

# **Investigating the Rates and Trends in Hypertension-related Cardiovascular Disease Mortality Among US Adults**

Group 18: Min Choi, Shakty Juarez, Meaghan Ramlakhan, Pinar Targil

## **Introduction**

Hypertension is a major risk factor for cardiovascular disease (CVD), contributing to a significant portion of CVD-related deaths in the United States. Despite public health efforts to manage hypertension through medication, lifestyle changes, and healthcare interventions, disparities in mortality rates persist across different demographic groups.

This study investigates the trends and factors influencing hypertension-related CVD mortality among US adults aged 35 and older from 2000 to 2019. Our analysis focuses on understanding:

- How demographic factors, such as **race/ethnicity, age, geography, and sex**, correlate with mortality rates.
- Whether public health initiatives have effectively reduced mortality rates over time.
- The geographic and demographic disparities that remain in 2019.
- The role of health insurance coverage in influencing mortality rates, to examine whether groups with higher insured rates experience lower CVD mortality.

## **Key Hypotheses**

- **Geographic disparities:** We expect southern US states to have higher mortality rates due to socioeconomic and healthcare access inequalities.
- **Racial/ethnic disparities:** We expect Black Americans to have higher mortality rates due to racially biased system inequities, including limited healthcare access and socioeconomic disadvantages.
- **Temporal trends:** We predict a decline in mortality over time, particularly in regions with strong public health initiatives.
- **Insurance coverage impact:** We anticipate that groups with higher health insurance coverage rates will have lower mortality rates, reflecting better access to preventive and medical care.

## **Data Description**

We used a data set from the **Centers for Disease Control and Prevention (CDC)**, covering hypertension-related CVD mortality from 2000 to 2019. The data includes over 1 million observations, classified by state, county, age group, race/ethnicity, sex, and year. To supplement the analysis, we incorporated an additional dataset containing health insurance coverage information, allowing us to explore the relationship between insurance status and CVD mortality.

## Key Variables

- **Year:** Ranges from 2000 to 2019.
- **Mortality Rate:** Age-adjusted hypertension-related mortality rate (deaths per 100,000).
- **Race/Ethnicity:** Stratified into **Black, White, Asian/Pacific Islander, American Indian/Alaska Native, and Hispanic** groups.
- **Sex:** Male and female.
- **Percent Insured:** The proportion of the population with health insurance coverage within each demographic group and region.
- **Percent Uninsured:** The proportion of the population without health insurance coverage, used to identify potential gaps in healthcare access.

## Methods

We analyzed the data using **dplyr** pipelines to create summaries, new variables, and filtered datasets. **SQL** was utilized for initial data organization, table creation, and merging operations to ensure consistency across the datasets. Our visualizations were primarily created using **ggplot2** to highlight trends and disparities across demographic and geographic factors. Our dual-dataset approach allowed us to identify patterns related to health insurance coverage and its potential influence on mortality rates, providing insights into how access to healthcare may affect outcomes.

## Data Cleaning

We began by filtering out irrelevant columns, including location details and data source identifiers, and removed any rows with missing values to ensure the data set was complete and reliable for analysis. For instance, we renamed key variables for clarity: “Stratification1” to “AgeGroup,” “Stratification2” to “Race,” and “Stratification3” to “Sex.” This cleaned data set facilitated our analysis on the effect of certain factors on mortality rates.

## Text Mining

We listed the tables in our database connection\_db and loaded the “combined” table, removing the MortalityRatePercentage column and updating “Ages 65 years and older” to “Ages 65+ years” in LabelGrouping. We renamed columns for clarity. For new\_data, we standardized terms in Topic and AgeGroup and converted Year to integers. We handled “NA” strings in MortalityRate by replacing them with actual NA values and then filtering and converting to integers. Finally, we created a DiffAvgMortalityRate column showing the difference from the average mortality rate.

## Investigating Data

We began by examining key demographic factors such as race, sex, and geographic location to understand their interaction and contribution to hypertension-related CVD mortality trends across populations.

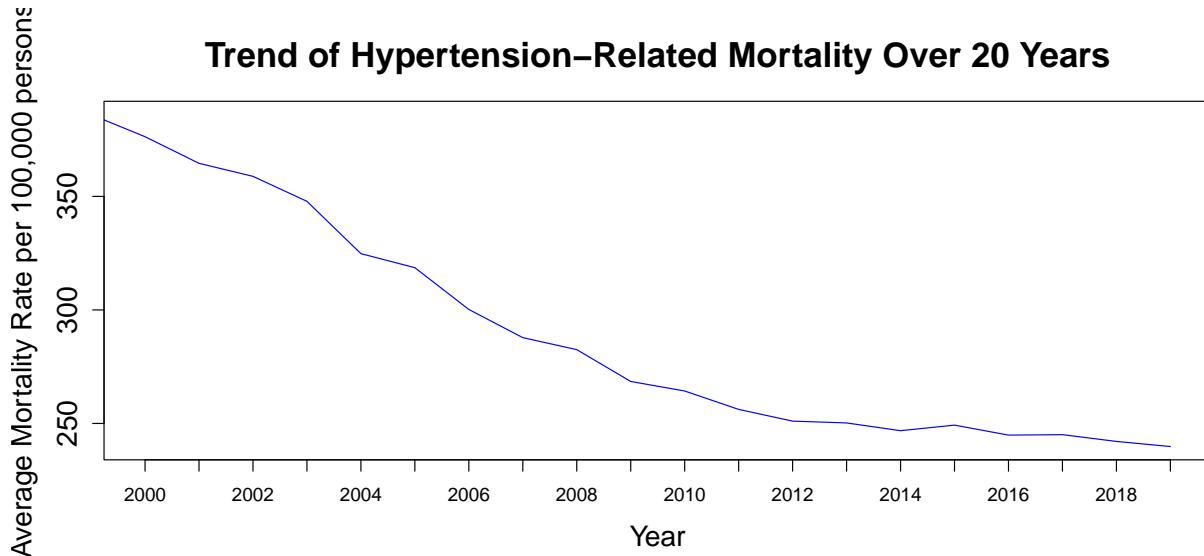
By analyzing two variables at a time, we can uncover patterns that indicate potential disparities or trends in CVD mortality. For instance, visualizing the relationship between mortality rates and race/ethnicity can help reveal whether specific racial groups are disproportionately affected. Similarly, examining geographic differences can highlight regions where hypertension-related deaths are concentrated.

