Factors Influencing Abundance & Virus Isolation in Mosquitoes

Which environmental, ecological, and temporal factors impact mosquito populations

Andrew Ruiz

2024-04-26

# Summary/Abstract

Eastern equine encephalitis (EEE) is a rare but serious disease caused by an Alphavirus transmitted to humans by infected mosquitoes. With a fatality rate of 30-50% and potential lifelong disability for survivors, understanding its transmission cycle is crucial. EEE transmission involves two cycles: an enzootic cycle between Culiseta melanura mosquitoes and birds, and an epizootic cycle where bridge vector mosquitoes transmit the virus to humans and other mammals. While C. melanura poses minimal direct threat to humans, isolations of EEE virus in these mosquitoes serve as early indicators of virus circulation. Conversely, isolations in bridge vector species signal heightened transmission risk. This research aims to identify significant relationships between variables to inform dissertation prospectus development and guide methodological approaches.

# Problem statement

Which environmental,ecological, and temporal factors influence the transmission cycles for EEE in southeastern Massachusetts and which models best describe the relationship between these factors?

# Introduction

## Background

Eastern equine encephalitis (EEE) is caused by an Alphavirus transmitted to humans by the bite of an infected mosquito. Human infections are rare, but serious (Lindsey, Staples, and Fischer 2018). Humans and other mammals are dead-end hosts for EEE. About 94% of human EEE infections are asymptomatic. While symptomatic cases are rare, the consequences can be severe. Mortality rates associated with symptomatic infections range from 30 to 50% (Morens, Folkers, and Fauci 2019). Those who survive often experience lifelong disability that includes debilitating neurological damage (Massachusetts Department of Public Health 2024). On average, there are 11 cases reported in the US annually. Most occur in Massachusetts and Florida (Vander Kelen et al. 2014). Until recently, nearly all human EEE cases in Massachusetts occurred in two counties: Bristol and Plymouth.

Since 2000, there have been 45 human cases recorded in Massachusetts, resulting in 22 deaths (Massachusetts Department of Public Health 2024). There are no human vaccines for EEE, and treatment consists of palliative care only. Prevention measures include personal behaviors to avoid mosquito bites and decreasing the mosquito population through pesticide use and environmental modifications like removing standing water (Centers for Disease Control and Prevention 2022).

The Bristol County Mosquito Control Project (BCMCP) coordinates mosquito surveillance and testing in the county from June to October. BCMCP has used the same sentinel collection sites and trapping methods for over 40 years. Once trapped, mosquitoes are sorted and counted by species, and vector species are submitted to the Massachusetts Department of Health’s (MDPH) State Lab for PCR virus testing. Results are available within 24 hours of submission. When rates of EEE mosquito infections are above a defined threshold, MDPH notifies the local boards of health and recommends preventive measures. Recommended measures include outreach and education to increase personal protective practices . Additionally, MDPH may recommend that cities and towns in the affected areas cancel evening outdoor events and discourage outdoor activities when mosquito vectors are most active (dawn and dusk). When infection rates are high enough to indicate imminent human transmission, the Commonwealth of Massachusetts will recommend and fund aerial pesticide applications over the affected areas.

## Study area

Bristol County, Massachusetts covers an area of 691 square miles with a population density of 962 people per square mile. It borders Rhode Island to the west, Plymouth County, MA to the east, and Norfolk County, MA to the north. Naushon Island and Martha’s Vineyard are across Buzzard’s Bay to the Southeast.

Atlantic White Cedar (AWC) swamps span over 4,300 acres within Massachusetts, primarily concentrated in Plymouth and Bristol counties. These swamps serve as critical breeding habitats for *Culiseta melanura* (MEL) (Commonwealth of Massachusetts Division of Fisheries & Wildlife 2007), which is recognized as the primary enzootic vector for Eastern Equine Encephalitis (EEE). AWC trees have large root systems that form “crypts” and hold water throughout the year. These crypts serve as a main habitat for MEL larval growth and provide shelter during winter months.

Between 2003 and 2022 Plymouth and Bristol counties were the epicenters for EEE transmission in Massachusetts, accounting for 24 out of the state’s 41 human EEE cases.The Bristol County Mosquito Control Project (BCMCP) conducts mosquito surveillance and abatement activities for the county and often coordinates activities with Plymouth County Mosquito Control Project.

## Questions/Hypotheses to be addressed

**Which environmental,ecological, and temporal factors influence the transmission cycles for EEE in southeastern Massachusetts and which models best describe the relationship between these factors?** From prior research we know that the mosquito lifecyle is influenced by a number of factors. Temperature plays a major role in the mosquito life cyle virus transmission dynamics. As temperature increases, the rate of larval development also increases and the time between when a mosquito ingests virus in a bloodmeal and when it can transmit it decreases. The result is that the mosquito population can grow at faster rate and they will have more time to transmit the virus. However, MEL larvae develop in well-insulated crypts that are less sensitive to ambient air temperature changes, so the effect of temperature changes on MEL development speed may be less pronounced than other vector species that develop in open, standing water, like ditches and containers. Precipitation increases the amount of standing water and increases the chances of larval survival during the mosquito season and also during the winter months for those species that overwinter in the larval stage. For EEE, the main enzootic vector, MEL, overwinters as larvae.

While these relationships are well known, they are not well described for EEE, especially in terms of quantifying the role these two factors play in EEE transmission. Specifically, there’s a gap in the understanding of how these factors influence the number of EEE vectors mosquitoes collected in traps and the rate of virus infection among those mosquitoes.

## Data collection and processing

### Mosquito surveillance data

The mosquito surveillance dataset encompasses counts by species from all mosquito traps deployed in Bristol County, MA, from 2007 to 2023. The raw dataset includes over 45,000 records and includes temporal and spatial variable. The BCMCP, perform mosquito surveillance from June through October annually.

During this interval, CDC light traps baited with carbon dioxide are placed at various sentinel sites each week. Designed to attract mosquitoes in search of blood meals, light traps serve as a reliable measure of the actively feeding mosquito population, thereby establishing them as the standard method for EEE surveillance. Some of the sentinel site have been used for over 50 years.

