Social Determinants of Health

A Major Qualifying Project Proposal submitted to the faculty of WORCESTER POLYTECHNIC INSTITUTE in partial fulfillment of the requirements for the degree of Bachelor of Science

Submitted By:

Demetre Doherty

Kevin McGonigle

Eric Murdza

Ngoc Pham

Project Advisors:

Professor Jon Abraham

Professor Barry Posterro

December 2022

Abstract

This project assisted Milliman, an actuarial consulting firm, by creating a PowerBI dashboard to inform them of the implications Social Determinants of Health (SDoH) have on health outcomes of people 65 and older on Medicare. The dashboard includes filters to change certain SDoH or percentile of data to see the change in the average hospitalizations/emergency room visits by county. To determine the SDoH included in the dashboard, the team utilized information from various models to select the most significant variables. The team designed and created dashboards available to Milliman on the potential impacts that SDoH have on people's health outcomes.

Executive Summary

Context

The focus of this project is Social Determinants of Health and their implications on the health outcomes of people. **Social Determinants of Health** (SDoH) are the “conditions in the environments where people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks” (Social Determinants of Health, 2022). They are often broken down into five major groups:

- Economic Stability
- Education Access and Quality
- Health care Access and Quality
- Neighborhood and Built Environment
- Social and Community Context

The team decided to concentrate on four of the five major SDoH groups to evaluate which included all the major groups except neighborhood and built environment (ibid).

Many entities in the healthcare industry have begun to identify the importance of researching SDoH and are incorporating this research into public policy, risk calculations, and more. Certain aspects of these SDoH are included in medical claims data, but because this approach is so new, many SDoH records are incomplete, making it difficult to analyze their impacts.

Our Sponsor

Milliman is an actuarial consulting company that is using their expert analysis to assist a plethora of organizations that need guidance. Their mission statement is, "Our expert guidance and advanced technology solutions empower leading insurers, healthcare organizations and employers to protect the health and financial well-being of people everywhere" (Milliman, 2022). Milliman works closely with their clients to assist them with both social and business challenges such as retirement, healthcare, effects from climate change or a pandemic, and the effect of low interest rates. Milliman continues to try and find new ways to ensure that they have access to as much data as possible, which led them to focus on the impacts of SDoH (ibid).

Approach

The goal of the project was to provide insight into Social Determinants of Health and their possible effects on certain populations' health. To sufficiently complete this task, data had to be collected and formatted to only show the most vital variables attributed to social determinants of health and population health measures. The team collected data from the Centers for Medicare and Medicaid Services (CMS), the US Census, and the Center for Disease Control and Prevention (CDC) to acquire health outcomes, SDoH variables, and COVID-19 data. The data was then integrated into models and dashboards to discover important variables and display findings.

Results and Analysis

Throughout the project, there were over 100 variables that the team had to investigate to determine their significance in predicting health outcomes and their relationship with other variables. The group created models and selected three variables from their models as the most significant variables:

- Household income of less than 25 thousand per year (percent of county)
- Bachelor’s degree or higher attainment (percent of county)
- West region of the US

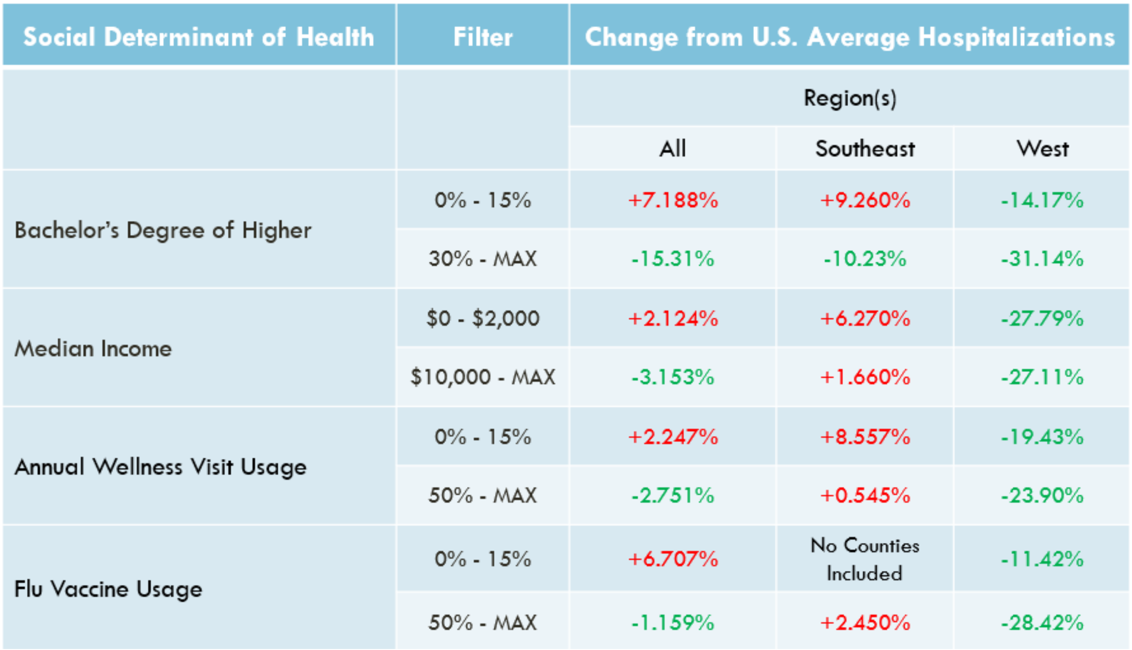
The team selected these variables because they were in the most models and were closely related to many of the other predictor variables, implying they provide the most information. In addition, creating a model to predict the number of hospitalizations reduced the most error compared to only using the mean whereas models predicting the number of ER visits reduced more error when looking at specific conditions.

The team prioritized the logistic models, creating several different models using different step functions and transformations of covariates. Overall, because hospitalizations have a smaller right tail, the model was simpler and had a lower true positive rate. When combining 2020 data and 2021 data, most of the variables remained significant, implying that the observed associations are more likely to continue in future years. The team used their final models to create confusion matrices for predicting if counties had high hospitalizations and high emergency room visit rates. The group members used the 2020 US Census data to create their models and then tried to predict whether a county from the 2021 Census data would be high or not. The models had an accuracy of 0.83 for both hospitalizations and ER visits high classification.

Milliman requested the results of the thorough data analysis be organized within a dashboard software, and this was achieved through PowerBI. It was crucial to show the data in a simplistic manner such that the team and eventually the sponsor could form opinions on the outcomes shown. The significant Social Determinants of Health variables used in the dashboards were:

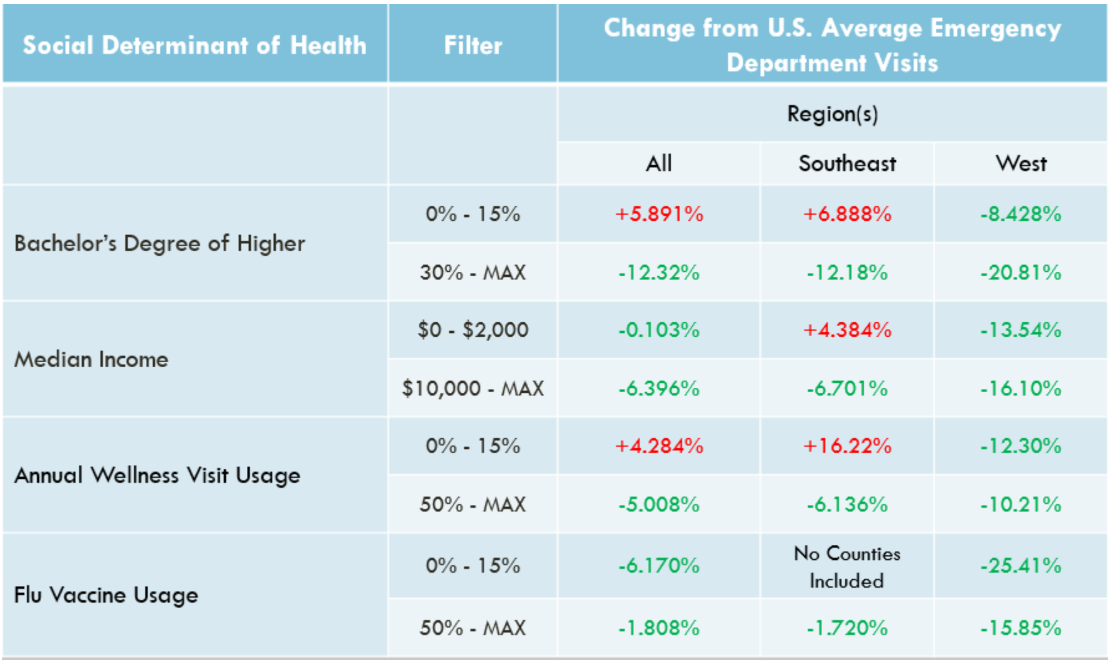
- Bachelor’s degree or higher (percent of county)
- Median income of county ($)
- Annual wellness visit usage (percent of county)
- Flu vaccine usage (percent of county).

The main discoveries about Social Determinants of Health and from the dashboards created are summarized in the following two figures:



Social Determinant of Health	Filter	Change from U.S. Average Hospitalizations		
		Region(s)		
		All	Southeast	West
Bachelor’s Degree of Higher	0% - 15%	+7.188%	+9.260%	-14.17%
	30% - MAX	-15.31%	-10.23%	-31.14%
Median Income	$0 - $2,000	+2.124%	+6.270%	-27.79%
	$10,000 - MAX	-3.153%	+1.660%	-27.11%
Annual Wellness Visit Usage	0% - 15%	+2.247%	+8.557%	-19.43%
	50% - MAX	-2.751%	+0.545%	-23.90%
Flu Vaccine Usage	0% - 15%	+6.707%	No Counties Included	-11.42%
	50% - MAX	-1.159%	+2.450%	-28.42%

Figure E.1: Change in Hospitalizations based on SDoH Filters



Social Determinant of Health	Filter	Change from U.S. Average Emergency Department Visits		
		Region(s)		
		All	Southeast	West
Bachelor's Degree of Higher	0% - 15%	+5.891%	+6.888%	-8.428%
	30% - MAX	-12.32%	-12.18%	-20.81%
Median Income	$0 - $2,000	-0.103%	+4.384%	-13.54%
	$10,000 - MAX	-6.396%	-6.701%	-16.10%
Annual Wellness Visit Usage	0% - 15%	+4.284%	+16.22%	-12.30%
	50% - MAX	-5.008%	-6.136%	-10.21%
Flu Vaccine Usage	0% - 15%	-6.170%	No Counties Included	-25.41%
	50% - MAX	-1.808%	-1.720%	-15.85%

Figure E.2: Change in Emergency Department Visits based on SDoH Filters

In both health outcome (hospitalizations/ER visits) summaries, as the lower percentages and dollar amounts of SDoH variables were selected, there were patterns of increased number of outcomes from the national averages. In addition, as the higher percentages and dollar amounts of SDoH variables were selected, there were patterns of decreased number of outcomes from the national averages. Aside from a few deviations, there seemed to be an underlying pattern or connection between significant SDoH variables and these health outcomes.

Conclusions

Overall, the educational attainment, income, and counties in the West region of the U.S. were the most significant variables for predicting the health outcomes of a county. Interestingly, it was observed that counties in the West have significantly fewer hospitalization and ER visit rates compared to other regions. The reason for the drastic difference is unclear, but the western region should be explored further on its own to deduce that reason. Some possible ideas that may be explored are that the West region imposed greater COVID-19 restrictions than other regions, thus limiting the number of hospitalizations or ER visits possible, or that the counties in Alaska and Hawaii may have pulled down the average in the West region. In the logistic models, language other than English and urban were more prevalent variables, implying that these variables are more common in worse counties, but each variable alone may not consistently contribute to worse health outcomes.

When modeling high hospitalizations/ER, the team decided to classify a health outcome as being "high" if the county was above the 80th percentile of that respective health outcome and "not high" if the county was below the 80th percentile. However, experimenting with different cut-off points could lead to the discovery of other important variables, and may make the models more accurate. In addition, looking into counties that are consistently classified as high could help identify variables in these regions that could be used for future predictions.

The team was able to analyze the significant variables mentioned above over a 3-year time span, 2019 to 2022. Therefore, for further analysis, observing how these variables impact health outcomes over a longer period using time series techniques can both validate the predictive quality of these variables and how these variables can be used to predict the changes in health outcomes over time. It is unknown whether there was simply a correlation between these variables and the change in health outcomes, or if there was actual causation. An analysis of counties with changes in SDoH over time could show the impact of improving or worsening SDoH in counties.

Finally, the analysis in this project only split up the data based on the counties' geographic areas. However, more can be done to divide these regions into groups with similar health outcomes trends. One approach is to use unsupervised cluster models like K-means and a

Gaussian Mixture Model. These models group data into different clusters based on their similar characteristics. Trying both these models and comparing their results can help to find and validate common health outcomes trends.

Acknowledgements

The team would first like to acknowledge the project's sponsor, Milliman, for the introduction to Social Determinants of Health and for the exhaustive list of skills and ideas the team learned while this research was completed. Specifically, without Danielle Rubin, Milliman actuary associate, this project would not have been possible. Her feedback, direction, and time spent with the team on the project was greatly appreciated.

The team would also like to thank our advisors, Jon Abraham and Barry Posterro, for their time spent consulting with the team and their helpful insight to keep the project precise and successful. The project was organized and coherent because of their observations and suggestions.

Table of Contents

Table of Figures

Chapter 1: Introduction

The focus of this project is Social Determinants of Health (SDoH) and their implications on the health outcomes of people. Social Determinants of Health are the conditions in the environments in which people live that may influence their health (CDC, 2022). This project was sponsored by Milliman, a company in the actuarial consulting space that provides actuarial products and services. Because Milliman provides health-related risk services to their clients, it is essential that they are experts in the social factors that affect their clients' risks. Recently through SDoH research, health care companies and consultants can more accurately assess a populations' health risks. Many entities in the healthcare industry have begun to identify the importance of researching SDoH and are incorporating this research into public policy, risk calculations, and more. Certain aspects of these SDoH are included in medical claims data, but because this approach is so new, many SDoH records are incomplete, making it difficult to analyze their impacts. As Milliman continues to develop their knowledge surrounding SDoH, one of their goals is to advise their clients on ways that they can allocate resources to improve health care outcomes. Therefore, under Milliman's guidance, the team examined the effects of multiple SDoH within public data to discern associations between SDoH and population health statistics.

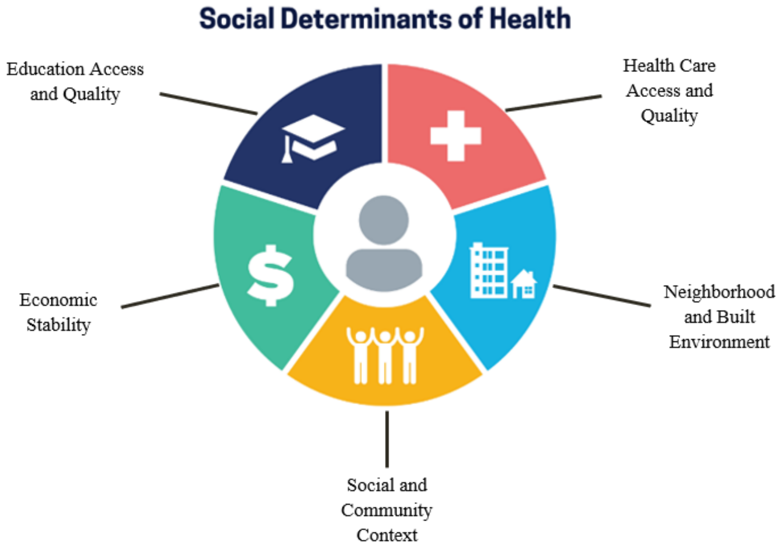

Social Determinants of Health
Education Access and Quality
Health Care Access and Quality
Economic Stability
Neighborhood and Built Environment
Social and Community Context

Figure 1.3: Social Determinants of Health

The public data used during the project came from the U.S. Medicare website, the U.S. Census, and the CDC. Due to this limitation to public data and Medicare data in general, the team reduced the scope of the project to only include people aged 65 and older. This decision made it possible to use the data accessible most effectively to the group, considering Medicare data is mostly composed of beneficiaries 65 and older (Statista Research Department, 2022).

The goal of this project was to analyze the potential impacts that Social Determinants of Health have on populations, specifically their health outcomes. First, the team gathered and merged data from three online databases:

- US Medicare (health statistics)
- US Census (SDoH data)
- CDC (COVID data)

Next, the group members coded models that selected the most significant variables using both step AIC algorithms and likelihood ratio tests. Using the information from the models, the team visualized the data by creating confusion matrices and bar graphs to better understand the impact that each significant variable had on hospitalizations and ER visits. Finally, the group created a dashboard using PowerBI to display the conclusions that they made throughout the project. This project would help provide Milliman with useful information about the impacts of SDoH to help with their goal to “protect the health and financial well-being of people everywhere.”

Chapter 2: Background

The team decided to concentrate on four major SDoH groups to evaluate given the constraints with available data and directed attention toward the SDoH that made the most sense to analyze. Those included:

- Economic Stability
- Education Access and Quality
- Health Care Access and Quality
- Social and Community Context

Even when concentrating on a few SDoH, many limitations arose because most acceptable public data was from the Centers for Medicare and U.S. Census websites.

The group also explored a fifth SDoH: Neighborhood and Built Environment. However, finding data that was relevant in this domain proved to be an arduous task. The data that would be relevant to this sector of SDoH may be available from other private sources, but the team did not have the resources or time to investigate them. Therefore, any important implications from this SDoH could only be assumed based on studies already done in the industry and will not be discussed in this project.

2.1 What are Social Determinants of Health?

Social Determinants of Health (SDoH) are the "conditions in the environments where people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks" (Social Determinants of Health, 2022). They are often broken down into five major groups as seen in Figure 2.2: economic stability, education access and quality, health care access and quality, neighborhood and built environment, and social and community context. The colored SDoH were the focal points of the team's project, with subsections that include some topics that the team examined throughout the project.

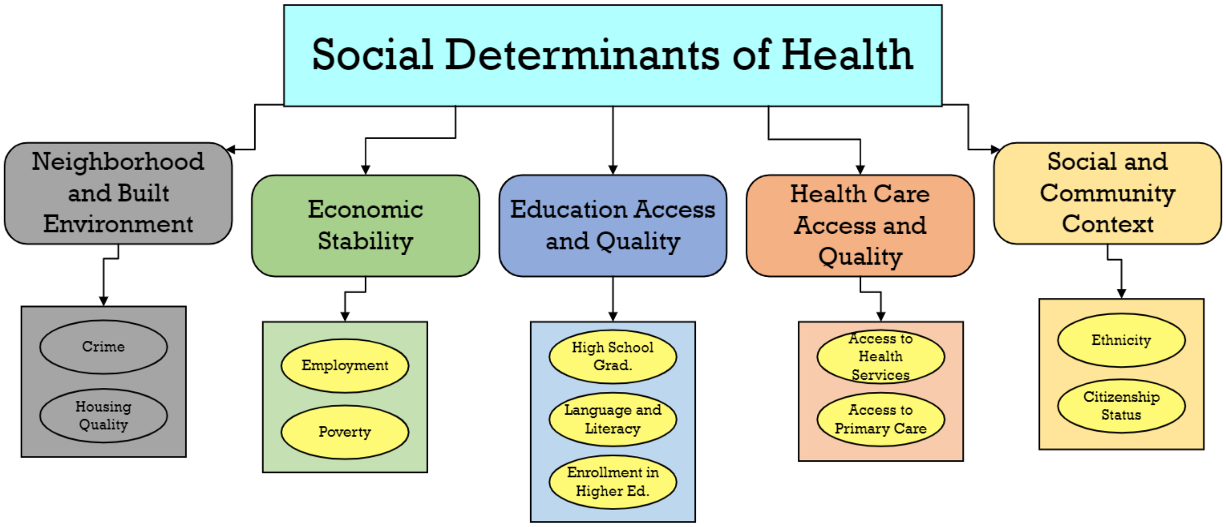

Social Determinants of Health
Neighborhood and Built Environment
Economic Stability
Education Access and Quality
Health Care Access and Quality
Social and Community Context
Crime
Housing Quality
Employment
Poverty
High School Grad.
Language and Literacy
Enrollment in Higher Ed.
Access to Health Services
Access to Primary Care
Ethnicity
Citizenship Status

Figure 2.4: Focal Social Determinants of Health in this Project

Some of these determinants are interconnected and those connections may explain why certain demographics of populations are more likely to have worse or better health. For example, "lack of education can impact employment opportunities which in turn constrain income. Low income reduces access to healthcare and nutritious food and increases hardship. Hardship causes stress which in turn promotes unhealthy coping mechanisms such as substance abuse and overeating of unhealthy foods" (NEJM, 2017). The main goal of gauging a populations' SDoH is to provide context for that environment and how it can impact the people in that area's health now and in the future.

2.2 Economic Stability

One main social determinant of health is economic stability. This determinant focuses on the impact that employment and income can have on someone's well-being. Some specific examples that can be derived from a person's economic stability are food security, housing stability, and poverty level. For this SDoH category, the team concentrated on searching for different health effects that related to employment and poverty levels since there was public data that attributed to those categories.

2.2.1 Employment

The jobs that people hold influence their health, not only because of the exposure to physical distress, but also because jobs may administer situations where healthy activities are endorsed (Egerter et al., 2008). In the U.S., many jobs provide Americans health care insurance in addition to the supply of income that enables them to live a healthier lifestyle. Obtaining employment can be a start to leading a healthy life, barring other factors that could negatively affect their health. For example, people who work in certain sectors, like a hospital or in construction, may make those workers more susceptible to occupational injuries over other less physical jobs and thus could risk having worse health in the future.

Research says that people who are unemployed are also linked to negative health results. It is reported that unemployed individuals "tend to suffer more from stress-related illnesses such as high blood pressure, stroke, heart attack, heart disease, and arthritis" (Employment, 2022). Unemployed people may also lack the income they need to live a healthy lifestyle, and subsequently could end up lacking the health care they need as well. In addition to these problems, unemployed people also are reported to have feelings of depression, anxiety, and other mental health issues that could lead to mental illness in their future (Employment, 2022). Knowledge of these findings directed us to research the qualitative effects of unemployment and the related costs for beneficiaries under Medicare.

2.2.2 Poverty

Poverty often occurs in concentrated areas for extended periods of time (Rural Poverty & Well-being, 2022). People that live in these impoverished areas are more likely to have reduced resources needed to support a healthy quality of life, and examples of these include stable housing, healthy foods, and safe neighborhoods (Singh & Siahpush, 2006). In addition to these detriments impacting their health, there are also connections between poverty and education since poverty has been linked to limiting the access of educational opportunities. This connection between education and poverty only further impedes people in poverty from acquiring a sufficient job and may create a perpetual cycle of poverty.

2.2.3 Food Insecurity

A portion of the population are exposed to food insecurity in their everyday lives which may negatively impact their health. Food insecurity is defined as a “household-level economic and social condition of limited or uncertain access to adequate food” and that can range from not having enough income for food, or simply not having a supermarket within reasonable distance to travel. In 2020, 28.6% of low-income households were food insecure, compared to the national average of 10.5% (Key Statistics & Graphics, 2022). In addition to low-income as a cause for food insecurities, neighborhood conditions may also prove to be a barrier for access to enough food. Certain areas in the U.S. may not provide full-service supermarkets or grocery stores. Furthermore, lack of transportation along with greater travel distances between households and supermarkets prove to hinder food security for people in those areas as well (Michele, 2009). Another consideration to take note of is that disabled adults may be at a higher risk for food insecurity since they have limited employment opportunities and may have health-care related expenses to take care of, reducing the income available for food.

2.2.4 Housing Instability

Housing instability can be best understood as a threat to a person’s housing security. It could also be explained as “the extent to which an individual's customary access to housing of reasonable quality is secure” (Frederick et al., 2014). Housing instability includes trouble paying rent, overcrowding, moving frequently, or spending too much on housing. These various traits are attributed to housing instability and can make it harder to get access to health care (Kushel et al., 2006). People with lower incomes may be forced to live in housing that is below the standard of reasonable living, which in turn could expose them to numerous health and safety risks while living there.

2.3 Education Access and Quality

A second main social determinant of health is education access and the quality of that education. Access to satisfactory education allows for people to pursue employment and the quality of that education can determine whether they can acquire a stable job with reasonable income. Therefore, in theory, aspects of economic stability and level of education should be positively correlated with one another, and as these are determinants of health, there may be a use to analyzing these statistics. Some specific examples of where education can have an impact on one’s health are language barriers or literacy issues, early childhood development, high school education, and enrollment in higher forms of education.

2.3.1 Educational Attainment

One attribute that the team concentrated on under education was the level of attainment that populations had by county in the United States. A high school education is commonly used as a general requirement for most jobs and leads to enrollment in higher education. When a person does not complete high school, there are a multitude of factors that can have disadvantageous effects on health including employment prospects, low wages, and poverty (High school graduation, 2022). Employment is a crucial aspect to maintaining a healthy life through a stable income, and since one’s education is associated with their level of employment, it is also a determinant that must be analyzed because it can impact health. Further illustrating the link between income and education, “students from low-income families often have less access to resources, and they tend to live in communities with underperforming schools” (High school graduation, 2022). Individuals who do not obtain a high school degree are more likely to self-report poorer health. Also, they more frequently report suffering from chronic conditions including asthma, diabetes, and heart disease over graduates of high school (High school graduation, 2022).

Enrollment in higher forms of education is another subset of education that should be analyzed for its effects on health outcomes. For example, research states that “graduation from college has a positive impact on employment options” and “the risk for underemployment or unemployment is higher for those with less education” (Enrollment in higher education, 2022). Specifically, as seen during the COVID-19 pandemic, people with more education were more likely to keep their job and work remotely, unlike some people that could not physically work from home (CEW Georgetown, 2022). Additionally, people with the income from these employment opportunities could indirectly improve their health by increasing their ability to gain resources that are linked to good health, such as high-quality housing. Lastly, individuals with more education are more likely to exercise, drink less alcohol, and seek preventative health care when necessary (Enrollment in higher education, 2022).

2.3.2 Language and Literacy

Language and literacy are also possible obstructions in establishing good health. Poor language skills and low literacy skills are connected to lower educational attainment and worse health outcomes (Language and literacy, 2022). Students who are unable to read proficiently "are especially unlikely to obtain a post-secondary degree" which could lead to adverse health outcomes (Foundation, the A. E. C., 2010). Specific segments of populations are more likely to have limited English language skills, with many of these segments comprised of families who do not speak English at home, immigrants, and individuals with lower levels of education. Language and literacy are especially important because inadequate English levels could become a restraint for individuals accessing health care services or understanding health care information (Language and literacy, 2022).