In the following sections, we present visual analyses for several key variables:

- Hypertension-Related Mortality Rates Over Time:** We explore the overall trend in mortality rates from 2000 to 2019, highlighting any declines and periods of stagnation.
- Mortality Rate Changes by State (2000 vs. 2019):** By comparing state-level mortality rates in 2000 and 2019, we can see where the greatest improvements have occurred.
- Mortality Rates by Age and Sex:** This analysis provides insights into whether older populations or different sexes experience higher mortality rates.
- Mortality Rates by Race:** Understanding the racial disparities in CVD mortality helps identify which ethnic groups are most vulnerable.
- Health Insurance Coverage and Its Impact on Mortality Rates:** We analyze whether higher insurance coverage reduces hypertension-related CVD risks and identify populations that may benefit from increased access.

These visualizations lay the groundwork for more complex analyses later in the report, offering a clear picture of how hypertension-related CVD mortality is distributed across different demographics and geographies, and how health insurance coverage might mitigate these risks.

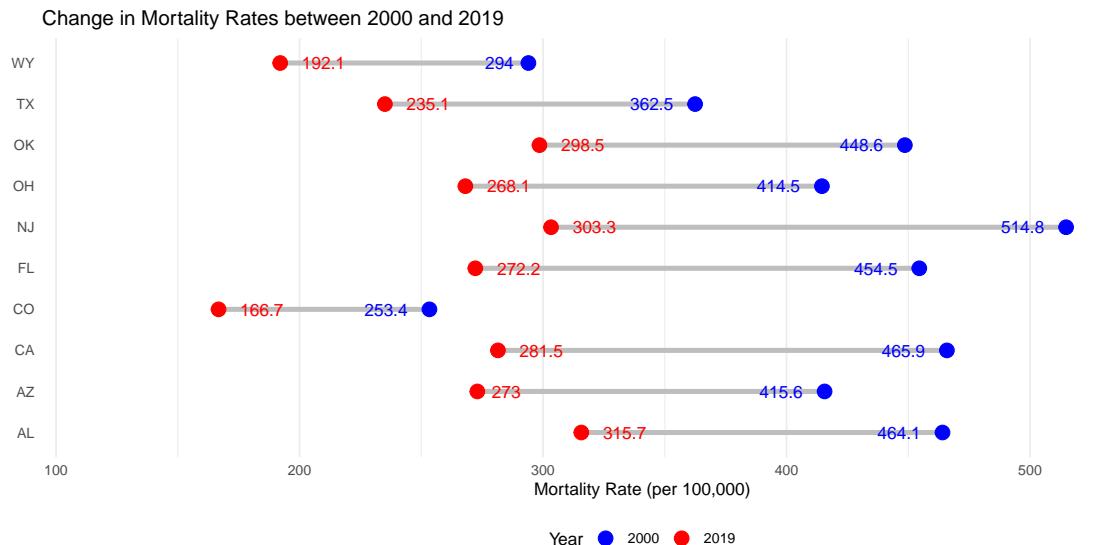
## 1. Line Plot of Mortality Over Time



This line plot illustrates the trend of hypertension-related mortality rates among U.S. adults aged 35 and older from 2000 to 2019. The y-axis represents the average mortality rate per 100,000 persons, while the x-axis shows the years of observation. Over this 20-year period, the graph highlights a clear overall decline in mortality rates, reflecting positive strides in managing hypertension. In 2000, the rate starts at around **350 deaths per 100,000 people** and steadily decreases to approximately **250** by 2019.

The most notable decline occurs between 2000 and 2010, where mortality rates drop significantly from around **350 to 280**. This decrease likely reflects improvements in healthcare interventions, increased awareness, and better management of hypertension. After 2010, the decline slows, with the mortality rate stabilizing around **250-270 per 100,000 people**. Although the trend continues downward, the flattening curve in recent years suggests that progress has slowed, possibly indicating the need for more focused public health measures to address remaining gaps.

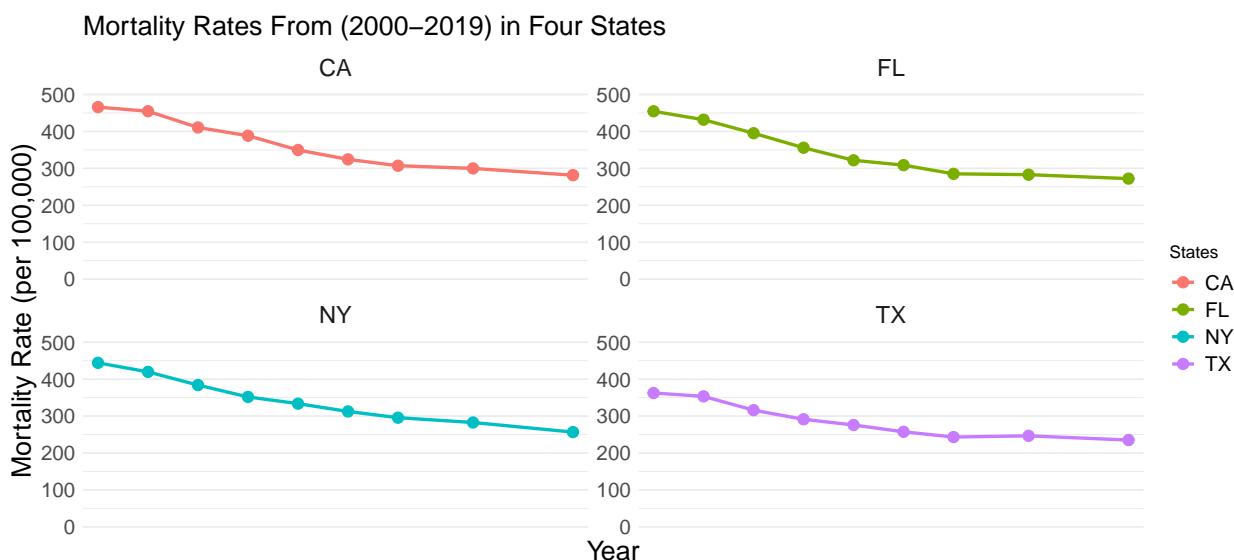
## 2. Dumbbell Plot on Change in Mortality Rates



This dumbbell graph helps visualize the change in mortality rates from the years 2000 to 2019 for various US states. The plot displays the red dot as being the average mortality rate for the year **2019** in the respective state, and the blue dot shows the average mortality rate for **2000**. Each gray line connecting the dots serves to highlight the extent of the decrease over time for each state.

