



# Fighting West Nile Virus in Chicago

Christopher, Elizabeth, Gareth



# Table of Contents

- 1) Problem Statement
  - 2) Data Cleaning and EDA
  - 3) Preprocessing
  - 4) Modelling
  - 5) Limitations & Recommendations
  - 6) Cost-Benefit Analysis
  - 7) Conclusion & Recommendations
-

# The West Nile Virus

- Leading cause of mosquito-borne disease
- Potentially fatal
- No vaccine



# Problem Statement



# Problem Statement

## **Project Aim**

- To find the most cost-effective way of eliminating West Nile Virus (WNV) in Chicago

## **Who are we?**

- Part of team at Disease and Treatment Agency

## **Stakeholders**

- Primary- Disease and Treatment Agency
- Secondary- Chicago residents



# Data Cleaning and EDA



# Data Cleaning

Train  
Test

- **Incorrect dtype**
- **Drop (Redundant)**
- **Drop (Duplicates)**
- **Drop (Not in Test)**

# Data Cleaning

Spray

- **Incorrect dtype**
- **Drop (Redundant)**
- **Drop (Duplicates)**



# Data Cleaning

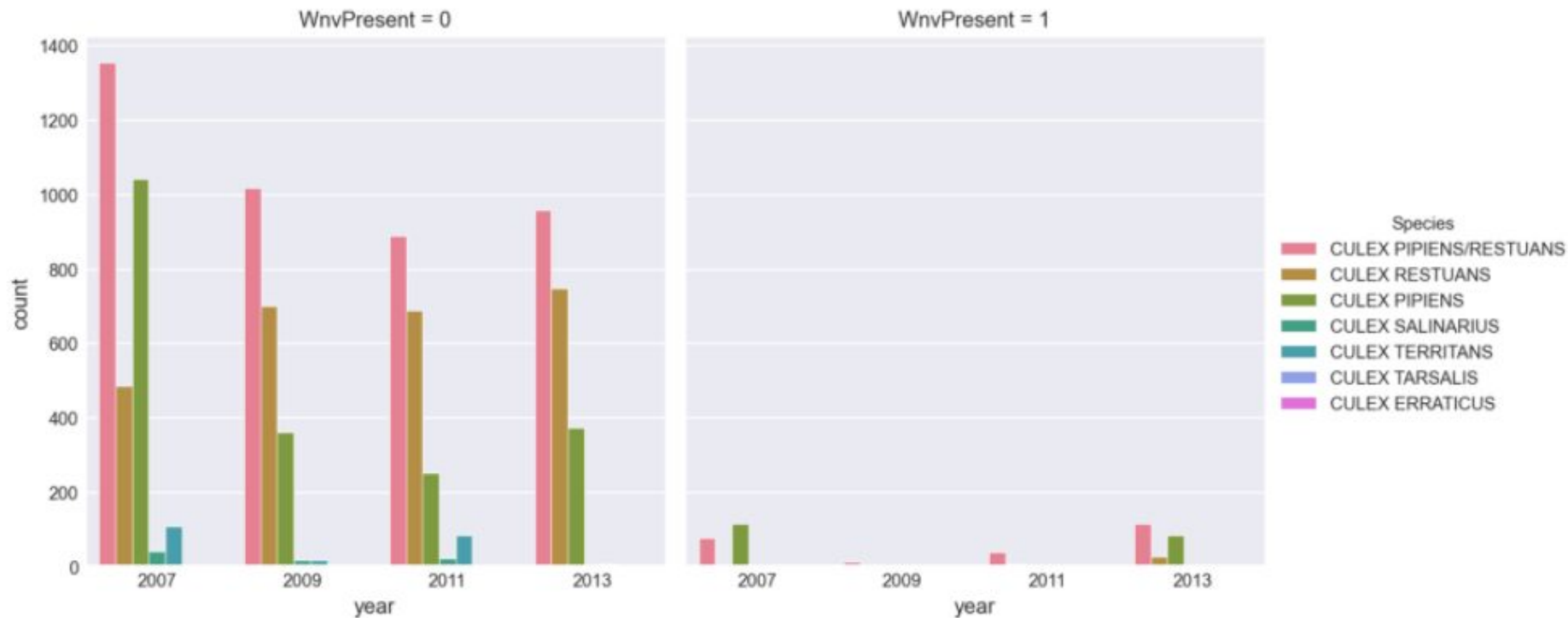
## Weather

- **Incorrect dtype**
- **Drop (Redundant)**
- **Drop (Duplicates)**
- **Impute ("M", "T")**