Those who indicate a limited English proficiency have also been, "less likely to have a usual place to go to when sick or have a preventative care visit in the past year" (Gulati & Hur, 2021). People with lower literacy skills also have problems, "following medication instructions, communicating with health care providers, and attaining health information," and all these effects can worsen health. Additionally, there are institutional barriers as well. The absence of well-trained interpreters and culturally qualified health care providers adversely affect the health of individuals with low literacy and limited English proficiency (Anderson et al., 2003).

2.4 Health Care Access and Quality

A third Social Determinant of Health is the access to health care and the quality of health care. Access to adequate health care and the quality of that health care gives people the opportunity to live a healthy life. Whether someone has access to health services, access to primary care, or even health literacy are examples of factors that may influence one's ability to have good health.

2.4.1 Access to Health Services

One of the largest barriers to health care access and thus a barrier to good health is inadequate health insurance (Access to health services, 2022). Without adequate health insurance, people must pay out of pocket for medical costs more frequently, which reduces their desire to go to health services even if they have a medical need. Usually, people with lower incomes are uninsured and can have a negative impact on their health. Uninsured individuals are "less likely to receive preventative services for chronic conditions such as diabetes, cancer, and cardiovascular disease" and kids are less likely to receive appropriate treatment for conditions like asthma (ibid.) Furthermore, a study showed that when people turn 65 and become eligible for Medicare, some "previously uninsured adults" began to use basic clinical services more often than when they were uninsured. Also, when Medicaid coverage is provided to previously uninsured adults, their chances of receiving a diabetes diagnosis increased. Another barrier that can reduce access to health services is the limited availability of resources. These resources could mean physicians in their respective area, poor transportation infrastructure, etc. Without the necessary resources to have sufficient health care, people can increase their risk of poor health outcomes.

2.4.2 Access to Primary Care

Additionally, there are geographical barriers to health services as well. Studies show that rural areas of the United States expose a lack of available medical services, including primary care physicians like family doctors or pediatricians (Douthit et al., 2015). For these service-deprived populations in rural areas, simply getting to a doctor may be an obstacle to accessing health care. People are less likely to travel to a doctor if the travel distance is too much of a burden, thus preventing them from getting sufficient care (ibid). Primary care providers can be a means to get care, early detection of diseases, chronic disease management, and preventative care. As such, people with usual sources of care can routinely get preventative services to positively affect their future health such as flu shots and cancer screenings. Likewise, as stated before, individuals without health insurance may delay seeking the care they need or neglect to have primary care, and this could lead to more hospitalizations for chronic conditions like diabetes (Access to primary care, 2022).

2.5 Social and Community Context

The final main group of Social Determinants of Health that the team researched encompasses social and community context. This SDoH is comprised of social aspects of the community that may influence a person’s health. The two main social aspects that the team looked at were ethnicity and citizenship status. Another part of this SDoH that may be worth researching is social cohesion. Social cohesion refers to “the strength of relationships and the sense of solidarity among members of a community” (Social Cohesion, 2022).

2.5.1 Ethnicity

Ethnicity is a population that is made up of people who share the same cultural background. A person's health can be greatly impacted based on their ethnicity. In the US, a study from 2019 shows that overall, many ethnicities such as Black and Hispanic people, "fared worse compared to White people across most examined measures of social determinants of health for which data were available" (Artiga et al., 2022). Asian and White people are more likely to attain a bachelor's degree or higher than many ethnicities which plays a crucial one's health as discussed earlier. White people also have the smallest percentage of people who have less than a high school education out of all the ethnicities looked at in the study. Hispanic and Asian individuals who are five and older are more likely to not be able to speak English very well (ibid). As explained earlier, not being able to communicate clearly impacts your health in a negative way. They are less likely to have a place to go to when they are sick, and it is a lot harder to communicate your health problems to a professional when you have trouble speaking English. The family median net worth for Black and Hispanic families was also much less than for White families. In 2019, Black families median net worth was $24,100 and Hispanic families was $36,050 while White families was $189,100 (Artiga et al., 2022). This is a significant difference, and it could have to do with people not working because of COVID but this could definitely have an impact on families being able to pay for hospitalizations and ER visits (ibid).

2.5.2 Citizenship Status

Citizenship status states whether an individual living in the US is a US citizen or not. The three ways a person can become a citizen is by being born within the territory of the US, being born to two US citizen parents, or by passing a citizenship test. There has not been a significant amount of research done on citizenship as a SDoH and most of it has to do with illegal immigrants which is a difficult population to get data from. There was research done on immigration policies and how they are known to discriminate (Hill et al., 2021). Also, there are people who receive work and student visas to come to the US. They are not citizens and therefore it is a lot harder for them to get medical insurance. The inability to quickly obtain medical insurance could lead to less hospitalizations if they do not want to be charged with expensive medical bills. Research showed that being a US citizen from birth does not tell us enough about a person because there are so many other SDoH that have a greater impact (Brou, 2019).

2.5.3 Social Cohesion

Social cohesion refers to the "connectedness" among people in a society (US Department of Health and Human Services, 2022). These positive relationships labeled as social cohesion are vital attributes of good physical and mental health. An indicator of social cohesion within a community is the amount of social capital they have, which refers to shared resources within a group. Social capital represents the positive product of human interaction, appearing in social network connections between individuals (Kenton, 2022). One study linked the four measures of social capital (perceived fairness, perceived helpfulness, group membership, and trust) and found that they were all connected to mortality. Although social networks spread social capital, they can also spread behaviors and health outcomes, known as "social contagion" (US Department of Health and Human Services, 2022). An example of a social contagion is if someone has a friend that smokes, then they are more likely to smoke, making the association between people negative or positive depending on others' attributes and actions. Opposite to these negative effects of socializing, there are also negative effects of social isolation. Social isolation is usually, "detrimental to health and increases mortality" (US Department of Health and Human Services, 2022). This is especially apparent with older adults since social contact decreases with old age. This phenomenon was significant during the COVID-19 pandemic when socializing in person became limited and many people had to isolate as a result. Since social isolation usually has negative impacts on health, there needs to be more research done regarding a community's social network and context.

2.6 Neighborhood and Built Environment

When determining factors of a population that influence their health, the environment around them can play a significant role as well. Concerns like crime and violence, and quality of the environment around them may be substantial to someone in that environment's health. Since this social determinant of health is tougher to measure for specific populations and may not be able to be matched one to one with the public data available, the team decided to omit this from their data analysis.

2.6.1 Environmental Conditions

Environmental conditions encompass a variety of characteristics that can have an impact on the daily lives of the people who live there. Water and air quality are two main aspects of the quality of the environment that can negatively influence a population's health. Populations with lower income tend to have poorer environmental conditions and thus a higher risk of exposure to health problems relating to these conditions. Water quality is an essential part of a healthy life especially since it is used for a multitude of reasons: bathing, drinking, or cleaning. Water can be contaminated through sewage leaks, pathogens, or chemicals, and studies show that the communities with lower incomes have higher risks of exposure to these contaminants. This is partly due to “city planning decisions, and the number of resources dedicated to managing the water system” in a particular region (US Department of Health and Human Services, 2022). Air quality is also a fundamental necessity to establishing good health. Things like dust, smoke, carbon monoxide, ozone, or nitrogen oxides are all air pollutants that can negatively affect regions in the world. Most of these pollutants are released from fires, vehicles, or industrial facilities, and some are linked to health problems like lung cancer and heart disease. Unlike most of the other Social Determinants of Health where living in an urban area positively impacts your health, urban areas are more likely to have worse air quality than their rural counterparts, especially urban areas with factories or industrial facilities which negatively impact your health (US Department of Health and Human Services, 2022).

2.6.2 Crime and Violence

Consistent with many of the other social determinants of health, crime and violence affect certain groups of people more often. “Low-income neighborhoods are more likely to be affected by crime and property crime than high-income neighborhoods” (US Department of Health and Human Services, 2022). On top of possibly experiencing physical trauma from violence, people can also experience mental distress and a reduced quality of life, which could then lead to other adverse health effects. Specific effects of the exposure to violence and crime include asthma, hypertension, stroke, cancer, and mental disorders (APHA, 2018). Additionally, “higher rates of neighborhood safety fears may lead to poorer self-rated physical and mental health,” which can be seen through a study that found a connection between gun-related crime and park usage (Han et al, 2018).

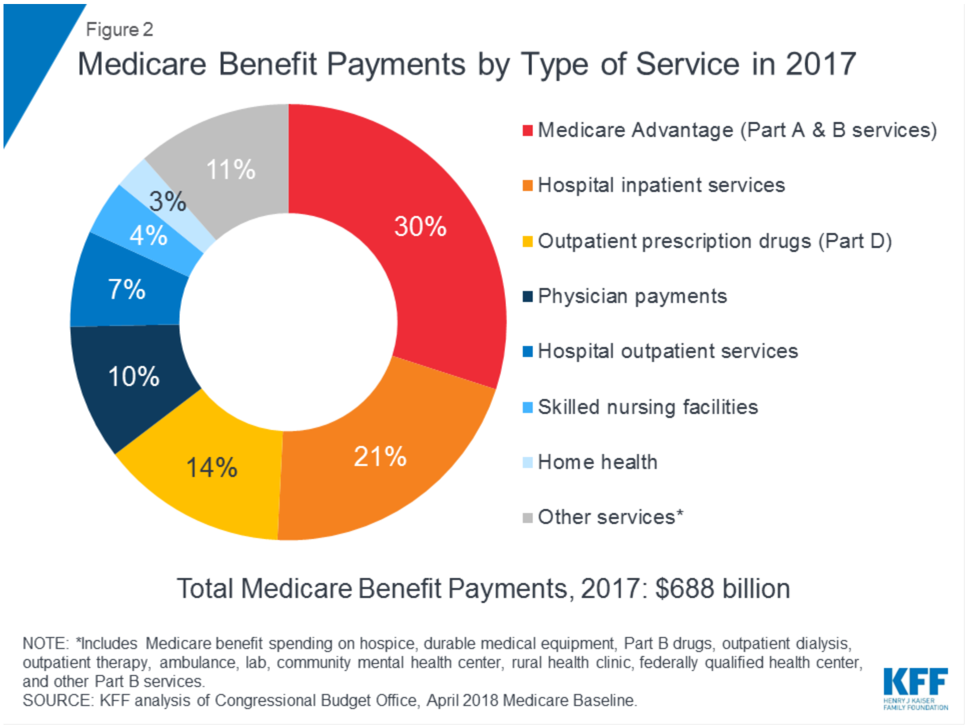
2.7 Medicare

President Lyndon B. Johnson signed Medicare into law on July 30, 1965, to ensure that senior citizens would receive health insurance (*Medicare Signed into* Law, 2019). Medicare is for people that are 65 and older, have certain disabilities, and people with end-stage renal disease. A person may buy Medicare to ensure that they have both health and financial security and so that basic health services are covered. Medicare consists of three distinct parts: Part A, Part B, and Part D.

Part A is hospital insurance, and it covers individuals staying at a hospital, nursing home, hospice care, and home health care. Most people pay for Part A before they start receiving benefits through Medicare taxes. If an individual pays the taxes for forty quarters or more (ten years) then they will not be charged premiums for Part A. If they pay from 30-39 quarters then their premiums will be $274 and for paying less than 30 quarters, premiums are $499.

Part B is medical insurance, and it covers specific types of doctors' services, outpatient care, medical supplies, and preventive services. Every individual is charged the same premium of $170.10 for Part B.

Part D covers prescription drug coverage which covers any prescription drugs that an individual may need which includes certain types of shots and vaccines (*An Overview of Medicare*, 2019).



Total Medicare Benefit Payments, 2017: $688 billion

NOTE: *Includes Medicare benefit spending on hospice, durable medical equipment, Part B drugs, outpatient dialysis, outpatient therapy, ambulance, lab, community mental health center, rural health clinic, federally qualified health center, and other Part B services.
SOURCE: KFF analysis of Congressional Budget Office, April 2018 Medicare Baseline.

Figure 2.5: Medicare Benefit Payment Percentages

2.8 U.S. Census

The US Census is one of the leading providers of quality data concerning the nation's people and economy. Their goal is to, "...provide the best mix of timeliness, relevancy, quality and cost for the data we collect and the services we provide." They collect data in several unique ways. The US Census uses the American Community Survey (ACS) for data on America's population, housing, and workforce. They utilize the Census of Governments for information on the outlook and quality of the nation's state and local government sector which includes public finance and employment as well as classifications. The Census uses the Decennial Census of Population and Housing to count every resident in the US and this takes place every 10 years. From the Economic Census they gather the official 5-year measures of American businesses which provides thorough statistics at national, state, and local levels of businesses. They also conduct their own surveys and programs that aim to provide periodic and comprehensive statistics about the nation (Bureau, 2022).

This wide variety of data is used for many different purposes. It is critical for government programs, policies, and decision-making. It also helps determine where to provide services for the elderly, build new roads and schools, and locate job training centers. The data contains information that says whether an individual qualifies for social security and other retirement benefits, passport applications, proving relationship in settling estates, and researching family history. The government uses the data to determine the distribution of Congressional seats to state. It is also used to distribute over $675 billion in federal funds to local, state, and tribal governments each year. This money goes towards providing states and communities with information about allocating funding for neighborhood improvements, public health, education, and transportation.

The team's main purpose of the US Census throughout the project was to collect data on certain nationwide statistics in various categories that included poverty, education, race, income, and employment. The team took advantage of the various filters that the Census provides and always selected US and all counties within the US and Puerto Rico because for the project every U.S. county made up the targeted area. The Census had data that was as current as 2020 when the team started the project. The group members used data from 2020 and 2019 when gathering information but two months into the project the Census came out with data from 2021 which they were then able to use later when using their models to try and predict what the 2021 data would look like (Bureau, 2022).

2.9 Milliman

The insurance world is a massive industry all over the world and many people often struggle to figure out which insurance plan is best for them. This struggle has led to an increase in the need for companies that can give expert advice on which insurance plan is best for specific needs. Milliman is one such company that is using their expert analysis to assist a plethora of organizations that need guidance. Their mission statement is, "Our expert guidance and advanced technology solutions empower leading insurers, healthcare organizations and employers to protect the health and financial well-being of people everywhere." Milliman is an independent risk management, benefits, and technology firm founded in 1947 that is comprised of actuaries, technologists, clinicians, economists, climate and data scientists, and benefits and compensation experts (Milliman, 2022).

Milliman works closely with their clients to assist them with both social and business challenges such as retirement, healthcare, effects from climate change or a pandemic, and the effect of low interest rates. Milliman continues to try and find new ways to ensure that they have access to as much data as possible and know how that data affects certain situations. A major area that they have been focusing on lately are social determinants of health. In October 2017, Milliman created an alliance with LexisNexis Risk Solutions, a corporation that sells data analytic products, to gain access to databases that consist of data on social determinants of health. Milliman has now been able to look through more than 400 attributes that deal with clinically validated SDoH. The LexisNexis databases contain information on a wide variety of topics including relatives and associates, assets, trends over time, education, and neighborhood and household characteristics. Milliman's focus with the new data from LexisNexis is to create identity profiles for people and then through a type of linking technology they can match a patient and member list that healthcare organizations send out to about 280 million people. After the healthcare organizations send out this list it gives the patient individual level socioeconomic attributes and scores. This identity profile is a major help to Milliman because it provides them with an accurate look at social determinants of an individual's health risk (LexisNexis Risk Solutions, 2017).

2.10 Generalized Linear Models

A linear model is used to explore associations between a response quantitative variable and an assortment of quantitative and categorical independent variables. If the association between the variables is strong enough, regression can be used to predict the mean value µ of an observed random variable Y (*Correlation and Regression with R*, 2016). The four assumptions for linear regression are:

1. Linearity of response variable
2. Homoscedasticity (constant variance of residuals)
3. Independence of predictions
4. Normality (E(Y|X=x) follows a normal distribution)

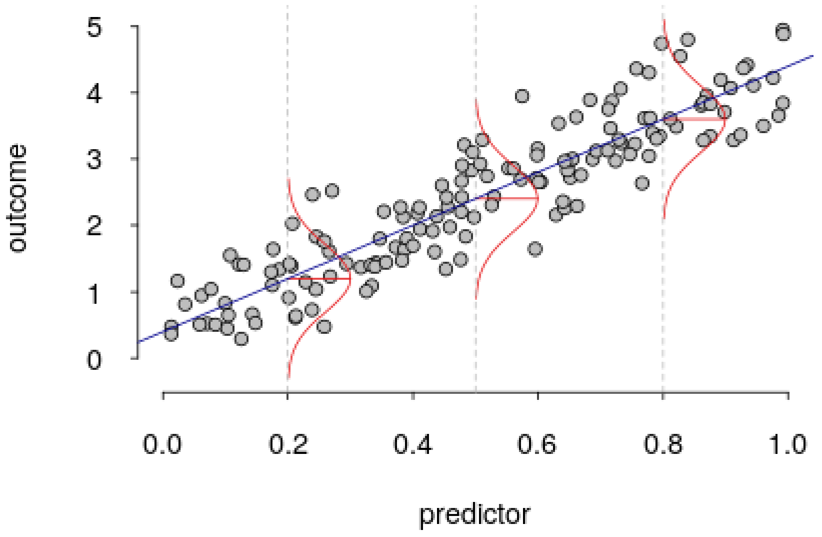


Figure 2.6: Scatter Plot with All Four Assumptions Met

Figure 2.6 is an example of a scatter plot where all four assumptions are appropriately met. It can be observed that the outcome variable Y is linearly associated with the predictor, the variance of residuals stays mostly constant, and residuals approximately follow a normal distribution. Although not explicit, it is assumed that the predictions are independent (Vanhove, 2019).

A simple linear model is not always appropriate depending on the nature of the response variable and the types of associations found. Generally, regression can be considered as estimating the mean or average of a random variable Y when considering additional effects. Another way to interpret regression is calculating E(Y|X=x) where the change in the conditional expectation is assumed to follow a certain function. Therefore, rather than using least squares, or attempting to minimize the sum of the squared residuals, a GLM fits an exponential family distribution to Y and calculates both the parameters and the coefficients for each X variable using maximum likelihood estimation. In many cases, the response variable is not linearly associated with its predictors and may have limitations on its domain. Consequently, a GLM can use a link function that transforms the response variable Y using a link function such that a linear model can be used. It is common to use a log link function for Poisson or negative binomial GLMs and a logit function for logistic regression. An example of a log link function would be as follows:

- $\log(\mu) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$
- $\mu = e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n}$

From these equations, it can be observed that the model can only predict positive values for μ (or λ), which is the expected value of the outcome variable Y that we are assuming follows a Poisson distribution. Additionally, the effects are multiplicative rather than additive, meaning that predictors predict percent change in μ rather than absolute change.

A GLM does not assume the variance of residuals is independent of the prediction for the mean of Y but is instead dependent on the chosen distribution and dispersion parameter. For example, when using Poisson regression, it is assumed that the variance of the prediction is equal to the prediction because the variance and expected value are equal in a Poisson distribution. However, in some cases, it makes sense to use a quasi-Poisson model which means changing the dispersion parameter from one to another value. Doing so changes the assumption that the variance and expected value are equal to the assumption that they are proportional. This change is especially useful when data is either over or under dispersed.

Therefore, a GLM is a more versatile tool than using only ordinary least squares with more flexible assumptions. For example, if the response variable is a count rather than a continuous variable, it would likely no longer have its mean follow a conditional normal distribution, but rather a Poisson or a negative binomial distribution. Changing the distribution that is fit to the mean alters how significant variables appear and reduces residuals. Below is an example where Poisson regression is appropriate because the values of Y are counts and there appears to be an exponential relationship between X and Y.

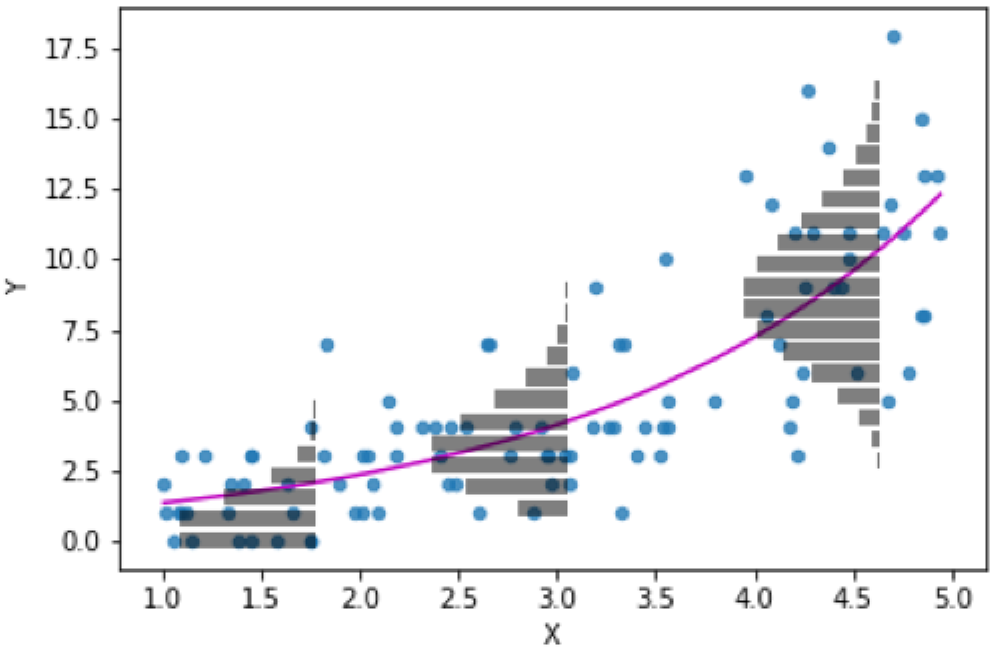


Figure 2.7: Scatterplot of Poisson Regression

In this scatterplot, it can be observed that the predictor and outcome variable are not linearly associated, the variance of residuals is no longer homogenous, and the response variable follows a mostly Poisson distribution with a lambda conditional on X (Turtureanu & Pananos, 1968).

Although GLMs are incredibly useful when dealing with more complex data, it can be harder to interpret how well they fit the data. With ordinary least squares, R^2 and the root-mean-square error are acceptable. When using GLMs, one of the main ways to compare the fits across different models using the same data is their **AIC** (Akaike Information Criterion) which can be defined as follows:

- $AIC = 2(p - \log(L))$

In this equation p is the number of predictors and L is the maximum value of the likelihood function. AIC is utilized to compare models such that each predictor should at least increase the log(L) by 1 if AIC is to be minimized.

When comparing models, any model with more predictors will have a higher maximum likelihood function value because it will always at least fit the data just as well when using 0 as a coefficient. Therefore, AIC penalizes the model for each predictor it adds, so if a predictor changes the log(L) by less than 1, then that predictor should not be used. However, only using AIC can cause overfitting, especially when there are many predictors. Therefore, a Likelihood Ratio test can be used to test if a reduced model has a significantly lower log likelihood using a chi-squared test statistic with degrees of freedom equal to the difference of predictors in the two models. If the null hypothesis of the two models having the same log likelihood fails to be rejected, then the reduced model can be used. Both AIC and the Likelihood Ratio Test can be used to minimize the number of predictors in a model. However, when comparing models that use different data, AIC can no longer be used. Instead, we can compare standardized residuals by comparing the differences in their coefficients of variations when just using the mean versus using the conditional mean given by the model for each prediction. When comparing model diagnostics for logistic regression models, it is better to instead use a confusion matrix which determines how many observations were correctly and incorrectly classified. For most applications, accuracy is the most important metric, or the total number of correctly classified observations divided by the total number of observations, followed by sensitivity and specificity.

Chapter 3: Methodology

The purpose of the project was to provide insight into social determinants of health and their possible effects on certain populations' health. To sufficiently complete this task, data had to be collected and cleaned to only show the most vital variables attributed to social determinants of health and population health measures. The public data found had a fair share of limitations, especially with some counties missing data, and this was dealt with as well as possible. It was important that the various data sources gathered represented the same populations to stay consistent during the analysis stage. After gathering and cleaning the public data, it was possible to formulate some generalizations about the interactions between SDoH and population health measures.

3.1 Data Collection

One of the main components of the project was data collection. To figure out the implications of SDoH and their implications on specific health outcomes it was necessary to gather mass amounts of data that may have relation to the project. The team collected this data through two main online databases. The main database utilized for health outcomes was the Centers for Medicare and Medicaid Services (CMS). The CMS had many filters that the team used to make sure that the data was recent and had the necessary measures for the project.

The CMS filters that were used include:

- Year: 2021, 2020, 2019
- Geography: County
- Measure: Average Total Cost, Average Principal Cost, Hospitalization, Emergency Department Visit Rate, Preventive Services, Readmissions
- Adjustment: Unsmoothed actual
- Analysis: Base measure
- Domain: Primary chronic conditions
- Condition/service: Asthma, Depression, Diabetes, Heart failure, COPD
- Sex: All
- Age: All
- Dual eligible: Medicare only
- Medicare eligibility: Old Age/Survivor's Insurance

The people within the 65 and older age group were the focal demographic for the project. The bottom two filters, Dual eligible and Medicare eligibility, were important because it selected people 65 and older with Medicare and omitted people with disabilities under 65. Electing "Medicare only" ensured that there were only people with Medicare and Old Age/Survivor's Insurance, which consists of mostly people aged 65 and older. For each condition, the team selected all the necessary filters that remained the same and then had to change the measure for each dataset so that they could collect data on every measure for every condition. The group did this for both 2019, 2020, and 2021 so they could investigate the year-to-year trends and see the impact that COVID-19 had on the data.

The second main online database the team used was the US Census which was mainly for data on SDoH. Each team member searched for data on different SDoH. The group also utilized filters when using the US Census which always included the United States as a location and all counties within the US and Puerto Rico. Then each member found important data on the three SDoH areas that they found the most important. This included data on employment and poverty status, people's highest level of education and their language and literacy skills, and individual's access to health services and primary care. Once data was found and downloaded as a csv file, the team had to search for the data that was relevant to the target population. The target population for the project was people 65 and older since those are the individuals that have access to Medicare. Most people that are 65 or older have Medicare and the individuals who are 65 and older without Medicare are so little that in the US Census data it is negligible. Many of the tables from the US Census have information on many specific age ranges so the group utilized Python to cut the US Census data so that the only data that was showing was data for people 65 and older.

The only other database that the team used was the Centers for Disease Control and Prevention (CDC). There was one table that was used from the CDC which had to do with the COVID-19 death counts by county in 2020. This helped the team look closer into if COVID drastically changed any SDoH statistics.