Additionally, the county employs gravid traps, which are specifically aimed at attracting and capturing gravid *Culex* mosquitoes looking for egg-laying sites. While these traps are primarily used for West Nile Virus (WNV) surveillance, they hold significant value for EEE monitoring as well. Given the role of *Culex* species as bridge vectors for EEE, their capture in gravid traps provides critical insight into the risk of virus transmission.

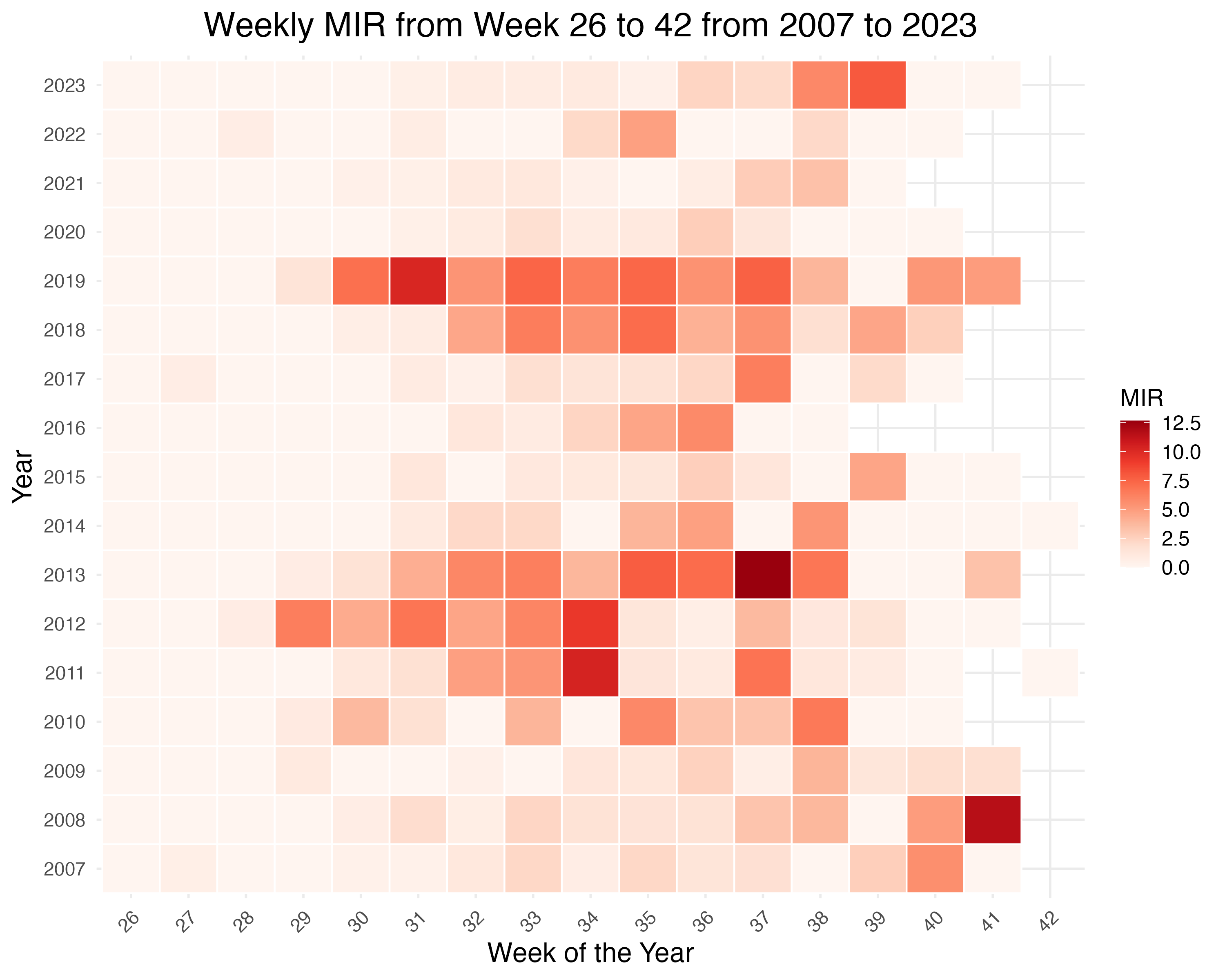
The traps are installed in the afternoon and retrieved the following morning. After the traps are collected, the next are stored in a freezer to preserve and kill the mosquitoes. The mosquitoes are sorted by species using a dissecting scope. Counts for each species are recorded in the database and a portion of the vector species are driven to the Massachusetts State Laboratory in Boston. The lab tests each mosquito pool (or batch) for WNV and EEEv using PCR. Results are shared with BCMCP on the same day of testing. Virus isolation results are recorded in the collection records.

In addition to mosquito counts by species, the dataset includes the town or city in which they were collected, the collection date, and whether the mosquitoes were submitted for virus testing.

### Minimun infection rate (MIR)

One indicator of the level of virus activity present in the mosquito population can be calculated from the information contained in the mosquito dataset. The minimum infection rate (MIR) can be calculated at various temporal and spatial scales from trap site over one day to all sites over an entire season. MIR is calculated as:

MIR provides an estimate of the prevalence of a specific pathogen within a mosquito population. While it does not directly measure the risk to humans, a higher MIR in mosquitoes that commonly bite humans can indicate a higher risk of transmission. A study pubulihsed in 2007 indicated that in years with human cases in Massachusetts between 1979 and 2004, MIR was ≥1.0 by week 31 (Hachiya et al. 2007). MIR allows for the comparison of infection rates across different geographical areas or time periods. This is crucial for identifying areas or times of higher risk and for evaluating the effectiveness of control measures.



Minimum infection rate by week

The surveillance dataset is structured to conform with state and federal reporting requirements. As a result, the format of the data must be transformed to in order to make it ready for analysis. Once the data structure is corrected, it can be joined to other relevant datasets using spatial or temporal attributes.

### Weather data

Weather plays a significant role in the lifecyle of the mosquito and influences the rate at which a pathogen can develop in a mosquito. Hourly and weather summaries were downloaded from NOAA’s Climate Data Portal. The datasets cover the years 2007-2023 at the Taunton Airport in Bristol County.

