

Project 4 - West Nile Virus Prediction

Eugene Khoo, Marah Moehammad
Rifqi, Daniel Seto, Mary-Anne Hia,
Florian Combelles



TABLE OF CONTENTS

01

02

03

04

05

BACKGROUND & PROBLEM STATEMENT

- Background /Problem Statement
- Virus information
- Workflow

EXPLORATORY DATA ANALYSIS

- Data Cleaning
- Visualisations

MODELLING

- Features Selection
- Model

SPRAY EFFECTIVENESS ANALYSIS

- Spray Analysis
- Cost and Benefit Analysis

RECOMMENDATIONS & CONCLUSIONS

- Moving Forward
- Recommendations
- Conclusion



BACKGROUND & PROBLEM STATEMENT

TO
↗

Disease And Treatment Agency, Division of Societal Cures In Epidemiology and New Creative Engineering (DATA-SCIENCE)

- Team of data scientists working with the Chicago Department of Public Health
- Focus on mosquitoes-borne diseases



PROBLEM STATEMENT

Due to the rising **seasonal epidemic** of West Nile Virus (WNV) in Chicago, an increasing number of people need medical attention. As such, there is a need for **better mosquito control**.

However, the city of Chicago can only **react** to new cases and **spraying pesticide is costly** and must be done regularly.

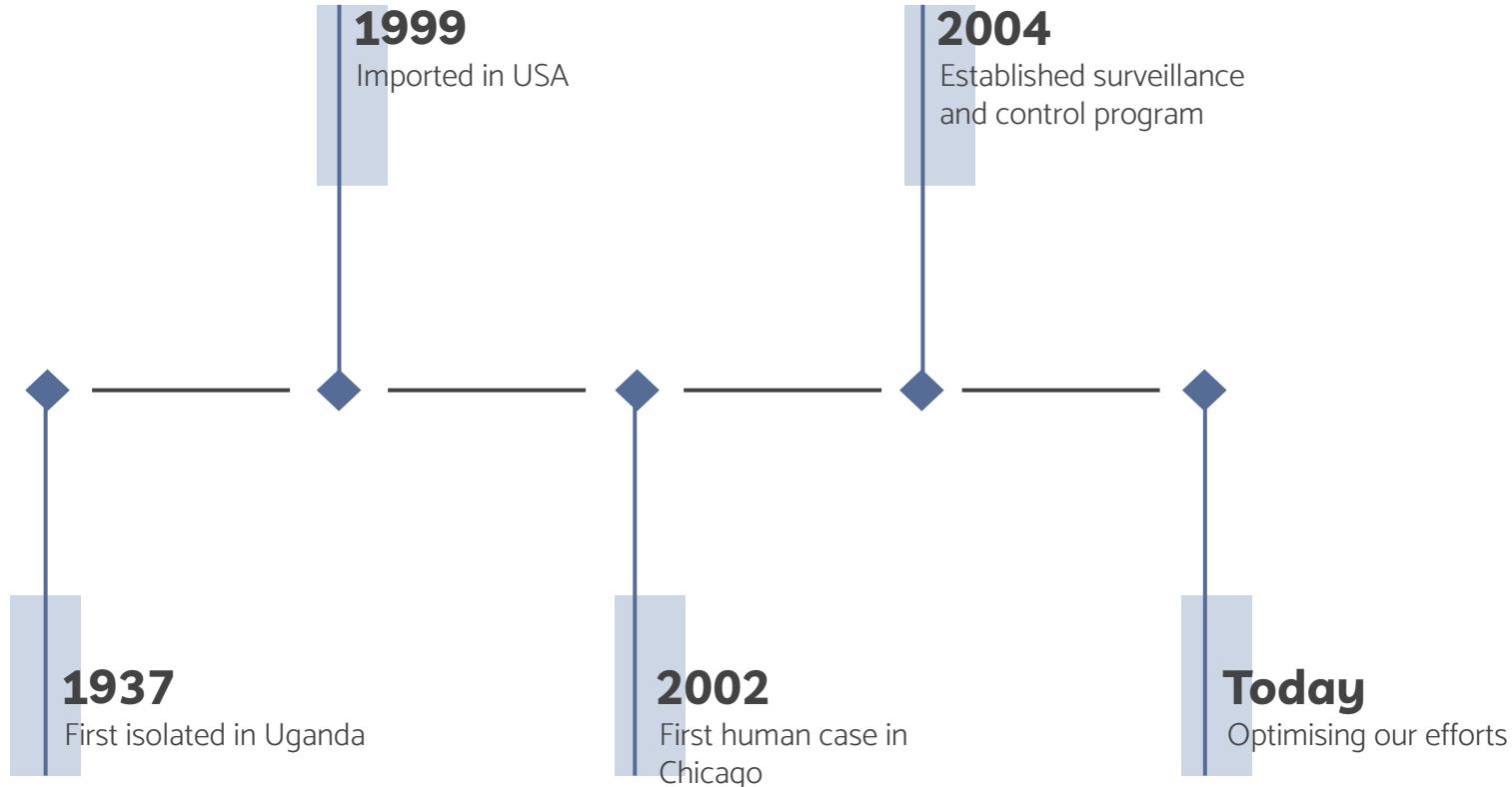
PREVENT

- Predict areas of the city where mosquitoes carrying the WNV will be detected
- Recall of at least **90%**

REMEDIATE

- Optimise pesticide spraying effort throughout the city based on predictions
- Cost/Benefit analysis

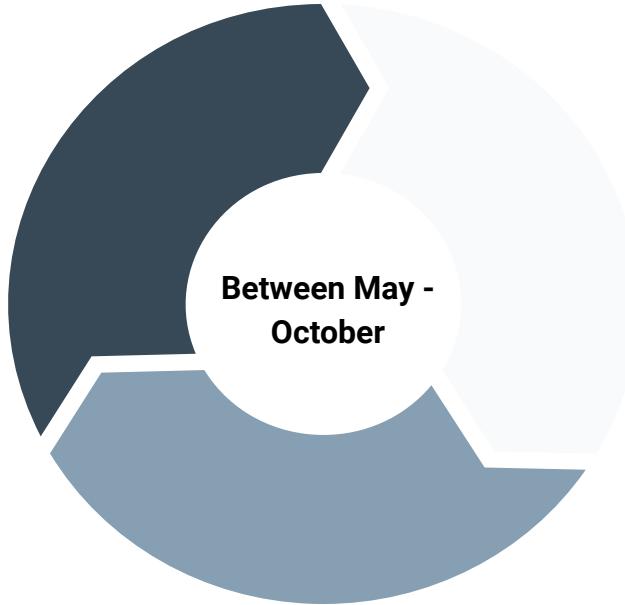
TIMELINE



HOW DOES THE VIRUS DEVELOP?