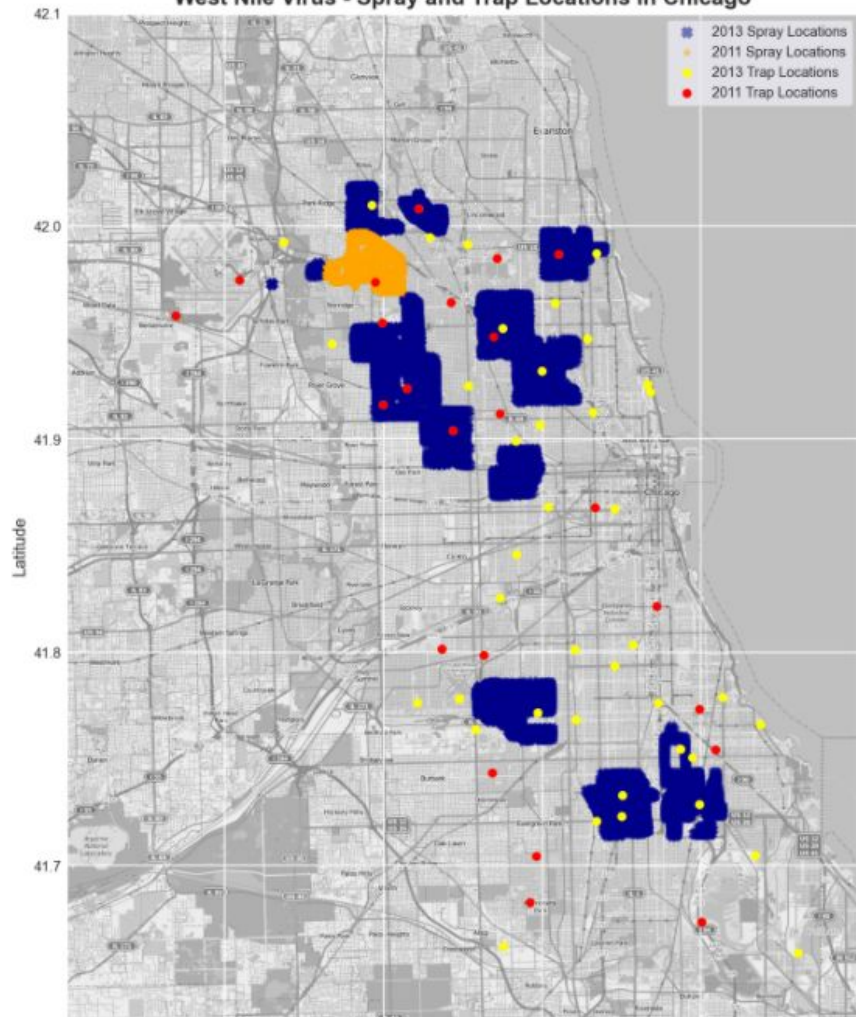


EDA

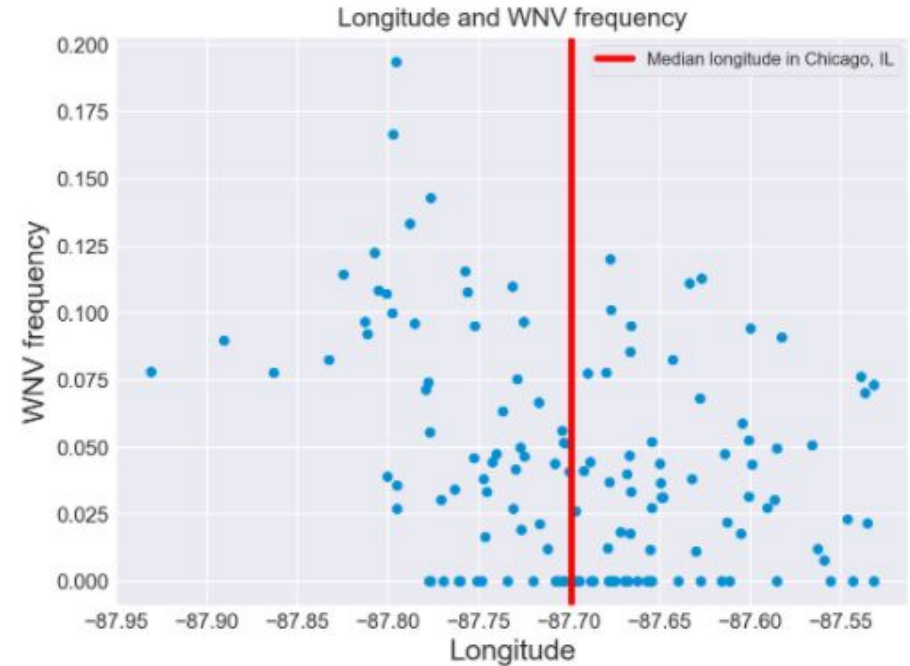
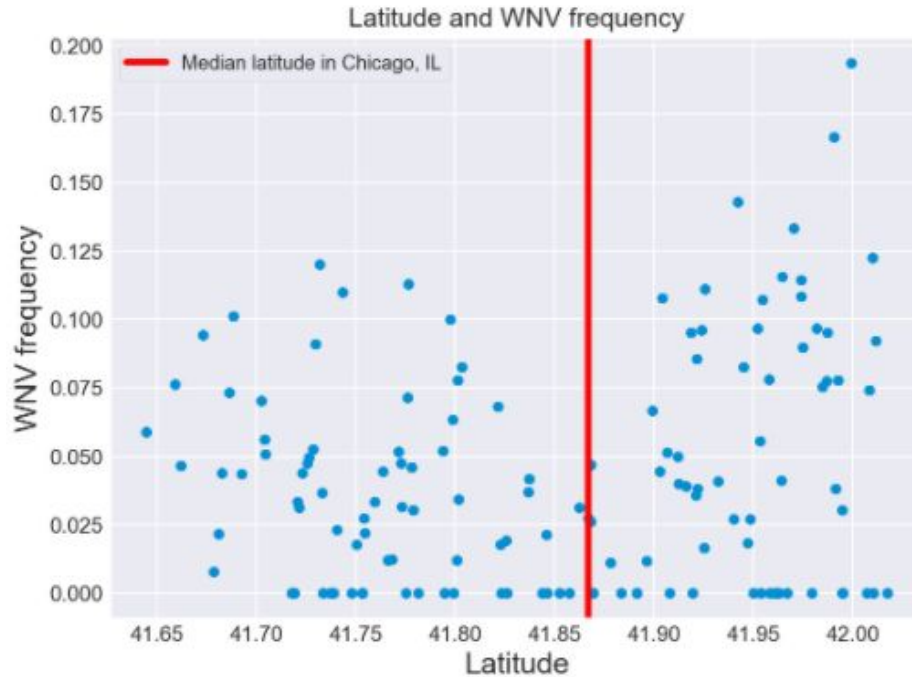