The hourly summaries were used to calculate weather variables, such as the daily maximum temperature and precipitation. While daily summaries provide this information will less processing, the hourly summaries provide more granular data that can be used to calculate additional variables, such as the number of hours above a certain temperature threshold or degree-days. Degree-days is a measure of the amount of heat accumulated over a period of time. It is calculated by summing the difference between the average daily temperature and a base temperature. Degree-days are used to estimate the rate of development of insects and other organisms that are sensitive to temperature. For mosquitoes, degree-days can be used to estimate the rate of larval development and the time between when a mosquito ingests a virus and when it can transmit it (Gu and Novak 2006). To simplify this variable, the number of hours the temperature was ≥50°F and ≤86°F. These thresholds were chosen based on the known temperature range for mosquito development and virus transmission.

In addition to the weather summaries, drought data will be included in the models. West Nile Virus (WNV) research suggests that drought has been a driver in WNV outbreaks. While the lack of water can lead to smaller mosquito populations, drought can lead to higher prevelance of mosquito infection (Paull et al. 2017). Drought data was obtained from the US Drought Monitor and the Federal Emergency Management Agency (FEMA). The data includes weekly drought severity ratings for the entire United States. The data is available at county and regional levels and will be used to calculate the number of weeks in drought conditions and the severity of the drought. For this project, drought data for Bristol County, MA, will be used and incorporated with FEMA East Coast regions (I-IV) drought indicators.

## Data aquisition

With the exception of the mosquito collection data, all other datasets are publicly available for use without restrictions.

## Data import and cleaning

Datasets will be loaded into R in a Quarto document to document and the processing methods will be documented.

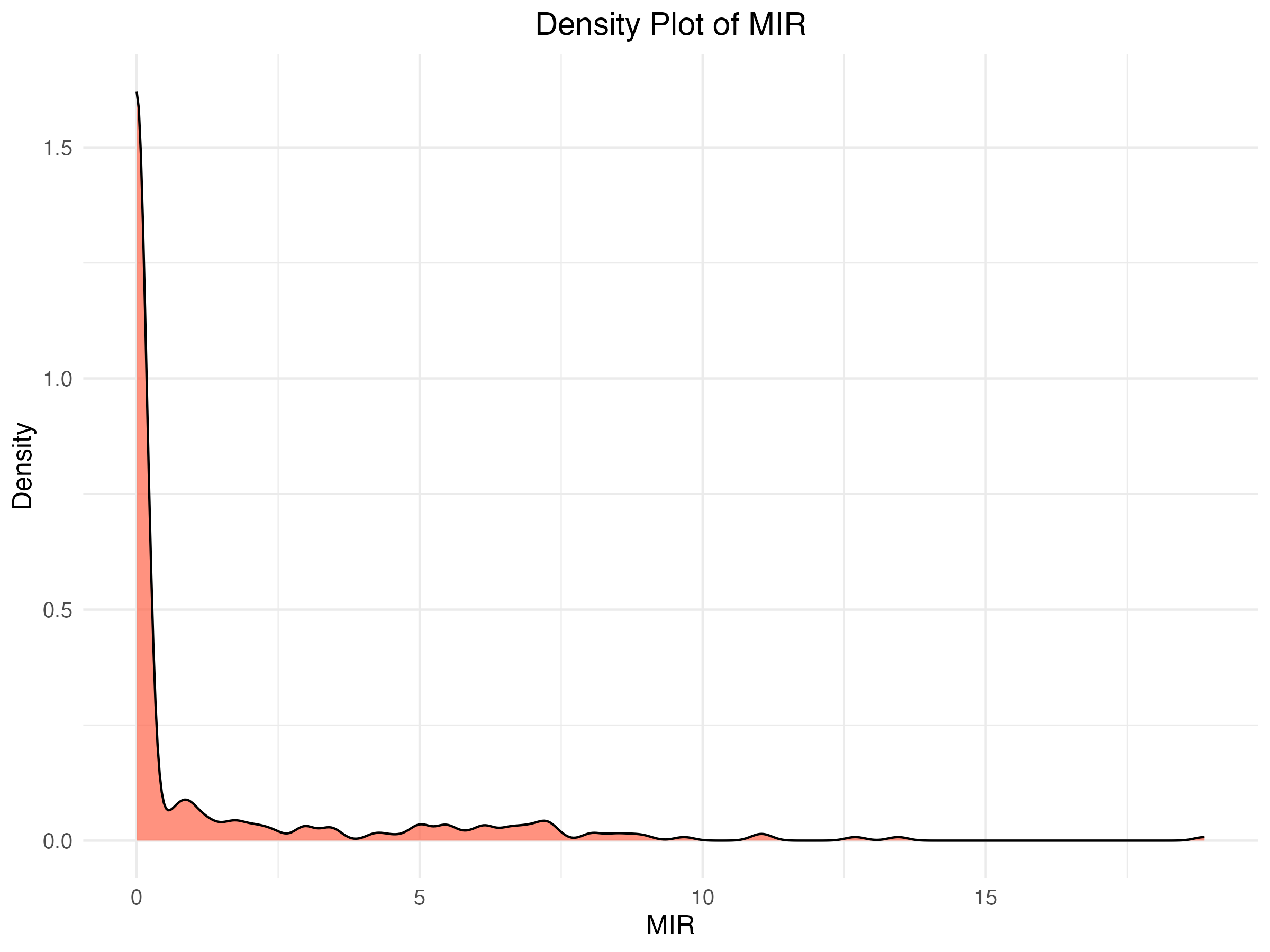
The Massachusetts Department of Public Health requires that all mosquito control districts submit their collection data in a specific format. This format divides each trap event into separate rows based on mosquito species and whether they were submitted for virus testing. However, this format is not useful for certain calculations where a zero count for a species is not explicitly recorded. For example, even when sorted by species to calculate the average count per trap, the average could be inflated since the records will not include trap events with a zero count of that species. In order to correct this, the data was pivoted to create a column for every unique “species code” then the “pool size” for each species code is summed for each trap event. A trap event is when “town”, “date of collection”, “trap type” are all the same. For species codes not included in the orginal table for a trap event, the field will be assigned a zero value. The resulting table will have one row for every trap event.

# Methods

Symptomatic EEE, and to a lesser extent WNV, human cases are rare. On average, the United States records only about 9 cases of EEE annually. Notably, Massachusetts reports more cases than any other state except Florida, highlighting regional disparities in disease occurrence. However, the incidence rate for neuroinvasive EEE cases in Bristol County, Massachusetts, is estimated to be only around 0.06 per 100,000 people (Centers for Disease Control and Prevention 2022). The low incidence of human cases poses significant challenges for using traditional statistical models, making it difficult to accurately estimate human transmission risk. This difficulty is largely due to the sparse nature of the data and the variability in transmission patterns, which complicates efforts to predict outbreaks and necessitates alternative modeling approaches.

