The Effects of Socioeconomics on the Spread of West Nile Virus in California Counties

Sharis Hsu, Alexander Mai, Eric Gavizon, Thomas Warkentine, Danaii Elizondo

Christopher Grizzaffi, Dr. Rusty Lowe, Cassie Soeffing, Peder Nelson, & Bill Lam

In partnership with NASA, the Texas Space Grant Consortium, & the University of Texas at Austin Center for Space Research

Abstract

The spread of mosquito-borne diseases—illnesses that use mosquito species as a primary vector—across human populations in a variety of environments is a growing concern in the domain of public health. Socioeconomic factors such as income and education create disparities in public health; therefore socioeconomic factors across various regions can dictate the uneven distribution of mosquito-borne diseases towards different groups. This study aims to determine the strength of this relationship between specifically West Nile cases, and different races in varying socio-regions in California counties. This study will have a scope that spans the state of California across the time period *2020.* The spread of the diseases is hypothesized to be most prevalent among minority groups in lower income regions due to increased rates of artificial breeding, and lower access to treatment. In our analysis, we will use data from the Census, the CDC, and WorldCover Map. Wwe will utilize statistical maps to determine how the different factors overlay in counties in California. Our study found that the factors of income and land cover do not play a role in influencing the prevalence of West Nile Virus in a given region. However, we did find a positive, linear relationship between white population percentage and the amount of virus cases. Thus, more resources for the treatment of West Nile Virus should be allocated to regions in which there is a high or rising percentage of white population. In the future, the relationship between land cover and the virus should be reexamined as land cover data is ever improving, both in quality and quantity, and while more socioeconomic factors should be considered.

Keywords- West Nile Virus, California Counties, Race, Income, Socioeconomic, Social Epidemiology

Research Question

1. How do the socioeconomic factors of race and income affect the spread of West Nile Virus per California county?
2. How does land cover reflect socioeconomic status and impact the spread of West Nile Virus?

Our research aims to establish a correlation between socioeconomic status and West Nile Virus cases in California counties. Regarding socio economics, we are specifically focusing on the median income level of a given county, and the majority race of the county. We want to establish if those who come from a lower socioeconomic status are more susceptible to infection of West Nile Virus. Although most people survive being infected with West Nile Virus, the virus does have long-term symptoms and neurological impacts. Therefore our research targets providing tailored medical resources and vector control for different counties in California depending on their socioeconomic status.

Introduction and Literature Review

The West Nile Virus, a prominent descendent from the family *Flaviviridae*, is an infectious pathogen spread via mosquitoes of genus *Culex*, originally discovered in Uganda in 1937. West Nile Virus made its American debut in New York during 1999, and from 1999 to 2015 there have been 44,000 confirmed and probable cases of West Nile Virus in the United States (Clark et al. 2021). However, infection of West Nile Virus can be asymptomatic, with the Central of Disease Control stating 70-80% of human infections are asymptomatic. In the United States over 65 mosquito species (Colpitts et al. 2012) are currently being used as vectors for West Nile Virus, increasing the spread of infection beyond the spectrum of birds, the original hosts for the virus. West Nile Virus is transferred to a human host when an infected vector probes the skin to gain access to the host's blood. All mosquito species have a proboscis, which is used to target dermal blood vessels in vertebrates. These proboscis release active saliva proteins which differ depending on the species of mosquito. However, differences in saliva have proved to alter the transmissibility of West Nile Virus, for example the reduction of splenocyte proliferation and Th1//Th2 cytokine production allows for more virus transmission, increasing the effectiveness of infection for certain species of mosquitoes (Wanasen et al. 2004).

Our research focuses on the state of California, home to one of the largest income gaps in the United States (Allegretto 2011). Various levels of socioeconomic status displayed in California create large disparities in access to resources and medical care, reflected in "diseases of poverty" such as *tuberculosis* and *malaria*. A similar study regarding socioeconomic status and *Aedes* mosquitoes was performed in 2020, with the results indicating too wide a spectrum of study to establish any sort of relationship (Whiteman et al. 2020). Large cultural, climate, and public health differences across the world made the study inconsistent, leading us to focus our research solely on counties in California.

Today, about 1% of all West Nile Virus cases are neuroinvasive, presenting themselves with symptoms such as acute flaccid paralysis and lesions such as *lymphoplasmacytic meningo-encephalomyelitis* (Habarugira et al. 2020). The absence of a definitive treatment for West Nile Virus, increases the need for knowledge surrounding the most susceptible populations. This study provides an analysis of populations from various socioeconomic backgrounds such as income and race to help allocate needed medical resources to treat West Nile Virus in these groups.

Research Method

Due to the statistical nature of our question, our research team began by gathering 2020 Census data for all 58 counties in California. Through the 2020 Census, we were able to identify the median household income in each county, majority race, and population. We additionally utilized race percentages from 2019 American Community Survey 5-year estimates. Unfortunately, race percentages from 2020 are not available, however, since race percentages typically do not change extreme amounts in a single year, we chose to stick with the 2019 survey results. To integrate land cover data, we utilized WorldCover 2020 to map the regional landscape for San Bernardino County, Orange County, Lassen County, Imperial County, and Stanislaus County. To quantify the land cover, we approximated the percentages of each land classification through image analysis of color with the 2013 GeoTests Color Proportions Program from the University of Toulouse-Mirail.

ANOVA of Median Income & West Nile Virus Cases

The 2020 Census divides median household income into 6 categories as follows:

- $22,901 to $44,972
- $44,973 to $54,076
- $54,077 to $67,339
- $67,340 to $76,105
- $76,106 to $96,775
- $96,775 to $160,305

While there are 6 categories provided by the 2020 Census, no California counties fall in the $22,901 to $44,972 category, creating an empty set. Therefore, for our analysis of median income and West Nile Virus cases note that we will only use the 5 highest categories of median income.

To establish statistical differences between our five categories of medium income we use One-Way Analysis of Variance (ANOVA). Due to varying population sizes for each county, we ran ANOVA on the percentage of West Nile Virus cases in each county. For decimal accuracy, these minute percentages were multiplied by 10^6.

In this case, we define our null hypothesis to be no difference in average percentage of West Nile Virus between income groups in California, whereas our alternative hypothesis is that there is a significant difference in average percentage of West Nile Virus between income groups in California. We also assume that our data is normally distributed, variance among the income groups is approximately equal, and that all observations are independent of each other.

Model Utility Test of White Population Percentage & Confirmed West Nile Virus Percentage

The 2019 American Community Survey (conducted by the Census Bureau) provides data on the percentage of white population for all California counties except: Merced, San Bernardino, Santa Cruz, Stanislaus, and Tuolumne. Consequently, these five counties have been removed from the model utility test.

