



# TACKLING THE WEST NILE VIRUS OUTBREAK WITH DATA

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



# TABLE OF CONTENTS

01

INTRODUCTION

02

FIRST LOOK AT DATA

03

DATA EXPLORATION &  
FEATURE ENGINEERING

04

PREDICTIVE MODELING

05

COST BENEFIT ANALYSIS

06

RECOMMENDATIONS



# 01 • INTRODUCTION



# DID YOU KNOW?

In 2002,

**>4,150**

cases of West Nile  
virus infection  
were reported

In 2002,

**284**

fatalities linked to  
West Nile virus  
infections

# DID YOU KNOW?

Total Productivity Cost of 284 Fatalities:

≈ USD 345.7 M

# IN ILLINOIS



# ENDEMIC

since 2002

**1,371**

cases reported from 2005 to 2016

**906**

cases from Chicago (66%)

# IN CHICAGO

In 2002,

**225**

cases of West Nile  
virus infection  
were reported

In 2002,

**22**

fatalities linked to  
West Nile virus  
infections

# IN CHICAGO

Total Productivity Cost of 22 Fatalities:

≈ USD 26.8 M

# MOSQUITO FACTS

It takes

**7-10 DAYS**

To develop from  
an egg to an adult

Thrive at around

**27 DEG C**

In humid climates

Most active at

**8PM-6AM**

Or the dusk to  
dawn hours



# CURRENT STRATEGY

**FIGHT THE BITE**  
MOSQUITOES SPREAD DISEASES - PROTECT YOURSELF

CITYOFCHICAGO.ORG/MOSQUITOES





# ENHANCED STRATEGY

- Actively detect at-risk sites
- Prompt spraying
- Reduce public exposure to virus

# PROBLEM STATEMENT

As part of the Disease And Treatment Agency's DATA-SCIENCE division, our goal is to create an effective pesticide deployment strategy in response to the West Nile Virus endemic through data analysis and modeling, allowing us to:

- allocate resources efficiently
- minimise pesticide cost
- prioritise public health and safety



## 02 • FIRST LOOK AT DATA





# DATA SOURCES

---



## TRAIN, TEST, SPRAY

Chicago Department of Public Health



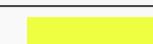
## WEATHER

National Oceanic and Atmospheric Administration



## CHICAGO COMMUNITY AREAS

City Government of Chicago



# DATA OVERVIEW

## TRAIN

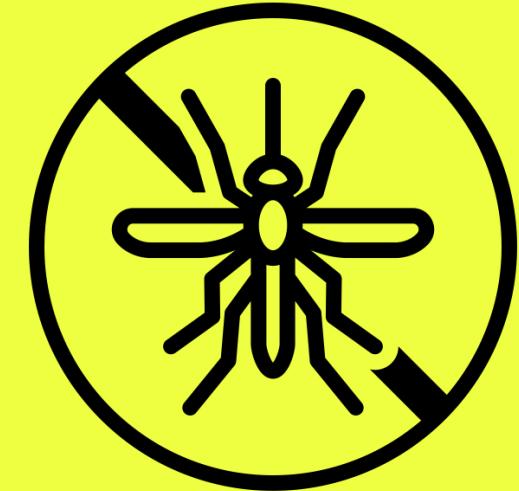
- **Data Points:**  
10,506
- **Features:**  
12
- **Period:**  
2007, 2009, 2011, 2013
- **Missing Values:**  
Observed in 0 features

## TEST

- **Data Points:**  
116,293
- **Features:**  
11
- **Period:**  
2008, 2010, 2012, 2014
- **Missing Values:**  
Observed in 0 features

# 'TRAP'

- Mosquito count records exceeding 50 are divided into separate entries
- Ensures number of mosquitoes does not exceed 50



# DATA OVERVIEW

## WEATHER

- **Data Points:**  
2,944
- **Features:**  
22
- **Period:**  
2007 - 2014
- **Missing Values:**  
Observed in 14 features

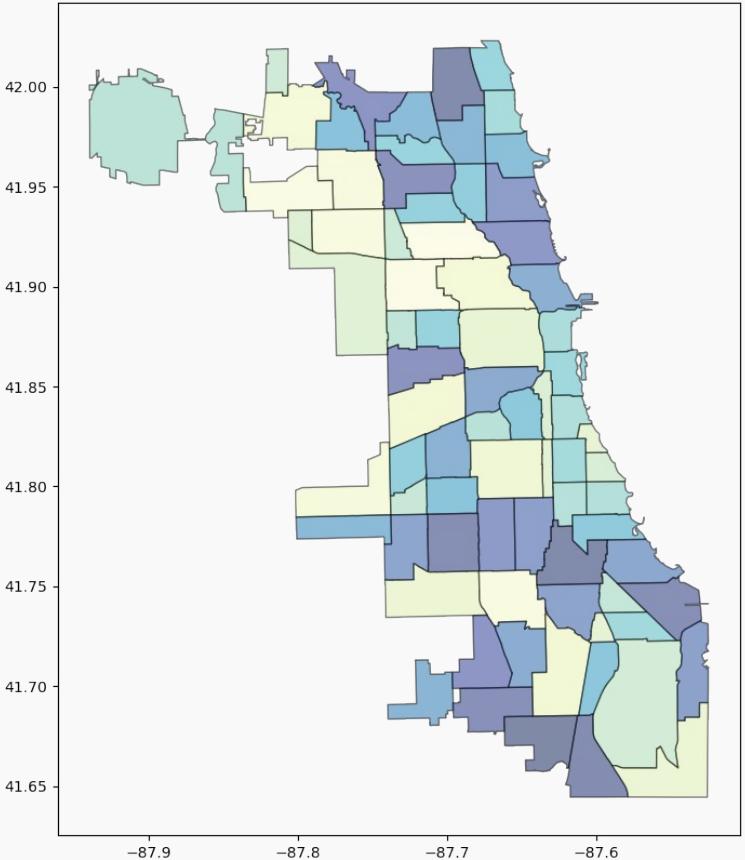
## SPRAY

- **Data Points:**  
14,835
- **Features:**  
4
- **Period:**  
2011, 2013
- **Missing Values:**  
Observed in 1 feature

