Kelly Cao and Kirsten Gilbert Report

Socioeconomic and geographic determinants of HIV Prevalence

Kelly Cao and Kirsten Gilbert

2024-12-01

# 1. Abstract

In this writing we will explore the questions of how Human Immunodeficiency Virus(HIV) testing rates and related behaviors varied across U.S. states over time, and are there observable disparities based on demographic factors such as age, gender, or race/ethnicity? Human immunodeficiency virus, is an illness that comprises the efficiency of the immune system, which can impair the body’s response to other diseases or infections (U.S. Department of Health & Human Services, 2023). In the long term, the illness can also cause the body to become more at risk of contracting heavy chronic illnesses such as cancer, cardiovascular disease, as well as neurological disease. HIV is an illness that requires quick and efficient treatment as if it were to be left untreated, it can progress to even more severe stages that form into conditions that have greater strain on the body(U.S. Department of Health & Human Services, 2023). With this in mind also brings the next idea of the overall stigma around Human Immunodeficiency Virus(HIV). The most common misconception that is around HIV is that of it having disproportionate impacts on a certain demographic of people. This idea is deeply rooted from a combination of historical as well as societal factors alongside years of the spread of misinformation, thus putting said demographic of people at a higher public health risk. This is one of the causes that allow for research literature such as this one is important in the name of highlighting and progressing the current state of the health of society as a whole. For this specific writing piece, the focus will be mainly on the United States of America, we also have the intention of putting an emphasis on the state of Georgia, the action of us presenting a lens that is more local allows for the relevance and sense of immediacy to be felt by whomever it is that is reading this. Writings such as these are very important in the name of creating a catalyst in the discussions concerning not only HIV but other illnesses that find their way into the communities of people around the world, thus allowing for the challenge of public health epidemiological issues to be overcome.

# 2. Introduction

In the fields of healthcare and epidemiology, it is imperative to consider the multitude of factors that may influence how individuals access and receive treatment for various disease states. There can be numerous demographic factors that play a part in this, them being race, gender, sexual orientation, as well as income. An illness that has a notable reputation for having strong effects towards those who are lower in the socioeconomic hierarchy would be that of Human Immunodeficiency virus(U.S. Department of Health & Human Services, 2023). For decades, HIV has been pervasive across many communities in the United States, often progressing to Acquired Immunodeficiency Syndrome (AIDS) if left untreated. In this report, said factors will be used as a means of analysis, which will incorporate demographic factors along with whether or not said people are facing different forms of treatment if any at all. It is commonly known that individuals from minority communities are at an elevated risk of contracting HIV, and these same communities often experience delays in receiving treatment, contributing to worse health outcomes. The primary objective of this analysis is to explore the relationship between demographic factors and HIV treatment access, specifically within the context of the United States, with particular attention to Georgia. By analyzing public health data, examining current legislation, and reviewing relevant academic and policy literature, this report seeks to shed light on the disparities in HIV care and treatment. The data is from the year of 2013, thus allowing for everything to be relevant in the name of time. This is important in the name of progressing public health in a direction that motivates change. Utilizing information that is recent can also allow for the credibility of this report to be supported by a factual foundation that provides support in the name of being able to easily illustrate the presented claim, which only allows for an even greater impact. In the name of implementing ideas and possible solutions to the issue at hand, this report will also dwell into analyzing the current policies that are currently being enforced, and comparing it to the reality that is being faced by those who are in the affected regions. An example being Atlanta, Georgia, which serves as a particularly pertinent case study due to its status as a hub of racial, economic, and cultural diversity. The city has long been a focal point for African American culture and continues to be an influential center for various socio-cultural movements. The diversity within the city means that a wide range of demographic groups are affected by HIV, and understanding how these factors interplay is essential for crafting effective public health strategies.The data that will be used during this analysis will also aid in answering the question of, what is the relationship between education level and HIV testing rates, along with other questions that will aid in the presentation of a conclusion. We will also be considering comparative data, by questioning exactly what demographic factors in Georgia are most strongly associated with HIV testing rates, and how these compare to national trends. After this, we will then be analyzing answers to conclusion questions such as can demographic and geographic factors predict HIV testing rates in Georgia? Through a careful and thorough analysis, this report hopes to provide a clearer picture of the disparities in HIV treatment and offer concrete recommendations for closing the gap in healthcare access, ultimately improving health outcomes for all communities, particularly those at greatest risk. Igniting a dialogue that reaches everyone who is affected as well as those who aren’t affected. Thus allowing for evidence-based recommendations that can influence policymakers, healthcare providers, epidemiologists, as well as new lessons being taught within a classroom. Thus addressing the complexities of the issue that surrounds Human Immunodeficiency virus, and fostering an inclusive health strategy that isn’t affected by race, gender, sexual orientation, and income.