MIR will be the main outcome of interest for this project. MIR values will be calculated for each week during the surveillance season over the entire county. To reduce bias, only mosquitoes captured using carbon dioxide-baited CDC light traps will be used to to calculate MIR. These traps are the standard EEE surveillance and are used to monitor the actively feeding mosquito population. Additionally, these trap types have been used in this area for over 40 years, many at the same sites currently in use.

After MIR was calculated and joined with weather and climate data, the distribution of the MIR was examined more closely. About 72% of the records were zero values, which is not uncommon for this type of data. The non-zero values were skewed to the left and there was strong evidence of over-dispersion.



The average MIR was 1.26, with an observed variance of 7.56, resulting in a Variance-to-Mean Ratio (VMR) of 6.00. This pronounced overdispersion suggests that mosquito infection rates were not uniformly distributed over time or space. Instead, they may have been influenced by a range of factors, such as environmental conditions or variations in amplifying host abundance.

The data was visualized to identify trends and relationships between the variables. There were no clear patterns in the data between MIR and the potential predictors. However, several varibles showed a non-linear relationship with MIR. For example, the relationship between MIR and average temperature was not linear. Instead, MIR increased with temperature up to a certain point, then decreased. This suggests that the relationship between temperature and MIR may be more complex than a simple linear model can capture.

In response to the observed non-linear relationships in the data, a two-stage Generalized Additive Model (GAM) analysis was used. GAMs provide a versatile framework for capturing complex, non-linear relationships between response variables and multiple predictors, which are often encountered in ecological data sets.

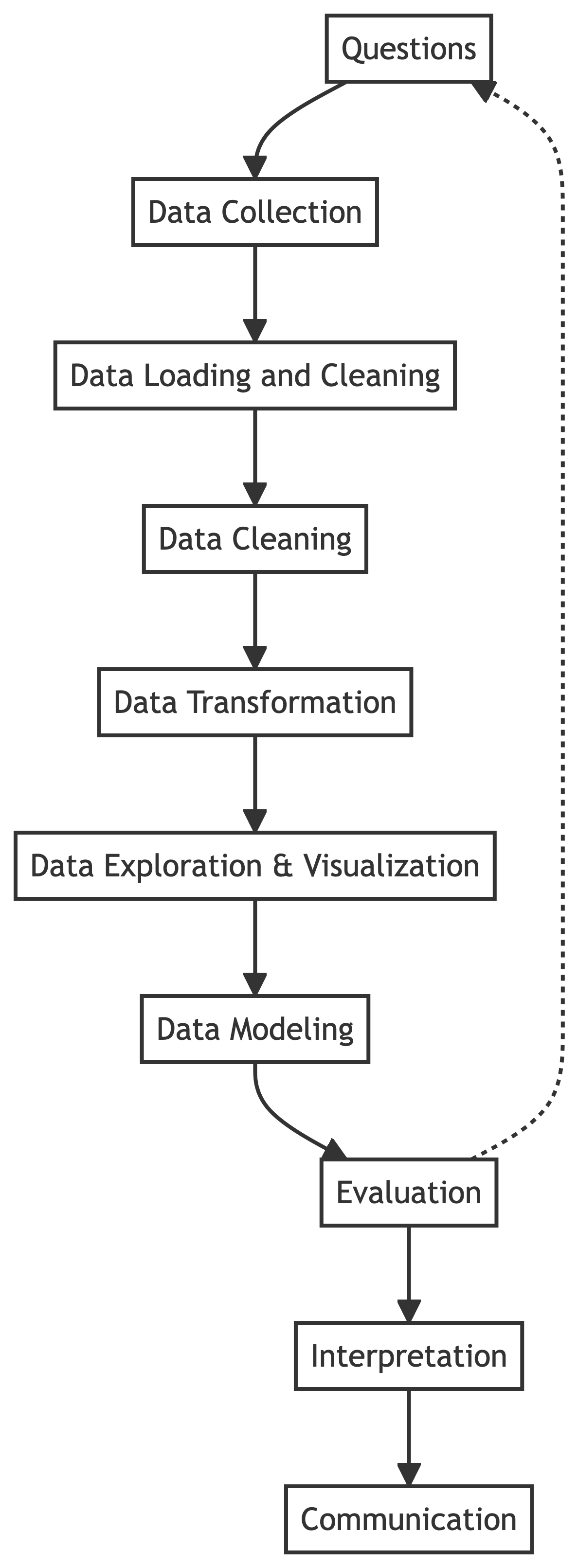
In the first stage of the analysis, the GAM was utilized to address the high incidence of zeros in the dataset—an indicator of overdispersion. This zero-inflated model allowed us to separately model the presence versus absence of the outcome variable, thereby appropriately accounting for the excessive zeros.

Subsequently, the second stage of the GAM focused on modeling the continuous part of the MIR data, relating it to the suite of potential predictors. This step incorporated smooth functions of the predictors, permitting the data to inform the shape of the relationship without imposing a predetermined form, as would be the case in a traditional linear model.

While the two-stage GAM appeared to be a good fit, there were too few observations to provide a robust estimate of the relationship between the predictors and MIR. The original dataset contained over 40,000 observations, but after reformatting the data to fit a rectangular shape needed for analysis, there were less than 400 observations in the final set. After the zeros were removed for part 2 of the GAM, the sample number of observations left was less than 100. Because of this, it was decided to use a single GAM with a dataset that included zero and non-zero observations. Additionally, all trap types were included in the analysis, not just the carbon dioxide-baited CDC light traps. While some precision was lost when the one part model was adopted, the gain in observations was deemed more valuable.