# DATA OVERVIEW

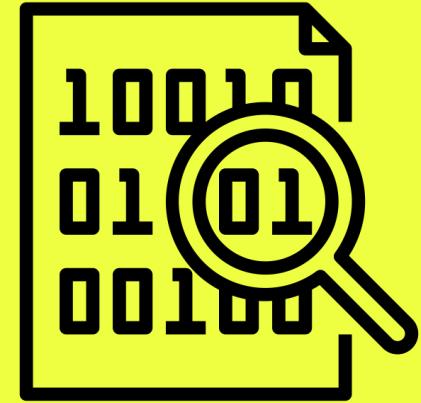
## CHICAGO COMMUNITY AREAS

- **Data Points:**  
77
- **Features:**  
10
- **Missing Values:**  
None

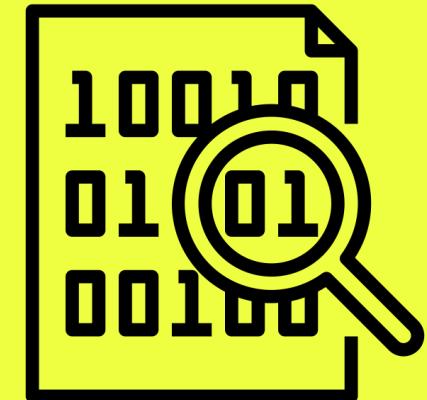
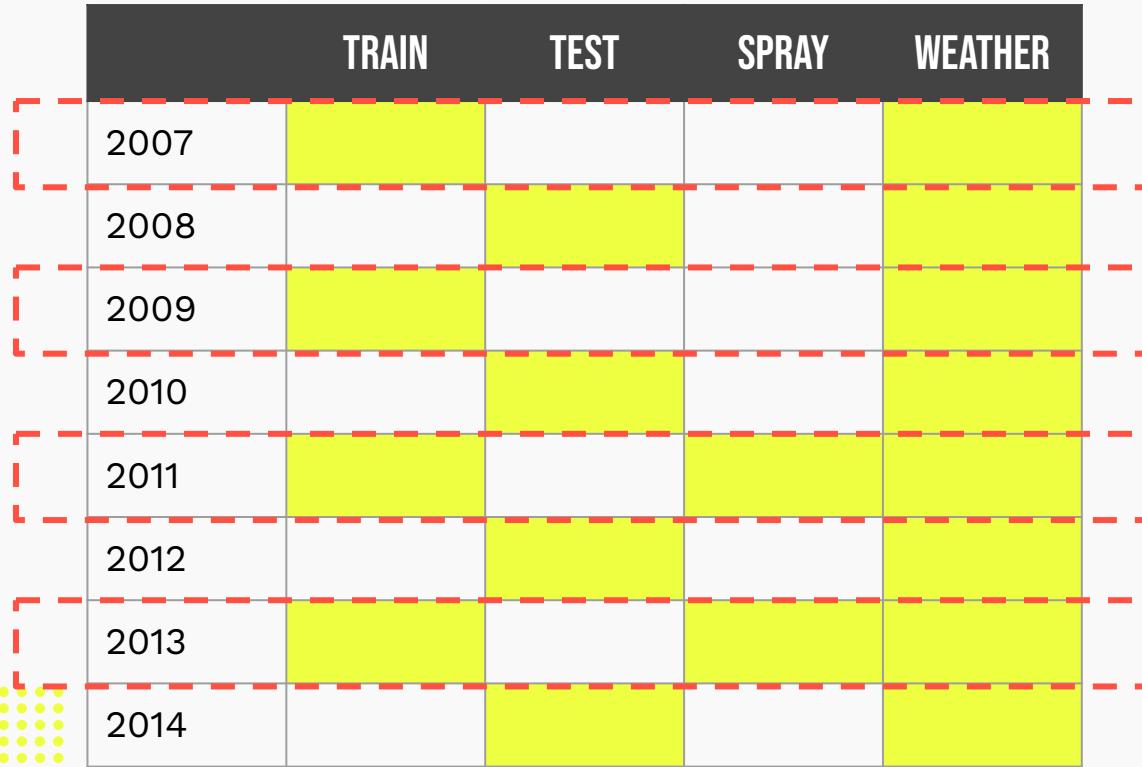


