

# **An Analysis of the Association Between Climate Change and West Nile Virus in the United States**

Emma Hardin-Parker & Makenzie Hicks

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr   1.5.1
v lubridate  1.9.3      v tibble    3.2.1
v purrr      1.0.2      v tidyr     1.3.1
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
Loading required package: carData
```

Attaching package: 'car'

The following object is masked from 'package:dplyr':

recode

The following object is masked from 'package:purrr':

some

Registered S3 method overwritten by 'quantmod':

```
method      from
as.zoo.data.frame zoo
```

Attaching package: 'reshape2'

The following object is masked from 'package:tidyr':

smiths

## **Abstract**

West Nile Virus (WNV) is a disease that was first isolated in 1937 but has continued to spread across the world, and since the outbreak from 1999-2010, it has spread across the United States.<sup>1</sup> As WNV has increased in recent years, it became significant to look at the possible causes, including climate change. Climate change can contribute to variations in WNV incidence by impacting vector and host factors in the spread of WNV. The analysis found no significant associations between WNV incidence and the climate change indicators chosen (drought severity, average rainfall, and average temperature). However, there were limitations to the available data, and many other contributing factors could be impacting this relationship. As climate change continues to be a prevalent issue in our world, it is important to explore how it could impact WNV incidence and, more broadly, other vector-borne disease incidence. Further research is necessary to examine the complex relationship between climate change and WNV incidence fully.

## Introduction

West Nile Virus (WNV) is a virus in the flavivirus family that is spread through the bite of an infected mosquito. The species of mosquito specifically that spreads WNV is the Culex mosquito. When a mosquito bites a bird infected with WNV, the virus enters the mosquito's system. It then accumulates in the salivary glands, which allows it to enter a human when it takes a blood meal. There has additionally been evidence of transmission from organ donation, blood transfusion, and from mother to child via breast milk. There has been no evidence of human-human transmission.<sup>1</sup>

The incubation period of WNV is approximately three to fourteen days long. Around 80% of infected people are asymptomatic, which can be concerning as asymptomatic individuals may not be aware that they could be transmitters of diseases. Those with symptoms can experience fever, headaches, fatigue, and nausea and vomiting. Approximately one in every 150 infected individuals experience severe disease, also known as neuroinvasive disease. Older individuals and immunocompromised individuals are at a higher risk of this. Symptoms of neuroinvasive disease include fever, tremors and convulsions, paralysis, and coma.<sup>1</sup>

Two of the most significant issues for WNV in public health are treatment and prevention. Currently, there is no vaccine available for WNV. Because of this, mosquito control is the best way to prevent the spread of WNV. This can occur on the individual level with behaviors such as wearing insect repelling and long clothing and on the community level by reducing mosquito breeding sites. Additionally, there is no specific treatment for WNV, but supportive care is helpful. This includes treatments such as pain control, nausea medications, and rehydration therapy.<sup>2</sup> Because of this, it is important to address what is manageable, such as mosquito control.

Currently, WNV has been reported across the United States, and it typically peaks in the summer and fall months. However, epidemics are periodic, so annual rates are not consistent.<sup>3</sup> Similarly, there is no uniform geographic distribution of human risk of WNV. Some geographic areas of higher risk include the west of the United States due to its relatively arid climate and high temperatures in the summer months. There are also areas east of the Mississippi River, although these are not responsible for as many cases.<sup>4</sup> Geographic distribution is essential when it comes to the consideration of transmission of WNV.

According to surveillance data from 2021, WNV was the leading cause of domestically acquired arboviruses in the US. All 49 states, as well as D.C., reported cases for a total of 2,911 cases. The incidence of neuroinvasive WNV was reported as 0.61 per 100,000, and of these, 95% were hospitalized, and 11% died. This was an increase as compared to the cases from 2010 to 2020. Arizona accounted for the highest incidence of neuroinvasive disease as well as WNV in general, with over half of the cases being reported in just three counties. The highest incidence was in adults over 70 years of age compared with younger adults and in males compared with females, with incidences of 2.4 and 0.8, respectively.<sup>5</sup> Given that WNV has been increasing in recent years in the US, it is important to consider the possible causes.

Climate change has been worsening in recent years, and it has been shown to be associated with WNV.<sup>6</sup> Firstly, increasing global temperature has been shown to lead to an acceleration of disease incubation,

mosquito development, and rates of biting all of which increase the opportunity for WNV transmission and severe disease. Additionally, WNV has been shown to correlate with drought and mild winters, but rainfall can also contribute to mosquito breeding sites. Due to climate change, drought and mild winters have become more likely. Lastly, climate change has affected bird migration and breeding patterns, which could impact the spread of the virus.<sup>7</sup> Because of these factors, it is necessary to further research the association between WNV and climate change.

Studies have shown that there is a need for further research on the impacts of climate change on WNV incidence and synthesized evidence.<sup>8</sup> This project aimed to directly compare WNV incidence and mortality and key climate change indicators across the United States, an area previously unexplored in literature. The indicators chosen were drought severity, average precipitation, and average temperature. It also aimed to explore further how climate change indicators have impacted the incidence of neuroinvasive WNV. These data sets could demonstrate the need for further research into WNV and climate change and increased prevention and intervention methods. In the future, it could potentially drive further research into other vector-borne diseases and guide climate change policy.