MEASURES TAKEN

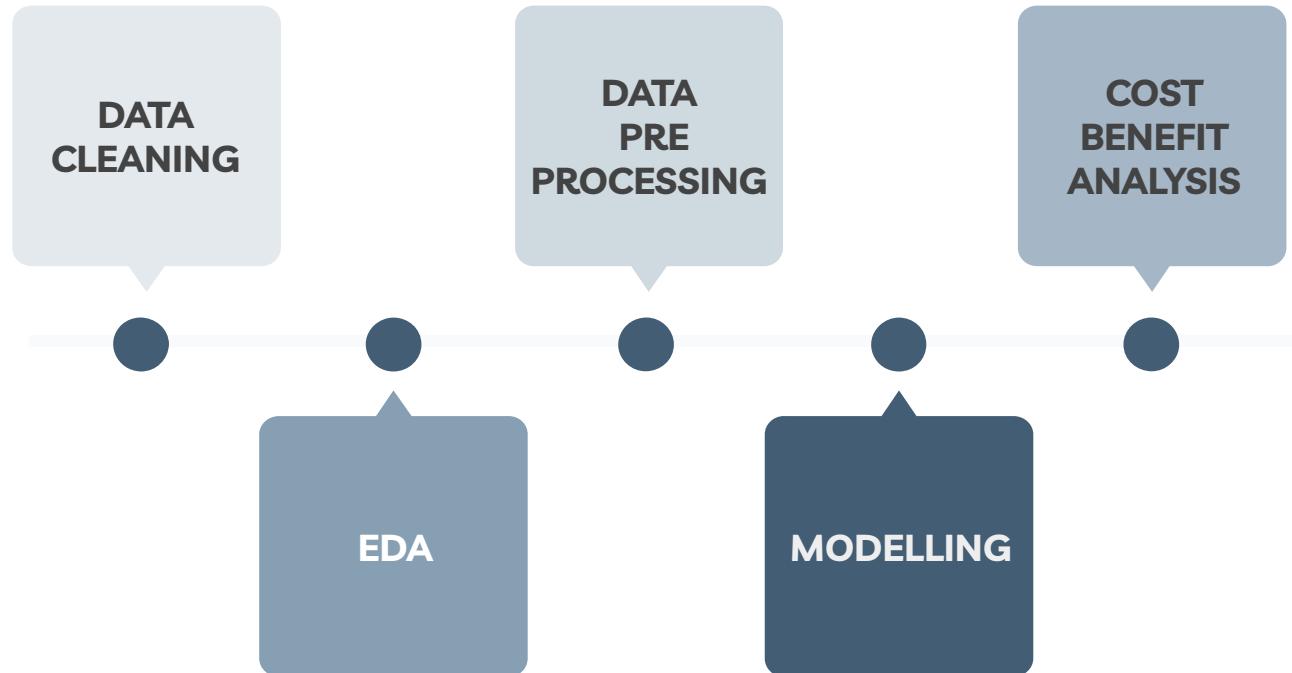


1. Trapping of Mosquitoes

2. Collection and testing of mosquitoes

3. Spraying of insecticide

WORKFLOW





EXPLORATORY DATA ANALYSIS

02

DATA CLEANING (WEATHER)

IMPUTING VALUES

Features with trace ('T') were assigned zero

Missing values from Station 2 were filled using values from Station 1

- Columns: ['Depart', 'Sunrise', 'Sunset', 'Heat', 'Cool', 'SeaLevel', 'WetBulb', 'StnPressure', 'AvgSpeed', 'PrecipTotal']

Average temperature was filled using the mean of minimum and maximum temperature

DROPPING FEATURES

Features were removed due to a lack of relevant data

- 'Water1': Only contained null values
- 'SnowFall': Lack of variability in data (values either zero or trace)
- 'Depth': Only zero values

MERGING WEATHER AND TRAIN DATA

Weather features merged based on each location's proximity to station

DATA CLEANING (TRAIN)

GROUPING DUPLICATE ENTRIES

“When the number of mosquitos exceed 50, they are split into another record (another row in the dataset), such that the number of mosquitos are capped at 50”

Date	Address	Species	Latitude	Longitude	WnvPresent	NumMosquitos
2007-07-18	3800 East 115th Street, Chicago, IL 60617, USA	CULEX PIPiens/RESTUANS	41.686398	-87.531635	1	50
2007-07-25	South Doty Avenue, Chicago, IL, USA	CULEX PIPiens	41.673408	-87.599862	1	50
2007-07-25	South Doty Avenue, Chicago, IL, USA	CULEX PIPiens	41.673408	-87.599862	1	50
2007-07-25	South Doty Avenue, Chicago, IL, USA	CULEX PIPiens	41.673408	-87.599862	1	50
2007-07-25	South Doty Avenue, Chicago, IL, USA	CULEX PIPiens/RESTUANS	41.673408	-87.599862	1	50
2007-07-25	South Doty Avenue, Chicago, IL, USA	CULEX PIPiens/RESTUANS	41.673408	-87.599862	1	50



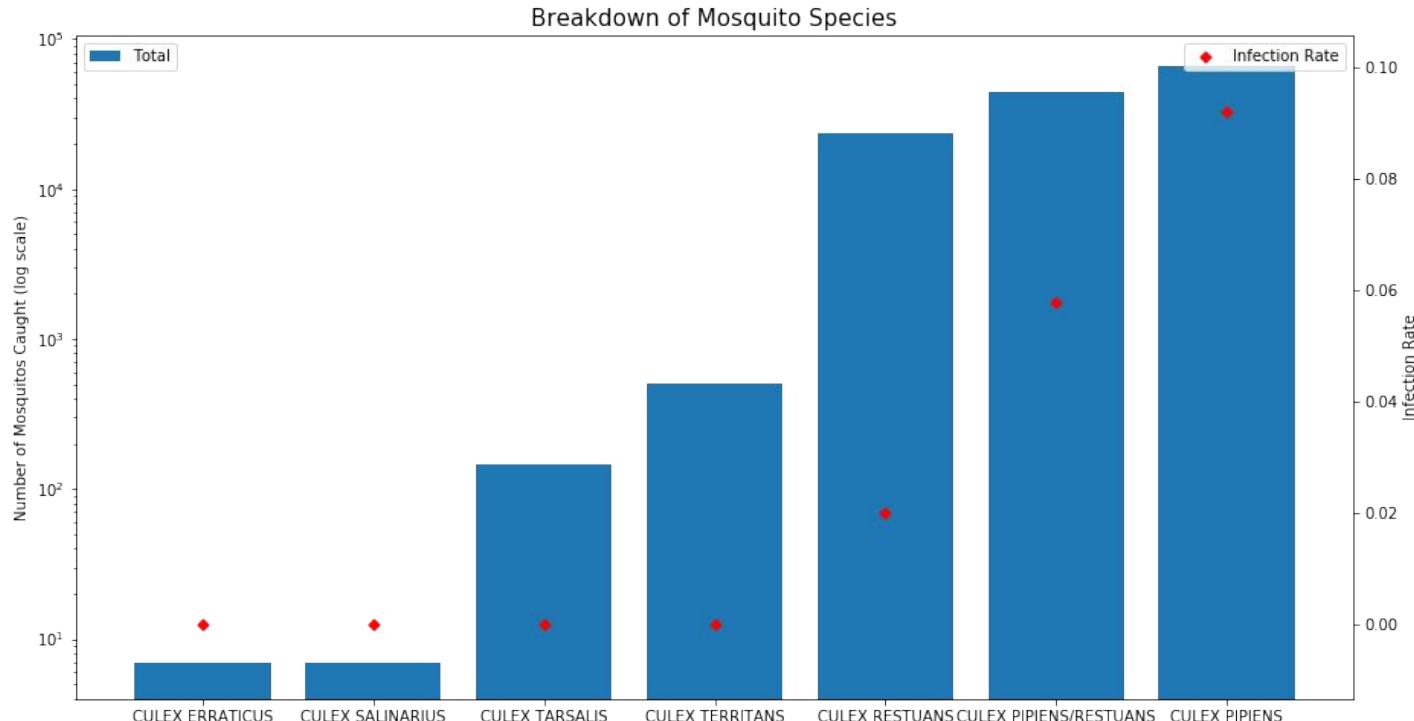
- Grouped based on EDA requirements
- Infection Rates are calculated based on the percentage of WnvPresent == 1

Date	Address	Latitude	Longitude	WnvPresent	NumMosquitos
2007-07-18	3800 East 115th Street, Chicago, IL 60617, USA	41.686398	-87.531635	1	50
2007-07-25	South Doty Avenue, Chicago, IL, USA	41.673408	-87.599862	1	250

Date	Address	Latitude	Longitude	Species	WnvPresent	NumMosquitos
2007-07-18	3800 East 115th Street, Chicago, IL 60617, USA	41.686398	-87.531635	CULEX PIPiens/RESTUANS	1	50
2007-07-25	South Doty Avenue, Chicago, IL, USA	41.673408	-87.599862	CULEX PIPiens	1	150
2007-07-25	South Doty Avenue, Chicago, IL, USA	41.673408	-87.599862	CULEX PIPiens/RESTUANS	1	100

