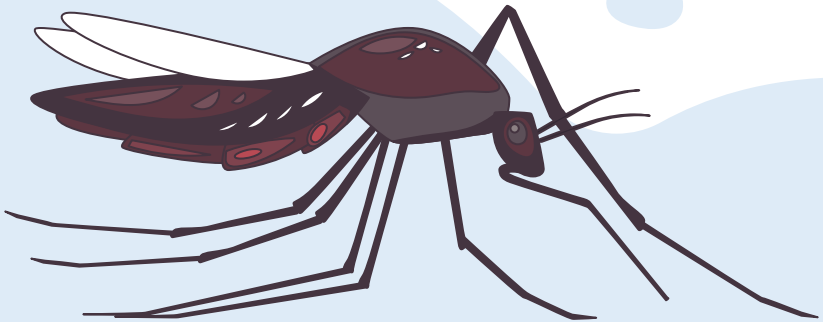
West Nile Virus Prediction

Big problems come in small packages



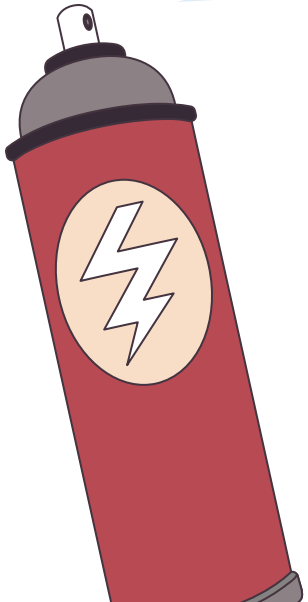
Members

Mubin

Weihan

Zhenming

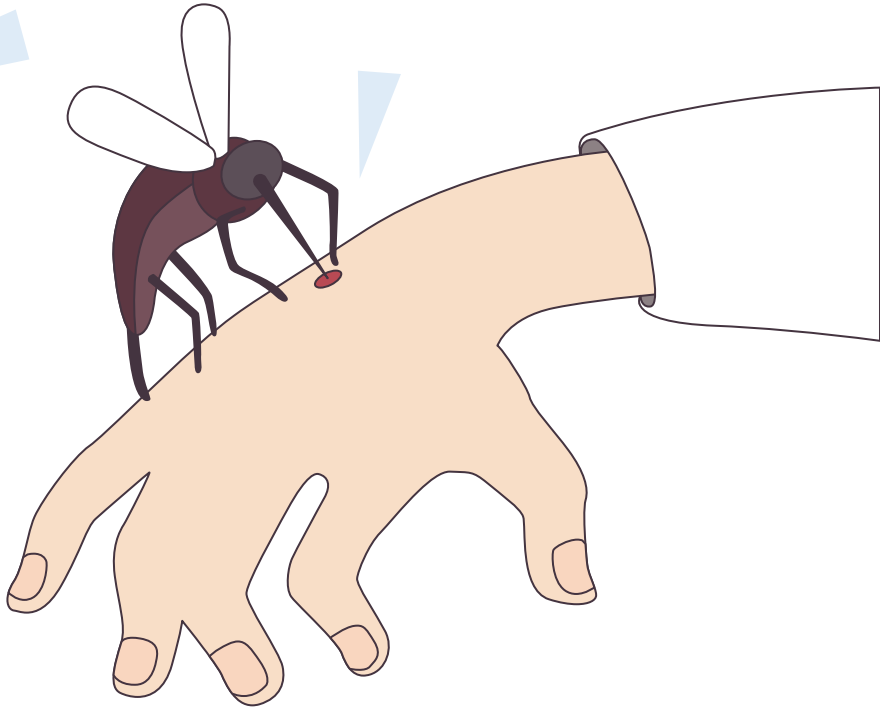
Contents



01

What’s the big deal?

Background & Motivations



2746 cases / year

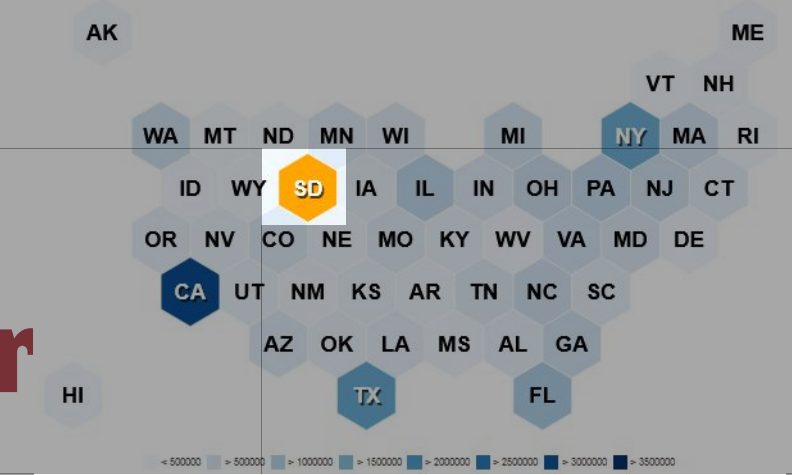
Across the United States of America[1]

1 in 5

Developed symptoms after being bitten

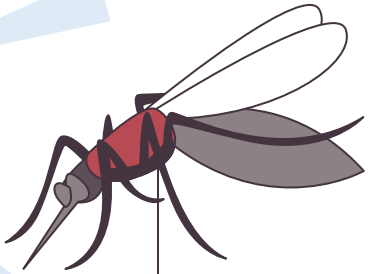
$56 million/year

In medical treatment and productivity losses [2]



1. Average data from 2005 to 2020. Source: https://www.cdc.gov/westnile/statsmaps/cumMapsData.html#three
2. Average cost from 1999-2012. Source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3945683/

Deadly Characteristics

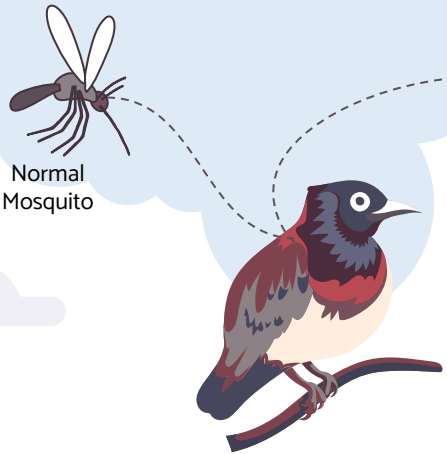


Transmission Mechanism

High Mobility Carriers

Cost of Treatment

Transmission Mechanism

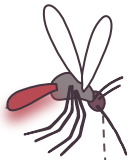


Normal
Mosquito

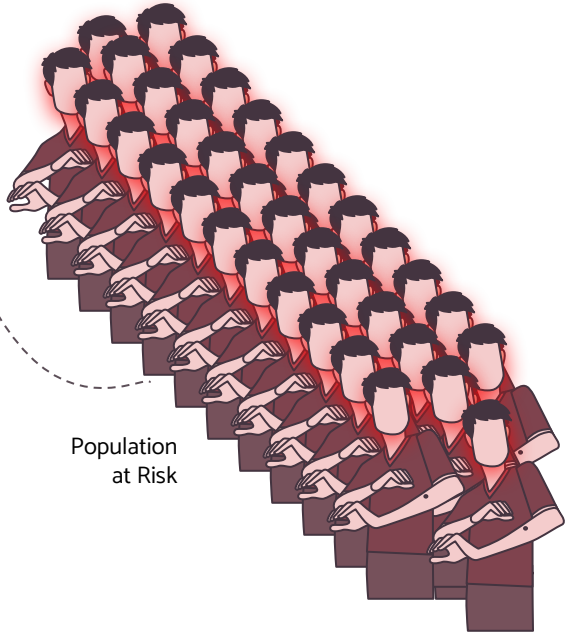
Infected
Bird

Infected
Mosquito

Healthy
Human



Infected
Mosquito



Population
at Risk

Project Objectives

Estimate the number of mosquitoes in an area

Predict the presence of WNV in an area

Estimate the costs of vector controls

Recommend Time and Place to conduct spraying

02

What do we know?

Data Collection, & EDA

Data Provided

Data Source

Chicago Department of Public Health

Date Range

Spring - Summer
2007 - 2014

50-50 split by time for training and validation

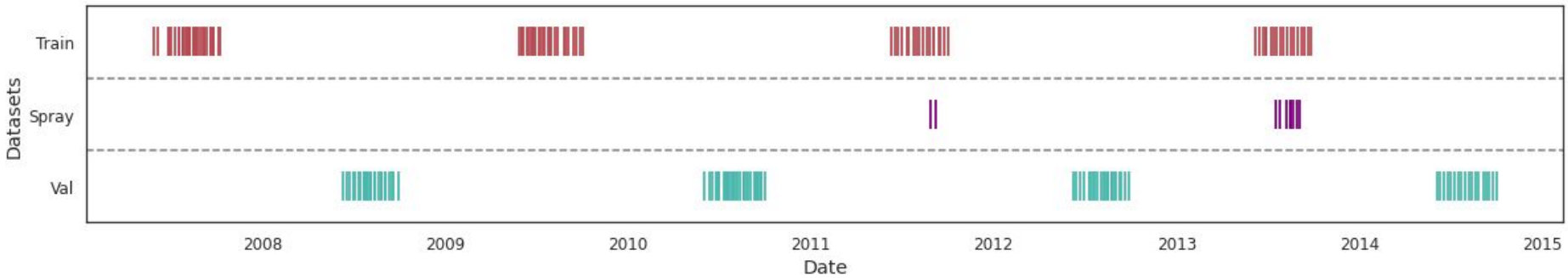
Traps

136 Mosquito Traps
Number of Mosquitoes
WNV Presence
Sprays Conducted

Weather

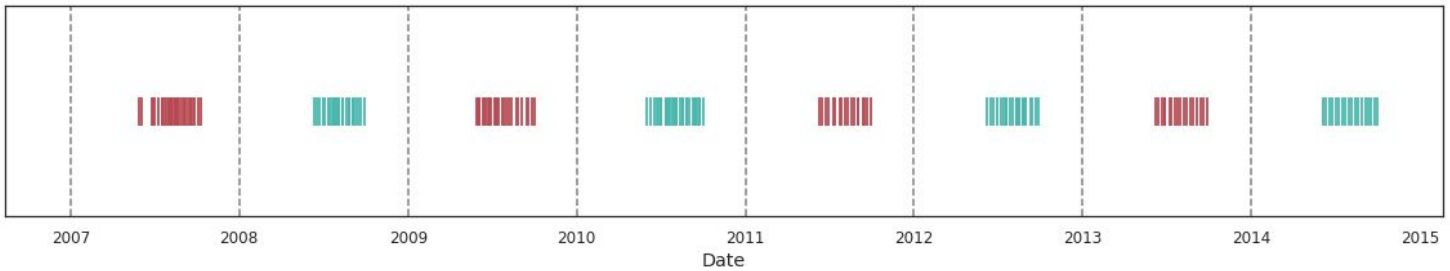
2 Weather Stations
Temperature
Wind Direction/Speed
Humidity

