

# CHICAGO WEST NILE VIRUS CONTROL PLAN

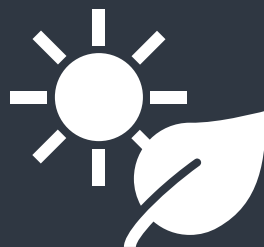


WESLEY  
RUSSELL  
SATHYA  
BOON JUN

# WHAT IS WEST NILE VIRUS (WNV)



Leading cause of  
mosquito borne  
disease in USA



Summer  
through fall



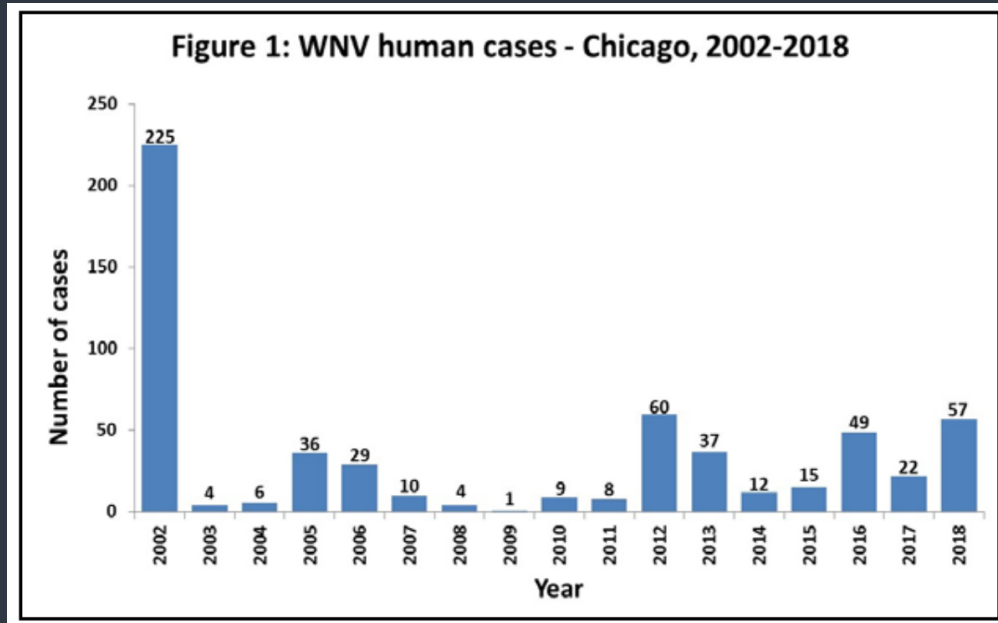
No vaccine,  
No medications

**1 IN 5**

Developed symptoms

## INTRODUCTION

# CHICAGO



Source: Chicago Department of Public Health (CDPH)

- First Wnv case: 225 in 2002
- Implemented city wide surveillance & mosquito control measures
- Continues to have one of the most robust mosquito control program in the US

# INTRODUCTION

**Predicting The Presence Of Wnv For Coming  
Mosquito Season**

**&**

**Providing Effective Spray Strategy For City  
Of Chicago**

**PROBLEM STATEMENT**

**TRAIN**

**10,506**

Observations

**2007,2009,2011,2013**

NUM MOSQUITO

WNV PRESENT

DATE  
ADDRESS

SPECIES

BLOCK

STREET

TRAP

ADDRESS NUM & STREET

LATITUDE

LONGITUDE

ADDRESS ACCURACY

**TEST**

**116,293**

Observations

**2008,2010,2012,2014**

**DATASETS**

# WEATHER

2007 - 2014

NOAA weather data

22 weather features

## 2 Stations

- 1) CHICAGO O'HARE INTERNATIONAL AIRPORT
- 2) CHICAGO MIDWAY INTL ARPT

# SPRAY

14,835

Observations

2011 (2 dates)

2013 (July - Sep)

Spray effort data by Chicago  
government

# DATASETS



Data  
Cleaning



EDA



Data  
Merging



Feature  
Engineering



Modelling



Cost-Benefit  
Analysis



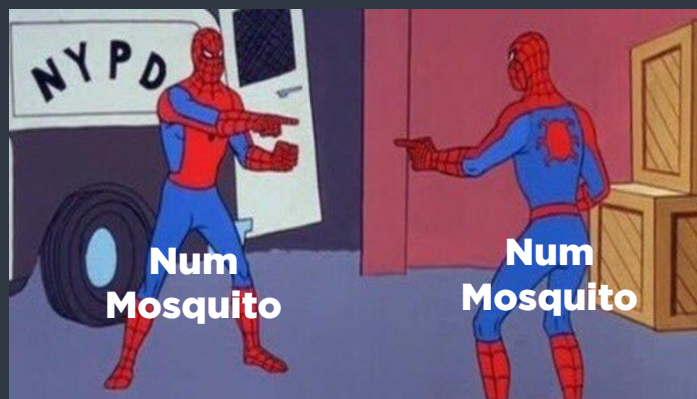
# WORKFLOW

# TRAIN & TEST

- Relatively clean with no null values

## Train

- Mosquitos count capped at 50 for each date, trap, species and Wnv present
- Sum up all mosquitos of same date, trap, species and Wnv present, drop duplicate



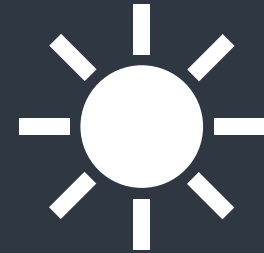
# CLEANING





# WEATHER

- Data had no null values there were many missing values and traces which were labelled M and T
- Making references to best fill missing data (e.g. WetBulb, StnPressure)
- Dropped columns with low variance or when there is no logical way to input values (e.g. SnowFall, Water1)