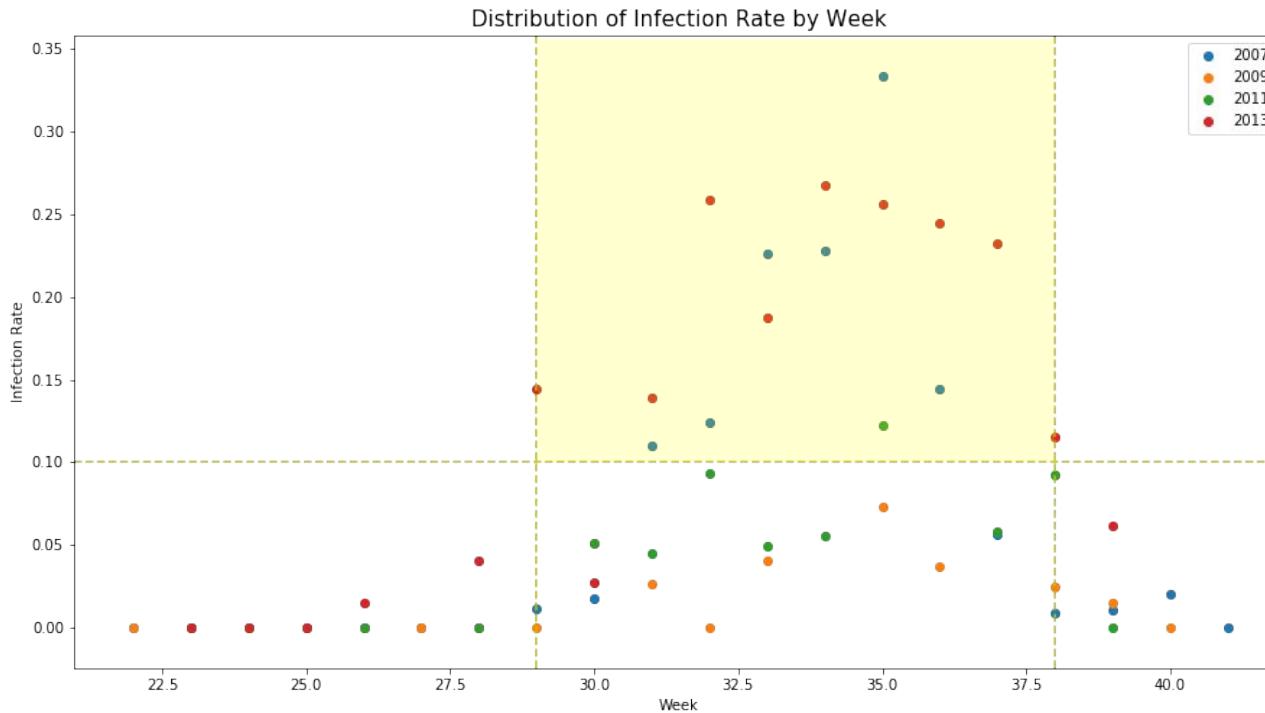
MOSQUITO SPECIES

Culex Restauns and **Culex Pipiens** seem to be the only West Nile Virus carriers



WEST NILE VIRUS INFECTION RATE

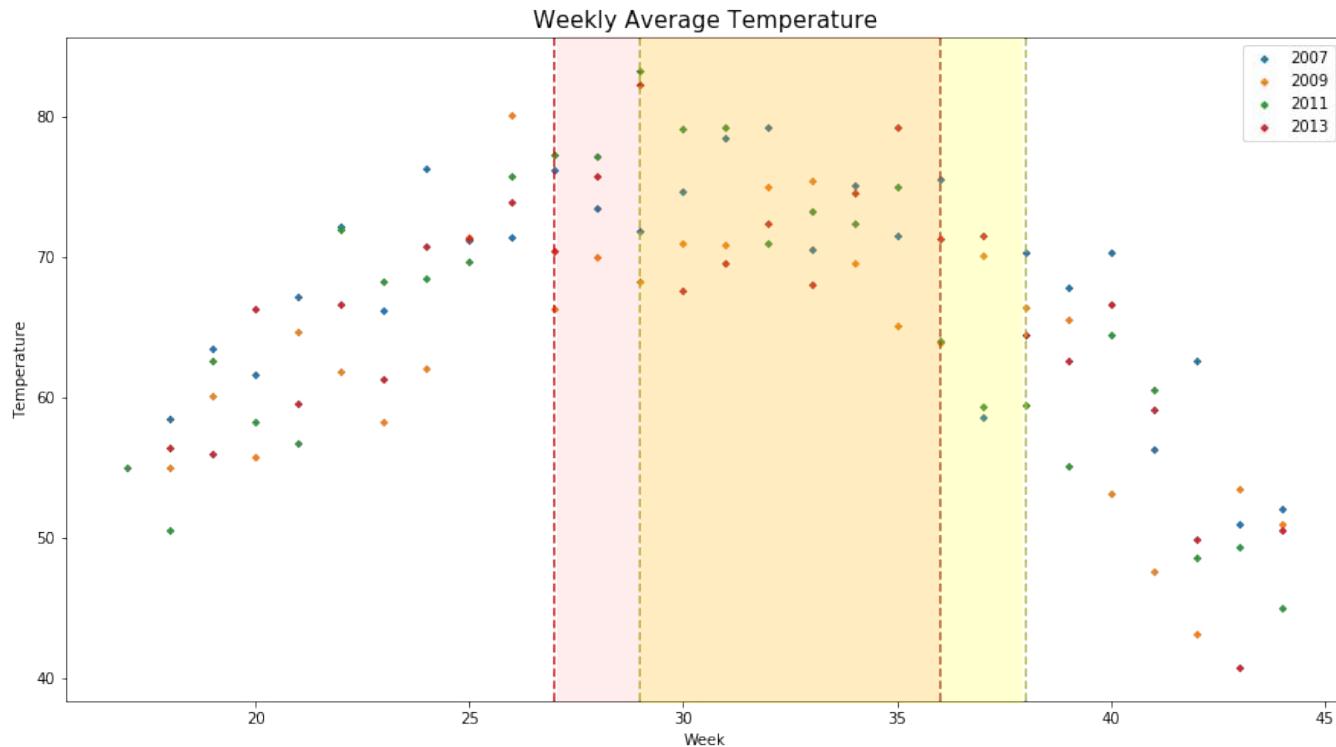
Infection rates above 10% are observed between **weeks 29-38** (mid-July to end-September)



TEMPERATURE

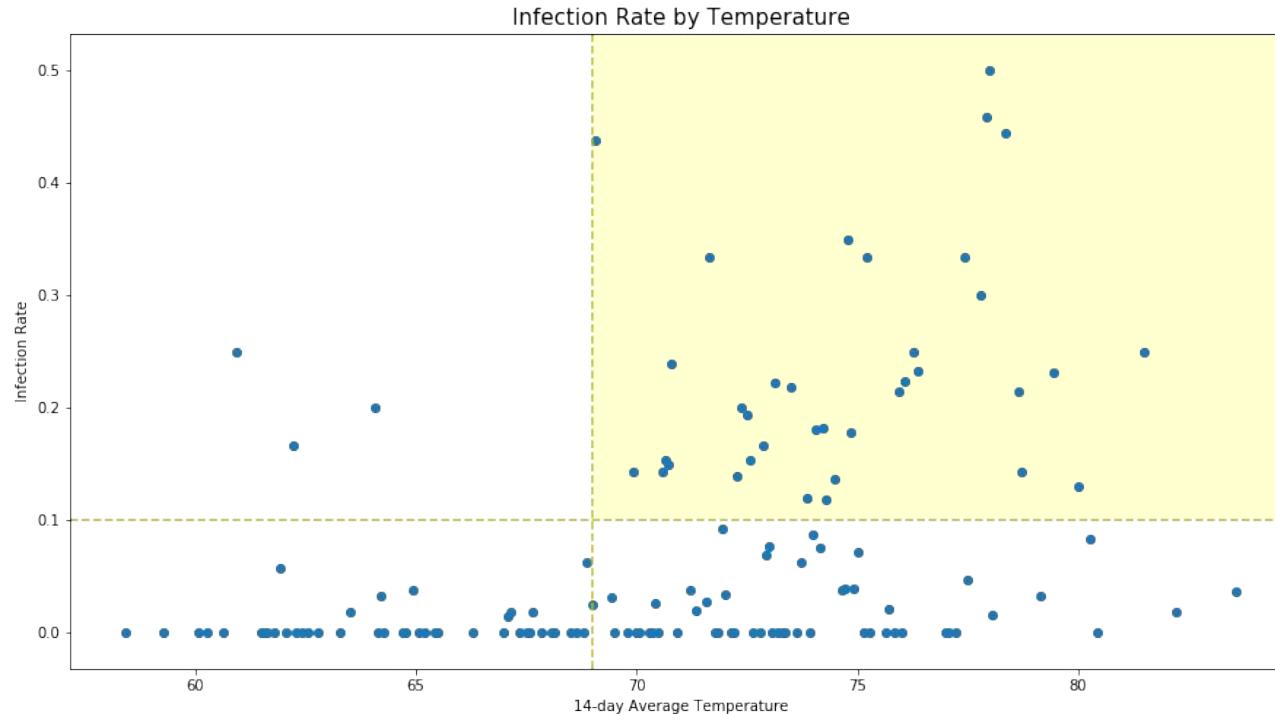
Red zone (week 27-36) - Highest temperatures