3.2 Data Cleaning

After the data collection process, the next step was data formatting and merging. The team found early on that for many tables there were counties that had no information for certain Medicare data. For example, 3215 counties had information on "All Emergency Department Visits" but only 3009 counties had information on "Average Principal Cost" so to be able to use both simultaneously, the 3009 common counties were used and the other 206 were not. The group members found that many of the counties that were missing large portions of data were from Puerto Rico and there were only a few that were from the US. Fortunately, all 3009 counties with data on “Average Principal Cost” also had data on every other category of the Medicare data so those 3009 counties could be used when creating models and merging data later.

The next step in the data cleaning process was to combine all the data from every table that had been collected into one table for efficient use. Each health statistic had its own csv file and columns, so the county and state columns were used as look up keys to merge the tables on. The team loaded the Medicare data from the CMS into Jupyter Notebook and then merged the data using the pandas package within Python. Next, the group edited the US Census data which became a bigger challenge than the Medicare data. Some of the county's names in the US Census had accent marks in their names which did not match with the Medicare data. Thus, to merge the Medicare and Census data, group members needed to remove the accent marks so that they would match the Medicare data. Lastly, the CDC data state names were abbreviated which also did not match with the Medicare data, so the team transformed that column into the right format. After these necessary changes, the Medicare and US Census data was merged into one table so that it was easier to analyze trends and correlations between variables. Another issue was that some of the data was in counts rather than percentages, which offered less meaning. To fix this problem every count was turned into a percentage so that it was easier to compare each county with each other.

Another big part of data cleaning was creating “buckets” that would act as variables. There was a lot of data that by itself wasn’t a variable or there were too many variables in a certain area, so they needed to be grouped together. For example, there are over 3000 counties in the US, and it wouldn’t make sense to have each county be its own variable so the group created regions that the counties would be allocated into. These regions included:

- Northeast
- Midwest
- Southeast
- Southwest
- West

Once the modeling process started it was easier to run the models and figure out which variables showed significance because of the various buckets for certain variables.

3.3 Modeling

The team explored many different models to see which variables correlate with each other and it was decided to create a correlation matrix of the data due to many aliased and colinear variables. The matrix was necessary to learn which variables had a high correlation with each other. There were certain instances where two variables had a correlation close to 1. Most of the time it was because one variable almost always occurred with another variable. When this happened, one of the variables needed to be removed so that when the team ran various models in R, it would not result in any issues with the output.

The group used various distributions in GLM such as negative binomial and Poisson distributions and then compared the AIC to choose which type of model would be the best fit for the data. This process was run for both a model with only main effects and a model with both main effects and interaction terms as the covariates. Main effects are predictors that only consider their own individual impact on the response variable. In other words, when retaining the same values for all the other predictors, a main effect will convey the average change for each unit change. An interaction term, however, combines multiple effects by multiplying predictors together which will further change the average change when both predictors are increased or decreased.

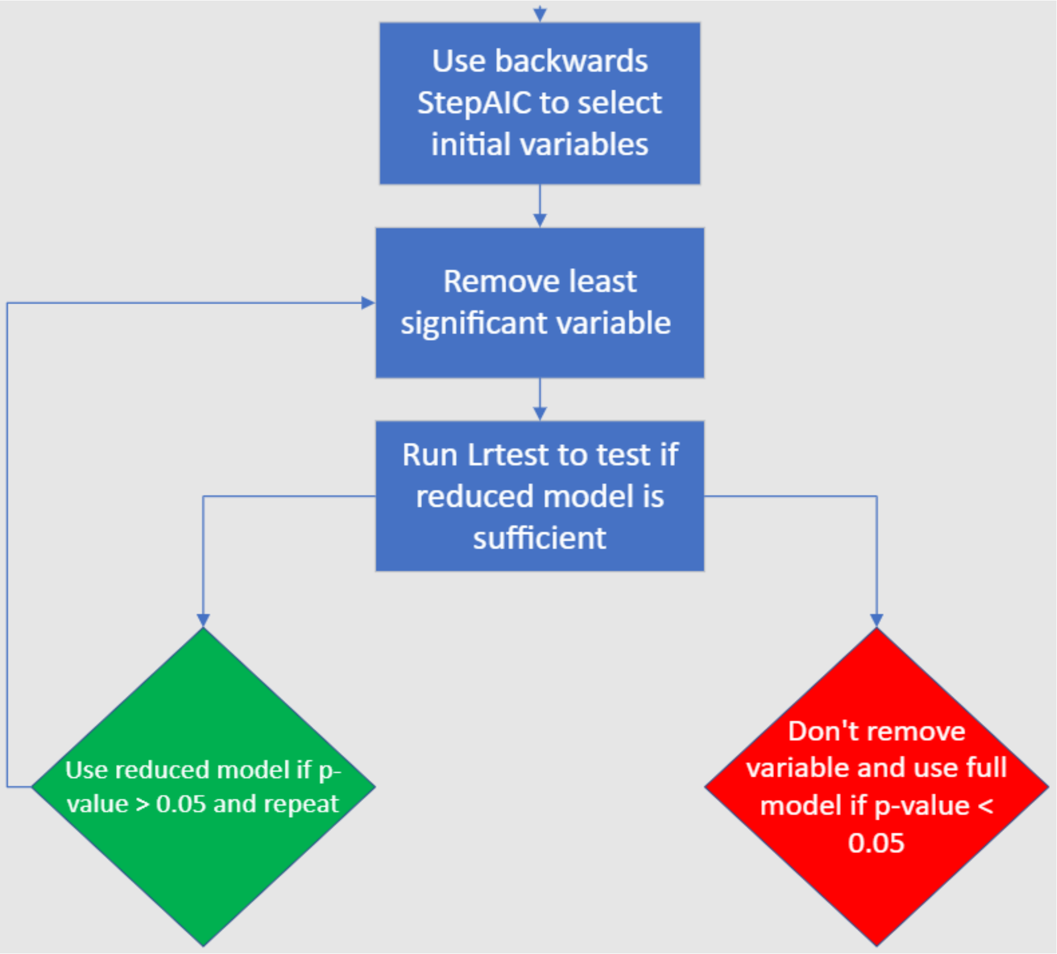

Use backwards StepAIC to select initial variables
Remove least significant variable
Run Lrtest to test if reduced model is sufficient
Use reduced model if p-value > 0.05 and repeat
Don't remove variable and use full model if p-value < 0.05

Figure 3.8: Model Creation Process

Team members used variations of the step AIC function in R to remove the least significant variables, so that they could better understand which variables were the most important. Another function used was the Likelihood-Ratio test, and the goal of this function was to see if the reduced models were acceptable after the removal of additional variables after the step AIC function. To determine whether the reduced model was sufficient or not, the p-value of the Likelihood-Ratio test was used. If the p-value of the test was greater than 0.05, the team would use the reduced model and repeat removing the least significant variable until the test returned a p-value less than 0.05. Once the p-value was less than 0.05, that model was selected to be used. This process was done for each health statistic. After the team knew that the reduced models were acceptable, they selected the final variables from each health statistic.

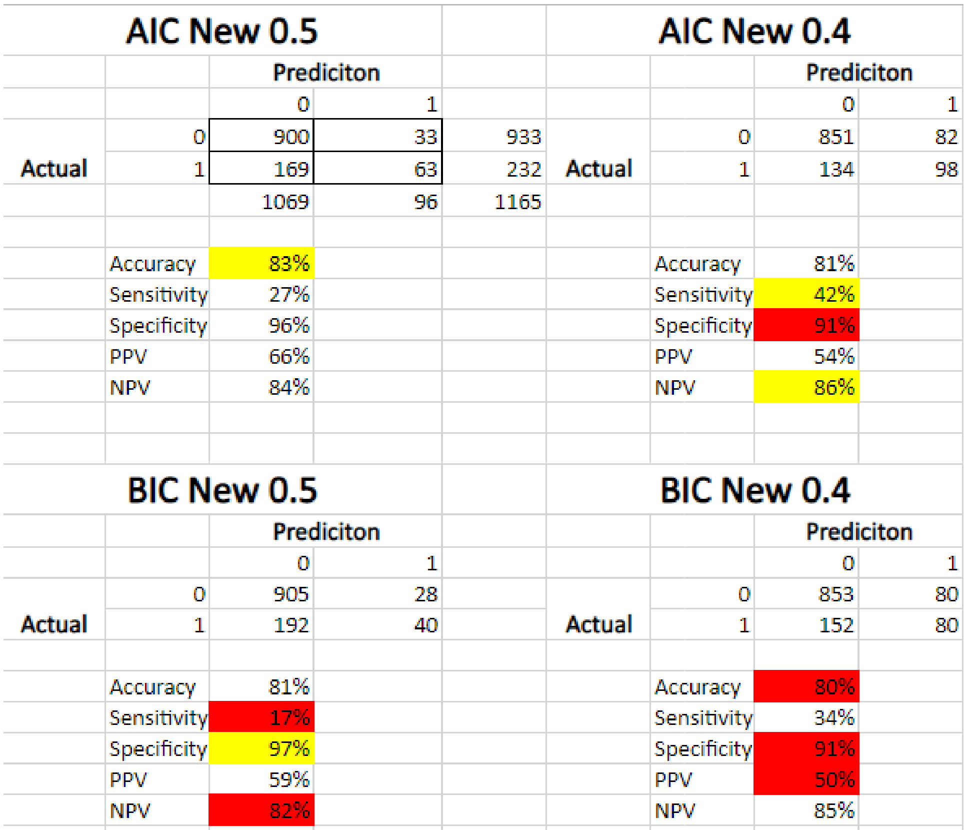
After talks with Milliman, they decided that the areas that interested them the most were counties with an elevated risk of high hospitalizations and emergency room visit rate. The team then made logistic regression models to determine which variables give the most information for determining high hospitalizations and emergency room visits and their associated estimated probabilities. Group members utilized step AIC and step BIC functions to look at the models in a variety of ways including:

- Main effects
- Main effects and interactions
- Main effects, interactions, and transformed variables (ex. x^2, $x^{\frac{1}{2}}$, $\log(x)$)

The team then used the 2020 CMS data as training for the models created. During the project the CMS released data from 2021. This allowed the group members to use the 2021 data on the models for testing in which they used the 80th percentile for emergency room visits and hospitalizations separately for 2020 and 2021 data. This came out to about 400 hospitalizations for 2020 and 500 for 2021.

3.3.1 Model Evaluation

The team decided to evaluate the models they created using confusion matrices. The confusion matrices the group members created showed how many observations/counties the model correctly identified as having high and low (non-high) hospitalizations, how many observations the model incorrectly identified as having high hospitalizations when it was low and vice versa.



AIC New 0.5

		Prediciton		
		0	1	
Actual	0	900	33	933
	1	169	63	232
		1069	96	1165
	Accuracy	83%		
	Sensitivity	27%		
	Specificity	96%		
	PPV	66%		
	NPV	84%		

AIC New 0.4

		Prediciton	
		0	1
Actual	0	851	82
	1	134	98
	Accuracy	81%	
	Sensitivity	42%	
	Specificity	91%	
	PPV	54%	
	NPV	86%	

BIC New 0.5

		Prediciton	
		0	1
Actual	0	905	28
	1	192	40
	Accuracy	81%	
	Sensitivity	17%	
	Specificity	97%	
	PPV	59%	
	NPV	82%	

BIC New 0.4

		Prediciton	
		0	1
Actual	0	853	80
	1	152	80
	Accuracy	80%	
	Sensitivity	34%	
	Specificity	91%	
	PPV	50%	
	NPV	85%	

Figure 3.9: Evaluation of Confusion Matrices

This process was repeated on the various models the team created until they determined the model with the highest accuracy, sensitivity, and specificity when testing on 2021 data. Accuracy is the percentage of how many times the model correctly indicates whether a county is high or low for hospitalizations. Sensitivity measures the percentage of counties that are high that the model correctly predicted. Specificity is the percentage of counties that are not high that the model correctly predicted. The team then combined the 2020 and 2021 data and used the same covariates to create the final model.

3.4 Power BI

Power BI is a Microsoft data visualization software that the team used to display the discoveries found in the data collected. Due to the vast amount of work with data cleaning to compile the data, it was vital to show results in a simplistic manner such that the team and eventually the sponsor could easily draw conclusions. After the team determined which variables were important for each health outcome through correlation analysis and modeling, those variables were used as focal points on each health outcome page.

There were two types of dashboards that were used for analysis in PowerBI. One of the dashboards had sliders to change the significant SDoH variables selected, and the other analyzed counties that were above or below a certain percentile of outcomes in the U.S. The two types of dashboards are shown in Figure 3.3 and Figure 3.4. The “slider dashboard” depicted how SDoH filters may impact the average outcome and the “percentile dashboard” showed changes in average SDoH variables as varying percentiles were used as inputs. As the focal point of the data visualization reduced to two health outcomes, all cause hospitalizations and all emergency department visits, the team made a slider dashboard and a percentile dashboard for both. These health outcomes had the most variability in the data collected and the team thought that they could be integrated well within PowerBI. Descriptions of both types of dashboards are below, using hospitalizations as the default health outcome.

3.4.1 Slider Dashboards

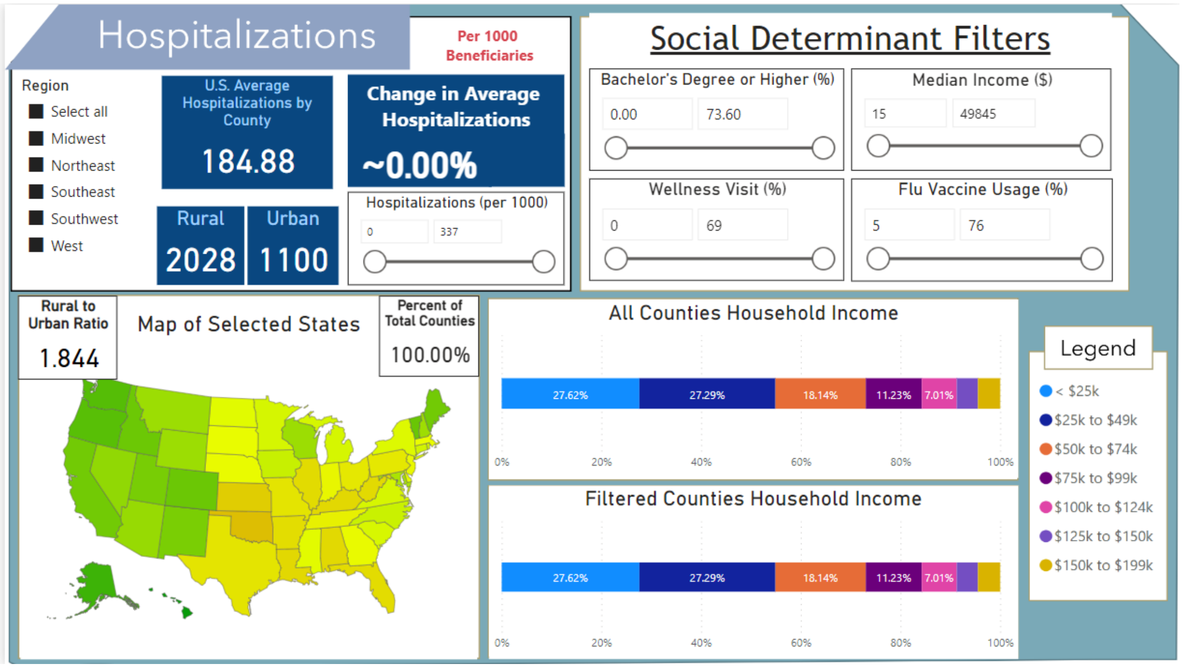
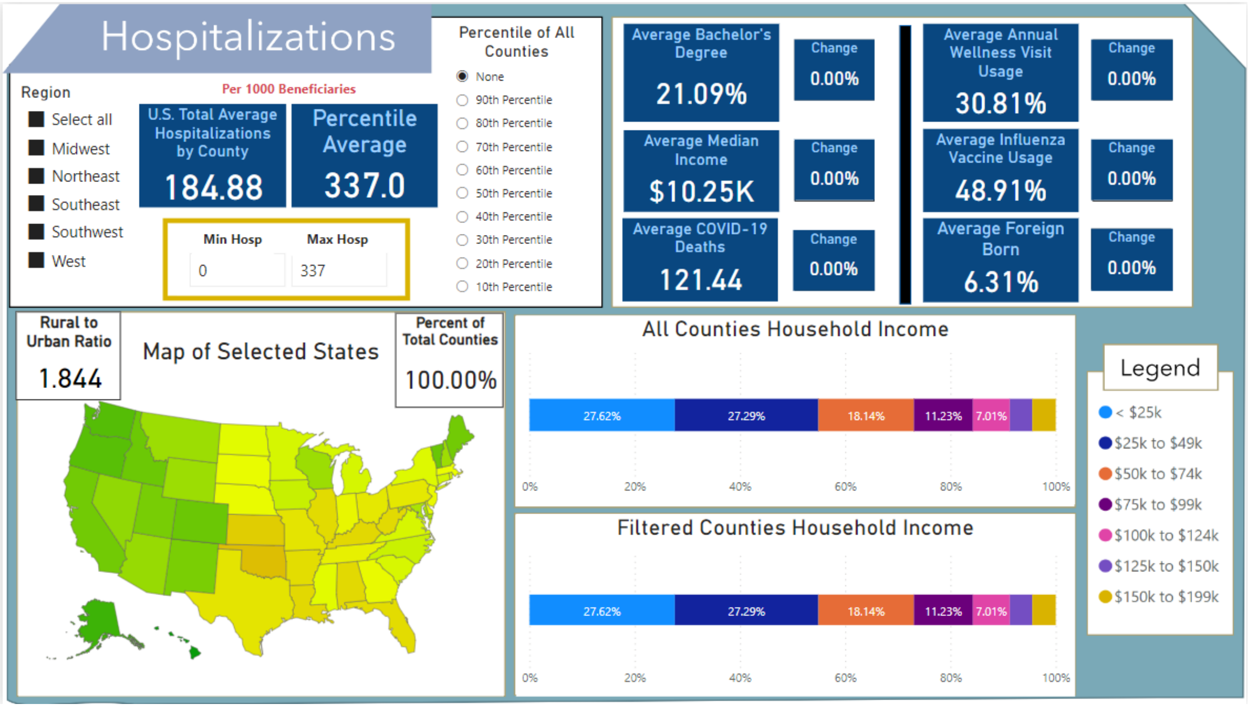


Figure 3.10: Hospitalizations Slider Dashboard

Each slider dashboard has a total U.S. average hospitalizations or emergency department visits in a big blue box in the upper left quadrant. Also in the upper left quadrant, there is another blue box that shows the change in average hospitalizations or emergency department visits from the total U.S. average after the sliders are manipulated. The counts of urban and rural counties were also included in this area to give a general idea of how many counties were selected. The sliders in the upper quadrants of the dashboard manipulate the counties that are filtered in and out of the total dataset. The variables that were used as "sliders" in the dashboard were: percentage with a bachelor's degree or higher, median income, annual wellness visit usage (%), and influenza vaccine usage (%). These variables were consistent with the rest of the project as they were composed of people that were 65 and older at the county level. The sliders allowed the team to only include counties in the dashboard with certain conditions, in which the team could then speculate what influenced the unfavorable or favorable outcomes seen. The bottom left quadrant showed the rural to urban county ratio, as well as the percentage of the total counties in the U.S. being filtered. The map was colored by average outcome per state, with lower average outcomes in a shade of green and higher average outcomes in red. Since mostly all the health outcomes analyzed were influenced significantly by household income brackets, the team decided to include those on each page to see how the breakdown of household income changed as other variables changed. Therefore, the lower right quadrant had a breakdown of the initial household income brackets of the whole U.S., and then the chart below it showed the difference in household income brackets as the SDoH sliders change.

3.4.2 Percentile Dashboards



Hospitalizations

Region
■ Select all
■ Midwest
■ Northeast
■ Southeast
■ Southwest
■ West

Per 1000 Beneficiaries

U.S. Total Average Hospitalizations by County: 184.88

Percentile Average: 337.0

Min Hosp: 0
Max Hosp: 337

Percentile of All Counties
● None
○ 90th Percentile
○ 80th Percentile
○ 70th Percentile
○ 60th Percentile
○ 50th Percentile
○ 40th Percentile
○ 30th Percentile
○ 20th Percentile
○ 10th Percentile

Average Bachelor's Degree	21.09%	Change	0.00%
Average Median Income	$10.25K	Change	0.00%
Average COVID-19 Deaths	121.44	Change	0.00%
Average Annual Wellness Visit Usage	30.81%	Change	0.00%
Average Influenza Vaccine Usage	48.91%	Change	0.00%
Average Foreign Born	6.31%	Change	0.00%

Rural to Urban Ratio: 1.844

Map of Selected States

Percent of Total Counties: 100.00%

Figure 3.11: Hospitalizations Percentile Dashboard

In the upper left quadrant of the percentile dashboards, a total U.S. average of hospitalizations or emergency department visits in a big blue box is shown. However, instead of the sliders in the previous dashboard, a percentile selection is shown. When a certain percentile was selected, the “Percentile Average” box would change, and the user then had to change the “Min Hosp” or “Max Hosp” in the gold box to determine what is shown on the page. The upper right quadrant had numerous SDoH conditional averages that depended on the “Min Hosp” and “Max Hosp” selected and they also included the change in that SDoH from the total dataset. The blue boxes shown had averages for these variables: percentage with a bachelor’s degree or higher, median income, annual wellness visit usage (%), influenza vaccine usage (%), COVID-19 deaths, foreign born (%). Most of variables were consistent with the rest of the project as they were composed of people that were 65 and older at the county level. However, the variable of COVID-19 deaths was the number of people who had died from COVID-19 in those counties and included all ages. The team decided to leave this variable in the dashboard because the data was from 2020, and the potential effects of COVID-19 should be acknowledged for that year.

Like the slider dashboard template, the bottom left quadrant displays the rural to urban county ratio, as well as the percentage of the total counties in the U.S. being filtered. The map was colored by average outcome per state, with lower average outcomes in a shade of green and higher average outcomes in red. Since mostly all the health outcomes analyzed were influenced significantly by household income brackets, the team decided to include those on each page to see how the breakdown of household income changed as other variables changed. Therefore, the lower right quadrant again had a breakdown of the initial household income brackets of the whole U.S., and then the chart below it showed the difference in household income brackets as the selected percentile changed.

Chapter 4: Results

The first set of results the group members found from the various models they created were the most significant variables when it came to predicting the number of hospitalizations and emergency room visit rates. The team created a variety of GLMs to make predictions and after many rounds of testing they were able to select a final model that was the most effective when predicting if a county in the US had high hospitalizations or not as well as high emergency room visit rates. The model predicted which counties would have an outcome worse than 80% of other counties in 2021 for the two categories being looked at using the 2020 data and confusion matrices were created for each. The most important deliverable was the PowerBI dashboards. These dashboards, using sliders and percentiles, showed how different variables affected the number of hospitalizations and emergency rate visits.

4.1 Most Significant Variables

Throughout the whole project, there were many variables that the team had to look at to see if they were significant or not. During the process, many variables were removed because they were deemed as having little significance in the presence of other variables when determining what affects hospitalization rates and emergency room visit rates. From the models that the team created, they were able to identify the most significant variables which were:

- Percent of county with household incomes less than $25,000 a year
- Percent of county having a bachelor's degree or higher
- Counties located in the West region of the United States

The group members selected these three variables because they were closely related to many of the other predictor variables, implying that those variables provided the most information compared to other variables. This discovery is useful because if you know a region's trend of highest education level attained, then it will give insight on the region's average income, poverty level, and whether it is more urban or rural. The team also used the logistic model that they created to find significant variables, narrowing down on variables specific to high hospitalizations and emergency rate visits.

From the logistic models, the group was able to identify three more significant variables which were:

- Urban counties
- Percent of county having a bachelor's degree or higher
- Percent of county that speaks a language other than English

The percentage of a county's bachelor's degree or higher population was the only variable that showed up in both the final model and the logistic models. That is not to say that it is the most significant variable, but it clearly shows its importance in determining a county's hospitalization and emergency room visit rates.

When the team investigated the most significant variables with more conditions, they were able to find even more information on their predictive power. Counties that are below the regionwide 25th percentile for having a bachelor's degree tend to have more hospitalizations and ER visits than the counties that are above the regionwide 75th percentile as shown in the graphs below.

