Predicting West Nile Virus

With Machine Learning

•••

Problem Statement:

How do certain weather events affect the presence of West Nile virus in the Chicago area? Can this problem effectively be tackled with Machine Learning?

The Data

Train/Test- GIS

- Trap
- Location
- Date
- Num Mosquitos
- WNV Present

Spray Data

- 2011-2013

Weather Data

- 2007-2014

Understanding the problem

Species

Not all mosquito species carry West Nile.

In Chicago these species are the vectors for disease.

- Culex Pipiens
- Culex Tarsalis
- Culex Restauns

Climate & Conditions

Certain conditions contribute to increased mosquito vector populations.

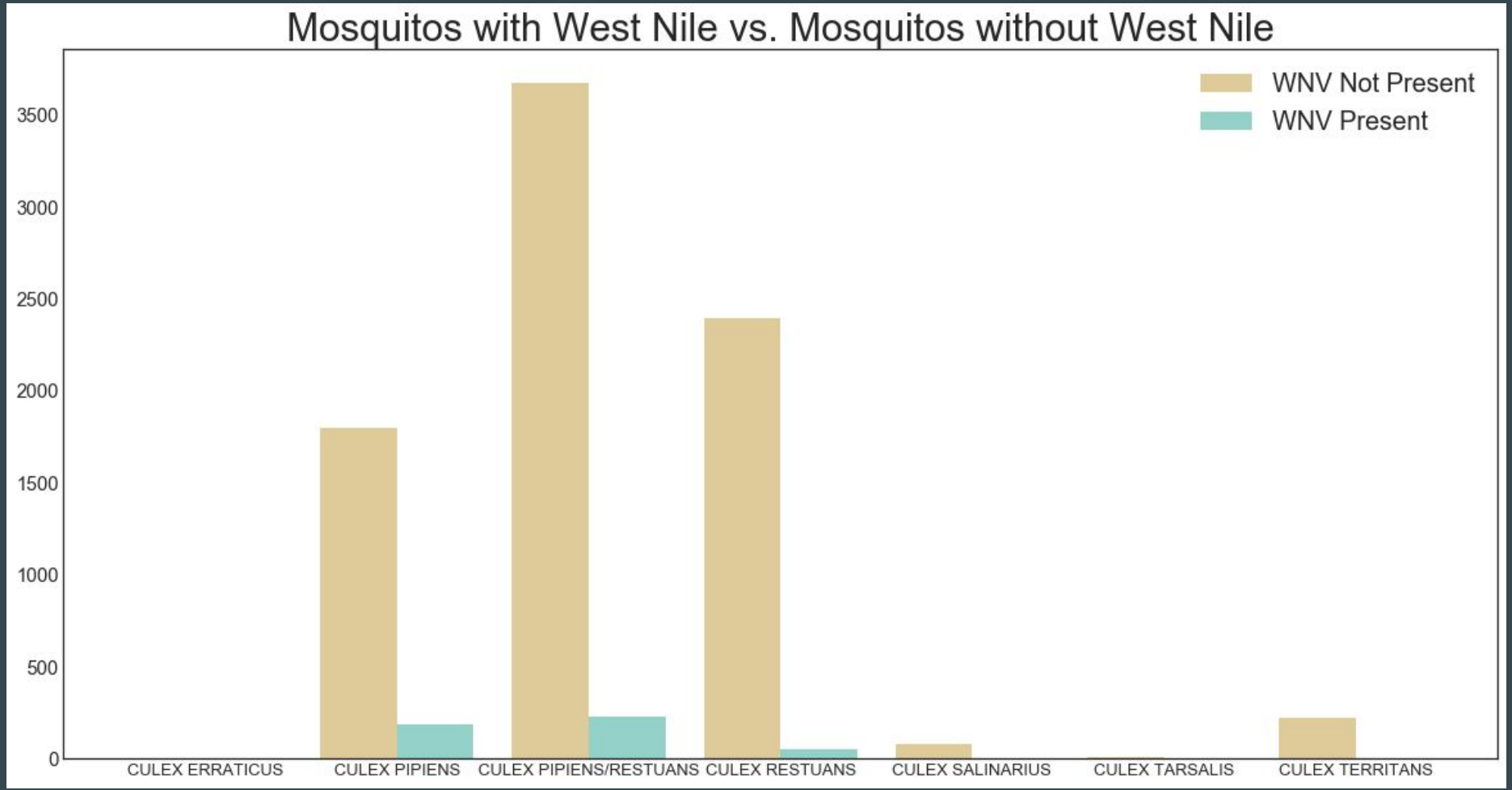
- Stagnant water
- Temperatures above 70 degrees (summer)
- Forests, marshes, tall grass, and weeds

Cost

West Nile costs the US ~56 million a year.. This takes into account long term medical treatment and lost productivity.