Yellow zone (week 29-38) - Highest infection rates



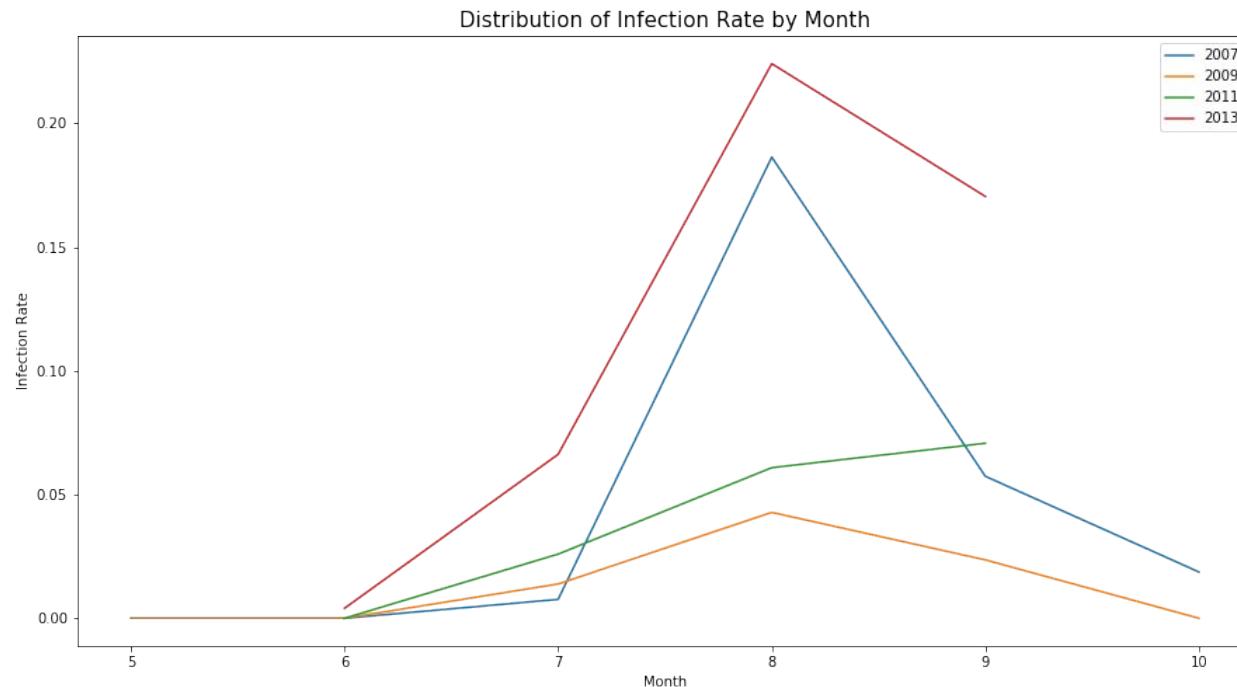
TEMPERATURE

Infection rates above 10% are observed mostly when the **14-day moving average temperature is above 69F**