For our first model utility test, we include all counties in California (other than the five with no data), and define β to be the true average change in percent of West Nile Virus cases for every 1% increase in percent of white population in a county. We assume normality, a linear relationship, independence, and prove homoscedasticity in the residual plot (**Figure 1**).

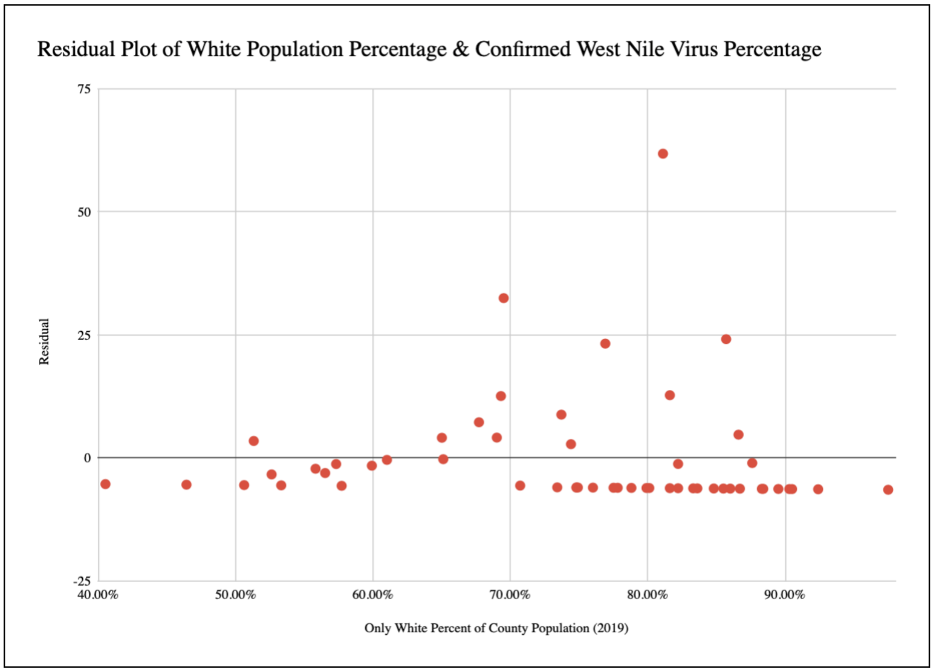

Figure 1: Residual plot of white population percentage & confirmed West Nile Virus percentage for all counties

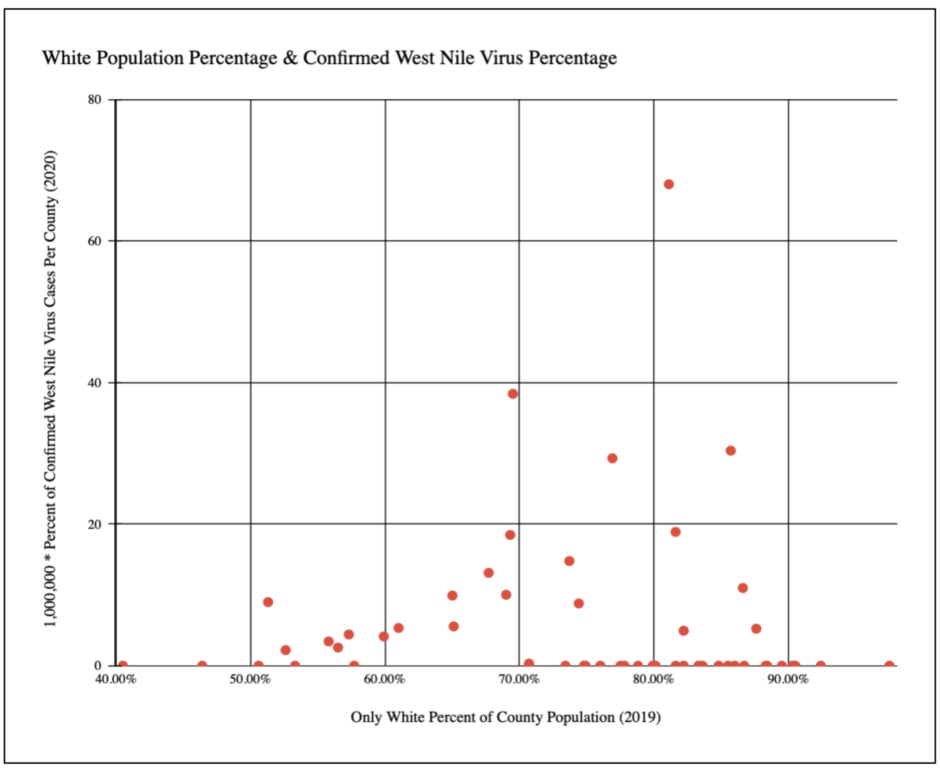

Figure 2: Scatterplot of white population percentage & confirmed West Nile Virus percentage for all counties

However, it is important to note that in 2020, 28 California counties failed to acquire confirmed cases of West Nile Virus. Because of this, our data was most likely skewed, requiring a second model utility test utilizing only the California counties that had confirmed West Nile Virus cases in 2020. Therefore, calculations were once again performed on the 23 California counties that have confirmed cases of West Nile Virus in 2020 and released white population percentages in 2019.

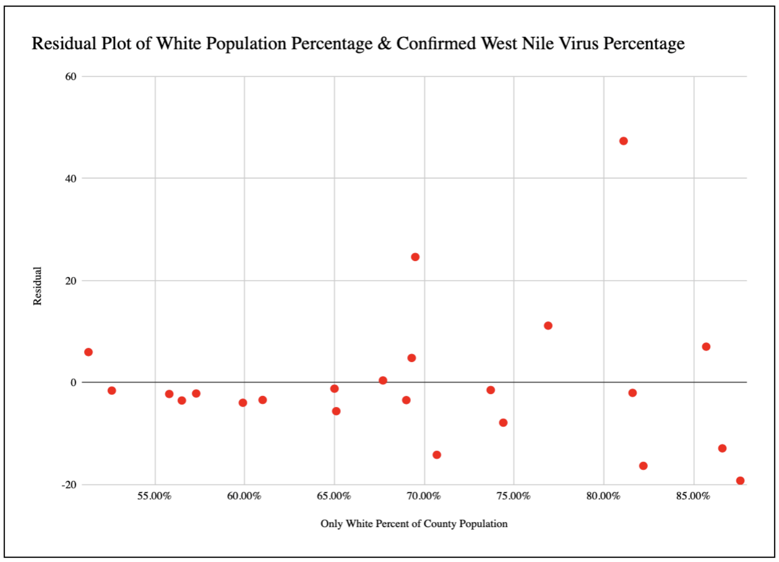

Figure 3: Residual plot of white population percentage & confirmed West Nile Virus percentage for counties with one or more confirmed West Nile Virus cases