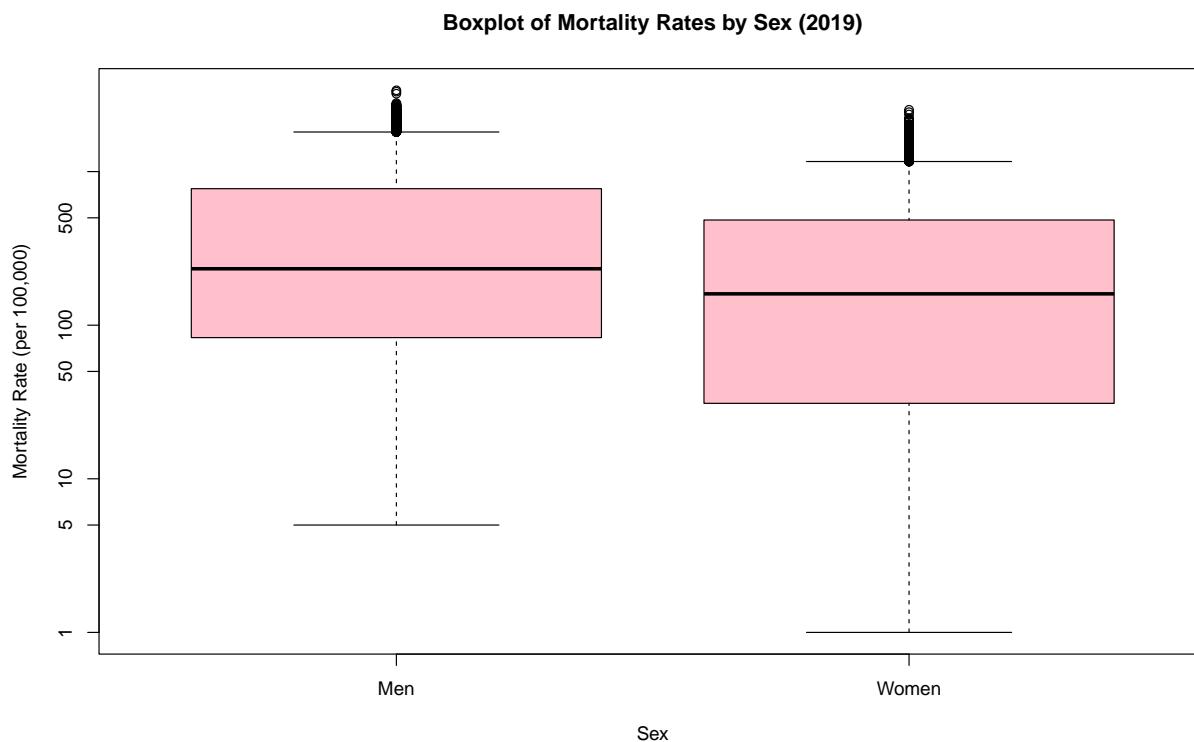
The plot highlights a consistent decrease in hypertension-related CVD mortality rates over the years. Upon further inspection of the graph, we can determine that **Wyoming (WY)**, **New Jersey (NJ)**, and **Colorado (CO)** show some of the most dramatic reductions in heart disease mortality. For instance, **Wyoming dropped from 451 to 358.8 per 100,000**, and **Colorado dropped from 380 to 298.5 per 100,000**. Similarly to the line plot, the consistent downward trend across all 10 states highlights the effectiveness of public health initiatives aimed at reducing heart disease mortality.

## 3. Faceting Plot for Mortality Across States



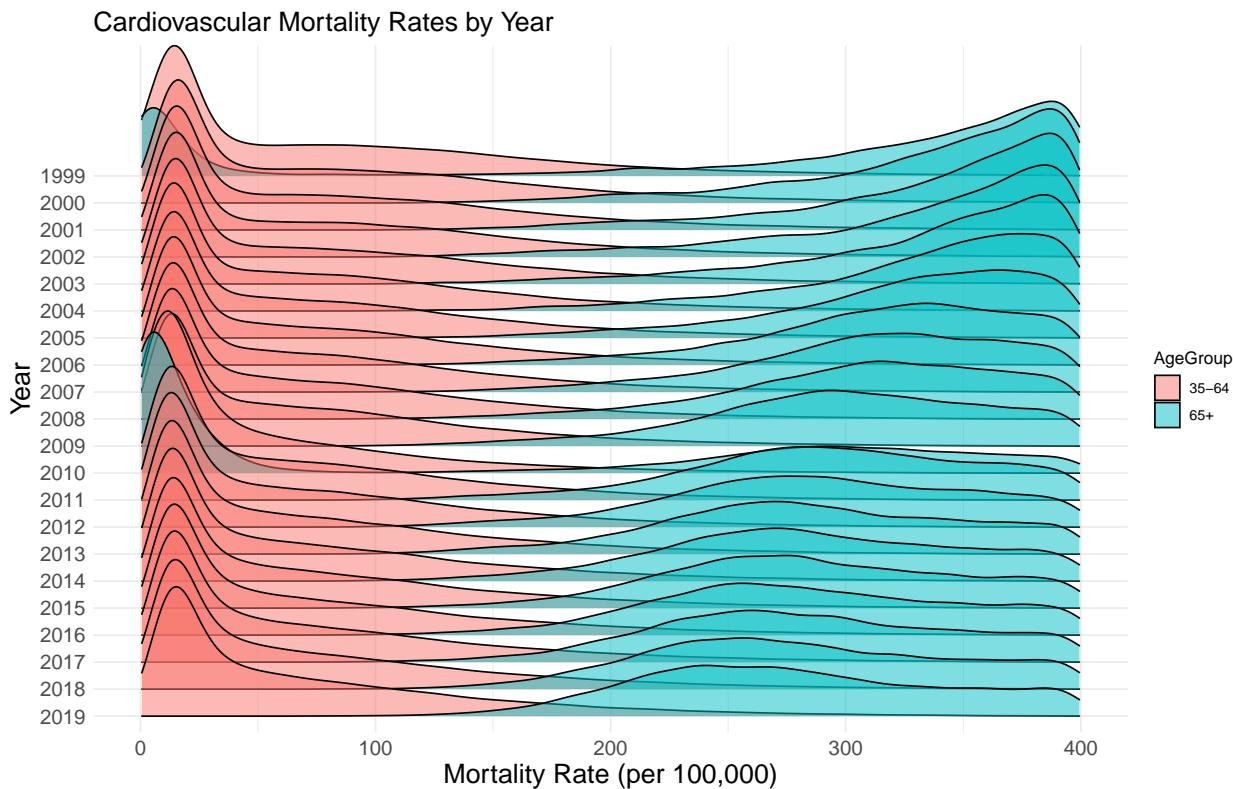
The faceting plot above helps visualize the hypertension-related CVD mortality rates per 100,000 individuals from **2000 to 2019** across **California (CA)**, **Florida (FL)**, **New York (NY)**, and **Texas (TX)**. Each subplot represents the trend for one state over the selected years, allowing for easy state-by-state comparisons all in one place. We can see that **all four states** have experienced an overall decline in heart disease mortality rates. Notably, New York and California have shown a more significant reduction in mortality compared to Florida and Texas. While all states demonstrate levels of improvement, the variation in the slopes of the lines suggests that there is a **disparity in heart disease-related public health care** among states. This could indicate differences in healthcare policies, preventive measures, and public health initiatives that contribute to varied outcomes in managing heart disease across these regions.