## Methods

### Data Collection & Manipulation

#### Data Transformation

Cumulative historic WNV incidence data were collected from the Center for Disease Control and Prevention's (CDC) ArboNET surveillance system for 1993-2023. The data includes reported average annual neuroinvasive WNV incidence by county.<sup>9</sup> High-risk criteria were implemented to determine counties and states of interest for further study. Counties with an average annual incidence greater than five and states with an average annual incidence greater than one were included in further analysis. After implementing the high-risk criteria, we created a high-risk county dataset including 114 counties and a high-risk state dataset with 12 states.

#### CCI Inclusion

The months of illness onset dataset was used to determine the months most appropriate for CCI analysis. We focused on July-September for our CCI data as they had the highest reported cases. CCI data points on average temperature (°F), precipitation (in), and drought severity (PDSI) were manually entered into the high-risk county dataset. The National Integrated Drought Information System defines the Palmer Drought Severity Index based on a simplified soil-water balance. Values greater than 4 indicate very wet conditions, and values less than -4 indicate arid conditions.<sup>11</sup> Due to data engine limitations, average precipitation and temperature data points were accessed from July to September, and drought severity data points were used for August.

#### Yearly WNV Data Collection

Yearly data were also collected from the ArboNET surveillance system for 2000-2023.<sup>9</sup> These datasets include stratified WNV case reports by activity type and county.<sup>9</sup> Activity type here is defined by the following: Non-human activity indicates that veterinary disease cases or infections in mosquitoes, birds, or sentinel animals have been reported to the CDC; Human infections indicate that human disease cases or infections in blood donors have been reported to the CDC, and Human infections and non-human activity: indicates that both human infections and non-human infections have been reported to the CDC.<sup>9</sup> Climate Change Indicator (CCI) data were sourced from the NOAA National Centers for Environmental Information (NCEI).<sup>10</sup> The Climate at a Glance Time Series Engine was utilized to individually source data points on average temperature, average precipitation, and drought severity for counties of interest.<sup>10</sup> Finally, we accessed a small .csv file from the ArboNET surveillance system that outlined reported WNV cases by month of illness onset.<sup>9</sup>

## **Statistical Analysis**

Three main statistical analysis techniques were conducted in RStudio v. 4.4.1 to assess potential relationships between CCIs and the incidence of WNV. K-means clustering analyses were calculated to assess the strength of the relationships between WNV incidence and environmentally similar countries. Simple and multiple linear regression models were built to predict WNV incidence, adjusting for CCIs. Correlation analyses were conducted to explore relationships between WNV incidence and various CCIs for our 114 high risk counties. Additionally, we conducted a novel Anomaly analysis to determine specific relationships between Temperature Anomalies and WNV incidence among high-risk counties.

## Results

### Descriptive Statistics & Definining High-Risk Criteria

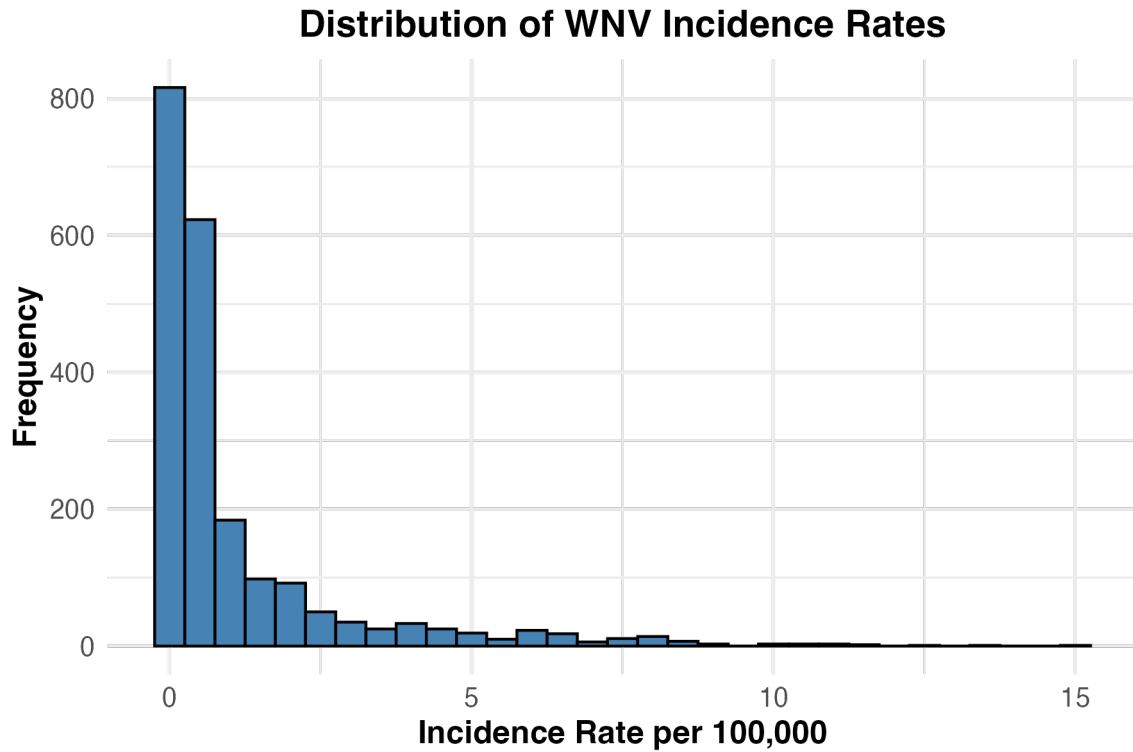


Figure 1: Distribution of WNV Incidence Rates

Figure 1 shows the distribution of WNV incidence rates by county from 1993-2023. This plot is incredibly right-skewed. The distribution of this plot motivated us to create high risk criteria to simplify our analyses and focus on counties that have pre-existing high WNV incidence. After exploring the data, we decided to include all counties with an incidence greater than 5.