Date Range



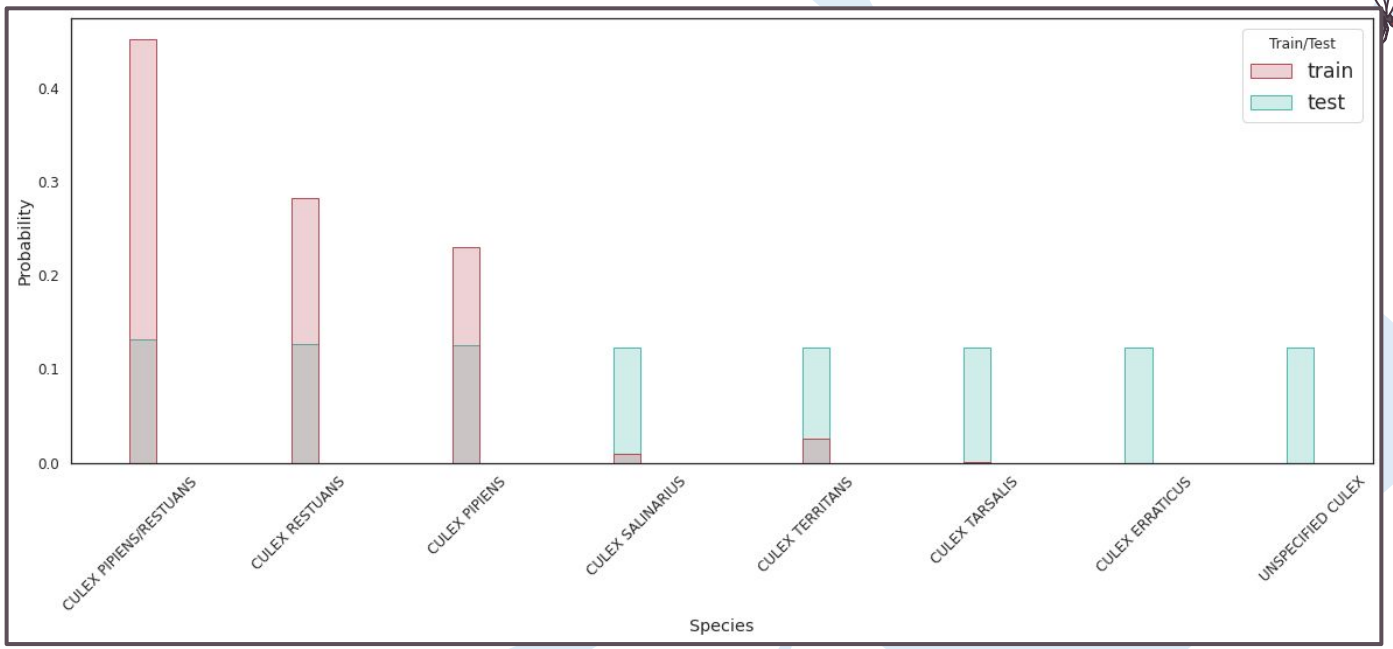
- Training and Validation data sets are split across alternate years
- Spray data was collected for only 2 of 4 years in the training data set

Date Range



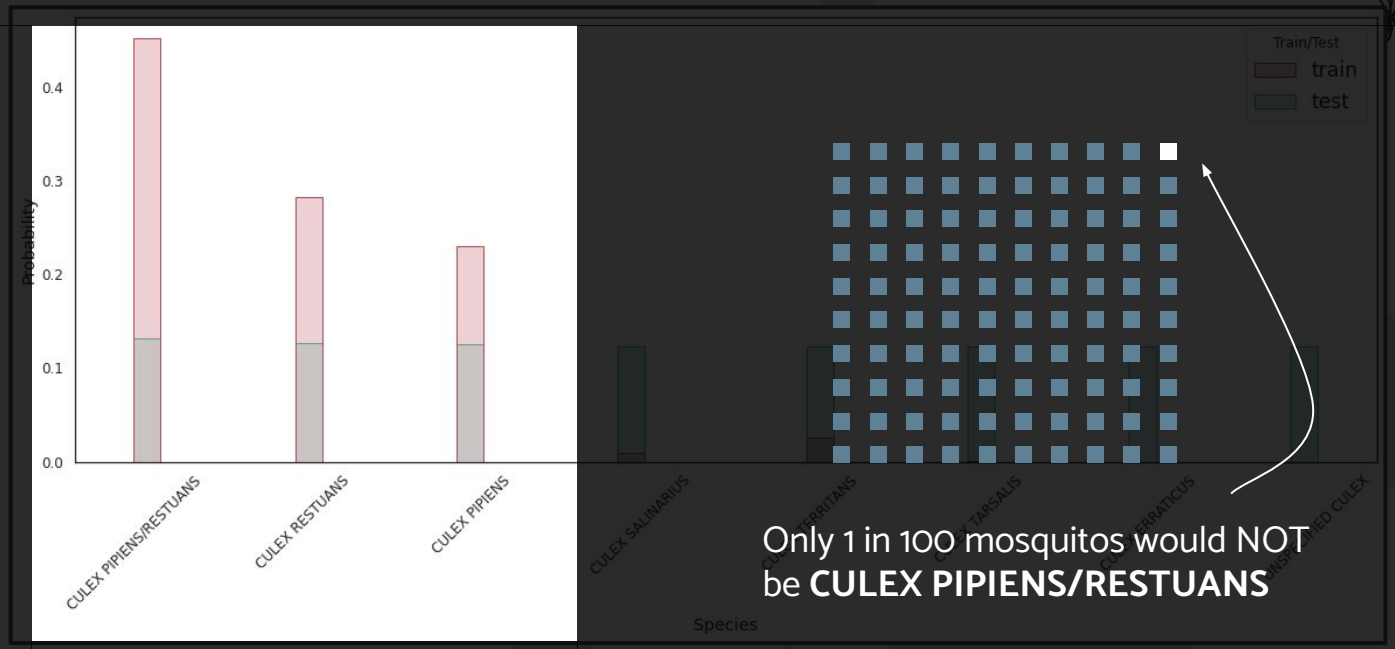
- Training and Validation data sets are split across alternate years
- Spray data was collected for only 2 of 4 years in the training data set

Species Distribution



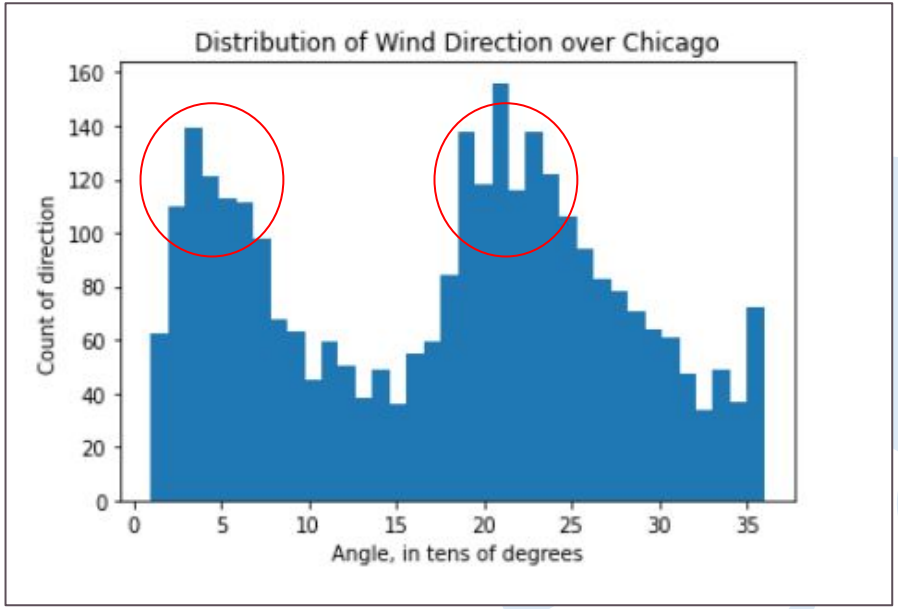
- CULEX PIPIENS & CULEX RESTUANS account for more than 99.5% of the data set

Species Distribution



- CULEX PIPIENS & CULEX RESTUANS account for more than 99.5% of the data set

Wind Direction over Chicago



- Data provided was rounded off to **nearest 10 degrees.**
- Prevailing winds are from predominantly from the **~20° and ~210° direction.**

chicago

Chicago

Illinois
USA

Partly cloudy · 7°C
2:43 AM

Directions | Save | Nearby | Send to your phone | Share

Quick facts

Chicago, on Lake Michigan in Illinois, is among the largest cities in the U.S. Famed for its bold architecture, it has a skyline punctuated by skyscrapers such as the iconic John Hancock Center, 1,451-ft. Willis Tower (formerly the Sears Tower) and the neo-Gothic Tribune Tower. The city is also renowned for its museums, including the Art Institute of Chicago with its noted Impressionist and Post-Impressionist works.