# Count by Mosquito Species



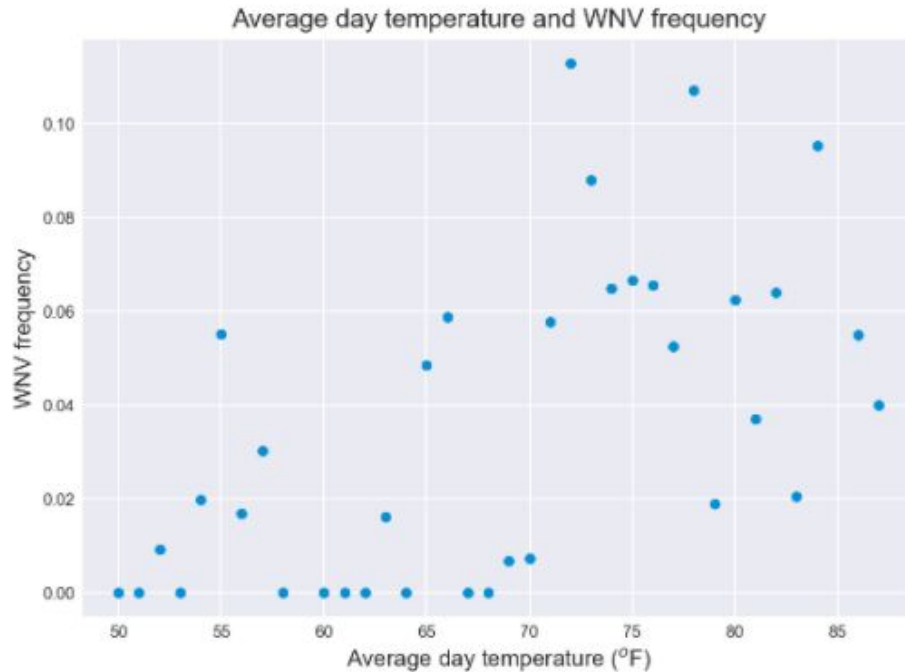
## West Nile Virus - Spray and Trap Locations in Chicago