# CLEANING



# TRAIN

Wnv Present = 1

5.3%

Wnv Present = 0

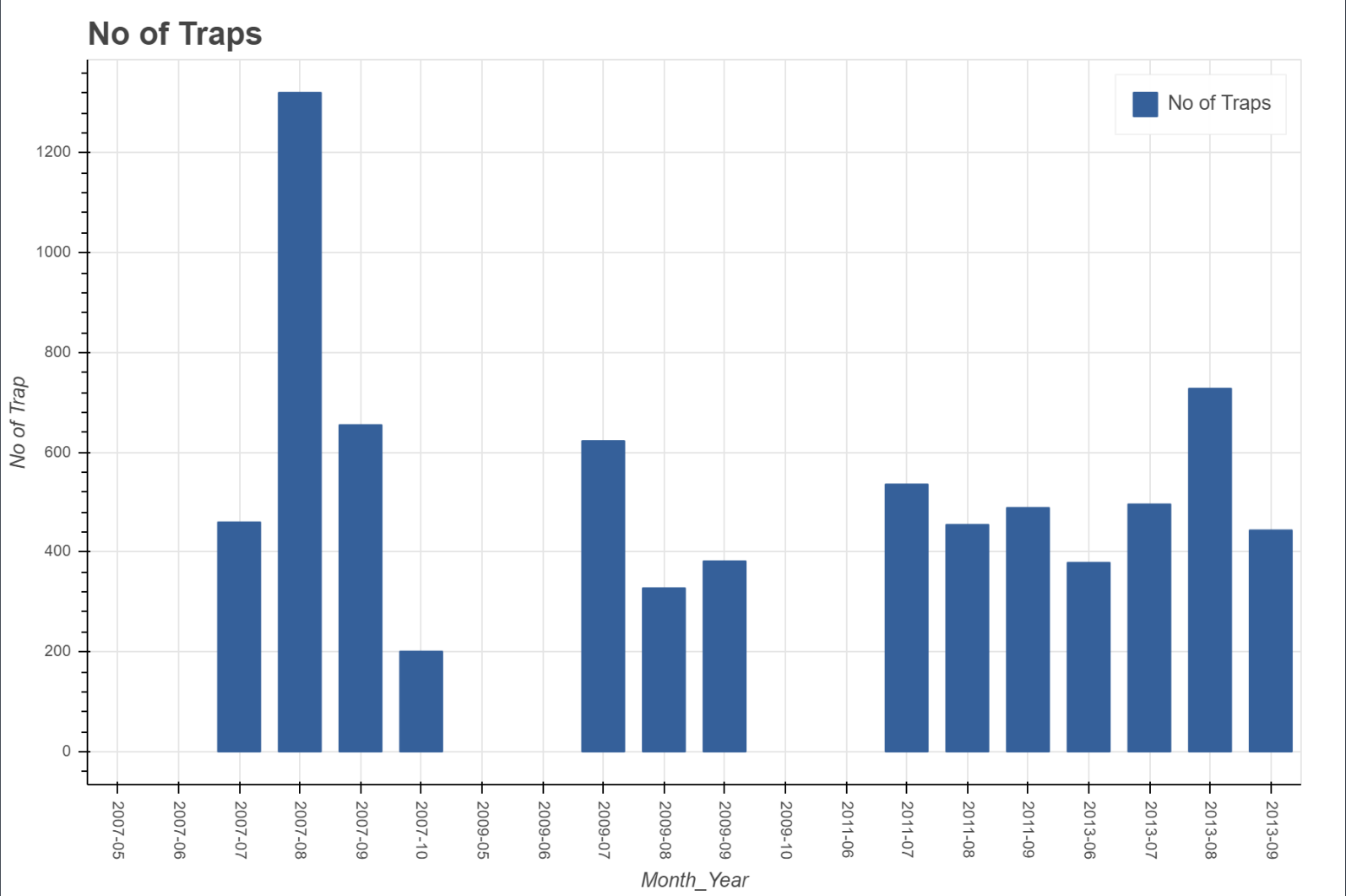
94.7%



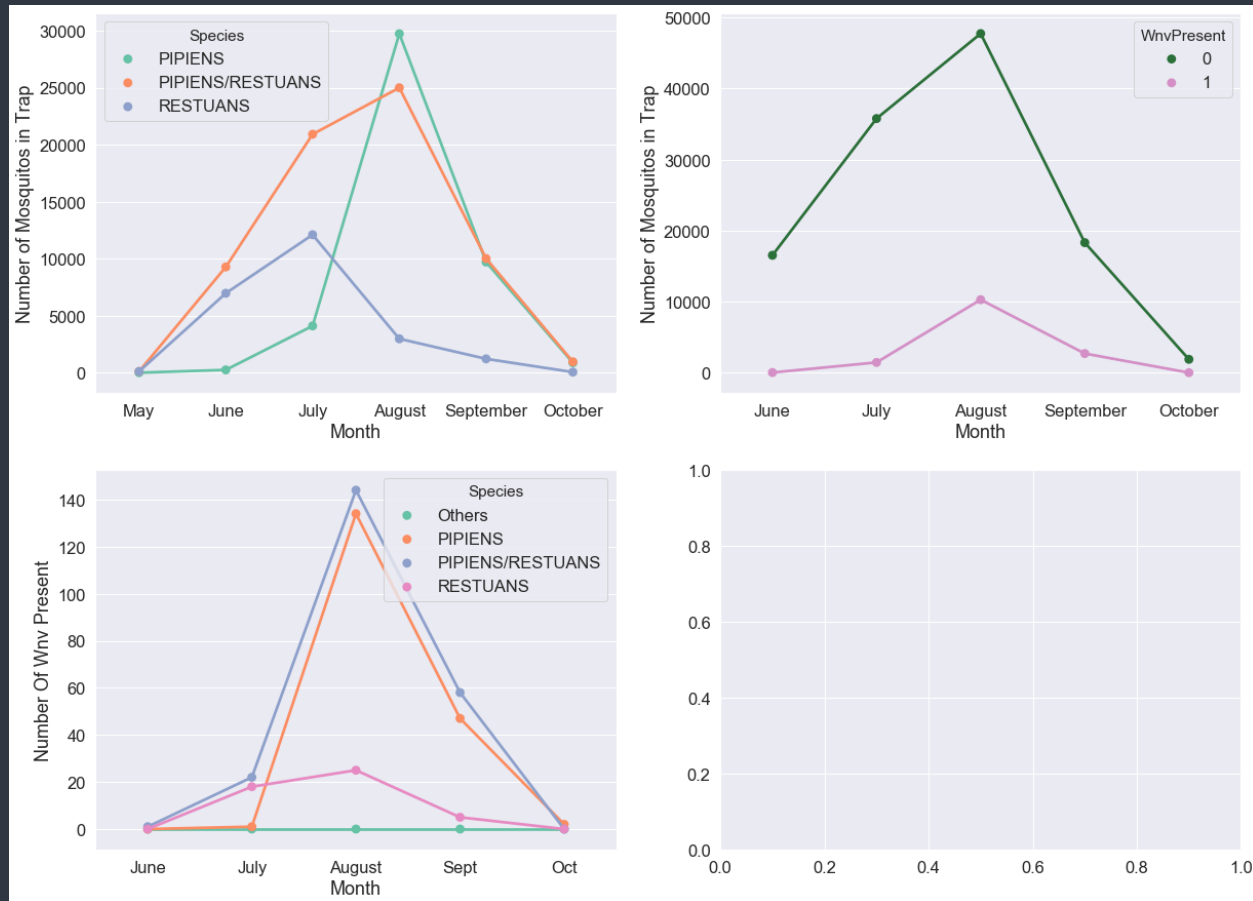
EDA



# TRAIN

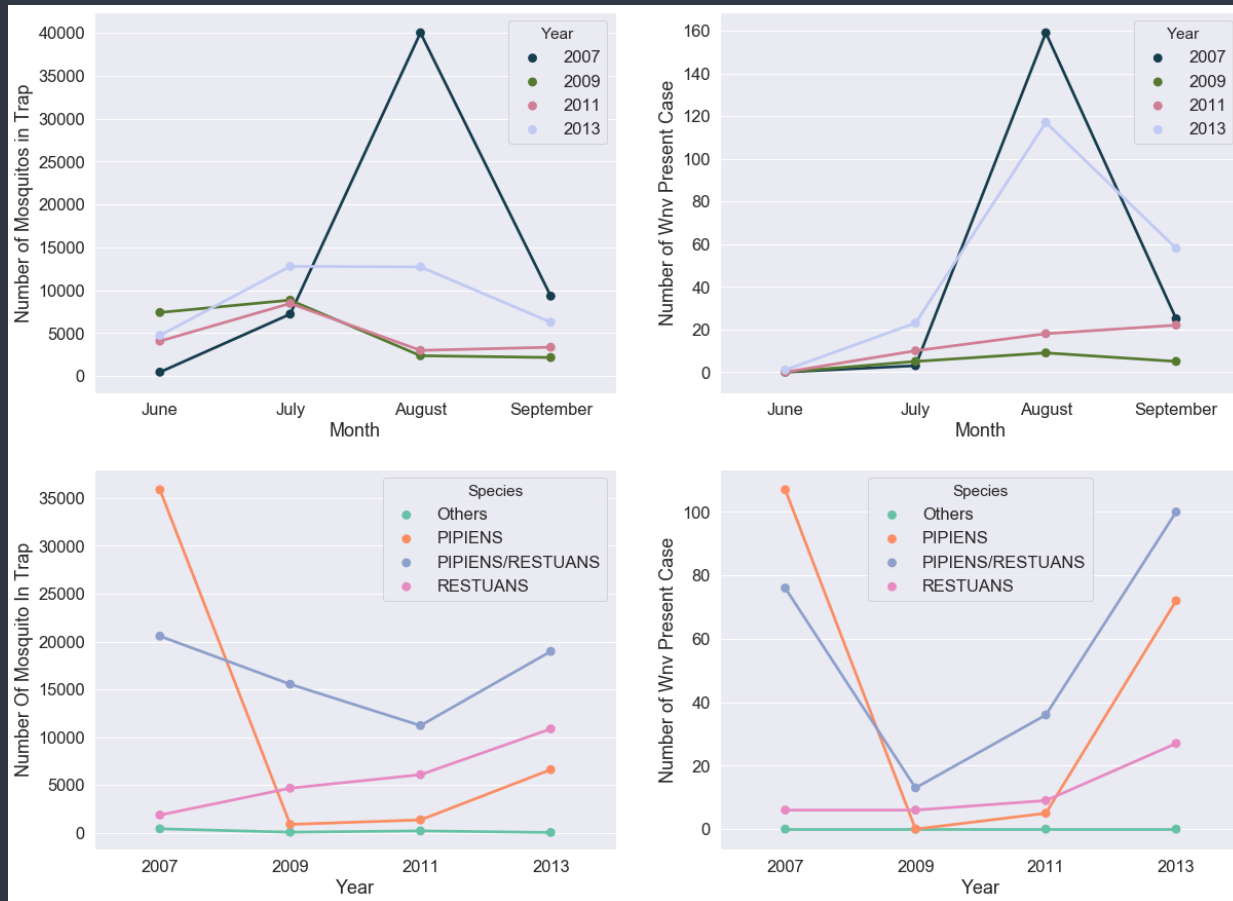


# TRAIN



- Most number of Wnv positive mosquitos in August
- Most number of mosquitos in August for species Pipens & Pipens/Restuans
- Pipens & Pipens/Restuans seem to be the main contributors