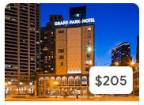
Iconic Chicago

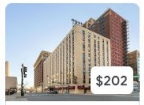
The Art Institute of Chicago
4.8 ★ (26,292)
Renowned art museum with global works

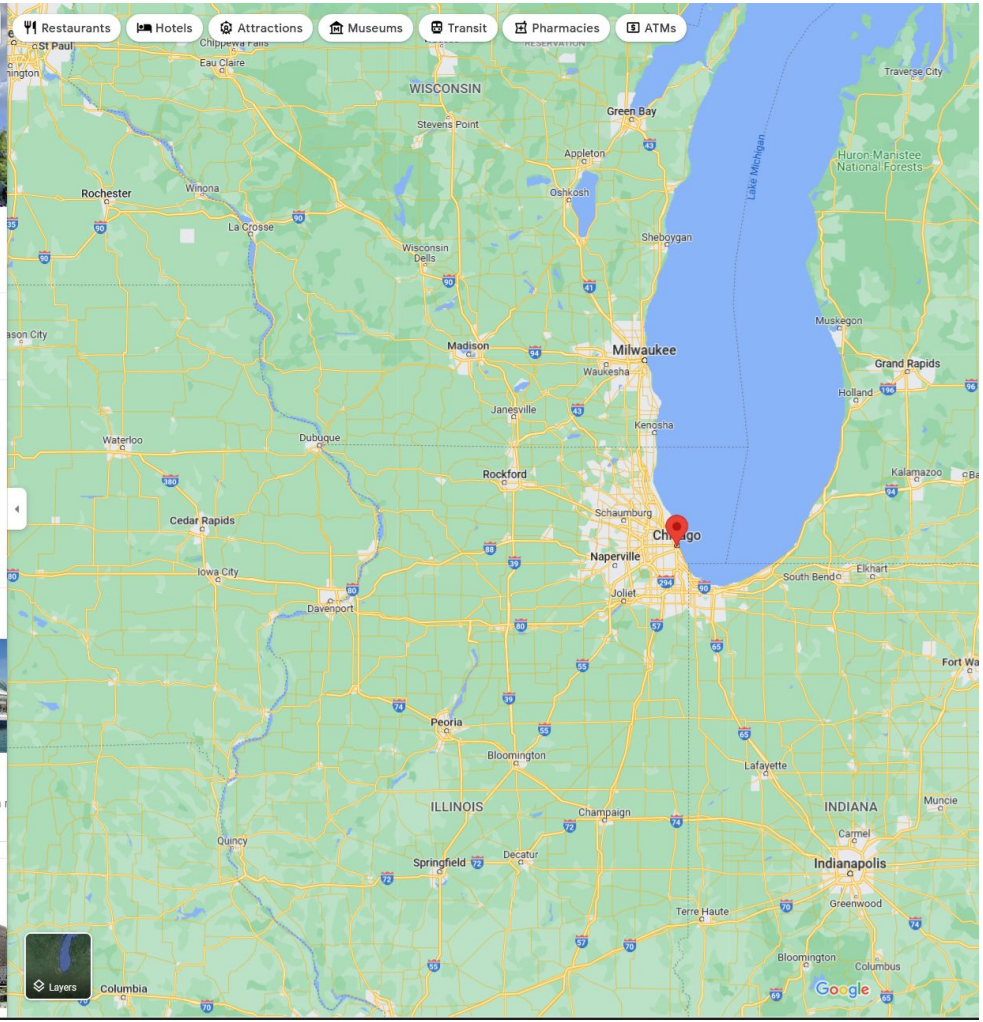
Navy Pier
4.6 ★ (60,360)
Destination with

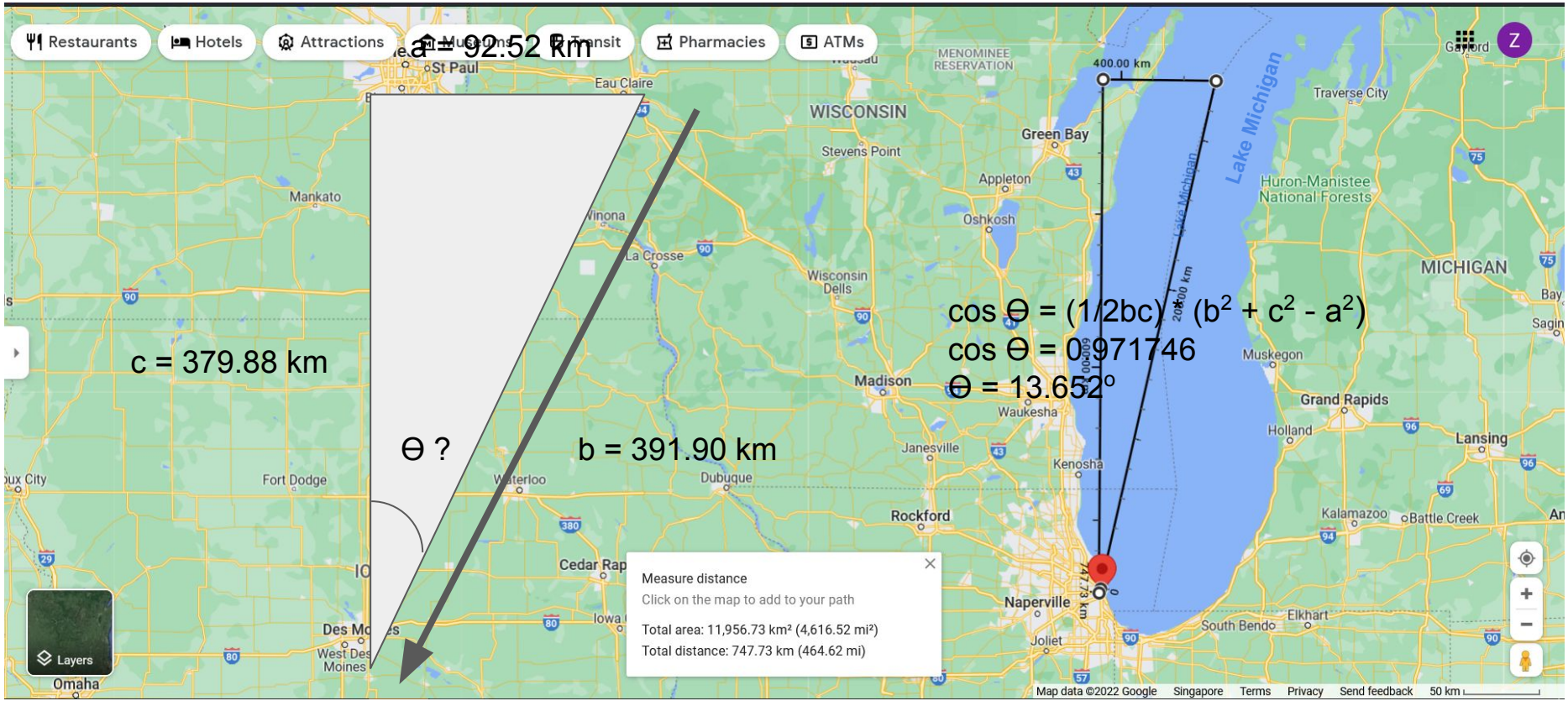
Hotels

About pricing

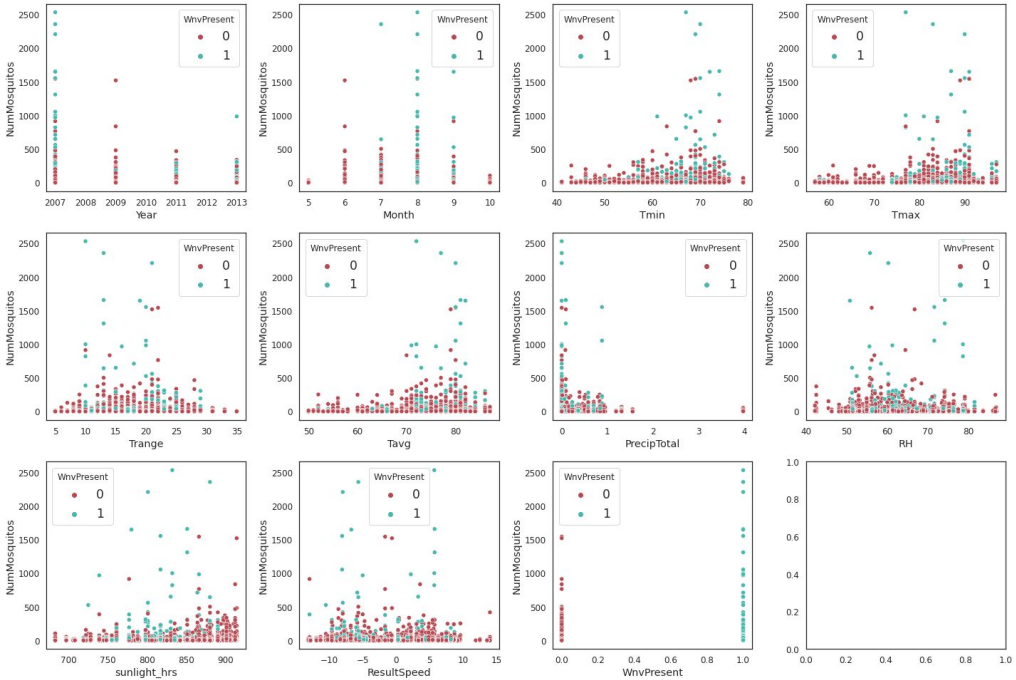


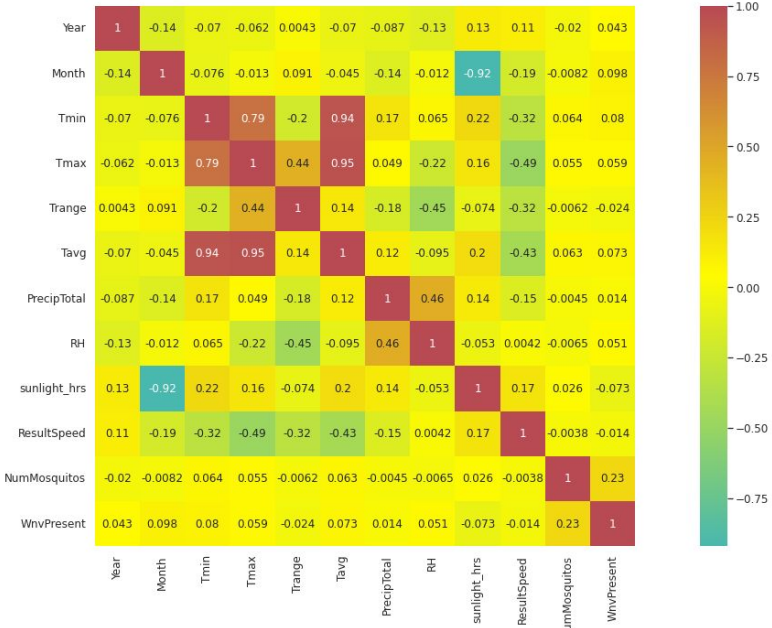





a = 92.52 km
c = 379.88 km
Θ ?
b = 391.90 km
cos Θ = (1/2bc) (b² + c² - a²)
cos Θ = 0.971746
Θ = 13.652°
400.00 km
200.00 km
600.00 km
747.73 km
Measure distance
Click on the map to add to your path
Total area: 11,956.73 km² (4,616.52 mi²)
Total distance: 747.73 km (464.62 mi)
Restaurants
Hotels
Attractions
Transit
Pharmacies
ATMs
Layers
Lake Michigan
WISCONSIN
MICHIGAN
Green Bay
Madison
Milwaukee
Chicago
Rockford
Naperville
Joliet
Kenosha
Waukesha
Janesville
Dubuque
Waterloo
Fort Dodge
Mankato
St Paul
Eau Claire
La Crosse
Wisconsin Dells
Stevens Point
Appleton
Oshkosh
Traverse City
Muskegon
Grand Rapids
Holland
Lansing
Kalamazoo
Battle Creek
Elkhart
South Bend
Omaha
Map data ©2022 Google
Singapore
Terms
Privacy
Send feedback
50 km

