Le, Cody
GEO 448
Fall 2022
Final Project

Spatial Exploration of the Intersection of HIV and COVID-19 Epidemics: Executive Summary

Introduction

In the last two years, public health efforts have been focused on the COVID-19 pandemic. Yet, another epidemic has been living among us and continues to spread widely. HIV is listed as a medical condition by the Center for Disease Control (CDC) that is a high risk and individuals with the pre-existing condition have a higher risk of hospitalization, intensive care, and potentially lead to death from COVID-19. This analysis seeks to understand the intersection of the HIV epidemic, COVID-19 pandemic, and socioeconomic factors specifically race and ethnicity through a spatial analysis of counties in California (AIDSVu). California is the focus of this analysis because the majority of new HIV cases overwhelming rise from this state. This analysis is inspired by an analysis conducted by Hui Luan et. al. in 2021, where they researched the intersection of HIV infection and COVID-19 infection based on area by U.S. County. This analysis will expand on this previous study to include treatments for the two epidemics, PrEP (pre-exposure prohylaxis) a preventive drug taken by HIV negative individuals to decrease the chance of HIV infection and vaccinations, either a two-dose or one-dose vaccine given to increase immune response to COVID-19. Although many studies have focused on COVID-19 or HIV separately, very few have focused on the intersection of the two. With the two epidemics occurring at the same time and with the CDC indicting higher vulnerability of COVID-19 infection in HIV individuals, this the main goal of this analysis is to spatially analyze new infection rates between the two epidemics and its specific impact on individuals of color. The analysis also seeks to determine the role of racial and ethnic minority status and socioeconomic status in the rate of infection of HIV and COVID-19. Are there specific areas or clusters in California that are more vulnerable to HIV or COVID-19 infection? Do the two epidemics affect each other and are their certain demographics with greater risk to the virus?

Data

Three datasets were used for the HIV data which comes from AIDsVu a partnership between Gilead Sciences, Inc., and the Center for AIDS Research at Emory University. The datasets include 2020 data on new HIV infections, 2020 data on current HIV statistics, and 2020 data on PrEP Users. Two datasets were used for the COVID data which comes from the California Department of Public Health Open Data Database. The first is the COVID cases, tests, and deaths dataset by date and the second dataset is the COVID vaccine progress dataset by date. The HIV dataset contains specific data on individual ethnic groups rate of infection, which was not available for the COVID data. Lastly, the 2020 Social Vulnerability Index (SVI) dataset for California by county comes from the CDC and contains census tracts as well as spatial geometry data. Extensive preprocessing was performed on the datasets. First, all datasets were cleaned and evaluated for data quality. All datasets contain county information, and the structure of each dataset was cleaned to ensure that each instance is a county in California with each column representing case counts or infection rates. The HIV datasets already contain infection counts and rates and by ethnicity. The COVID datasets were transformed extensively from each row as a date, to each row representing a county with an aggregation of case counts and vaccinations. In addition, feature engineering was performed to generate rates of infection or vaccination for the COVID data based on county population. Lastly, spatial join was performed to join the HIV, COVID, and SVI datasets. The final dataset has 58 rows, and the following features were used for the analysis:

HIV-Features: All data represents observations from 2020
- PrEP Rate – Rate of New PrEP User
- PrEP Users – Number of New Prep Users
- New Diagnoses Rate – Rate of New HIV Infection
- New Diagnoses Cases – Number of new HIV Cases
- New Diagnoses Black Rate – Rate of New HIV Infection Among Blacks
- New Diagnoses Black Cases – Number of new HIV Infection Among Blacks
- New Diagnoses White Rate – Rate of New HIV Infection Among Whites
- New Diagnoses White Cases - Number of new HIV Infection Among Whites
- New Diagnoses Hispanic Rate – Rate of New HIV Infection Among Hispanics
- New Diagnoses Hispanic Cases – Number of new HIV Infection Among Hispanics
- New Diagnoses Asian Rate - Rate of New HIV Infection Among Asians
- New Diagnoses Asian Cases - Number of new HIV Infection Among Asians
- New Diagnoses American Indian/Alaska Native Rate – Rate of New HIV Infection Among American Indian/Alaska Natives