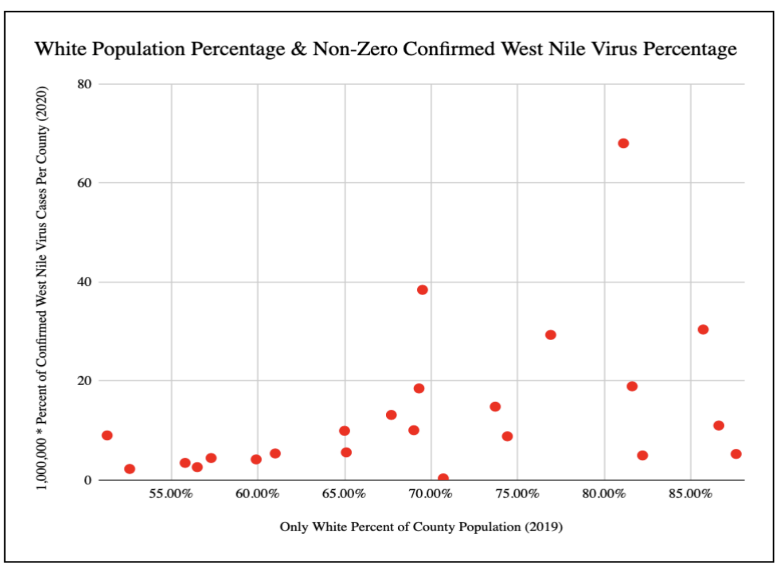

Figure 4: Scatterplot of white population percentage & confirmed West Nile Virus percentage for counties with one or more confirmed West Nile Virus cases

Land Cover Approximation Through Color Proportion Analysis

In order to search for a correlation between land cover and West Nile cases within a county, it was necessary to quantify the land cover in the area. To do this, five counties: Stanislaus, San Bernardino, Orange, Lassen, and Imperial were selected to be a simple random sample via a number generator, to acquire a sample size representative of the state's other counties. Since California has 58 counties, we utilized a sample of 5 counties as this is less than 10% of all California's counties, allowing for independence. Afterwards, the WorldCover Map of California as well as the state's county map, were overlaid and the corresponding counties were isolated for analysis.

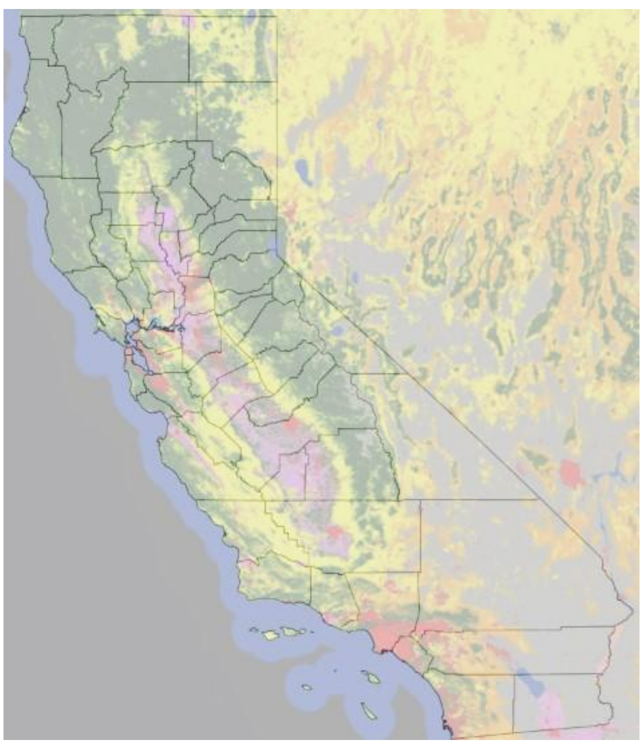

Figure 5: Overlay map of the California counties dividing the WorldCover map of the region's land cover

