DRAFT: Factors Influencing Abundance & Virus Isolation in Mosquitoes

Which environmental, ecological, and temporal factors impact mosquito populations

Andrew Ruiz

2024-03-07

# 1. Summary/Abstract

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. About 30-50% of symptomatic cases lead to death and those who survive are left with life-long disability. The transmission cycle for EEE includes two distinct cycles: an amplifying enzootic cycle in which the virus is transmitted between the Culiseta melanura mosquito and birds and an epizootic cycle in which bridge vector mosquitoes transmit the EEE virus (EEEv) from birds to humans and other mammals. Because C. melanura feeds almost exclusively on birds, it is not considered a direct human threat. However, EEEv isolations in C. melanura are an early warning sign that EEEv is circulating in the ecosystem. Isolations in bridge vector mosquito species indicate heightened transmission risk to humans.

This project will lay the foundation for my dissertation research. **The goal for this phase of the project is to answer which environmental, ecological, and temporal factors influence the transmission cycles for EEE in southeastern Massachusetts.**

# 2. Introduction

## 2.1 General Background Information

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%. Those who survive often experience lifelong disability that includes debilitating neurological damage. On average, there are 11 cases reported in the US annually. Most occur in Massachusetts and Florida. 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. 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. 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 20 years. Once trapped, mosquitoes are sorted 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.

## 2.2 Description of data and data source

*The cornerstone of this research will be the mosquito collection data provided by a mosquito control project in SE Massachusetts. A formal data-sharing agreement will be in place by January 2024. The dataset encompasses nearly two decades of detailed mosquito surveillance. The datasets include spatial identifiers. Mosquito collection data will be integrated with other datasets using spatial or temporal attributes. These datasets will include biological and ecological data, environmental and spatial data, and demographic data, which are crucial in understanding the transmission dynamics of EEE. Potential datasets with their sources are outlined in the table below. Other datasets may be included based on availability and need.*

| Data | Description | Source |
| --- | --- | --- |
| Mosquito Surveillance Data | Records of mosquito species counts, locations, and dates of collection. Data on the presence of EEE in mosquito populations is particularly valuable. | XXX County Mosquito Control |
| Weather Data | Temperature, humidity, rainfall, and wind speed influence mosquito activity, population dynamics, and virus transmission. | NOAA |
| Bird Population Data | Since birds are a natural reservoir for EEE, information on bird populations and migration patterns could be relevant. | eBird |
| Human Case Reports | Data on confirmed human cases of EEE, including location, date of onset, and clinical outcomes. | MA Dept of Public Health publications |
| Veterinary Surveillance Data | Since horses and other mammals can also be affected by EEE, veterinary records could provide early warning signs of virus activity in an area. | MA Dept of Public Health |

## 2.3 Questions/Hypotheses to be addressed

\*This project will lay the foundation for my dissertation research. **The goal for this phase of the project is to answer which environmental, ecological, and temporal factors influence the transmission cycles for EEE in southeastern Massachusetts.** While the following questions will not be addressed in this phase, this project will hopefully inform research that will answer the following questions:

1. Can an AI-driven model that leverages historic mosquito surveillance, ecological, and environmental data accurately quantify the risk of human EEEv infection in southeastern Massachusetts?

2. What is the potential of machine learning algorithms to identify early warning signals for EEE outbreaks, enabling timely public health interventions in Massachusetts to prevent human infections?

3. How can machine learning models use real-time data on mosquito activity, environmental factors, and weather to accurately predict when the risk of EEE transmission to humans is low enough to lift evening outdoor activity restrictions? \*

# 3. Methods

*Specific methods will be determined when all data sources are confirmed.*

## 3.1 The process will involve creating a relational database in which all datasets are joined using spatial or temporal attributes. This will lead to a thorough examination of the datasets to understand the distributions, identify missing data and outliers, and uncover potential correlations between variables. The EDA will involve using visualizations, geographic information systems, and statistical tests to understand how biological, ecological, environmental, and demographic forces influence the EEE transmission cycles. The EDA will inform the preprocessing steps, feature engineering, and selection of appropriate AI models.

### 3.1.1 Schematic of workflow

*This section will be updated before analysis begins*

## 3.2 Data aquisition

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

## 3.3 Data import and cleaning

Datasets will be loaded into R in a Quarto document to document and describe cleaning methods. There are two datasets that will require more extensive cleaning: the mosquito trap data and the crowd-sourced bird data. The process is outlined in detail in the MADAproject\_pt2.qmd file. Less significant cleaning and data management steps not included in this manuscript are also outlined in that file.

The mosquito dataset spans 2007 to 2023. However, the virus and case data only covers the years 2014 to 2020. Because mosquito populations can be influences by processes of the previous year. We will set the analytical horizon will span 2013 to 2021.

Additionally, the Massachusetts Department of Public Health requires that all mosquito control district submit their colelction data in a specific format. This format divides each trap event into separate rows based on mosquito species. 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.