# TRAIN



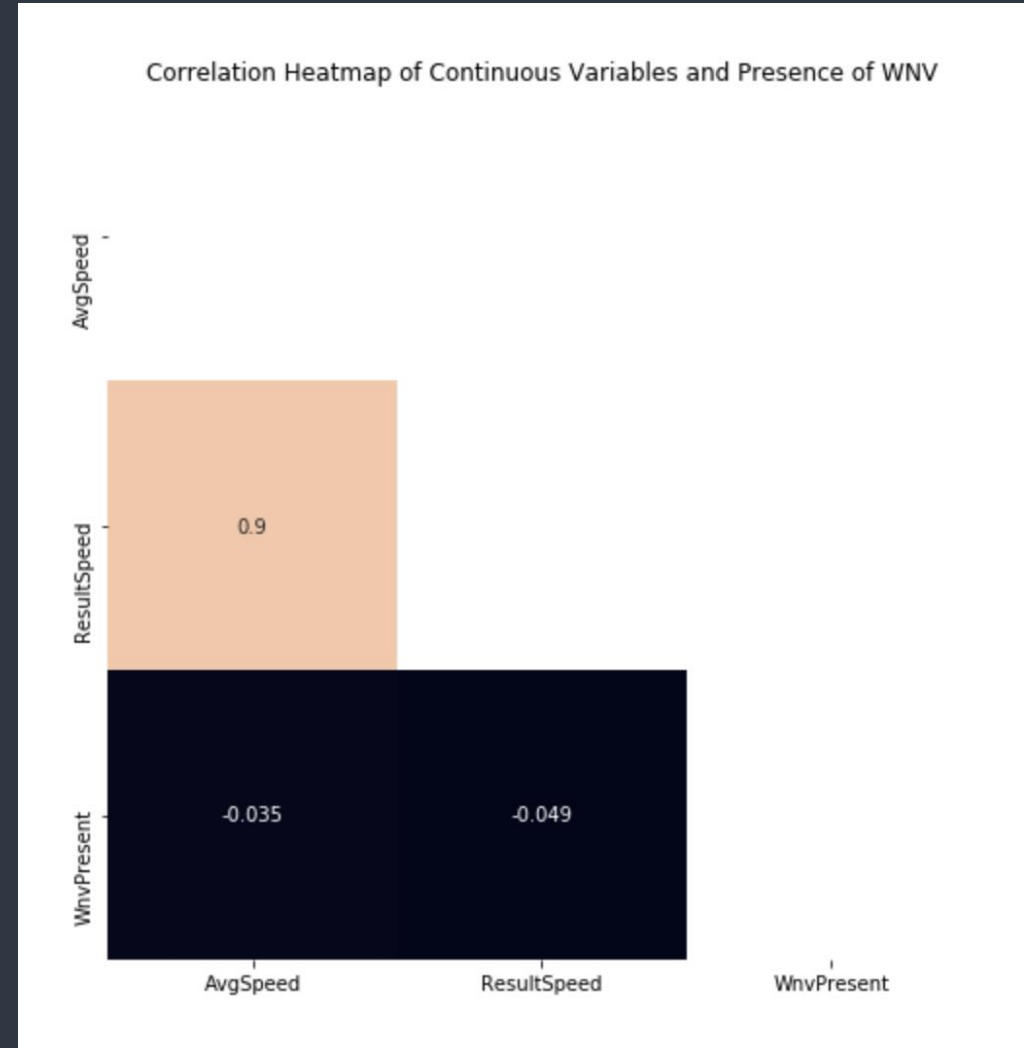
- Sharp drop in no. of Wnv Present cases between 2007 to 2009
- Sharp increase between 2011 and 2013
- Sharp drop in PIPENS mosquito from 2007 to 2009 and increase from 2011 to 2013
- PIPENS mosquito seems to be a major factor affecting whether Wnv is present