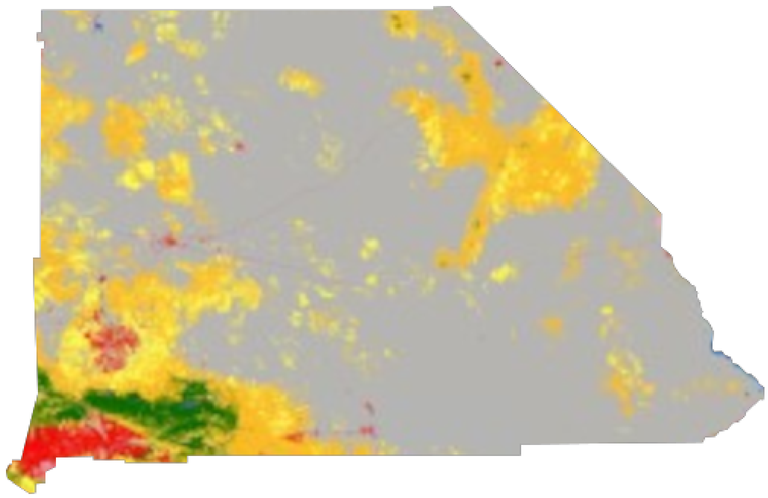

Figure 6: Regional land cover of San Bernardino County

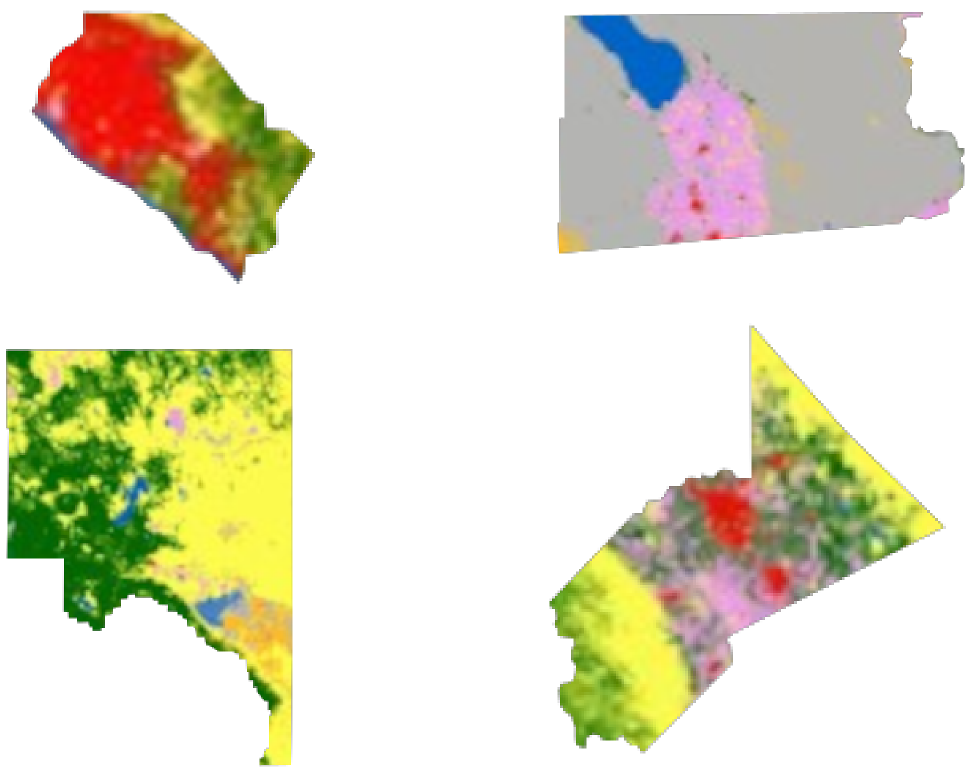

Figure 7: Regional land cover of Orange County (top-left), Imperial County (top-right), Lassen County (bottom-left), and Stanislaus County (bottom-right)

Once the images were prepared. They were placed within the 2013 GeoTests Color Proportions Program developed by L. Jèqou of the University of Toulouse-Mirail. To approximate the land cover makeup of a region, the program observes a map and takes many samples of the image, noting the color, and the number of samples with the specific color. By noting the total color samples, as well as the corresponding land cover classification for each color, the percentage approximation of land cover classes within a county is derived through dividing the number of samples for a particular color, over the total sample size.

The program sampling step, the area size of each color sample, was set to 7, and the color difference threshold, the degree of difference needed to classify a color as different from others already found, was set to 30. The YUV color space format was chosen to format the colors, and the data was presented in an HSL color spectrum circle. In doing so, the program collected a suitable number of samples and organized color data from the county samples to approximate land cover. It must be noted that the different colors had to be organized manually afterwards due to many different variations of a single hue being found from color mixing during sampling. Thus, some marginal error and variation from the true value must be taken into account and acknowledged in conclusion.