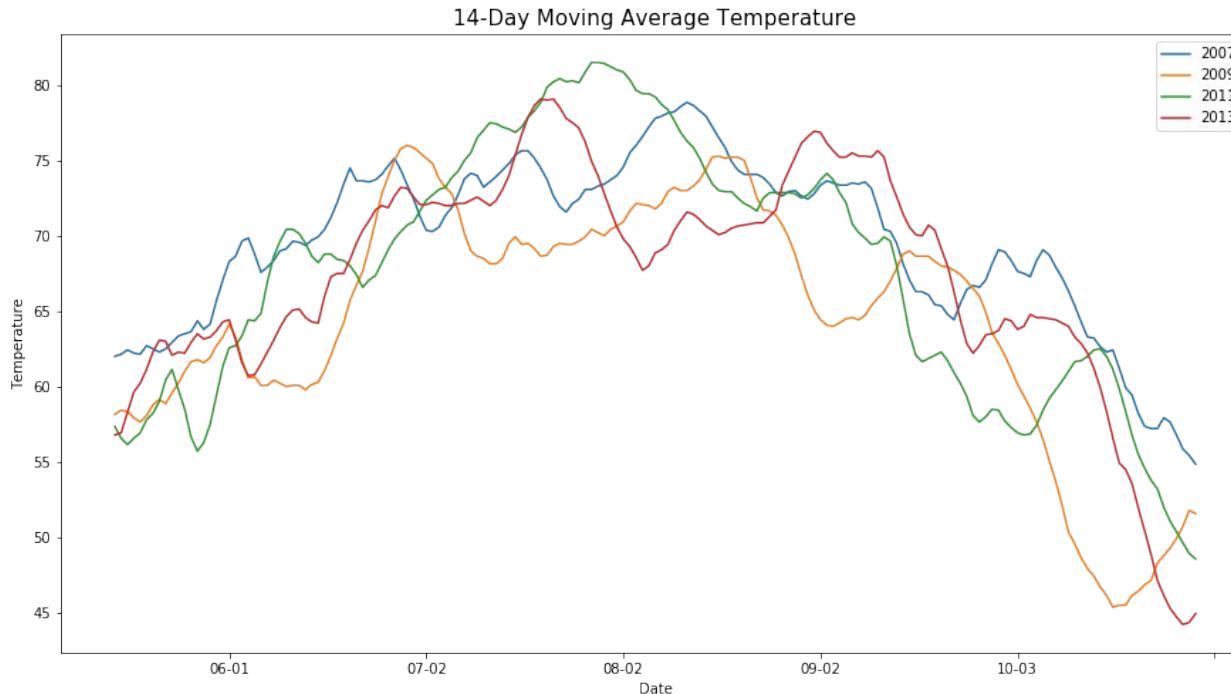
WEST NILE VIRUS INFECTION RATE

2007 and 2013 saw higher infection rates than 2009 and 2011



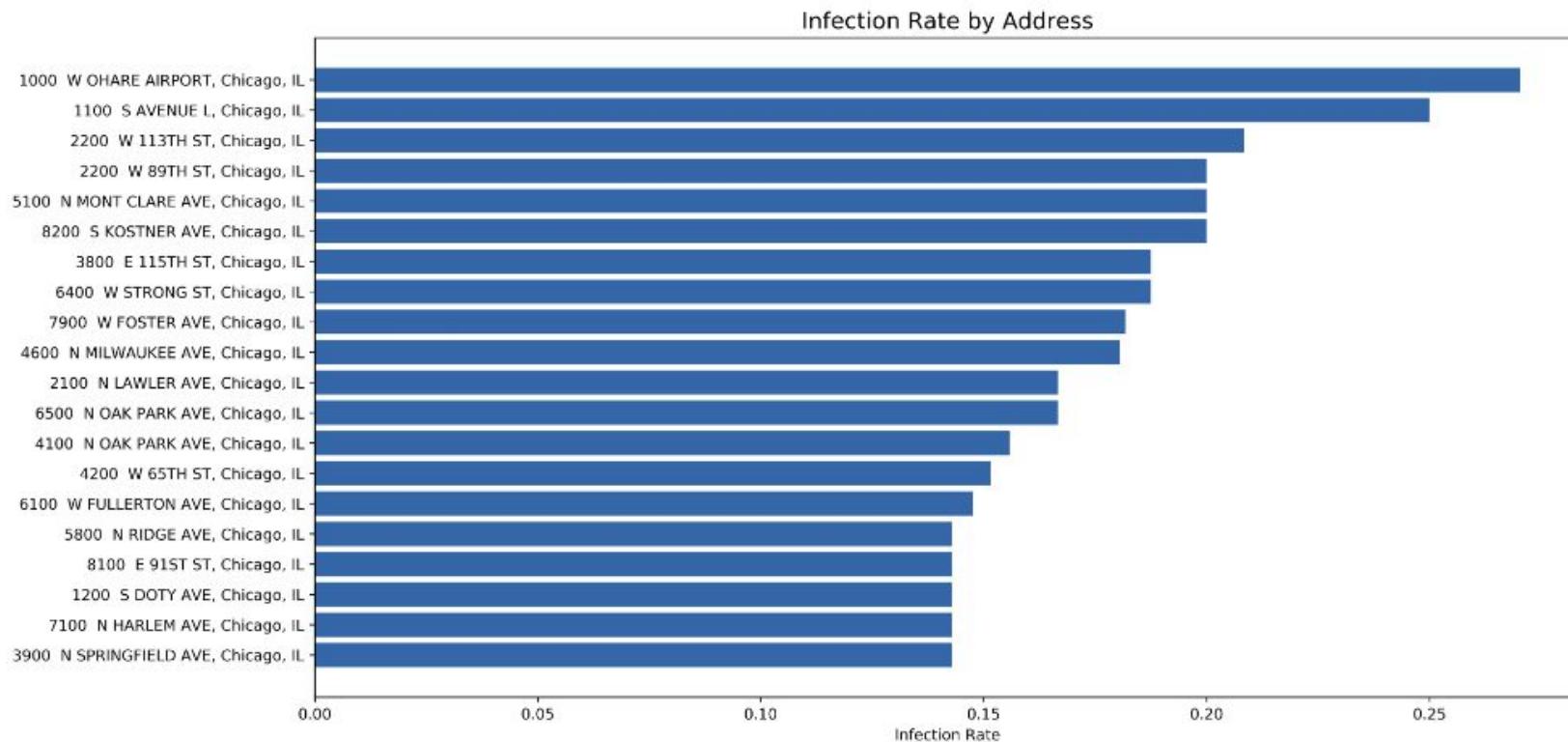
TEMPERATURE

No distinct difference in temperature trends between (2007, 2013) and (2009, 2011)



INFECTION RATE BY LOCATION

Top 20 addresses by infection rate



MODELLING

CO

FACTORS SELECTION

ADDRESS

MOSQUITO
SPECIES

MONTH

WEATHER
CONDITIONS

(e.g. average temperature, wind speed and humidity
of the past 14 days)

DATA PREPARATION & MODELLING



TRAINING MODELS & MEASURE OF SUCCESS

Classification Models:

- Logistic Regression
- K-Nearest Neighbors
- Random Forest
- AdaBoost
- Gradient Boosting

Evaluation Metrics:

- Recall

$$\text{recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

- AUC-ROC score
Measures how well a model is able to distinguish between the labels

What we're looking for:

- ↑ Higher Recall
- ↑ Higher AUC-ROC score
- ↑ Better Model Performance

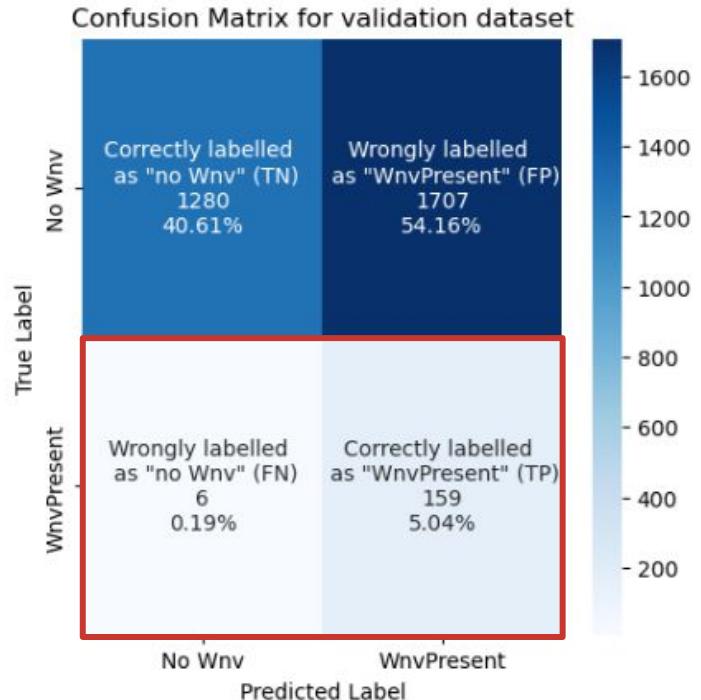
ADABOOST CLASSIFIER ACHIEVED 96% RECALL SCORE

	model	management of imbalanced data	scaler	Recall for train	Recall for val	ROC-AUC Score for train	ROC-AUC Score for val
0	KNN (baseline)	oversampling of minority class using replacement	MinMaxScaler	1.000000	0.478788	0.971668	0.712855
1	LogReg	oversampling of minority class using replacement	MinMaxScaler	0.757032	0.793939	0.817600	0.825725
2	Random Forest Classifier	oversampling of minority class using replacement	-	0.992824	0.272727	0.995104	0.779056
3	Adaptive Boosting Classifier	oversampling of minority class using replacement	-	0.842997	0.793939	0.891544	0.845907
4	Gradient Boosting Classifier	oversampling of minority class using replacement	-	0.942164	0.757576	0.928106	0.850787
5	LogReg	SMOTE	MinMaxScaler	0.787600	0.793939	0.817330	0.824197
6	KNN	SMOTE	MinMaxScaler	0.874713	0.442424	0.937850	0.773796
7	Random Forest Classifier	SMOTE	-	0.987945	0.309091	0.986025	0.794374
8	Adaptive Boosting Classifier	SMOTE	-	0.922646	0.654545	0.882579	0.840767
9	Gradient Boosting Classifier	SMOTE	-	0.950057	0.587879	0.900633	0.847292
10	Adaptive Boosting Classifier with hyperparameter tuning	oversampling of minority class using replacement	MinMaxScaler	0.987047	0.963636	0.853649	0.837977

Selected model for tuning

Best result post-tuning

ADABOOST CLASSIFIER ACHIEVED 96% RECALL SCORE



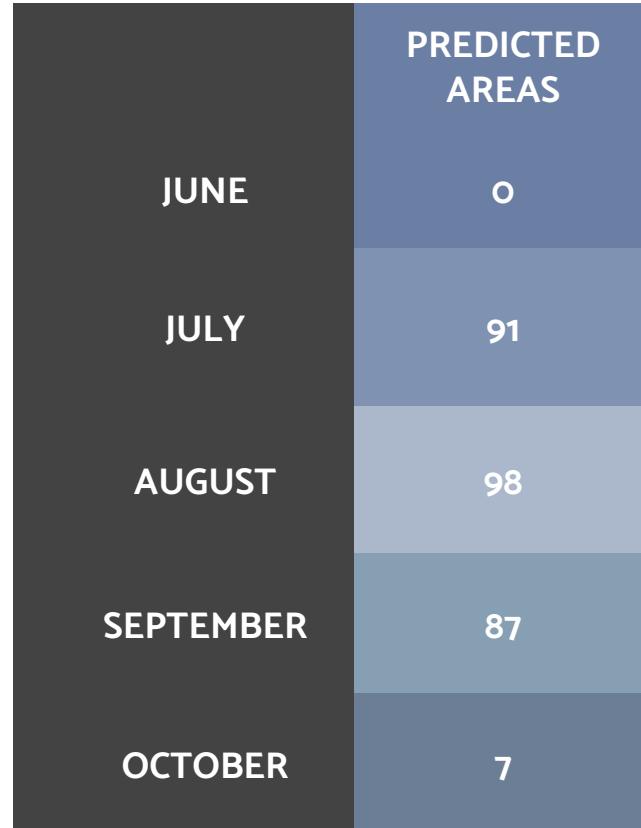
Highly accurate in predicting the presence of WNV

- **Total positive observations: 165**
- **Correct positive classifications: 159**

HOW MANY PREDICTED AREAS TO SPRAY ?