#### 4. Boxplot for Mortality Rates by Sex



This boxplot illustrates the mortality rates per 100,000 people in 2019, stratified by sex, and the y-axis is plotted on a logarithmic scale to better visualize our wide range of value and our outliers. Men tend to have higher mortality rates, as indicated by a higher median. Women exhibit a lower median mortality rate than men, suggesting that cardiovascular disease mortality is generally lower among females in 2019. Both distributions show a significant number of outliers (black circles), particularly for men. The clear difference in the median mortality rates between men and women can possibly be attributed to factors such as biological differences (i.e., earlier development of CVD, blood pressure, and cholesterol), lifestyle (smoking, alcohol, diet), and delayed health care for men.

## 5. Ridgeline Plot for Mortality over Time by Age Group



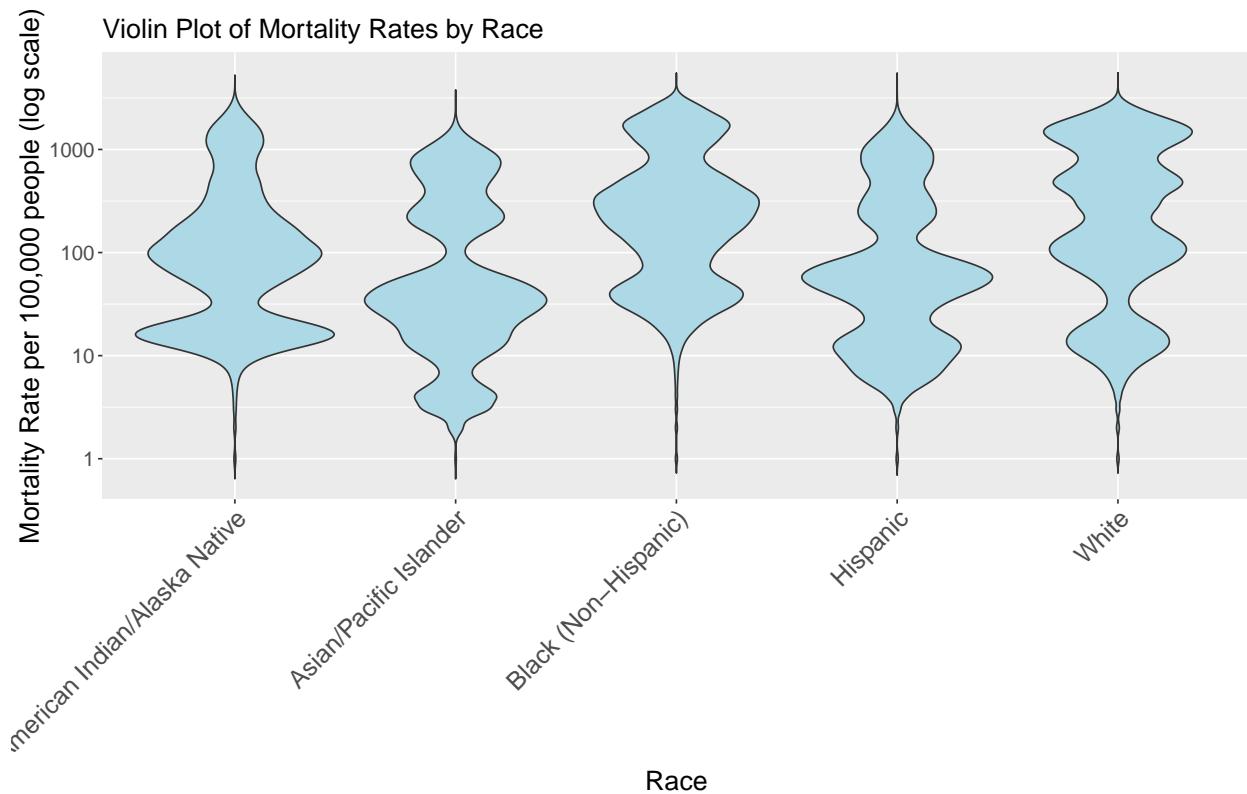
This ridgeline plot displays the distribution of hypertension-related CVD mortality rates (per 100,000) from **1999 to 2019**, stratified by year and categorized into two age groups: **Red (Ages 35–64 years)** and **Teal (Ages 65+ years)**. Each ridge represents the distribution of mortality rates for a specific year.

The plot illustrates that the cardiovascular mortality rate is consistently **higher for the 65+ age group** compared to the 35–64 age group, emphasizing that CVD is more prevalent among older adults. Over the 20-year period, we observe a decline in mortality rates for both age groups, as the peaks of the distributions gradually shift to the left.

The **65+ age group** exhibits a broader distribution, particularly in the earlier years, signifying greater variability in mortality rates across different states or regions. In contrast, the **35–64 age group** shows a narrower distribution, indicating more consistent mortality rates across states.

As the timeline progresses from **1999 to 2019**, the overall height and spread of the ridges diminish, showcasing that CVD mortality rates have generally decreased over time for both age groups. This downward trend highlights progress in public health initiatives and medical interventions aimed at reducing hypertension-related CVD deaths.