- New Diagnoses American Indian/Alaska Native Cases – Number of New HIV Infection Among American Indian/Alaska Natives
- New Diagnoses Native Hawaiian/Pacific Islander Rate – Rate of New HIV Infection Among Native Hawaiian/Pacific Islanders
- New Diagnoses Native Hawaiian/Pacific Islander Cases – Number of New HIV Infection Among Native Hawaiian/Pacific Islanders

COVID-Features: All data represents observations from 2021
- COVID Cases – Number of Confirmed COVID Cases
- COVID Deaths – Number of Confirmed COVID Deaths
- Total COVID Tests – Number of COVID Tests Administered
- Positive COVID Tests – Number of COVID Tests with Positive Results
- Partially Vaccinated – Number of individuals who completed at least one shot out of a two shot vaccine series
- Fully Vaccinated – Number of individuals who completed either the 2nd shot of two shot vaccine series or a one-shot vaccine series
- Boosted – Number of individuals who completed a booster shot (an additional shot after the initial 2-shot or 1 shot vaccine series)

SVI-Features: All data represents publication from 2020
- FIPS – Federal Information Processing Standard code used to identify counties
- County – Name of the County
- Geometry – Location Geometry
- RPL_THEME1 – Overall percentile ranking for socioeconomic status which includes no health insurance, unemployment, below poverty, housing cost burden, and no high school diploma.
- RPL_THEME3 – Overall percentile ranking for racial and ethnic minority status.
- EP_MINRTY – Percentage of racial and ethnic minority calculated from 2016-2020
- EP_UNINSUR – Percentage of uninsured (without health insurance) calculated from 2016-2020

Additional Engineered Features
- Location – County Location derived from Geocoder
- Lat – Latitude derived from Geocoder
- Long – Longitude derived from Geocoder
- COVID-19 Cases Rate – COVID Cases Divided by County Population multiplied by 100
- Total COVID Tests Rate – Total COVID Tests Divided by County Population multiplied by 100
- Positive COVID Tests Rate – Positive COVID Tests Divided by County Population multiplied by 100
- Partially Vaccinated Rate – Partially Vaccinated Divided by County Population multiplied by 100
- Fully Vaccinated Rate – Fully Vaccinated Divided by County Population multiplied by 100
- Boosted Rate – Boosted Divided by County Population multiplied by 100

Methods

The methods for this analysis include data preprocessing, exploratory data analysis, exploratory spatial data analysis, and agglomerative cluster analysis. For the analysis, using rates instead of case counts provided more beneficial insights since due to the volume of cases for HIV and COVID-19, case counts would only result in increasing patterns instead of specific key patterns. Most of the preprocessing involved transforming the data structures, merging the data, feature engineering, and spatially joining the data. During the exploratory data analysis, it was determined through correlation matrix that the HIV features were not highly correlated to COVID features. From this discovery, the focus of the analysis shifted to focus on spatial clustering, spatial outliers, local indicator of spatial autocorrelation (LISA), and agglomerative clustering. Based on the exploratory data analysis, the following features were selected to calculate spatial clusters and outliers: New Diagnoses Rate, PrEP Rate, Total COVID Test Rates, Fully Vaccinated Rates, EP_MINRTY, and EP_UNINSUR. Using Moran's Plot, statistical significance of 0.05 was applied for the spatial cluster and outliers and a LISA interactive map with all features was created. The LISA interactive map was used to compare the spatial overlap between significant clusters or outliers between the selected features. Lastly, agglomerative non-spatial and spatial clusters were explored using the following features: New Diagnoses Black Rate, New Diagnoses White Rate, New Diagnoses Hispanic Rate, New Diagnoses Asian Rate, New Diagnoses American Indian/Alaska Native Rate, New Diagnoses Native Hawaiian/Pacific Islander Rate, COVID Cases Rate, Total COVID Tests Rate, Fully Vaccinated Rate, RPL_THEME1, and RPL_THEME3. By clustering these features, spatial clusters were created with counties in California showing distinct characteristics of the features.