# WEATHER

Removing Collinear Terms

<b>AvgSpeed</b>	0.034605
<b>ResultSpeed</b>	0.048893

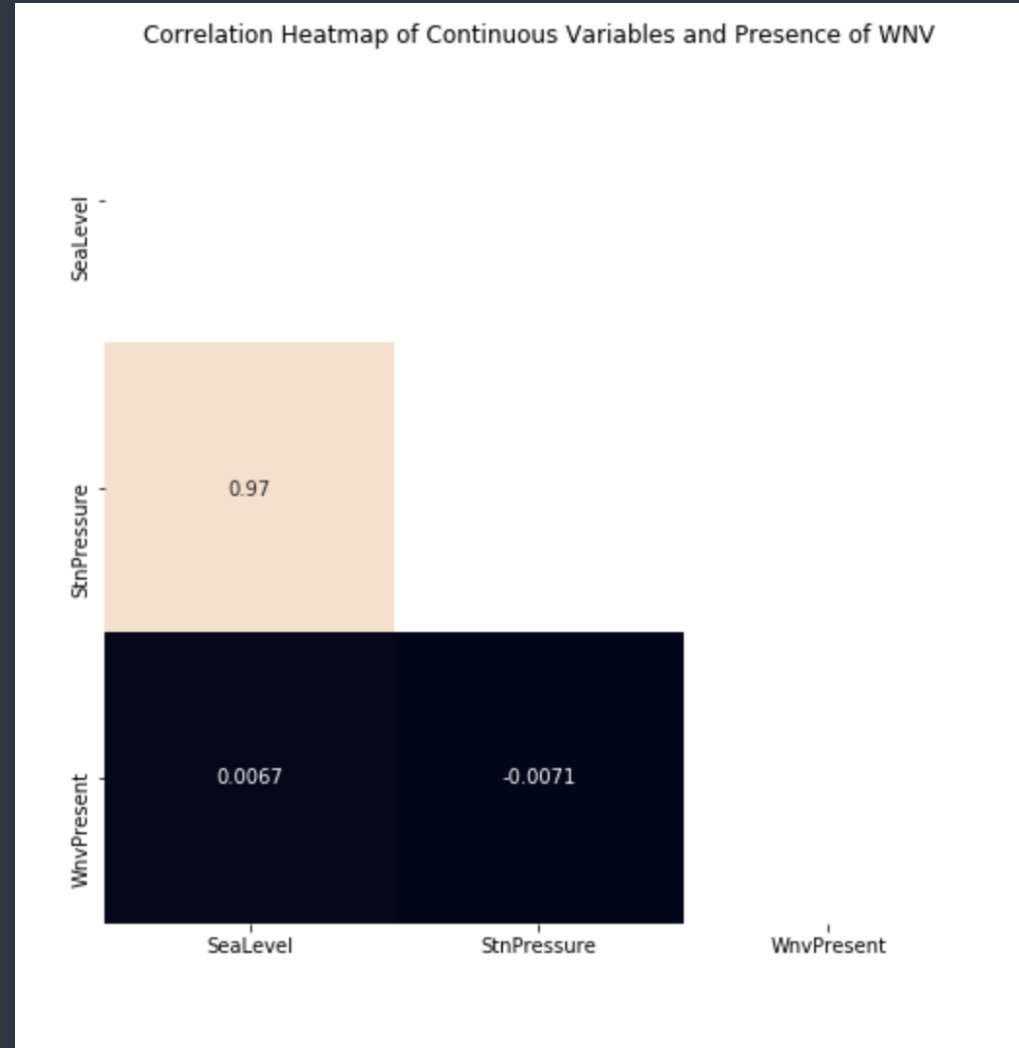
EDA



# WEATHER

Removing Collinear Terms

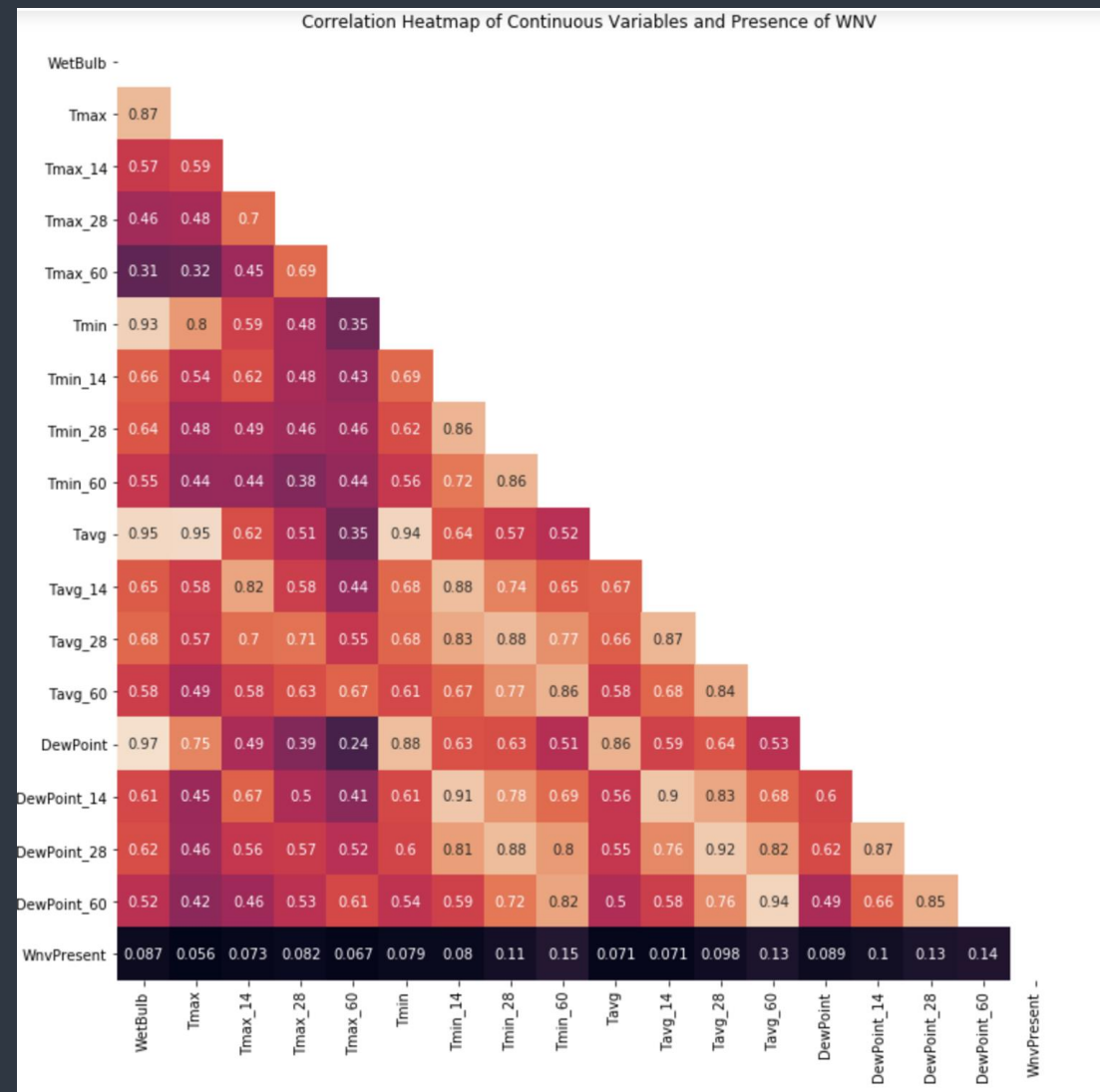
SeaLevel	0.006738
StnPressure	0.007149



# WEATHER

## Removing Collinear Terms

WetBulb	0.087295
Tmax	0.056156
Tmax_14	0.073485
Tmax_28	0.082459
Tmax_60	0.067110
Tmin	0.078749
Tmin_14	0.080009
Tmin_28	0.110154
Tmin_60	0.154747
Tavg	0.070603
Tavg_14	0.070628
Tavg_28	0.097595
Tavg_60	0.130006
DewPoint	0.088737
DewPoint_14	0.103341
DewPoint_28	0.132170
DewPoint_60	0.142533



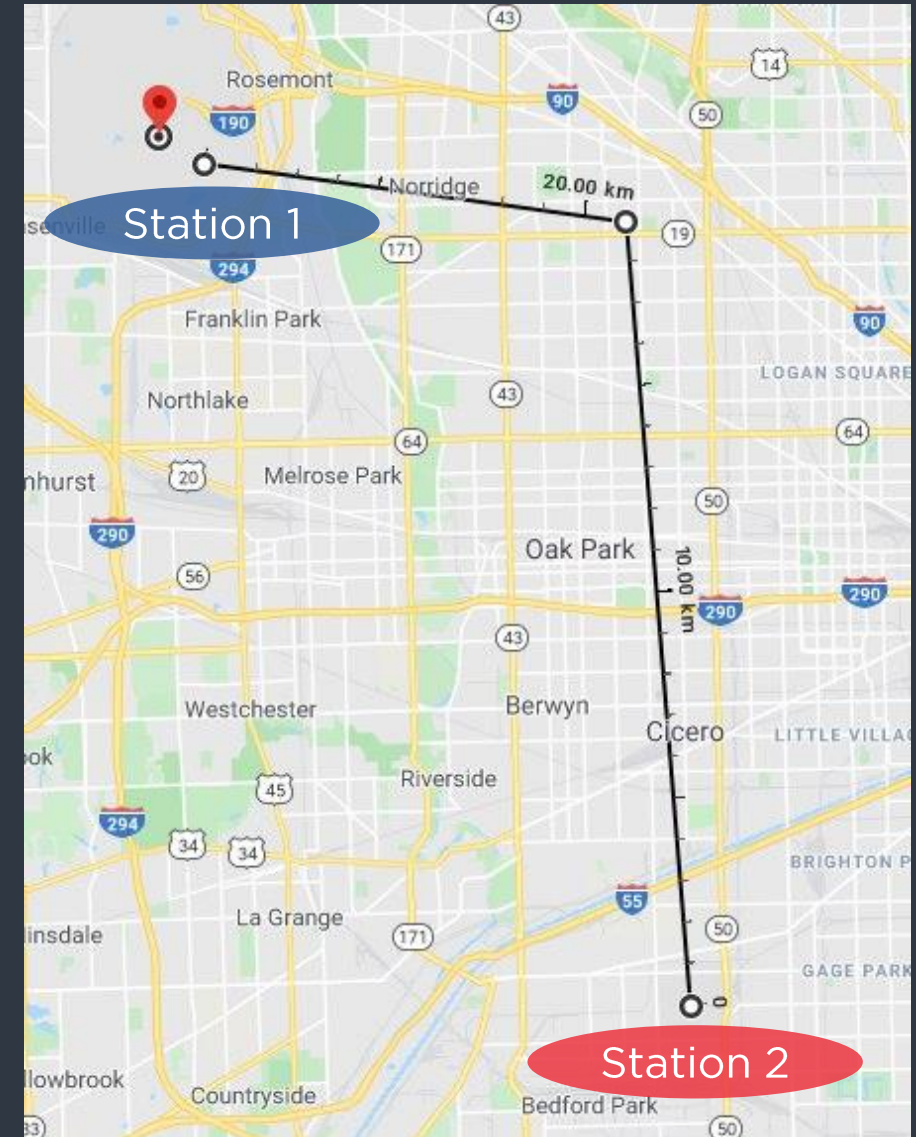


# WEATHER + TRAIN/TEST

Two main weather stations:

- Station 1: Chicago O'Hare International Airport
- Station 2: Chicago Midway International Airport

Calculate displacement to the weather stations and take information from nearest station when merging to train/test data.



# MERGING





Identify main locations with highest WnvPresent



Calculate the distance from these locations to the rest of the train/test data



Get rolling mean of different periods for temperature, dewpoint and precipitation

