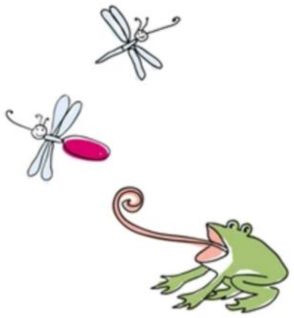
The West Nile Virus Prediction Model

Edward, Sook-Yee



Background

Our client Chicago Department of Public Health has requested us to provide suggestions of **when** and **where** to perform aerial spray to control mosquito population.

This will help **decrease the city's** west nile virus (WNV) **infection and death rates**.

Aim

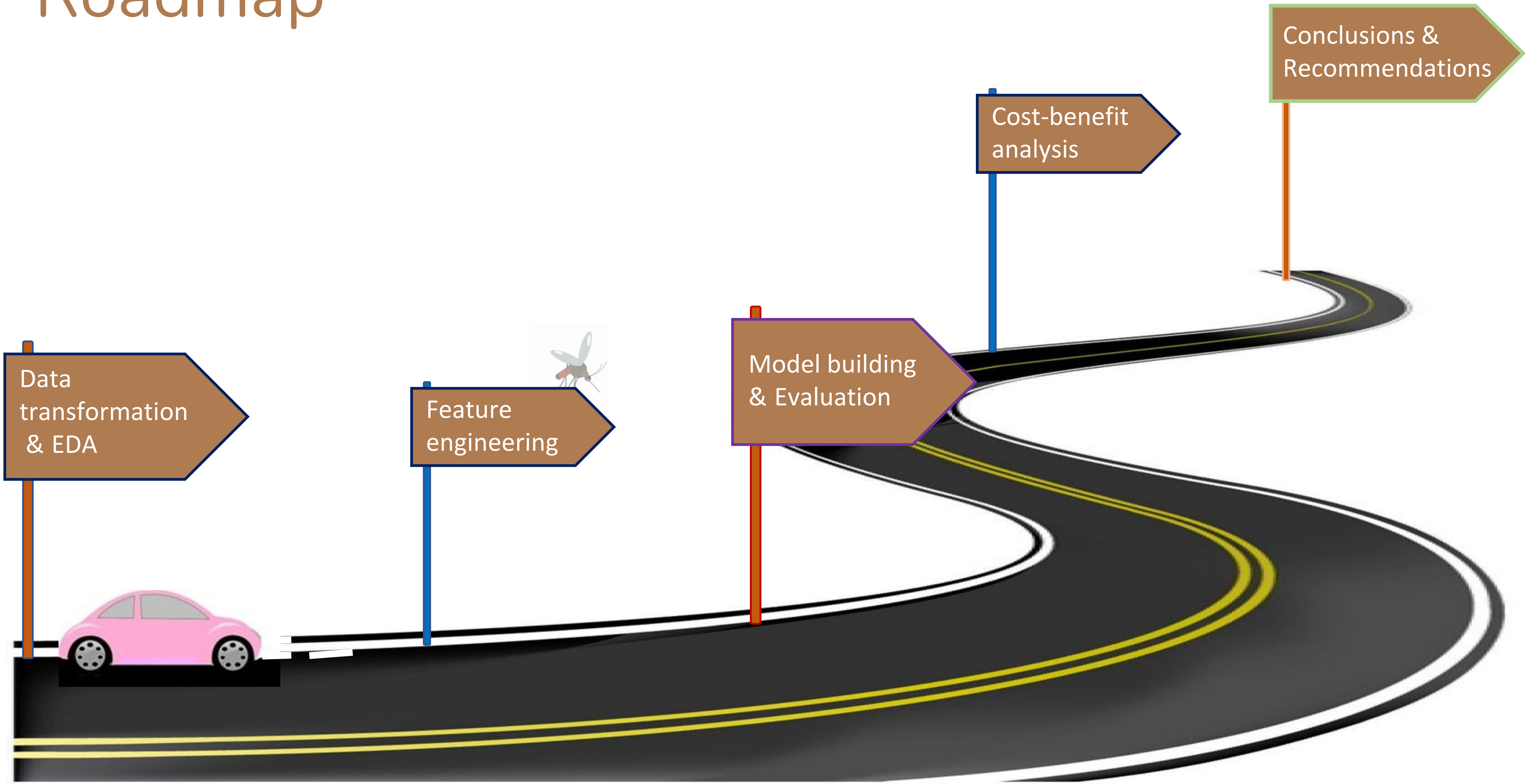
Provide recommendations and conduct a cost-benefit analysis to control mosquito population, and reduce the number of west nile virus (WNV) cases in Chicago.

Approach:

Determine the important features for predicting presence of wnv.

Focus on **accurate predictions**, and **maximizing sensitivity**.