# DATA SUMMARY

	TRAIN	TEST	SPRAY	WEATHER
2007				
2008				
2009				
2010				
2011				
2012				
2013				
2014				



# DATA SUMMARY



# 135,039

mosquitoes collected

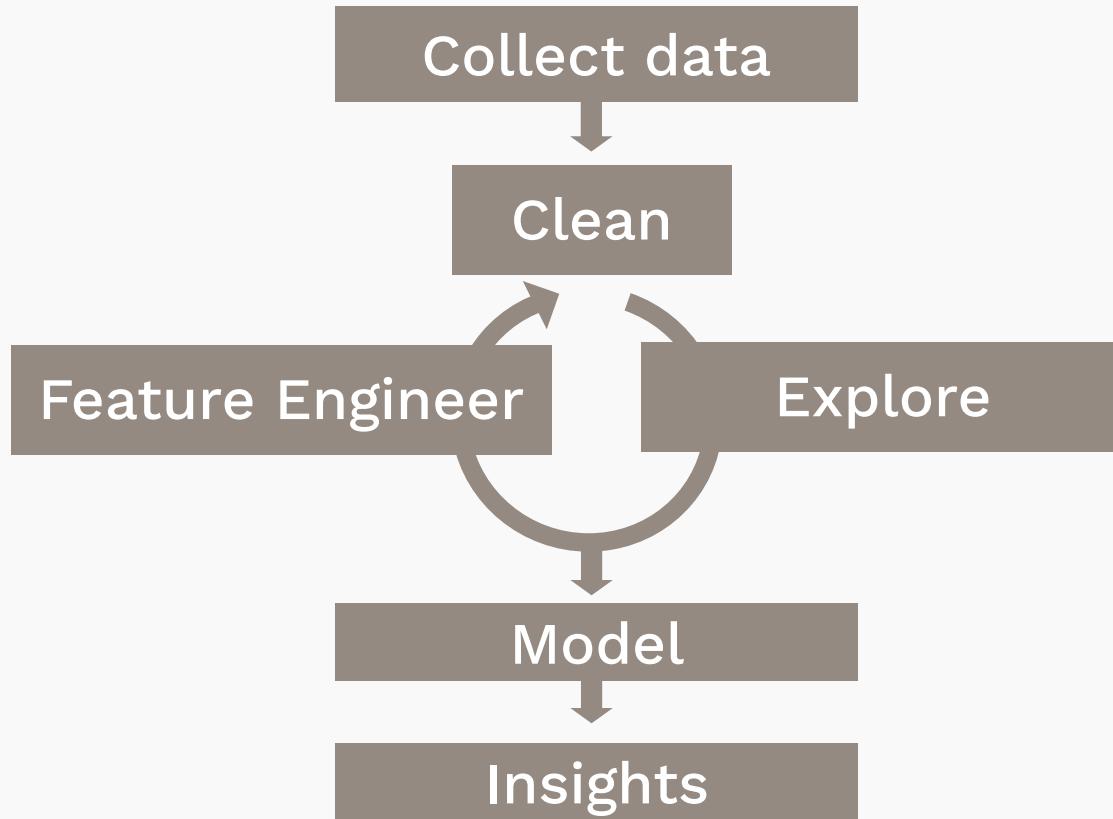


# 14,519

were WNV positive



# DATA WORKFLOW

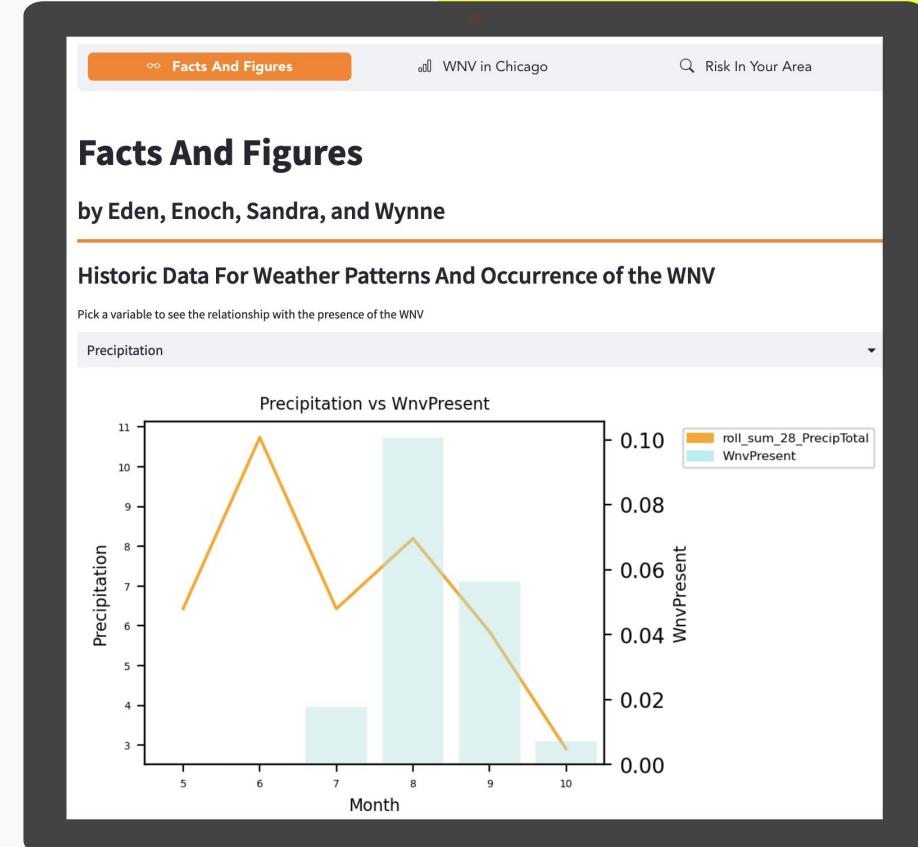