# ENGINEERING

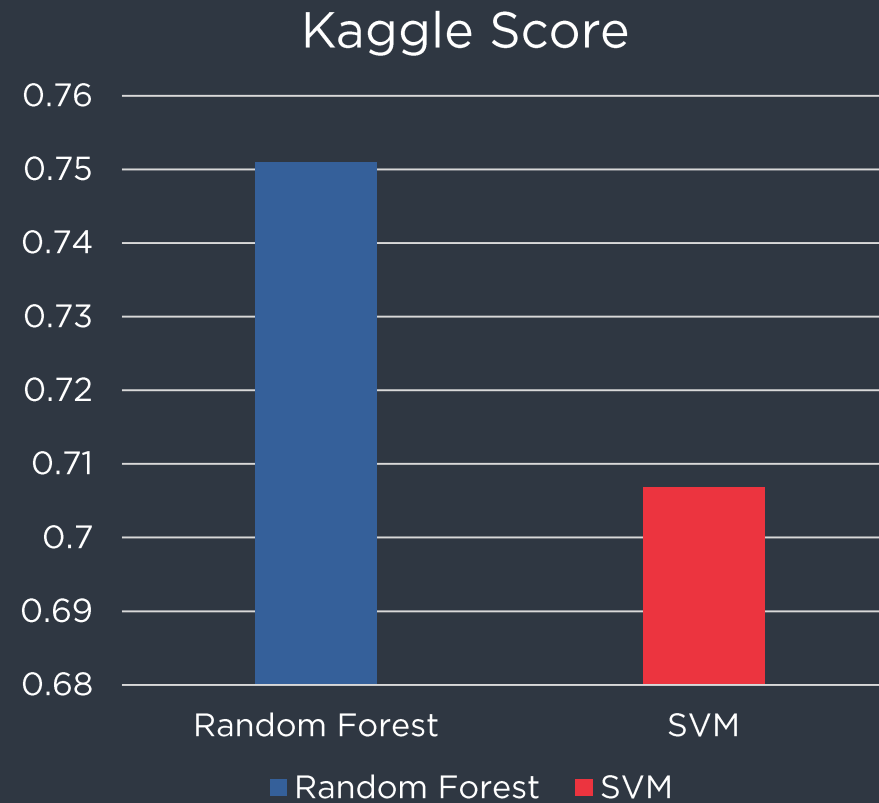


## RANDOM FOREST CLASSIFIER

**0.75095**

## SUPPORT VECTOR CLASSIFIER

**0.70679**

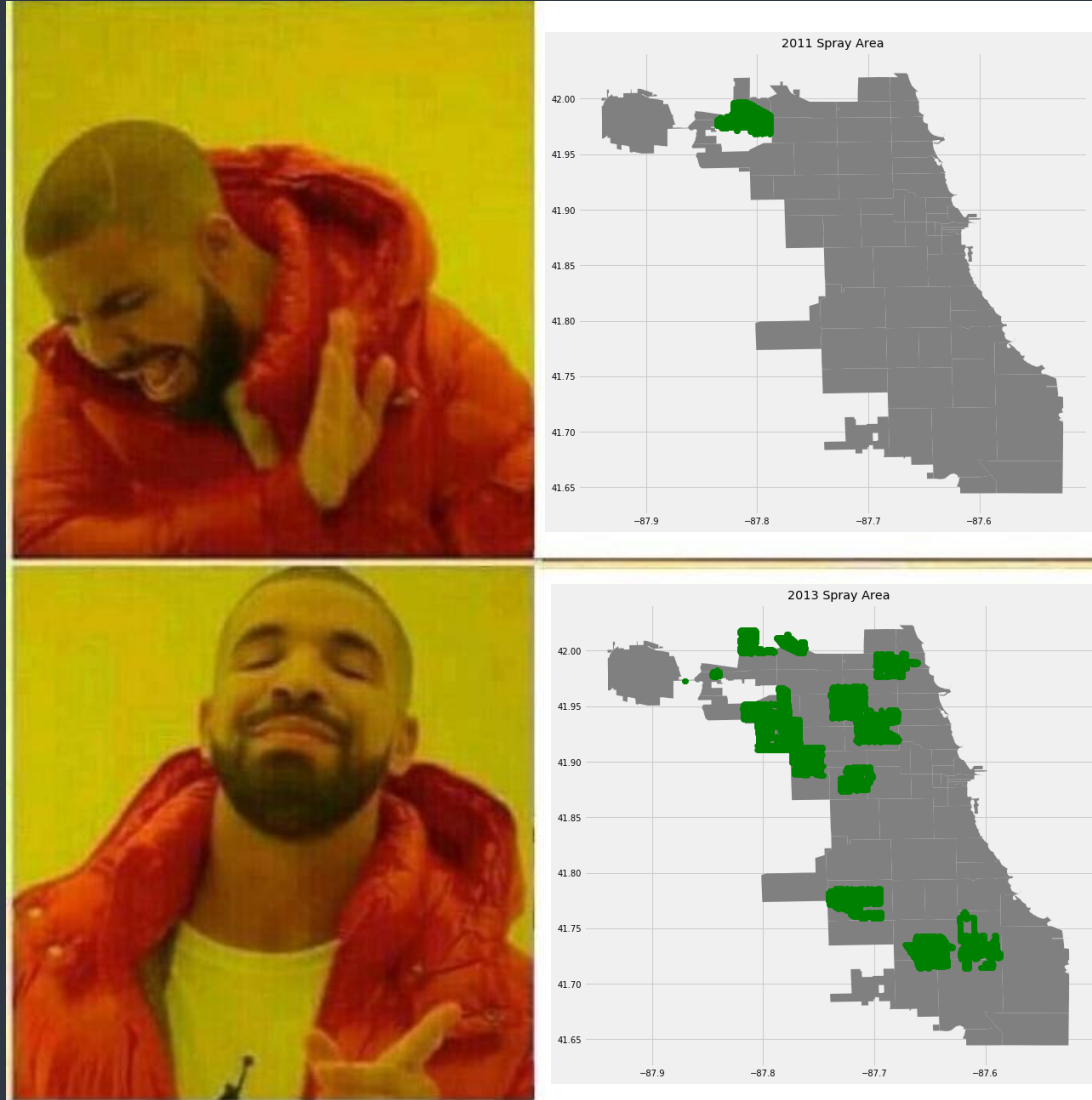


# MODELLING



# SPRAY DATA

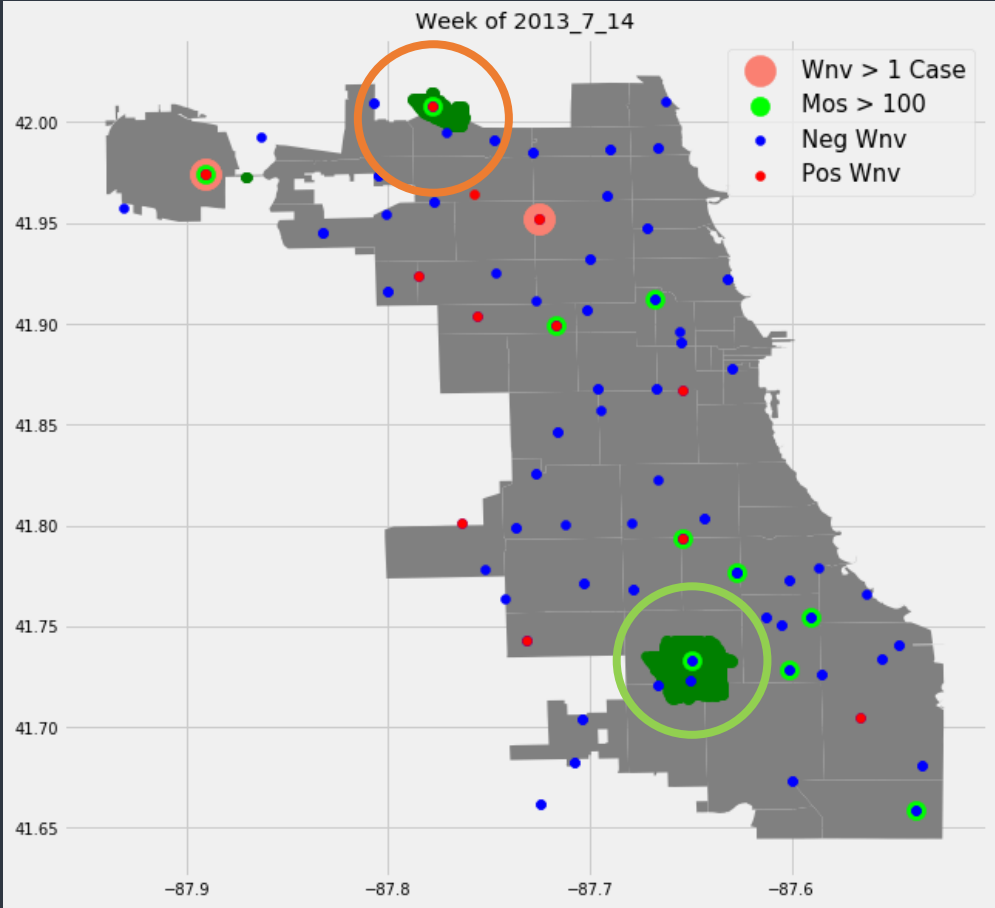
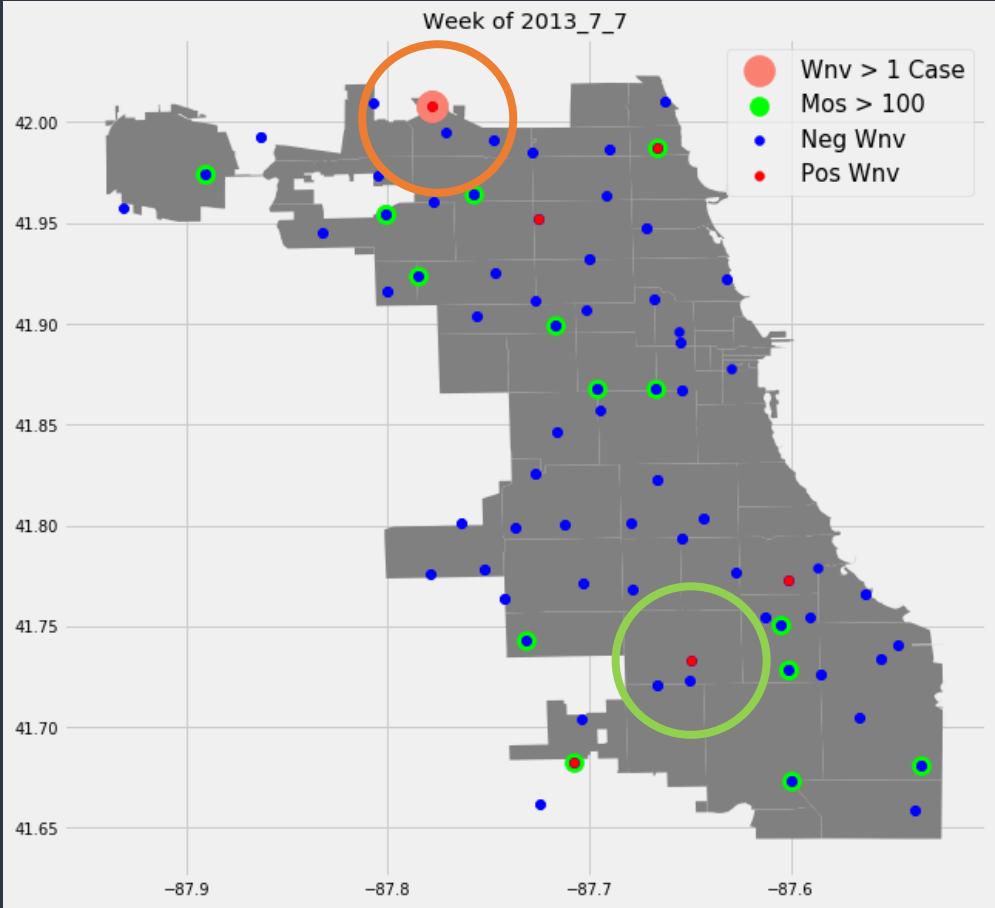
- 2011 data limited
- Analyse only on 2013