Results

The interactive LISA map shows that significant spatial clusters and outliers appear in different areas for HIV infection rate compared to COVID-19 infection rate. HH clusters appear in counties in Southern Califor

nia such as Los Angeles, Ventura, Orange, San Bernadino, Riverside, and San Diego as well as Bay Area counties including Marin, Contra Costa, and Alameda. In contrast, for COVID infection rate, HH clusters appear in Central California including Kern, Tulare, Madera, and Mariposa counties. The analysis shows that there are no significant spatial clusters or outliers between the two epidemics. In otherwards, there is no interaction spatially between the two. This falls in line with the correlation matrix during the exploratory data analysis which showed that the two features were not highly correlated. HH clusters do appear in the exact same counties for Racial and Ethnic minority status as HIV infection, showing that there is significant overlap spatially between the two in these clusters. Likewise, for percentage of uninsured health insurance, HH clusters appear in Kern and Mariposa counties in Central California which are the same counties with HH clusters for COVID-19 infection. Percentage of uninsured health insurance also has spatial overlap in Orange County in Southern California with HIV infection. For the agglomerative cluster analysis (*see Appendix for figures*), 7 clusters revealed that with spatial weights, the clusters were distinct and represented key areas in California. Based on the cluster map the main cluster areas include Los Angeles, the Bay Area and Sacramento, Siskyou, Modoc, and Trinity, Lassen, Del Norte, and the rest of California. The results show that by applying spatial weights, we can clearly identify key areas in California that may be experiencing a higher social vulnerability regarding racial and ethnic minority status, socioeconomic status, and the spread of exasperation of the HIV and COVID epidemics. Moreover, HIV and COVID may not directly impact each other, but there may be indirect impacts by spatial area, socioeconomic status, and racial and ethnic minority status as shown through cluster analysis.

Discussion

       The analysis shows that HIV and COVID-19 are not significantly impacting each other and there are not significant areas where the two directly influence one another. Instead, there are specific areas that have higher infection of HIV among racial and ethnic demographics. Key counties in Southern California and the Bay Area appear as significant areas of infection for HIV. Whereas key counties in Central California appear as significant areas of infection for COVID. Moreover, racial and ethnic minority status and socioeconomic status pose a greater risk and burden for HIV and COVID-19 infection. Individuals of color have HIV infection rates that are three times higher than Whites. The areas where race, ethnicity, socioeconomic status, HIV and COVID meet are in Los Angeles, the Bay Area, and counties in upper Northern California. From this analysis, targeted efforts for public health promotion and advocacy can be implemented. Using the interactive LISA map and building on cluster analysis, key areas and location can be determined that experience the most social vulnerability and public health risk. From this analysis, separate campaigns should be implemented for HIV versus COVID, as the two epidemics are drastically different. Different areas and focus should be implemented when advocating for high risk individuals including consideration for race, ethnicity, health insurance, among other socioeconomic factors.