# Latitude/Longitude and WNV Frequency



# Temperature and WNV Frequency



# Preprocessing

# Preprocessing

## **Feature Engineered**

- Dummy variables for species created in both train and test set.
- Parsed dates into year, month and week of year.
- Merged weather dataset to train and test set on Date.
- Dropped features with collinearity.
- Spray set is omitted due to lack of information over the years.



# Modelling

# Modeling

- Scaled the features
- Used smote to deal with imbalanced classes

## **Models Used**

- Logistic Regression
- Gradient Boosting
- AdaBoost
- K Nearest Neighbours
- Random Forest
- Decision Tree

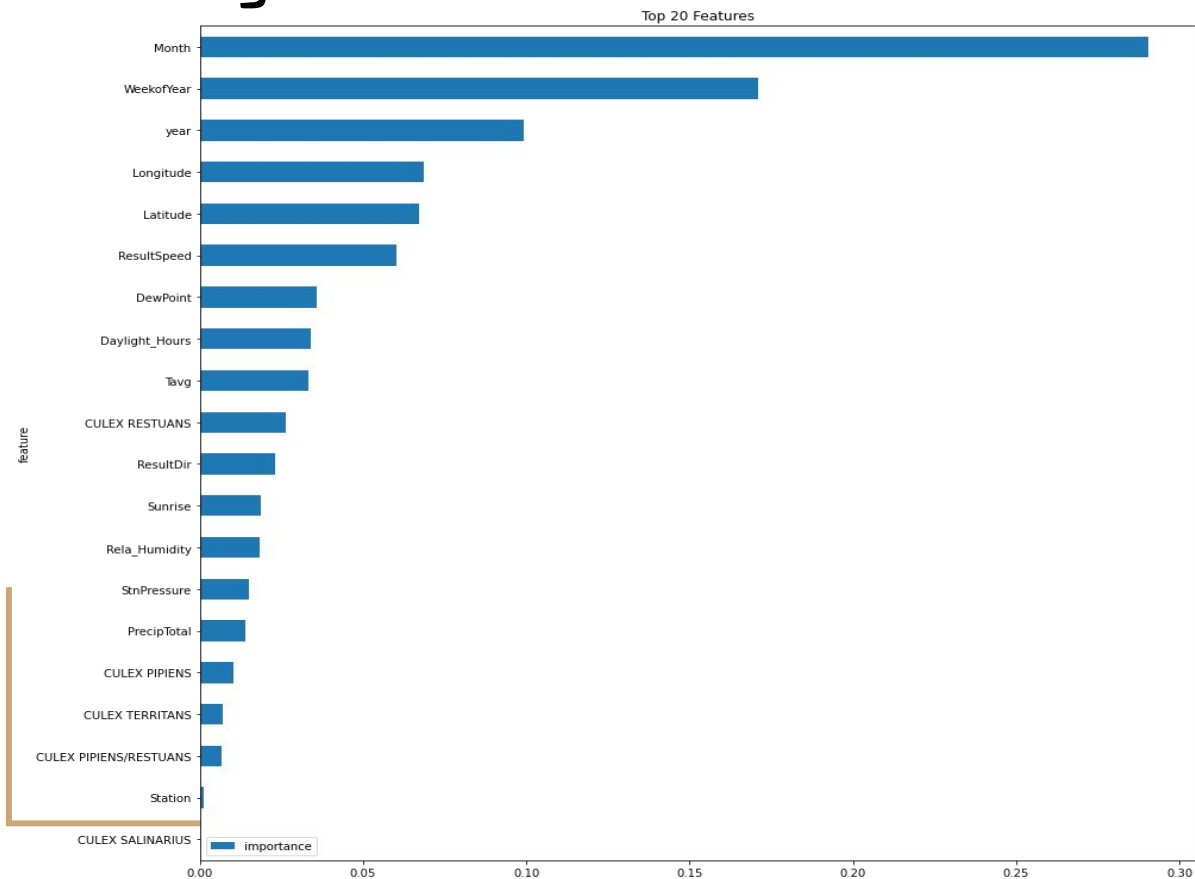