PREDICTED AREAS WITH WNV : 98 out of 138

- Using our Machine Learning we predict that 90+ areas will have a WNV positive mosquito each month from July to September
- Necessity to factor in these predicted areas during spray efforts



7 AREAS TO SPRAY IN OCTOBER

Address	7	8	9	10
3900 North Springfield Avenue, Chicago, IL 60618, USA	0.567349	0.808627	0.701159	0.516486
5100 North Mont Clare Avenue, Chicago, IL 60656, USA	0.631653	0.818448	0.686384	0.516486
8200 South Kostner Avenue, Chicago, IL 60652, USA	0.600009	0.808627	0.801848	0.516486
9100 West Higgins Road, Rosemont, IL 60018, USA	0.543492	0.808627	0.801848	0.516486
ORD Terminal 5, O'Hare International Airport, Chicago, IL 60666, USA	0.619090	0.808627	0.686384	0.516486
O'hare Court, Bensenville, IL 60106, USA	0.593319	0.817206	0.684588	0.516486
4800 West Montana Street, Chicago, IL 60639, USA	0.544875	0.707553	0.597851	0.502944



SPRAY EFFECTIVENESS ANALYSIS

04

A photograph of a dense urban cityscape, likely New York City, showing a street filled with traffic and tall buildings. The scene is slightly hazy, suggesting a busy, sunlit day.

SPRAY ANALYSIS

HOW IS SPRAYING CONDUCTED ?



PESTICIDE

Zenivex E4 RTU mosquito adulticide

- Ground and aerial applications
- Agricultural and urban areas



APPLICATION

Truck mounted fogger

- Ultra Low Volume fogging
- Mosquitoes killed by contact



MANPOWER

People needed to perform spray activity

- 2 Technicians
- 2 Supervisors from Chicago Department of Street and Sanitation

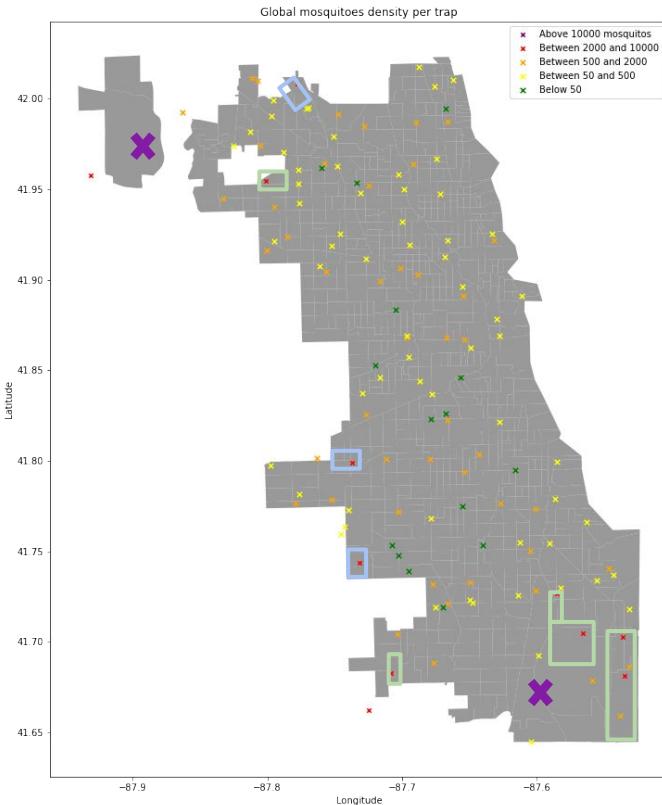


TIMING

Spraying during Wnv period

- July to September
- Spray done after dusk till early morning

WHERE ARE MOSQUITOES MOST PRESENT?



2 Areas with much higher mosquito number

- Lake Calumet and South Deering Industrial Area
- O'Hare Airport



Concentration of mosquitos around parks...

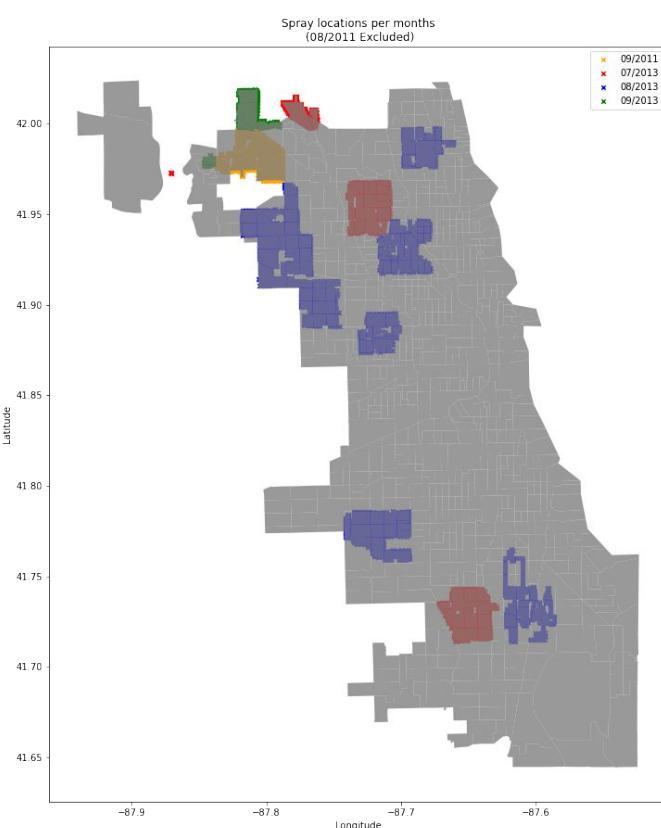
- Wolf Lake and Eggers Grove
- Vet's Park
- Marian Byrnes Park and Natural Area
- Wildwood Park
- Park 601



... Or in greener neighbourhood

- Scottsdale
- Archer Heights
- South Mount Greenwood

WHERE DID CHICAGO FOCUS SPRAY EFFORTS?



Spraying done during 2011 and 2013

- September 2011
- July, August and September 2013



Spray Locations

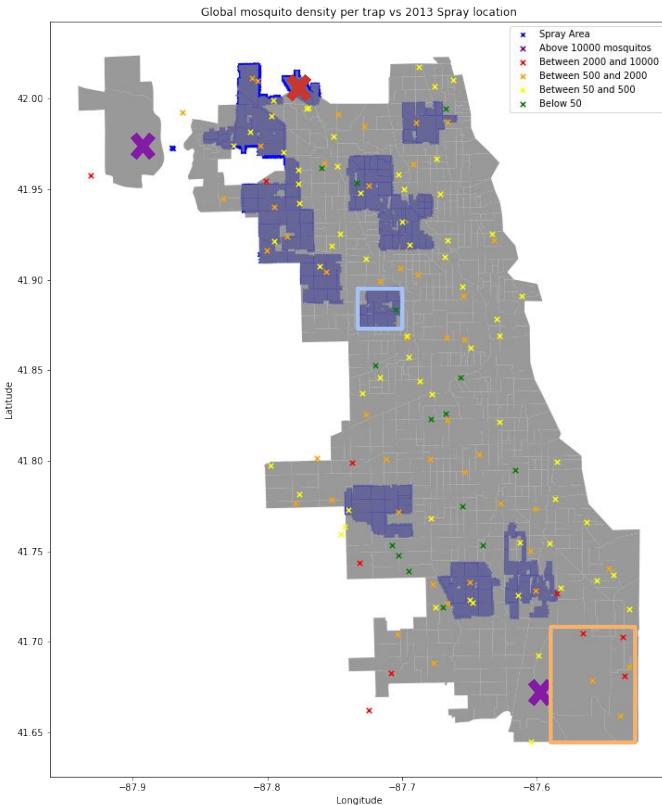
- Urban areas with high density of habitations
- Parks and green areas are excluded from the spraying



92.92 Square Kilometers sprayed

- 2011: 8.48 km/sqr - 2095.45 acres
- 2013: 84.44 km/sqr - 20865.57 acres

ARE THE CORRECT AREAS SPRAYED ?



