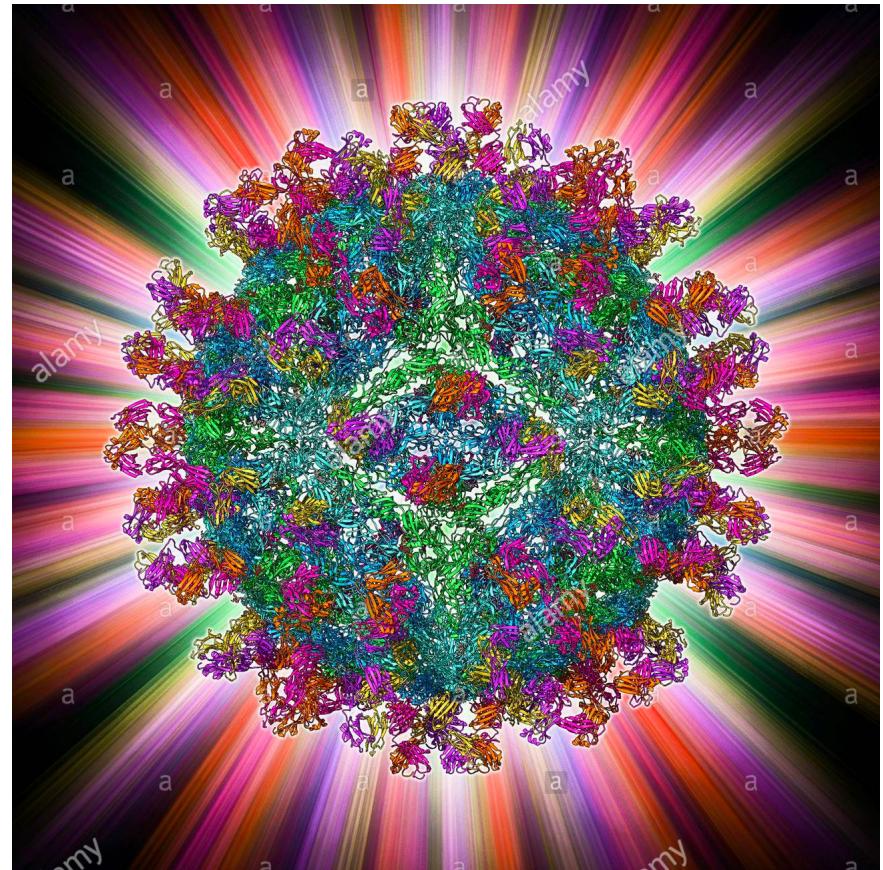


West Nile Ballers

Friday, August 10th 2017



Agenda

Overview	3
Approach	4
Model Specification	6
Findings	7
Recommendations	10

Overview

Our team modeled the presence of West Nile to understand where it might develop next.

THE CHALLENGE

Predict where the City of Chicago should spray to prevent West Nile Outbreaks through an efficient and cost-effective approach

THE APPROACH

EXPLORATION

Conduct basic analysis to visualize trends, find missing information, and understand the nature of the dataset

DATA CLEANING

Reformat data, complete/drop missing data, and structure data in preparation for further analysis.

FEATURE ENGINEERING

Generate new features based on existing columns to reframe our understanding and aid our analysis

MODELING

Create models of the data to predict outcomes and evaluate model performance to incrementally improve predictions