No Strong Linear Correlation with Response





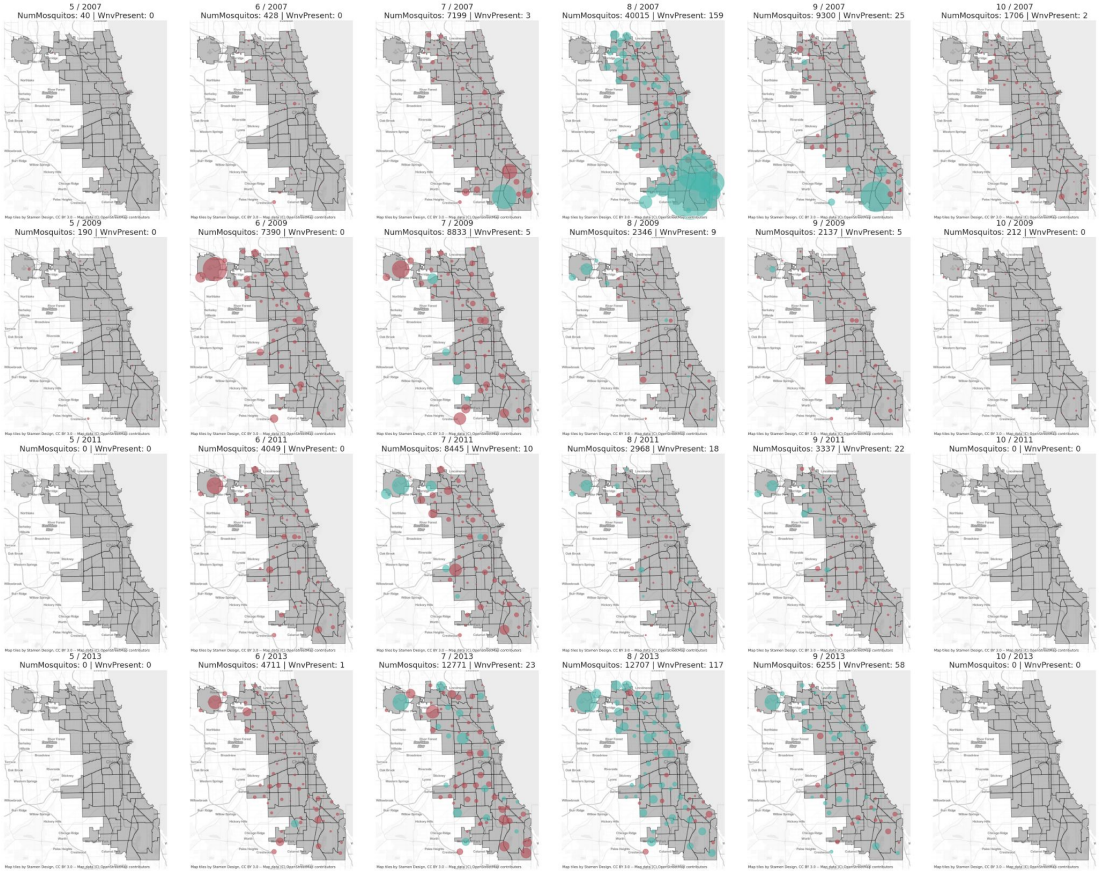
Hotspots

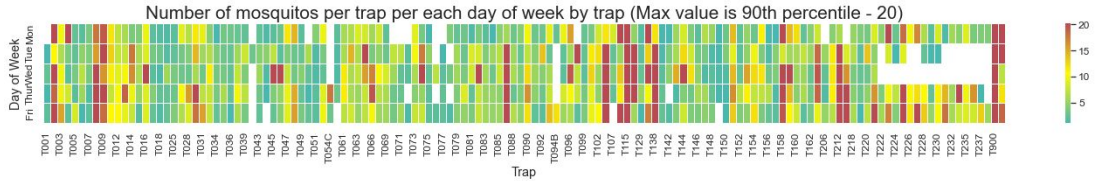
- Hotspots tend to result in rapid growth rate of the number of mosquitoes
- The high number of mosquitoes puts the area at a higher risk of WNV prevalence

Size: Higher NumMosquitos

Wnv Present

No Wnv





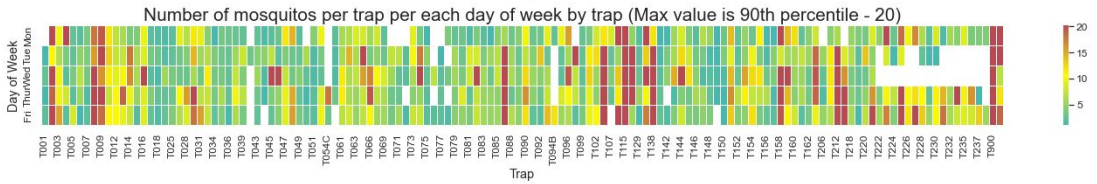
Hotspots

- Hotspots tend to result in rapid growth rate of the number of mosquitoes
- The high number of mosquitoes puts the area at a higher risk of WNV prevalence

Size: Higher NumMosquitos

Wnv Present

No Wnv

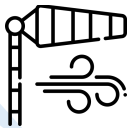


02

What do we know?

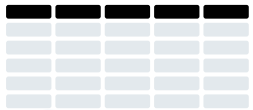
Feature Engineering & Data Modelling

Feature Engineering



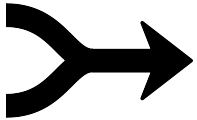
Wind Direction

Resolving wind direction towards **Lake Michigan**



One-Hot-Encode

One hot encoding of categorical features



Species Aggregation

Aggregating based on ***Culex Pipen/Restuans***

Lag Weather

Lagging weather observations by **5 to 14** days

Distance Matrix

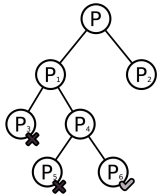
Inverse weighted distance from neighbouring traps