Wrong areas sprayed

- Areas sprayed are not among the highest mosquito density areas
- Belmont Cragin/Garden** area has less than 50 mosquitos



High density areas and parks not targeted

- Only 1 area with more than 2000 mosquitos sprayed - **Wildwood**
- Lake Calumet** and **O'Hare Airport** ignored



South East Chicago forgotten

- Highest density of mosquitos in the city
- No trap below 500 mosquitos

A wide-angle photograph of a city street, likely New York City, during the day. The scene is filled with tall, multi-story buildings of various architectural styles. In the foreground, several vehicles are visible, including a large white truck, a red car, and a dark SUV. A bridge structure is visible in the distance across a body of water. The overall atmosphere is busy and urban.

COST BENEFIT ANALYSIS

HOW MUCH DOES SPRAYING COST ?

	SURFACE	PESTICIDE COST	MANPOWER COST	LOGISTICS COST
2011	2095.45 Acres	\$2688	\$400*	\$18000**
2013	20865.57 Acres	\$24192	\$4400*	\$18000**
TOTAL	22961.02 Acres	\$26880	\$4800	\$36000
				67680\$ \$2.95/acre

* Based on 4 people per spray for 5h night time work (20\$/hr wage)

** Based on Truck and Spray Equipment cost of 180000\$ with 10 year amortization

HOW CAN WE OPTIMIZE SPRAYING ?



1

LEVERAGE DATA SCIENCE

- Use Machine Learning to predict areas where virus will be detected
- Allow for better preparation and anticipation

AGRI DRONES

- Adapted for use in rural and green areas
- Cheaper than managing fleet of vehicles

2



3

2 STEPS SPRAYING PROGRAM



- Yearly spray for high density
- Targeted spray for positive clusters

HOW CAN WE OPTIMISE SPRAYING ?



SPRAY HIGH DENSITY AREA

Main aim: Control mosquito population

- Yearly spray of areas with high mosquitos density
- Spray when weather is conducive to mosquito breed



SPRAY POSITIVE BLOCKS

Second aim: Prevent virus spread

- Targeted spray of blocks when positive virus is predicted
- Fast and precise spraying

HOW MUCH CAN WE SAVE ?

	SURFACE	PESTICIDE COST	MANPOWER COST	LOGISTICS COST
Yearly Spray	13810.72 Acres	\$16128	\$200*	\$8400***
Targeted Spray	0.15 Acre	\$0.16	\$40**	\$0
TOTAL****	13825.72 Acres	\$16176	\$14000	\$8400

**\$36576
\$2.64/acre**

* Based on 2 people per spray for 5h night time work (20\$/hr wage)

** Based on 2 people per spray for 1h night time work (20\$/hr wage) for 472 spray a year

*** Based on Drone and Equipment cost of 25000\$ with 3 year amortization

**** Based on an estimation of 100 targeted sprays per month over 3 months

HOW MUCH CAN WE SAVE ?

CURRENT COST

\$ 2.95 /
Acre

PROJECTED COST

\$ 2.64 /
Acre

BENEFIT INCREASE



11 %



CONCLUSIONS & RECOMMENDATIONS

SO

MOVING FORWARD

IRREGULARITY IN DATA COLLECTION

- Data collected from traps irregular across months
- Not all traps recorded each time

IMBALANCED DATA

- More data on class 1 (positive West Nile Virus)

MORE SUPPORTING DATA

- Natural disasters
 - Flooding
- Population density
- Construction sites

RECOMMENDATIONS



Monitor Weather

- Higher temperature
- Higher humidity
- Target areas high in mosquitoes



Virus Prediction

- Spray the whole block based on virus prediction



Use of Drones

- Cut manpower and logistic costs
- Inaccessible areas for vehicles
- More precise
- Trap collection



Eco-Friendly Insecticide

- Parks and greenery
- Less toxic for all

CONCLUSION



PREVENT

2 out of 6 mosquito species carry the virus

- Culex Pipiens & Culex Restuans

Model accuracy of **96%**

- Recall - True positive clusters

Identify areas to apply control measures

- Lake Calumet, O'Hare Airport, and Southeast Chicago



REMEDIATE

Save \$\$\$ with 2-step spraying program

- Yearly spray in high mosquito density areas
- Targeted spray for positive clusters

Reduce overall mosquitoes to reduce virus

THANK YOU

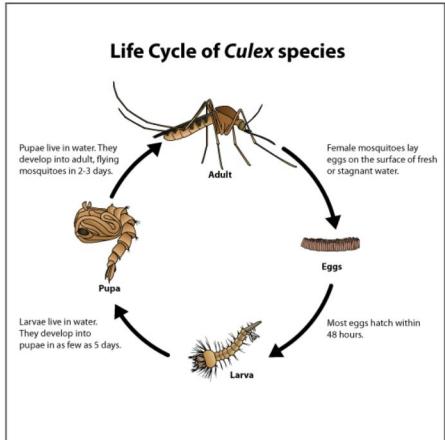
Does anyone have any questions?

SOURCES

- <https://www.chicagotribune.com/news/ct-xpm-2003-03-09-0303090283-story.html>
- https://www.chicago.gov/city/en/depts/cdph/provdrs/healthy_living/news/2021/august/city-to-spray-insecticide-wednesday-to-kill-mosquitoes.html
- <https://www.chicago.gov/content/dam/city/depts/cdph/Mosquito-Borne-Diseases/Zenivex.pdf>
- <https://www.cdc.gov/westnile/index.html>
- <https://www.health.nd.gov/sites/www/files/documents/Files/MSS/WNV/Resources/North%20Dakota%20Mosquito%20Control%20Product%20List%20Tabulation%202022.pdf>
- <https://www.empirepestcontrol.co.uk/5-ways-drones-are-used-in-pest-control/#:~:text=Agricultural%20Use&text=Drone,s%20are%20also%20sometimes%20equipped,easy%20reach%20of%20traditional%20methods.>
- <https://www.extension.purdue.edu/extmedia/ppp/ppp-58.pdf>
- <https://www.dronenerds.com/products/drones/enterprise-drones/dji-agras-series/agras-t30/dji-agras-t30-ready-to-fly-kit-agrast30rtf-dronenerds.html>