       This analysis is limited in scope due to the following reasons: the timeline of the data is taken at different times with COVID-19 data in 2021, HIV data in 2020, and SVI data in 2020. Unfortunately, at this time, HIV and SVI data for 2021 is not available and in 2020, vaccinations were not yet available and thus, the analysis used data from different years. In this way, an exact overlay of the infection rate and case counts was not performed and instead, it is estimated that the data from 2020 for HIV and SVI would remain the same for 2021. In addition, there is no direct data on individuals living with HIV and number of positive cases as that information is not recorded. The analysis could be different if it focused on data of individuals with confirmed HIV infection, who also tested positive for COVID-19. Lastly, the COVID and vaccination rates were calculated by simply dividing the case counts by county population. However, the CDC and other public health agencies now use a 7-day average for every 100,000 cases as the rate of infection. Since the COVID dataset was obtained with case counts only, results may be different if different metrics for COVID infection rate was analyzed. This analysis focused on spatial data exploration and clustering, for future analysis, advanced spatial regression could be implemented, and the analysis could be focused more locally on Los Angeles County only. By focusing on a county with high transmission rate, and perhaps by focusing on individuals living with HIV instead of new HIV infections, there may be additional intersections between HIV and COVID-19 waiting to be discovered.
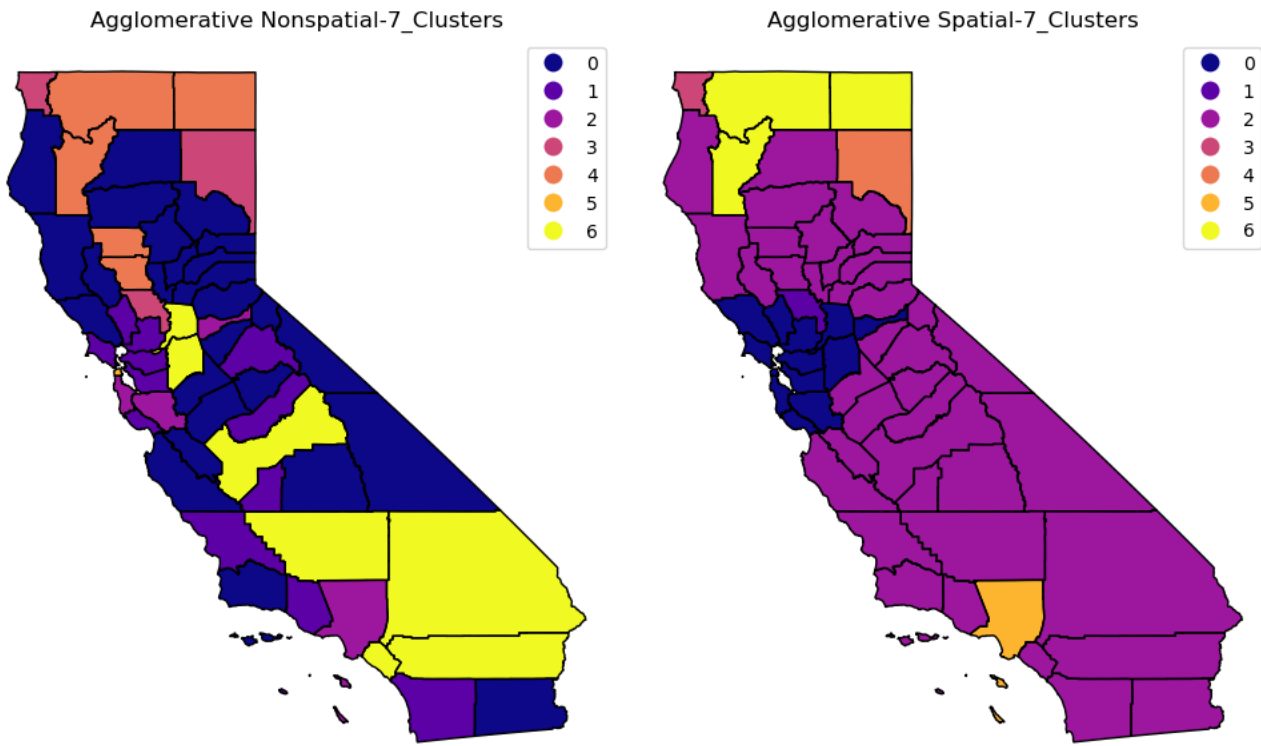
Appendix
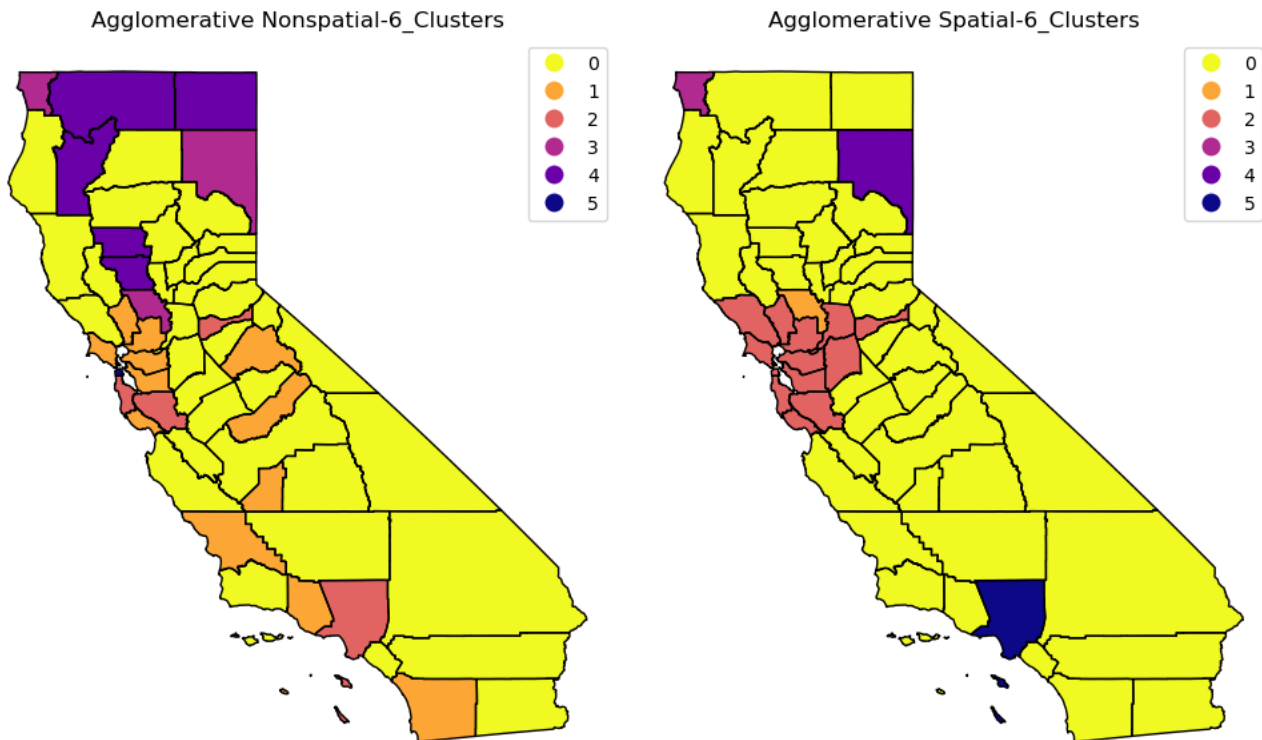

Figure 1.1: Agglomerative Clustering with 7 Clusters

Figure 1.2: Agglomerative Clustering with 6 Clusters

Works Cited

CDC/ATSDR Social Vulnerability Index (SVI). Centers for Disease Control and Prevention, 16 November 2022. https://www.atsdr.cdc.gov/placeandhealth/svi/index.html. Accessed 22 November 2022.

covid19.ca.gov. covid19.ca.gov. https://covid19.ca.gov/. Accessed 22 November 2022

COVID-19 Data Review: Update on COVID-19–Related Mortality. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/science/data-review/index.html. Accessed 22 November 2022

Datasets. California Health and Human Services Open Data Portal. https://data.chhs.ca.gov/dataset. 22 November 2022

HIV. Centers for Disease Control and Prevention, 04 November 2022. https://www.cdc.gov/hiv/default.html. Accessed 22 November 2022

Local HIV and AIDS Data - AIDSVu: Understanding HIV Where You Live. AIDSV, 21 August 2019. https://aidsvu.org/local-data/. 22 November 2022

Luan, Hui, et al. "HIV infection prevalence significantly intersects with COVID-19 infection at the area level: a US county-level analysis." *Journal of Acquired Immune Deficiency Syndromes (1999)* 88.2 (2021): 125.