ANALYSIS

Review the outputs of models and apply models to predict results

INTERPRETATION

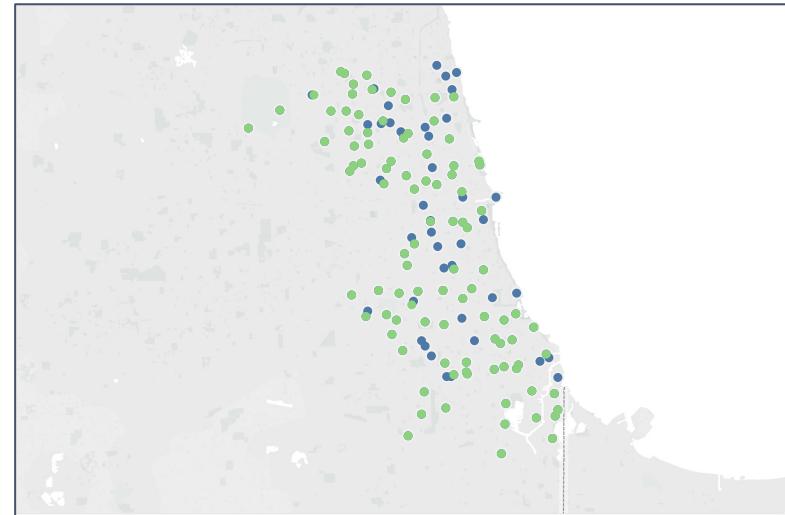
Interpret the results of model predictions to make determinations and recommendations

Exploration

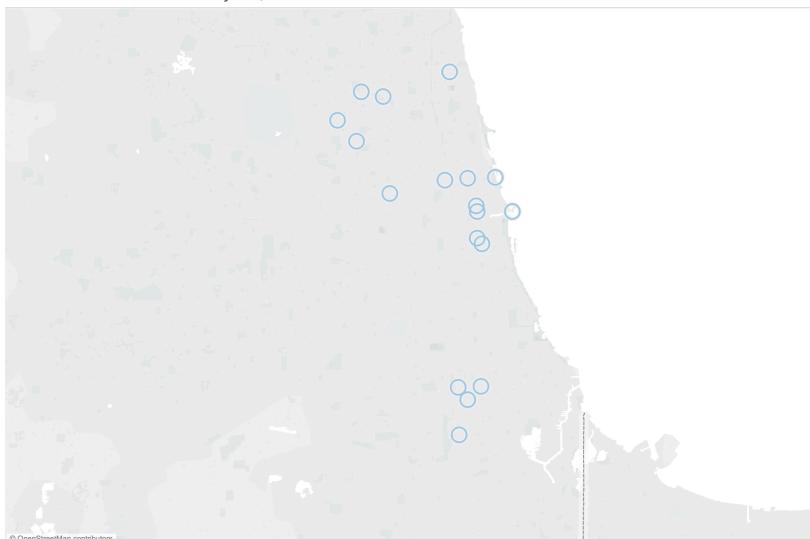
We uncovered some high level trends to help inform our approach and analysis

Data

- **West Nile Testing for 2007, 2009, 2011, 2013**
 - Species
 - WNV status
 - Trap and Location
 - Date
- **Spray day for 2011 and 2013**
- **Weather data for 2007 – 2014**
 - Temperature
 - Rainfall
 - Daylight



West Nile Presence - May 29, 2007



All Traps for 2007, 2009, 2011, 2013

All Traps for 2007, 2009, 2011, 2013

Weather is Important

- Previous temperatures
- Previous rainfalls
- Daylight hours
- Temperature above average

Assumptions and Initial Steps

Based on our exploratory analysis, we made a few assumptions and Features

ASSUMPTIONS

- We used a **training dataset** for 2007, 2009, 2011, 2013 and a **test dataset** for 2008, 2010, 2012 and 2014
- **95% of observed mosquitos test negative for West Nile**, so we need to evaluate our models differently – AUC score
- **Spray data should not factor** into our predictions
- We limited our analysis to only **features that are available for both the training and test data**

FEATURES

- **Weather**
 - 7, 14, and 30 day past averages/totals for
 - Temperature
 - Rainfall
 - Departure from average temperature
 - Daylight Hours
 - Dew Point/Humidity
 - Windspeed
- **Training Data**
 - Species factorization
 - Prior large mosquito populations
 - Trap factorization

Model Specification

Our team tested multiple models to inform our approach

RANDOM FOREST

- Gridsearched parameters
- Included all available features
- Selected features by importance



71%

LOGISTIC REGRESSION

- Used all features to try understand feature importance
- Looked at odds ratios



N/A

XGBOOST

- Used important features from random forest/logistic regression
- Used generated features
- Tuned parameters

