCTPOV",
                 "UNEMPLOY","BPMED","CHOLSCRE","HEALTHI",
                 "HEARTDI","HOME","ENEBUR")
colnames(totalWomen2) <- c("CountNum","Display", "County", "CVDDEATH","Range",
                 "CORHEART","HIBP","STROKE","HICHOL","DIABET","OBESE",
                 "LIFESTYLE","SMOKE","HISDIP","CODIP","INCOME","PCTPOV",
                 "UNEMPLOY","BPMED","CHOLSCRE","HEALTHI",
                 "HEARTDI","HOME","ENEBUR")
#####*****Value = Age Standardized Rate per 100,000
##Linear Regression
#Total
Model1 <- lm(CVDDEATH ~ ENEBUR + CORHEART + HIBP +
          STROKE + HICHOL + DIABET + OBESE +
          LIFESTYLE + SMOKE + HISDIP + CODIP +
          INCOME + PCTPOV + UNEMPLOY + CHOLSCRE +
          HEALTHI + HEARTDI + HOME,
         data =Total2)
summary (Model1)
#Men
MensModel <-lm(CVDDEATH ~ ENEBUR + CORHEART + HIBP +
             STROKE + HICHOL + DIABET + OBESE +
             LIFESTYLE + SMOKE + HISDIP + CODIP +
             INCOME + PCTPOV + UNEMPLOY + CHOLSCRE +
             HEALTHI + HEARTDI + HOME,
            data =totalMen2)
summary (MensModel)
#WomensTotal
WomensModel <- lm(CVDDEATH ~ ENEBUR + CORHEART + HIBP +
             STROKE + HICHOL + DIABET + OBESE +
             LIFESTYLE + SMOKE + HISDIP + CODIP +
             INCOME + PCTPOV + UNEMPLOY + CHOLSCRE +
             HEALTHI + HEARTDI + HOME,
            data =totalWomen2)
summary (WomensModel)
##################CRUDE ASSOCIATIONS########################
crude <- lm(CVDDEATH ~ ENEBUR, data=Total2)
summary(crude)

crudeMens <-lm(CVDDEATH ~ ENEBUR, data=totalMen2)
summary(crudeMens)

crudeWomens <-lm(CVDDEATH ~ ENEBUR, data=totalWomen2)
summary(crudeWomens)



crude2 <- lm(CVDDEATH ~ ENEBUR + INCOME, data=Total2)
summary(crude2)
#apply an SES variable as a confounder here
#remove variables as necessary
#make note of how correlated each SES variable is (income,pctpov)
#control for covariates (risk factors like obesity)
#run crude model for men/women and then run with precision cov included (single or multiple)
```

```
#should we include income/SES, run SES covariates, may erase data (talk about in discussion)

############TO-DO##############

#crude association "Value~name"
# no_hsdip + no_college + income + povpct + unemploy
#summarize results
#thesis outline/results section


#############DECIDING WHICH COVARIATES TO USE###################
#SES VARAIBLES

##VARIABLE NAMES
#Value = Deaths
#Name = Energy Burden Percentage (%)
#pl_chd = Coronary Heart disease(%)
#pl_bphigh = High Blood Pressure(%)
#pl_stroke = Stroke(%)
#pl_highchol = High cholesterol(%)
#dm_prev_adj = Diabetes(%)
#ob_prev_adj = Obesity(%)
#ltpia_prev_adj = physical inactivity (%)
#pl_csmoking = current Smoking status(%)
#no_hsdip = No high school diploma (Less than HS (%))
#no_college = (Less than college (%))
#income = Median Household Income ($)
#povpct = percent poverty (%)
#unemploy = unemployment rate
#pl_cholscreen = cholesterol screening (%)
#pctui = Health insurance status (%)
#prev_hd = Heart Disease Prevalence (%)
#home = Home value ($)

##SES##
#Total


dat <- Total2
cor(dat$HISDIP, dat$CODIP, #highly correlated
    method = "spearman")

cor(dat$INCOME, dat$PCTPOV,
    method = "spearman")

cor(dat$INCOME, dat$UNEMPLOY,
    method = "spearman")

cor(dat$INCOME, dat$HOME, ####highly correlated
    method = "spearman")

cor(dat$PCTPOV, dat$UNEMPLOY, #pretty high correlation
    method = "spearman")

####SES relationship
dat2 <- total %>%
  as.data.frame() %>%
  select (HISDIP,CODIP,INCOME,
          PCTPOV,UNEMPLOY,HOME
          ) %>%
  as.matrix() %>%
  cor()
print (dat2)

install.packages("ggcorrplot")
library(ggcorrplot)

ggcorrplot(dat2, method = "circle")
ggcorrplot(dat2, hc.order = TRUE,
           type = "lower",
           col = colorRampPalette(c("white", "pink", "blue4"))(5),
           lab = TRUE)
#######health relationship

dat3 <- Total2 %>%
  as.data.frame() %>%
  select (CORHEART,
          HIBP,
          STROKE ,
          HICHOL,
         DIABET,
         OBESE,
         LIFESTYLE,
         BPMED
  ) %>%
  as.matrix() %>%
  cor()
print (dat3)


ggcorrplot(dat3, method = "circle")
ggcorrplot(dat3, hc.order = TRUE,
           type = "lower",
           col = colorRampPalette(c("white", "purple", "blue4"))(5),
           lab = TRUE)


####Final Model?######
#Overall
ModelTotal<- lm(CVDDEATH ~ ENEBUR + HIBP  + OBESE + DIABET +
                  LIFESTYLE + SMOKE + PCTPOV + HOME, data =Total2)
summary (ModelTotal)

#Men
MenModelTotal<- lm(CVDDEATH ~ ENEBUR + HIBP  + OBESE + DIABET +
                     LIFESTYLE + SMOKE + PCTPOV + HOME, data =totalMen2)
summary (MenModelTotal)
#Women
WomenModelTotal<- lm(CVDDEATH ~ ENEBUR + HIBP  + OBESE + DIABET +
                       LIFESTYLE + SMOKE + PCTPOV + HOME, data =totalWomen2)
summary (WomenModelTotal)

#Men and women got VERY different results using this model
#should i run a different model between the genders?