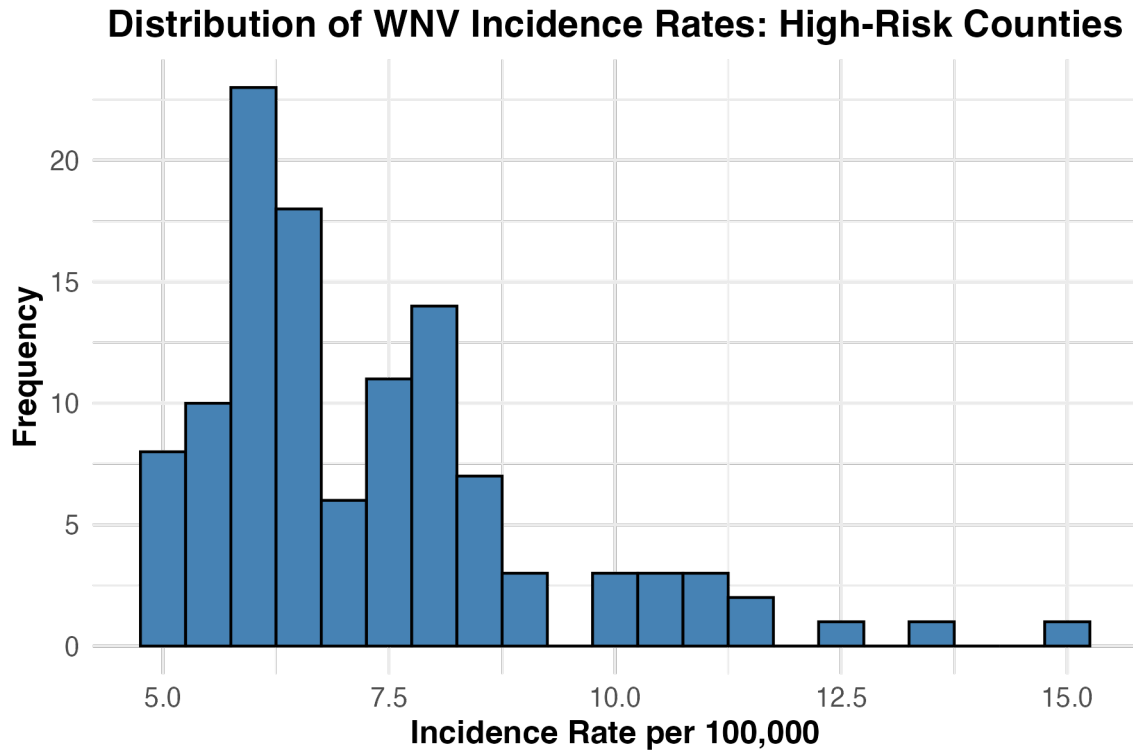


Figure 2: Distribution of High Risk Counties WNV Incidence Rates

Figure 2 shows the distribution of WNV incidence rates by high-risk counties from 1993-2023. Though the plot is still positively skewed, we felt more comfortable about the distribution moving forward. We were left with 114 total counties. Moving forward, we wanted to find our high-risk criteria for state-level data, as well as cross-reference our two newly created datasets to see if there were any discrepancies of counties determined to be high-risk existing in states not considered high-risk.