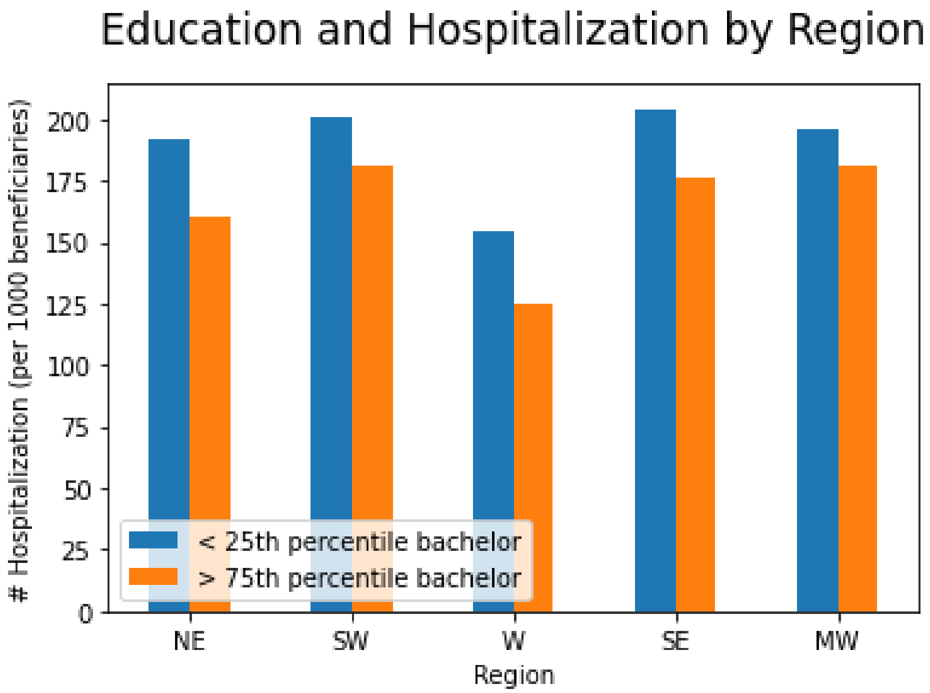


Figure 4.12: Bachelor's Degree Percentile to Hospitalization Rate by Region

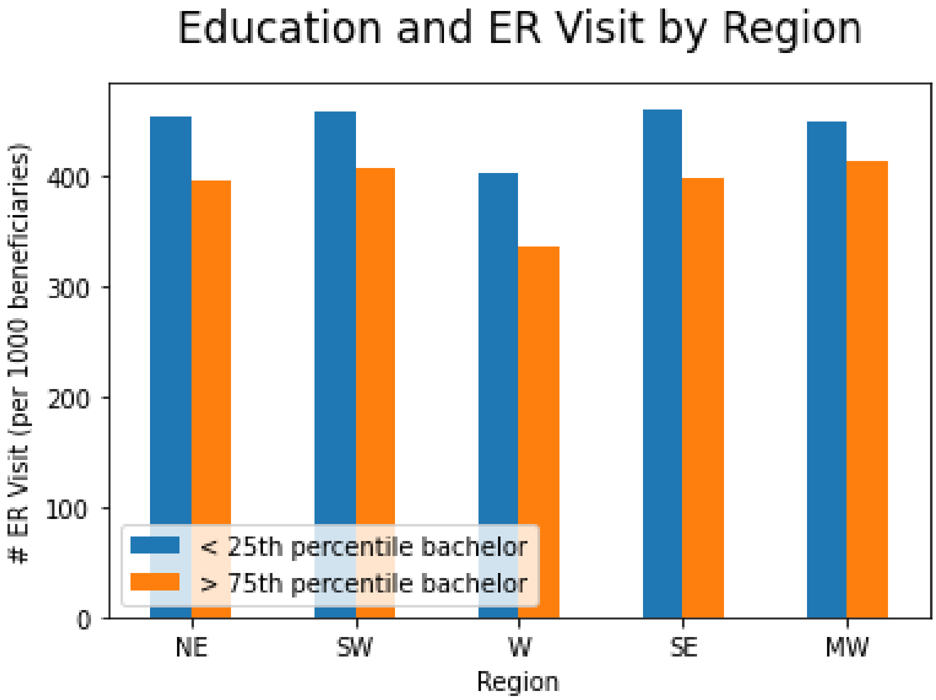
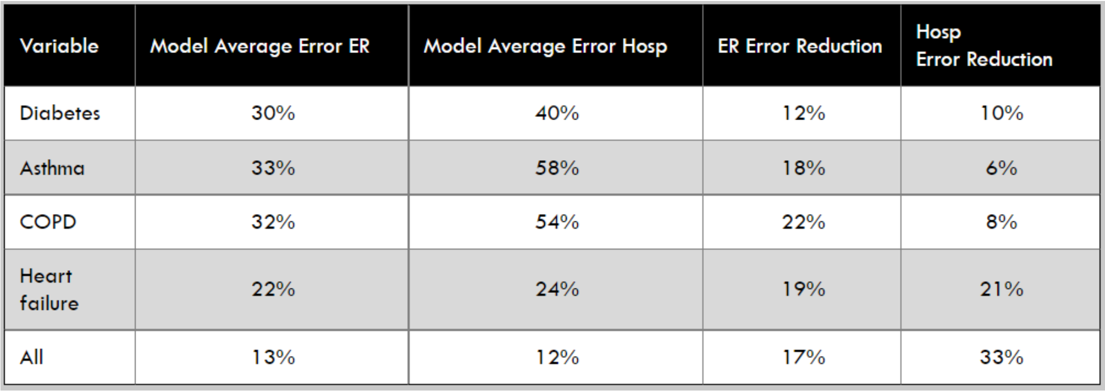


Figure 4.13: Bachelor’s Degree Percentile to ER Visit Rate by Region

The group members also discovered a trend between rural and urban hospitalizations in the southeast and having an income of less than 25 thousand per year. For urban counties in the southeast, hospitalizations gradually increased as the percentile of people making less than 25 thousand per year increased. For rural counties in the southeast, the number of hospitalizations peaked at the 50th percentile and declines while it goes towards both the 25th percentile and the 75th percentile.

4.2 Final Models

The team prioritized the logistic models, creating several different models using different step functions and transformations of covariates. Overall, because hospitalizations have a smaller right tail, the model was simpler and had a lower true positive rate. When combining 2020 and 2021 data, most of the variables remained significant, implying that the observed associations are more likely to remain in future years. The group also observed that SDoH added the most information when predicting all hospitalizations. The models' average errors and error reductions can be seen in figure 4.14.



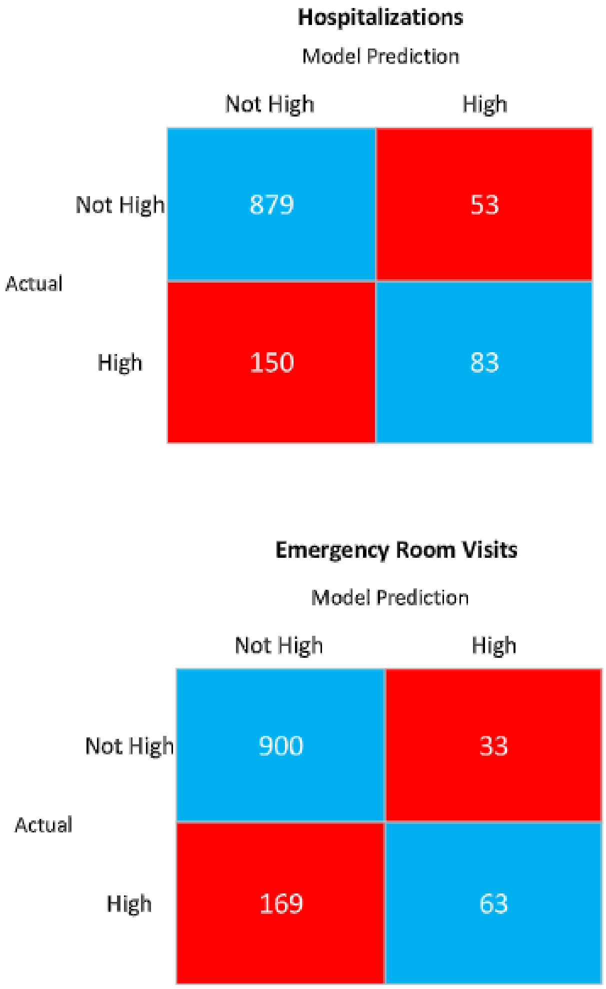
Variable	Model Average Error ER	Model Average Error Hosp	ER Error Reduction	Hosp Error Reduction
Diabetes	30%	40%	12%	10%
Asthma	33%	58%	18%	6%
COPD	32%	54%	22%	8%
Heart failure	22%	24%	19%	21%
All	13%	12%	17%	33%

Figure 4.14: Model Statistics

It is also evident from the table that ER visit statistics were easier to predict for all the other metrics in the table expect “All”. With this information the team concluded that SDoH are better at predicting more specific ER visit reasons than specific hospitalization reasons. However, SDoH have more predictive power when it comes to predicting all hospitalizations than all ER visits.

4.3 Confusion Matrix Results

The team used their final models to create confusion matrices for predicting if counties had high hospitalizations or not and if they had high emergency room visit rates. The group members used the 2020 US Census data to create their models and then tried to predict whether a county from the 2021 Census data would be high or not. The results from the model are shown in the charts below.



Hospitalizations

Model Prediction

		Not High	High
Actual	Not High	879	53
	High	150	83

Emergency Room Visits

Model Prediction

		Not High	High
Actual	Not High	900	33
	High	169	63

Figure 4.15: Confusion Matrices Results for Predicting Hospitalizations and ER Visits

There are many different values that can be found from the data represented by this matrix. The five main values that the team looked at were accuracy, sensitivity, specificity, positive PTPR), and true negative rate (TNR). The hospitalization version of these values was determined using these formulas:

- Accuracy: $\frac{(879+150)}{1165} = 0.88$
- Sensitivity: $\frac{83}{(83+150)} = 0.36$
- Specificity: $\frac{879}{(879+53)} = 0.94$
- PPV: $\frac{83}{(53+83)} = 0.61$
- NPV: $\frac{879}{(879+150)} = 0.85$

The ER version of these values are calculated by the formulas shown below:

- Accuracy: $\frac{(900+63)}{1165} = 0.83$
- Sensitivity: $\frac{63}{(63+169)} = 0.27$
- Specificity: $\frac{900}{(900+33)} = 0.96$
- PPV: $\frac{63}{(33+63)} = 0.65$
- NPV: $\frac{900}{(900+169)} = 0.84$

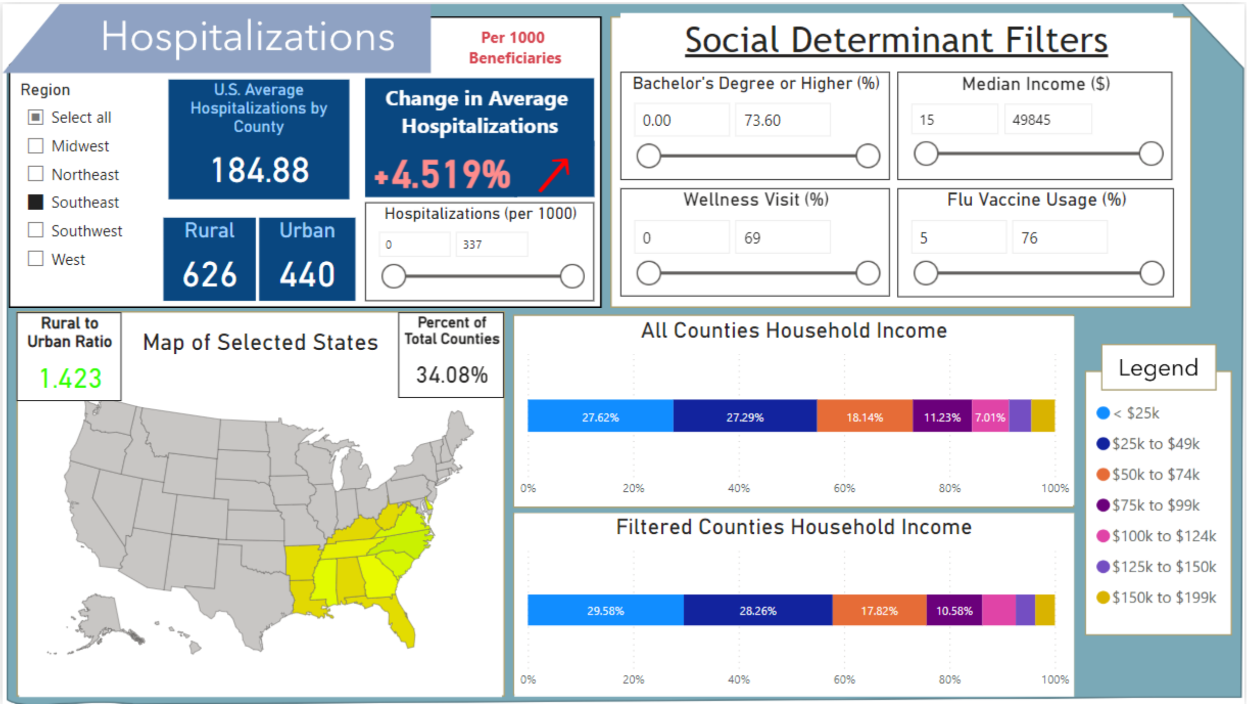
The two main values that the group looked at when deciding on a final model were accuracy and sensitivity because accuracy measures how many times the model correctly guesses if a county is high or not high and sensitivity measures the number of times the model correctly guesses that a county is high out of all the counties that were truly high. The goal of the model was to maximize accuracy and sensitivity. The sensitivity is significantly lower than all the other values but that is to be expected because the probability of a county being recognized as a county with high hospitalizations is 20% when there is no other additional information. The team attempted to change the cutoff probabilities to increase the sensitivity, but both the specificity and accuracy would decrease too much.

4.4 PowerBI Findings

PowerBI was a useful tool to visualize the general discoveries of the project and to observe the changes that happened in the data when different filters were applied. After the team discovered the most significant variables that were influential in the models, the PowerBI dashboard allowed for further exploration and analysis of the effects that those variables had on the data. The discoveries found in PowerBI for hospitalizations and emergency department visit rates are outlined below, with nuance from the slider dashboards and the percentile dashboards.

4.4.1 Hospitalizations – Slider Dashboard

As a first step, an investigation of average SDoH and hospitalizations by region was done by filtering out one region at a time. The regions with average hospitalizations above the national average were Midwest, Southeast, and Southwest. On the other hand, the Northeast and West regions had average hospitalization rates well below the national average. The West region had the best average hospitalization rate with a rate about 25% below the national rate.



Hospitalizations

Per 1000 Beneficiaries

Region

☒ Select all
□ Midwest
□ Northeast
■ Southeast
□ Southwest
□ West

U.S. Average Hospitalizations by County: 184.88

Change in Average Hospitalizations: +4.519%

Rural: 626

Urban: 440

Hospitalizations (per 1000): 0 – 337

Social Determinant Filters

Bachelor's Degree or Higher (%): 0.00 – 73.60

Median Income ($): 15 – 49845

Wellness Visit (%): 0 – 69

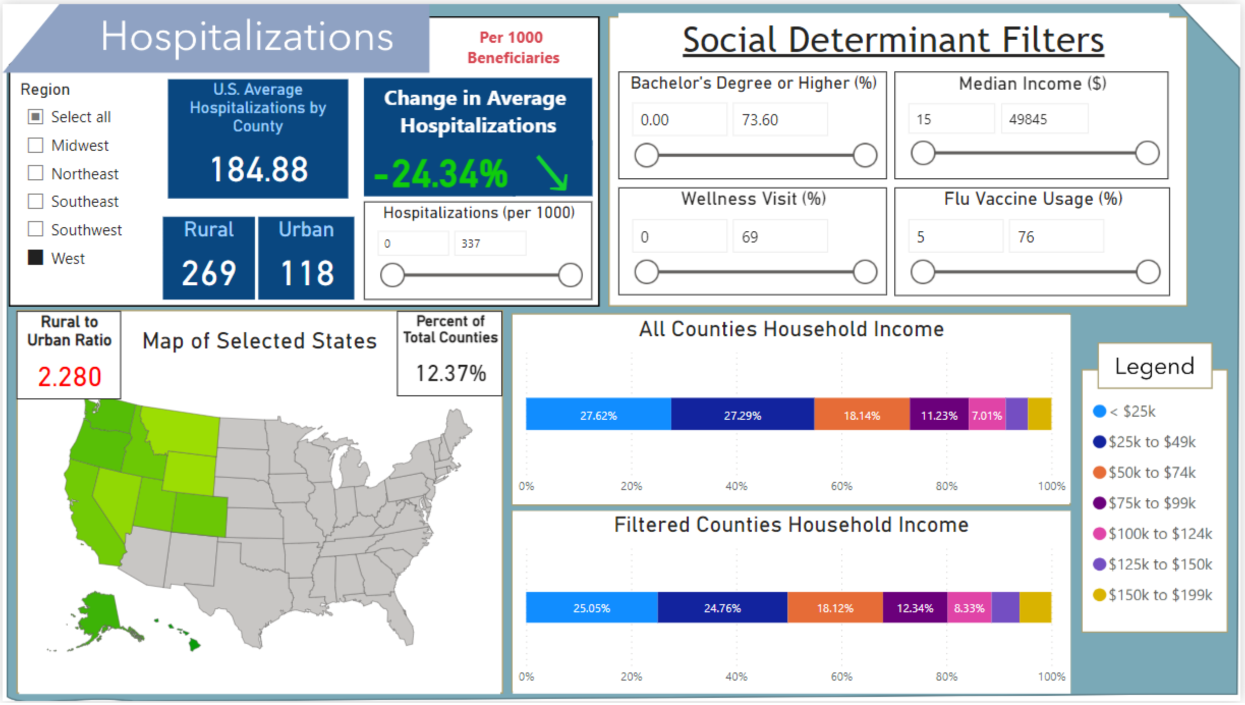
Flu Vaccine Usage (%): 5 – 76

Rural to Urban Ratio: 1.423

Map of Selected States

Percent of Total Counties: 34.08%

Figure 4.16: Hospitalizations Southeast Region



Hospitalizations

Per 1000 Beneficiaries

Region

☒ Select all
□ Midwest
□ Northeast
□ Southeast
□ Southwest
■ West

U.S. Average Hospitalizations by County: 184.88

Change in Average Hospitalizations: -24.34%

Rural: 269

Urban: 118

Hospitalizations (per 1000): 0 – 337

Social Determinant Filters

Bachelor's Degree or Higher (%): 0.00 – 73.60

Median Income ($): 15 – 49845

Wellness Visit (%): 0 – 69

Flu Vaccine Usage (%): 5 – 76

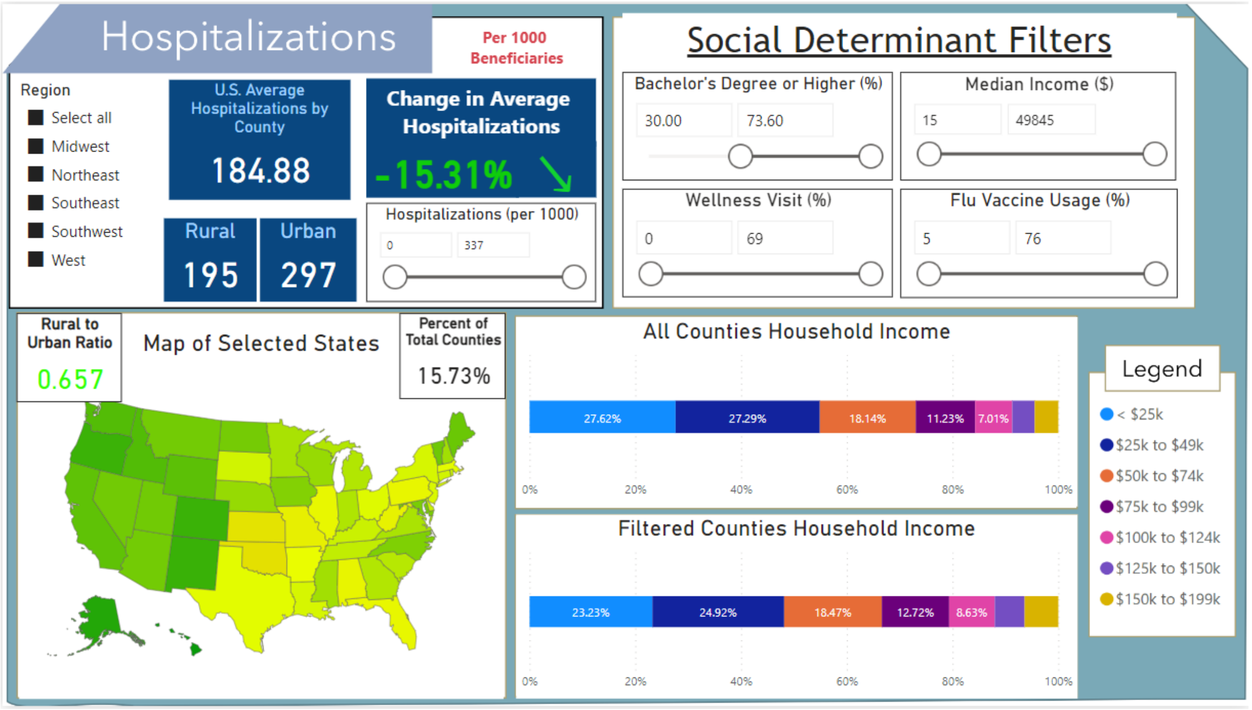
Rural to Urban Ratio: 2.280

Map of Selected States

Percent of Total Counties: 12.37%

Figure 4.17: Hospitalizations West Region

The team determined that an analysis of the Southeast and West regions was the most valuable for the reasons previously stated. The Southwest region had the worst average hospitalization rate, with a rate 8% above the national average, however, the Southeast region also had a strong increase in average hospitalization rate of 4.5%. The Southeast region was of more significance because it had the greatest number of people over the age of 65. Between the Northeast and West regions, the West region was significantly below the national hospitalization average. Additionally, the household income brackets began to spread out more and the higher brackets began to increase in size. The bottom two income brackets summed to about 55% of the 65 and older population before any filters were applied, but after the team looked at only the West, the bottom two brackets only accounted for about 49%.



Hospitalizations

Per 1000 Beneficiaries

Region
- ■ Select all
- ■ Midwest
- ■ Northeast
- ■ Southeast
- ■ Southwest
- ■ West

U.S. Average Hospitalizations by County: 184.88

Change in Average Hospitalizations: -15.31%

Rural: 195

Urban: 297

Hospitalizations (per 1000): 0 – 337

Social Determinant Filters

Bachelor's Degree or Higher (%): 30.00 – 73.60

Median Income ($): 15 – 49845

Wellness Visit (%): 0 – 69

Flu Vaccine Usage (%): 5 – 76

Rural to Urban Ratio: 0.657

Map of Selected States

Percent of Total Counties: 15.73%

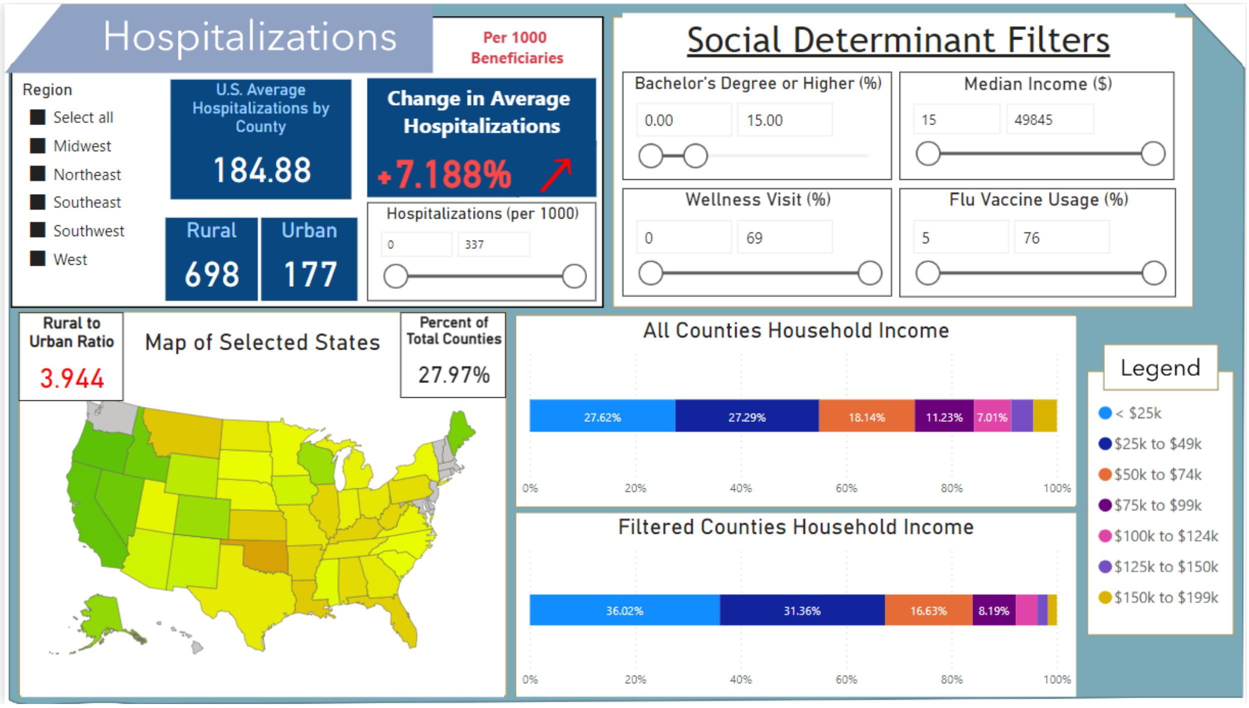
All Counties Household Income

Filtered Counties Household Income

Figure 4.18: Hospitalizations and Bachelor's Degree Over 30%

The bachelor's degree or higher slider is the percentage of the 65 and older population in that county with a bachelor's degree or higher. As the lower bound is increased, the number of counties with low percent values for this variable are omitted. When the lower bound was 10%, the number of counties only decreased by about 5%, so there was still a large number in the selection, and thus fewer effects throughout the filtered data. However, as seen in figure 4.18, as the lower bound was increased to 30%, there was a significant drop in counties. After this filter was placed on the data, around 500 counties were in the dataset and a decrease of 15.31% in the average hospitalization rate was seen in this group. In addition to this, the household income brackets started to move toward the higher brackets and the rural to urban ratio decreased from its U.S. average of 1.844 to 0.657. The research done throughout the project led the team to expect that as the higher education attainment rate increased in a county, the rate of adverse health outcomes would decrease, so this analysis began to verify that research.

Furthermore, as the lower bound was increased further to 30%, 40%, and 50%, the same trend was seen, with an even lower average hospitalization rate and a higher urban county ratio in the filtered dataset. From these filters on the data, it seemed that the counties with a higher percentage of their 65 and older population with a bachelor’s degree may also have fewer hospitalizations per 1000 beneficiaries.



Hospitalizations

Per 1000 Beneficiaries

Region
■ Select all
■ Midwest
■ Northeast
■ Southeast
■ Southwest
■ West

U.S. Average Hospitalizations by County: 184.88

Change in Average Hospitalizations: +7.188% ↗

Rural: 698

Urban: 177

Hospitalizations (per 1000): 0 – 337

Social Determinant Filters

Bachelor's Degree or Higher (%): 0.00 – 15.00

Median Income ($): 15 – 49845

Wellness Visit (%): 0 – 69

Flu Vaccine Usage (%): 5 – 76

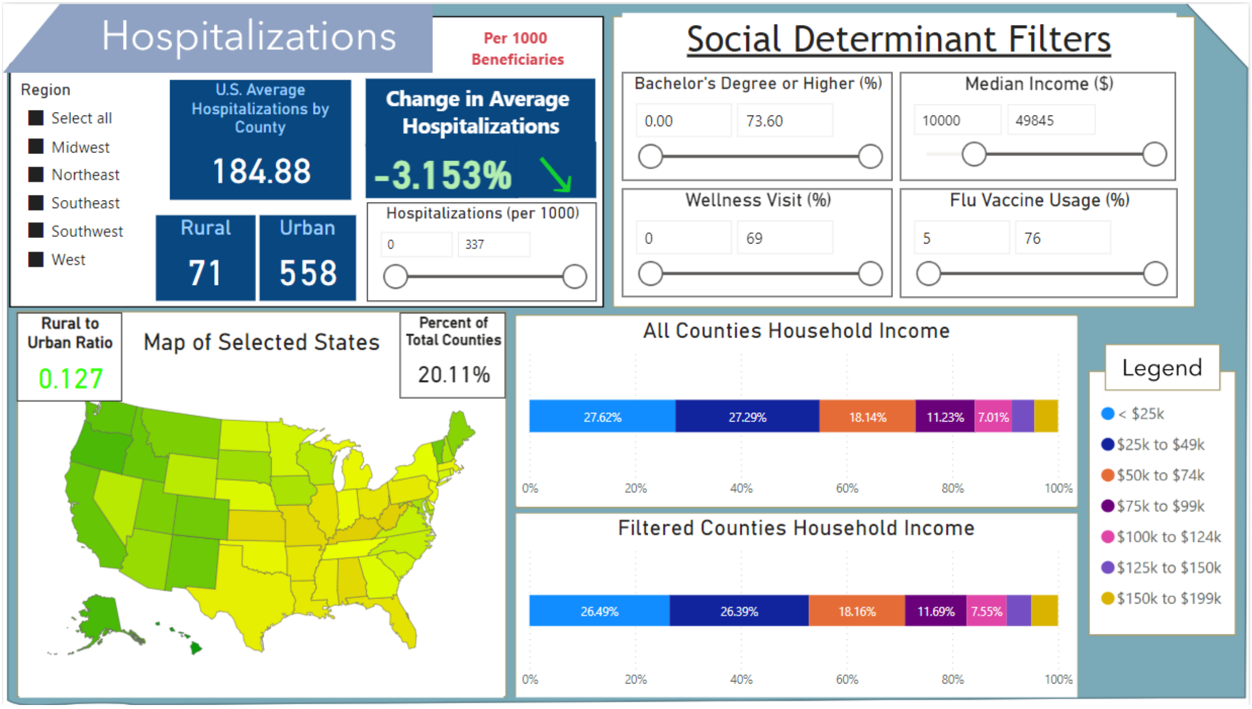
Rural to Urban Ratio: 3.944

Percent of Total Counties: 27.97%

Figure 4.19: Hospitalizations and Bachelor's Degree Under 15%

When the upper bound was set to 40%, there were still 95% of the counties still in the dataset, with little to no change in the average hospitalization rate. However, as the upper bound was set to 25%, the average hospitalizations increased by about 4%, the rural to urban ratio increased, and the household income brackets were now concentrated more in the lower brackets. After lowering the upper bound to 15%, the average hospitalization rate increased by about 7.2% and the number of rural counties outweighs the urban counties significantly. The income brackets substantially move toward the lower brackets, with about 67% of households in the bottom two brackets as opposed to the roughly 55% households initially in those brackets for the total U.S. These trends continue as the upper bound was decreased, however, as the number of counties got low (less than 5% of total counties), outliers begin to emerge and did not follow the exact trends previously seen.