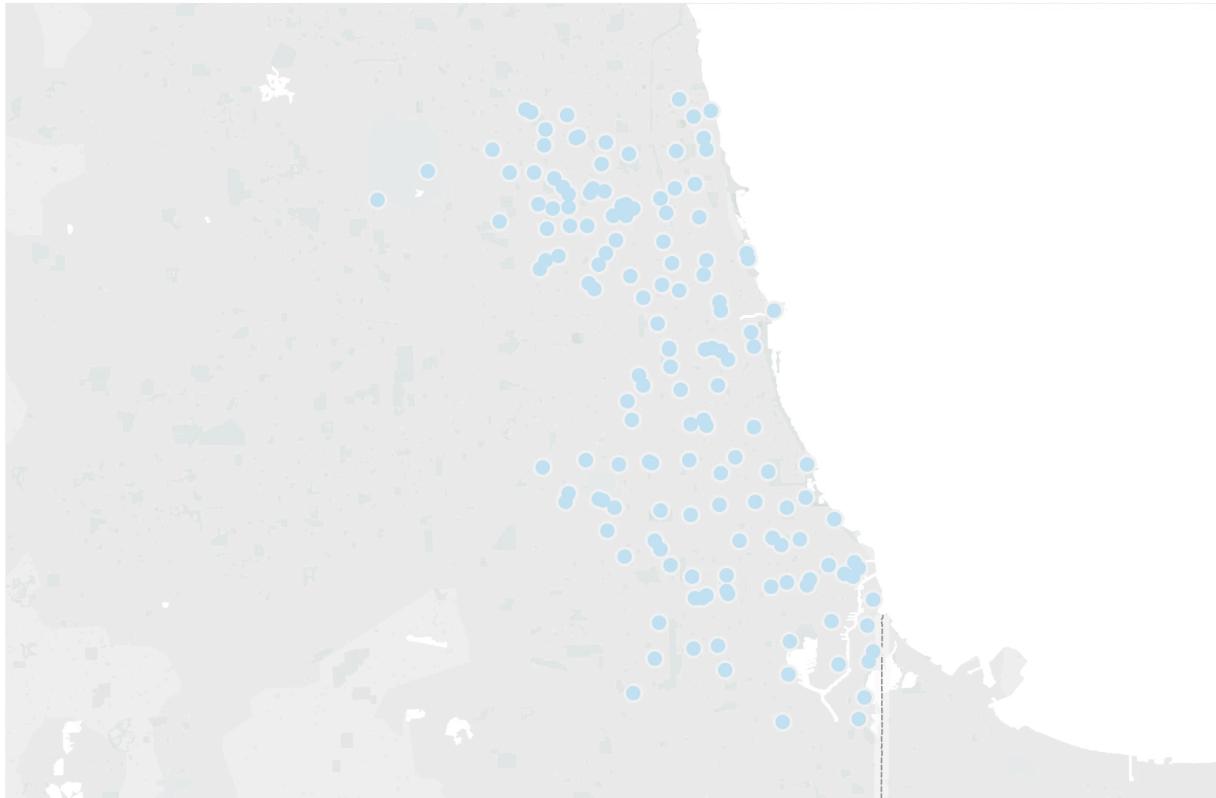


75%

Findings

Based on the performance of our model, our team identified locations susceptible to WNV