The bird observation dataset was obtained from Cornell University’s eBird program. The original dataset included all reported bird observations for Southeast Massachusetts and was over 600 MB of data. While all birds are potential amplifying hosts for EEE and West Nile Virus (WNV), birds from the Genus *Corvus* are of particular concern. The bird dataset processed to include records of *Corvus spp* sightings after 2014. The final dataset included 8 of the 50 variables, including location data.

## 3.4 Statistical analysis

This exercise will serve as the beginning of the exploratory data analysis phase. One of the main objectives for of this phase is describe the data, understand its structure, clean data, and identify associations within the datasets.

# 4. Results

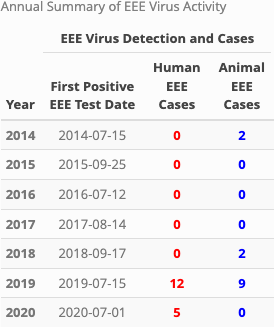
## 4.1 Exploratory/Descriptive analysis

The first phase of the exploration is complete. The next phase will include statistical tests to determine if there are possible associations between the environmental variables and the mosquito abundance as measured by trap counts or EEE virus activity.

It is important to know where the mosquitoes infected with EEE are captured. Below is a table showing the virus isolations for EEE in trapped mosquitoes by county between 2014 and 2020.



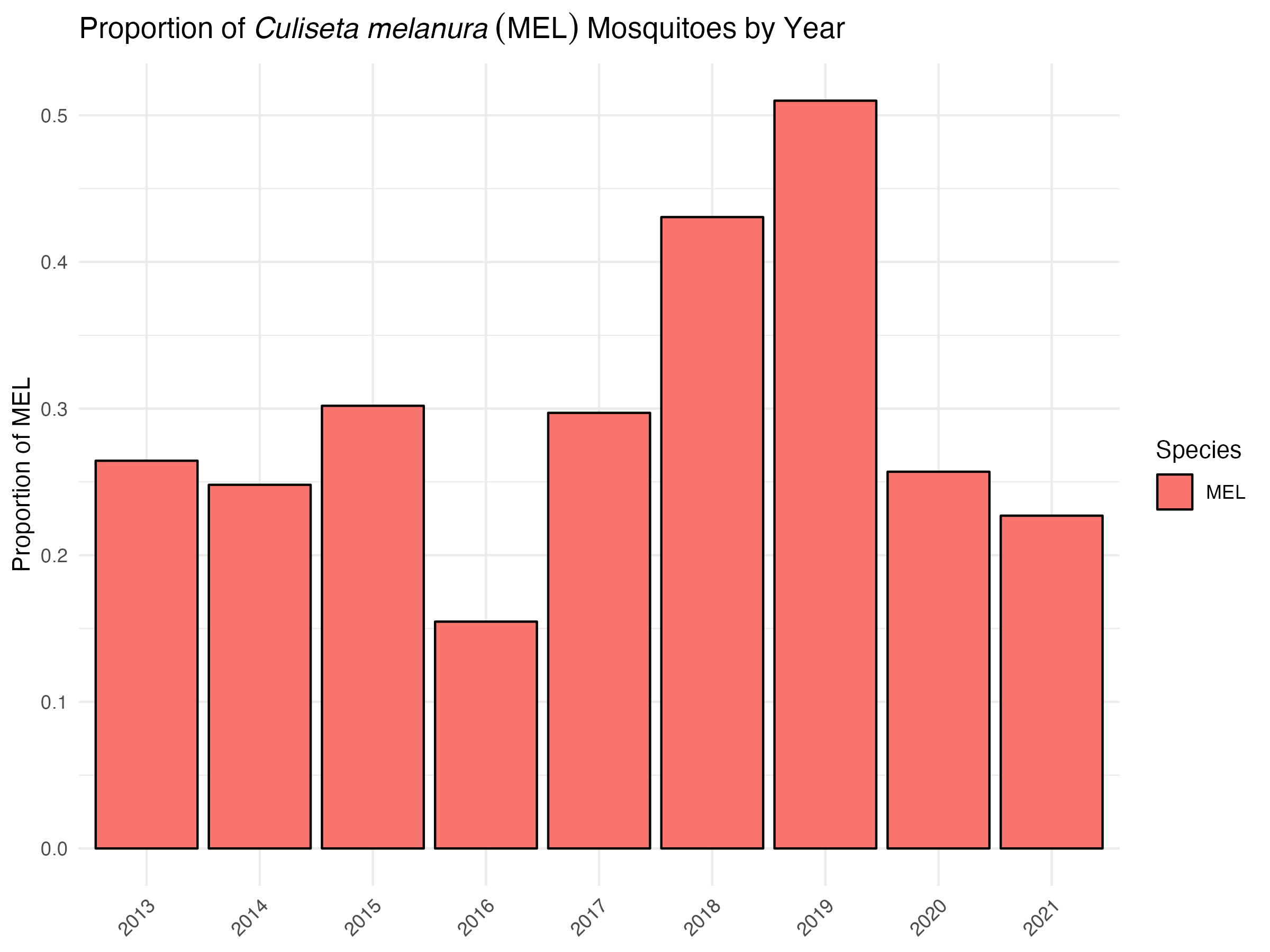
The geographic focus for this study is Bristol County. One factor to consider is if early virus isolations are associated with cases in humans and other mammals. The table below shows the date of the first EEE virus isolation in mosquitoes and the number of human of other animal cases by year.