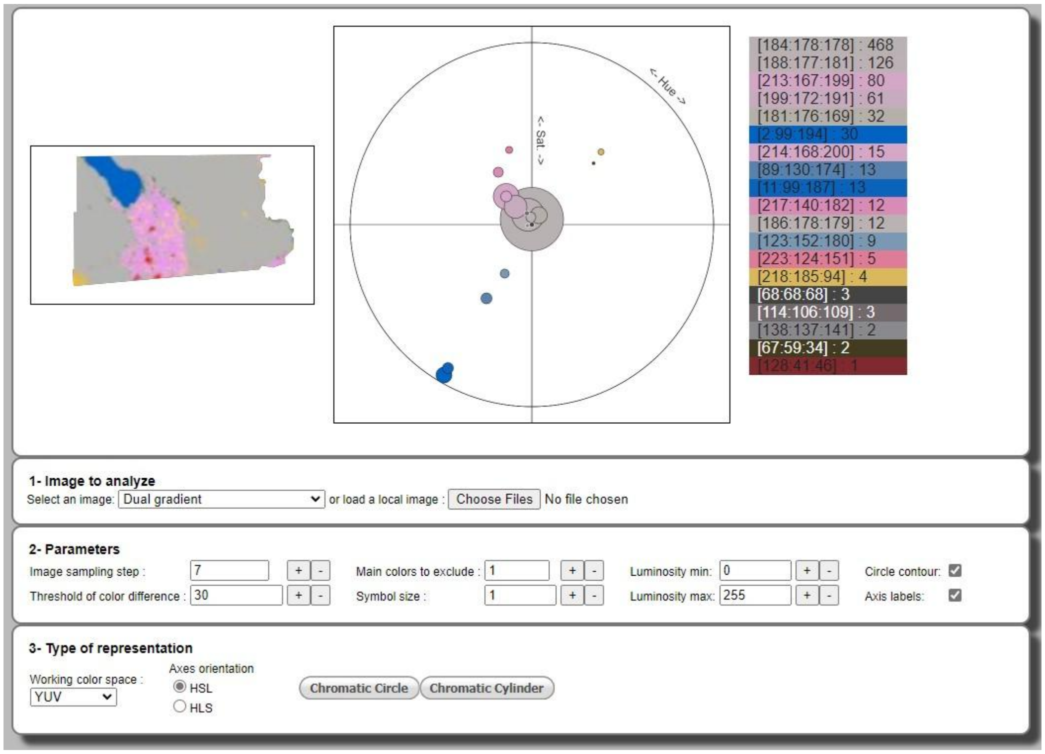

Figure 8: Color proportion analysis and chromatic circle for Imperial County

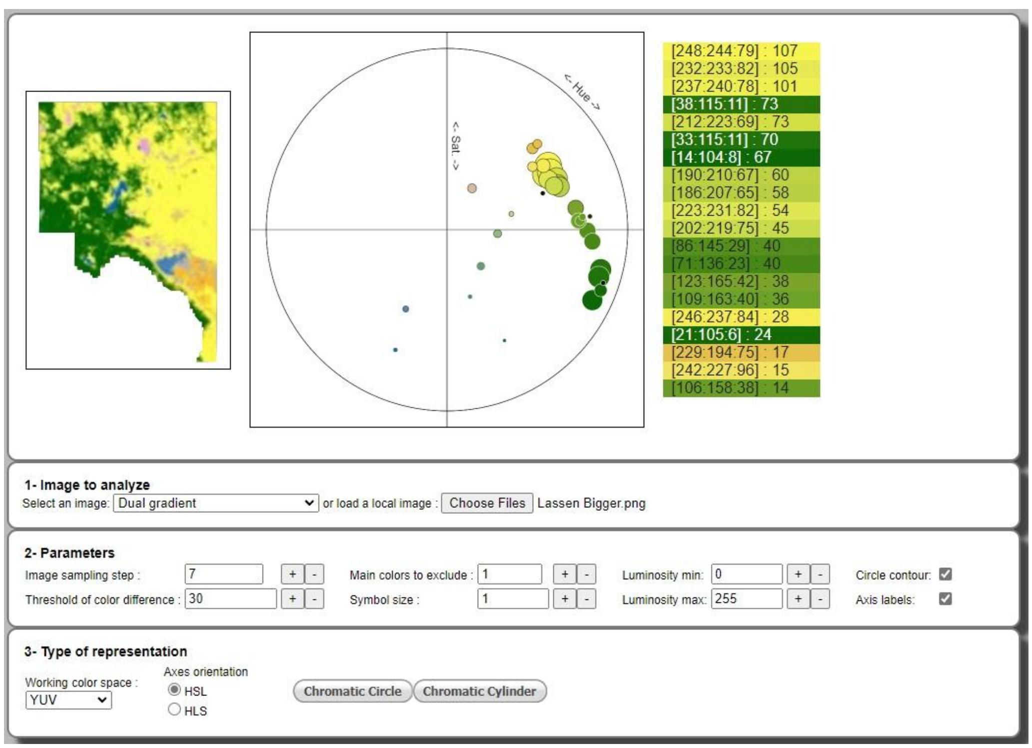

Figure 9: Color proportion analysis and chromatic circle for Lassen County

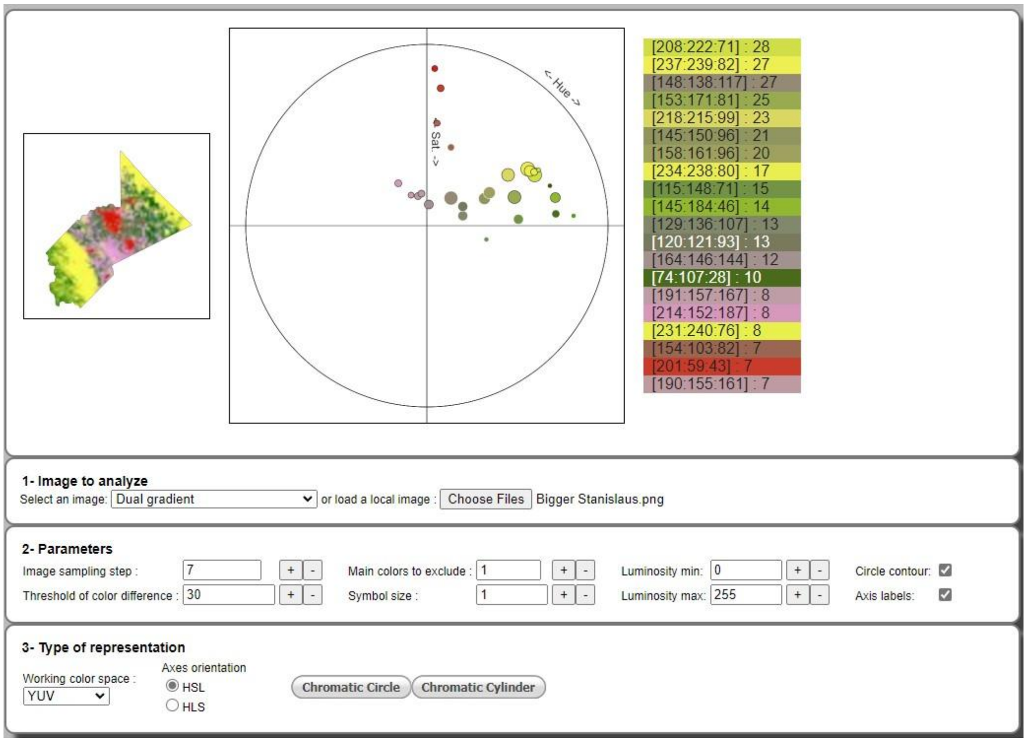

Figure 10: Color proportion analysis and chromatic circle for Stanislaus County

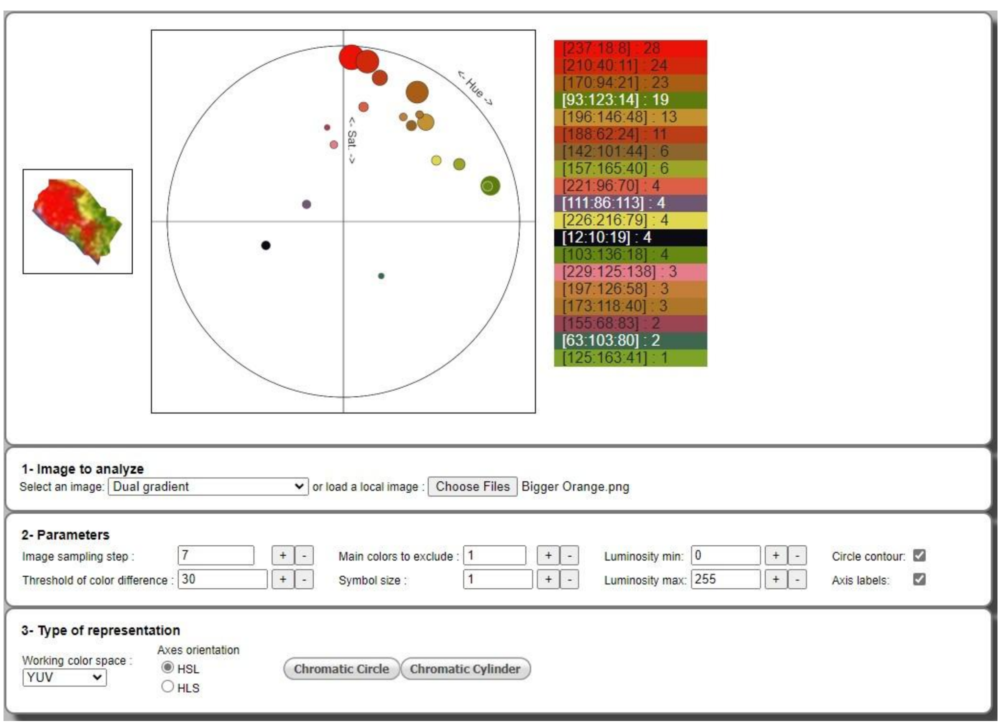

Figure 11: Color proportion analysis and chromatic circle for Orange County

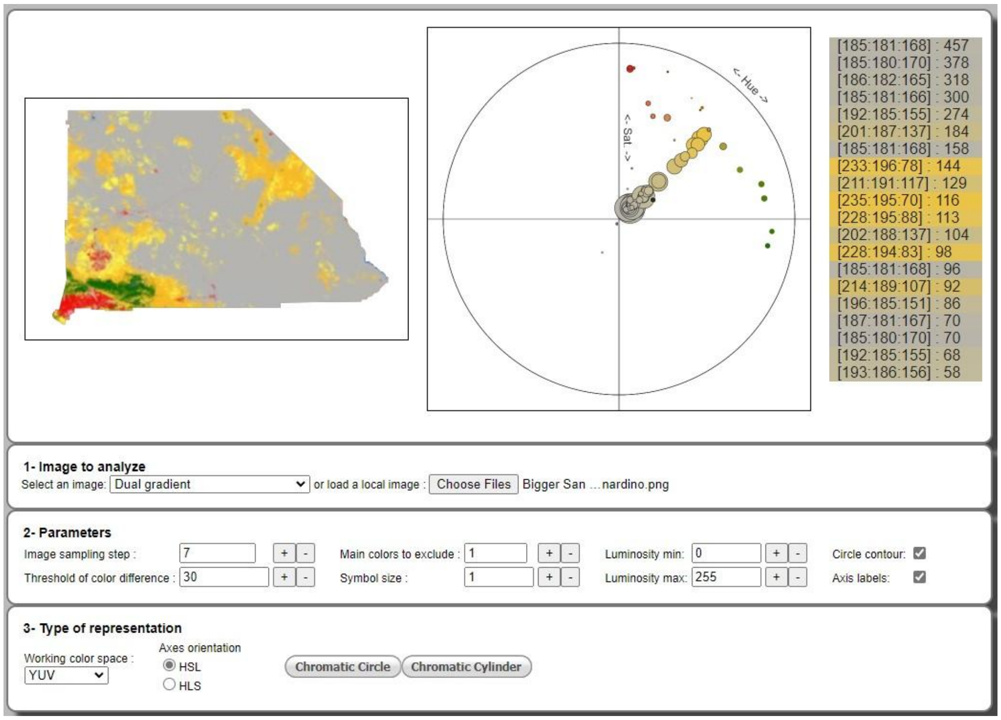

Figure 12: Color proportion analysis and chromatic circle for San Bernardino County

Model Utility Test on Land Cover & Confirmed West Nile Virus Percentage

After establishing the land cover for our given 5 counties (Figure 13), we ran model utility tests on built up land cover, tree cover, grassland cover, bare/sparse vegetation cover, and permanent water body land cover. These types of land cover are most similar to biomes that mosquitoes prefer, and reflect socioeconomic factors.

County	1,000,000 * Confirmed West Nile Virus Percentage (2020)	Built Up	Tree	Grassland	Bare/Sparse Vegetation	Permanent Water Bodies
Imperial	5.56%	0.11%	0%	0%	72.73%	7.30%
Lassen	0%	0%	38.49%	56.77%	0.00%	1.05%
Orange	5.33%	71.88%	20.00%	6.25%	0%	0%
San Bernardino	1.38%	2.35%	3.27%	9.49%	74.56%	0%
Stanislaus	63.31%	7.89%	33.92%	30.70%	15.50%	0%

Figure 13: Summary of land cover data for our 5 counties

Results

ANOVA of Median Income & West Nile Virus Cases

Upon running ANOVA on our data, we receive a p-value of 0.204662869, greater than our significance level of $\alpha = 0.05$, so we fail to reject the null hypothesis. There is not statistically significant evidence that West Nile Virus cases vary between the different income brackets from the 2020 Census.