Answering the problem

- If we can create a model that is comprised of weather features and predicts the presence of west nile…
- We can infer from the model's resulting decisions what weather features are contributing to the presence of West Nile

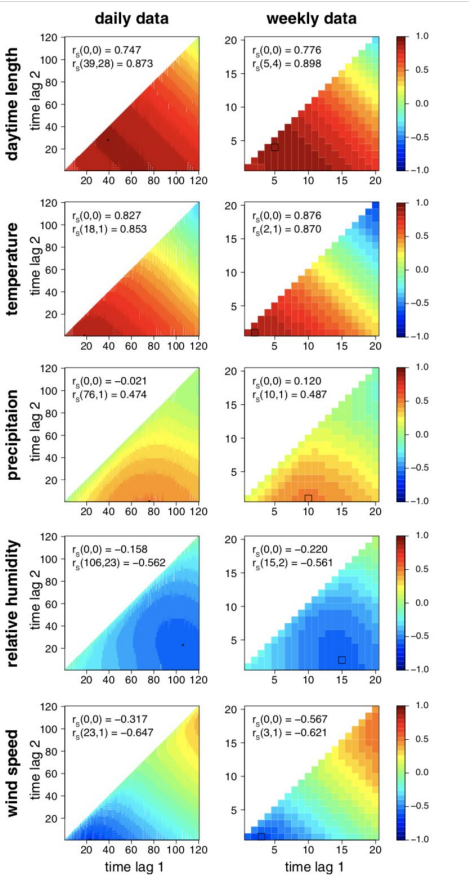

Mosquitos with West Nile vs. Mosquitos without West Nile
WNV Not Present
WNV Present
3500
3000
2500
2000
1500
1000
500
0
CULEX ERRATICUS
CULEX PIPIENS
CULEX PIPIENS/RESTUANS
CULEX RESTUANS
CULEX SALINARIUS
CULEX TARSALIS
CULEX TERRITANS

Methodology:

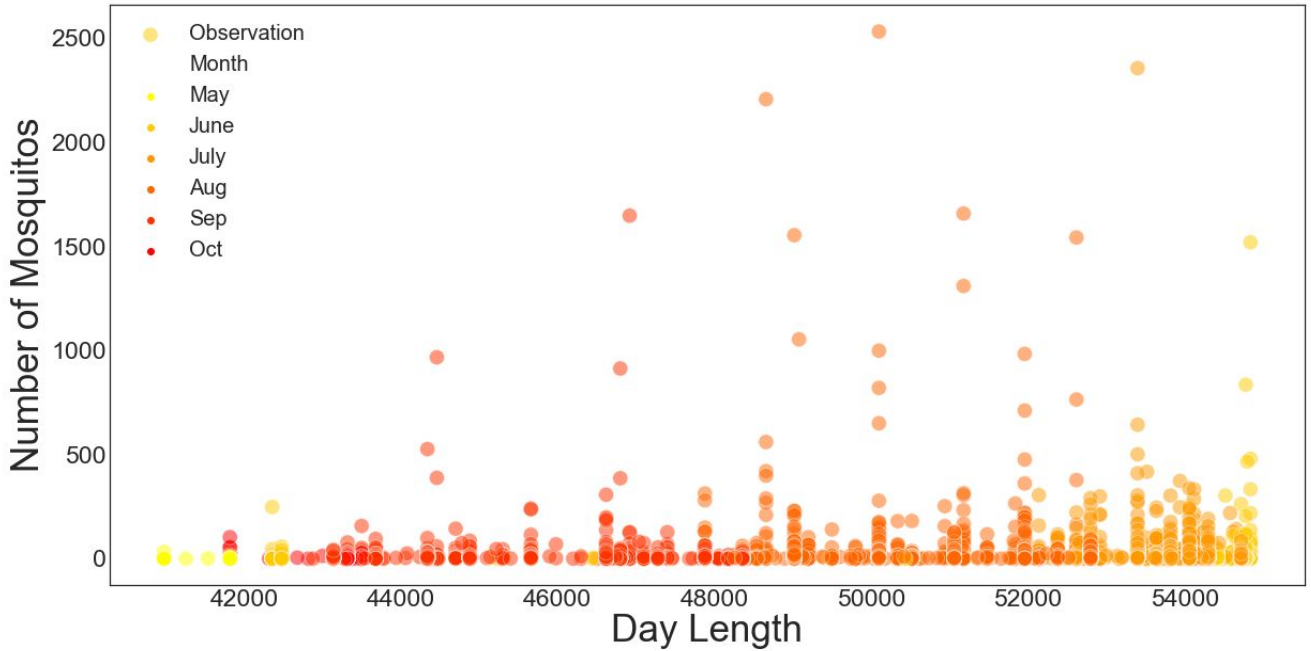
- Examining correlation
  - Through research and graphing
- Feature Engineering
- Modeling