# Modeling

- Based on the models, get the best params
- Used the best params on model training
- Choose the model with the best AUC ROC score

## Top 3 Models

|   | Models            | Local validation score | Kaggle score |
|---|-------------------|------------------------|--------------|
| 1 | Gradient Boosting | 0.854                  | 0.635        |
| 2 | Random Forest     | 0.851                  | 0.653        |
| 3 | Adaboost          | 0.843                  | 0.651        |

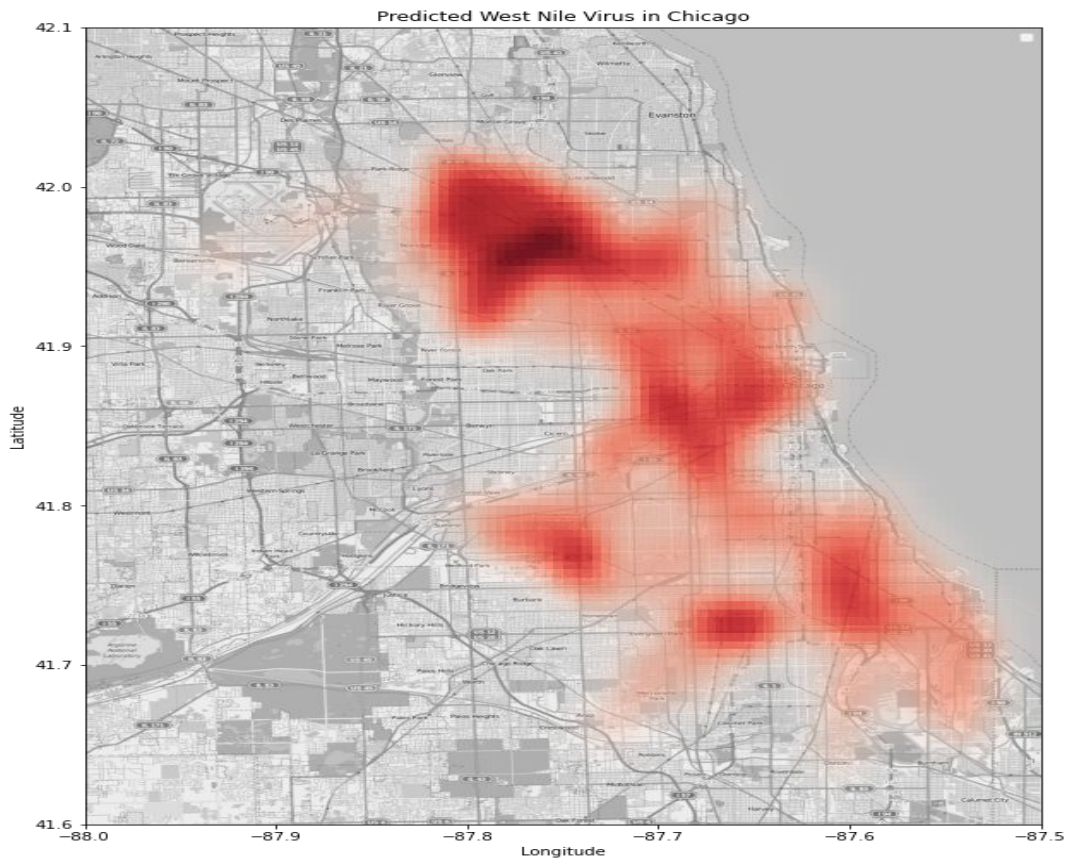
# Modeling



## Top Features

- Month, Week of Year, Year
- Longitude, Latitude
- Daylight Hours
- ResultSpeed
- Average Temperature
- Culex Restuans
- Culex Papiens