# COST ANALYSIS



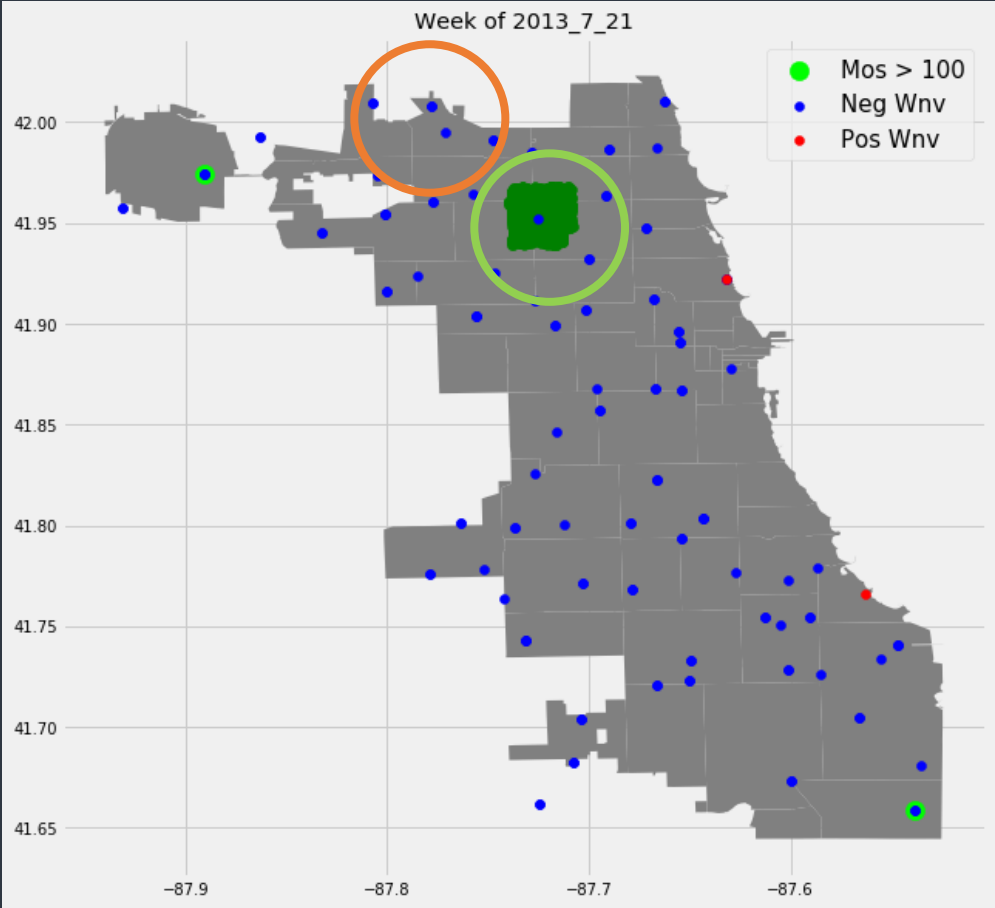
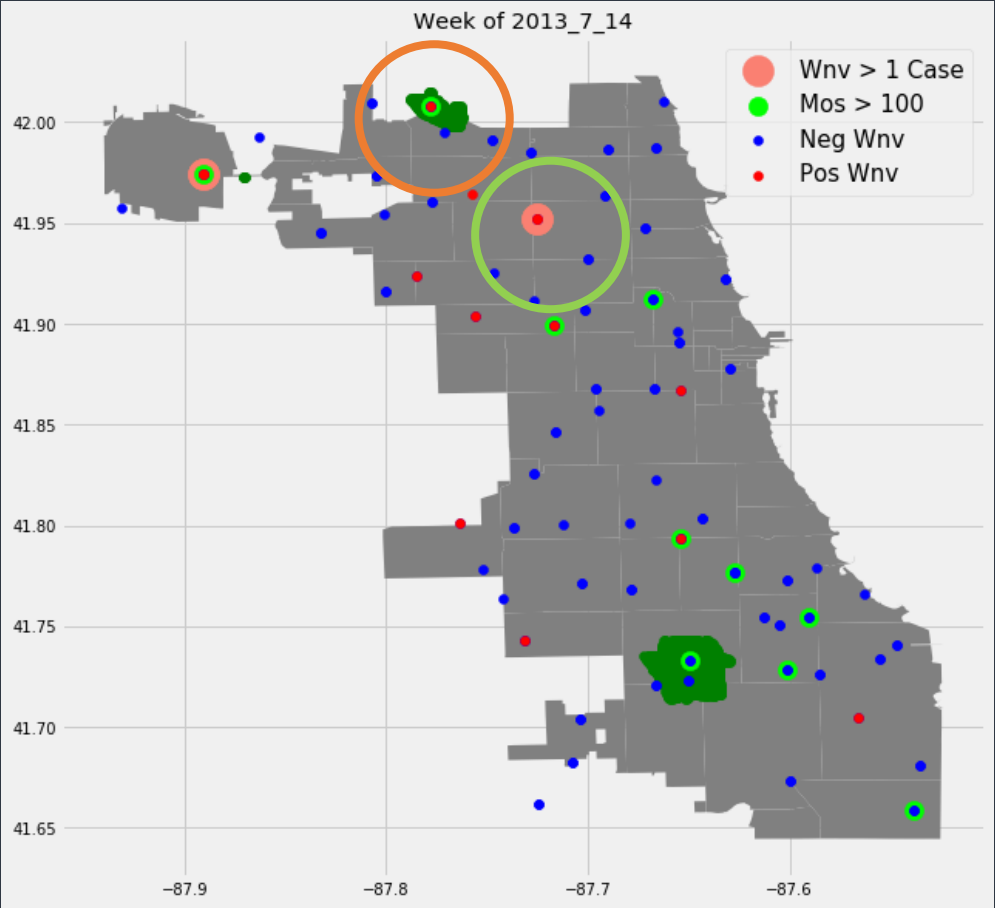
# 2013 SPRAY DATA



# COST ANALYSIS



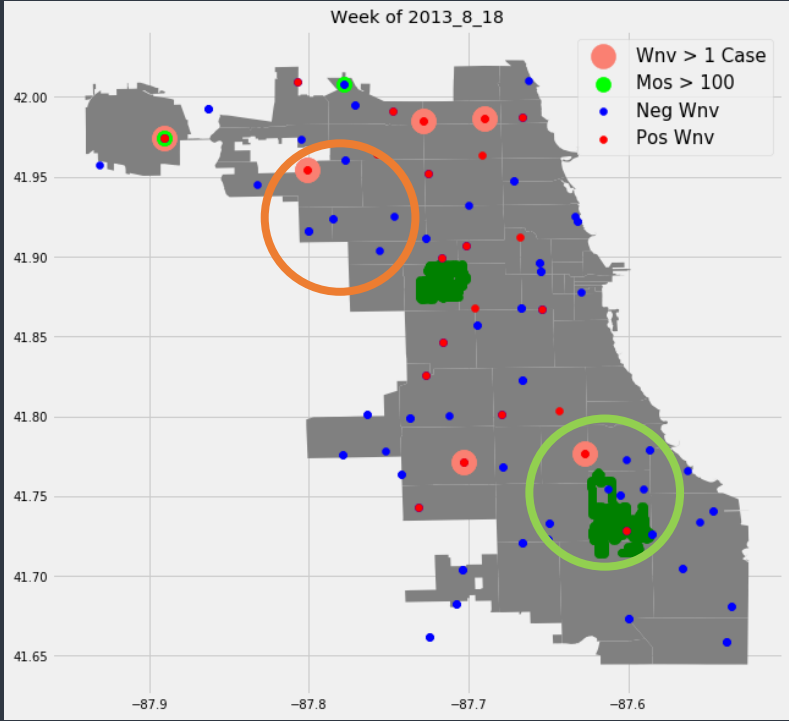
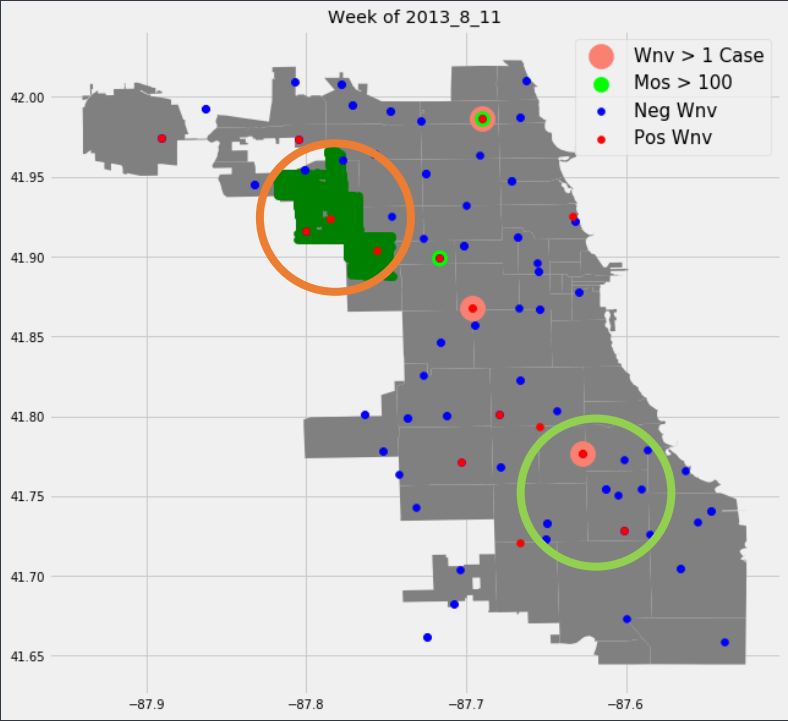
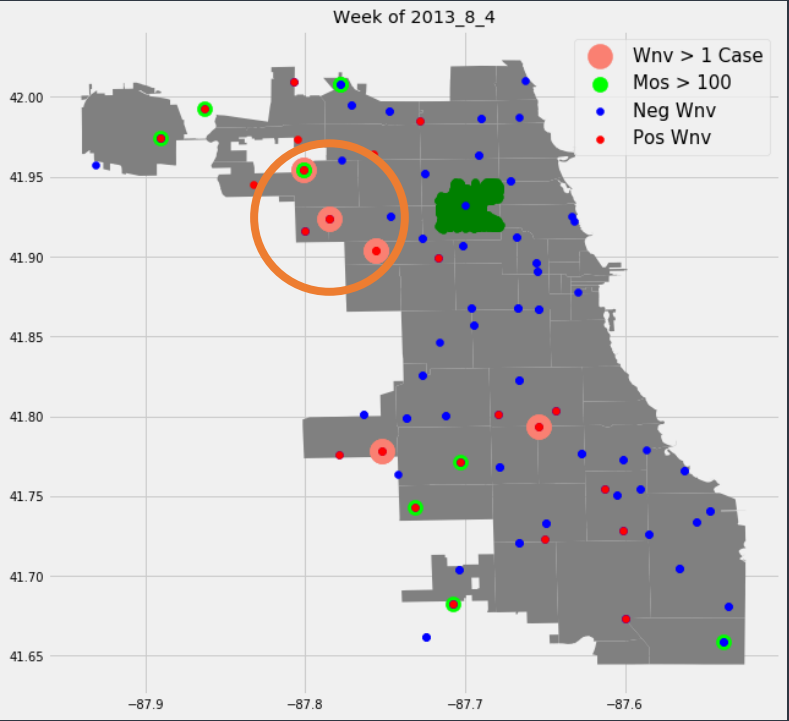
# 2013 SPRAY DATA