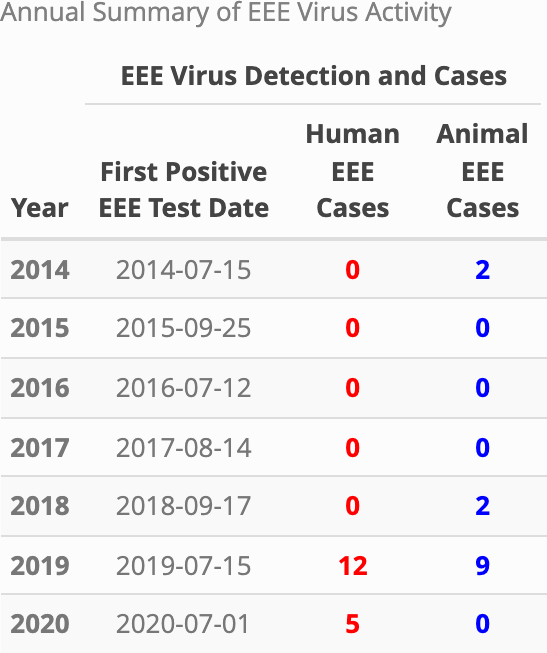
### Schematic of workflow

*Schematic will be refined in later steps. The doc will also be reformatted to move the schematic to an appendix* *The image should appear upon rendering*



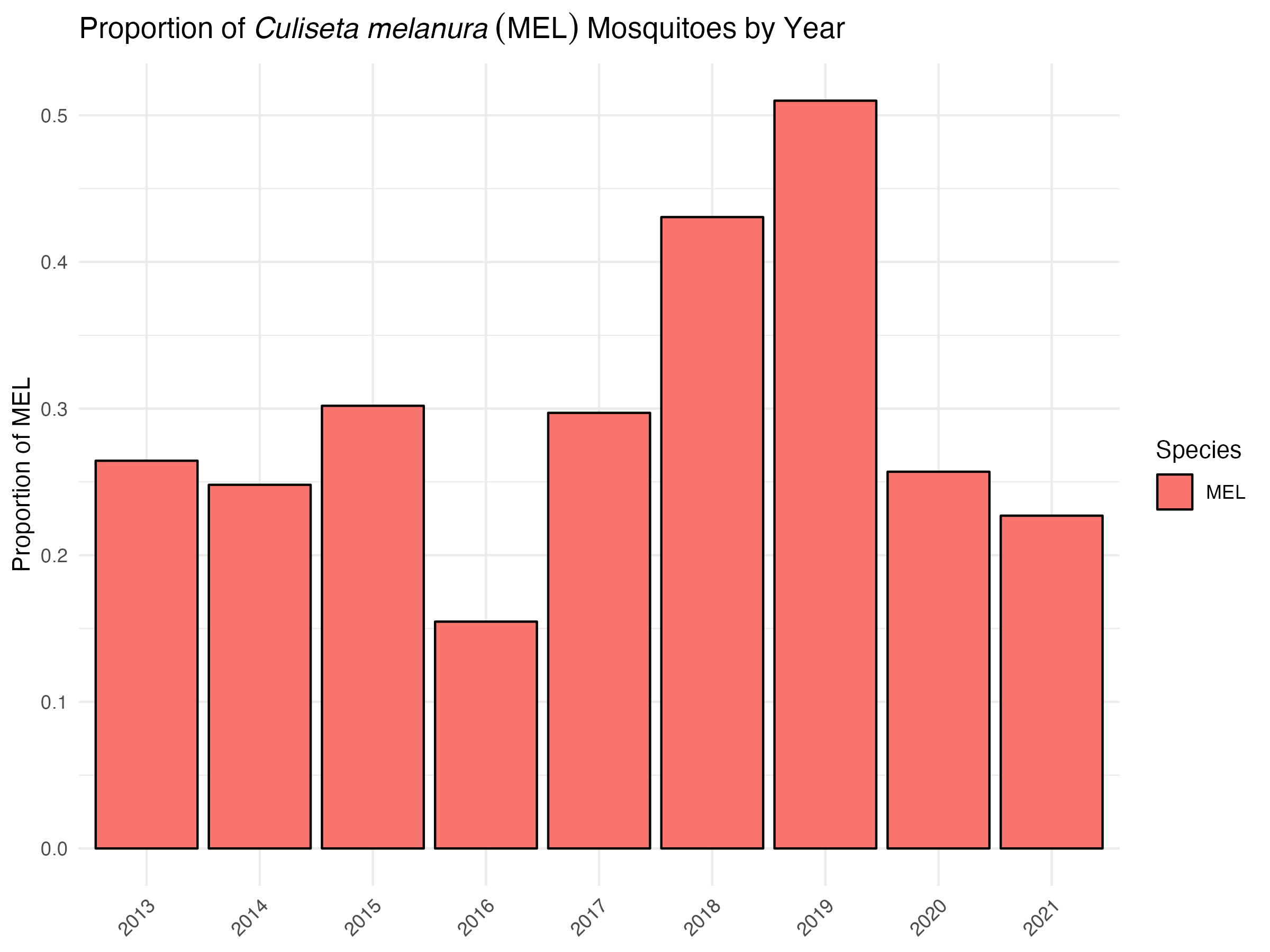
## Exploratory/Descriptive analysis

One factor to consider is if early virus isolations are associated with cases in humans and other mammals. The table below shows the date of the first EEE virus isolation in mosquitoes and the number of human of other animal cases by year.