Roadmap


Data transformation & EDA
Feature engineering
Model building & Evaluation
Cost-benefit analysis
Conclusions & Recommendations

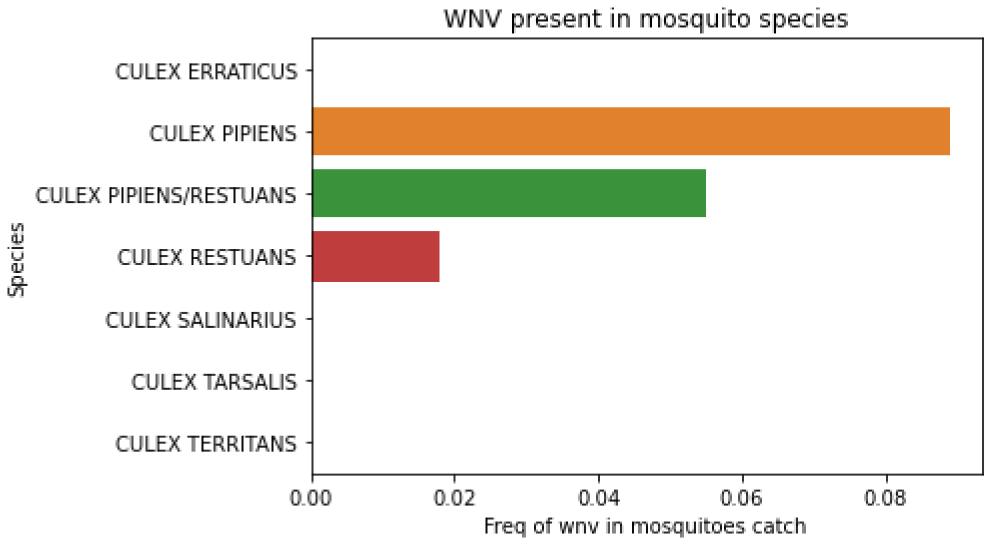
Surveillance data

	YEAR							
	2007	2008	2009	2010	2011	2012	2013	2014
Train	X		X		X		X	
Test		X		X		X		X
Weather	X	X	X	X	X	X	X	X
Spray					X		X	

Ordinalise mosquito species and traps

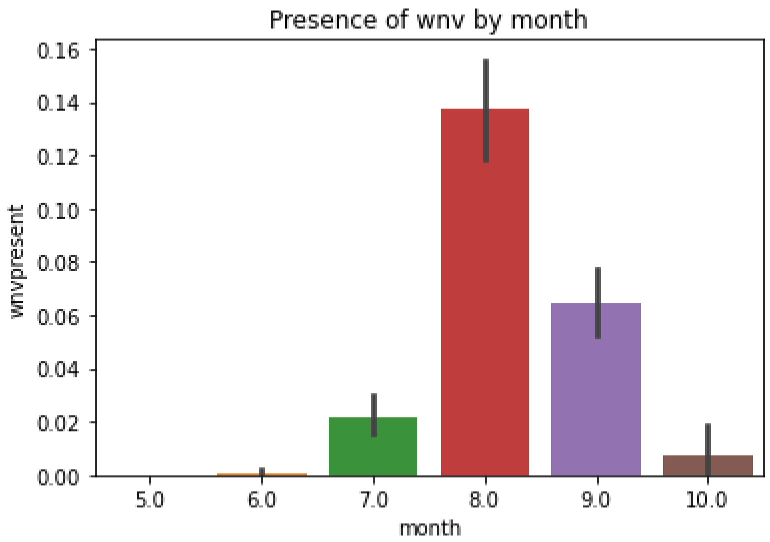
Culex Pipiens/Restuans species is the primary transmission vector of west nile virus (WNV).