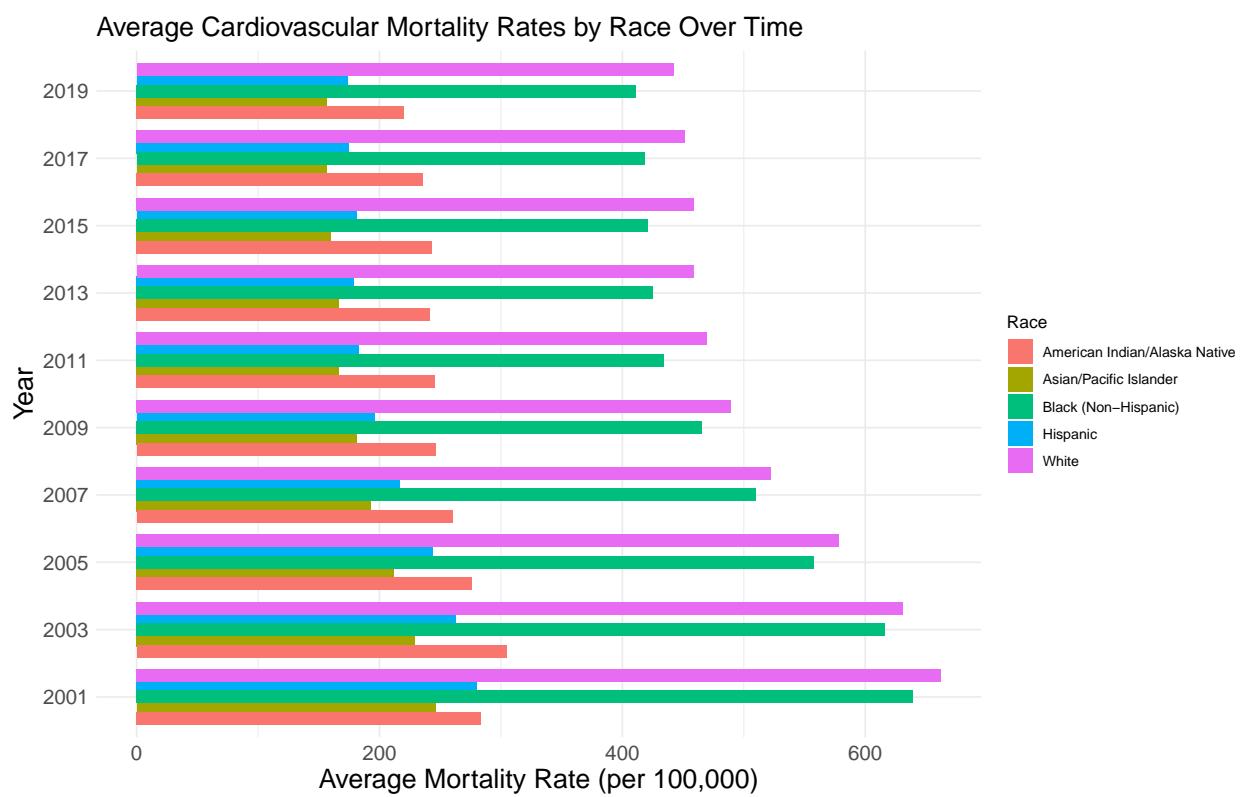
## 6. Violin Plot for Distribution of Mortality Rates by Race



This violin plot offers a view of the distribution of mortality rates per 100,000 people across various ethnic groups, using a logarithmic y-axis to handle the wide range of values. The width of each violin at a given y-value reflects the density of observations, emphasizing where mortality rates are most concentrated for each group.

The differences in violin widths at various points on the log scale highlight important disparities in mortality rates between ethnicities. Asian/Pacific Islander and Hispanic groups show relatively lower and more consistent rates, with concentrated mortality near 100. On the other hand, Black (Non-Hispanic) and White populations display broader distributions of higher mortality rates. The American Indian/Alaska Native group, with its dual peaks around 30 and 100, stands out for its more diverse range of mortality rates. These observations underscore the varying public health challenges faced by different ethnic groups, with some experiencing more concentrated and lower mortality rates, while others have wider distributions and higher overall rates.

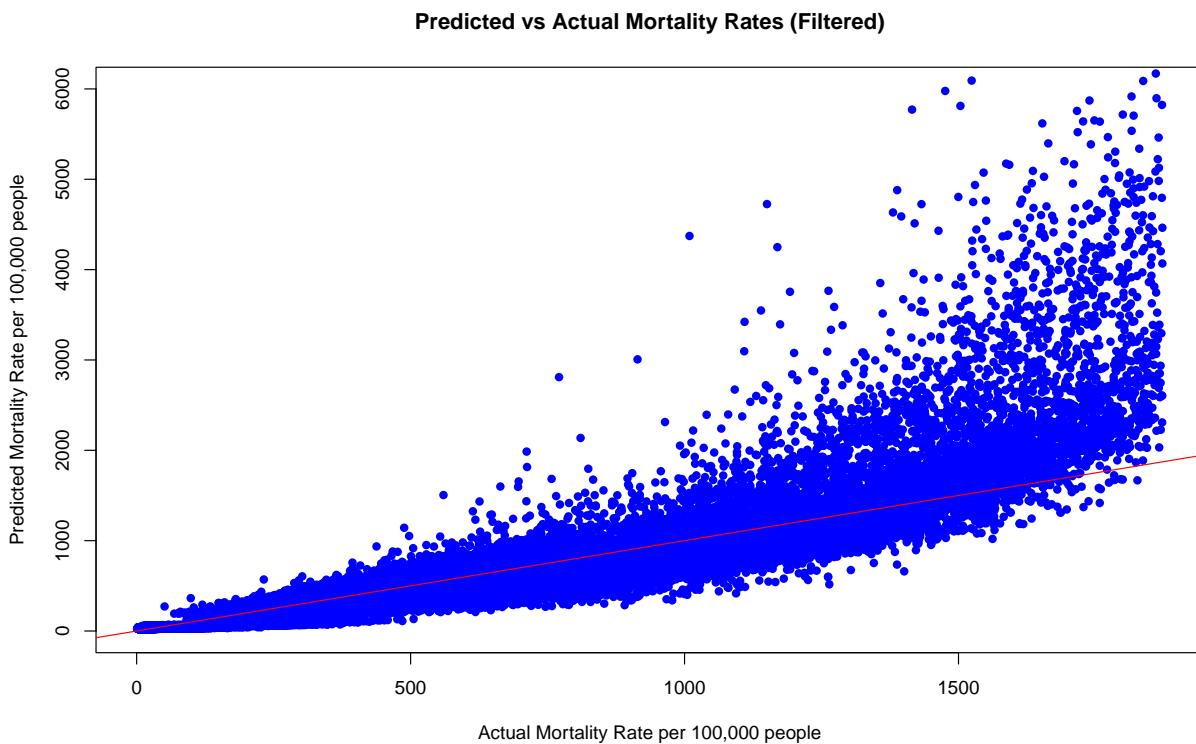
## 7. Clustered Bar Plot for Mortality Rates By Race