Unfortunately, the time horizon is limited by data availability. Additionally, EEE human infections are rare, so traditional statistical methods may be inadequate to suggest any relationships. What is clear is that 2019 was an exceptional year for EEE.

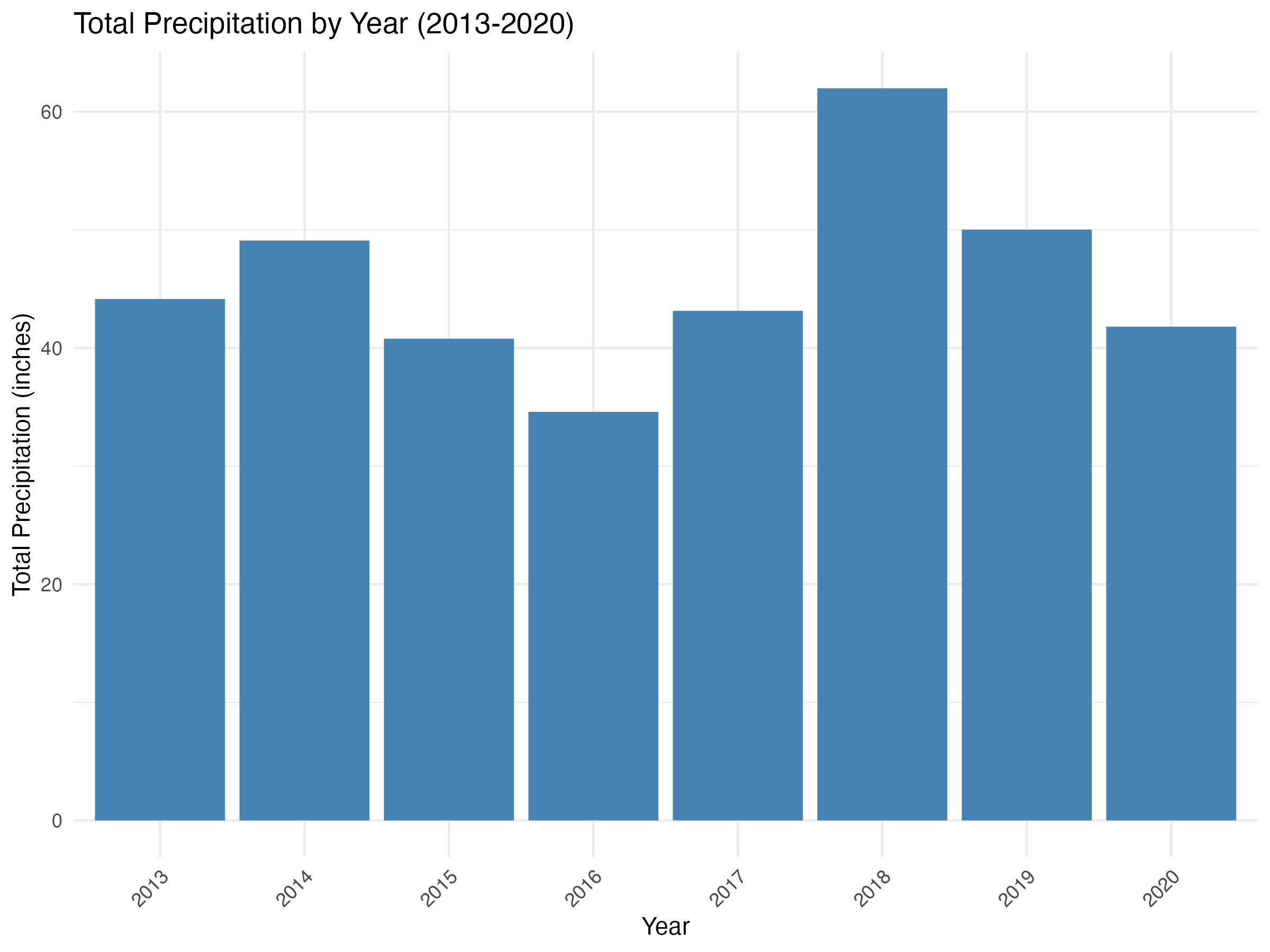
*Culiseta melanura* (MEL) plays an important role in EEE virus amplification in the enzootic cycle. Understanding the population dynamics might provide more insight into what is driving the 2019 spike.

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.

## 4.2 Basic statistical analysis

*These will be generated in later phases of the exercise*

## 4.3 Fullanalysis

*These will be generated in later phases of the exercise* # Discussion *this section will be completed when analysis is done*

## 4.4 Strengths and Limitations

*this section will be updated after analysis is complete*

While the reliability of the mosquito trap data is high, the trap data cannot be directly associated with the virus isolation data. The information this scale is not released by the State. Spatial scale is also 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.

## 4.5 Conclusions *this section will be completed when analysis is done*

# 5. References

References will be added in part 3 of the project