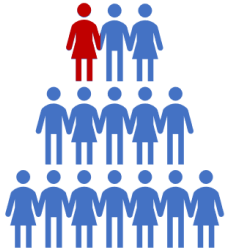
WEST NILE VIRUS

Project 4
Jiansheng | Le-Anne | Shaun

Intro/Background

West Nile Virus (WNV) is the leading cause of mosquito-borne disease in the United States (Centers for Disease and Control Prevention, 2019). It is most commonly spread to people by the bite of an infected mosquito.

1 out of **5**
Fever/other symptoms



1 out of **150**
Serious/fatal illness

Risk of WNV can be reduced by controlling the mosquito population to prevent the spread of WNV.

Problem Statement

Which part of Chicago do we want to deploy pesticides?

Methodology:

1. **Predict the presence of the West Nile Virus.**
2. **Evaluate the effectiveness of spray**
3. **Recommend a deployment plan for the pesticides throughout the city of Chicago.**

Workflow

Dataset

Cleaning & Considerations

EDA

Feature Engineering

Model Evaluation & Selection

Cost-benefit Analysis

Dataset

Weather

Station
Date (2007-14)
T.max
T.min
T.avg
Depart
Dewpoint
Wetbulb
Heat
Cool

Sunrise
Sunset
Codesum
Depth
Water1
Snowfall
PreciTotal
StnPressure

Sealevel
Resultspeed
Resultdir
Avgspeed

Spray

Date (2011,13)
Time
Latitude
Longitude

Test	Train
Id	
(2008,10,12,14) Date	Date (2007, 09, 11, 13)
Address	Address
Species	Species
Block	Block
Street	Street
Trap	Trap
AddressNumberAndStreet	AddressNumberAndStreet
Latitude	Latitude
Longitude	Longitude
Address Accuracy	Address Accuracy
	NumMosquitos
	WnvPresent

Cleaning & Considerations

Weather

- Changing date format
- Check and remove for 'M' (missing) values
- Dropping Tmin, Tmax, Depart, Heat, Cool, Sunrise, Sunset, Codesum, Depth, Water1, Snowfall, StnPressure, Avgspeed

Spray

- Changing date format
- Removing 541 duplicate entries

Train

- Removing 813 duplicate entries
- Changing date format
- Merging the insignificant species that contribute to the WNV into 'others'
- Dropping block, address, street, addressnumberandstreet & address accuracy
- Check and removing null values
- Convert all satellite trap to same trap name as parent trap

Test

- Changing date format
- Merging the species to match Train dataset
- Dropping block, address, street, addressnumberandstreet & address accuracy
- Convert all satellite trap to same trap name as parent trap

Feature Engineering

Train

- *Tot_mos_species*: Sum of the species that have WvnPresent within each trap.
- Converting Months (in Date) into Cos/Sin function

Weather

- Using Dewpoint & Wetbulb, calculate Saturated Vapour Pressure (SVP) & Vapour Pressure (VP)
- Using SVP & VP to calculate *Relative Humidity*
- Adding *Longitude* & *Latitude* for all entries:
  Station 1: (41.995, -87.933)
  Station 2: (41.786, -87.752)

Filtered Predictors

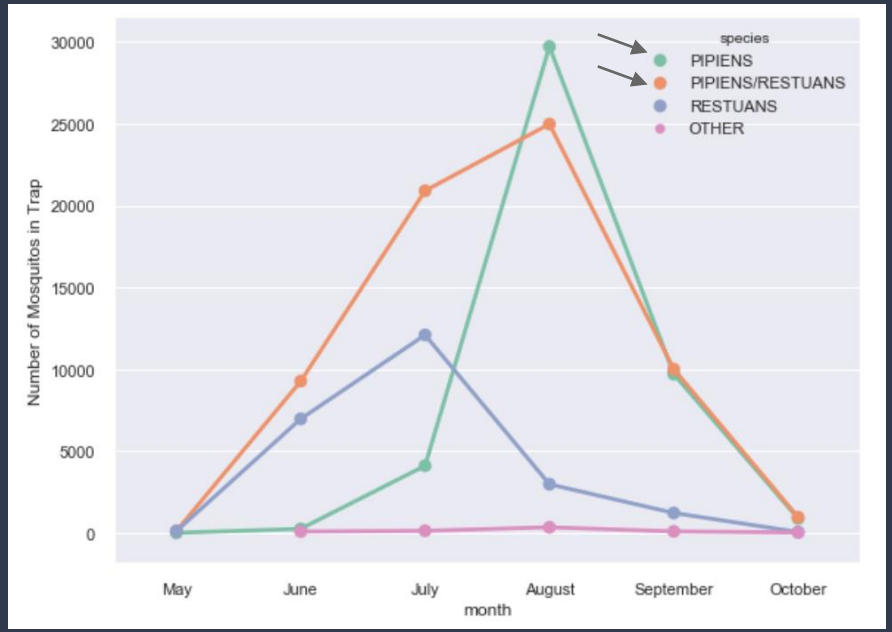
Weather
Station
Date (2007-14)
T.avg
PreciTotal
Sealevel
Resultspeed
Resultdir
Relative Humidity
Longitude
Latitude