This clustered bar plot displays the average hypertension-related CVD mortality rate (per 100,000 people) by race/ethnic group over time from 2001 to 2019. From the plot, we can see that **Black and White individuals** consistently exhibit the **highest mortality rates** across all years, showing notably longer bars compared to others, especially in the early 2000s. In contrast, **Asian/Pacific Islanders** have consistently maintained the lowest cardiovascular mortality rates over the entire period.

Over the years, the mortality rates for most racial/ethnic groups show a general decline, with the most pronounced reductions observed in the Black and White populations. Despite the overall decline and the gap between Black and White populations and other racial groups narrowing over time, the plot underscores that disparities remain, as the **Black and White populations** continue to face higher average mortality rates. This persistent difference highlights the importance of targeted public health interventions to address cardiovascular health disparities among racial and ethnic groups.

## 8. Regression Analysis



This plot illustrates the relationship between the actual and predicted mortality rates from a linear regression model for the year 2019. The model was built using log-transformed data and focuses on predicting county-level mortality rates based on factors such as age and race and the provided confidence limits. The data was cleaned and filtered to remove outliers.

In the context of this project, the linear regression analysis provides a starting point for understanding how well the selected variables explain variations in mortality rates across different counties. The plot shows that while the predicted mortality rates generally follow the actual rates (as seen in the cluster of points near the reference line), the model tends to overpredict in regions with higher actual mortality rates (e.g., counties with rates exceeding 800 per 100,000 people). This indicates that the model may not fully capture the complexities associated with these high-mortality regions, possibly due to unaccounted socioeconomic or healthcare-related factors.

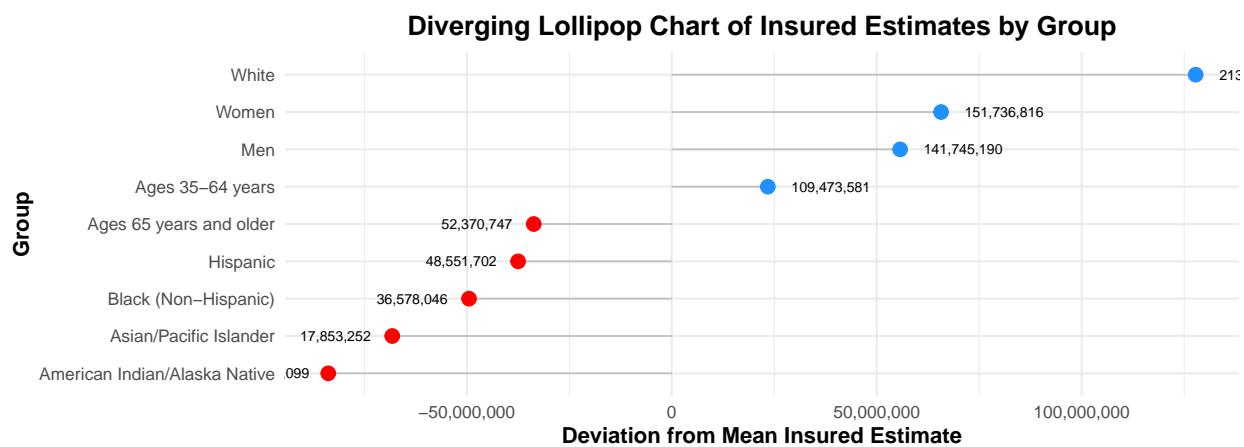
The red reference line represents a perfect fit where the predicted rates would match the actual mortality rates exactly. However, the divergence of points, especially at the higher end of the mortality spectrum, suggests the model might not be accurately predicting the highest-risk counties.

Overall, this regression analysis offers valuable insights into how to proceed with our analysis by highlighting limitations in the model's ability to account for extreme cases. The insights from this regression analysis could inform future public health interventions by identifying areas where the current model underperforms and where more targeted analysis is needed to understand the underlying causes of higher mortality rates.

## Relationship of Health Insurance with CVD

In this study, we also aimed to understand how health insurance coverage intersects with CVD outcomes, focusing on hypertension-related mortality. Understanding the distribution of insured estimates is not only vital for assessing overall coverage but also plays a critical role in examining health outcomes, including those related to CVD. Health insurance coverage is often linked to the accessibility and quality of healthcare services, which can directly influence the management and prevention of chronic conditions like CVD. To deepen our analysis, we integrated health insurance data with our primary dataset on hypertension-related CVD mortality. These insights set the stage for the following visual analyses, which delve deeper into the relationships between health insurance coverage and hypertension-related CVD mortality across different demographics.

### 9. Diverging Lollipop Chart for Insured Estimates by Group



The diverging lollipop chart provides insights into the distribution of insured estimates across demographic groups. The **White**, **Women**, and **Men** groups show positive deviations from the mean, indicating higher insured estimates compared to the average. This suggests that these groups benefit from better access to employment-related insurance, higher income levels, or other socioeconomic advantages.

Conversely, groups such as **American Indian/Alaska Native**, **Asian/Pacific Islander**, **Black (Non-Hispanic)**, **Hispanic**, and **Ages 65 years and older** exhibit negative deviations from the mean, highlighting potential disparities in insurance coverage. The **American Indian/Alaska Native** group shows the largest negative deviation, indicating significant challenges in accessing insurance. These trends may reflect social and economic barriers, such as employment in sectors without comprehensive insurance or systemic inequalities.

The age analysis shows that while the **Ages 35-64 years** group has insured estimates above the mean, the **Ages 65 years and older** group falls below, suggesting potential gaps in senior coverage despite programs like Medicare.

Overall, these findings highlight the need for targeted policies to address insurance disparities, particularly for underserved groups. Ensuring equitable access to insurance is crucial for improving healthcare access and reducing disparities. Further analysis of these deviations could inform effective strategies to promote insurance equity, especially when considering how insurance coverage impacts health outcomes like CVD mortality.