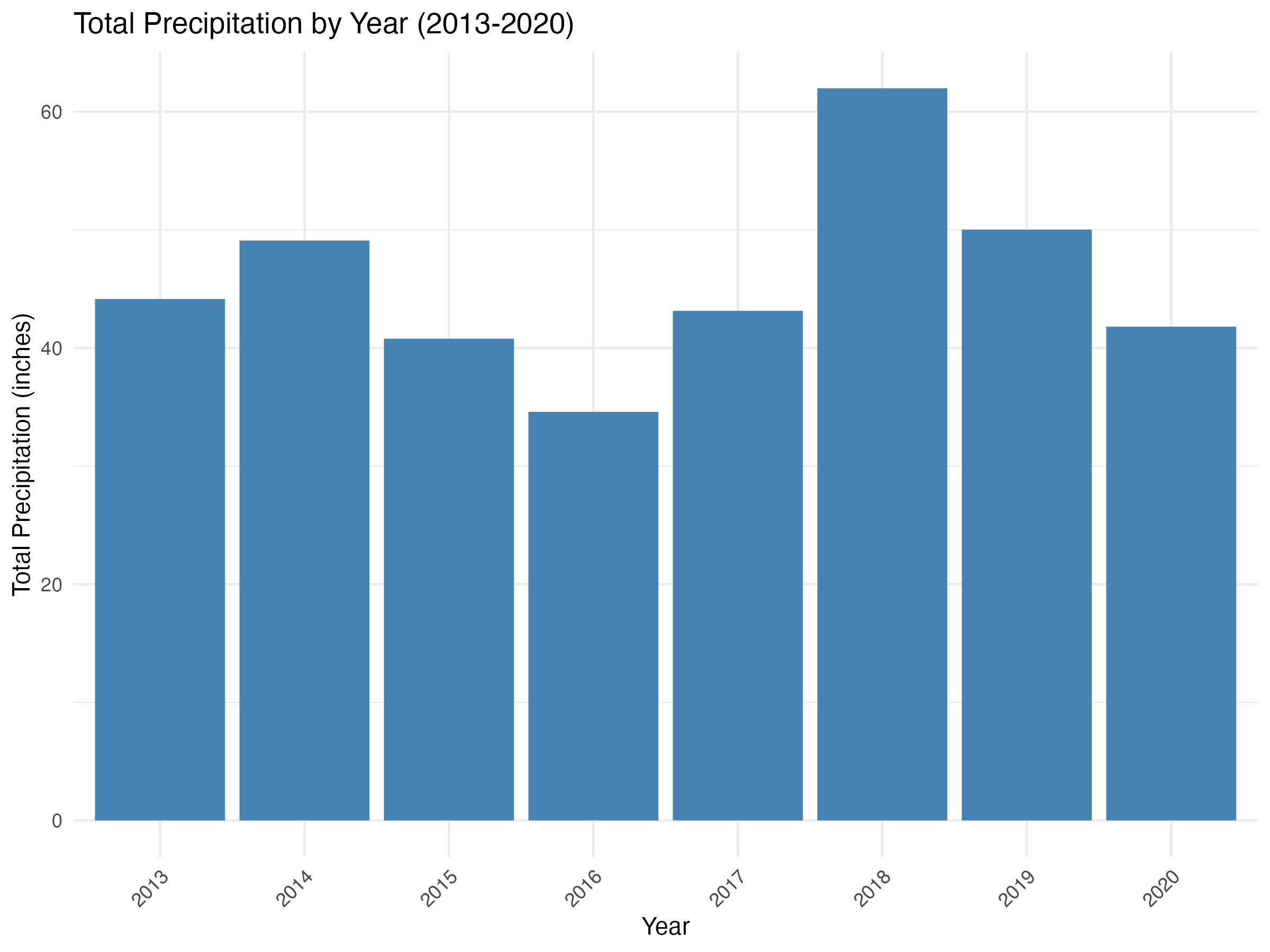
*Culiseta melanura* (MEL) plays an important role in EEE virus amplification in the enzootic cycle. Understanding the population dynamics might provide more insight into what is driving the spikes, like in 2019.

The graph below shows the proportion of MEL among all mosquito vector species trapped in Bristol County, MA.



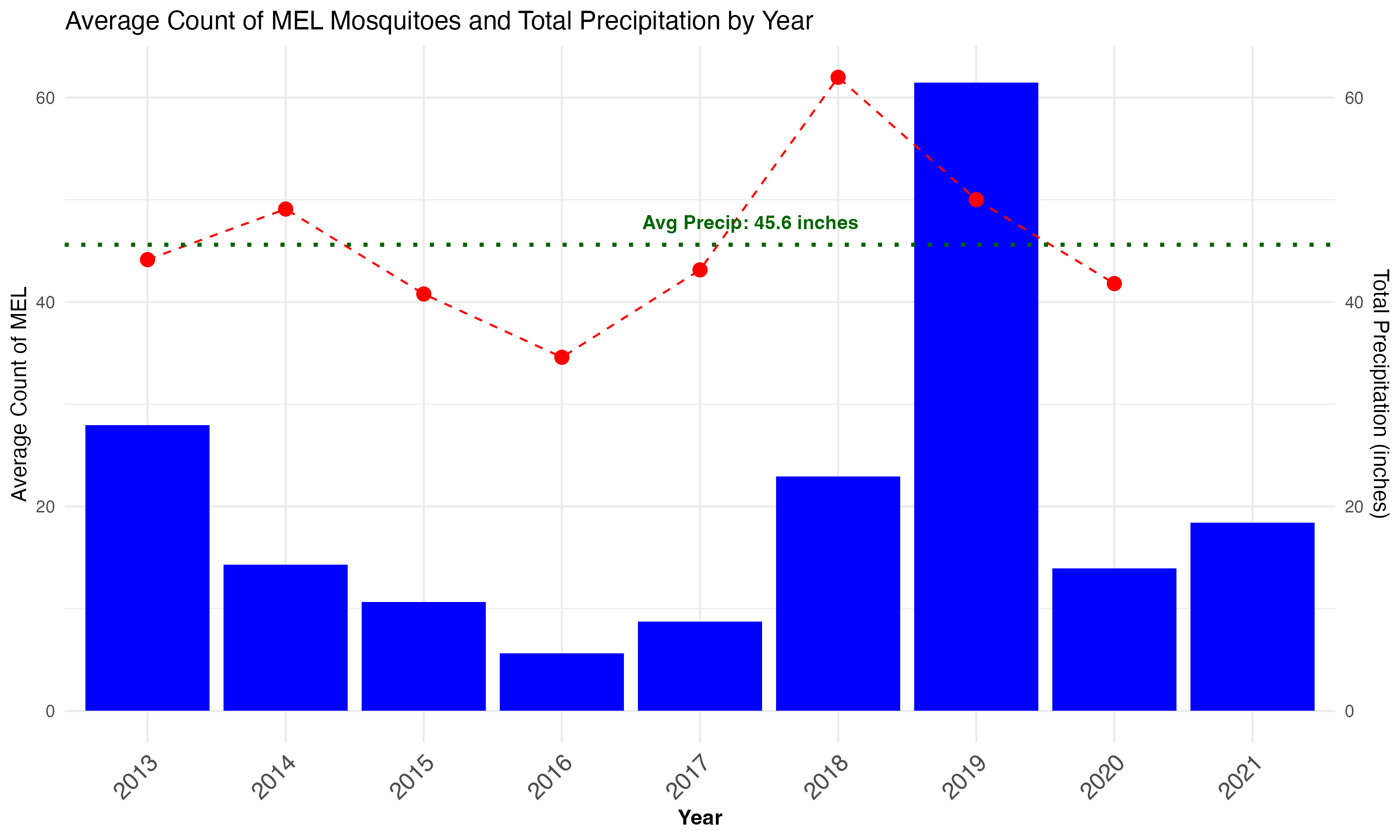
From this we see a large proportion of MEL in the traps for 2019. In fact, they comprise more than half of all vector species for that year. Another interesting aspect is the high proportion in 2018. MEL overwinters as larvae and as long as there is sufficient standing water to last the winter, MEL hatched in the previous fall would survive to emerge the next year. This suggests it might be important to look at conditions in the previous year or season for a better understanding.