Our Research

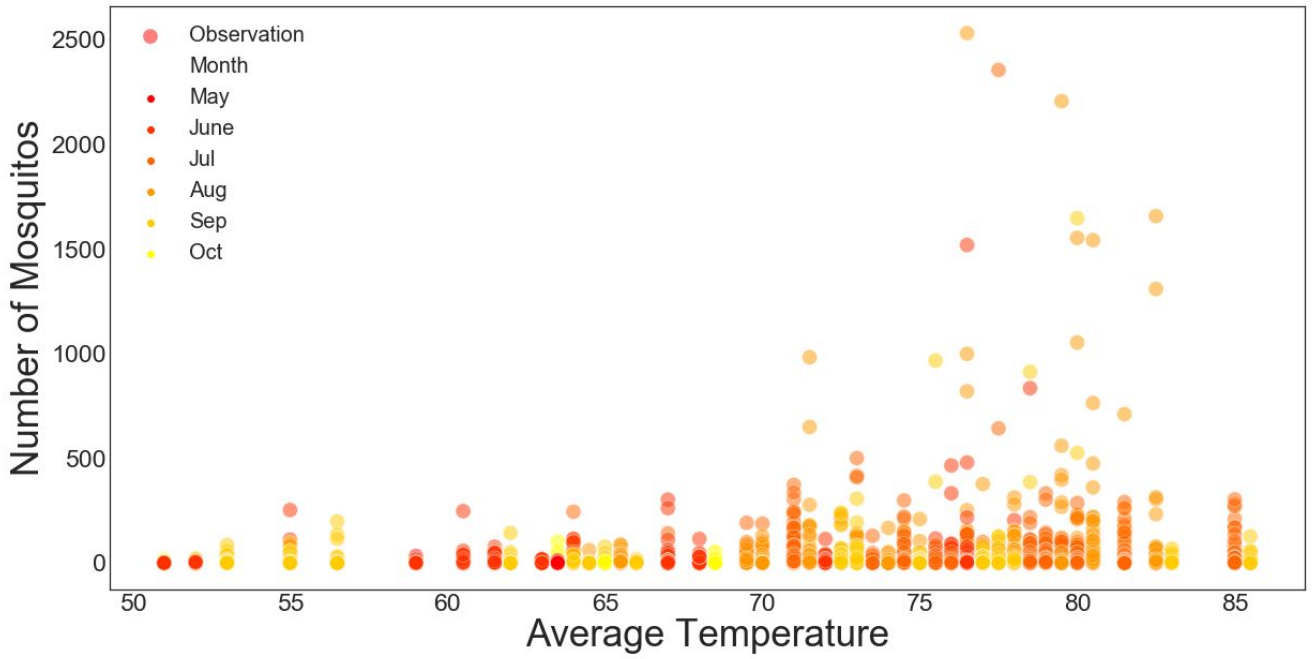
Bakran-Lebl, Karin & Brugger, Katharina & Rubel, Franz. (2013). Predicting Culex pipiens/restuans population dynamics by interval lagged weather data. Parasites & vectors. 6. 129. 10.1186/1756-3305-6-129.