Model Utility Test of White Population Percentage & Confirmed West Nile Virus Percentage

Our model utility test for all counties, provides a p-value of 0.8738888892, once again greater than our significance level of $\alpha = 0.05$, so we fail to reject the null hypothesis. We do not have convincing statistical evidence of a linear relationship between white population percentage and confirmed West-Nile Virus cases. However, upon examining our data, it does appear that there is a linear relationship, but it has been altered due to the large number of counties with no confirmed West Nile Virus cases.

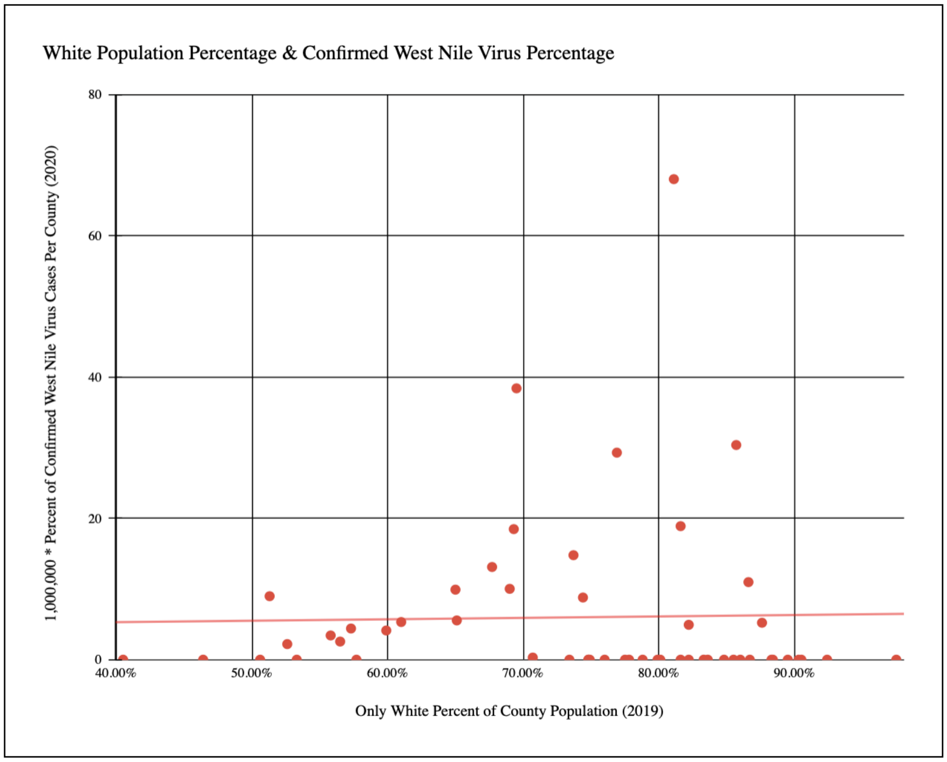

Figure 14: Scatterplot of white population percentage & confirmed West Nile Virus percentage for all counties along with the linear regression $y = 0.0202303191 + 4.504788658$

Running another model utility test on only counties that have confirmed West Nile Virus establishes a p-value of 0.0391891246 less than our significance level of 0.05, so we reject the null hypothesis. There is convincing statistical evidence of a moderate linear relationship between white population percentage and confirmed West-Nile Virus cases for counties with at least 1 confirmed West-Nile Virus case. This linear relationship is described by the equation $y = 0.597963359x + 27.75961643$, where y is the percentage of West-Nile Virus cases multiplied by 10^6 and x is the white population percentage. For this linear regression, $r = 0.4326918961$ and $df = 21$.