The median income slider is the median income of the 65 and older population in each county in the dataset. One major aspect to note about the median income slider was how it affected the rural to urban ratio of the data. This slider had a massive influence on whether the dashboard showed a majority of rural or urban counties, significantly more than the other sliders.


Hospitalizations

Per 1000 Beneficiaries

Region
■ Select all
■ Midwest
■ Northeast
■ Southeast
■ Southwest
■ West

U.S. Average Hospitalizations by County: 184.88

Change in Average Hospitalizations: -3.153%

Rural: 71

Urban: 558

Hospitalizations (per 1000): 0 – 337

Social Determinant Filters

Bachelor's Degree or Higher (%): 0.00 – 73.60

Median Income ($): 10000 – 49845

Wellness Visit (%): 0 – 69

Flu Vaccine Usage (%): 5 – 76

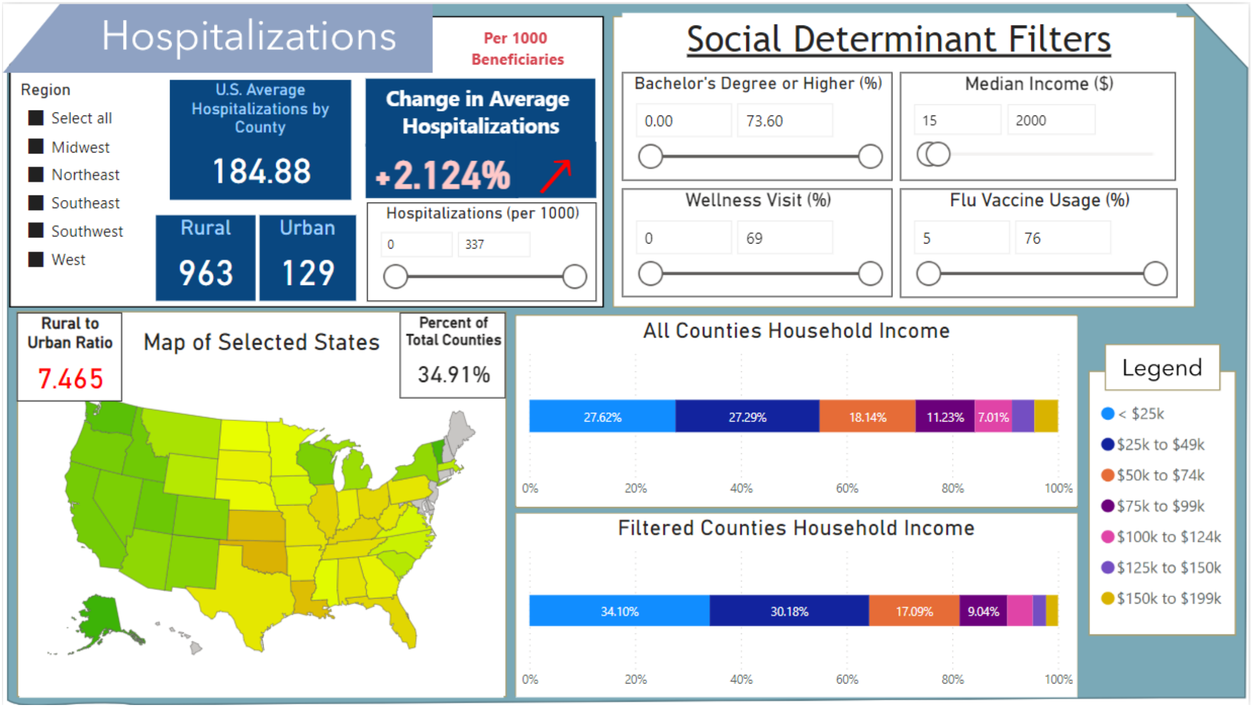
Rural to Urban Ratio: 0.127

Percent of Total Counties: 20.11%

Figure 4.20: Hospitalizations and Median Income Over $10k

For example, if the lower bound was increased to $10,000, practically all the rural counties in the dataset were filtered out of the dashboard. With this lower bound set, over 500 urban counties were included, and the average hospitalization rate had gone down by over 3%. This led the team to question whether there was a connection between urban counties and a lower hospitalization rate.

After increasing the lower bound to $30,000, the team saw the same trends, but more extreme. There were 227 urban counties to just 3 rural counties with over a $30,000 median income, with a greater decrease in average hospitalizations of 4.7%.



Hospitalizations

Per 1000 Beneficiaries

Region

■ Select all
■ Midwest
■ Northeast
■ Southeast
■ Southwest
■ West

U.S. Average Hospitalizations by County 184.88

Change in Average Hospitalizations +2.124%

Rural 963

Urban 129

Hospitalizations (per 1000): 0 – 337

Social Determinant Filters

Bachelor's Degree or Higher (%): 0.00 – 73.60

Median Income ($): 15 – 2000

Wellness Visit (%): 0 – 69

Flu Vaccine Usage (%): 5 – 76

Rural to Urban Ratio 7.465

Map of Selected States

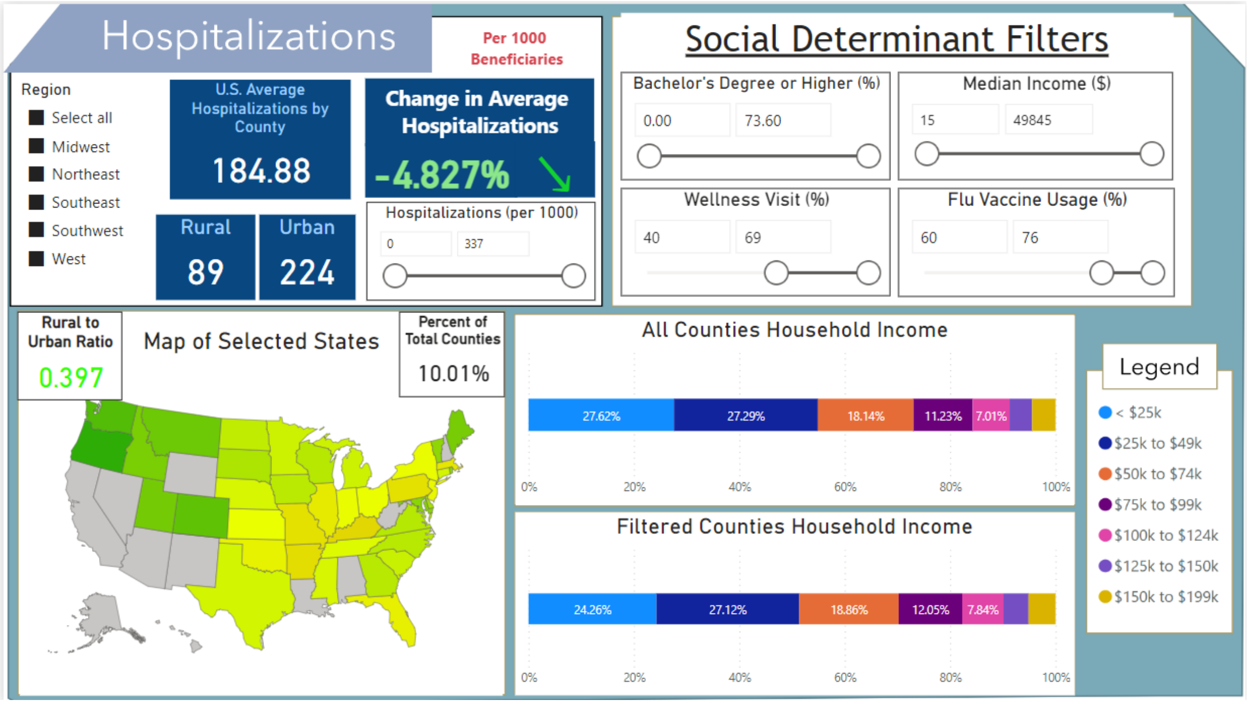
Percent of Total Counties 34.91%

Figure 4.21: Hospitalizations and Median Income Under $2k

On the other hand, if the upper bound was decreased, there was not much of a change until around a maximum median income of $2,000. With this parameter, for every urban county, there were over seven rural counties, and the household income brackets were concentrated in the lower brackets as expected. There was not much of a change regarding the average hospitalization rate, but there was a minor increase from the total average. Overall, it seemed as if the median income of a county had greater effects on the average hospitalization rate when high median incomes were observed. There seemed to be fewer effects on the average hospitalization rate when counties had lower median incomes.

Lastly, the team observed the possible effects of annual wellness visit and influenza vaccine usage among the 65 and older population. With these sliders, the team observed each variable’s individual effects on the dashboard and then their combined effects since they are both preventive measures.

When the lower bound of annual wellness visits was increased to 40%, there was an average hospitalization rate decrease of 1% and an increased urban concentration of counties. After that slider was reset, the lower bound of influenza vaccine usage was increased to 60%, and that showed a similar pattern of lower average hospitalization rates and urban counties. Since these were preventive measures and seemed to affect the dashboard in similar ways, the team wanted to see how they would affect the dashboard when combined.



Hospitalizations

Per 1000 Beneficiaries

Region
■ Select all
■ Midwest
■ Northeast
■ Southeast
■ Southwest
■ West

U.S. Average Hospitalizations by County: 184.88

Change in Average Hospitalizations: -4.827%

Rural: 89

Urban: 224

Hospitalizations (per 1000): 0 – 337

Social Determinant Filters

Bachelor's Degree or Higher (%): 0.00 – 73.60

Median Income ($): 15 – 49845

Wellness Visit (%): 40 – 69

Flu Vaccine Usage (%): 60 – 76

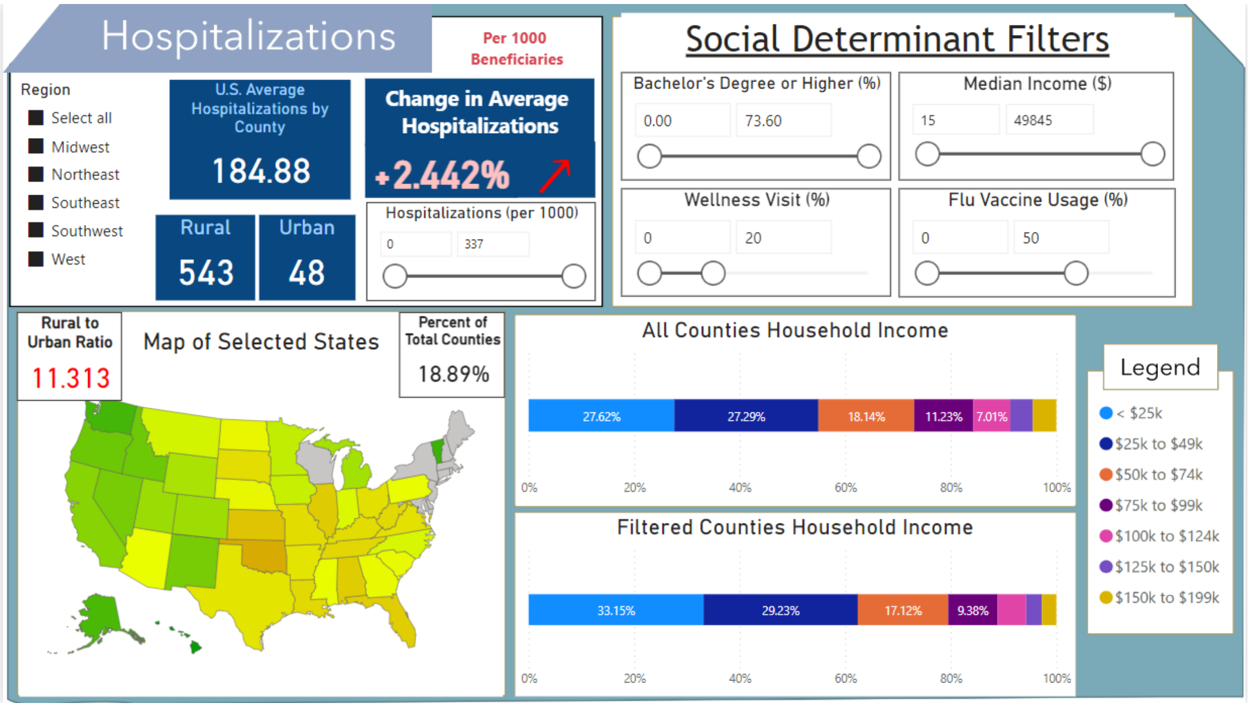
Rural to Urban Ratio: 0.397

Map of Selected States

Percent of Total Counties: 10.01%

Figure 4.22: Hospitalizations and High Preventive Measures

After a filter was set to only see counties with greater than 40% annual wellness visit usage and 60% flu vaccine usage showed a more amplified effect of the two individual sliders. The change in average hospitalizations was about -5% and the counties were once again mostly urban. After both sliders were reset and the upper bound of wellness visits was adjusted to 10%, mostly all the counties were rural and there was a much greater percentage of people in lower income brackets than average. When the upper bound for flu vaccine usage was lowered, there were multiple varying effects that did not follow a trend. However, when manipulating both upper bounds down, there was a trend of filtering out urban counties and increases in the average hospitalizations.



Hospitalizations (per 1000): 0 – 337

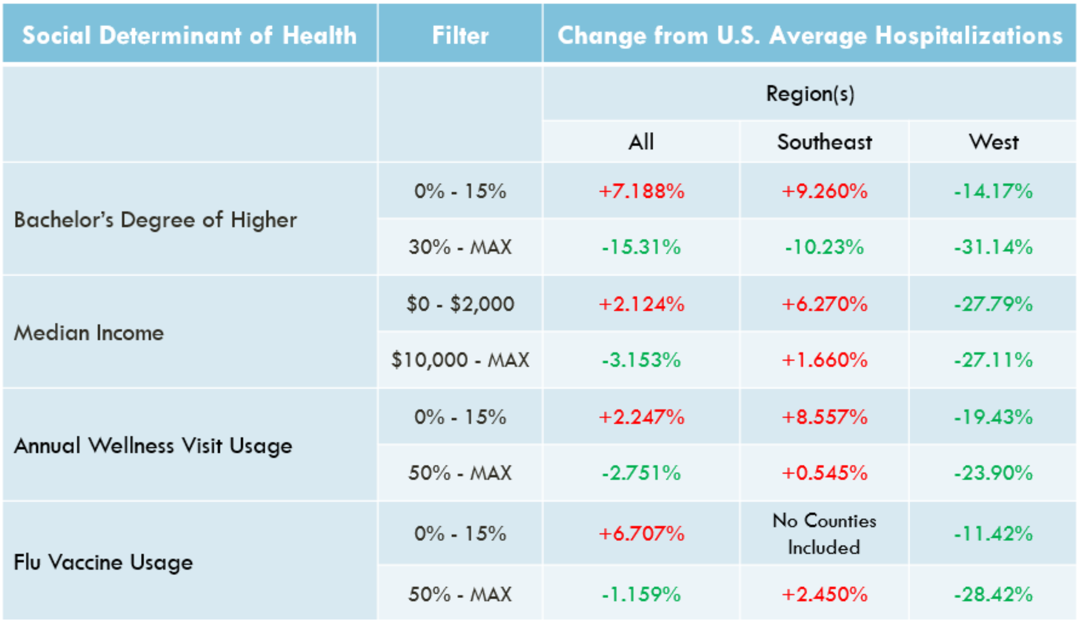
Bachelor's Degree or Higher (%): 0.00 – 73.60

Median Income ($): 15 – 49845

Wellness Visit (%): 0 – 20

Flu Vaccine Usage (%): 0 – 50

Figure 4.23: Hospitalizations and Low Preventive Measures



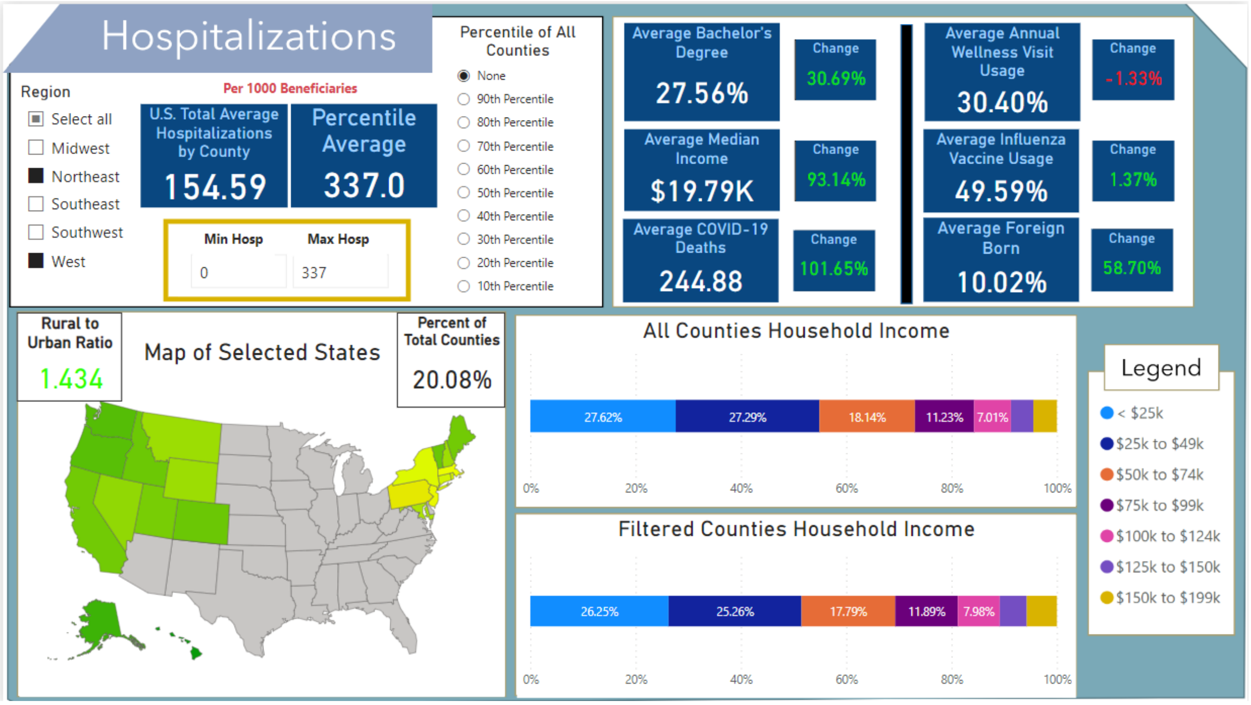
Social Determinant of Health	Filter	Change from U.S. Average Hospitalizations		
		Region(s)		
		All	Southeast	West
Bachelor's Degree of Higher	0% - 15%	+7.188%	+9.260%	-14.17%
	30% - MAX	-15.31%	-10.23%	-31.14%
Median Income	$0 - $2,000	+2.124%	+6.270%	-27.79%
	$10,000 - MAX	-3.153%	+1.660%	-27.11%
Annual Wellness Visit Usage	0% - 15%	+2.247%	+8.557%	-19.43%
	50% - MAX	-2.751%	+0.545%	-23.90%
Flu Vaccine Usage	0% - 15%	+6.707%	No Counties Included	-11.42%
	50% - MAX	-1.159%	+2.450%	-28.42%

Figure 4.24: Hospitalizations Summary

In the hospitalizations summary in figure 4.24, as the lower percentages and dollar amounts of SDoH variables were selected, there were patterns of increased number of outcomes from the national averages. In addition, as the higher percentages and dollar amounts of SDoH variables were selected, there were patterns of decreased number of outcomes from the national averages. Aside from a few deviations, there seemed to be an underlying pattern or connection between significant SDoH variables and hospitalization rates.

4.4.2 Hospitalizations – Percentile Dashboard

When the hospitalizations percentile dashboard was explored, the team focused mostly on counties with high counts of hospitalizations, but also examined the counties with lower hospitalizations as well. As a first step, an investigation of average SDoH variables by regions was done. The Midwest, Southeast, and Southwest regions were of importance because they had the worst average hospitalizations in the previous dashboard.



Hospitalizations

Per 1000 Beneficiaries

Region
- ☒ Select all
- □ Midwest
- ■ Northeast
- □ Southeast
- □ Southwest
- ■ West

U.S. Total Average Hospitalizations by County: 154.59

Percentile Average: 337.0

Min Hosp: 0 | Max Hosp: 337

Percentile of All Counties
- ● None
- ○ 90th Percentile
- ○ 80th Percentile
- ○ 70th Percentile
- ○ 60th Percentile
- ○ 50th Percentile
- ○ 40th Percentile
- ○ 30th Percentile
- ○ 20th Percentile
- ○ 10th Percentile

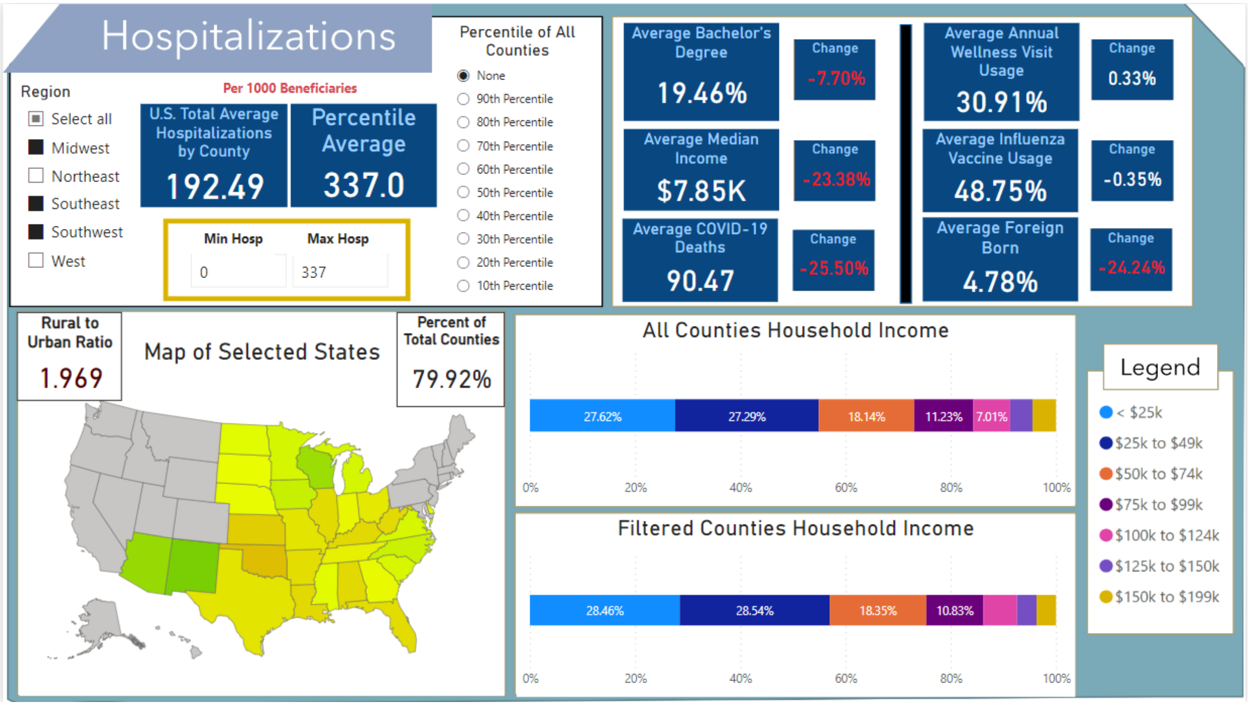
		Change			Change
Average Bachelor's Degree	27.56%	30.69%	Average Annual Wellness Visit Usage	30.40%	-1.33%
Average Median Income	$19.79K	93.14%	Average Influenza Vaccine Usage	49.59%	1.37%
Average COVID-19 Deaths	244.88	101.65%	Average Foreign Born	10.02%	58.70%

Rural to Urban Ratio: 1.434

Map of Selected States

Percent of Total Counties: 20.08%

Figure 4.25: Northeast and West SDoH Averages



Hospitalizations

Per 1000 Beneficiaries

Region
- ☒ Select all
- ■ Midwest
- □ Northeast
- ■ Southeast
- ■ Southwest
- □ West

U.S. Total Average Hospitalizations by County: 192.49

Percentile Average: 337.0

Min Hosp: 0 | Max Hosp: 337

Percentile of All Counties
- ● None
- ○ 90th Percentile
- ○ 80th Percentile
- ○ 70th Percentile
- ○ 60th Percentile
- ○ 50th Percentile
- ○ 40th Percentile
- ○ 30th Percentile
- ○ 20th Percentile
- ○ 10th Percentile

		Change			Change
Average Bachelor's Degree	19.46%	-7.70%	Average Annual Wellness Visit Usage	30.91%	0.33%
Average Median Income	$7.85K	-23.38%	Average Influenza Vaccine Usage	48.75%	-0.35%
Average COVID-19 Deaths	90.47	-25.50%	Average Foreign Born	4.78%	-24.24%

Rural to Urban Ratio: 1.969

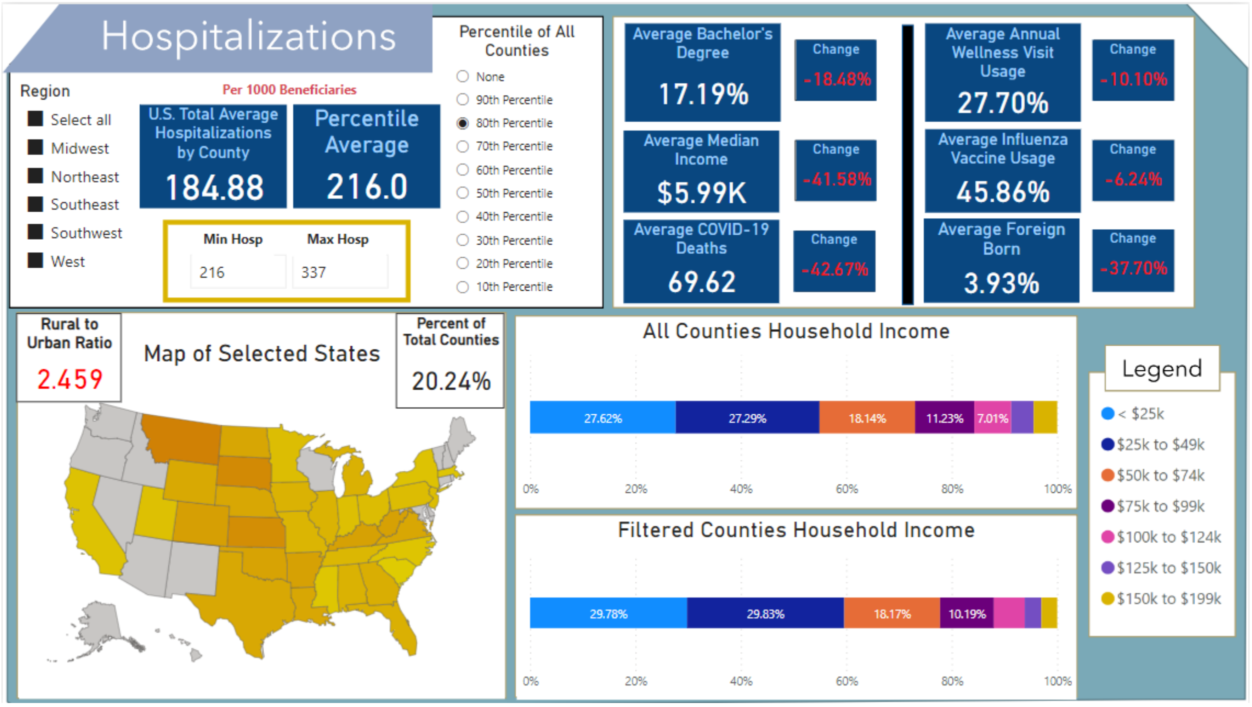
Map of Selected States

Percent of Total Counties: 79.92%

Figure 4.26: Midwest, Southeast, and Southwest SDoH Averages

The team looked for any commonalities between the three regions below the national average of hospitalizations and found that all three had a bachelor's degree or higher rate below the national average, whereas the Northeast and West were above the national average. Additionally, the Midwest, Southeast, and Southwest regions had an average median income less than the national average where Northeast and West were both well above average.