Modelling Approach

Data Preprocessing

Regression Modelling

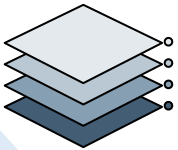
To obtain
NumMosquitos



Classification Modelling

To determine
Wnv Presence

Preprocessing



Species Balancing

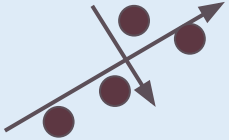
Add rows for unfound species

Oversampling

Oversampling on minority response class using SMOTE

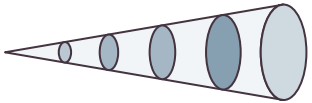
Cross validation

5-fold Cross-validation on training set



PCA

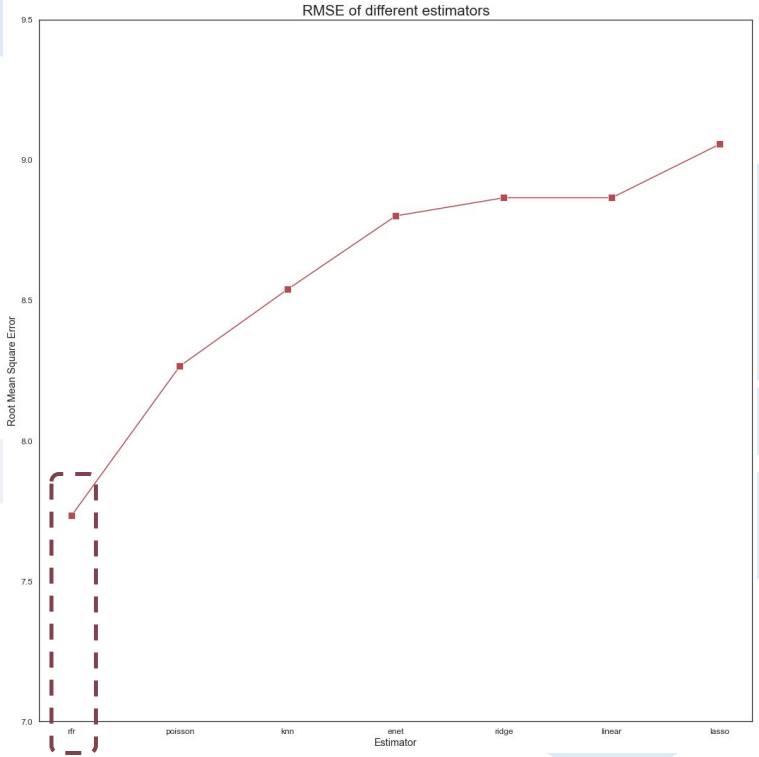
Conduct PCA for linear models

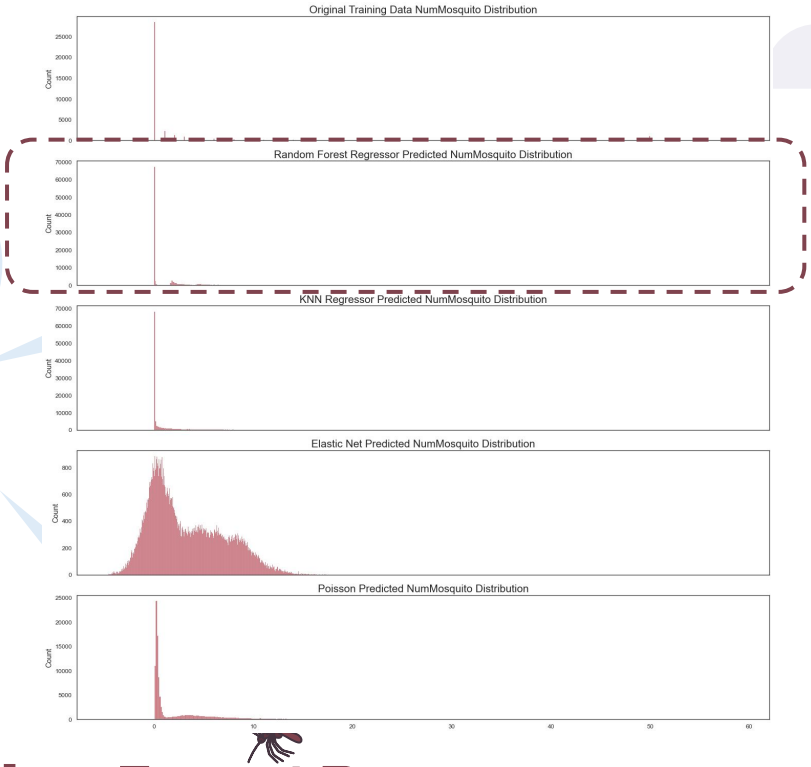


Scaling

Either Standard Scaling or MinMax Scaling

Model Performance (Regression)





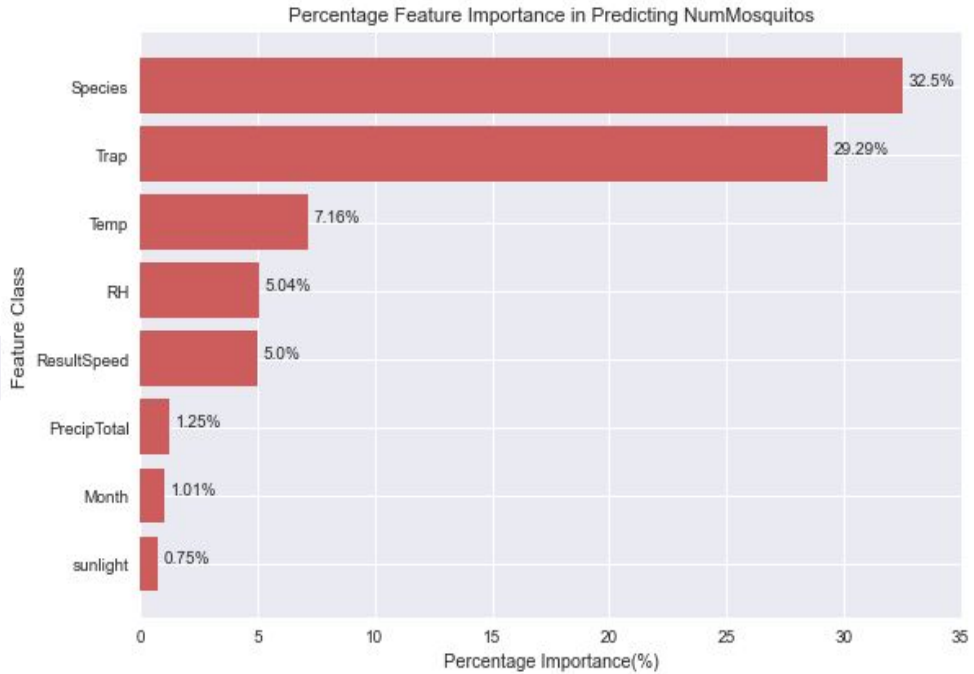
Chosen Model: Random Forest Regressor

Model Performance (Classification)

Cross Validated Scores

Model	ROC-AUC	Accuracy	Recall	Precision
Gradient Boosting	0.93	0.93	0.55	0.13
Random Forest	0.91	0.93	0.70	0.14
Logistics Regression	0.77	0.88	0.52	0.07

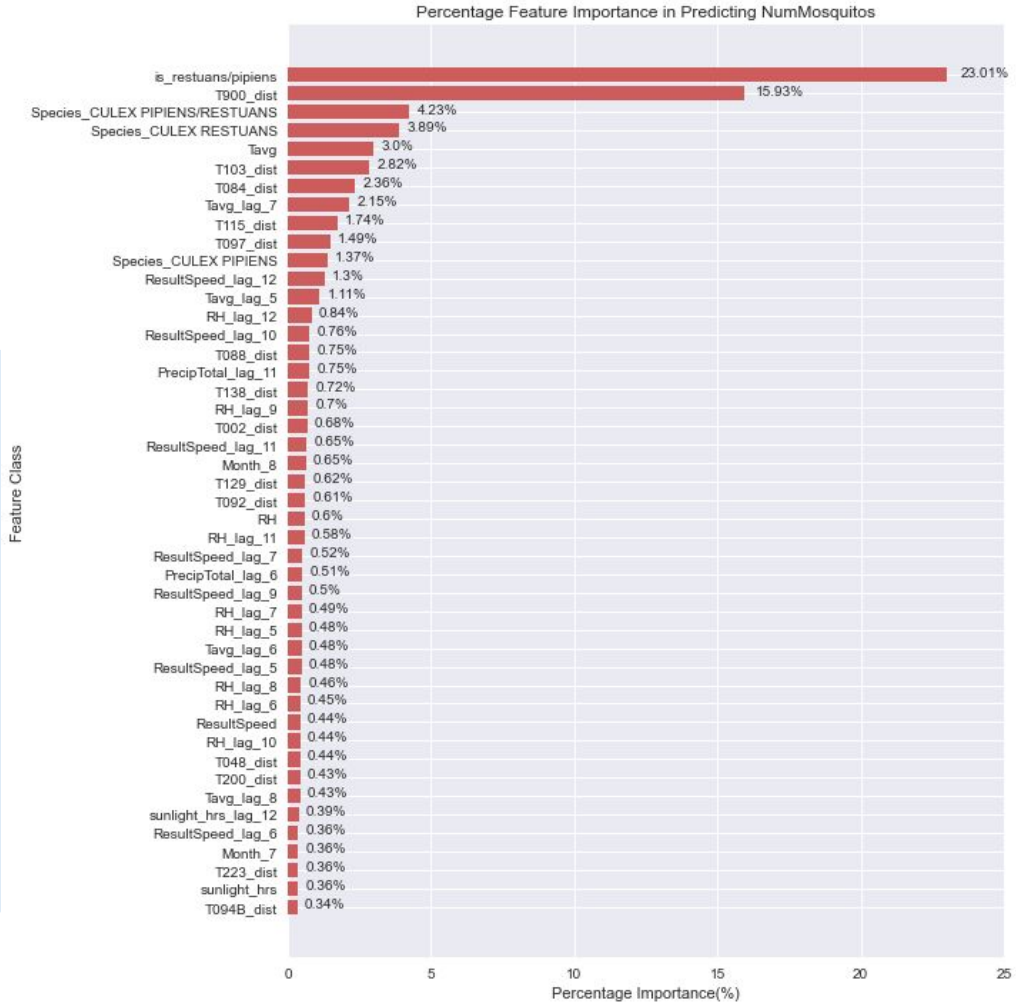
Feature Importance: NumMosquitos



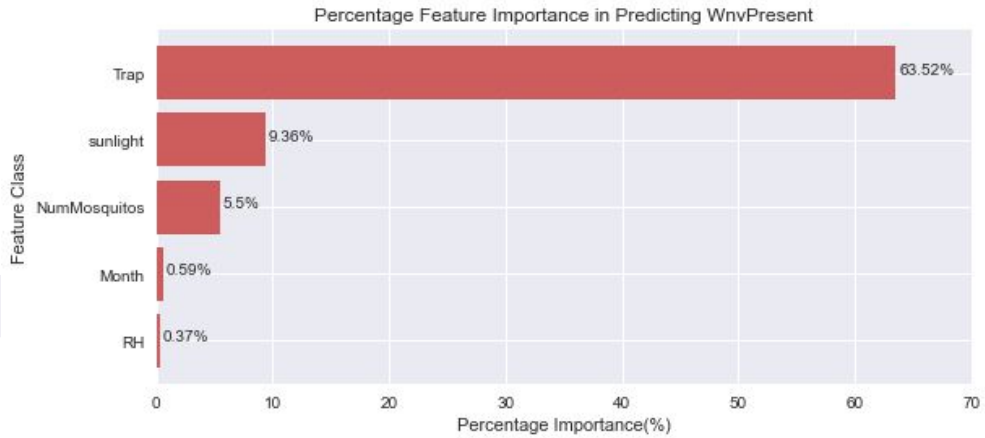
- Species type (32.5%)
  - Pipiens / Restuans
- Trap location (29.3%)
- Weather Data (18.5%)
  - Temperature
  - Humidity
  - Wind Speed/Direction
  - Total Rainfall