# 03. DATA EXPLORATION & FEATURE ENGINEERING



# WHAT FACTORS ARE ASSOCIATED WITH WNV+ TRAPS?

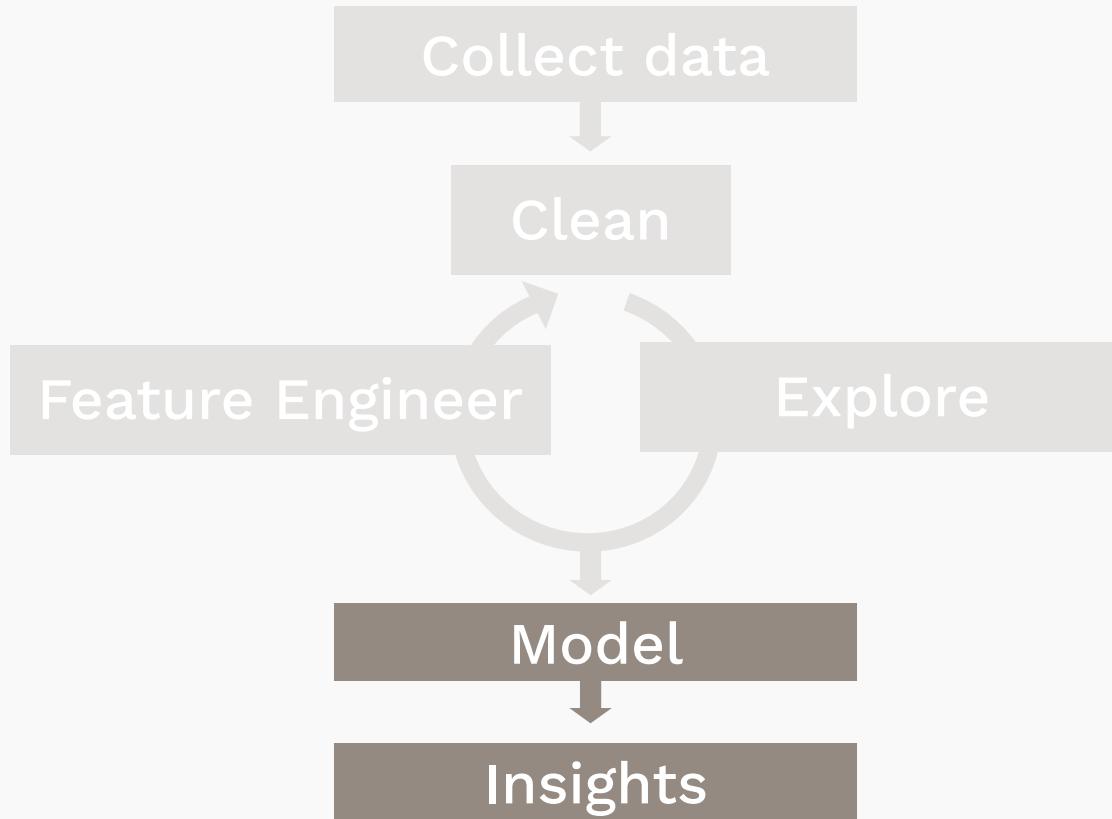




# 04 ▪ PREDICTIVE MODELING



# DATA WORKFLOW



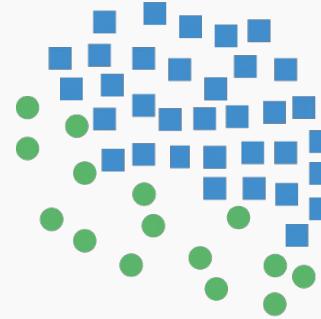
# CLASS IMBALANCE



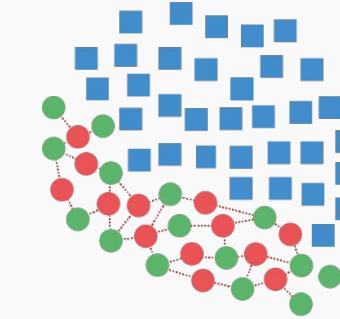
**Model will overtrain on  
negative observations**