Table 1: Mean Incidence of High-Risk States

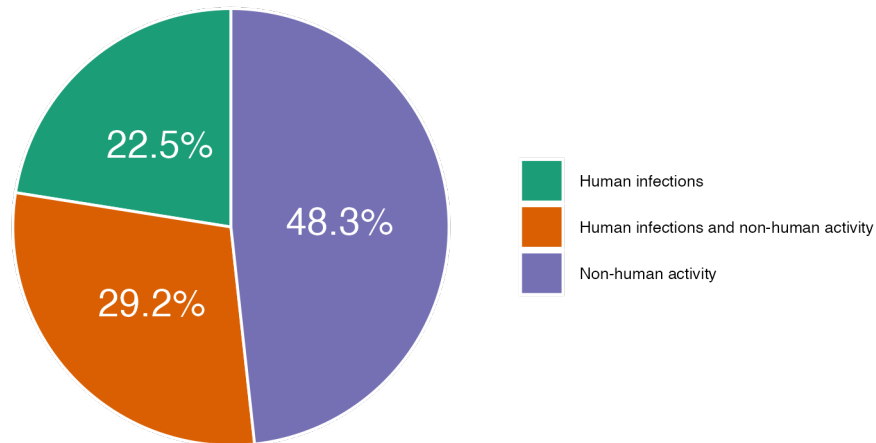
High Risk States: WNV Incidence		
Summary of High Risk States (Incidence > 1)		
State	Mean Incidence	total_population
SD	5.2919	826733
ND	4.3494	687041
NE	4.0398	1835527
MT	2.9159	899488
CO	2.4929	5082308
WY	2.2730	503178
KS	1.9127	2570120
ID	1.6266	1382020
TX	1.3745	25534378
NV	1.1411	2705322
NM	1.1345	2060350
MS	1.0645	2749604

Table 1 shows the distribution of WNV incidence rates by states defined as high-risk from 1993-2023. The initial WNV Incidence values from all 50 states ranged from 0.01 to 5.3. We elected to subset the data to only include states with an incidence value greater than one. The table above outlines the mean WNV incidence spanning the 12 high-risk states. The mean incidence values range from approximately 1 to 5.3 cases per 100,000. After cross-referencing the newly formed high-risk counties dataset and the high-risk states dataset, we found that all high-risk counties are within the 12 high-risk states listed above. All analyses were conducted using these 12 states and the 114 individual high-risk counties within these states.

## Case Activity

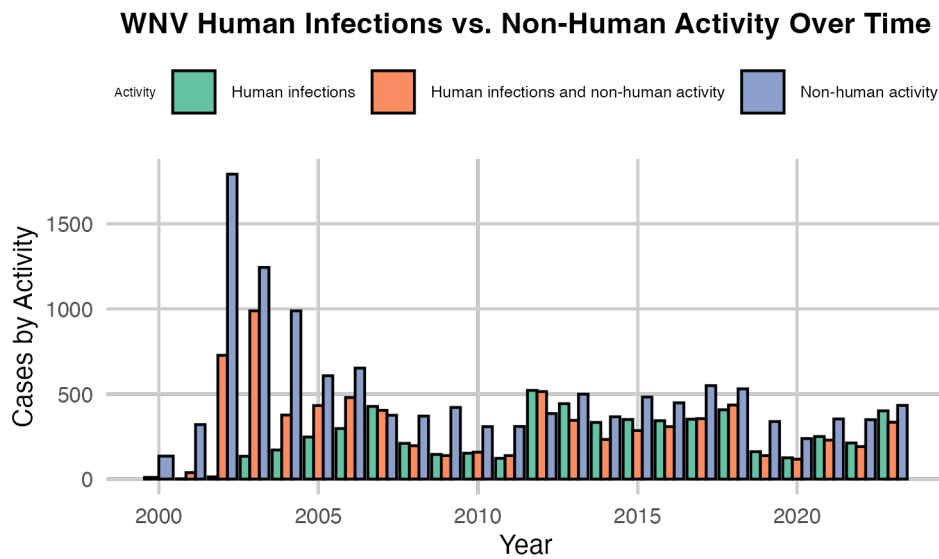
These data were individually downloaded from the [CDC Historic Data](#). Each individual data set contained information on the number of cases reported, their activity type, case type, and corresponding county. Activity type here is defined by the following: **Non-human activity**: indicates that veterinary disease cases or infections in mosquitoes, birds, or sentinel animals have been reported to the CDC, **Human infections**: indicates that human disease cases or infections in blood donors have been reported to the CDC, and **Human infections and non-human activity**: indicates that both human infections and non-human infections have been reported to the CDC. These data sets were individually aggregated and written into a .csv file for future use.

**Distribution of Case Reports by Type**



**Figure 3: Case Activity Type**

**Distribution of Case Activity Over Time**



**Figure 4: Distribution of Case Activity Type over Time**

Figures 3 and 4 shown above were created to show the distribution of case activity over time and determine the percentages of each activity type. Due to the role of mosquitoes and other animals in

human contraction of WNV, we expected non-human activity to have the highest rate of all cases and the most cases over time.<sup>9</sup> The CDC broadly defines these categories; however, we found distinguishing case counts between the activity types essential.