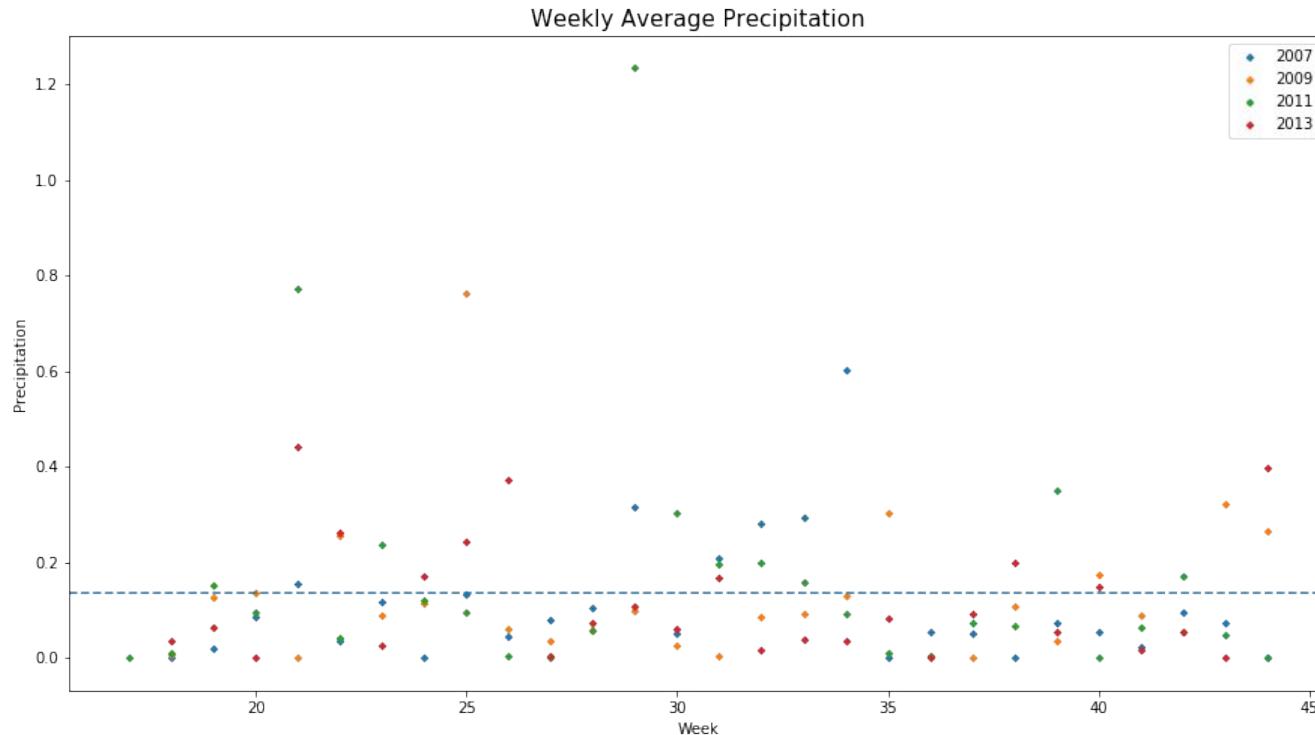
Relative Influence of Land Use, Mosquito Abundance, and Bird Communities in Defining West Nile Virus Infection Rates in *Culex* Mosquito Populations ([link](#))

About 7-10 days for an egg to develop into an adult mosquito



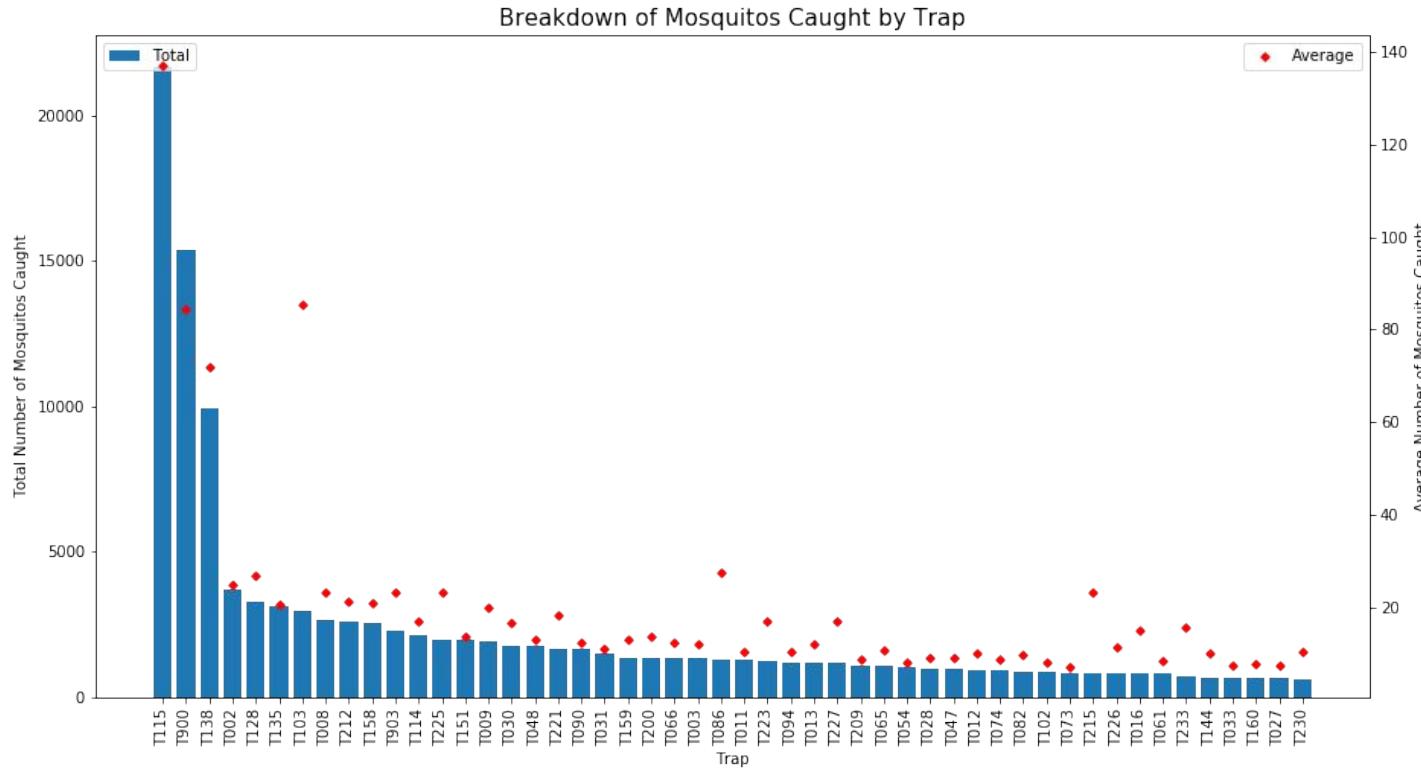
PRECIPITATION

Chicago experiences consistent rain/snowfall on with an average of 0.135 inches per day
No clear trends in precipitation



MOSQUITO TRAP DEPLOYMENT

Traps **T115, T900, T138 and T103** caught the most mosquitos on average





Use our editable graphic resources...

You can easily resize these resources, keeping the quality. To change the color, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Don't forget to group the resource again when you're done.



...and our set of editable icons

You can resize these icons, keeping the quality.

You can change the stroke and fill color; just select the icon and click on the paint-bucket/pen.

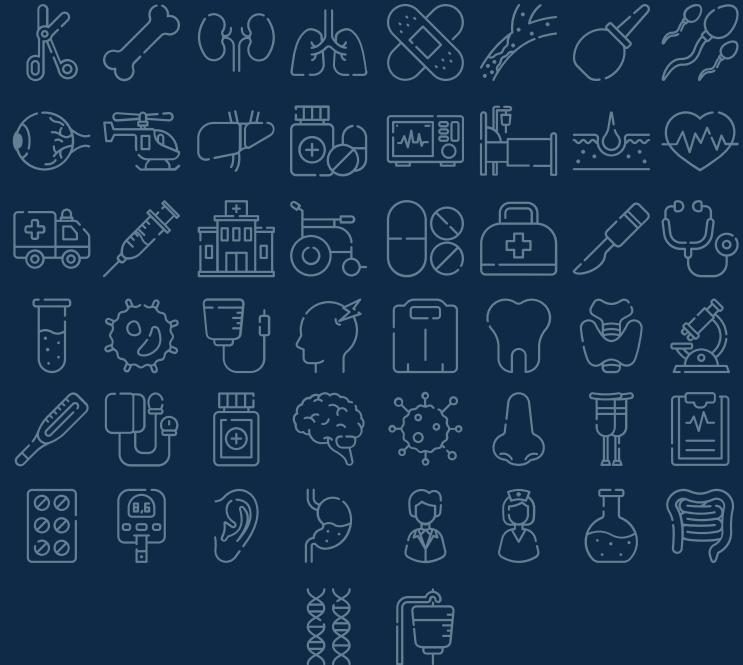
In Google Slides, you can also use Flaticon's extension, allowing you to customize and add even more icons.



Educational Icons



Medical Icons



Business Icons



Teamwork Icons



Help & Support Icons



Avatar Icons



Creative Process Icons



Performing Arts Icons



Nature Icons



SEO & Marketing Icons