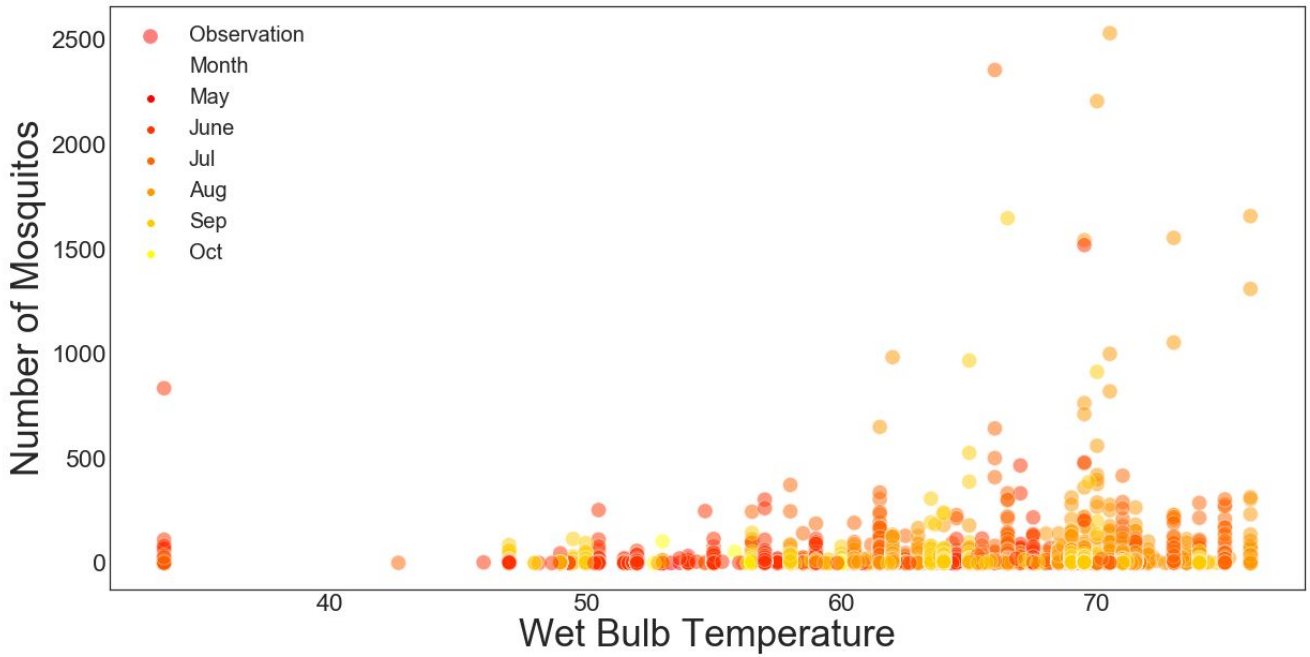
Day Length


Observation
Month
May
June
July
Aug
Sep
Oct
Number of Mosquitos
2500
2000
1500
1000
500
0
42000
44000
46000
48000
50000
52000
54000
Day Length

Average Temperature



Wet Bulb


Observation
Month
May
June
Jul
Aug
Sep
Oct
2500
2000
1500
1000
500
0
Number of Mosquitos
40
50
60
70
Wet Bulb Temperature

Optimal shifted values:

Feature	Rolling Mean	Shifted Value
Day Length	3	28
Temp Avg	3	14
Temp Max	3	14
Temp Min	3	14
Departure	3	14
Heat	3	14
Cool	1	14
Result Speed	1	21
Result Dir	1	21
Wet Bulb	3	14