### Motivation for CCI Analysis

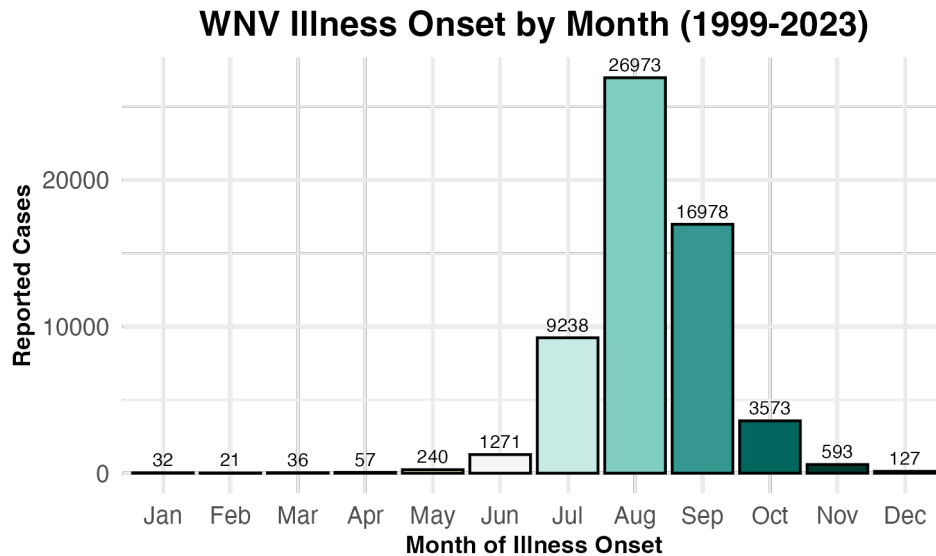


Figure 5: Years of WNV Illness Onset

Figure 5 shows the distribution of the Months of Illness Onset data file accessed from the CDC. The distribution justifies our decision to proceed with the July-September time frame for CCI analysis, given that the data are approximately normally distributed around August and September. Due to time constraints and complexities with accessing CCI data points, we decided it was most appropriate to focus on environmental conditions data for the months where WNV illness onset was highest from 1993 to 2023.

## Correlation Analysis

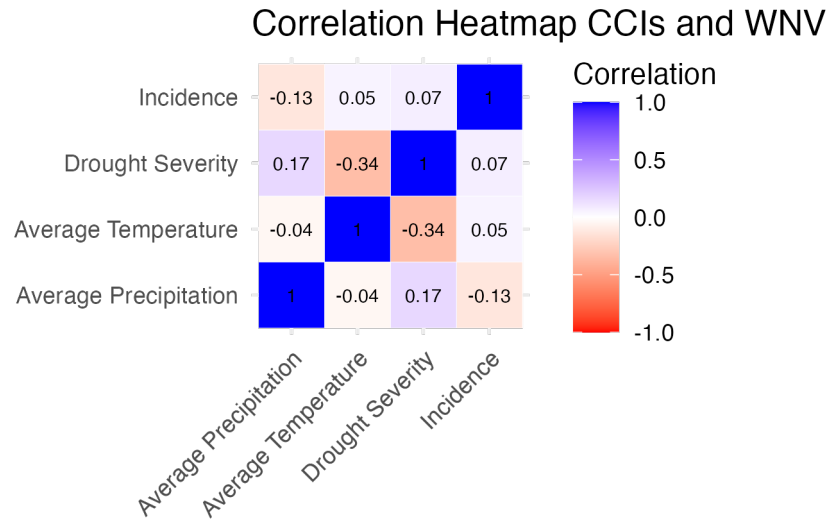


Figure 6: Correlation Analysis of CCIs by County

The above correlation plot shows no evidence of significant correlation between CCI indicators and WNV incidence. Moving forward, we elected to try other analysis techniques to see if we can find more nuanced connections between these indicators of Climate Change and WNV incidence for our 114 high risk counties.

## Clustering Analysis

We performed a k-means clustering analysis on this data to explore patterns among similar counties, relationships between drought severity, average temperature, and precipitation with WNV incidence, and identify risk groups among three clusters. These clusters were created and assigned directly through R.

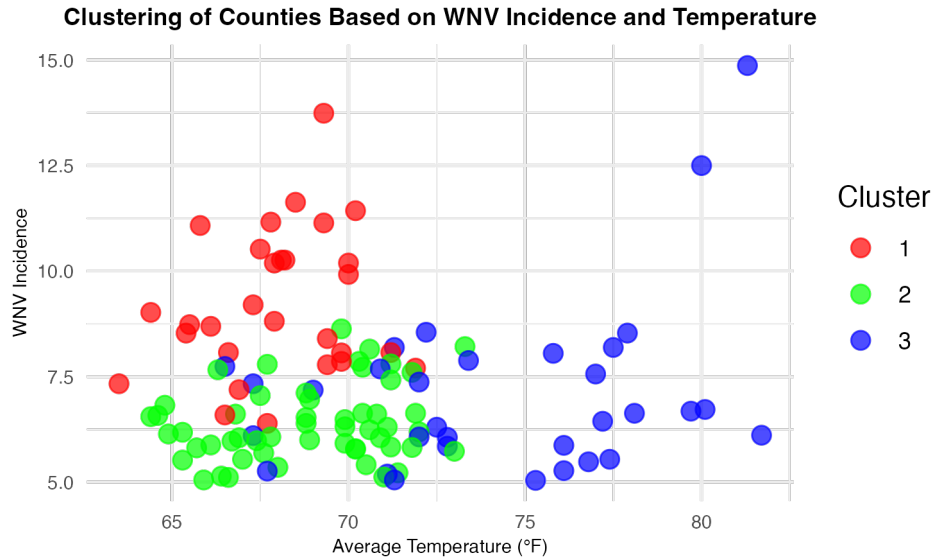


Figure 7: Incidence v. Temperature with Clustering

Figure 6 above shows clustering between the relationship of WNV incidence and average temperature. From the plot above, we can determine that counties with higher average temperatures (cluster three) between July and September experience more variability in WNV incidence, while counties with lower average temperatures consistently show lower WNV incidence, except for the variability seen in cluster one. The distinct cluster separations highlight patterns that indicate that temperature may be essential in explaining the variability of WNV incidence across different regions.