- Three categories sum to >80%

Feature Importance: NumMosquitos

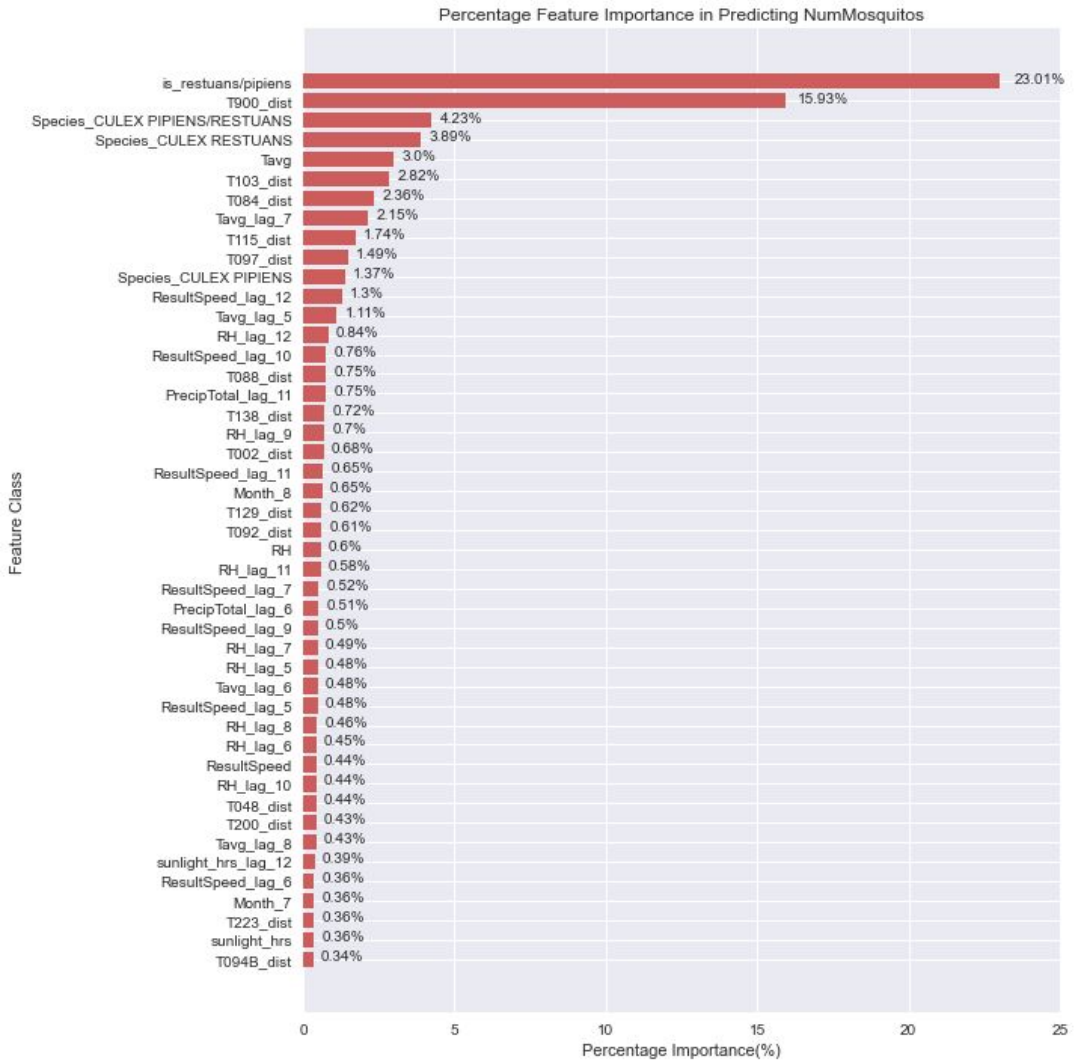


Feature Importance: WnvPresent



- Trap location (63.52%)
- Sunlight hours (9.36%)
- NumMosquitos (5.5%)
- Month Number (0.59%)

Feature Importance: WnvPresent



03

How much will it cost?

Cost Benefit Analysis

Primary Cost Drivers

Spraying

Cost associated with
conducting vector controls

Healthcare & Productivity

Costs associated with
medical coverage & productivity losses

Spray Cost Model

Spray Costs

Cost associated with conducting vector controls

f	**Frequency**	Number of sprays in a given period
A	**Area**	Area covered
UC_s	**Unit Cost**	Cost per spray

$$C_s = f \times A \times UC_s$$

Healthcare Cost Model

N_M	**NumMos**	Number of mosquitoes trapped
T	**%Trapped**	Percentage of mosquitoes trapped
N_B	**NumBites**	Number of human bites, per mosquito
V	**WnvPresent**	Binary toggle; Indicates if WNV is detected in mosquitoes trapped
UC_H	**Unit Cost**	Averaged healthcare and productivity costs, per person

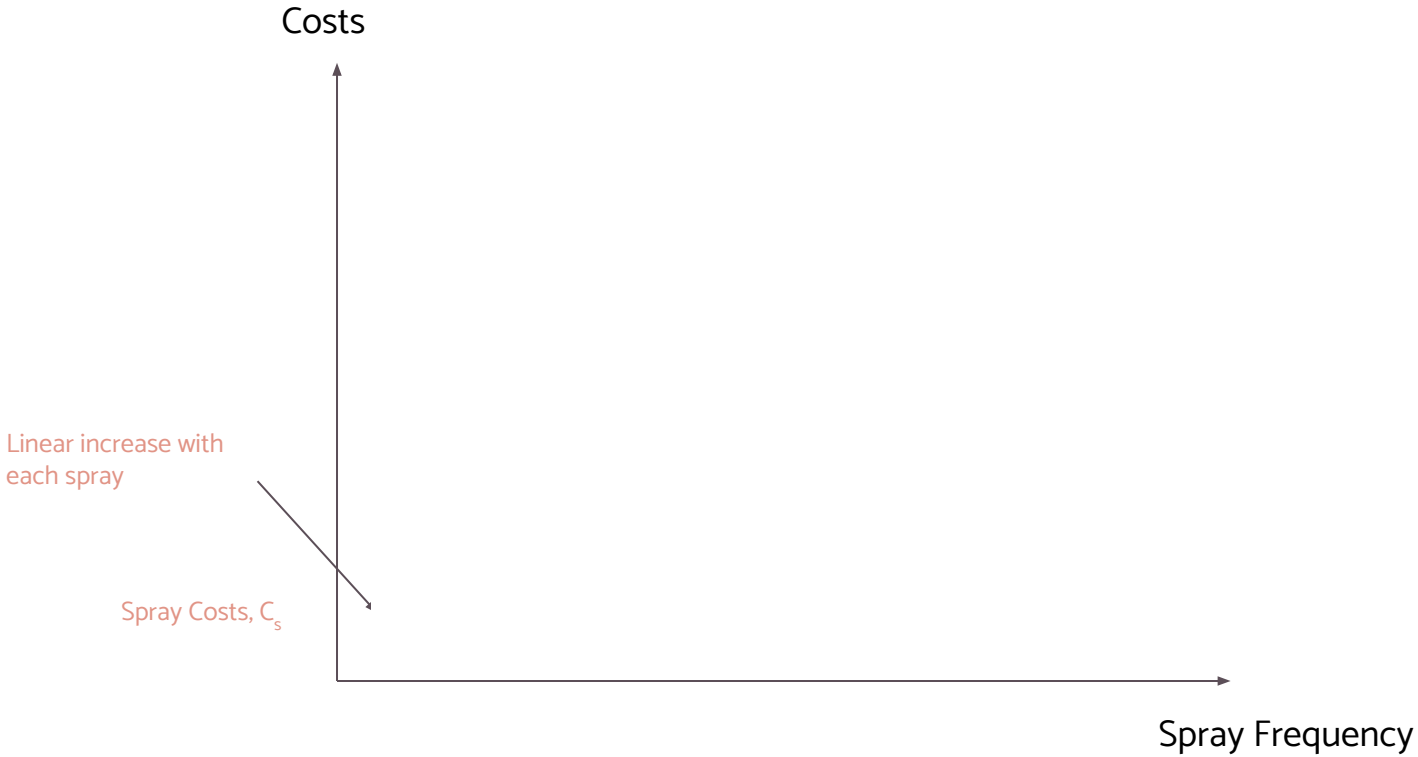
Healthcare & Productivity Costs

Costs associated with medical coverage & productivity losses

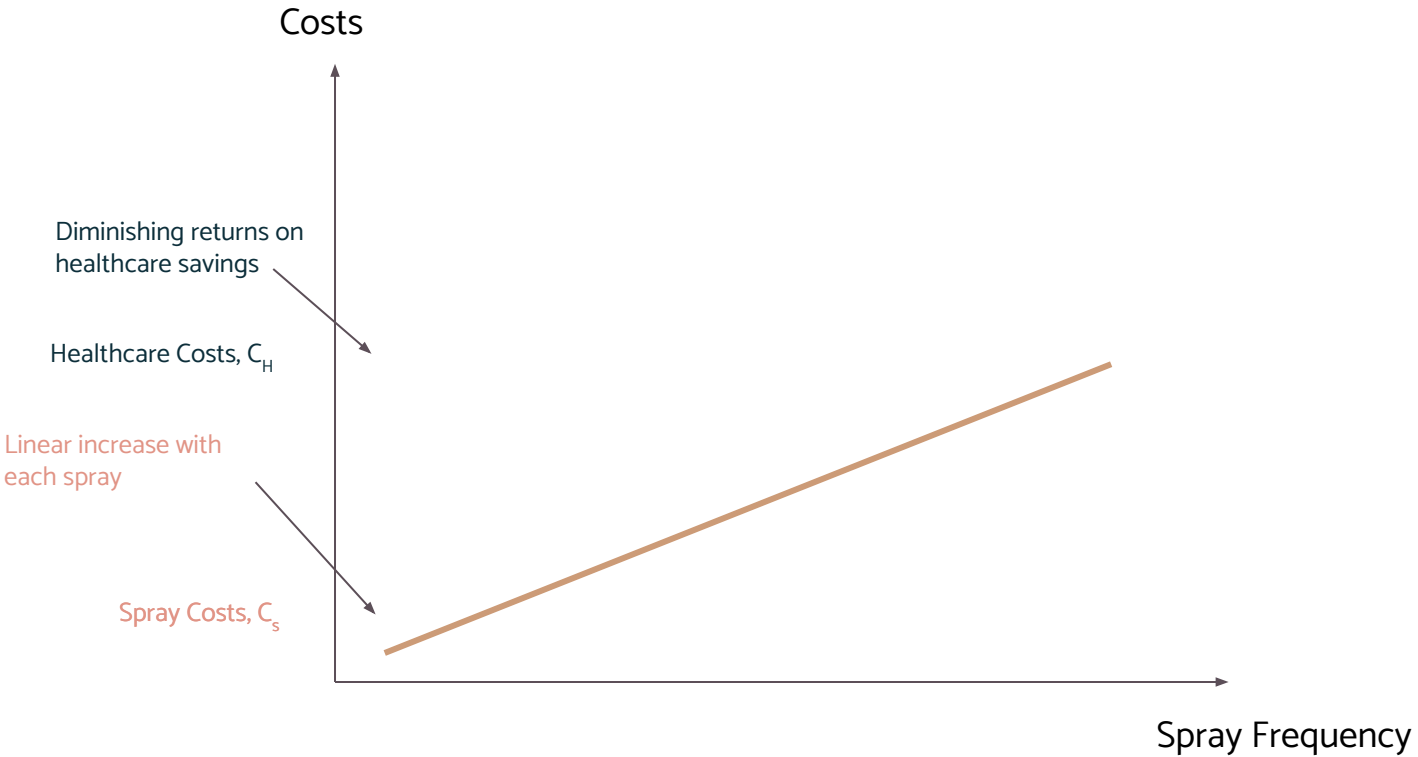
$$C_H = (N_M \times T) \times (N_B \times V) \times UC_H$$

$$C_H = \text{Mosquito Population} \times \text{Human Infection Rate} \times \text{Per Head Cost}$$