Rank traps based on how they are likely to trap mosquitoes with WNV

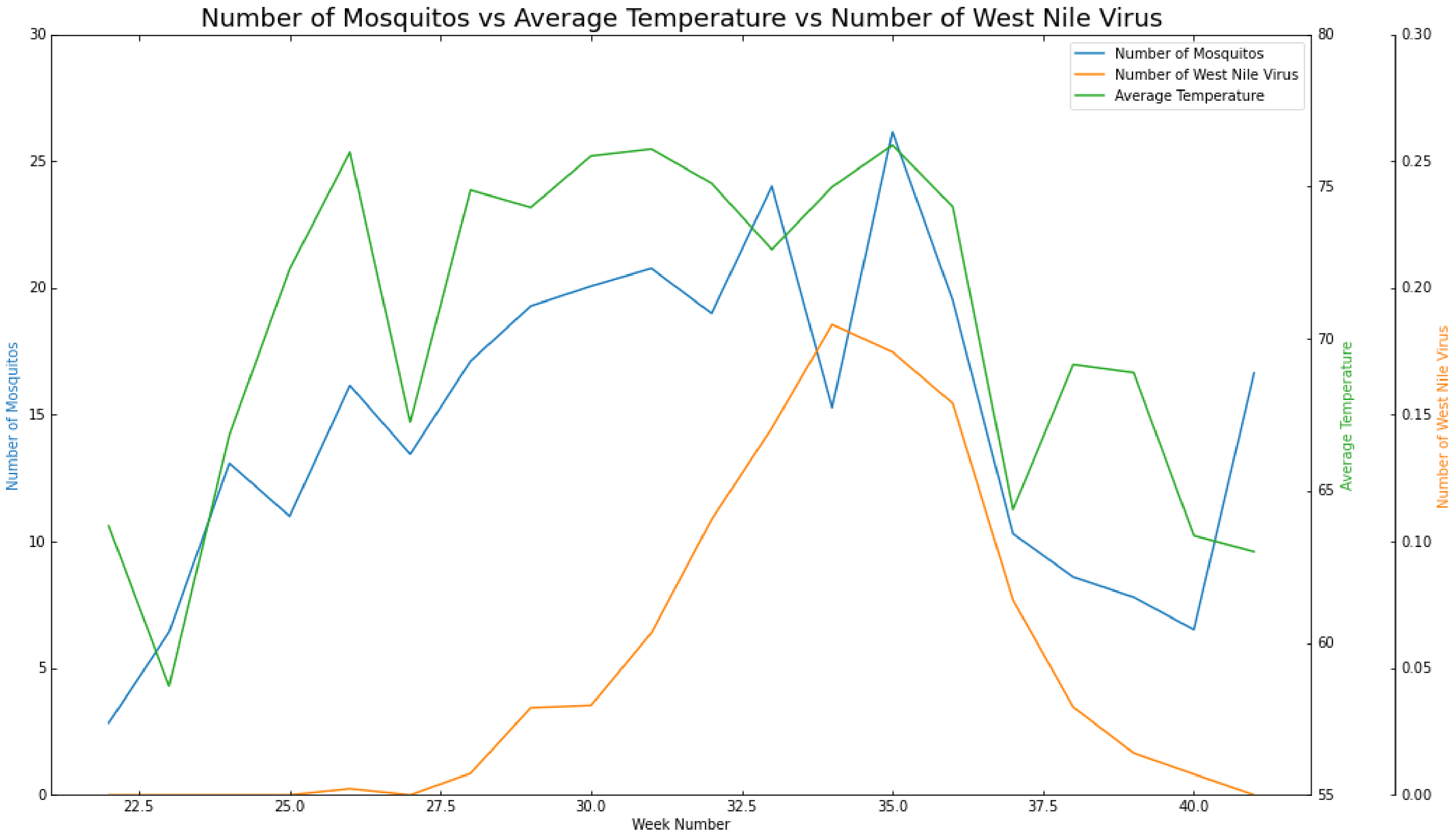


Seasonal differences

Number of mosquitoes trapped and WNV detection was highest in the month of August where the weather is hot and humid.