# BALANCING CLASSES USING SMOTETOMEK

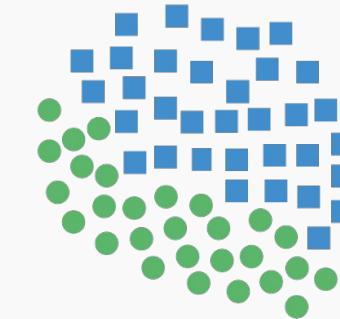
**Smote**



Original Dataset

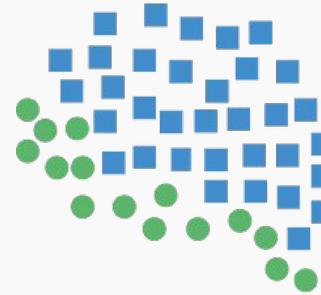


Generating Samples

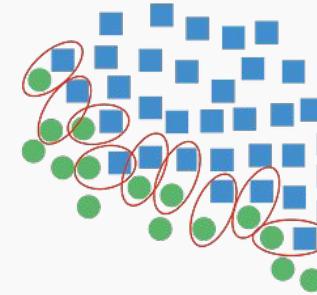


Resampled Dataset

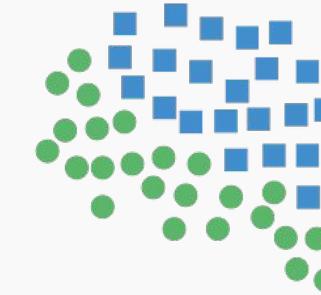
**Tomek  
links**



Original Dataset

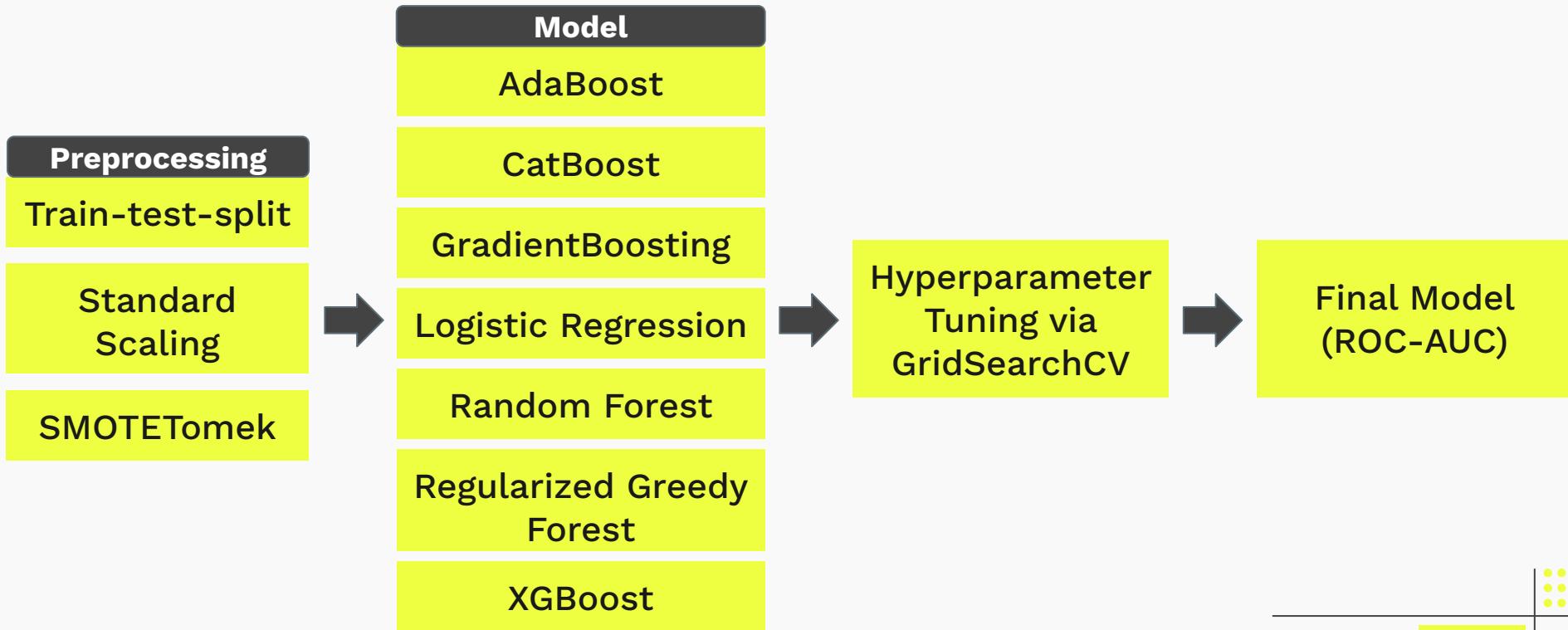


Finding TomekLinks



Resampled Dataset

# MODEL APPROACH

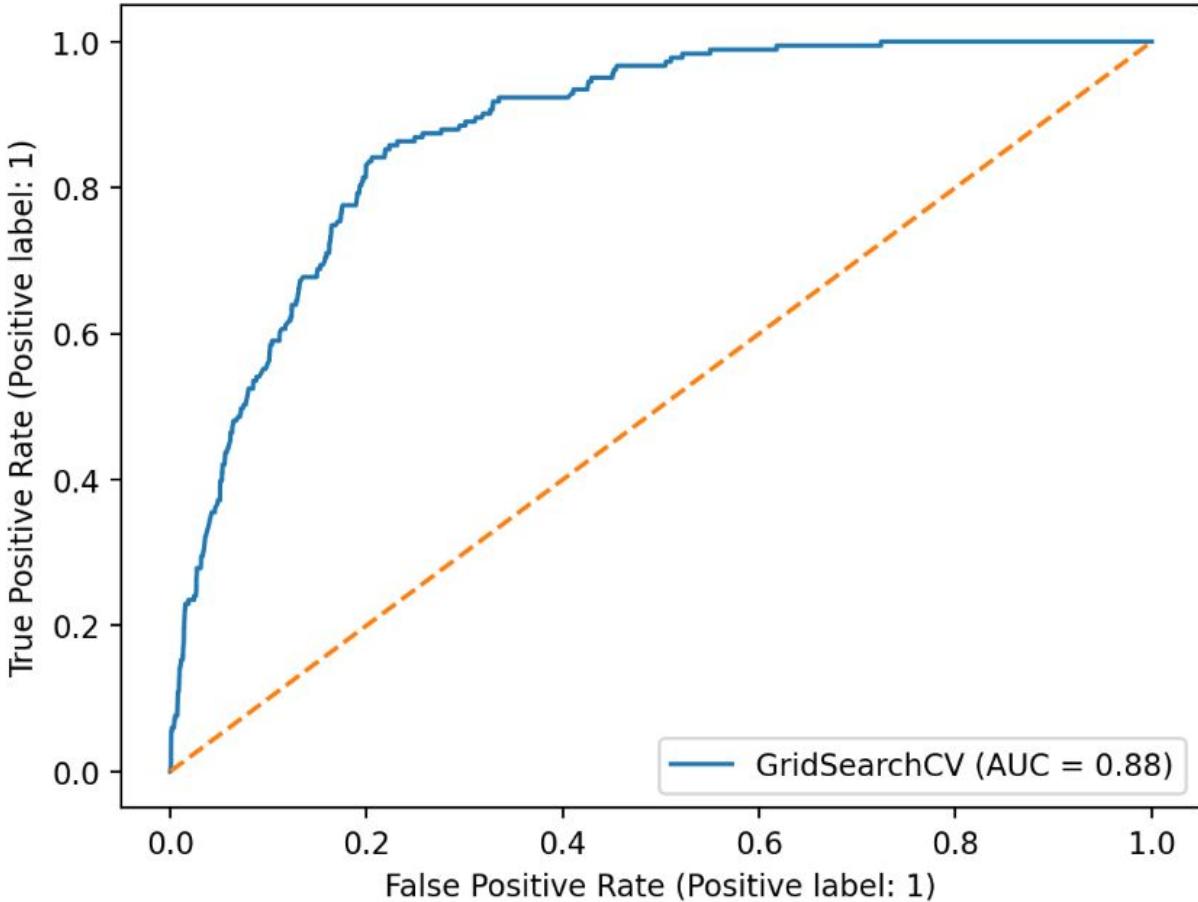


# MODEL PERFORMANCE

	ROC-AUC (TRAIN)	ROC-AUC (TEST)	RECALL (TRAIN)	RECALL (TEST)
RANDOM FOREST (BASELINE)	0.99	0.84	1.0	0.35
ADABOOST	0.85	0.86 (↑ 2.4%)	0.77	0.80 (↑ 229%)
CATBOOST	0.96	0.87 (↑ 3.6%)	0.64	0.34 (↓ 2.9%)
LOGISTIC REGRESSION	0.85	0.86 (↑ 2.4%)	0.80	0.81 (↑ 231%)
GRADIENTBOOSTING	0.91	0.87 (↑ 3.6%)	0.61	0.52 (↑ 149%)
REGULARIZED GREEDY FOREST	0.89	0.88 (↑ 4.8%)	0.68	0.64 (↑ 183%)
XGBOOST (FINAL)	0.86	0.88 (↑ 4.8%)	0.80	0.86 (↑ 246%)