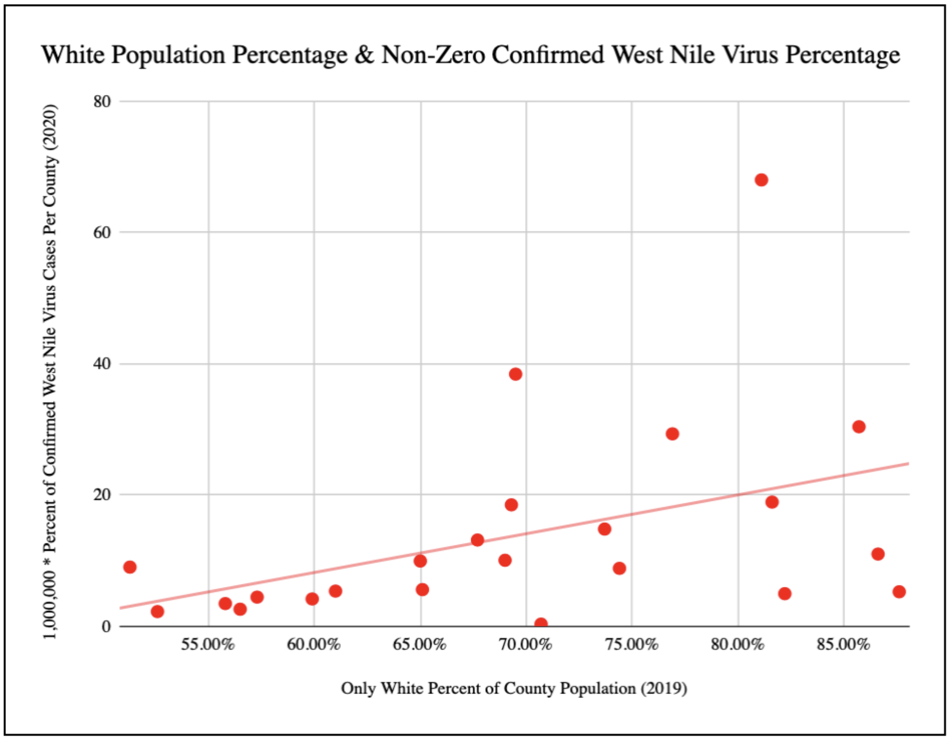

Figure 15: Scatterplot of white population percentage & confirmed West Nile Virus percentage for all counties along with the linear regression $y = 0.597963359x + 27.75961643$

Model Utility Test on Land Cover & Confirmed West Nile Virus Percentage

All five of our model utility tests run on land cover, resulting in extremely large p-values (Figure 16). Therefore, we do not have convincing statistical evidence of a linear relationship between any of the types of land cover and percentage of confirmed West Nile Virus cases per county.

Land Cover	P-value	Reject or Fail to Reject	Correlation Coefficient	Equation
Built Up	0.8655957341	Fail to Reject	-0.1057583438	y = -0.0918295952 + 16.62622952
Tree Cover	0.4683764491	Fail to Reject	0.4313091848	y = 0.6703121649 + 2.288906412
Grassland	0.7835676783	Fail to Reject	0.1708199615	y = 0.198662839 + 11.01520168
Bare/Sparse	0.7021144617	Fail to Reject	-0.2361730316	y = -0.1679323635 + 20.58354189
Permanent Water Bodies	0.6864289442	Fail to Reject	-0.2488716071	y = -2.117053995 + 18.65148017

Figure 16: Summary of p-values, correlation coefficients, and equations for all five model utility tests run between types of land cover and percentage of confirmed West Nile Virus cases per county

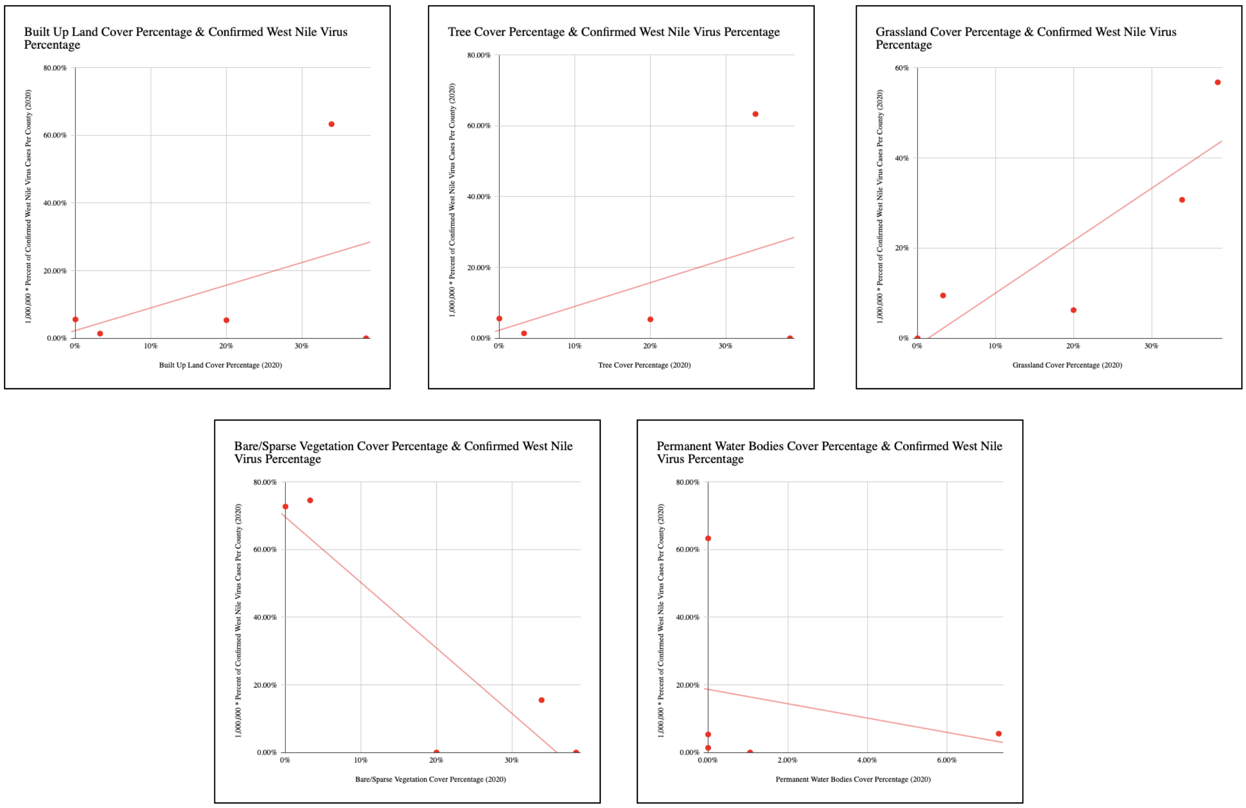

Figure 16: Scatterplots and linear regressions for types of land cover and percentage of confirmed West Nile Virus cases per county