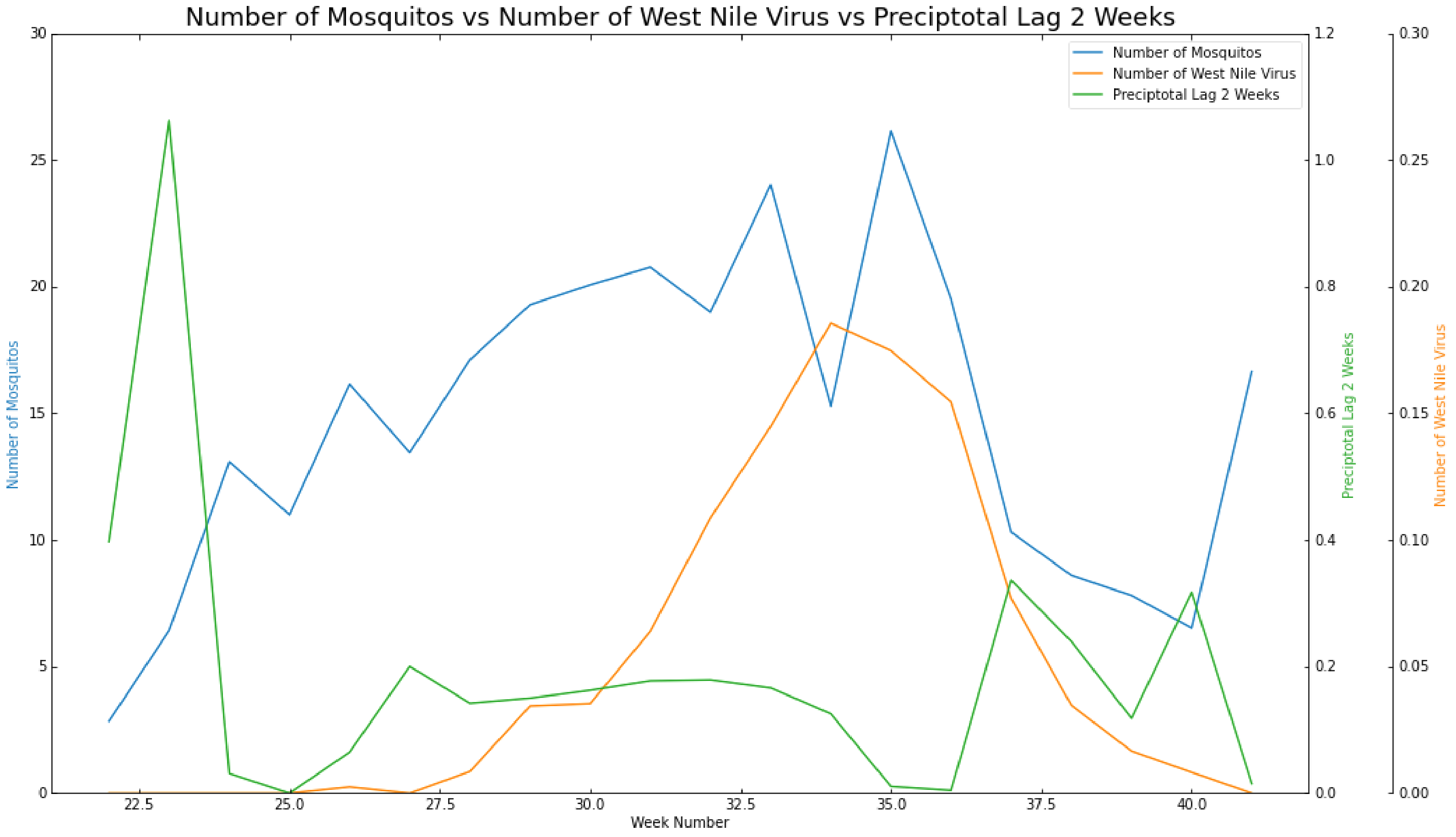
Effect of Temperature



$\uparrow$ Temperature $\Rightarrow$ $\uparrow$ Mosquitoes

$\uparrow$ Mosquitoes $\Rightarrow$ $\uparrow$ WNV Mosquitoes

Effect of Lagged Precipitation



↓ Precipitation ⇒ ↑ Mosquitoes

↑ Mosquitoes ⇒ ↑ WNV Mosquitoes

Model

Feature reduction

1. Features selected based on correlation

2. Removed features that are collinear
   Tavg = (Tmin and Tmax)/2

3. Reduction of feature based on human-based calendar grouping
   Date/Month/Week = Mosquitoes do not abide by human calendars

4. Removed target y
   Nummosquitos/wnvpresent = Data leakage

Feature extraction

- Reduce dimensionality of the data set

- PCA-transform

Metrics Evaluation

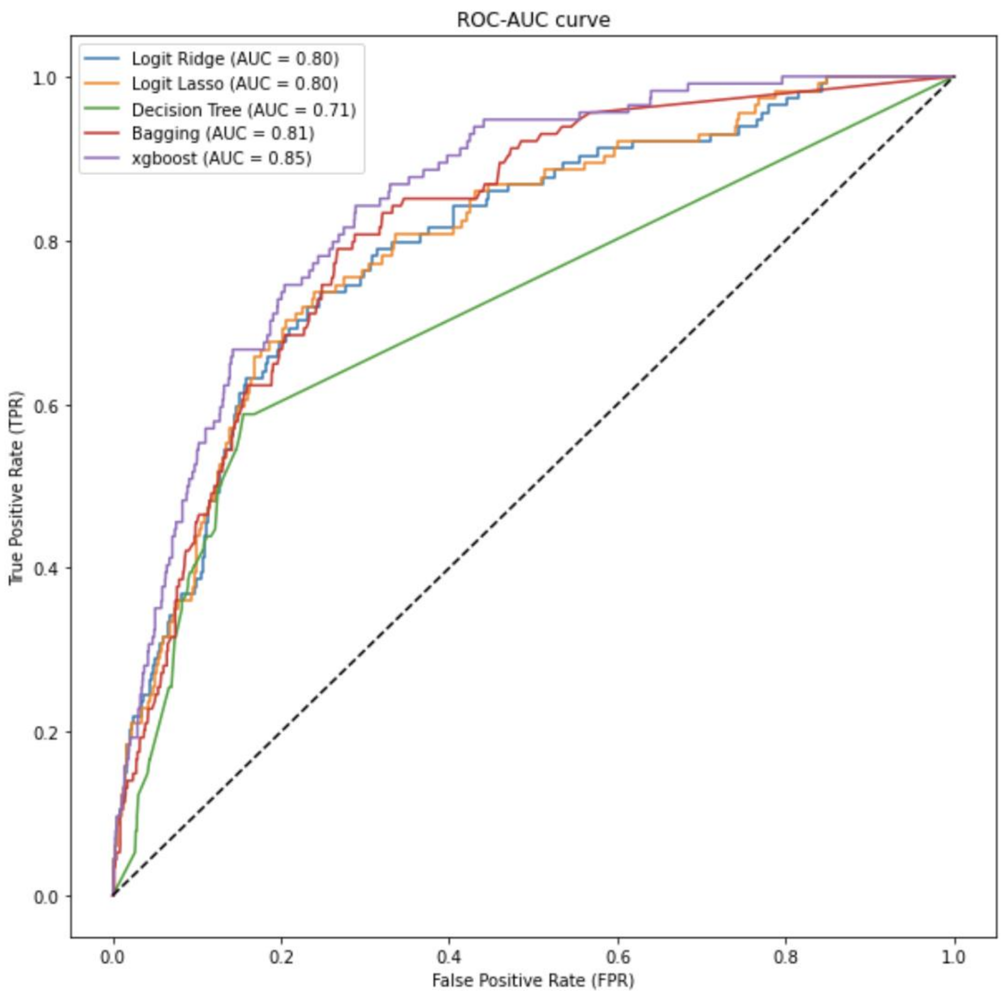
Metrics:

1. Sensitivity
2. ROC_AUC

Predicting mosquitoes not to have WNV, when they actually do (false negatives) is crucial.