Since water plays a significant role in the mosquito life cycle, let’s look at the annual precipitation.



The graph indicates that the annual precipitation total for 2018 is higher than the other years. This suggest that there was enough water to support large MEL larval populations though out the winter.

Another view plots the average annual MEL catch against total precipitation.



## Basic statistical analysis

Basic linear models assessed environmental factors’ impact on the EEE virus transmission cycle. Initial models investigated the link between virus isolation in mosquitoes and weather conditions (average temperature by week and total precipitation). The outcome was whether at least one positive test occurred for a pool in a given week. The predictor in the first model was the average maximum daily temperature from the previous week. The logistic model revealed a significant relationship. Higher temperatures were positively associated with the likelihood of a positive result (odds ratio ≈ 1.069 for each unit increase in temperature; Estimate = 0.06695, SE = 0.02033, p < 0.001). Incorporating Lagged\_Avg\_TMAX as a predictor improved the model, reducing residual deviance to 405.79 from the null deviance of 417.24, indicating a better fit. However, residual deviance on 299 degrees of freedom suggests unexplained variability, indicating other influential factors. The model’s AIC of 411.59 suggests a balance between complexity and predictive power, supporting its utility. Yet, this relationship warrants closer examination due to likely autocorrelation of temperature with progressing weeks during summer months

While precipitation has been demonstrated to influence the population dynamics it did not show to have any effect on a positive result. This was initailly surprising, but upon after considering the specific habitat of the main EEE enzootic vector, *Culiseta melanura*, it makes sense. MEL larvae develop in well insulated crypts that are less sensitive evaporative processes than surface water. The effect of precipitation on MEL larval habitat abundance may be less pronounced than other vector species that develop in open, standing water, like ditches and containers. Another model assessed whether the temperature precipitation in April, May, and June of each year effected the date of the first positive pool test (R file: PT3\_edacode). The date of the first positive pool was identified and assigned a day number from January 1st of that year. The predictors in this linear model were the average maximum temperature for April, May, and June and the sum of precipitation for the same time period. None of these predictors were significant.

There was one other significant result. When MIR is set as the dependent, continuous variable in a linear regression and the average maximum temperature for the previous week is used as the predictor, the model suggests there is a significant inverse relationship. As average temperature increases, the MIR decreases (model\_mir\_temp.R). This relationship is counterintutive, especially given the result of the logistic regression and requires more examination.

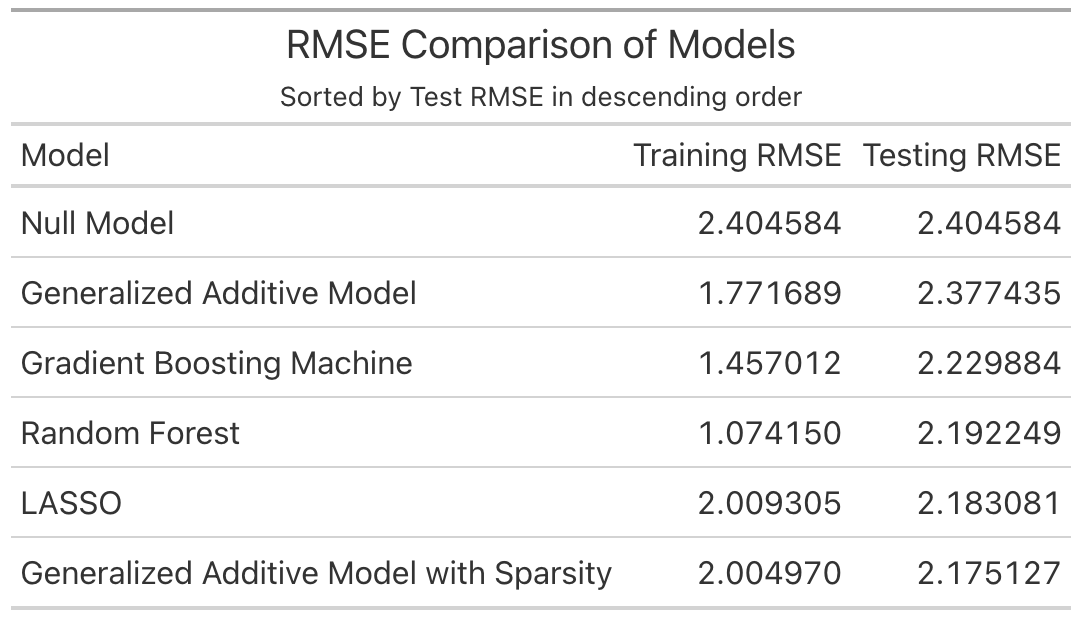
## Statistical analysis

Before any models were assessed, the RSME for the null model was calculated (2.40). This was used as the basis for assessing model improvement.

Analysis began with a Generalized Additive Model (GAM) with predictors that were selected based on the exploratory single variable GAM. The variables were chosen if their p=values were <0.05. Eight variables met that condition including: the number of hours below 50°F the week before the mosquito collection week (Hrs\_Below\_50\_lag), the number of hours below 50°F the week of collection (Hrs\_Below\_50\_current), the drought severity index for East Coast FEMA regions (DSCI\_fema), the average hourly relative humidity the week of collection (Avg\_HourlyRelativeHumidity\_current), the average hourly windspeed the week of collection (Avg\_HourlyWindSpeed\_current), the average number of daylight hours per day (average\_daylight\_hours), the average number of MEL captured by trap (Average\_MEL), and the proportion of MEL captured out of the total number of mosquitoes trapped (MEL\_Prop).