Modeling

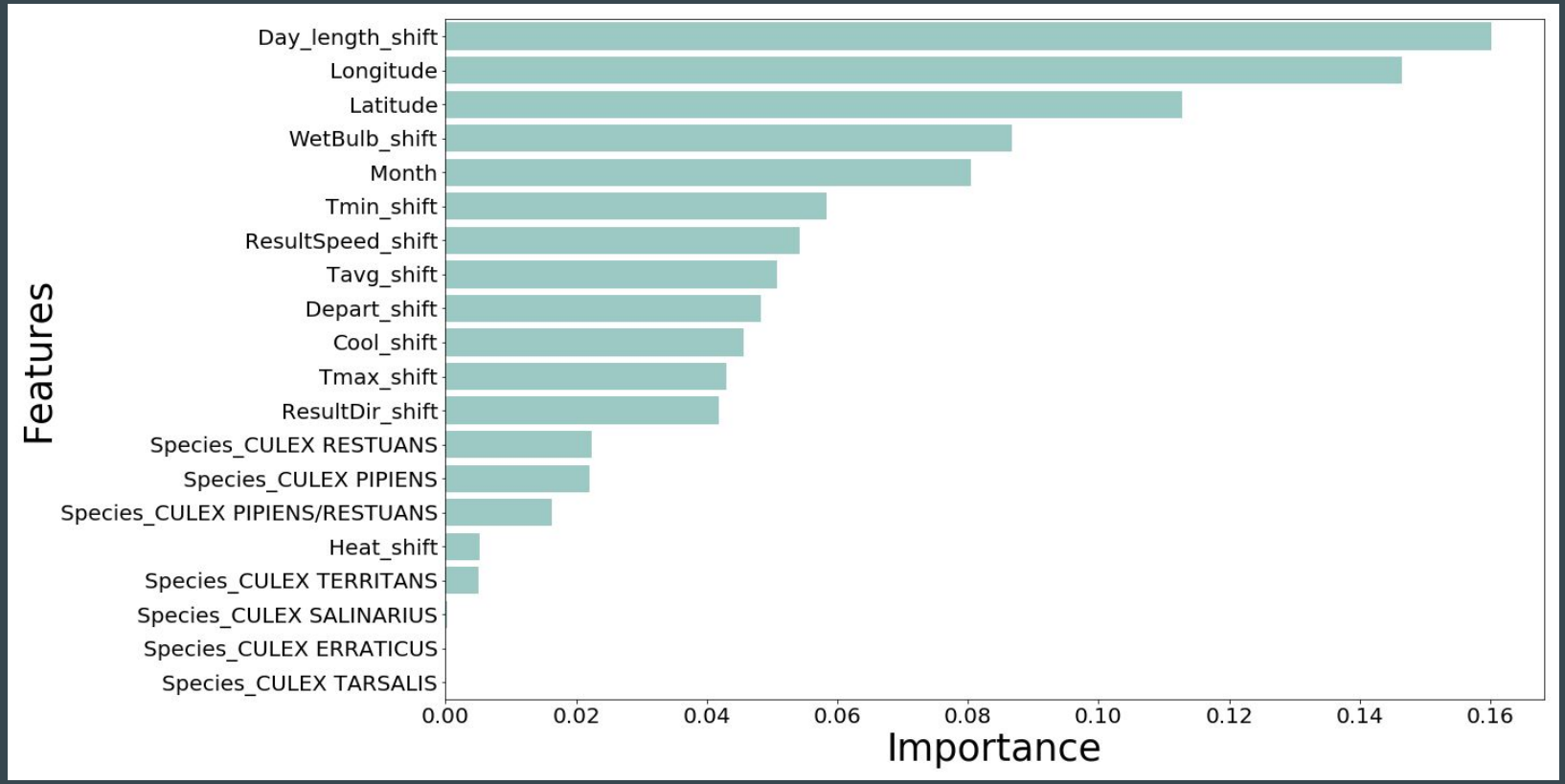
- Attempted several variants
- Logistic Regression scored 74.2% on Kaggle
- Random Forest scored 70%

Chosen Model: Random Forest

Advantages:

1. Zero assumptions about the spread of data
2. Easy to interpret
3. Generalizability

Random Forest - Feature Importance



Cost Benefit Analysis

- 2005 West Nile outbreak in Sacramento, California.
- 163 human cases
- Estimated economic impact: $2.28 million
- Vector control cost: $701,790
- Only 15 cases of West Nile would need to be prevented in order for the emergency spray to be cost effective

- 2002 West Nile outbreak in United States
- 4,156 cases reported, 329 of those were in Louisiana
- Estimated economic impact in Louisiana : $20.1 million (June - February)
- Increased aerial and ground spray lead to an 86% decrease in West Nile carrying species