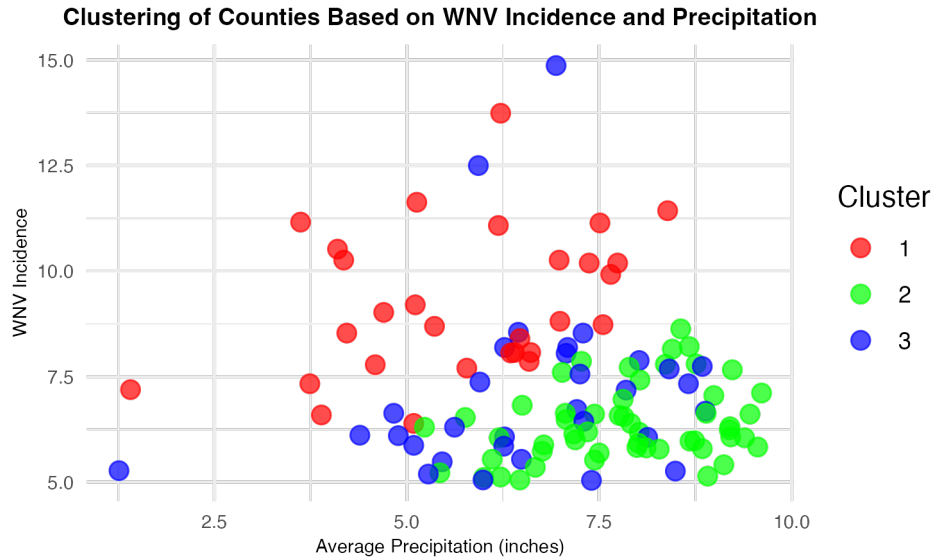


Figure 8: Incidence v. Precipitation with Clustering

The clustering above suggests that precipitation alone may not fully explain the variation in WNV incidence. The three clusters show varied relationships between precipitation levels and WNV incidence, indicating that multiple environmental or demographic factors might be driving the conflicting patterns.

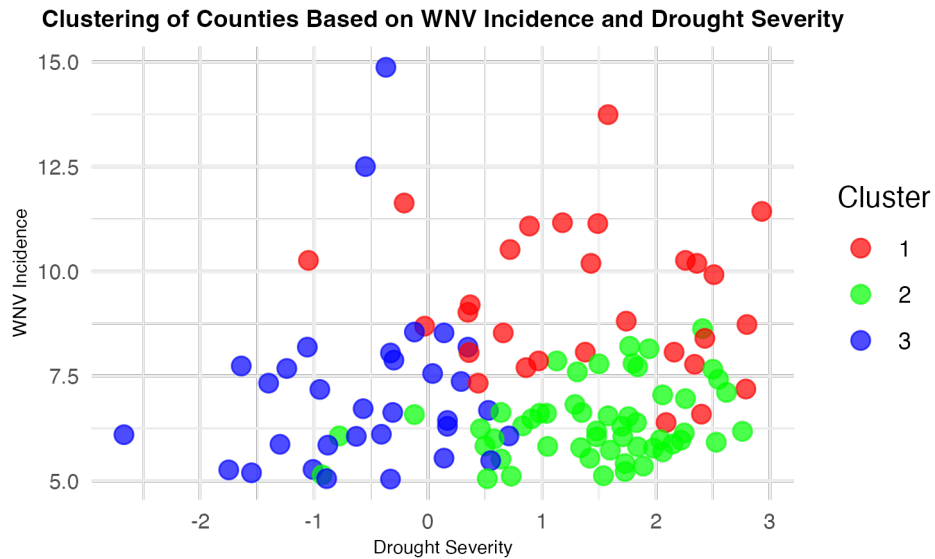


Figure 9: Incidence v. Drought with Clustering

From the above plot, we can conclude that higher drought severity indicated by negative values is clustered with moderate to lower WNV incidence, suggesting that extreme drought does not always correlate with an increase in WNV incidence. Additionally, counties that experience less severe or no drought around zero show a wider range of WNV incidence, indicating no significant relationship between drought severity and WNV incidence. However, it should be noted that drought severity data was only accessed for August, and data limitations may have affected the significance of this particular analysis.