Model	Accuracy	Specificity	Sensitivity	ROC_AUC
SS, Smote, Logit-Ridge	72.5	72.5	73.7	0.731
SS, Smote, Logit-Lasso	73.1	73.0	74.6	0.738
SS, Smote, DTree	87.6	90.3	40.4	0.653
SS, Smote, Bag	86.6	88.9	46.5	0.677
SS, Smote, XGBoost	88.9	91.2	49.1	0.702

ROC Curve Evaluation



Highest AUC value: XGBoost

Pros:

- Less overfitting due to training shallow trees
- Greater predictive value from training weaker models into stronger models

Cons:

- Computational time
- Sequential vs Parallel processing

Kaggle Score / Feature importance

Model	Sensitivity	ROC_AUC	AUC	Kaggle Score
Logistic-Lasso	0.746	0.738	0.80	0.72469
XGBoost	0.491	0.702	0.85	0.75119

Top 5 Logistic Features

Later sunrise: Higher chance of WNV mosquito present

	feature_importance
sunrise	13.020677
temp_diff	5.511193
tavg	5.511193
stnpressure	3.191687
TS	1.981297

Top 5 XGBoost Features:

Sunset/Sunrise are the most important determinants

	feature_importance
sunset	0.265326
sunrise	0.162277
species_nr	0.085531
addressaccuracy	0.050645
year	0.048792

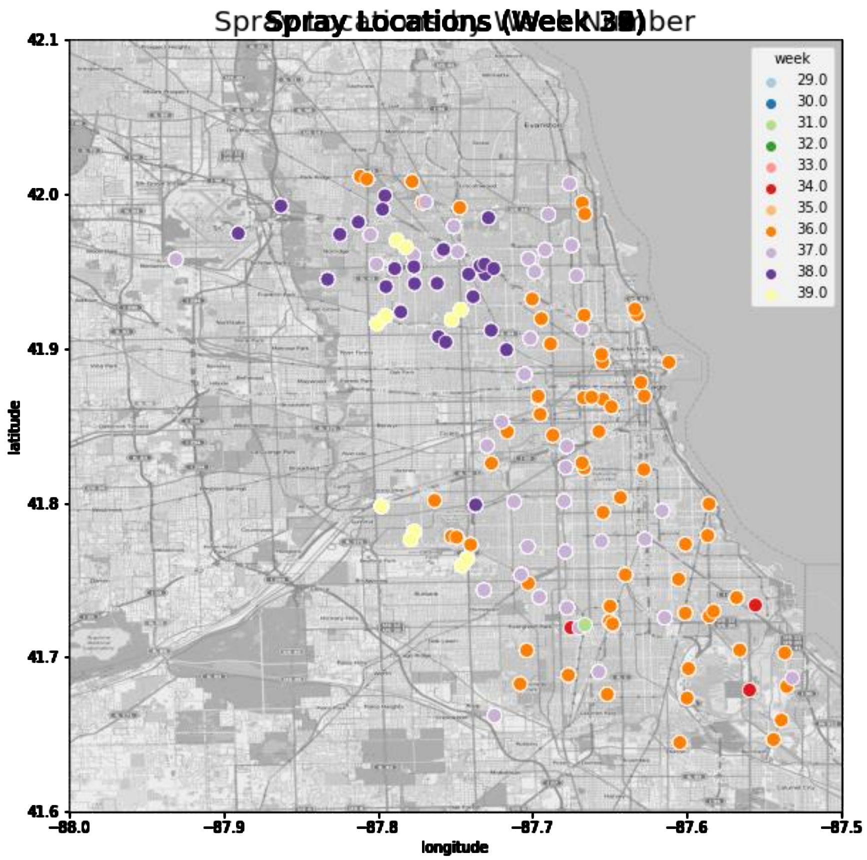
Bottom 5 Logistic Features

Higher precipitation: Less chance of WNV mosquito present

	feature_importance
preciptotal	0.665960
longitude	0.570535
dewpoint	0.414168
sealevel	0.362066
depart	0.072401

Model Predictions

Spray locations by week



Cost-benefit analysis

	Economic costs of WNV outbreak/ US$/ per person	
	Louisiana (Zohrabian et al., 2004)	Sacramento County (Barber et al., 2010).
Inpatient costs	17,680	33,143
Outpatient costs	392 – 7,391	333 - 4,037
Productivity loss/ day	225	125 - 191
Total	15,297 – 22,296	33,601 – 37,371

	Spraying costs	
	Sacramento County	Chicago
Sprayed area/km^2	477	606
Costs/ US$	641,790	815,355
Costs/km^2	1,345	1,345

Cost-benefit analysis

Economic costs per person requiring medical care due to WNV is approx. $26,334 ($15,297 - 37,371).
In year 2013, that would be $1,027,026 if ⅓ of 117 reported cases require medical treatment.