Predicted WNV Presence - June 11, 2008



Highest Likelihood WNV Locations

- Locations in the Northwest of Chicago
- Southern Chicago

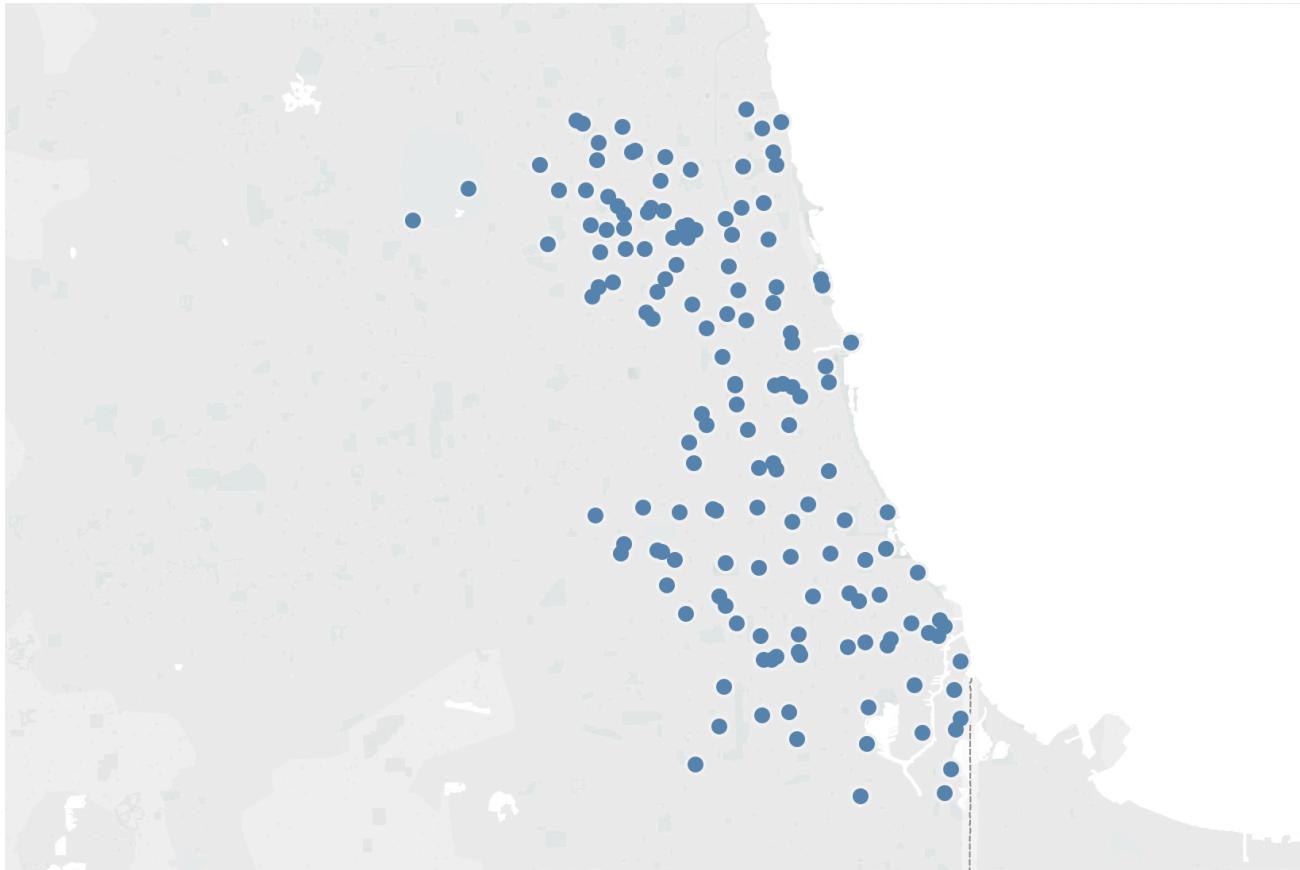
Factors influencing likelihood of WNV

- Hotter months
- Later in the summer
- Higher presence of the Culex Pipiens and Culex Restuans

What does it mean?