# 3. Method

The following steps were performed using R version 4.3.1, RStudio, and various R packages to perform data exploration and analysis. Github and git are used for version control. ## Data Procurement The following data used for analysis was procured as a dataset provided from the Center for Disease Control (CDC). It contains HIV-AIDS prevalence data procured from lans-line and cell-phone use through state based surveillance systems known as the Behavioral Risk Factor Surveillance System (BRFSS). The data has been collected from 2011 to present and is continuously updated annually. The data is downloaded from the CDC database as a CSV file with over 30,000 rows and 27 columns. Some notable columns include: HIV prevalence, location, stratification group, and stratification value.

## 3.1 Data Processing

The data is loaded into RStudio as a csv file with base R functions. The dplyr and stringr packages are used to clean the data for further analysis. Out of the 27 possible columns, 13 columns were selected and filtered out for empty or NA rows. The column names were remapped and renamed for ease of future use and greater clarity. From that, every data point in which the ‘state\_name” == ‘Georgia’ is used and filtered further as a separate object in the RStudio environment. A summary of the cleaned US data and the Georgia data is reviewed and then saved as a separate CSV files under the ‘data’ directory.

## 3.2 Exploratory Data Analysis

When performing data exploration, R packages, such as here, dplyr, and ggplot were commonly used to determine distribution of predictors for the HIV prevalence outcome. This was performed in two main parts: data manipulation and visualization through ggplot graphs.