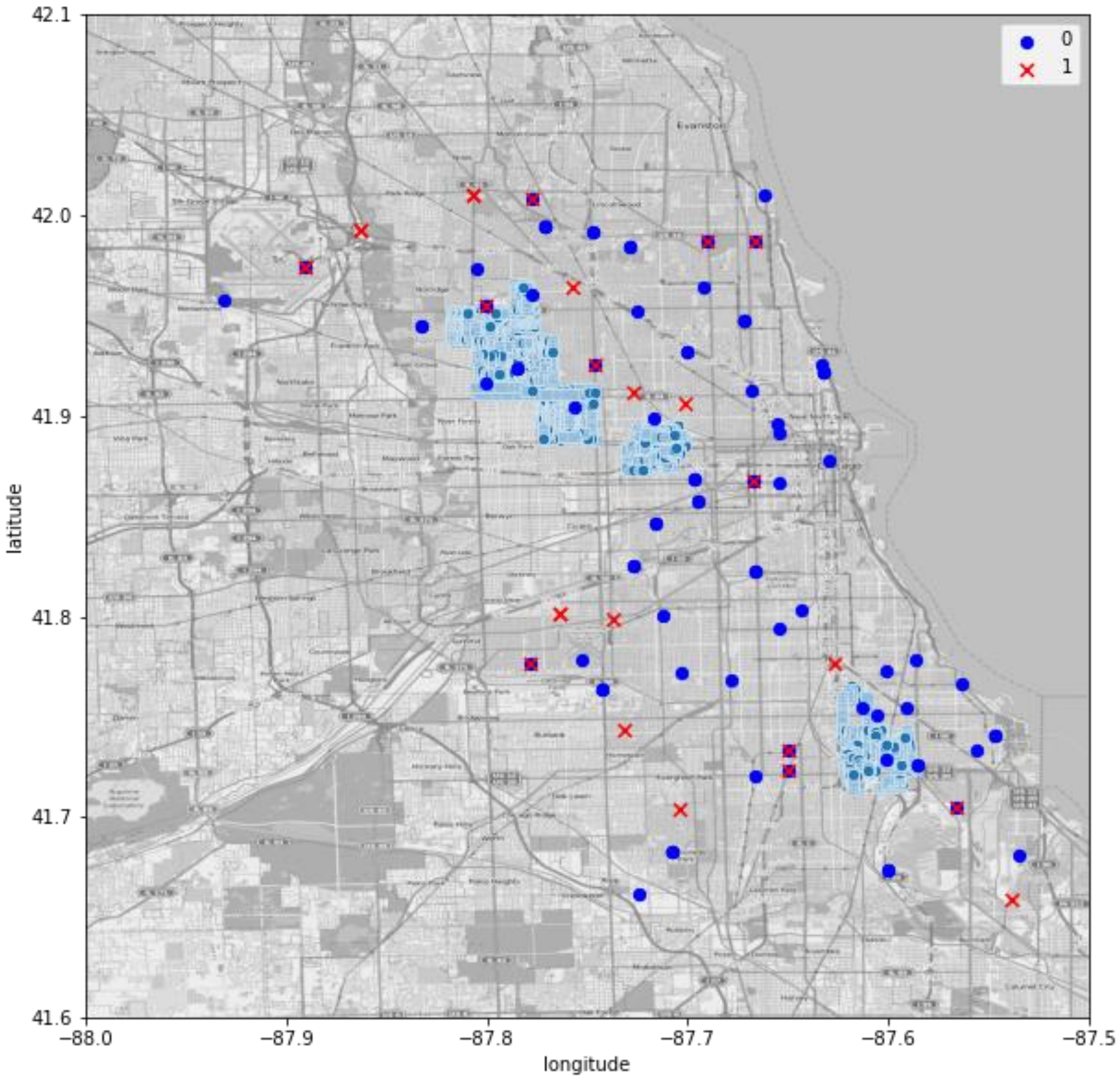
Costs to spray entire Chicago is $815,355.

Prioritise specific areas based on predictions and previous WNV detection.

Effect of spray

Previous studies (Carney et al., 2008; Holcomb et al., 2020) found spraying to be effective to prevent the WNV transmission

Deploying pesticides may have a lagged effect on WNV incidence.