## Relationship of Behavioral Trends and CVD

From our analysis thus far, we determined that insurance coverage and basic demographic factors alone do not fully explain CVD mortality rates. To better understand these trends in relation to lifestyle and health, we incorporated data from **County Health Rankings & Roadmaps**, which provides behavioral statistics for each U.S. county, such as obesity rates, unemployment rates, insurance rates, smoking rates, education rates, and particulate matter rates. To integrate this with our analysis, we added a column to the data set showing the average CVD mortality rate for each county. To explore the relationships between each variable and mortality rates, we created scatter plots with regression lines for each variable. Additionally, to identify the variables most strongly correlated with mortality, we generated a table ranking these correlations.

### 10. Scatter Plots of Variables vs. Average Mortality Rate per County

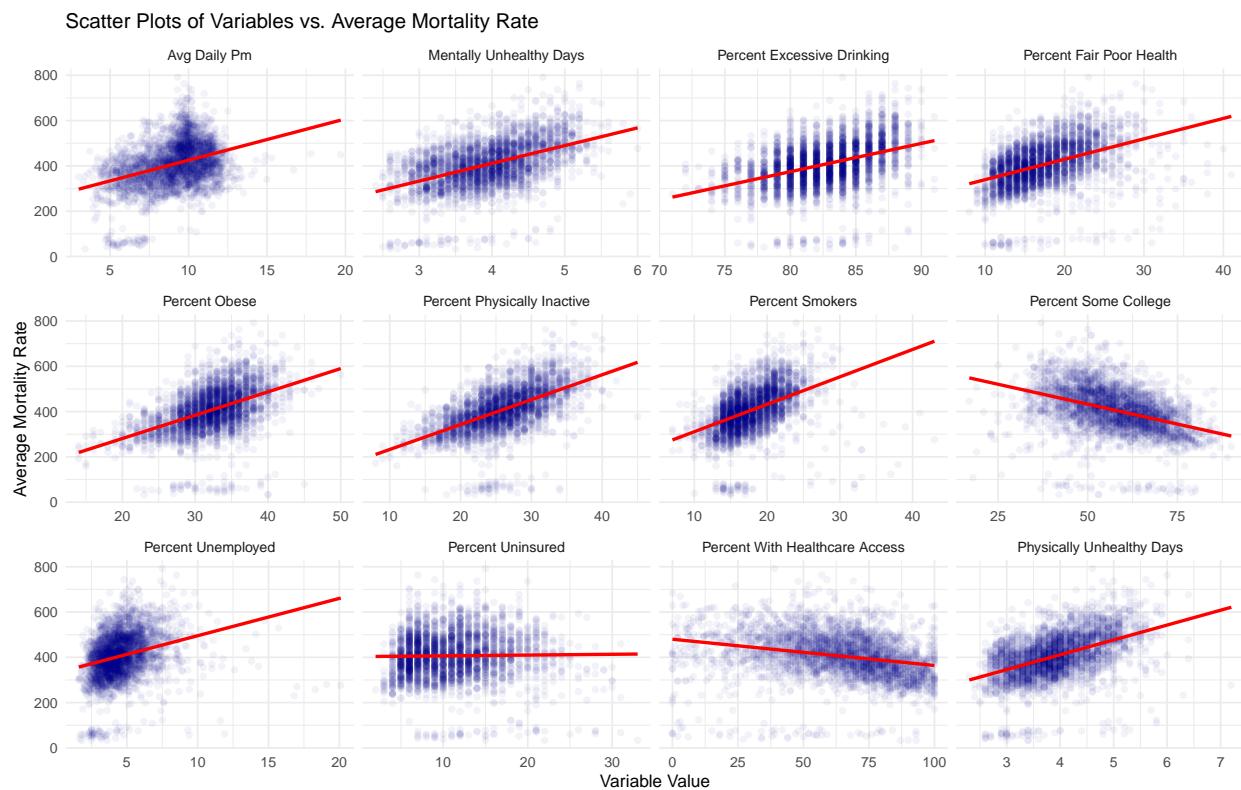
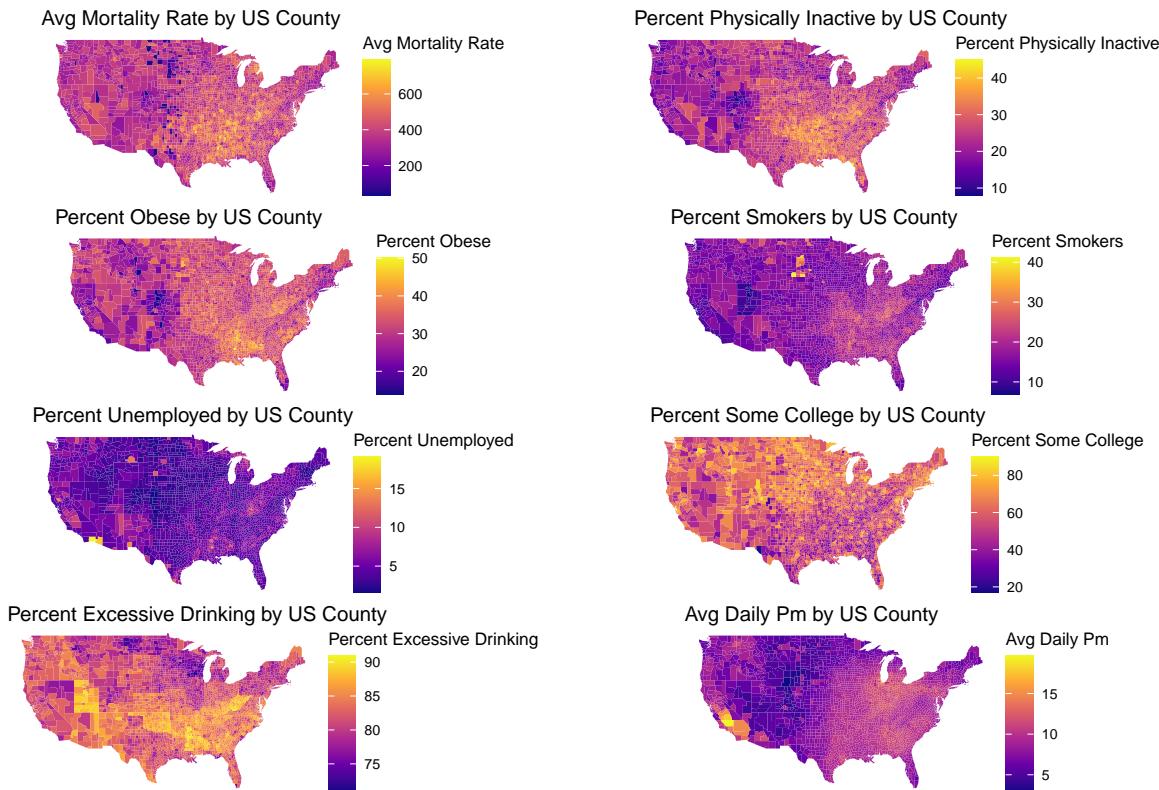