The SDoH variables that the team paid most attention to were bachelor's degree or higher, median income, and the preventive measures. The Midwest, Southeast, and Southwest regions all had large decreases SDoH variables on average in the bachelor's degree and median income variables. The preventive measures were minimally changed.



Hospitalizations

Region
■ Select all
■ Midwest
■ Northeast
■ Southeast
■ Southwest
■ West

Per 1000 Beneficiaries

U.S. Total Average Hospitalizations by County: 184.88

Percentile Average: 216.0

Min Hosp: 216
Max Hosp: 337

Percentile of All Counties
○ None
○ 90th Percentile
● 80th Percentile
○ 70th Percentile
○ 60th Percentile
○ 50th Percentile
○ 40th Percentile
○ 30th Percentile
○ 20th Percentile
○ 10th Percentile

Average Bachelor's Degree: 17.19% — Change: -18.48%

Average Median Income: $5.99K — Change: -41.58%

Average COVID-19 Deaths: 69.62 — Change: -42.67%

Average Annual Wellness Visit Usage: 27.70% — Change: -10.10%

Average Influenza Vaccine Usage: 45.86% — Change: -6.24%

Average Foreign Born: 3.93% — Change: -37.70%

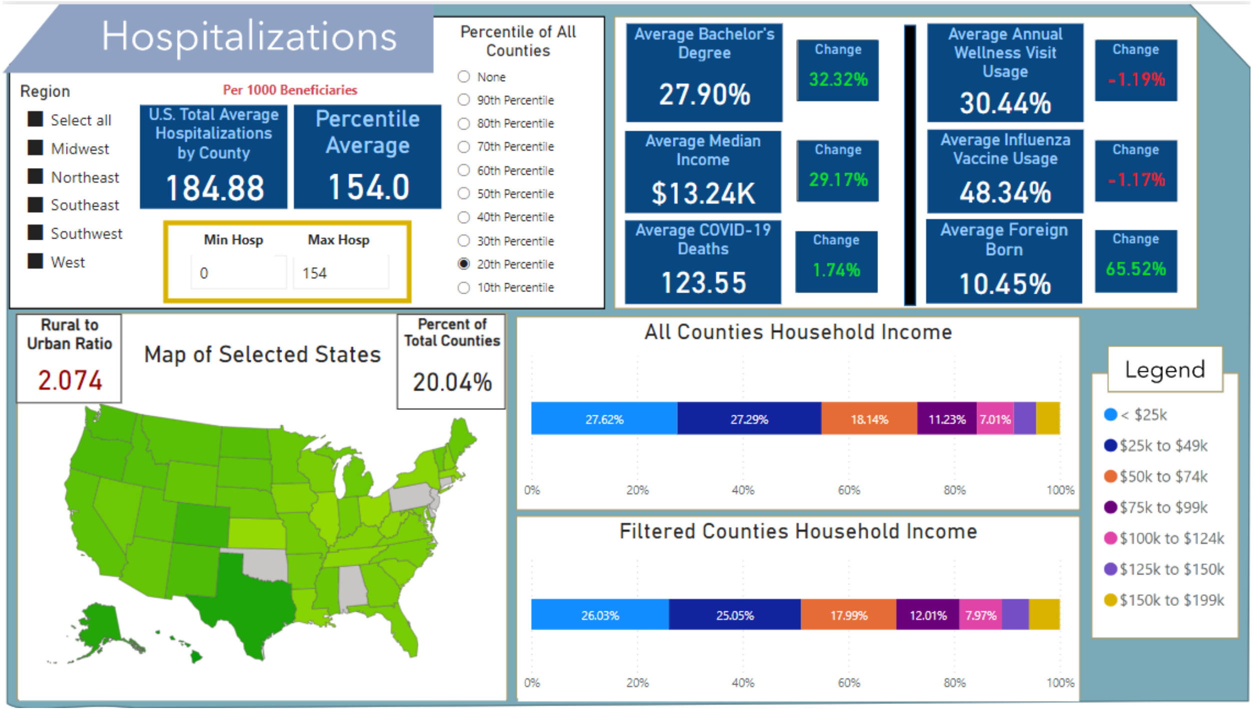
Rural to Urban Ratio: 2.459

Map of Selected States

Percent of Total Counties: 20.24%

Figure 4.27: Above 80th Percentile Hospitalizations

After the team focused on the worst and best regions, the top and bottom percentiles were explored. When the 80th percentile was selected and the number of hospitalizations was inputted into the "Min Hosp," the average SDoH variables shown decreased more than when the team only looked at the worst regions. The 80th percentile was used to see the counties that have more hospitalizations than 80% of the counties in the U.S. The bachelor's variable was decreased to 17% of the counties on average, and the average median income of these counties decreased by about 43%. The preventive measures also had a considerable decrease in their averages. Since the core aspect of this project was to determine how SDoH could have possible effects on health outcomes, the expectation was that as the team focused on counties above a certain outcome level, that the average of the SDoH would change accordingly.



Hospitalizations

Region
■ Select all
■ Midwest
■ Northeast
■ Southeast
■ Southwest
■ West

Per 1000 Beneficiaries

U.S. Total Average Hospitalizations by County: 184.88

Percentile Average: 154.0

Min Hosp: 0
Max Hosp: 154

Percentile of All Counties
○ None
○ 90th Percentile
○ 80th Percentile
○ 70th Percentile
○ 60th Percentile
○ 50th Percentile
○ 40th Percentile
○ 30th Percentile
● 20th Percentile
○ 10th Percentile

Average Bachelor's Degree: 27.90% — Change: 32.32%

Average Median Income: $13.24K — Change: 29.17%

Average COVID-19 Deaths: 123.55 — Change: 1.74%

Average Annual Wellness Visit Usage: 30.44% — Change: -1.19%

Average Influenza Vaccine Usage: 48.34% — Change: -1.17%

Average Foreign Born: 10.45% — Change: 65.52%

Rural to Urban Ratio: 2.074

Percent of Total Counties: 20.04%

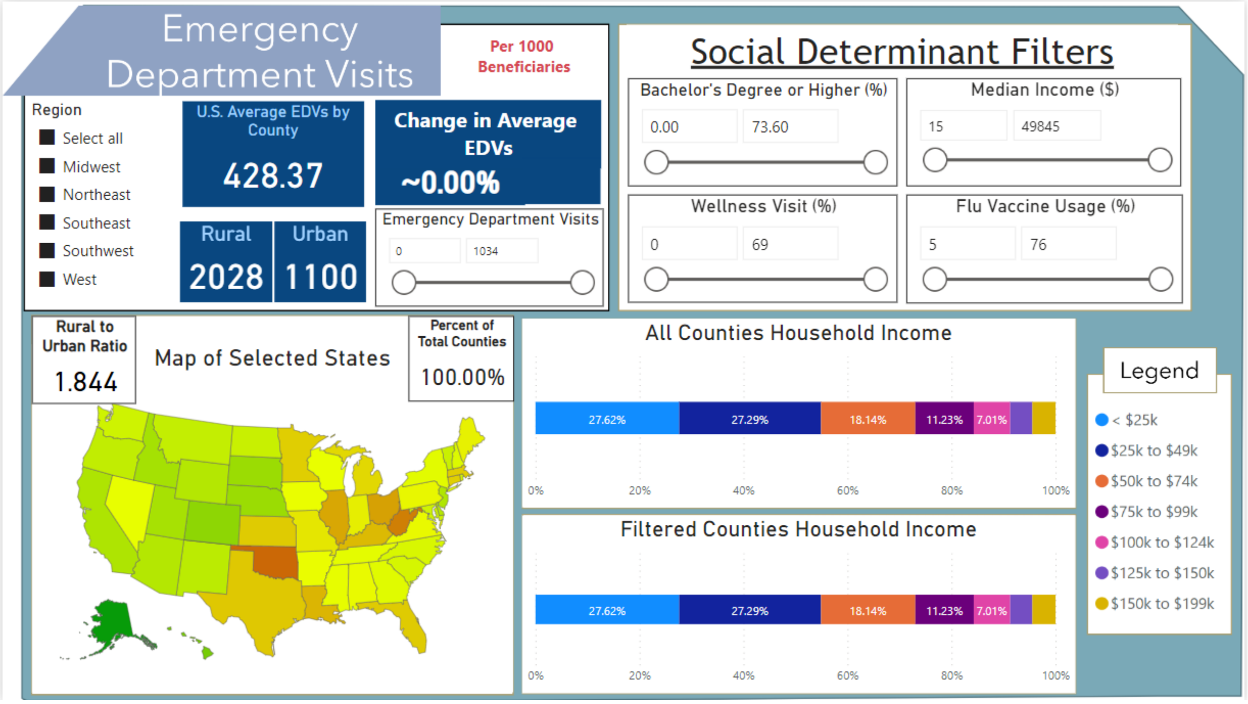
Figure 4.28: Below 20th Percentile Hospitalizations

When the 20th percentile was selected and the number of hospitalizations was inputted into the "Max Hosp," the average SDoH variables shown increased more than when the team only looked at the worst regions. The 20th percentile was used to see the counties that have fewer hospitalizations than 80% of the counties in the U.S. The bachelor's variable increased to 28% of the counties on average, and the average median income of these counties increased by about

30%. The preventive measures, however, showed patterns that the team did not expect since higher preventive measures should lower average hospitalizations. The preventive measures did show a slight decrease, but the decreases were too small from the national average to show a significant impact.

4.4.3 Emergency Department Visits – Slider Dashboard

Since the sliders only act as filters on the dataset and the data is the same for every county besides their health outcomes, the information on this dashboard was mostly the same when comparing hospitalizations and emergency department visits. The main thing to note was the change in the average emergency department visit rate as the same SDoH filters from the hospitalizations dashboard were applied. Also, the team investigated if these changes followed the same trends as hospitalizations or if emergency department visits were different.



Emergency Department Visits

Per 1000 Beneficiaries

Region
■ Select all
■ Midwest
■ Northeast
■ Southeast
■ Southwest
■ West

U.S. Average EDVs by County: 428.37

Change in Average EDVs: ~0.00%

Rural	Urban
2028	1100

Emergency Department Visits: 0 – 1034

Social Determinant Filters

Bachelor's Degree or Higher (%): 0.00 – 73.60

Median Income ($): 15 – 49845

Wellness Visit (%): 0 – 69

Flu Vaccine Usage (%): 5 – 76

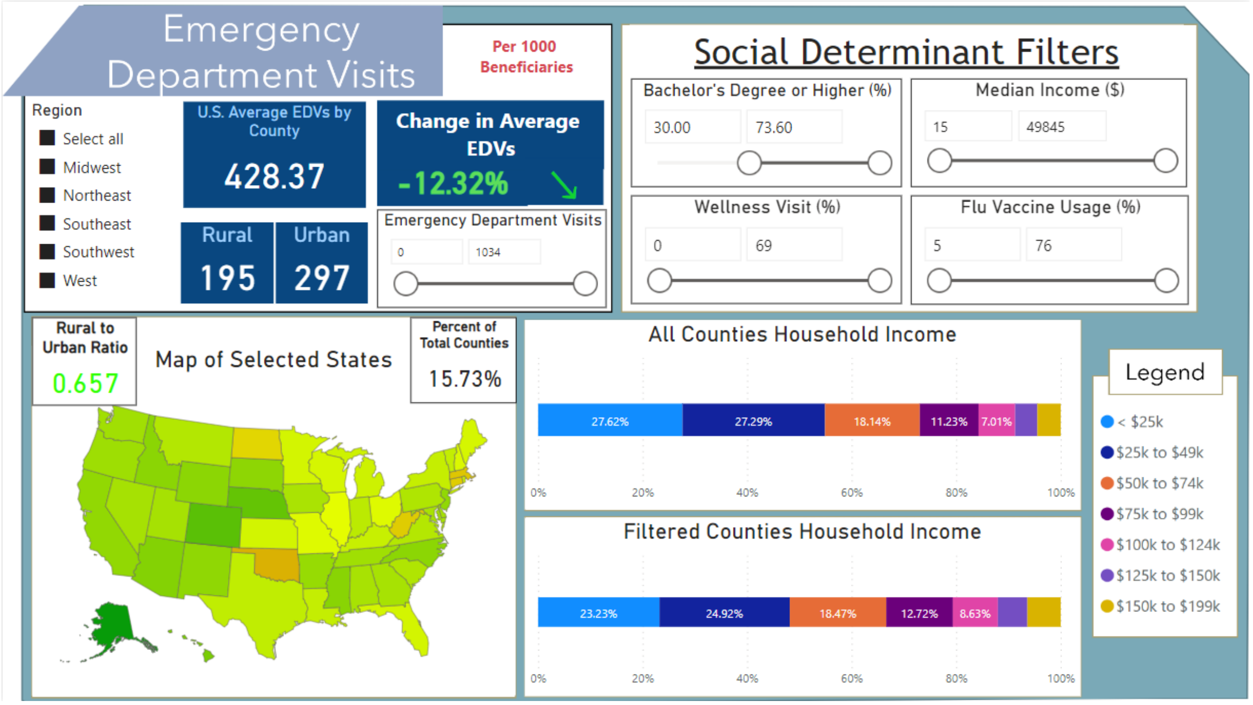
Rural to Urban Ratio: 1.844

Map of Selected States

Percent of Total Counties: 100.00%

Figure 4.29: Emergency Department Visits Slider Dashboard

As a first step for emergency department visits, the team wanted to see if the same patterns found in hospitalizations would be seen in the Southeast and West regions. The regions with average emergency department visits above the national average were Midwest, Southeast, and Southwest again, and Southwest was again the worst region on average with an increase of 6.5% in the average emergency department visit rate. The Southeast region was then investigated and a 1.696% increase in the emergency department visit rate was seen. It was worth noting that since the emergency department visit rates were of a greater magnitude than hospitalization rates, even a small percentage change could significantly alter the number of emergency department visits seen. Additionally, the Northeast and West regions had average emergency department visit rates below the national average. The West region had the greatest decrease once again with an emergency department visit rate about 13% below the national rate.



U.S. Average EDVs by County: 428.37

Change in Average EDVs: -12.32%

Rural: 195 | Urban: 297

Emergency Department Visits: 0 – 1034

Bachelor's Degree or Higher (%): 30.00 – 73.60

Median Income ($): 15 – 49845

Wellness Visit (%): 0 – 69

Flu Vaccine Usage (%): 5 – 76

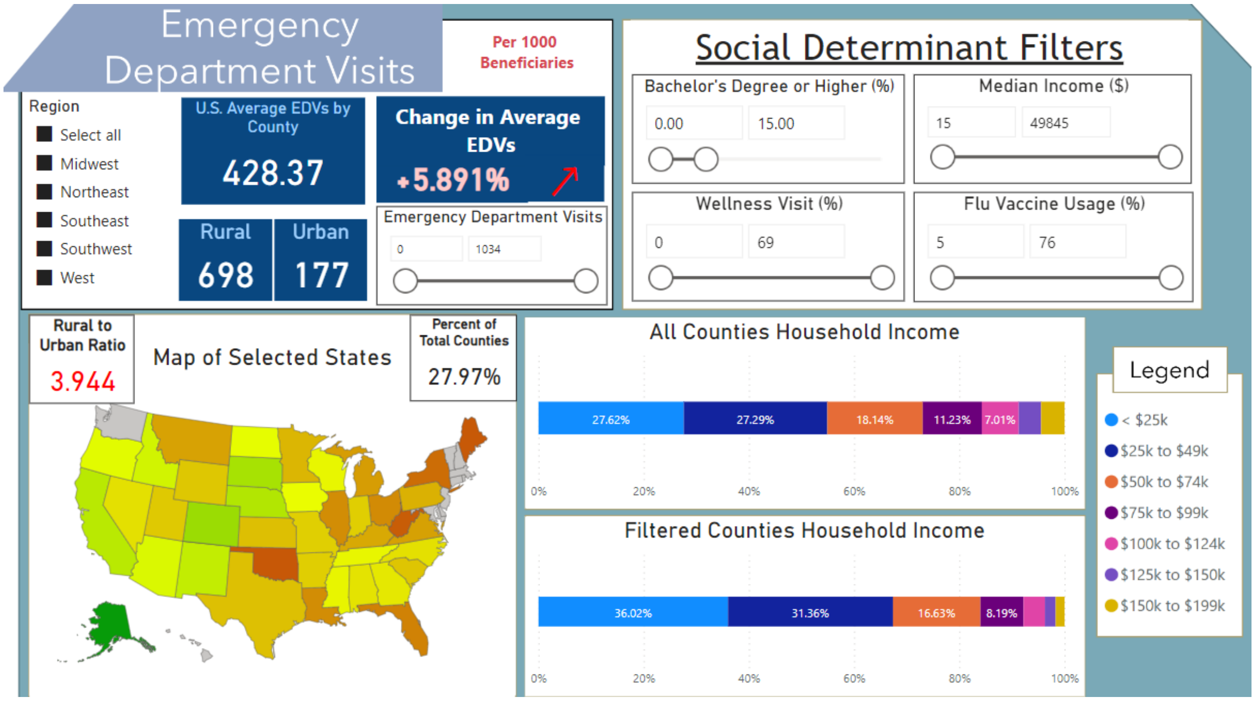
Rural to Urban Ratio: 0.657

Percent of Total Counties: 15.73%

Figure 4.30: Emergency Department Visits Bachelor's Degree Over 30%

As the lower bound for the bachelor's degree or higher slider is increased, the number of counties with low percent values for this variable are omitted. When the lower bound was 10%, the number of counties only decreased by about 5%, so there was still a large number in the selection, and thus fewer conclusions to be made about the filtered data. However, as the lower bound was increased to 30%, there was a significant drop in counties and a decrease of 12.32% in the average emergency department visit rate. Furthermore, as the lower bound was increased further to 30%, 40%, and 50%, the trend became extremely apparent, with a dramatically lower average emergency department visit rate. From these filters on the data, it seemed that the counties with a higher percentage of their 65 and older population with a bachelor's degree may also have fewer emergency department visits per 1000 beneficiaries as well.

The research done throughout the project led the team to expect that as the percentage of people with higher education increased in a county, the rate of unfortunate health outcomes would decrease, and this data analysis seemed to support that argument.



Emergency Department Visits

Per 1000 Beneficiaries

Region
■ Select all
■ Midwest
■ Northeast
■ Southeast
■ Southwest
■ West

U.S. Average EDVs by County: 428.37

Change in Average EDVs: +5.891%

Rural: 698

Urban: 177

Emergency Department Visits: 0 – 1034

Social Determinant Filters

Bachelor's Degree or Higher (%): 0.00 – 15.00

Median Income ($): 15 – 49845

Wellness Visit (%): 0 – 69

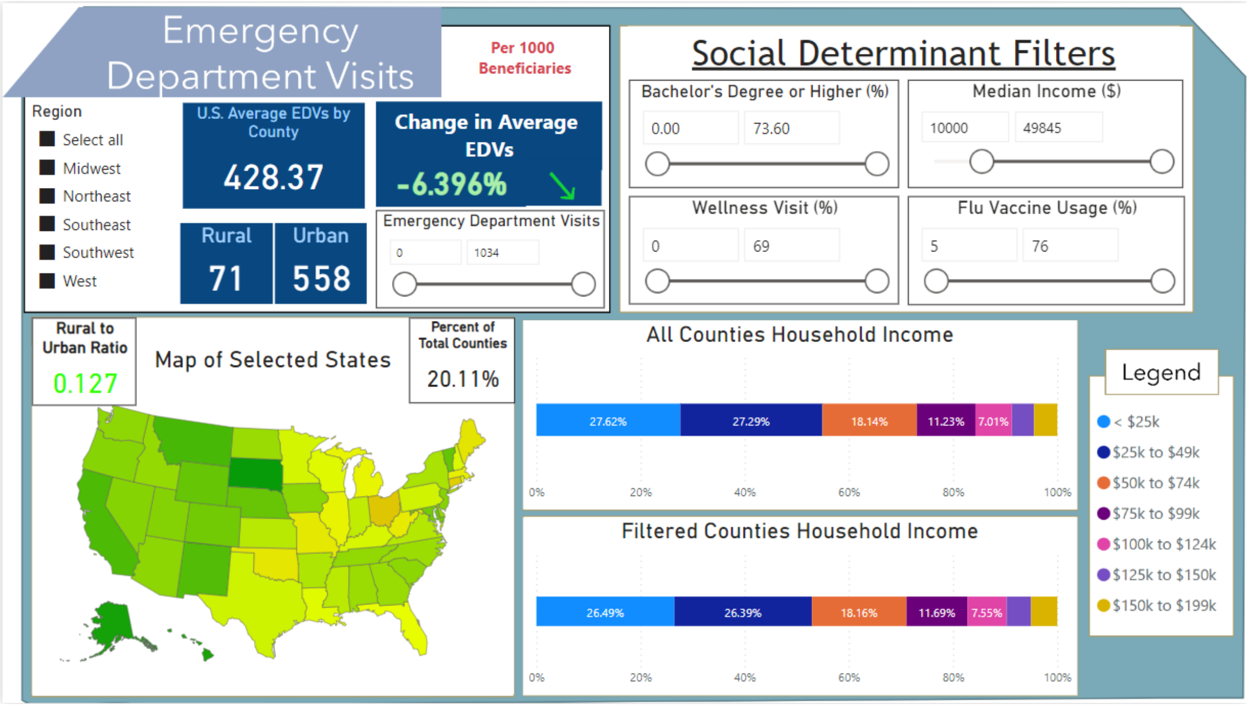
Flu Vaccine Usage (%): 5 – 76

Rural to Urban Ratio: 3.944

Percent of Total Counties: 27.97%

Figure 4.31: Emergency Department Visits Bachelor's Degree Under 15%

After resetting the bachelor’s degree slider by dragging it fully to the left, the upper bound was then manipulated. The upper bound was lowered to 15%, and the average emergency department visit rate increased by about 7% and the number of rural counties outweighed the urban counties significantly. This result confirmed the team’s notion that bachelor’s degree attainment is significantly tied to health outcomes in general. These trends continue as the upper bound was decreased, however, as the number of counties got low (less than 5% of total counties), outliers began to emerge as they did in the hospitalizations dashboard and at times did not follow the exact trends previously seen.