**0.88**

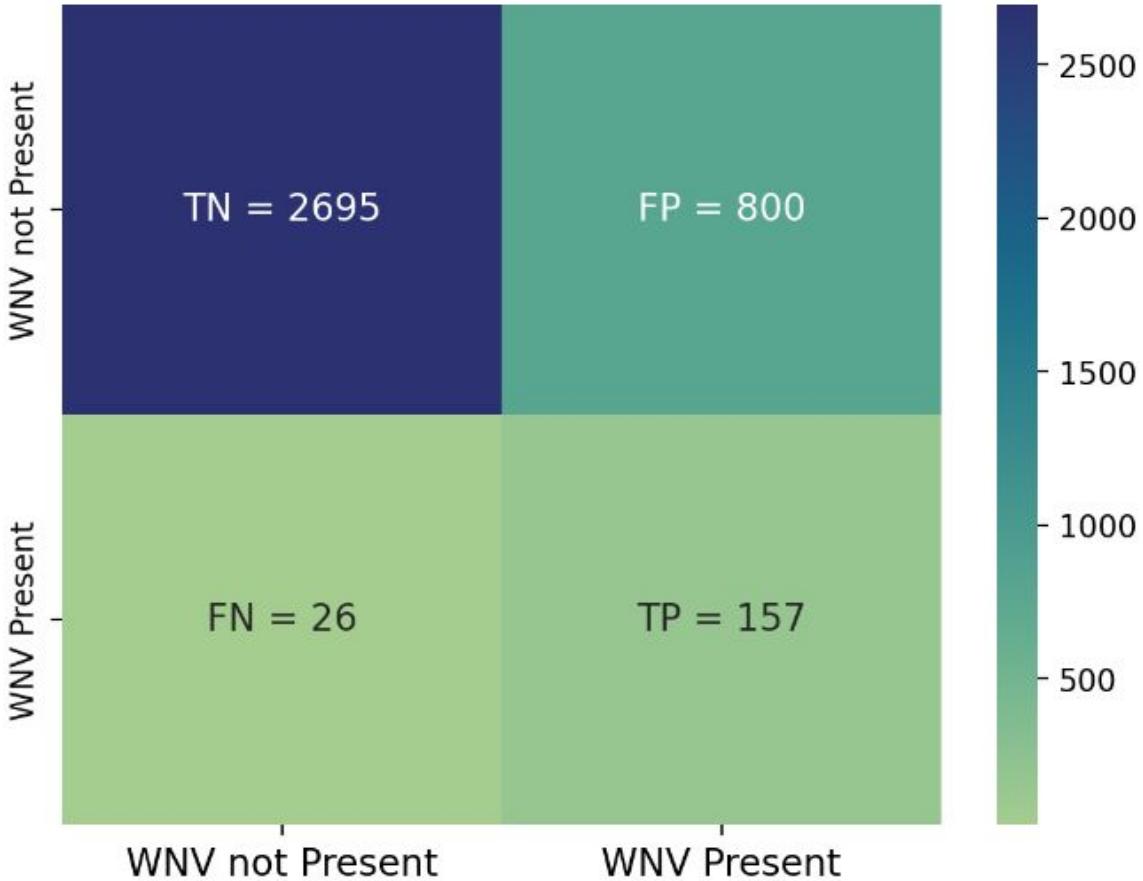
ROC AUC



# 0.86

Recall

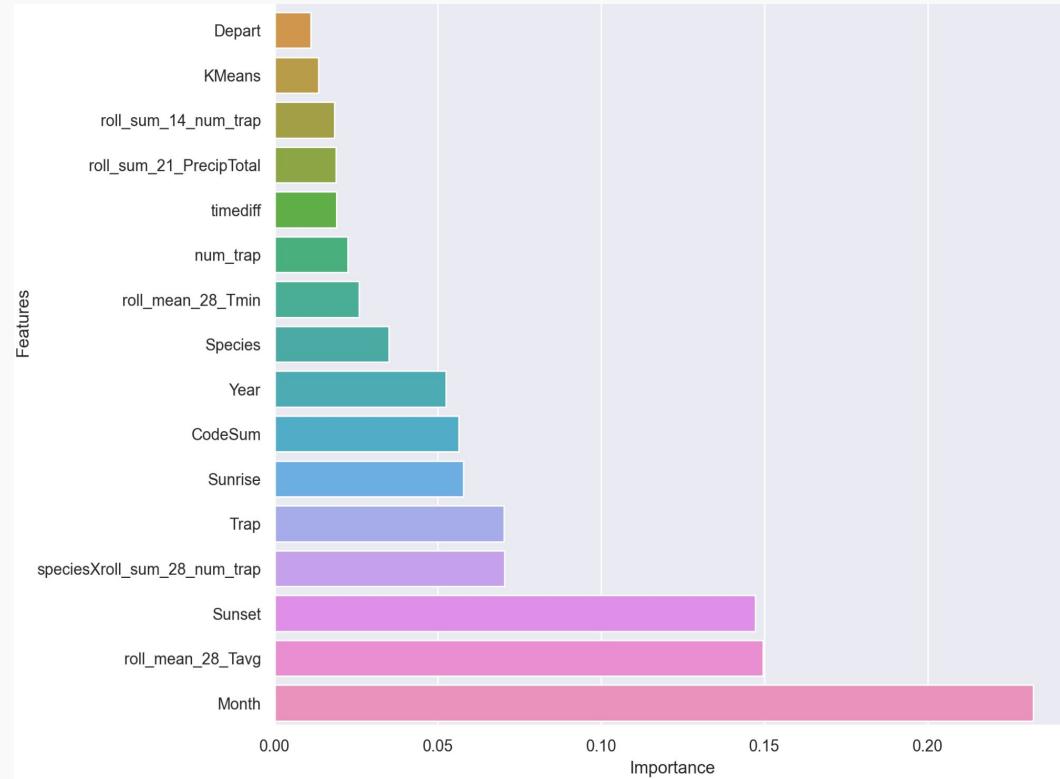
(Low False Negatives)