Cost Optimization

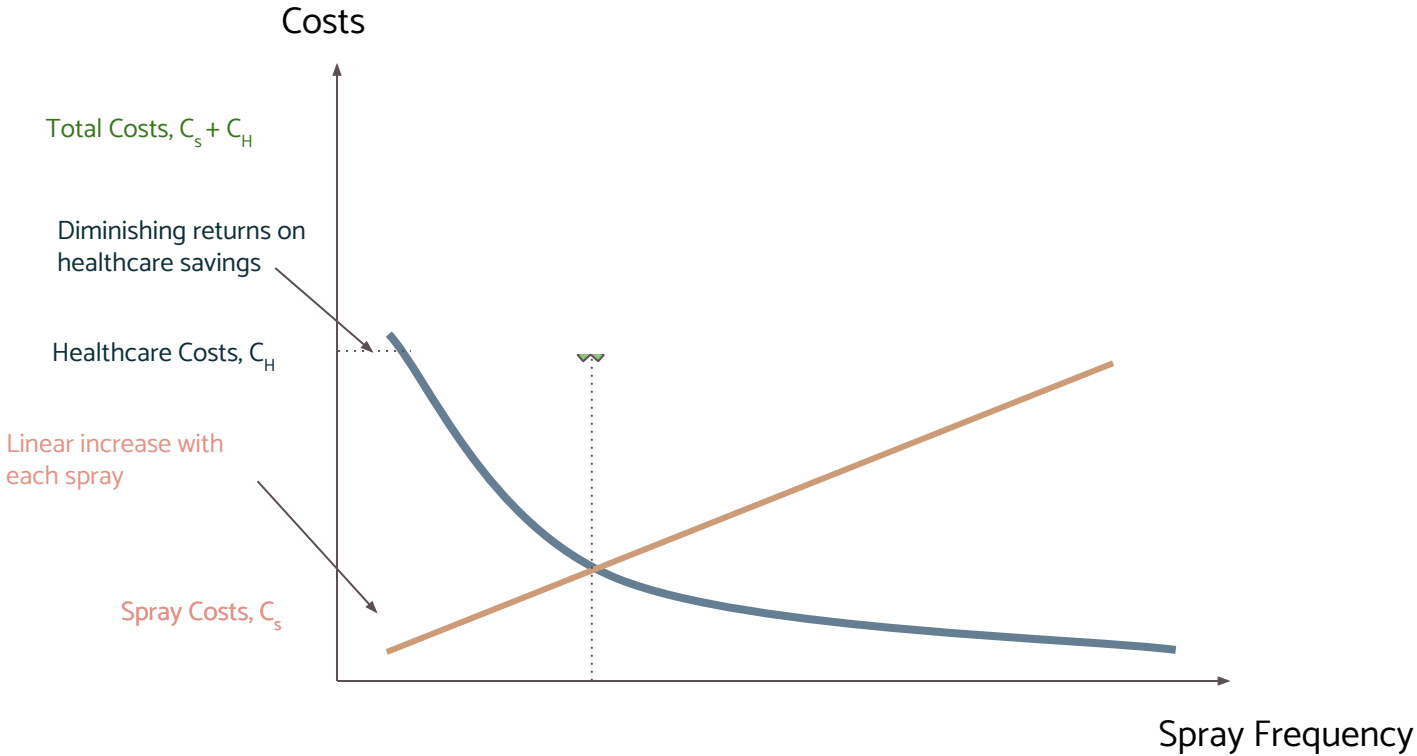


Cost Optimization

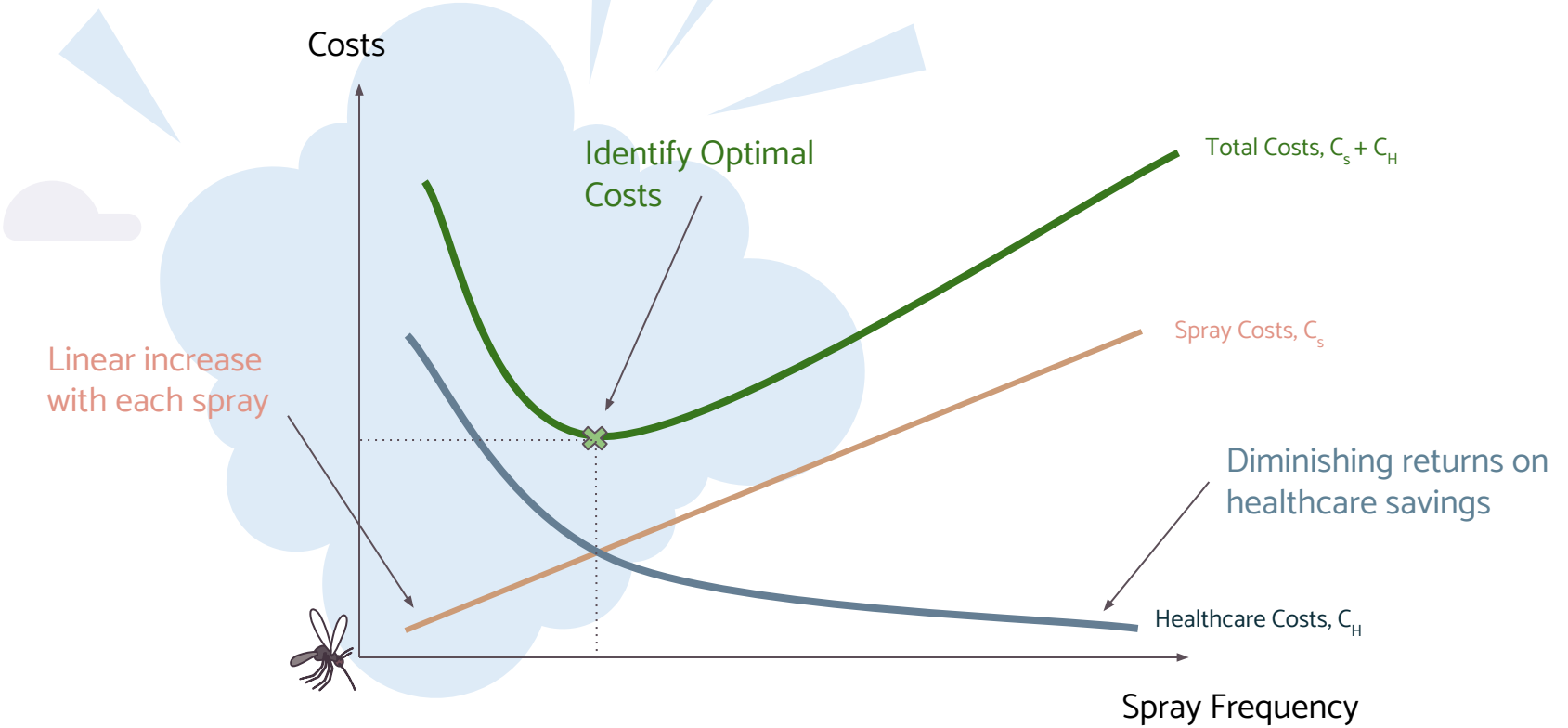


Cost Optimization

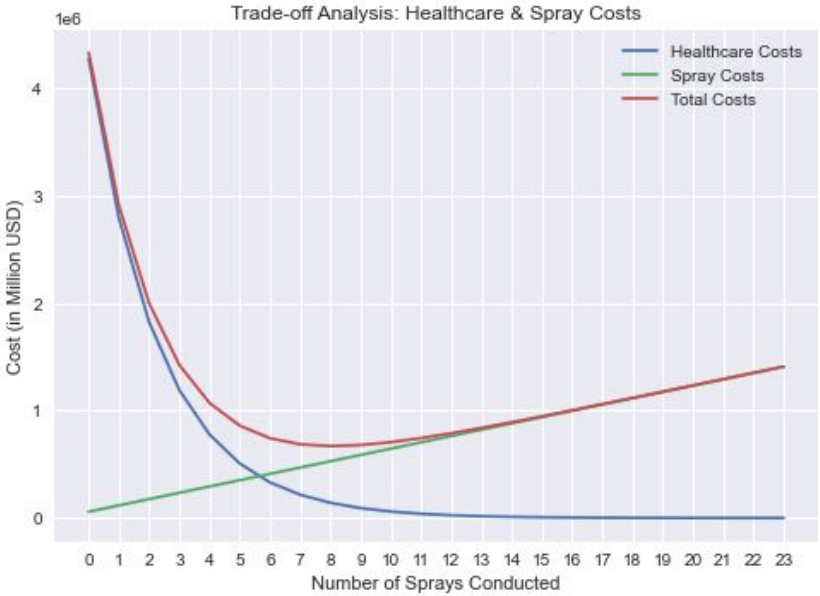
Identify Optimal Costs



Cost Optimization



Simulation Results

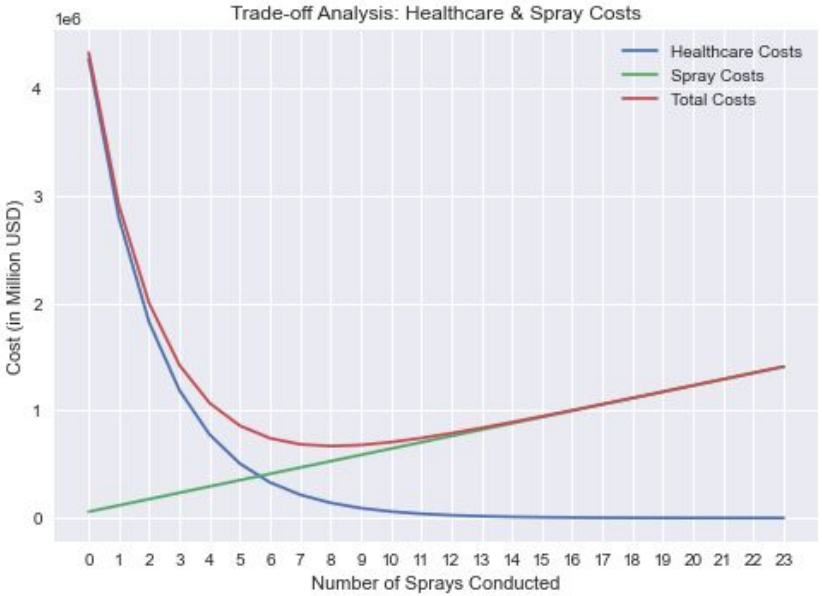


Min Total Costs	$669,321
Optimal Spray Freq	9 sprays

Healthcare Costs: $6,317 USD per person
%Population Infected: 0.1% (1,000 per million)

Spray Costs: $0.92 / acre
%Infected Reduction: 65.3%

Simulation Results



Healthcare Costs: $6,317 USD per person
%Population Infected: 0.1% (1,000 per million)

Spray Costs: $0.92 / acre
%Infected Reduction: 65.3%

04

What should we do?