Emergency Department Visits

Per 1000 Beneficiaries

Region
- ■ Select all
- ■ Midwest
- ■ Northeast
- ■ Southeast
- ■ Southwest
- ■ West

U.S. Average EDVs by County: 428.37

Change in Average EDVs: -6.396%

Rural: 71

Urban: 558

Emergency Department Visits: 0 – 1034

Social Determinant Filters

Bachelor's Degree or Higher (%): 0.00 – 73.60

Median Income ($): 10000 – 49845

Wellness Visit (%): 0 – 69

Flu Vaccine Usage (%): 5 – 76

Rural to Urban Ratio: 0.127

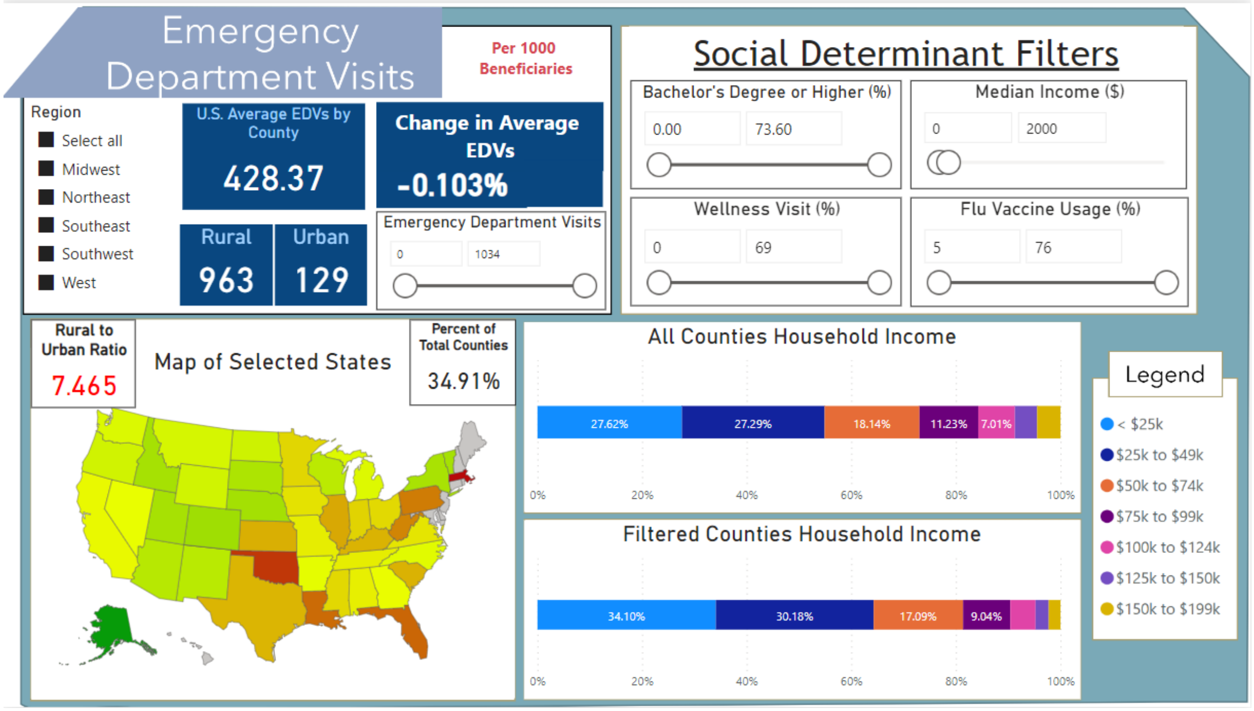
Percent of Total Counties: 20.11%

Map of Selected States

All Counties Household Income: 27.62% | 27.29% | 18.14% | 11.23% | 7.01%

Filtered Counties Household Income: 26.49% | 26.39% | 18.16% | 11.69% | 7.55%

Figure 4.32: Emergency Department Visits Median Income Over $10k



Emergency Department Visits

Per 1000 Beneficiaries

Region
■ Select all
■ Midwest
■ Northeast
■ Southeast
■ Southwest
■ West

U.S. Average EDVs by County: 428.37

Change in Average EDVs: -0.103%

Rural: 963
Urban: 129

Emergency Department Visits: 0 – 1034

Social Determinant Filters

Bachelor's Degree or Higher (%): 0.00 – 73.60

Median Income ($): 0 – 2000

Wellness Visit (%): 0 – 69

Flu Vaccine Usage (%): 5 – 76

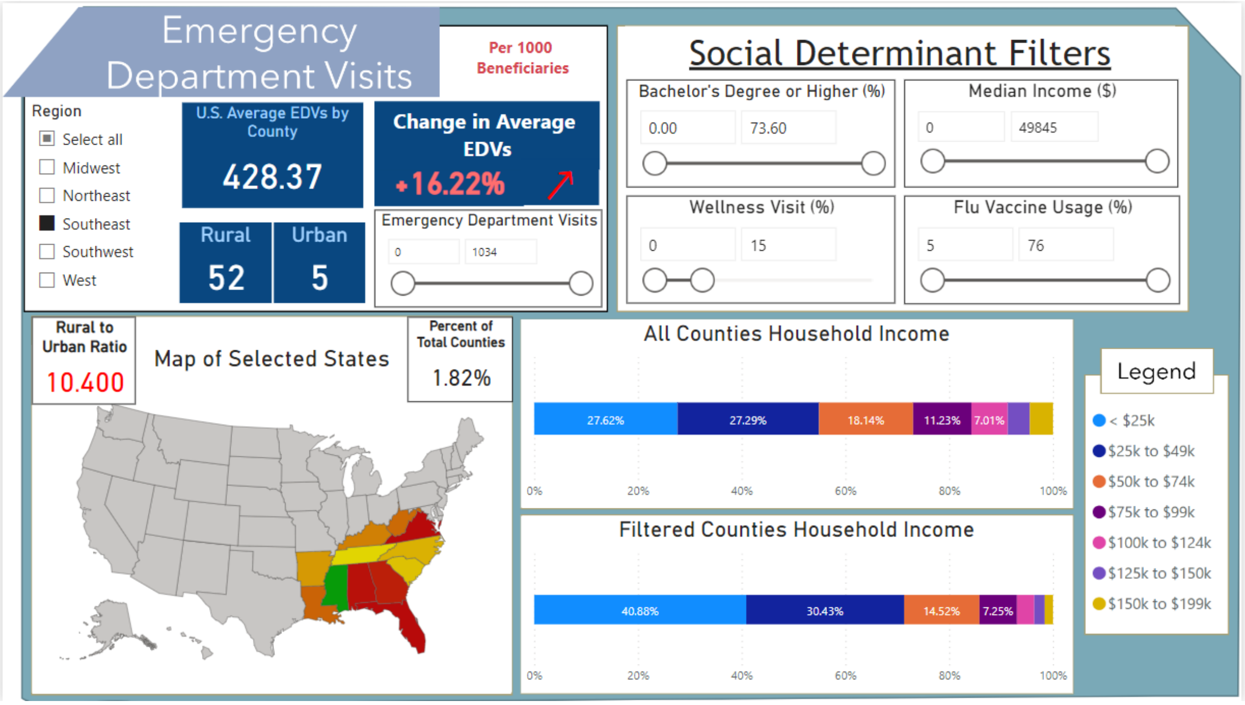
Rural to Urban Ratio: 7.465

Map of Selected States

Percent of Total Counties: 34.91%

Figure 4.33: Emergency Department Visits Median Income Under $2k

The adjustments made to the median income slider in emergency department visits also has similar results to hospitalizations, with higher median incomes resulting in lower emergency department visit rates. However, when the team looked at lower median incomes, there were relatively no significant patterns seen as the median income filter changed. This further demonstrates that SDoH are not always significant on their own, but rather provide context to that county's health outcomes in general. In conclusion, it seemed as if the median income variable had possible effects on emergency department visits when counties with relatively high median incomes were observed. There were no consistent effects on the average emergency department visit rate when focusing on counties with lower median incomes.



Emergency Department Visits

Per 1000 Beneficiaries

Region
▣ Select all
☐ Midwest
☐ Northeast
■ Southeast
☐ Southwest
☐ West

U.S. Average EDVs by County: 428.37

Change in Average EDVs: +16.22% ↗

Rural: 52

Urban: 5

Emergency Department Visits: 0 – 1034

Social Determinant Filters

Bachelor's Degree or Higher (%): 0.00 – 73.60

Median Income ($): 0 – 49845

Wellness Visit (%): 0 – 15

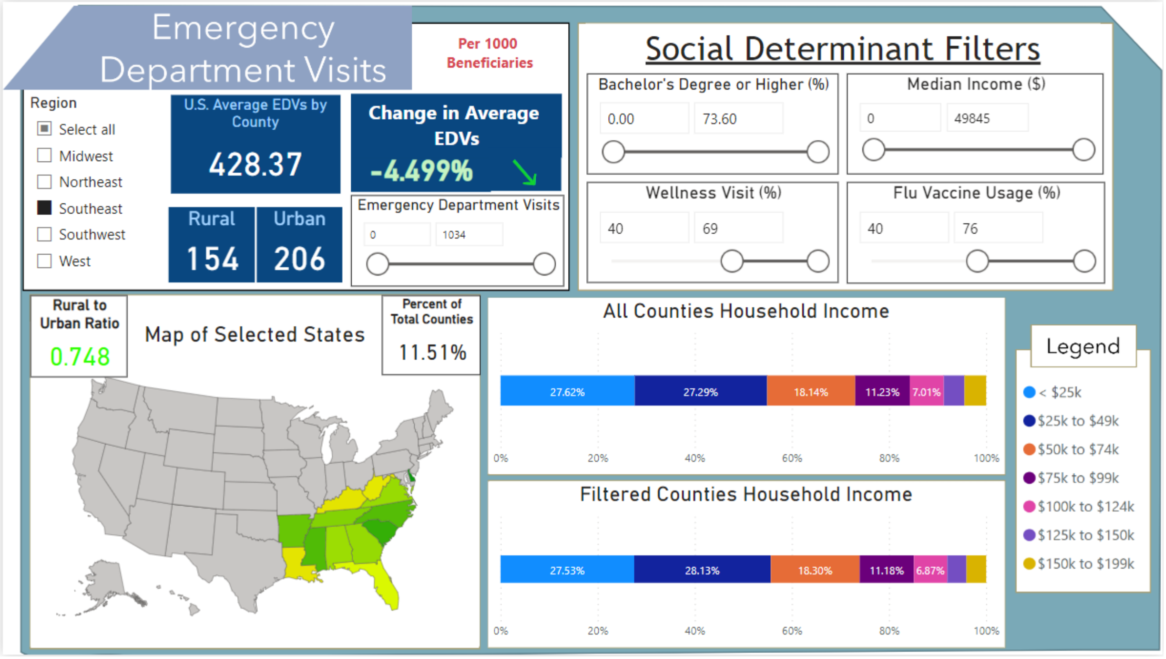
Flu Vaccine Usage (%): 5 – 76

Rural to Urban Ratio: 10.400

Percent of Total Counties: 1.82%

Figure 4.34: Emergency Department Visits Wellness Visit Usage Under 15% for Southeast

One significant aspect of preventive measures that was documented was that the Southeast region was particularly worse than the other regions in emergency department visit rates after some filters were applied. When the annual wellness visit slider was set to less than 15%, the dashboard showed that there was an increase of 16% in emergency department visit rates in the Southeast when compared to the national average. Specifically, counties in Florida with less than 15% of the county using annual wellness visits had relatively high emergency department visit rates.



Emergency Department Visits

Per 1000 Beneficiaries

Region
- ▣ Select all
- □ Midwest
- □ Northeast
- ■ Southeast
- □ Southwest
- □ West

U.S. Average EDVs by County: 428.37

Change in Average EDVs: -4.499% ↘

Rural	Urban
154	206

Emergency Department Visits: 0 – 1034

Social Determinant Filters

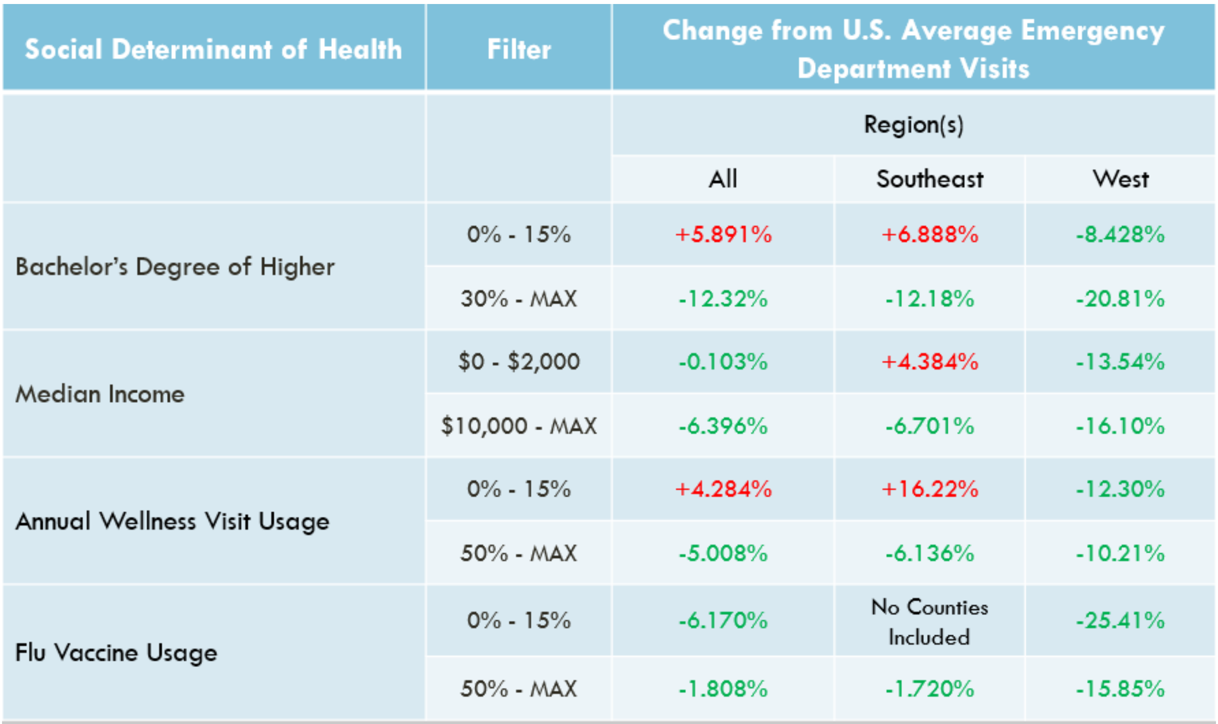
Filter	Min	Max
Bachelor's Degree or Higher (%)	0.00	73.60
Median Income ($)	0	49845
Wellness Visit (%)	40	69
Flu Vaccine Usage (%)	40	76

Rural to Urban Ratio: 0.748

Percent of Total Counties: 11.51%

Figure 4.35: Emergency Department Visits Preventive Measures Over 40% for Southeast

On the other hand, when annual wellness visit and flu vaccine usage were filtered to be over 40%, the Southeast had a decrease in average emergency department visit rate in comparison to the national average.



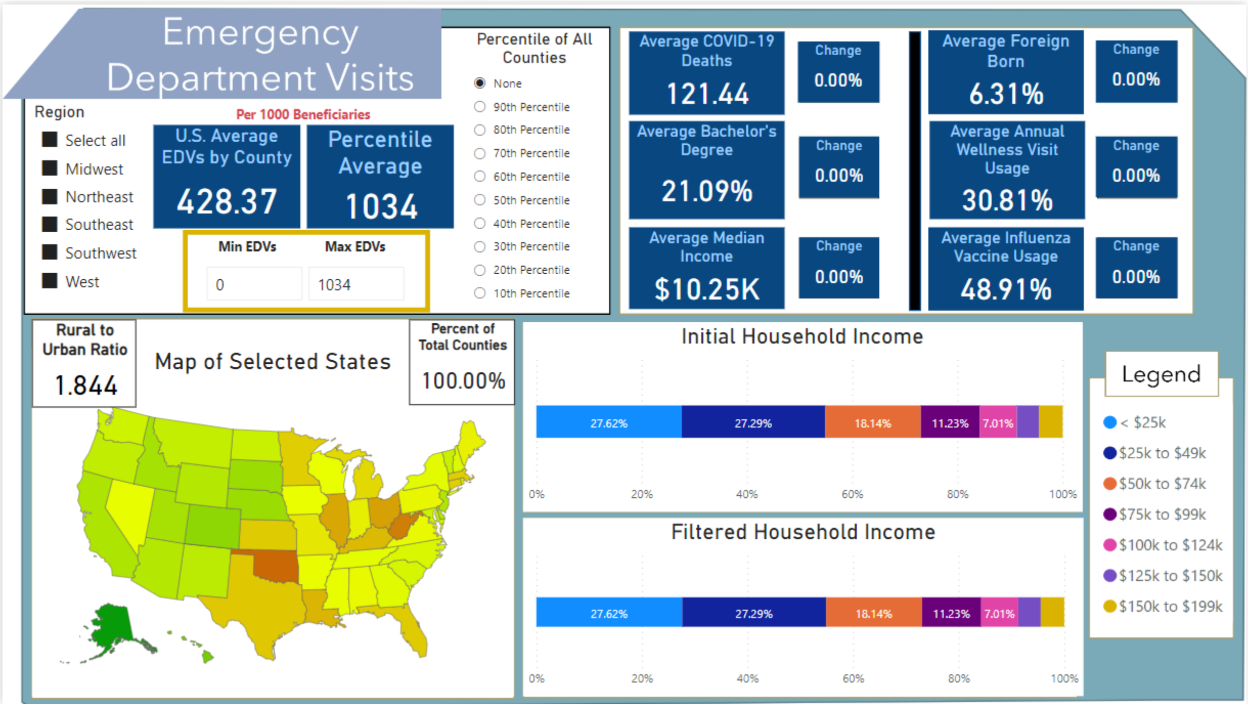
Social Determinant of Health	Filter	Change from U.S. Average Emergency Department Visits		
		Region(s)		
		All	Southeast	West
Bachelor's Degree of Higher	0% - 15%	+5.891%	+6.888%	-8.428%
	30% - MAX	-12.32%	-12.18%	-20.81%
Median Income	$0 - $2,000	-0.103%	+4.384%	-13.54%
	$10,000 - MAX	-6.396%	-6.701%	-16.10%
Annual Wellness Visit Usage	0% - 15%	+4.284%	+16.22%	-12.30%
	50% - MAX	-5.008%	-6.136%	-10.21%
Flu Vaccine Usage	0% - 15%	-6.170%	No Counties Included	-25.41%
	50% - MAX	-1.808%	-1.720%	-15.85%

Figure 4.36: Emergency Department Visits Summary

In the summary of emergency department visit rates in figure 4.36, there were some trends here that the team had not seen before in hospitalizations. The West region performed like how it had in the hospitalization dashboard, but the Southeast region was rather different. In the Southeast, when the team selected counties with higher SDoH variables, there was now a steady decrease in the average emergency department visit rates, whereas in hospitalizations, there was not such an effect. As for all regions, there was an effect that seemed erroneous in this summary. When the flu vaccine rate was low, counties showed a decrease in average rates of 6.170%, but when higher flu vaccine rates were selected, a decrease of only 1.808% was shown. This could most likely be explained due to outliers in the dataset, especially since many counties were within a flu vaccine usage rate of 20-45%.

In conclusion, there also seemed to be an underlying pattern or connection between significant SDoH variables and emergency department visit rates, but when outliers began to emerge in the dataset, it became hard to distinguish whether there was actual change or not.

4.4.4 Emergency Department Visits – Percentile Dashboard



Emergency Department Visits

Region
■ Select all
■ Midwest
■ Northeast
■ Southeast
■ Southwest
■ West

Per 1000 Beneficiaries

U.S. Average EDVs by County: 428.37

Percentile Average: 1034

Min EDVs: 0
Max EDVs: 1034

Percentile of All Counties
● None
○ 90th Percentile
○ 80th Percentile
○ 70th Percentile
○ 60th Percentile
○ 50th Percentile
○ 40th Percentile
○ 30th Percentile
○ 20th Percentile
○ 10th Percentile

Average COVID-19 Deaths: 121.44 — Change: 0.00%

Average Bachelor's Degree: 21.09% — Change: 0.00%

Average Median Income: $10.25K — Change: 0.00%

Average Foreign Born: 6.31% — Change: 0.00%

Average Annual Wellness Visit Usage: 30.81% — Change: 0.00%

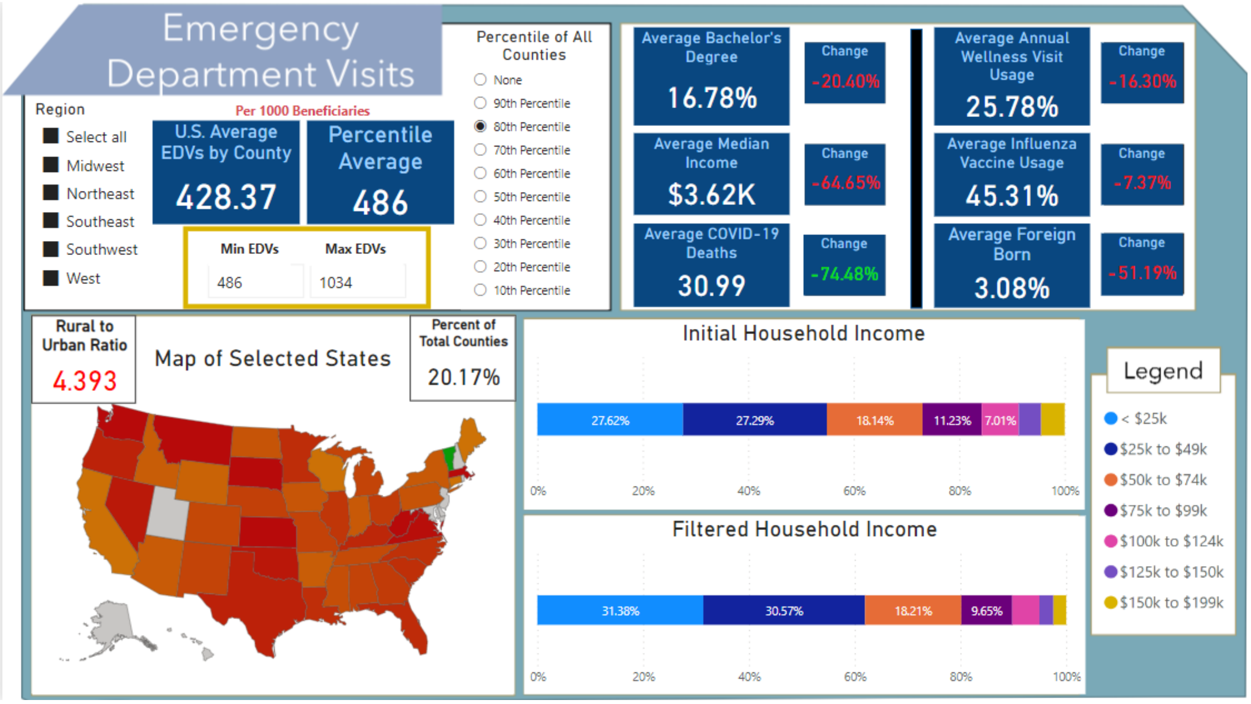
Average Influenza Vaccine Usage: 48.91% — Change: 0.00%

Rural to Urban Ratio: 1.844

Percent of Total Counties: 100.00%

Figure 4.37: Emergency Department Visits Percentile Dashboard

When the emergency department visits percentile dashboard was explored, the team focused mostly on counties with high counts, but also examined the counties with lower counts as well.


Emergency Department Visits

Region
- ■ Select all
- ■ Midwest
- ■ Northeast
- ■ Southeast
- ■ Southwest
- ■ West

Per 1000 Beneficiaries

U.S. Average EDVs by County: 428.37

Percentile Average: 486

Min EDVs: 486

Max EDVs: 1034

Percentile of All Counties
- ○ None
- ○ 90th Percentile
- ● 80th Percentile
- ○ 70th Percentile
- ○ 60th Percentile
- ○ 50th Percentile
- ○ 40th Percentile
- ○ 30th Percentile
- ○ 20th Percentile
- ○ 10th Percentile

Average Bachelor's Degree 16.78%	Change -20.40%	Average Annual Wellness Visit Usage 25.78%	Change -16.30%
Average Median Income $3.62K	Change -64.65%	Average Influenza Vaccine Usage 45.31%	Change -7.37%
Average COVID-19 Deaths 30.99	Change -74.48%	Average Foreign Born 3.08%	Change -51.19%

Rural to Urban Ratio: 4.393

Map of Selected States

Percent of Total Counties: 20.17%

Figure 4.38: Over 80th Percentile Emergency Department Visits

The counties above the 80th percentile of emergency department visits had worse SDoH averages than their hospitalizations counterparts from the other percentile dashboard. There were more extreme downturns in all the variables that were of importance. The median income average went down 65%, and the preventive service usage was also worse on average. The number of COVID-19 deaths decreased on average, which may be a result of rural counties with lower populations.

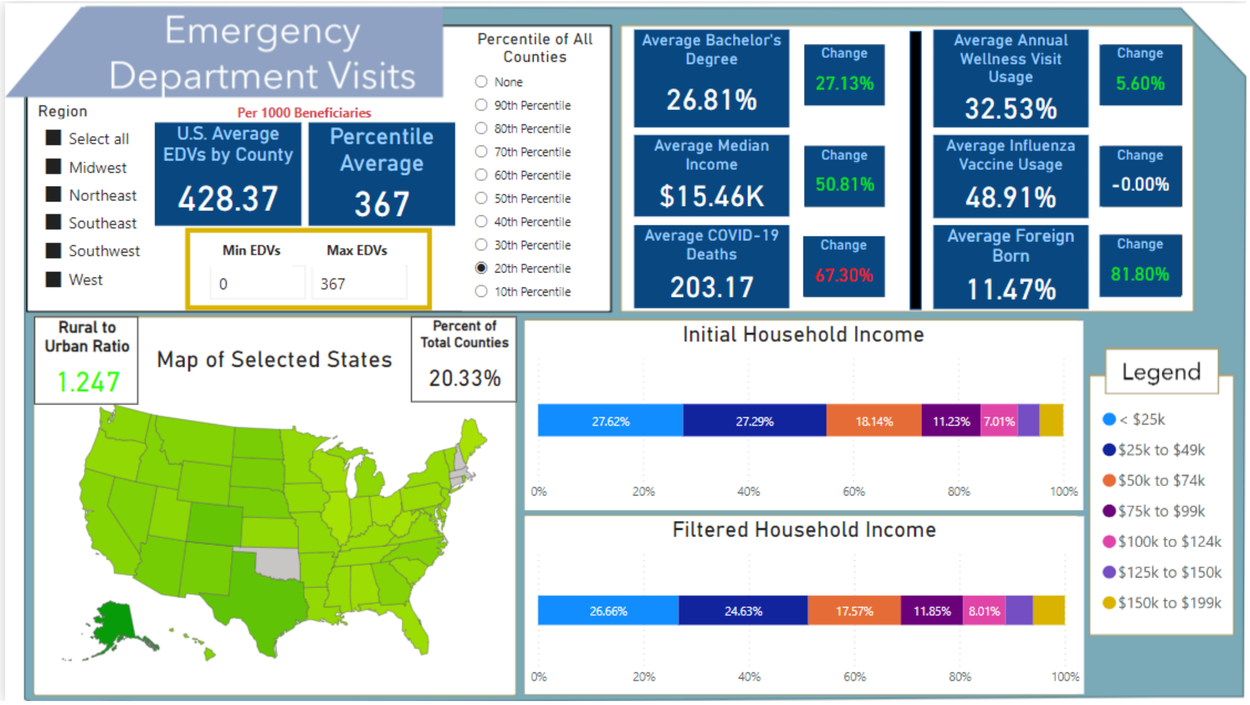


Figure 4.39: Under 20th Percentile Emergency Department Visits

The counties below the 20th percentile of emergency department visits had better SDoH averages than when hospitalizations were analyzed in the other percentile dashboard. Specifically, the median income of the counties below the 20th percentile increased by over 50% on average. The preventive service usage either increased or stayed roughly the same, but the COVID-19 deaths increased by over 60% on average. The increase in COVID-19 deaths and the lower Rural to Urban ratio possibly revealed that since there were more urban counties below the 20th percentile,

Chapter 5: Recommendations and Conclusion

Overall, the educational attainment, income, and counties in the western region were the most significant variables for predicting the health outcomes of a county. The information about a county's bachelor's degree attainment level and income breakdown gave information about other important variables like geography (urban or rural) and poverty level. Interestingly, it was observed that regions in the West have significantly fewer hospitalizations and ER visits compared to other regions, and the reason for the drastic difference is unclear. A possible factor was that the West region included Hawaii and Alaska, and both states had some of the lowest county-level hospitalization rates in the country, which may have pulled the average down in the West. In the logistic models, language other than English and urban were more prevalent variables, implying that these variables are more common in worse counties, but each variable alone may not consistently contribute to worse health outcomes.