# Modeling



# Pesticide Spraying - Cost Benefit Analysis

## Costs

- Economic Costs
- Non - Economic Cost

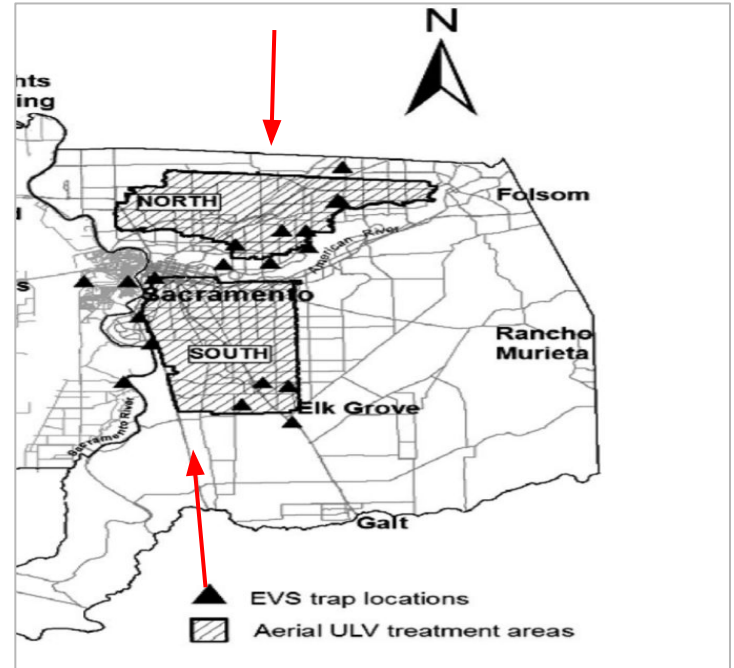
## Benefits

- Efficacy of Spray
- Effectively reduce West Nile Virus



# Sacramento, 2005

- Emergency aerial spray conducted in Sacramento County
- 2 main areas: 477km<sup>2</sup> across 6 nights
- Size of Chicago: 606km<sup>2</sup>
- Costs amounted to \$701,790
- Price of a single helicopter starts at \$100,000 ,  
operating costs of \$80,000 per year



# Results of Aerial Spray

After 12hrs

- Greatest mortality (100%) for cages in open fields - dead after 30 mins
- Mortality among mosquitoes placed in exposed or partially exposed sites - 77.1%
- Mortality among mosquitoes placed in protected places - 24.9%

Overall(Study in 2008)

- Reduced mosquito abundance and the number of infective bites
- Resulted in an approximately six-fold decrease in the relative risk of infection in humans
- No new human WNV cases in either of the treated areas
- 18 new cases reported in adjacent untreated area



# To spray or not to spray?

- Total cost of the 2005 Sacramento County WNV epidemic was around \$2,979,037.
- Costs for treating patients alone exceeded costs of emergency vector control by 3:1 ratio
- Benefits outweigh costs
- Spray event would need to prevent only 15 WNV cases to breakeven

# But...

Long-term adverse health effects that it may have on public health

- Pregnant women and children have a greater risk of getting sick from pesticides
- In NYC in 2000 more people were reported to have gotten sick from pesticide exposure from spraying than from WNV
- Other adverse outcomes include acute asthma attacks, other respiratory problems, and/or dermatological problems

# Conclusion and Recommendation

- Proceed with the vector control measure
- Focus on key areas
  - Areas in Chicago our model predicted WNV is likely to occur
  - Areas where mosquito species like Culex Restuans are commonly found
  - Certain time of the year where temperature is higher and windier: Summer
- Cheaper alternative: Truck Mounted Sprayer - \$85,000
- Long term adverse effects: use vector control as last resort
- Prevention of WNV through public education
  - Remove all potential breeding areas
  - Monitor ponds and sources of water regularly for signs of mosquito larvae