## 3.3 Statistical Data Analysis

### 3.3.1 Linear Regression

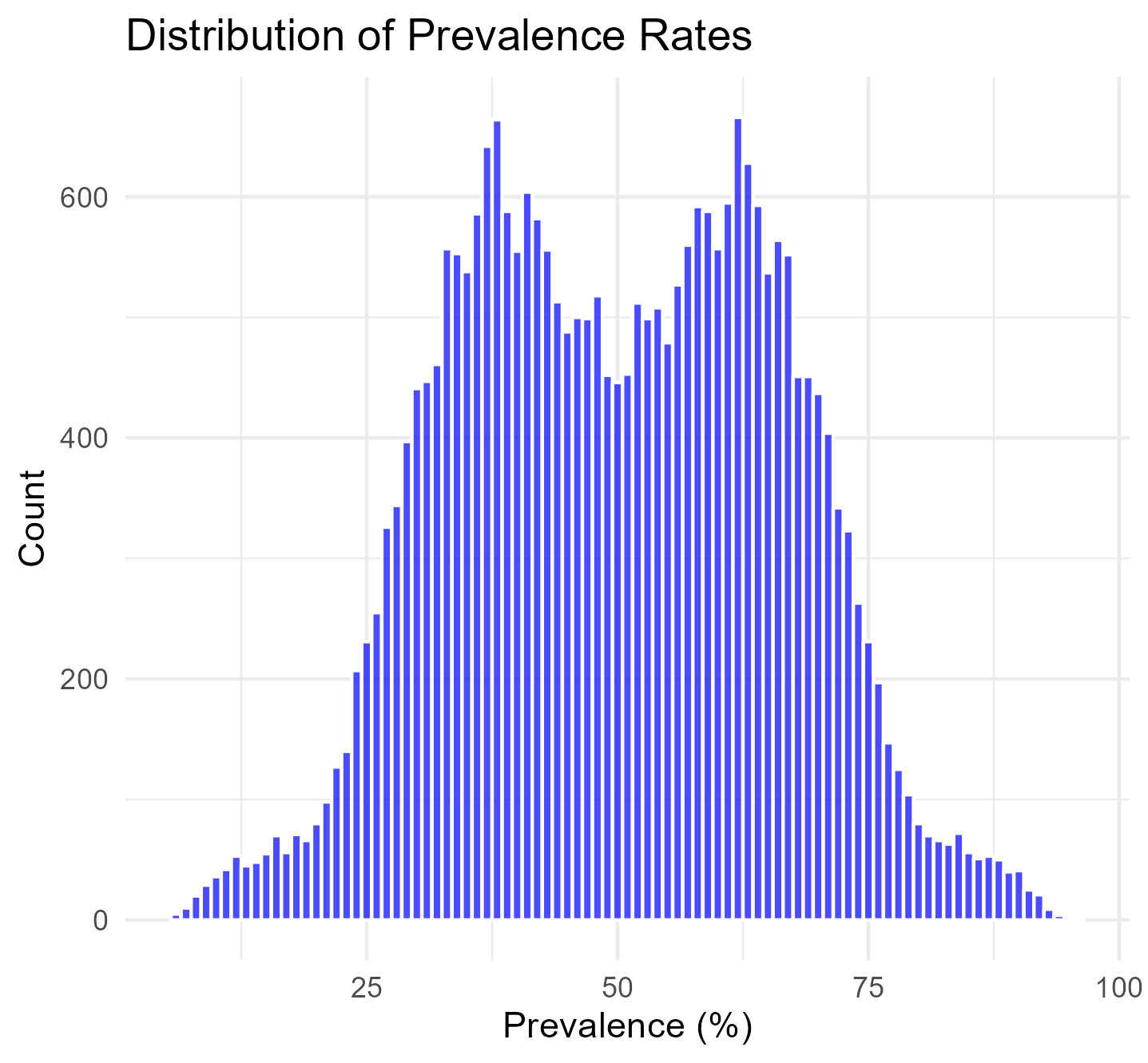
### 3.3.2 ANOVA

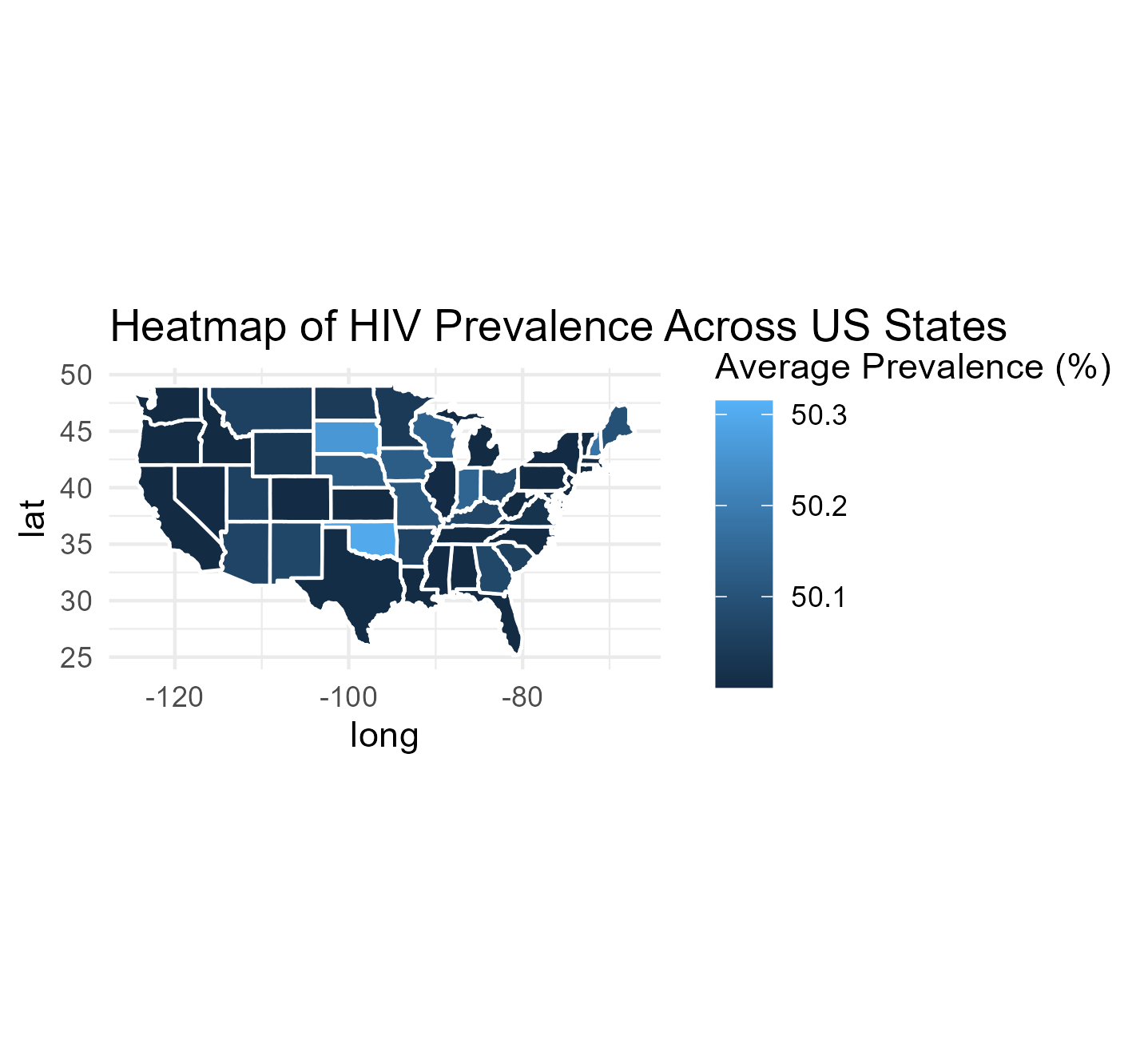
### 3.3.3 LASSO

# 4. Result

## 4.1 Exploratory Data Analysis

Following that, the distribution of the outcome was determined and graphed as a histogram graph to show the data distribution.Upon visualizing it, it becomes apparent that the data set is binomially distributed.

 The outcome variable was then graphed as a heatmap in the US. Upon producing this, it suggests that the variance in the outcome variable is limited or much lower than expected.

 Using base R functions and dplyr package, the data frame was manipulated and filtered to summarize by averaging the prevalence across the years for both the US and the state of Georgia. This was then plotted to create a time series plot, to clearly show the HIV prevalence from 2011 to present day, as shown in Figure 3. Something to note in figure three, is the sharp increase in HIV prevalence 2019 and 2022 in Georgia. Whether it is statistically significant is yet to be determined.

|  |
| --- |
| Figure 3. Time Series of HIV Prevalence between Georgia and the United States |

|  |
| --- |
| Figure 4. Boxplot showing prevalence of HIV in the US and GA startified by major determinants |

|  |
| --- |
| Figure 5. Time trend of HIV for Each Determinants/Stratification Group |

## 4.2 Statistical Data Analysis

### 4.2.1 Linear Regression

|  |
| --- |
| Table 1. Linear Regression Results |

### 4.2.2 ANOVA

|  |
| --- |
| Table 2. ANOVA Results |

### 4.2.3 LASSO

# 5. Discussion/Conclusion

# 6. References

# 7. Supplement

|  |
| --- |
| Supplement 1. Summary Statistics of each Variable in the data set |