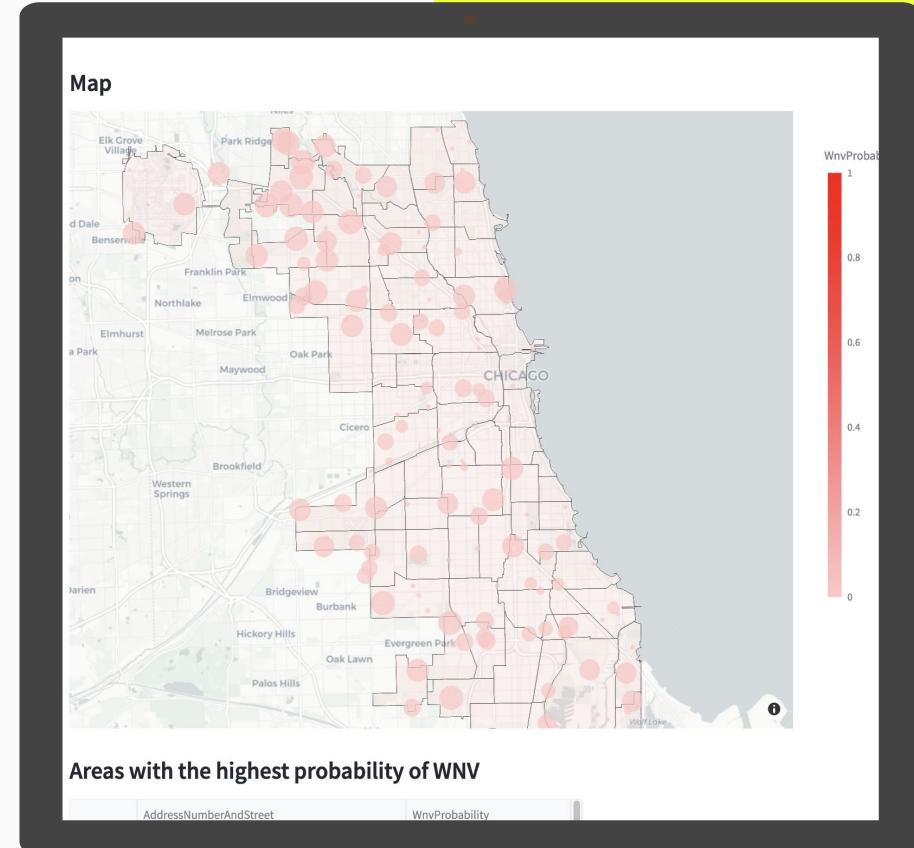
# WHICH FEATURES PREDICT WNV+ THE BEST?

**MONTH**      **TEMPERATURE**  
28-day rolling mean

**SUNSET**      **SPECIES AND NUMBER OF TRAPS**



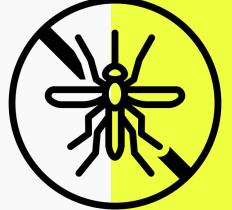
# VISUALISING PREDICTIONS





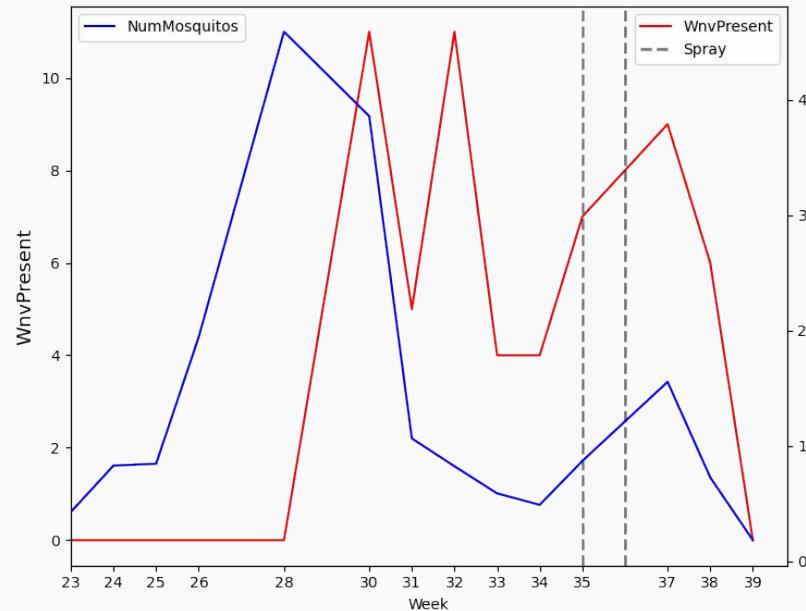
# 05 • COST BENEFIT ANALYSIS



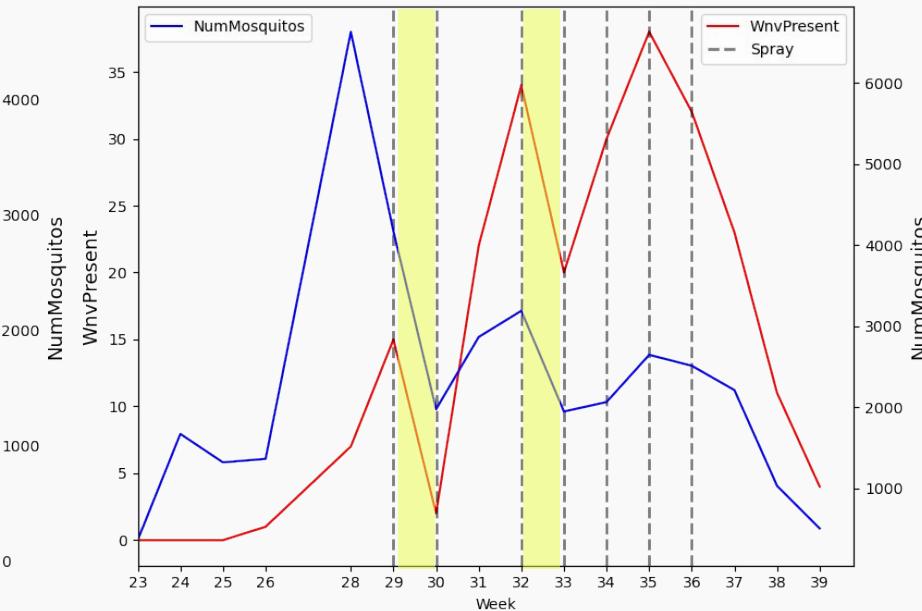


# EFFECTS OF SPRAYING

WnvPresent and NumMosquitos in 2011



WnvPresent and NumMosquitos in 2013



# COST ANALYSIS

(estimated using 2013 as scenario year)

## SPRAY WITH CURRENT APPROACH

(benefit-cost ratio: 87)