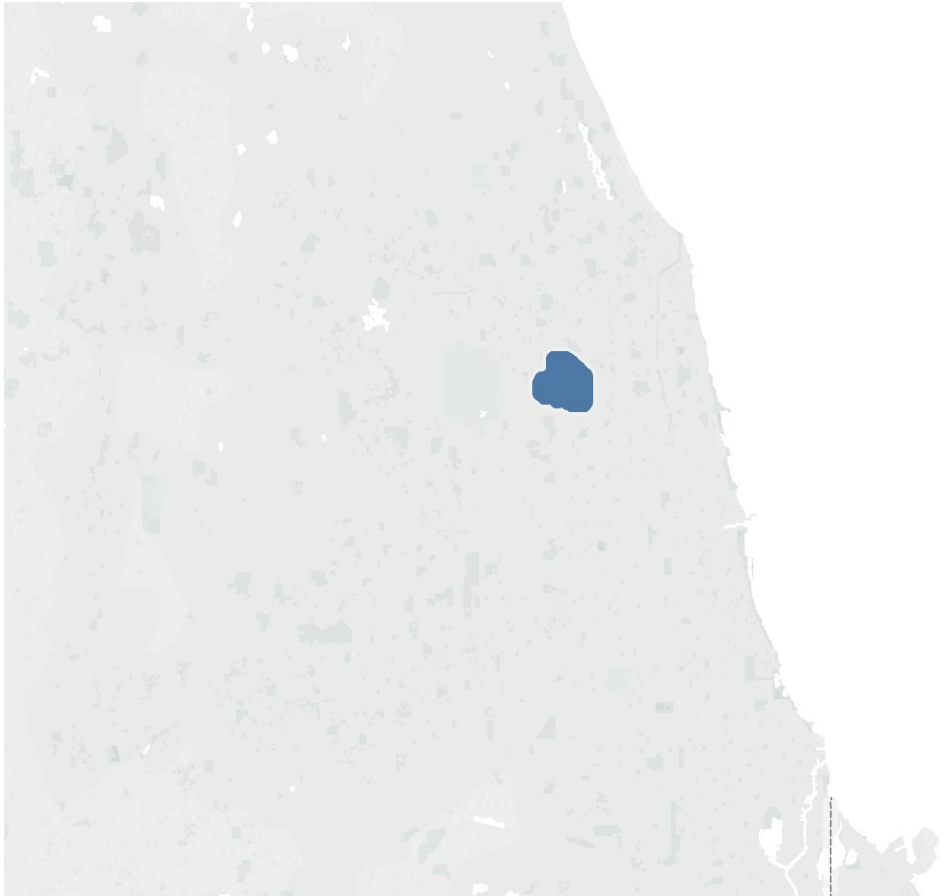
Cost Benefit Analysis: Chicago

- $5.3 million spent on Mosquito vector control in 2015
- Chicago population: 2.7 million
    - Sacramento county population: 1.5 million
    - Louisiana population: 4.68 million

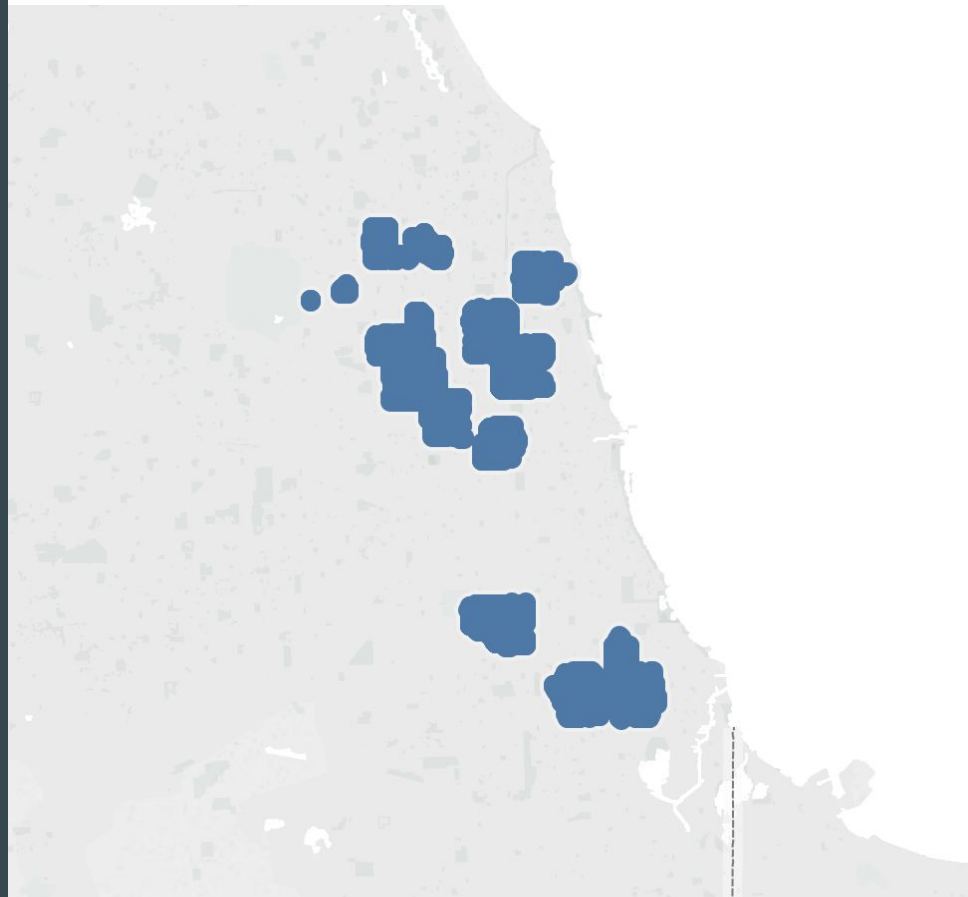
- If the 3:1 ratio holds than economic impact would be around $15 million