#plotting the results

#Total
plot(CVDDEATH ~ ENEBUR, data = Total2)
abline(ModelTotal)
```

```
library(ggplot2)

ggplot(Total2, aes ( x = ENEBUR, y = CVDDEATH)) +
  geom_point() +
  stat_smooth(method = "lm", col = "purple")

#Men
plot(CVDDEATH ~ ENEBUR, data = totalMen2)
abline(MenModelTotal)
library(ggplot2)

ggplot(totalMen2, aes ( x = ENEBUR, y = CVDDEATH)) +
  geom_point() +
  stat_smooth(method = "lm", col = "blue")

#Women
plot(CVDDEATH ~ ENEBUR, data = totalWomen2)
abline(WomenModelTotal)
library(ggplot2)

ggplot(totalWomen2, aes ( x = ENEBUR, y = CVDDEATH)) +
  geom_point() +
  stat_smooth(method = "lm", col = "pink")


#install.packages("rmarkdown")
```

Appendix C: DAGitty Model Code

DAGitty Model Code for
Figure 1.0

```
dag {
bb="0,0,1,1"
"Access to Healthcare" [pos="0.219,0.771"]
"Annual Income" [pos="0.432,0.431"]
"Female-Led Households" [pos="0.867,0.592"]
"Hypertension/High Blood Pressure" [pos="0.677,0.789"]
"Lifestyle/Diet" [pos="0.431,0.815"]
"Old Home" [pos="0.234,0.077"]
"Race/Ethnicity" [pos="0.488,0.109"]
CVD [outcome,pos="0.293,0.684"]
EB [exposure,pos="0.302,0.202"]
Obesity [pos="0.083,0.646"]
Renters [pos="0.688,0.616"]
SES [pos="0.717,0.287"]
Unemployment [pos="0.884,0.455"]
Unknown [pos="0.735,0.076"]
"Access to Healthcare" -> "Lifestyle/Diet"
"Access to Healthcare" -> CVD
"Annual Income" -> "Female-Led Households"
"Annual Income" -> "Lifestyle/Diet"
"Annual Income" -> EB
"Annual Income" -> Renters
"Annual Income" -> SES
"Female-Led Households" -> Renters
"Hypertension/High Blood Pressure" -> CVD
"Lifestyle/Diet" -> "Hypertension/High Blood Pressure"
"Lifestyle/Diet" -> CVD
"Lifestyle/Diet" -> Obesity
"Old Home" -> EB
"Race/Ethnicity" -> "Annual Income"
"Race/Ethnicity" -> "Old Home"
"Race/Ethnicity" -> CVD
"Race/Ethnicity" -> EB
"Race/Ethnicity" -> SES
"Race/Ethnicity" -> Unknown
CVD <-> EB
Obesity -> CVD
Renters -> "Lifestyle/Diet"
Renters -> EB
SES -> Renters
Unemployment -> Renters
Unemployment -> SES
}
```

DAGitty Model Code for
Figure 1.1

```
dag {
bb="0,0,1,1"
"Access to Healthcare" [pos="0.219,0.771"]
"Annual Income" [pos="0.432,0.431"]
"Female-Led Households" [pos="0.867,0.592"]
"Hypertension/High Blood Pressure" [pos="0.677,0.789"]
"Lifestyle/Diet" [pos="0.431,0.815"]
"Old Home" [pos="0.234,0.077"]
"Race/Ethnicity" [pos="0.488,0.109"]
CVD [outcome,pos="0.293,0.684"]
EB [exposure,pos="0.302,0.202"]
Obesity [pos="0.083,0.646"]
Renters [pos="0.688,0.616"]
SES [pos="0.717,0.287"]
Unemployment [pos="0.884,0.455"]
Unknown [pos="0.735,0.076"]
"Access to Healthcare" -> "Lifestyle/Diet"
"Access to Healthcare" -> CVD
"Annual Income" -> "Female-Led Households"
"Annual Income" -> "Lifestyle/Diet"
"Annual Income" -> EB
"Annual Income" -> Renters
"Annual Income" -> SES
"Female-Led Households" -> Renters
"Hypertension/High Blood Pressure" -> CVD
"Lifestyle/Diet" -> "Hypertension/High Blood Pressure"
"Lifestyle/Diet" -> CVD
"Lifestyle/Diet" -> Obesity
"Old Home" -> EB
"Race/Ethnicity" -> "Annual Income"
"Race/Ethnicity" -> "Old Home"
"Race/Ethnicity" -> CVD
"Race/Ethnicity" -> EB
"Race/Ethnicity" -> SES
"Race/Ethnicity" -> Unknown
CVD <-> EB
EB -> "Hypertension/High Blood Pressure"
EB -> "Lifestyle/Diet"
EB -> Obesity
Obesity -> CVD
Renters -> "Lifestyle/Diet"
Renters -> EB
SES -> Renters
Unemployment -> Renters
Unemployment -> SES
}
```