Discussion

Our research brings together both socioeconomic factors and statistical data to form several possible models and relationships. In regards to our work on median income and West Nile Virus, it is important to note that the majority of the time West Nile Virus is not life threatening. West Nile Virus cases are only counted if an individual chooses to enter a hospital and undergo a blood test. Those of lower income may avoid entering a hospital for their symptoms due to the extreme cost of healthcare in California. Alternatively, those of higher income may be more inclined to visit the hospital “just in case” since they can pay the bill without much difficulty.

As for the relationship between white population and West Nile Virus, a 2021 study reflects that the mosquito *Anopheles stephensi* prefers type B blood, with over 70% of mosquitos being attracted by it (Khan, 2021). The American Red Cross states that 2% of Americans have type B- blood, which is more than the 1% 's for African-American & Latin American, plus 0.4% for Asians. Hence it is possible that as the white population increases the amount of type B- blood also increases, attracting more mosquitos, and resulting in more bites that could transfer West Nile Virus. However, this theory fails to take into account that B+ blood is much more common in every other ethnicity than white. Note that *Anopheles stephensi* is a primary vector in Africa, but has little to no presence in California. It is possible that the species of mosquitoes found primarily in California have preferences from other blood types more prominent in whites.

Finally, for the lack of linear relationships between land cover and West Nile Virus cases. Our land cover data fails to take into account how well maintained an area is. For example, built up land cover refers to areas that have been developed with buildings. These buildings could be worn down with exposed bodies of stagnant water that are used for mosquito breeding. Previous studies indicate that there are specific house characteristics that can prevent mosquitoes, but extraneous factors such as these are not included in land cover data (Ngadjeu, 2020).

Our research was designed and conducted to avoid possible error and bias. However, despite our best efforts our data is still subject to census bias. Census data was used extensively in this project to gather population sizes for California counties and to determine the percentage of each county’s population that had contracted West Nile Virus. Additionally, census data was used to record the median household income and the White percentage of the population. Nevertheless, the census is the most accurate source for our data, and was therefore used in this study. We also would like to acknowledge possible human error in analyzing our land cover data. Landsat images were processed and fed into the 2013 GeoTests Color Proportions Program by hand.

Conclusion

ANOVA confirms the socioeconomic factor of median household income appears to have little to no effect on the spread of West Nile Virus. However, the lack of effect may just be a reflection of severe financial insecurity plus inability to afford and access healthcare. Despite our research, income most likely does still play a factor via the ability to afford better housing that reduces contact with mosquitos or other forms of vector control. The second socioeconomic factor of our study—race—has a moderate impact on the spread of West Nile Virus. Employing model utility testing, we have established that as the percentage of whites in a county increases, so does the percentage of the population that contract West Nile Virus. As California's demographic continues to shift in the upcoming years, extensive medical training on West Nile Virus should be concentrated on areas with high or increasing white populations. Additionally, better vector control whether that be by improved housing, insecticide, or another method should be utilized.

Land cover data has proved to have little implications or insight into socioeconomics. While one can claim that more built-up counties are wealthier, and better off a similar argument can be made for counties that are less built-up. Via several model utility tests we can conclude that built up land cover, tree cover, grassland cover, bare/sparse vegetation cover, and permanent water body land cover have no linear relationship with West Nile Virus cases. However, as GLOBE Earth System Explorer Data continues to improve and possibly become more specific, it is certainly a possibility to return to the model utility test.

In the future, we plan to continue our research by developing land cover approximations for all 58 counties in California, allowing us more data points to work with and analyze. We hope to work with more socioeconomic factors such as employment and education, with the possibility of developing a theory on the intersectionality of these factors and the spread of West Nile Virus cases. As of now, we have no plans to perform research on another state, but we will take it into consideration in the future.

Citations

Arcgis Hub. ArcGIS Hub. (n.d.). Retrieved July 13, 2022, from https://sees2022-igestrategies.hub.arcgis.com/maps/35075ab1a77544dfbe536b92f8a5ab20/explore?location=34.802096%2C-92.178442%2C4.35

Bureau, U. S. C. (2021, October 8). *2020 census demographic data map viewer*. Census.gov. Retrieved July 13, 2022, from https://www.census.gov/library/visualizations/2021/geo/demographicmapviewer.html

Centers for Disease Control and Prevention. (2020, March 11). *Potential range of Aedes aegypti and Aedes albopictus in the United States, 2017*. Centers for Disease Control and Prevention. Retrieved July 13, 2022, from https://www.cdc.gov/mosquitoes/mosquito-control/professionals/range.html

Centers for Disease Control and Prevention. (2021, July 7). *Clinical Evaluation & Disease*. Centers for Disease Control and Prevention. Retrieved July 19, 2022, from https://www.cdc.gov/westnile/healthcareproviders/healthCareProviders-ClinLabEval.htm l

Centers for Disease Control and Prevention. (2022, June 15). *Statistics & Maps*. Centers for Disease Control and Prevention. Retrieved July 20, 2022, from https://www.cdc.gov/westnile/statsmaps/index.html

Color proportions of an image. (n.d.). Retrieved July 27, 2022, from https://www.geotests.net/couleurs/frequences_en.html#ce

Explore scientific, technical, and medical research on ScienceDirect. ScienceDirect.com | Science, health and medical journals, full text articles and books. (n.d.). Retrieved July 20, 2022, from https://www.sciencedirect.com/

SA;, S. A. R. S. (n.d.). *Diseases of poverty and lifestyle, well-being and human development*. Mens sana monographs. Retrieved July 11, 2022, from https://pubmed.ncbi.nlm.nih.gov/22013359/

Welcome to the Committee on Accountability and Administrative Review. Welcome to the Committee on Accountability and Administrative Review | Committee on Accountability and Administrative Review. (n.d.). Retrieved July 13, 2022, from https://aaar.assembly.ca.gov/

Worldcover. WORLDCOVER. (n.d.). Retrieved July 26, 2022, from https://esa-worldcover.org/en