Table 1: Variables Most Correlated with Average Mortality Rate

	Correlation
percent_physically_inactive	0.57
percent_obese	0.49
mentally_unhealthy_days	0.48
percent_smokers	0.48
percent_excessive_drinking	0.40
avg_daily_PM	0.36
percent_unemployed	0.32
percent_uninsured	0.04
percent_with_healthcare_access	-0.26
percent_some_college	-0.42

Our analysis revealed that the strongest predictors of CVD mortality are physical inactivity (0.57), obesity (0.49), mentally unhealthy days (0.48), smoking (0.48), and college education (-0.42). To further understand how these factors vary geographically, we created U.S. county maps for each variable, highlighting their distribution across the country.

## 11. US Maps for Variables by County



These maps reveal distinct geographic patterns, with Southern states—particularly the **South East and Central regions**, including Mississippi and Alabama—standing out for their high rates of CVD mortality, physical inactivity, obesity, and smoking, along with their low college education rates. While these variables are strongly correlated with CVD mortality, it is important to note that correlation does not imply direct causation. Factors such as smoking and obesity are well-established risk factors, yet the elevated mortality rates point to deeper systemic issues beyond individual behaviors. These include longstanding **economic disparities, limited access to healthcare, and insufficient education** on healthy lifestyles, all of which compound the public health challenges in these regions. The consistently high mortality rates in these areas reflect the urgent need for comprehensive public health reforms that address both the **social and structural determinants of health**. Policies that focus on improving healthcare access, increasing education on healthy habits, and addressing underlying economic inequalities are crucial steps in reducing CVD mortality and closing health disparities.

## Killer Plot



Figure 1: Killer Plot (State: Alabama, Mortality Rate Range: 300-800, Health Factor: All)

Our **Killer Plot** is an interactive visualization of our county-level CVD mortality data, designed to highlight regional health disparities and key behavioral and socioeconomic health factors. Users can explore state and county trends, filter by mortality rates, and focus on predominant health factors such as smoking, obesity, inactivity, or alcohol use. Each county is represented by one of four distinct symbols,—a cigarette, couch, beer bottle, or hamburger—corresponding to the leading health factor, with icon size reflecting the mortality rate. Larger icons signify higher mortality, offering an intuitive way to assess health risks and patterns.

### Regional Highlights:

- **Southern States** (e.g., Alabama, Mississippi): These states consistently show elevated mortality rates. Counties like DeKalb and Lamar in Alabama or Panola and Lawrence in Mississippi feature large icons, signaling severe health challenges. Key contributing factors include high rates of obesity and inactivity, and in some areas, elevated smoking rates. Cultural and dietary habits, coupled with insufficient preventive care, can exacerbate these issues.
- **California:** Throughout our analysis, California has consistently stood out for its low CVD-related mortality rates, likely due to robust healthcare access, public health initiatives, and healthier lifestyle habits. Decades of aggressive anti-smoking campaigns have significantly reduced smoking prevalence, contributing to lower mortality rates associated with CVD.
- **Alaska:** While Alaska has fewer counties, distinct patterns emerge. Smoking prevalence remains a key health concern in many parts of Alaska, particularly in rural regions where healthcare access is limited. Despite this, the overall mortality rates in some areas remain relatively low due to lower population density and, in some cases, traditional diets that may mitigate some CVD risks. However, geographic and socioeconomic isolation in regions like Nome and Bethel can exacerbate smoking-related health outcomes.

## **Impact:**

The Killer Plot helps identify unexpected trends, such as regions with lower-than-expected mortality despite challenges or areas with disproportionately high rates. By visualizing these disparities, our interactive plot can support data-driven decisions, directing resources to regions in need and emphasizing the importance of tailoring public health strategies to address specific regional challenges. It bridges the gap between raw data and actionable solutions, contributing to the fight against cardiovascular disease across diverse communities.

## **Conclusion**

This analysis confirms that while there has been significant progress in reducing hypertension-related CVD mortality over the past two decades, racial and geographic disparities persist. **Black Americans** and residents of **southern states** continue to experience disproportionately higher mortality rates, suggesting that targeted public health interventions are needed.

The integration of health insurance data provided valuable insights into its relationship with mortality rates. However, as shown by our analysis of the top correlated variables—such as **physical inactivity**, **obesity**, and **smoking**—insurance coverage alone cannot address the multifaceted drivers of CVD mortality. For instance, **Black Americans**, despite having relatively high insurance coverage, still face elevated mortality rates. This highlights the critical influence of **socioeconomic conditions**, **access to education on healthy lifestyles**, **healthcare quality**, and **systemic inequities**. Similarly, groups with lower insurance coverage, such as **American Indian/Alaska Native** populations, also experience significant risks, reinforcing the need for policy solutions that address both coverage and equitable access to care.

Future research should incorporate a more comprehensive analysis of **socioeconomic factors**, **regional health behaviors**, and **education levels** to better understand the root causes of these disparities. Evaluating the effectiveness of interventions targeting behaviors like smoking and obesity and improving access to high-quality care in regions with high mortality rates will be essential. Ultimately, healthcare policies must focus not only on expanding coverage but also on improving the equity and quality of care to reduce hypertension-related CVD mortality and address the persistent racial and geographic disparities in health outcomes.

## **References**

Centers for Disease Control and Prevention, <https://catalog.data.gov/dataset/rates-and-trends-in-heart-disease-and-stroke-mortality-among-us-adults-35-by-county-a-2000-45659>

County Health Rankings & Roadmaps, <https://www.countyhealthrankings.org/health-data/methodology-and-sources/data-documentation/national-data-documentation-2010-2022>

United States Census Bureau, <https://data.census.gov/table/ACSST1Y2019.S2701?q=health%20insurance%202019>