# COST ANALYSIS



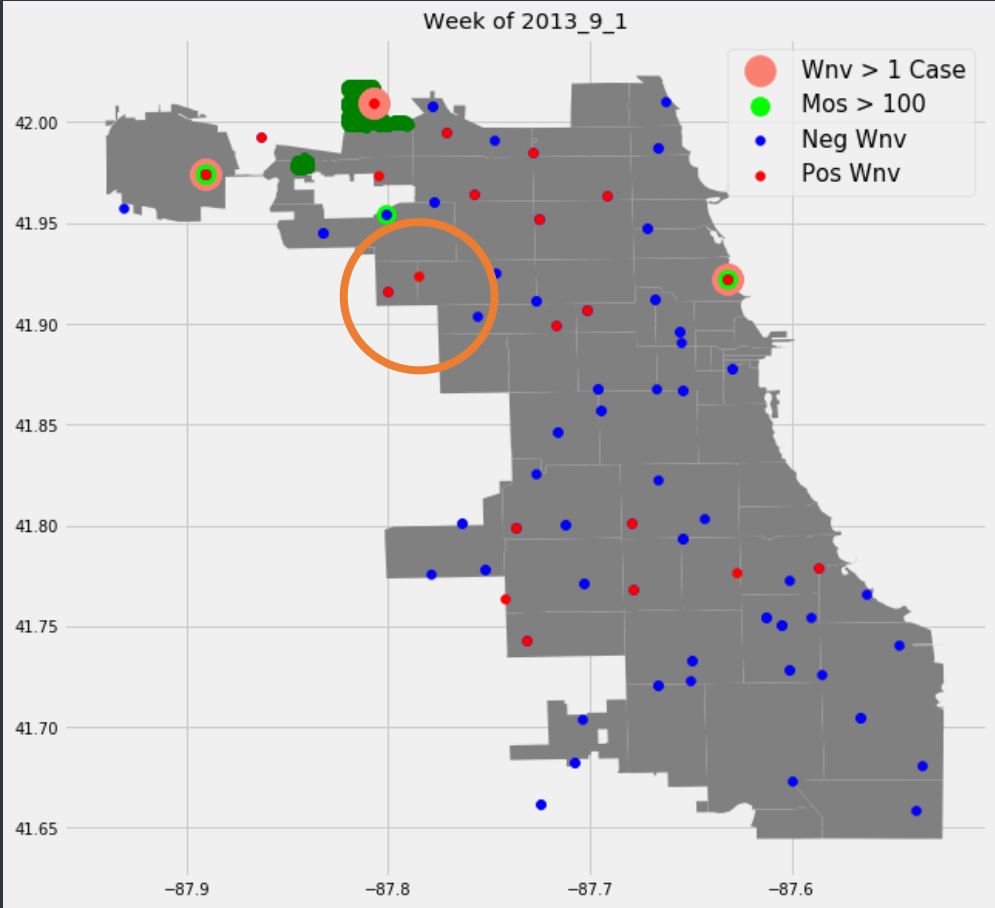
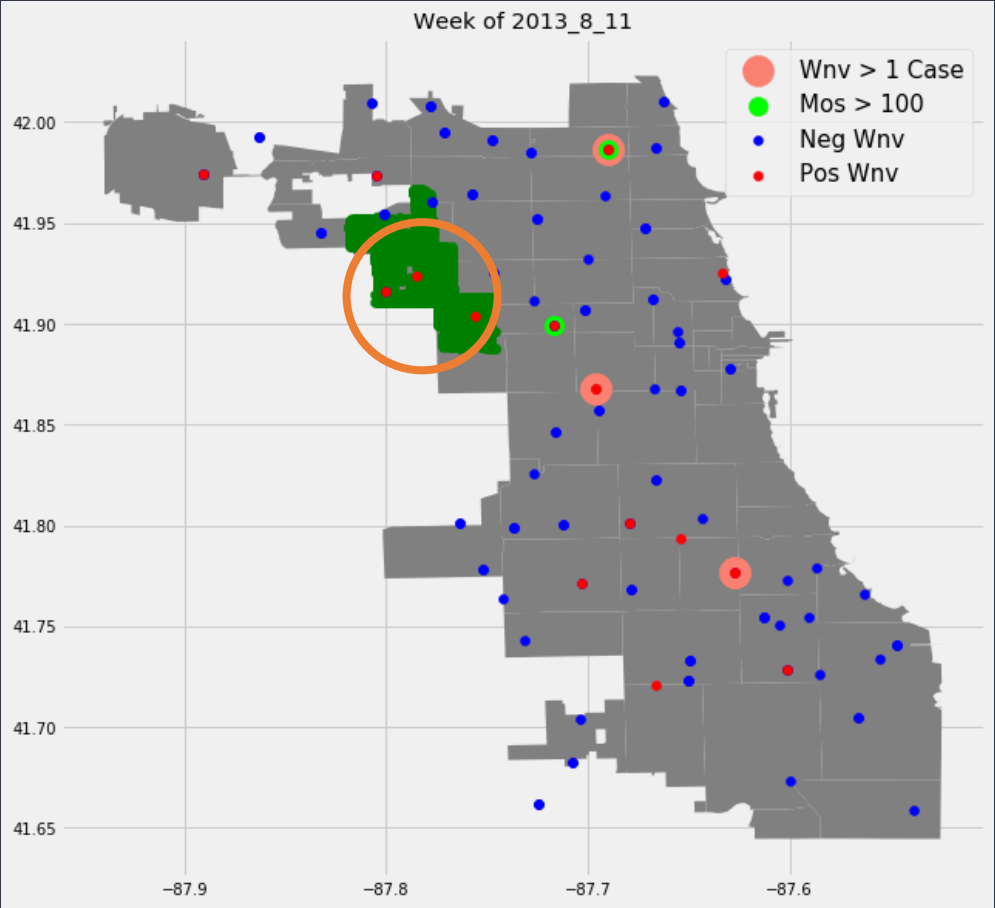
# 2013 SPRAY DATA



# COST ANALYSIS



# 2013 SPRAY DATA



# COST ANALYSIS





# 2013 SPRAY DATA

- Spraying worked but with short term effect
- Sprayed areas often lagged by at least 1 week

Year	Total Spray Count	Total Mosquitos with Wnv \+Ve	No\ of Wnv Human case*	Ratio of Wnv Mos : Human
2011	1668	50	8	6.25
2013	12626	199	37	5.38
Differences (2013/2011)	7.57	3.98	4.63	0.86

\*Source: Chicago Department of Public Health (CDPH)

## COST ANALYSIS



# COST ESTIMATION

- 12626 spray to eliminate 14 traps with +ve Wnv

$$12626/14 = \mathbf{902 \text{ spray/+ve Wnv trap}}$$

- Average price per spray: USD 100
- Visually inspect 10 weeks plot or calculate no. of trap require to spray if option is chosen in 2013