Although locations may show West Nile, it might not be likely to break out

Predicted WNV Presence - June 17, 2008



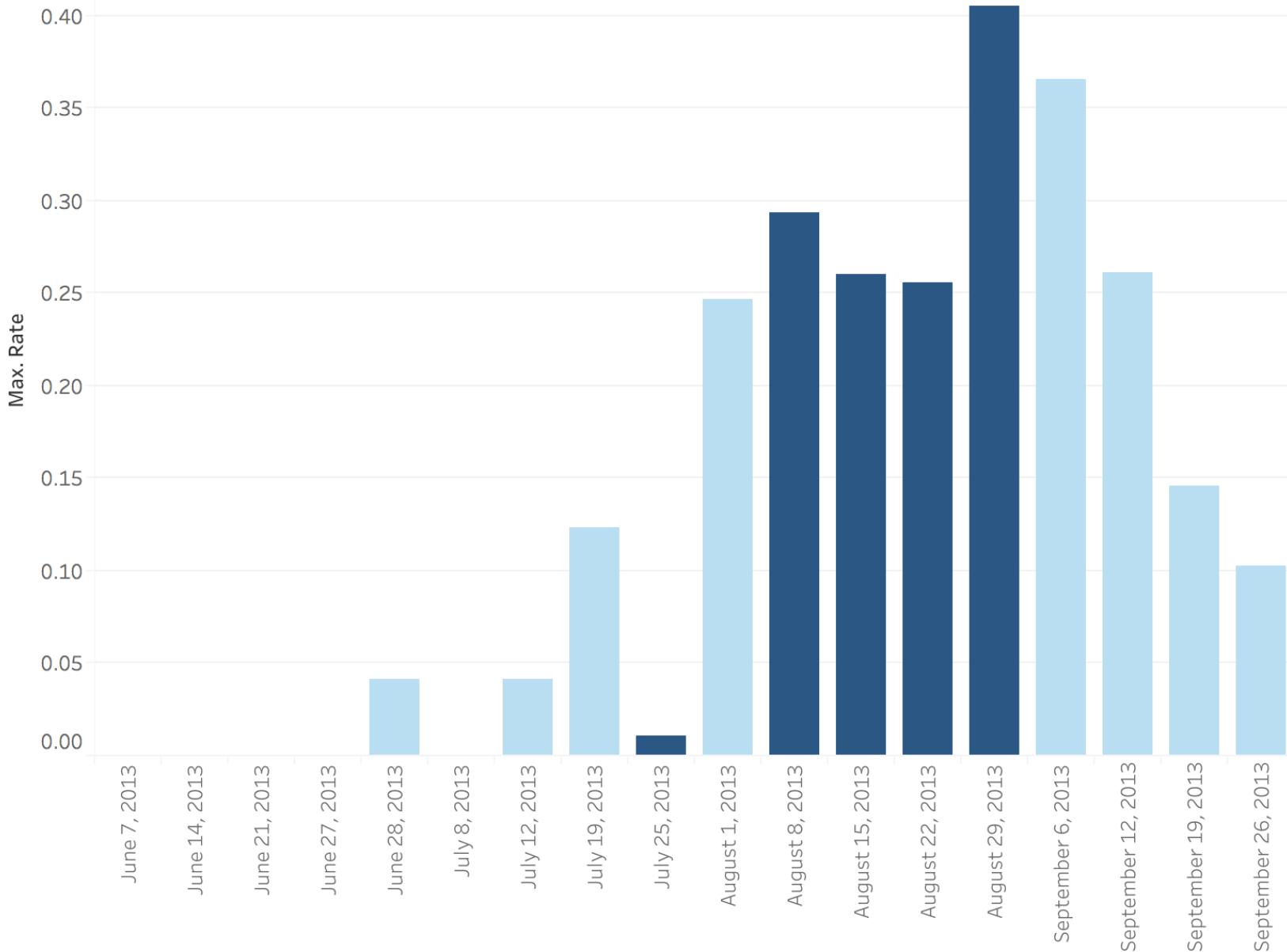
Deciding When to Spray

WNV Infection Rate of mosquitoes can be used to inform decisions to spray

Integrated Pest Management

- Prevention
 - Public information
 - Personal protection methods (avoidance, repellants, clothing)
 - Source reduction of mosquito breeding sites
 - Control of mosquito larvae
- Environmental Surveillance
 - Collect data (traps, field surveys, WNV positive birds, human cases)
 - Provide indicators of the threat to humans and identify geographic areas of high risk
 - Support decision for intervention
 - Monitor effectiveness
- Based on 2013 data, Minimum Infection Rate is approximately 0.25:
 - # of WNV Positive Pools/Total # of Mosquitoes tested