When modeling high hospitalizations/ER, the team decided to classify high as being above the 80th percentile. However, experimenting with different cut-off points could lead to the discovery of different variables, and could make the models more accurate. Furthermore, the models produced were only trained on 2020 data, so looking into if models trained on other years retained similar variables would help narrow down on the most important variables to have information on. In addition, looking into counties that are consistently classified as high could help identify variables in these regions that could be used for future predictions.

Due to time constraints, the team was only able to analyze the significant variables mentioned above over a 3-year time span, 2019 to 2022. Therefore, for further analysis, observing how these variables impact health outcomes over a longer period using time series techniques can both validate the predictive quality of these variables and how these variables can be used to predict the changes in health outcomes over time. Additionally, it is unknown whether changes in these variables would change health outcomes, so an analysis of counties with changes in SDoH over time could show the effectiveness of improving SDoH in counties.

Finally, the analysis in this project only split up the data based on the counties' geographic areas. However, more can be done to divide these regions into groups with similar health outcomes trends. One approach is to use unsupervised cluster models like K-means and a Gaussian Mixture Model. These models group data into different clusters based on their similarity. The difference between K-means and Gaussian Mixture is that K-means assign the membership of each datapoint to a cluster using Euclidean distance whereas Gaussian Mixture uses probability. Trying both these models and comparing their results can help to find and validate common health outcomes trends.

Bibliography

Anderson, L. M., Scrimshaw, S. C., Fullilove, M. T., Fielding, J. E., & Normand, J. (2003). Culturally competent healthcare systems. *American Journal of Preventive Medicine*, *24*(3), 68–79. https://doi.org/10.1016/s0749-3797(02)00657-8

APHA. (2018, November 13). *Violence is a public health issue: Public Health is essential to understanding and treating violence in the U.S.* American Public Health Association. Retrieved September 10, 2022, from https://apha.org/policies-and-advocacy/public-health-policy-statements/policy-database/2019/01/28/violence-is-a-public-health-issue

Artiga, S., Hill, L., & Haldar, S. (2022, April 19). *Key facts on health and health care by race and ethnicity - social determinants of health*. KFF. Retrieved December 10, 2022, from https://www.kff.org/report-section/key-facts-on-health-and-health-care-by-race-and-ethnicity-social-determinants-of-health/

Brou, L. (2019, June 7). *Immigration as a Social Determinant of Health*. University of Colorado School of Medicine. Retrieved December 9, 2022, from https://medschool.cuanschutz.edu/family-medicine/about/news/disrupting-the-status-quo-blog/disrupting-the-status-quo/immigration-as-a-social-determinant-of-health

Bureau, U. S. C. (2022, August 3). *What We Do*. Census.gov. Retrieved October 25, 2022, from https://www.census.gov/about/what.html

Catalyst, N. E. J. M. (2017, December 1). *Social Determinants of Health (SDOH)*. NEJM Catalyst. Retrieved September 10, 2022, from https://catalyst.nejm.org/doi/full/10.1056/CAT.17.0312

Centers for Disease Control and Prevention. (2022, December 8). *Social Determinants of Health at CDC*. Centers for Disease Control and Prevention. Retrieved September 10, 2022, from https://www.cdc.gov/about/sdoh/index.html

Correlation and Regression with R. Simple linear regression. (2016). Retrieved October 25, 2022, from https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/R/R5_Correlation-Regression/R5_Correlation-Regression4.html

Dan Turtureanu, Demetri Pananos. (1968, October 1). *How the poisson distribution is used in regression?* Cross Validated. Retrieved October 25, 2022, from https://stats.stackexchange.com/questions/542155/how-the-poisson-distribution-is-used-in-regression

Douthit, N., Kiv, S., Dwolatzky, T., & Biswas, S. (2015). Exposing some important barriers to health care access in the Rural USA. *Public Health*, *129*(6), 611–620. https://doi.org/10.1016/j.puhe.2015.04.001

Egerter, S., Dekker, M., An, J., Grossman-Kahn, R., & Braveman, P. (2008, December). *The commission to build a Healthier America*. The Commission to Build a Healthier America. Retrieved December 14, 2022, from http://www.commissiononhealth.org/

Foundation, the A. E. C. (2010, January 1). *Reading by third grade*. The Annie E. Casey Foundation. Retrieved December 14, 2022, from https://www.aecf.org/resources/early-warning-why-reading-by-the-end-of-third-grade-matters

Frederick, T. J., Chwalek, M., Hughes, J., Karabanow, J., & Kidd, S. (2014). How stable is stable? defining and measuring housing stability. *Journal of Community Psychology*, *42*(8), 964–979. https://doi.org/10.1002/jcop.21665

Gulati, R. K., & Hur, K. (2021). Association between limited English proficiency and healthcare access and utilization in California. *Journal of Immigrant and Minority Health*, *24*(1), 95–101. https://doi.org/10.1007/s10903-021-01224-5

Han, B., Cohen, D., Derose, K., Li, J., & Williamson, S. (2018, January 12). *Violent Crime and Park Use in Low-Income Urban Neighborhoods*. American Journal of Preventive Medicine. Retrieved September 10, 2022, from https://doi.org/10.1016/j.amepre.2017.10.025

Hill, J., Rodriguez, D. X., & McDaniel, P. (2021, October 29). *Immigration status as a social determinant of health during covid-19*. Public Health Post. Retrieved December 9, 2022, from https://www.publichealthpost.org/research/immigration-status-covid-19/

Kenton, W. (2022). *What is social capital?* Investopedia. Retrieved September 10, 2022, from https://www.investopedia.com/terms/s/socialcapital.asp

Kushel, M. B., Gupta, R., Gee, L., & Haas, J. S. (2006). Housing instability and food insecurity as barriers to health care among low-income Americans. *Journal of General Internal Medicine*, *21*(1), 71–77. https://doi.org/10.1111/j.1525-1497.2005.00278.x

LexisNexis Risk Solutions, L. N. R. (2017, October 24). *Milliman MedInsight to use LexisNexis risk solutions socioeconomic health attributes to help enhance healthcare intelligence*. Milliman MedInsight to Use LexisNexis Risk Solutions Socioeconomic Health Attributes to Help Enhance Healthcare Intelligence. Retrieved October 25, 2022, from https://www.prnewswire.com/news-releases/milliman-medinsight-to-use-lexisnexis-risk-solutions-socioeconomic-health-attributes-to-help-enhance-healthcare-intelligence-300541930.html

Medicare Signed into Law. U.S. Senate: Medicare Signed into Law. (2019, July 12). Retrieved September 15, 2022, from https://www.senate.gov/artandhistory/history/minute/Medicare_Signed_Into_Law.htm#:~:text=On%20July%2030%2C%201965%2C%20President,Harry%20Truman%20had%20proposed%20it

Michele Ver Ploeg, V. B. (2009, June 25). *Access to affordable and nutritious food-measuring and understanding food deserts and their consequences: Report to Congress*. USDA ERS. Retrieved October 12, 2022, from http://www.ers.usda.gov/publications/pub-details/?pubid=42729

Milliman (2022). *The Work Matters.* Retrieved October 25, 2022, from https://www.milliman.com/en

Published by Statista Research Department, & 20, J. (2022, June 20). *Distribution of Medicare beneficiaries by age 2020*. Statista. Retrieved September 15, 2022, from https://www.statista.com/statistics/248035/distribution-of-medicare-beneficiaries-by-age/

Published: Feb 13. (2019, February 13). *An Overview of Medicare*. KFF. Retrieved December 14, 2022, from https://www.kff.org/medicare/issue-brief/an-overview-of-medicare/#:~:text=%23Medicare%20plays%20a%20key%20role,physician%20services%2C%20and%20prescription%20drugs

Singh, G. K., & Siahpush, M. (2006). Widening socioeconomic inequalities in US life expectancy, 1980–2000. *International Journal of Epidemiology*, *35*(4), 969–979. https://doi.org/10.1093/ije/dyl083

The college payoff: More education doesn't always mean more earnings. CEW Georgetown. (2022, May 16). Retrieved October 10, 2022, from https://cew.georgetown.edu/cew-reports/collegepayoff2021/

United States Government. (2022, November 29). *Rural Poverty & Well-being*. USDA ERS - Rural Poverty & Well-Being. Retrieved September 16, 2022, from https://www.ers.usda.gov/topics/rural-economy-population/rural-poverty-well-being/

United States Government. (2022, October 17). *Key Statistics & Graphics*. USDA ERS - Key Statistics & Graphics. Retrieved November 3, 2022, from https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/key-statistics-graphics.aspx

US Department of Health and Human Services. (n.d.). *Access to health services*. Access to Health Services - Healthy People 2030. Retrieved September 10, 2022, from https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/access-health-services#cit4

US Department of Health and Human Services. (n.d.). *Access to primary care*. Access to Primary Care - Healthy People 2030. Retrieved September 10, 2022, from https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/access-primary-care

US Department of Health and Human Services. (n.d.). *Crime and Violence*. Crime and Violence - Healthy People 2030. (2022). Retrieved September 10, 2022, from https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/crime-and-violence

US Department of Health and Human Services. (n.d.). *Employment*. Employment - Healthy People 2030. Retrieved December 14, 2022, from https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/employment#cit1

US Department of Health and Human Services. (n.d.). *Enrollment in higher education*. Enrollment in Higher Education - Healthy People 2030. Retrieved September 10, 2022, from https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/enrollment-higher-education

US Department of Health and Human Services. (n.d.). *Environmental Conditions*. Environmental Conditions - Healthy People 2030. (2022). Retrieved December 14, 2022, from https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/environmental-conditions

US Department of Health and Human Services. (n.d.). *High school graduation*. High School Graduation - Healthy People 2030. Retrieved September 10, 2022, from https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/high-school-graduation

US Department of Health and Human Services. (n.d.). *Language and literacy*. Language and Literacy - Healthy People 2030. Retrieved September 10, 2022, from https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/language-and-literacy#cit5

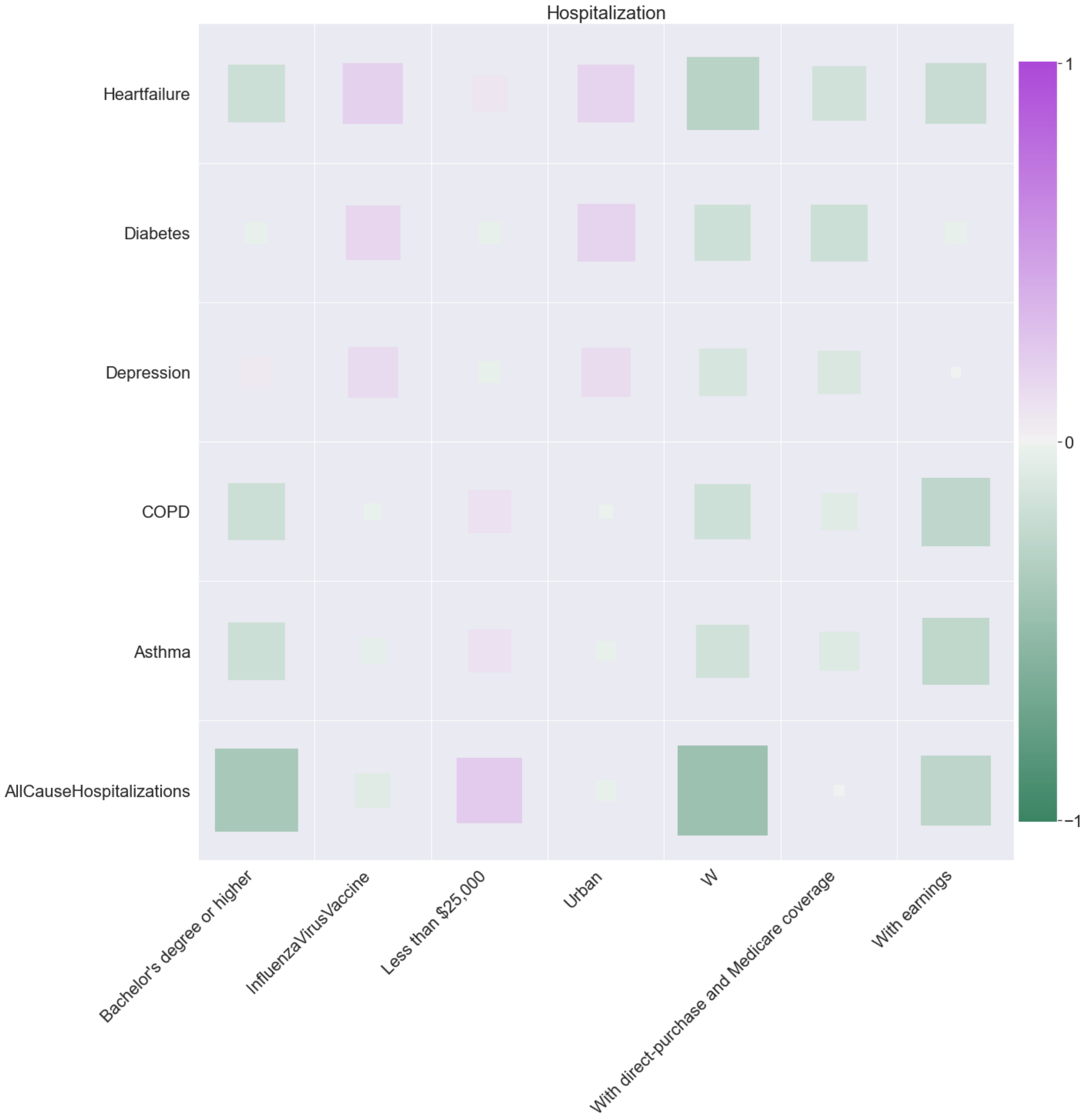
US Department of Health and Human Services. (n.d.). *Social cohesion*. Social Cohesion - Healthy People 2030. Retrieved September 10, 2022, from https://health.gov/healthypeople/priority-areas/social-determinants-health/literature-summaries/social-cohesion

US Department of Health and Human Services. (n.d.). *Social Determinants of Health*. Social Determinants of Health - Healthy People 2030. Retrieved September 10, 2022, from https://health.gov/healthypeople/objectives-and-data/social-determinants-health

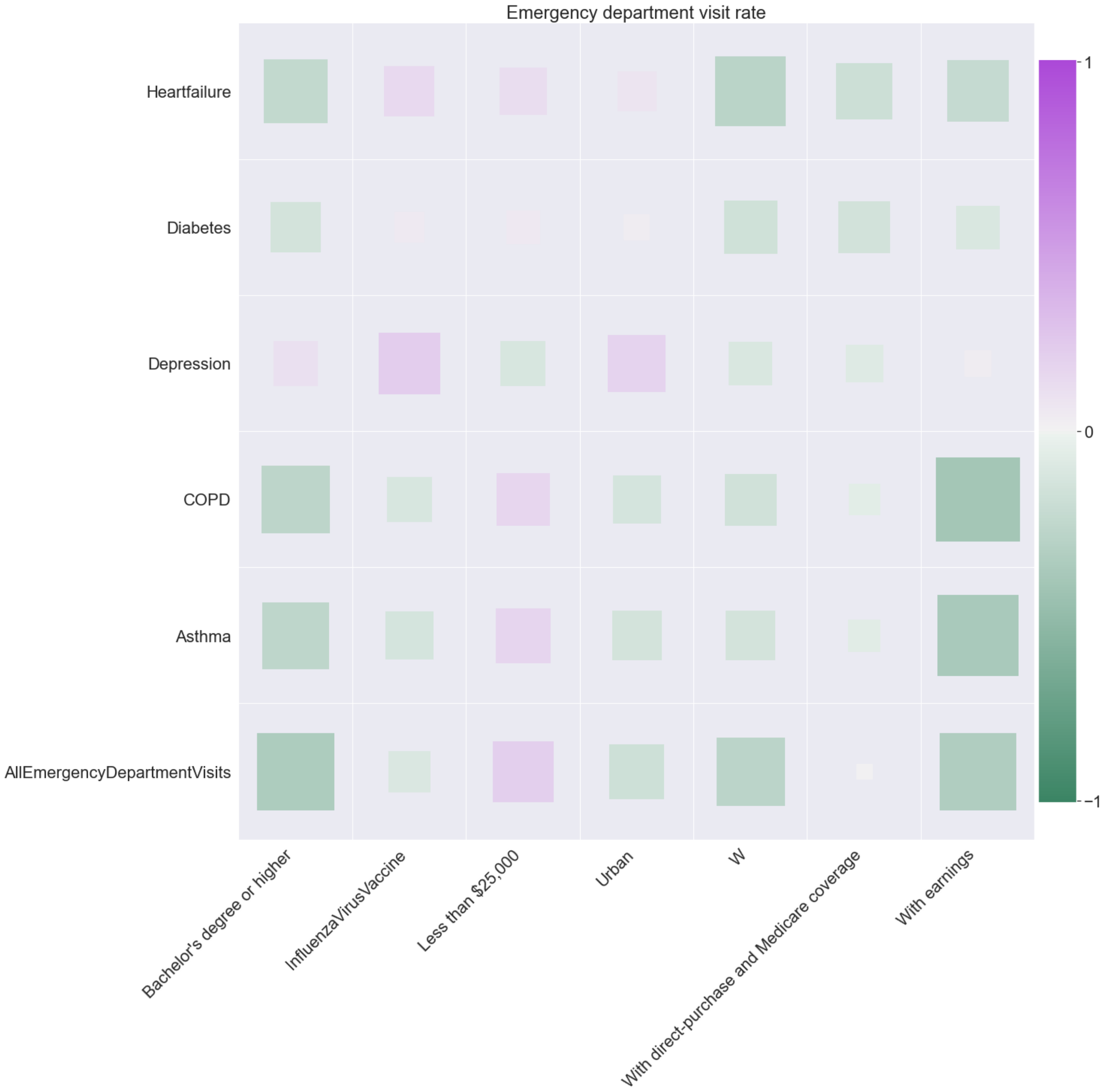
Vanhove, J. (2019, April 11). *Before worrying about model assumptions, think about model relevance*. Jan Vanhove :: Before worrying about model assumptions, think about model relevance. Retrieved October 25, 2022, from https://janhove.github.io/analysis/2019/04/11/assumptions-relevance#:~:text=The%20general%20linear%20model's%20assumptions,)%2C%20normality%2C%20and%20independence

Appendices

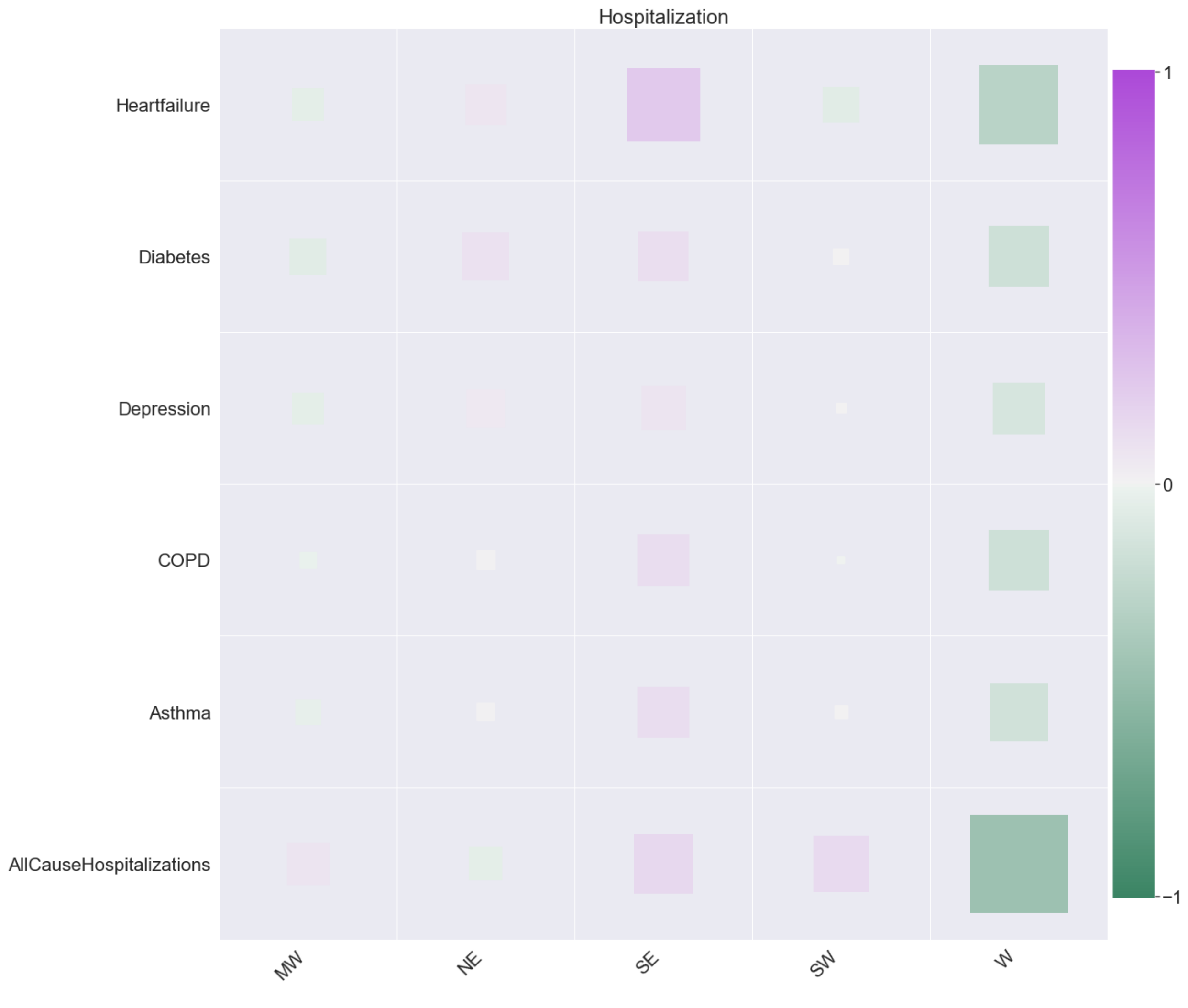
Appendix A: Significant variables correlation with hospitalization



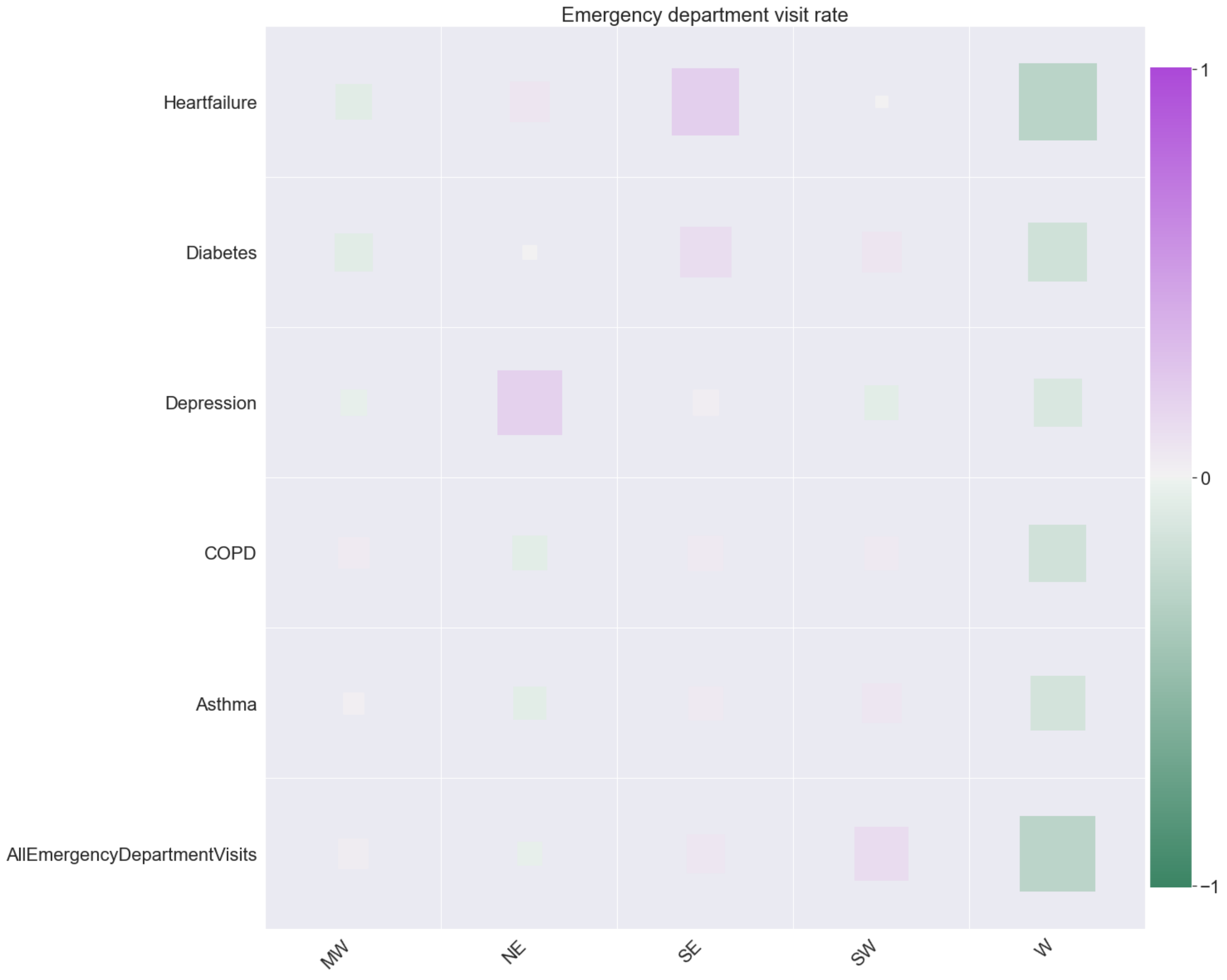
Appendix B: Significant variables correlation with ER visit



Appendix C: Region correlation with hospitalization



Appendix D: Region correlation with ER visit


Emergency department visit rate
Heartfailure
Diabetes
Depression
COPD
Asthma
AllEmergencyDepartmentVisits
MW
NE
SE
SW
W