Effect of spraying in weeks 33 and 34 on wnv in week 35, 2013

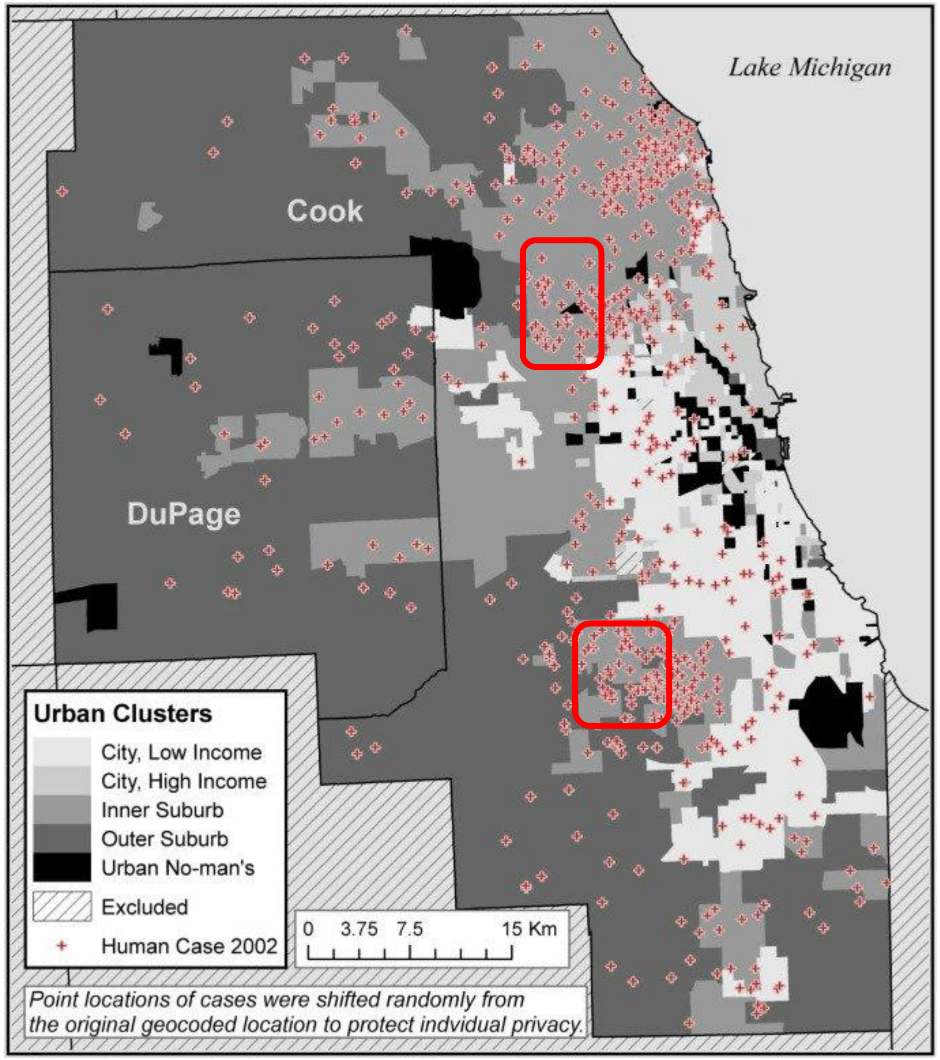
Urban ecology characteristics

locations:

Inland areas (near to Chicago Midway Int'l Airport)

Prioritise areas

- inner suburbs
- with older housing (40s-60s)
- land use allowing sustained interaction of birds and mosquitoes



Urban map of Chicago in 2002 during the WNV outbreak (Ruiz et al., 2007)

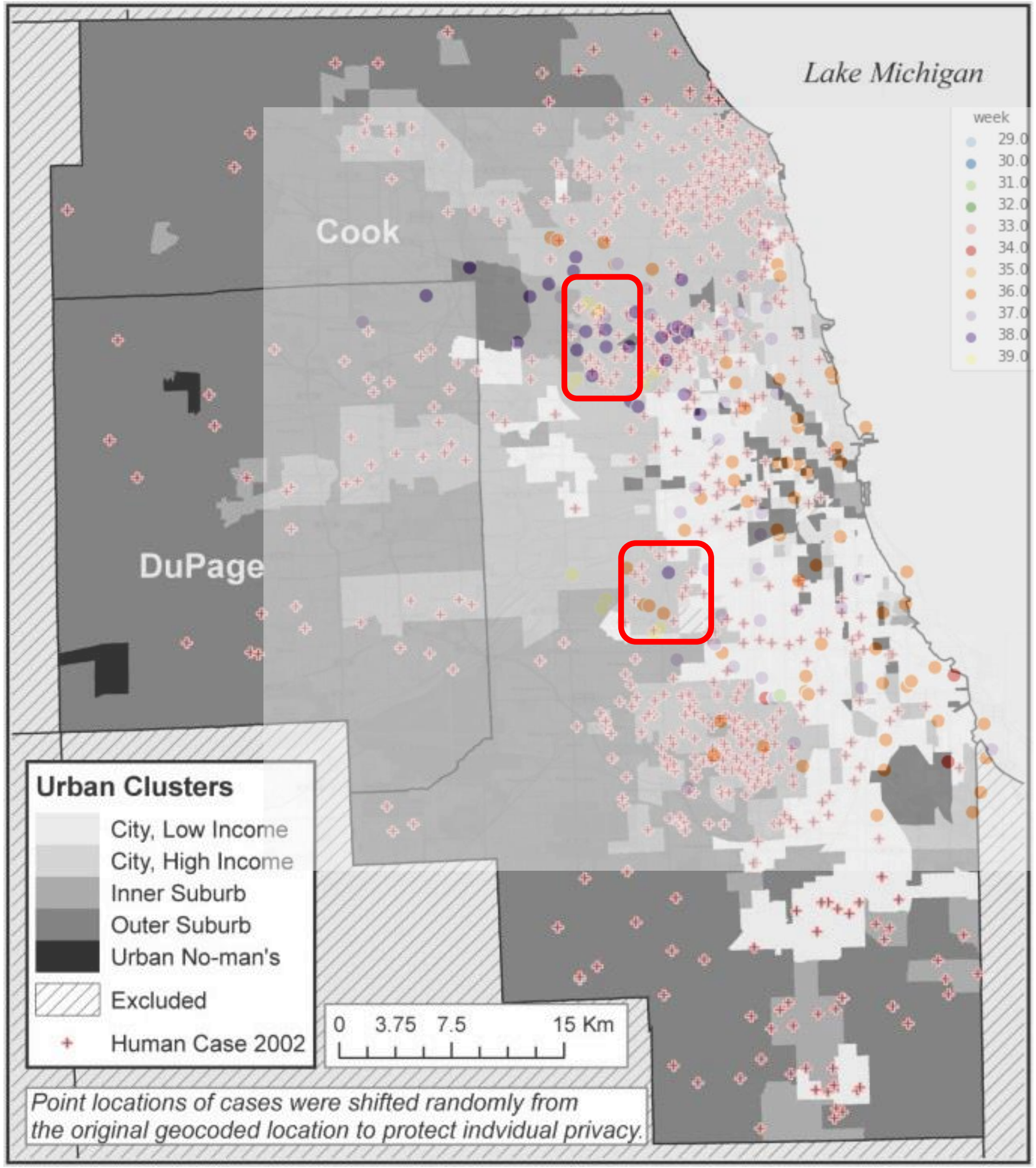
Urban ecology characteristics

locations:

Inland areas (near to Chicago Midway Int’l Airport)

Prioritise areas

- inner suburbs
- with older housing (40s-60s)
- land use allowing sustained interaction of birds and mosquitoes



Conclusion

Spraying should be prioritise -

locations:

Inland areas (near to Chicago Midway Int'l Airport)

inner suburbs

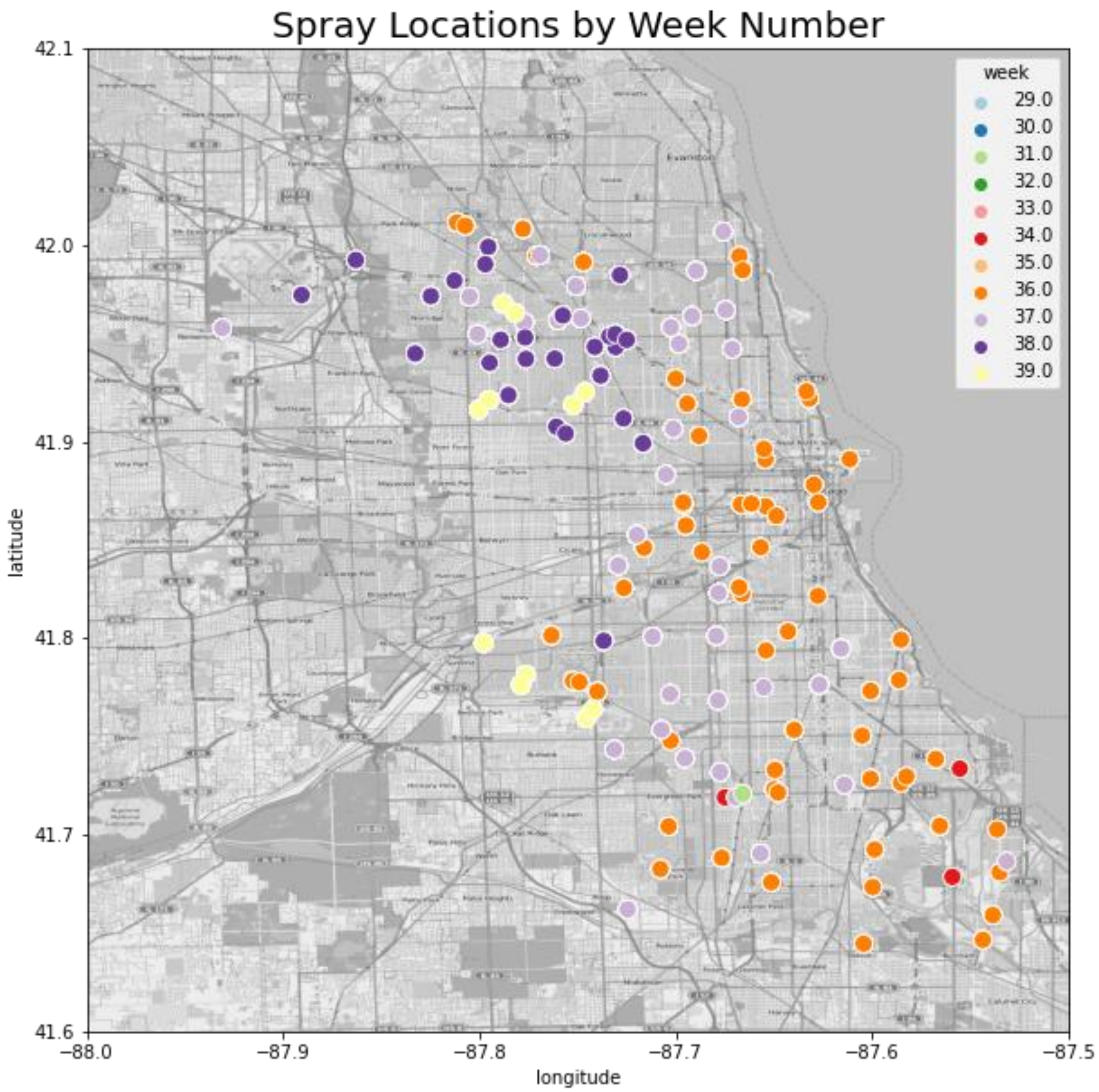
with older housing (40s-60s)

land use allowing sustained interaction of birds and mosquitoes

time:

right before sunrise where the mosquitoes are most active

max 14 days earlier



Short Term Recommendations

1. Increased number of lagged features; Include humidity, sunlight duration

2. More data on spray

3. Cross analysis of other states with WNV

4. Collect data on:

- Population density
- Housing density
- Housing year built
- Sewage flow
- Increased number of traps
- Migratory pattern of birds
  → Amplifier hosts for WNV.

Long Term Recommendations

“Fight Fire with (Genetically Modified) Fire”

Pros:

1. No pesticide usage
2. Lower recurring cost after implementation

Cons:

1. Community pushback
2. Gestation time

NEWS | 03 May 2021

First genetically modified mosquitoes released in the United States

The lab-grown mosquitoes developed by MosquitoMate seek to eliminate the disease-carrying Asian tiger mosquito (*Aedes albopictus*, shown here). Credit: Dennis Kunkel Microscopy/SPL