2013 Minimum Infection Rate of WNV



Costs and Benefits

Choosing locations to spray depends on the costs and benefits

Cost Breakdown

- Spray Costs
 - >= \$60 per Linear Mile
 - Estimate to spray all of Chicago (227.34 Sq Miles) - \$818,424
 - 2013 spray costs estimated at \$520,051
- Hospitalization Costs
 - Median cost Hospitalized for Meningitis or Fever - \$7,500
 - Median cost Hospitalized for Acute Flaccid Paralysis- \$25,000
 - Median Work Days Missed by Hospitalized Patient – 42
 - 2013 Medical and Hospitalization- \$575,000 to \$1.5 Million
 - 60 reported Cases, 7 Deaths associated with WNV

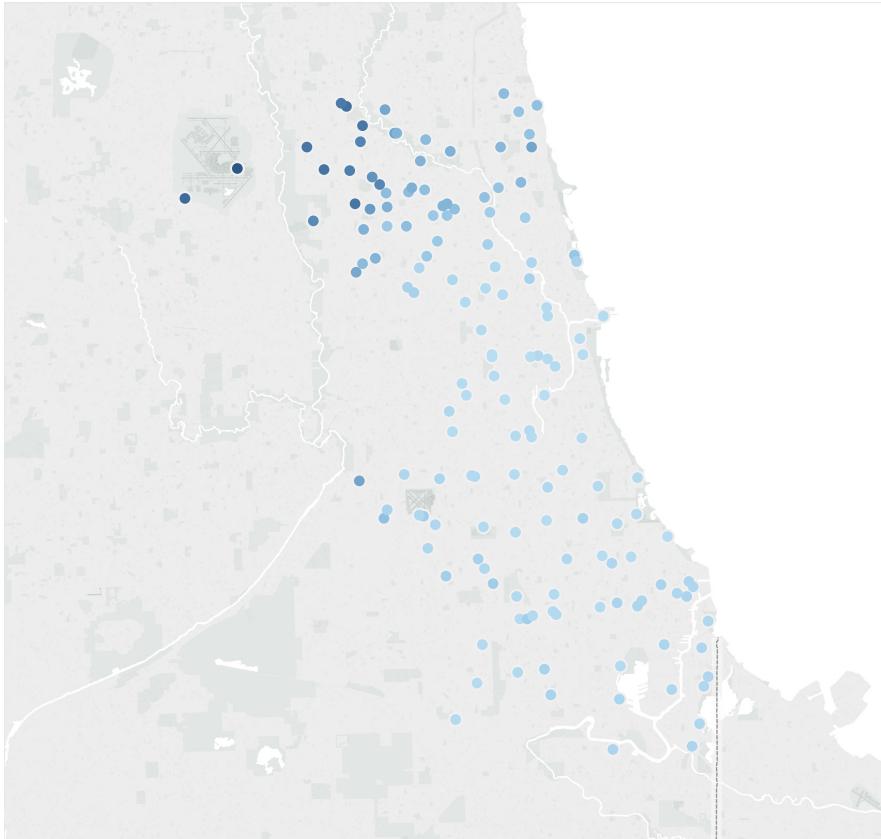
Benefits Breakdown

- WNV Outbreak After 14 Days, 0 New Human Cases.
- Less Expensive to Spray vs Possible Medical Costs
- Risk of illness from West Nile virus is greater than the small risk associated with adulticiding.

Recommendations

Based on our analysis, we recommend spraying in these areas

Predicted WNV Presence 2008, 2010, 2012, 2014



RATIONALE

- West Nile seems likely to become **more prevalent later in the summer**
- Spraying should take place about **14 days and 30 days in advance of predicted WNV probability**
- An **infection rate of mosquitos above ~25%** indicates a likelihood that spraying should take place
- **Above average temperatures can lead to higher probability of WNV, 14 days and 30 days in advance**

NEXT STEPS

- Include data on Human Infection rates to understand how **WNV presence increases the likelihood of infections in humans**
- Include more robust weather forecasting metrics to help predict WNV

Questions

Appendix