COST OF ADULTICIDE &  
LARVICIDE DEPLOYMENT<sup>1</sup>

deployed after WNV+ traps ↑

ACUTE MEDICAL COSTS,  
LONG-TERM CARE COSTS,  
LOST PRODUCTIVITY<sup>2</sup>

— USD\$180K

— USD\$830K

SOCIAL BENEFITS<sup>3</sup>

+ USD\$88mil

## MODEL-DRIVEN SPRAY APPROACH

(BCR: 353)

deployed before WNV+ traps ↑

— USD\$160K

— USD\$89K

+ USD\$88mil

1. Est ≈USD\$6,000 per truck-spray-area, assuming an area would take 5h to spray, based on locations in 2013

2. Staples, J Erin et al. "Initial and long-term costs of patients hospitalized with West Nile virus disease." *The American journal of tropical medicine and hygiene* vol. 90,3 (2014): 402-9. doi:10.4299/ajtmh.13-0206

3. Halasa, Yara A et al. "Quantifying the impact of mosquitoes on quality of life and enjoyment of yard and porch activities in New Jersey." *PLoS one* vol. 9,3 e89221. 6 Mar. 2014, doi:10.1371/journal.pone.0089221



# 06 ▪ RECOMMENDATIONS



# TIERED RESPONSE ON PREDICTED PRIORITY AREAS

Adulticiding



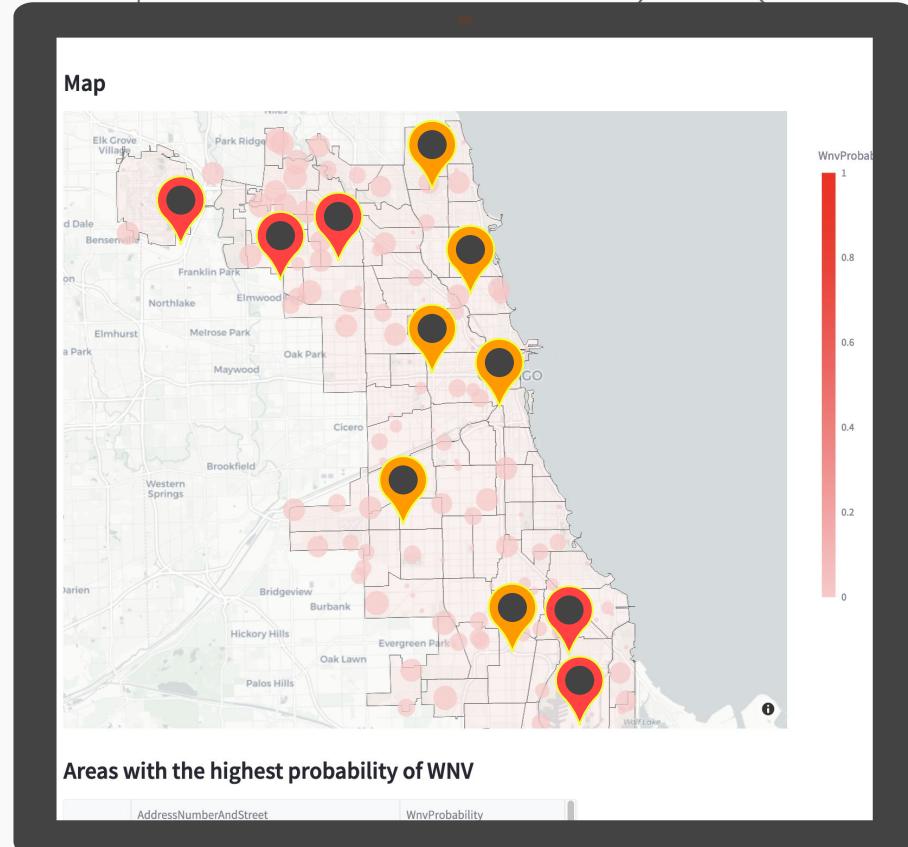
Larviciding



Public education



Public comms



# CONTINUED SURVEILLANCE AND RESEARCH

## Urban landscape features



Ruiz, Marilyn O et al. "Association of West Nile virus illness and urban landscapes in Chicago and Detroit." International journal of health geographics vol. 6 10. 12 Mar. 2007, doi:10.1186/1476-072X-6-10

## Innovative and more effective treatment against adult mosquitoes



New technology  
to knock down mosquitoes  
without using insecticides

kao

# OUR CALL TO ACTION

Today:

Weekly trap surveillance



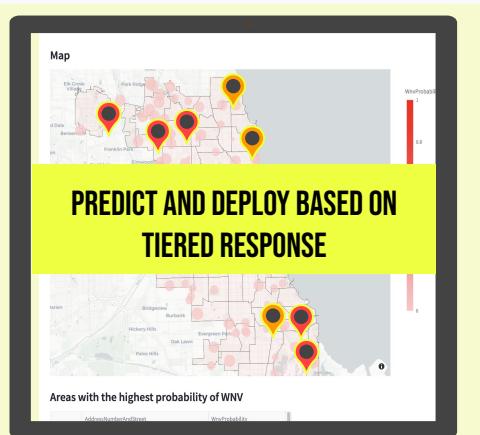
Test for WNV+ traps



Deploy adulticide/larvicide

Our data-driven predictive approach:

Weekly trap & weather surveillance



- ✓ Preemptive
- ✓ Reliable (ROC AUC: 0.88)
- ✓ Resource-effective ( $\downarrow$  costs)

# THANK YOU.

Q&A

CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), and infographics & images by [Freepik](#)

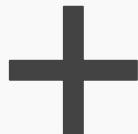
# ANNEX



# TOTAL PRODUCTIVITY

## MARKET PRODUCTIVITY

includes gross annual personal labor earnings adjusted for employer-paid benefits



## NON-MARKET PRODUCTIVITY

includes time devoted to household tasks, caregiving responsibilities, and volunteer services

# ILLINOIS



**1,371**

## total human WNV cases reported in Illinois from 2005 to 2016

906

cases (66%) were from the Chicago region

in Chicago

1

## case in 2009 (lowest)

225

## **cases in 2012 (highest)**

22

## **fatalities in 2012 (highest)**