Conclusions & Recommendations

Conclusions

High Risk in July-Sept

Consistently High Wnv Occurrence in these months

Hotspots Identified

Traps were ranked based on relative importance

Recommendations

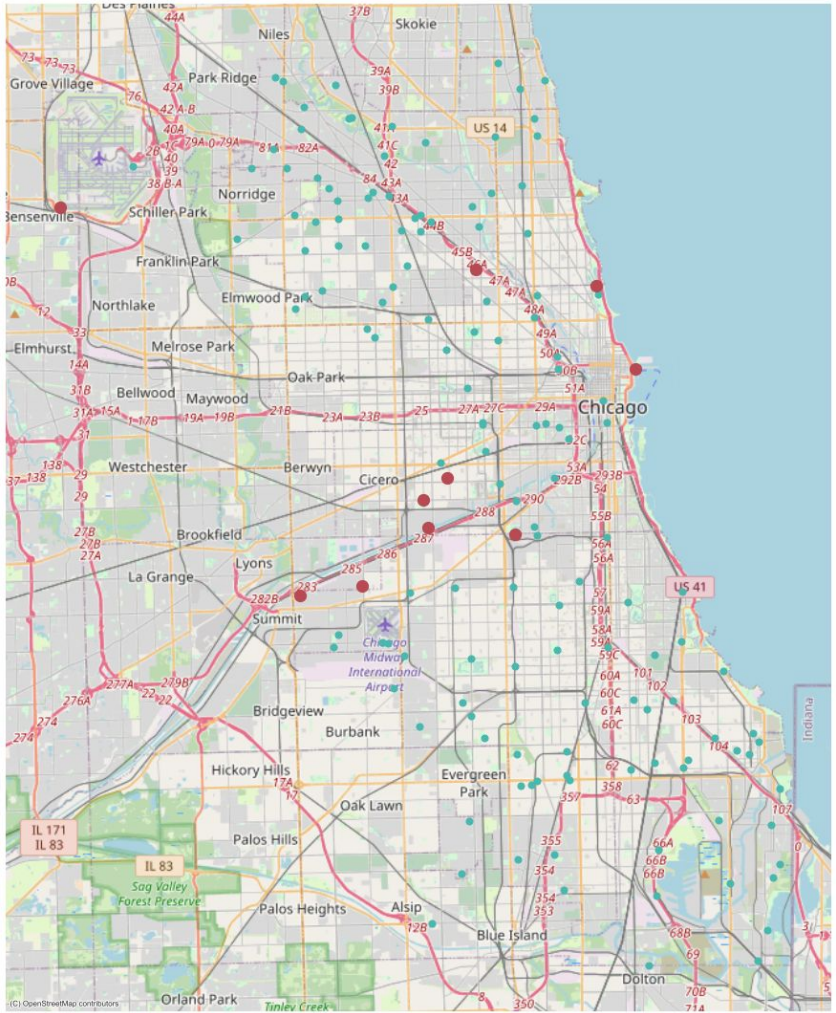
Preventive Sprays

Aggressive spraying in earlier months

Increased Surveillance

Increasing surveillance near hotspots

Location of Critical Monitoring Traps



Thanks

Do you have any questions?

Future Improvements

- From the plots, we visually see that the spray areas are concentrated in distinct clusters
- Incorporate Local Population as inputs
  - Has a key role in outbreak mechanisms
- Incorporate Bird Population & Migratory Paths
  - Leverage robust population modelling techniques

Logistic growth equation:

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right)$$

Competitive Lotka–Volterra equations:

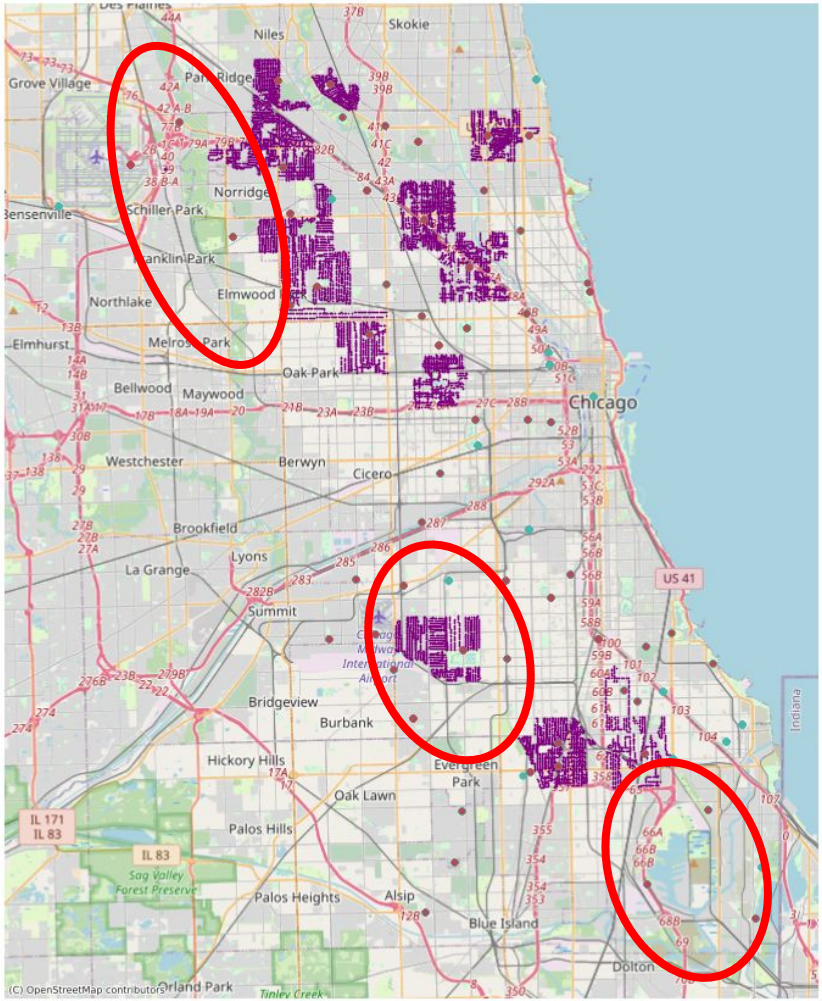
$$\frac{dN_1}{dt} = r_1 N_1 \frac{K_1 - N_1 - \alpha N_2}{K_1}$$

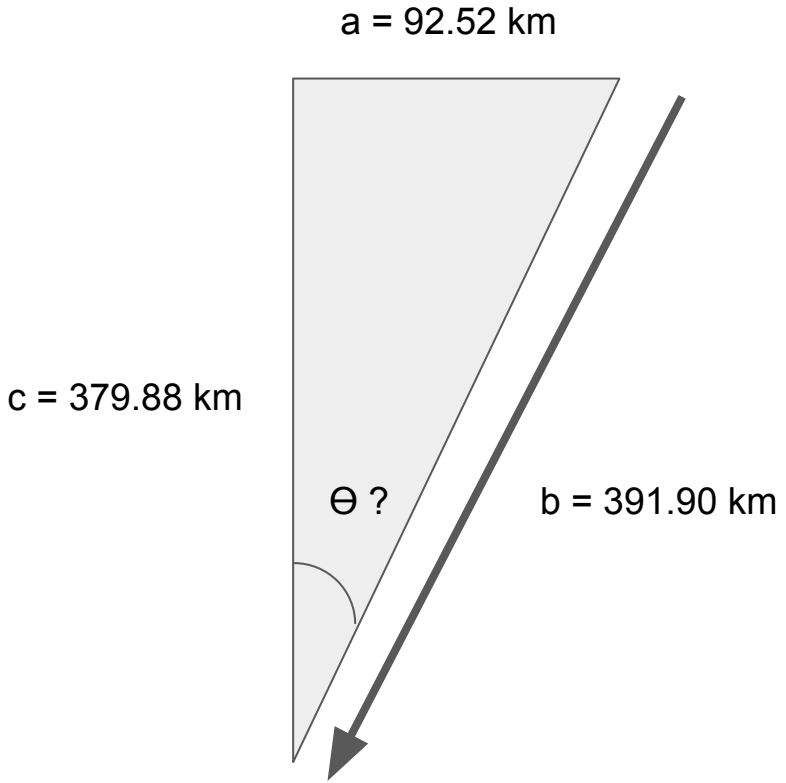
Island biogeography:

$$S = \frac{IP}{I + E}$$

Species–area relationship:

$$\log(S) = \log(c) + z\log(A)$$





$\cos \Theta = (1/2bc) * (b^2 + c^2 - a^2)$
$\cos \Theta = 0.971746$
$\Theta = 13.652^o$

Fonts & colors used

This presentation has been made using the following fonts:

Chivo
(https://fonts.google.com/specimen/Chivo)

Catamaran
(https://fonts.google.com/specimen/Catamaran)

#443440 #5c4f58 #5d3742 #8c8184 #76515b #7f434e

#b84a54 #cfe2f3 #ffffff