Chicago 2011

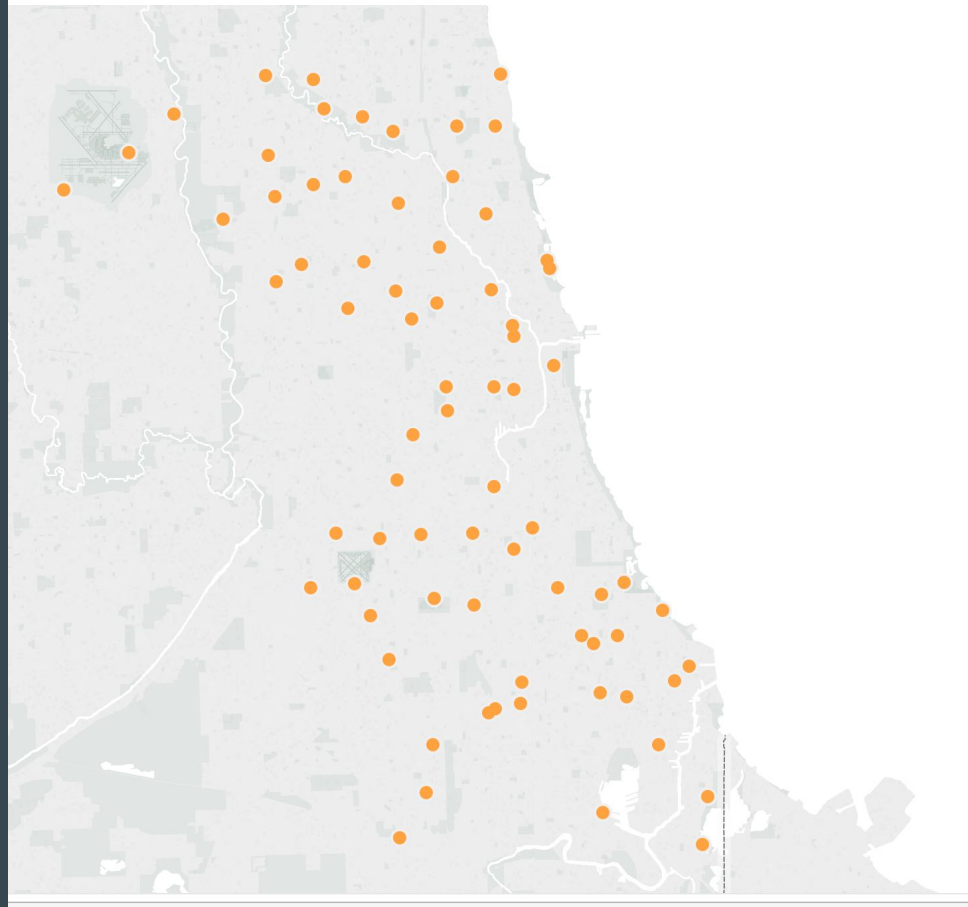
Spray Density



Chicago 2013



Chicago 2013



Our Predictions