## COST ANALYSIS



# RECOMMENDATIONS

Total: 199 Traps

## OPTION 1

Total cost:

$$199 * 902 * 100 = \$17,946,957$$

# COST ANALYSIS



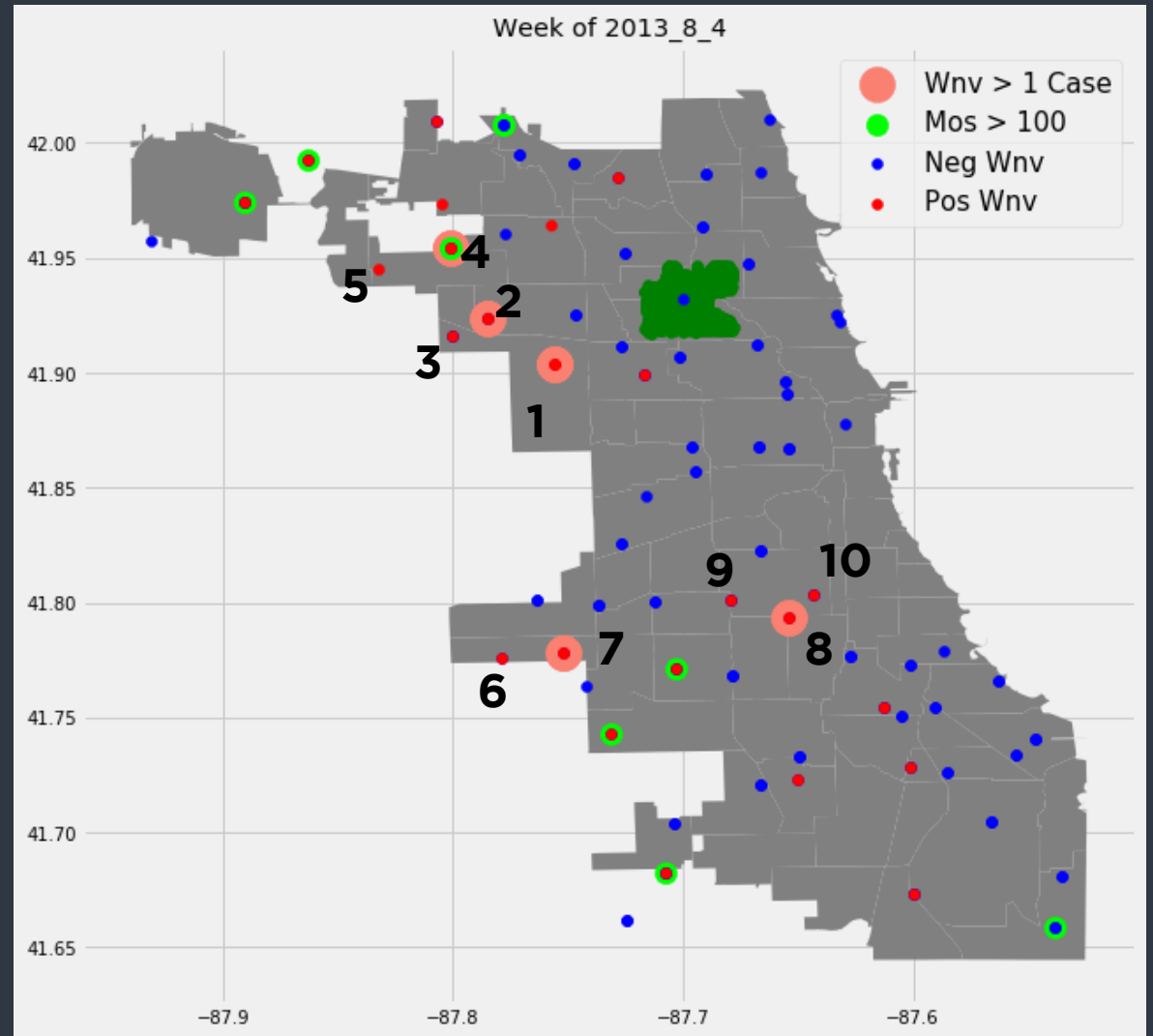
# RECOMMENDATIONS

## OPTION 2

Total: 41 Traps

Total cost:

$$41 * 902 * 100 = \$3,697,614$$



# COST ANALYSIS



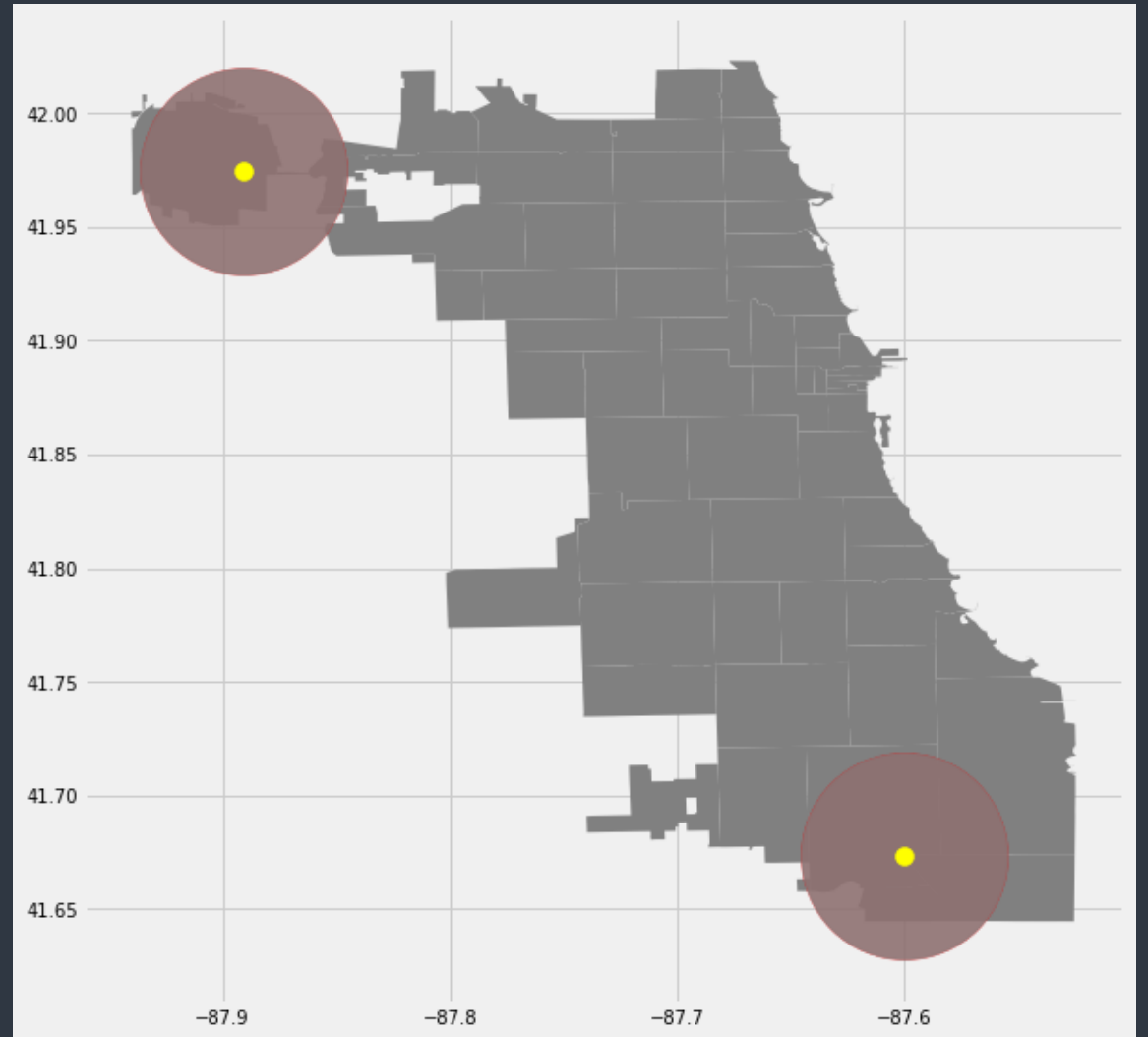
# RECOMMENDATIONS

## OPTION 3

Total: 23 Traps

Total cost:

$$23 * 902 * 100 = \$2,074,271$$



# COST ANALYSIS



# RECOMMENDATIONS

Estimated actual 2013 spending: **\$1,262,600**

	Option 1	Option 2	Option 3	Option 4
<b>Method</b>	Spray all Wnv Positive Areas	Spray based on community areas	Spray based on distance from 2 key points	Release genetically modified mosquitos
<b>Pros</b>	<ul style="list-style-type: none"><li>Reduce Wnv +ve mosquito can be dramatically</li></ul>	<ul style="list-style-type: none"><li>Reduce Wnv +ve mosquitos in high risk area</li><li>Cost saving</li></ul>	<ul style="list-style-type: none"><li>Focus effort in 2 specific areas</li><li>Spray on lesser areas</li></ul>	<ul style="list-style-type: none"><li>Long term fixed solution</li></ul>
<b>Cons</b>	<ul style="list-style-type: none"><li>Costly</li></ul>	<ul style="list-style-type: none"><li>Missing out some Wnv +ve traps</li></ul>	<ul style="list-style-type: none"><li>Missing out some Wnv +ve traps</li></ul>	<ul style="list-style-type: none"><li>Require significant R&amp;D investment cost &amp; time to implement</li></ul>
<b>Cost</b>	\$17,946,957	\$3,697,614	\$2,074,271	\$3,599,880

## COST ANALYSIS





Random Forest Classifier

**0.75095**



Spray Recommendations

**4 Options**



More Info

**No. Mosquito in test**

**No. of Wnv human case & area**

**More spray data**

# SUMMARY

THANK YOU