## Analysis of Variance

ANOVA Results for WNV Incidence Across Clusters					
Term	Degrees of Freedom	Sum of Squares	Mean Squares	F Value	p-value
factor(cluster)	2.000	153.607	76.803	33.4110772021975	4.37733383226797e-12
Residuals	111.000	255.160	2.299	-	-

Figure 10: ANOVA Results post-Clustering

Following the k-means cluster analysis, we performed an analysis of variance test to determine whether there was a statistically significant difference in WNV incidence across the three clusters. From this ANOVA test, we can determine that although average temperature, participation, and drought severity were used to cluster counties in similar environmental conditions, the three clusters do not significantly differ regarding WNV incidence ( $F = 2.077$ ,  $p = 0.13$ ).

## Regression Analysis

### Model 1: WNV Incidence ~ Average Temperature

Model 1: WNV Incidence ~ Average Temperature				
Term	Estimate	Std. Error	t Value	p-Value
(Intercept)	5.554	3.191	1.741	0.084
Average_Temperature	0.025	0.045	0.553	0.581
Residual standard error: 1.908   R-squared: 0.003   Adjusted R-squared: -0.006   F-statistic: 0.306   p-value: 0.581				

Figure 11: Model 1: WNV Incidence ~ Average Temperature



### Model 2: WNV Incidence ~ Average Precipitation

Model 2: WNV Incidence ~ Average Precipitation				
Term	Estimate	Std. Error	t Value	p-Value
(Intercept)	8.372	0.773	10.837	0.000
Average_Precipitation	-0.152	0.108	-1.402	0.164
Residual standard error: 1.894   R-squared: 0.017   Adjusted R-squared: 0.008   F-statistic: 1.967   p-value: 0.164				

Figure 12: Model 2: WNV Incidence ~ Average Precipitation

### Model 3: WNV Incidence ~ Drought Severity

Model 3: WNV Incidence ~ Drought Severity				
Term	Estimate	Std. Error	t Value	p-Value
(Intercept)	7.220	0.220	32.795	0.000
Drought_Severity	0.111	0.145	0.760	0.449
Residual standard error: 1.906   R-squared: 0.005   Adjusted R-squared: -0.004   F-statistic: 0.578   p-value: 0.449				

Figure 13: Model 3: WNV Incidence ~ Drought Severity

### Model 4: WNV Incidence ~ All Variables

Model 4: WNV Incidence ~ All Variables				
Term	Estimate	Std. Error	t Value	p-Value
(Intercept)	5.345	3.490	1.532	0.128
Average_Temperature	0.043	0.048	0.889	0.376
Average_Precipitation	-0.172	0.110	-1.560	0.122
Drought_Severity	0.195	0.156	1.251	0.214
Residual standard error: 1.895   R-squared: 0.033   Adjusted R-squared: 0.007   F-statistic: 1.258   p-value: 0.292				

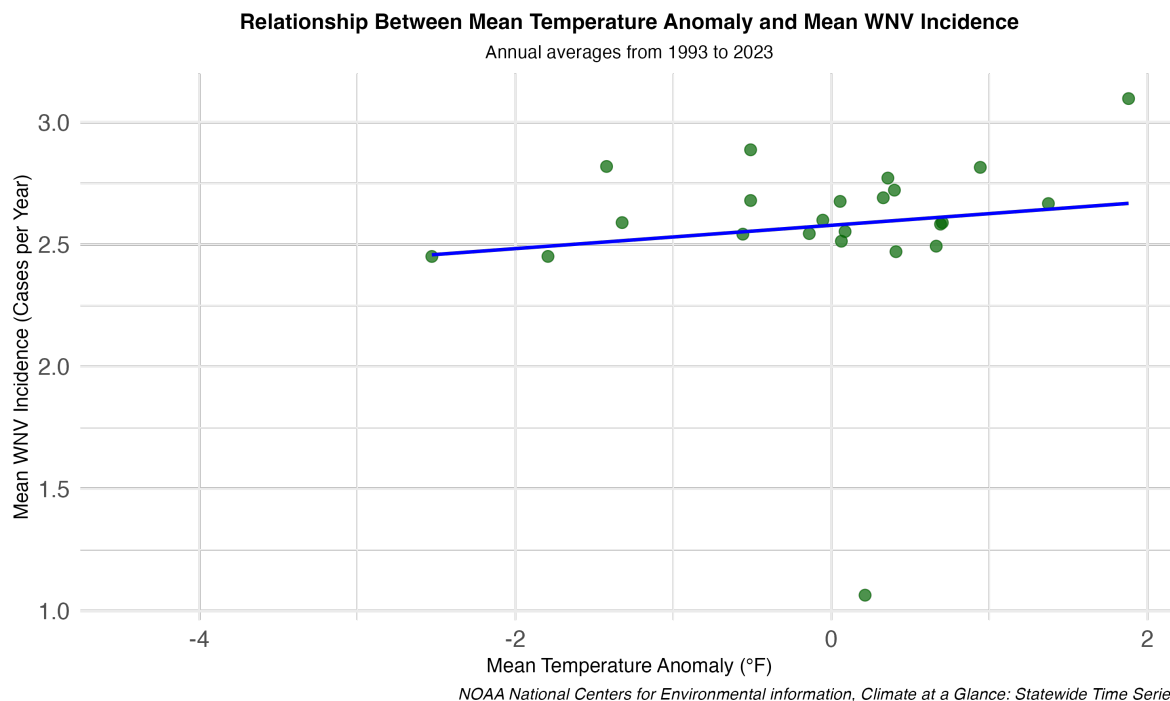
Figure 14: Model 4: WNV Incidence ~ All Variables

We conducted simple linear regression models for average temperature, precipitation, and drought severity on WNV incidence and found no statistically significant effects. We then performed a mul-

multiple regression model to determine whether multiple CCI predictors significantly predicted WNV incidence. However, the model above suggests otherwise. None of the CCI displayed above appear significant, with p-values well above the significance threshold of 0.05. Additionally, the R-squared value of 0.03317 indicates that the three CCIs together only explain 3.3% of the variation in WNV incidence.