All models included training (.75) and a testing (.25) dataset that were generated using the random seed 123. All analysis can be found in the analysis.qmd file. The first model was a GAM with all eight predictors, followed by a Random Forest, LASSO, Generalized Additive Model with Sparsity (GAMS), and a Gradient Boosting Machine (GBM) model.

As expected, the RSME is lower for the training dataset than the testing dataset for all models. The RMSE values are displayed in the table below.



# Discussion

The null model RMSE was 2.40. All models showed improvement over the null model on both the training and testing datasets. The GAM shows significant improvement over the null model for the training set, but only slightly improves over the null for the testing set. This suggests that beyond capturing the mean response, this GAM is not well generalized. The substantial improvement in the training set RSME for GBM and Random Forest coupled with less improvement over the null for the testing set, suggest a tendency for overfitting. These models may be capturing the training data’s patterns that do not gerneralize well to unseen data. The the gap between the training and testing RMSE for the LASSO and GAMS models suggest there is less overfitting and are slightly better at generalizing than the others. However, given the non-linear relationships, the LASSO model is likely not appropriate for this analysis.

## Strengths and Limitations

The accuracy and reliability of the mosquito collection data is excellent. The same procedures and sentinel sites have were in place for the entire analytical horizon. Further, there were only two people responsible for sorting, identifying and counting species over the entire period. However, the data is not without limitations. Mosquito surveillance is inherently biased. Traps are set in locations that are most likely to capture infected mosquitoes. While this is a positive for early warning of virus activity, it is not necessarily representative of the surrounding area. Probably the biggest limitation is the small number of observations. The mosquito season is short in Massachusetts and funding for trapping and testing mosquitoes is limited. Mosquito control efforts, like pesticide applications, are missing from this analysis. These efforts are designed to lower the levels of virus transmitting mosquitoes and would likely provide more insight into the relationship between the predictors and virus activity.

MIR is not a perfect measure of virus activity in mosquitoes. It tends to underestimate the level of virus circulating in the environment. However, other measures require equal pool numbers for testing. There is a correction factor for variable pool numbers and I will explore this option in later phases of this project. Ideally, the outcome would be human or animal cases. However, these cases are rare and modeling them given the limited amount of surveillance data may not be possible.

Initially, the lack of significance for most temperature and precipitation variable was surprising. However, upon closer examination, it makes sense. The main enzootic vector of EEE and the one targeted for surveillance, *Culiseta melanura*, develops in well insulated crypts that are less sensitive to ambient air temperature changes. The effect of temperature changes on MEL development speed may be less pronounced than other vector species that develop in open, standing water, like ditches and containers. Precipitation increases the amount of standing water and increases the chances of larval survival during the mosquito season and also during the winter months for those species that overwinter in the larval stage. For EEE, the main enzootic vector, MEL, overwinters as larvae.

The spatial scale is an issue with the human and veterinary cases data. Human case data is released only on a county scale to ensure patient privacy. Some veterinary case data is available on the town-level, but most remains only on the county-level. These limitations are barriers to accurate spatial analysis. The project will rely on virus isolation in mosquitoes to model risk. The available bird data relies heavily on crowd-sourced reports. There is one standardized survey (Breeding Bird Survey) that seems to be more accurate and reliable. However, this data is not available for every year.

Likely the biggest limitation ## Conclusions While all models demonstrated improvement over null model, none of them were robust enough to be used for reliable prediction. In the next phases of this project, the data will be expanded to include more years and cover a larger portion of Massachusetts. In additon to increasing the the number of observations, other sources of data will be incorporated. Including virus activity from surrounding regions and areas from which birds migrate to Massachusetts.

# References

Centers for Disease Control and Prevention. 2022. “Prevention Eastern Equine Encephalitis CDC.” https://www.cdc.gov/easternequineencephalitis/prevention/index.html.

Commonwealth of Massachusetts Division of Fisheries & Wildlife. 2007. “Natural Community Fact Sheet: ATLANTIC WHITE CEDAR SWAMPS.” Westborough, MA.

Gu, Weidong, and Robert J. Novak. 2006. “Statistical Estimation of Degree Days of Mosquito Development Under Fluctuating Temperatures in the Field.” *Journal of Vector Ecology* 31 (1): 107–12. <https://doi.org/10.3376/1081-1710(2006)31[107:SEODDO]2.0.CO;2>.

Hachiya, Masahiko, Matthew Osborne, Cynthia Stinson, and Barbara G. Werner. 2007. “[Human Eastern Equine Encephalitis in Massachusetts: Predictive Indicators from Mosquitoes Collected at 10 Long-Term Trap Sites, 1979-2004](https://www.ncbi.nlm.nih.gov/pubmed/17297037).” *Am J Trop Med Hyg* 76 (2): 285–92.

Lindsey, Nicole P, J Erin Staples, and Marc Fischer. 2018. “Eastern Equine Encephalitis Virus in the United States, 2003-2016.” *Am J Trop Med Hyg* 98 (5): 1472–77. <https://doi.org/10.4269/ajtmh.17-0927>.

Massachusetts Department of Public Health. 2024. “Arbovirus Surveillance Plan and Historical Data Mass.gov.” https://www.mass.gov/lists/arbovirus-surveillance-plan-and-historical-data.

Morens, David M., Gregory K. Folkers, and Anthony S. Fauci. 2019. “Eastern Equine Encephalitis Virus — Another Emergent Arbovirus in the United States.” *N Engl J Med* 381 (21): 1989–92. <https://doi.org/10.1056/NEJMp1914328>.

Paull, Sara H., Daniel E. Horton, Moetasim Ashfaq, Deeksha Rastogi, Laura D. Kramer, Noah S. Diffenbaugh, and A. Marm Kilpatrick. 2017. “Drought and Immunity Determine the Intensity of West Nile Virus Epidemics and Climate Change Impacts.” *Proc. R. Soc. B.* 284 (1848): 20162078. <https://doi.org/10.1098/rspb.2016.2078>.

Vander Kelen, Patrick, Joni A Downs, Thomas Unnasch, and Lillian Stark. 2014. “A Risk Index Model for Predicting Eastern Equine Encephalitis Virus Transmission to Horses in Florida.” *Appl Geogr* 48 (March): 79–86. <https://doi.org/10.1016/j.apgeog.2014.01.012>.