### Novel Anomaly Analysis

According to the NCEI, temperature anomalies are more important than the average baseline temperature as positive values indicate warmer than baseline and negative values indicate cooler than baseline. We elected to analyze the temperature anomalies of high risk states and WNV incidence to account for the broad geographic region of these states. We posited that analyzing the “out of the norm” temperature trends would provide greater insight into the relationship between Temperature and WNV Incidence.



As shown by the figure above, there is a weak positive relationship between Temperature Anomaly and WNV Incidence. This further suggests that higher average temperature and subsequent positive temperature anomalies may influence WNV incidence.

## Discussion

Initial descriptive results of this study indicate that both WNV incidence and specifically neuroinvasive incidence have been increasing in recent years. However, further statistical tests also indicated that there was no significant correlation between WNV and the CCIs chosen, which were drought severity, average precipitation, and average temperature. Our results did show that counties with higher average temperatures during peak months of WNV experienced a higher variability in incidence, which may indicate some seasonal fluctuations in mosquito activity and/or viral transmission. Additionally, results showed a slight positive relationship between temperature anomaly and WNV incidence. Overall, our studies did not find any substantial relationship between CCIs and WNV, but they did show some possibility of a relationship between temperature and WNV incidence. While these results were unexpected, they are consistent with our current knowledge of WNV and climate change.

Our study results do not entirely contradict current research surrounding WNV incidence and climate change, but current research can also help to explain part of why our results were not significant. Current research has indicated that climate change has impacted mosquitoes (vector) and birds (host),<sup>7</sup> which both have the potential to affect WNV transmission and could confound the relationships we examined. Additionally, many other factors can affect these two important aspects of WNV transmission including community and individual behaviors. Another aspect is that WNV incidence is highly variable from year to year due to factors such as the seasonality of vectors, varying surveillance, and changing human activity<sup>12</sup>, so it can be difficult to truly assess the relationship across years without consistent WNV incidence rates and the introduction of multiple confounders.

## Research Implications

While our results did not show any significant associations, they are important in that they contribute to the current knowledge surrounding WNV and climate change. Our research serves as a basis that can guide other research by displaying the need for further exploration of possible confounders. We also adequately display the complexity of the relationship between WNV and climate change as both topics individually require exploration of several factors and their interactions are equally as difficult to establish. However, this study was subject to several limitations that affect the implications of our research.

## Strengths and Limitations

It should be noted that the CDC itself reported limitations of using ArboNET surveillance data. The following limitations were accessed from the CDC Historic Data WebPage and are reported here for inclusion in our limitation section. Firstly, the CDC reports that under-reporting is a common limitation to all surveillance systems that solely rely on healthcare providers to consider the disease (WNV) as a possible diagnosis, obtain the proper laboratory tests, and report confirmed diagnoses to public health authorities. Secondly, mild illness (non-neuroinvasive disease) is more likely to be under-reported than

more severe (neuroinvasive) cases. Thirdly, data are reported using the county of residence of the individual, not necessarily the county or state of exposure, possibly skewing any state and county-level data. Finally, non-human surveillance is variably conducted across the country and the absence of non-human activity data should not be interpreted as no risk.

Unfortunately, the impact of Hurricane Helene in Asheville, North Carolina, affected our original data sources for climate change indicators. We were unable to access some data we intended to use in our analyses, which resulted in the use of other data sources and analysis techniques. While we find our analyses to be thorough and possess the chance of having a long-term impact on climate change policy and recognition of the impact of climate change on vector-borne diseases, we find it appropriate to mention these limitations. The majority of the strengths in our study lie in their ability to set the standard for future research into these topics. We were able to identify and explore several key indicators as well as establish a basic study design including methodology and an initial hypothesis.

### **Future Research Recommendations**

In order to further explore these possible associations in the future, vector and host factors should additionally be included in the analysis as they are impacted by climate change and then WNV incidence is impacted by them. This can include factors such as bird migration. Additionally, it is important to expand research into diagnostics and surveillance, community behaviors, and individual behaviors, as all of these can significantly impact WNV reporting and incidence and were unable to be characterized due to the scope of this study. This research can be used as a starting point for future research as the relationship between the CCI and WNV is more complex than we were able to fully explore.