Train
Date (2007, 09, 11, 13)
Species
Trap
Latitude
Longitude
NumMosquitos
{WnvPresent}
Tot_mos_species

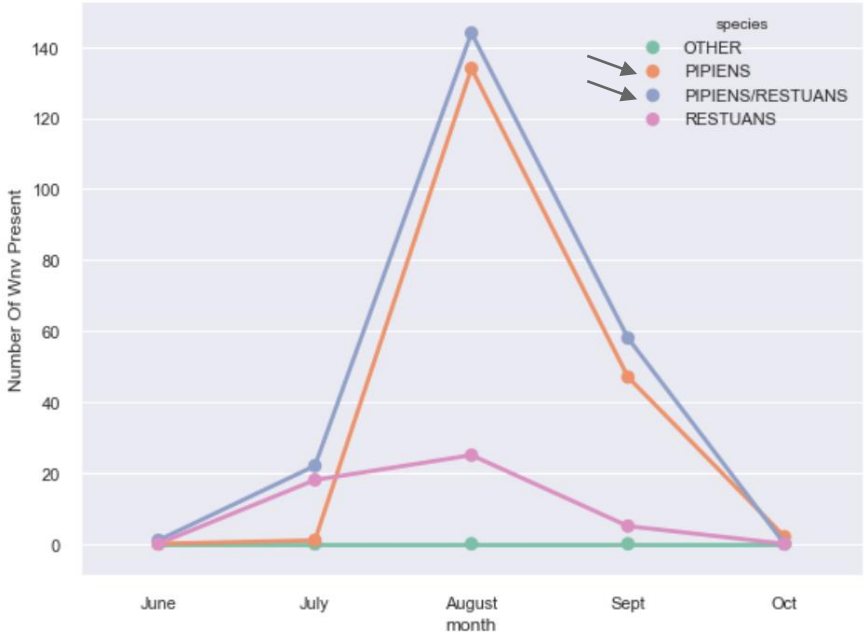
Spray
Date (2011,13)
Time
Latitude
Longitude

Test
Id
Date (2008,10,12,14)
Species
Trap
Latitude
Longitude

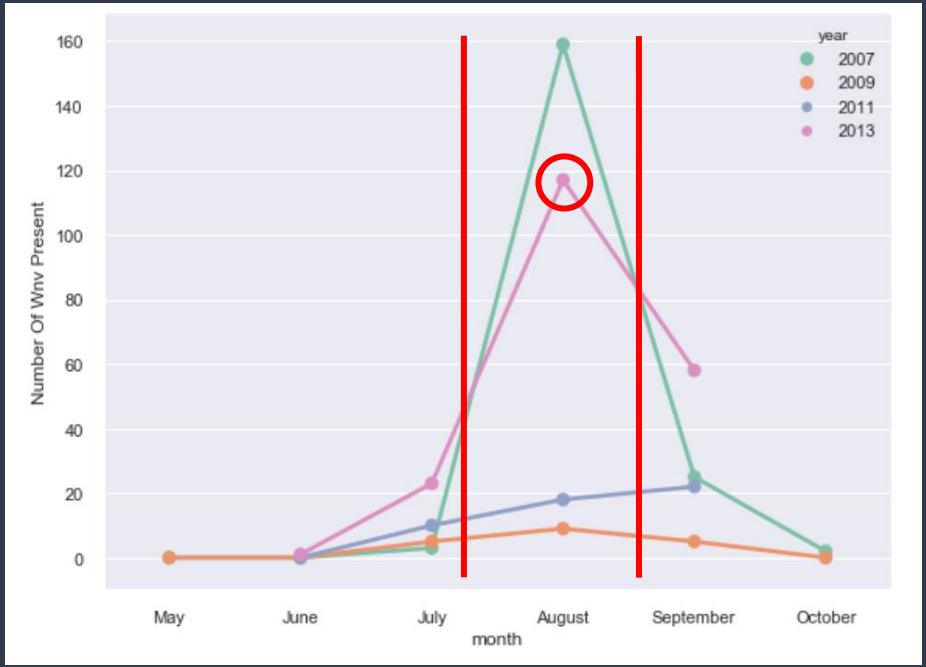
EDA - Species of Mos. Trapped



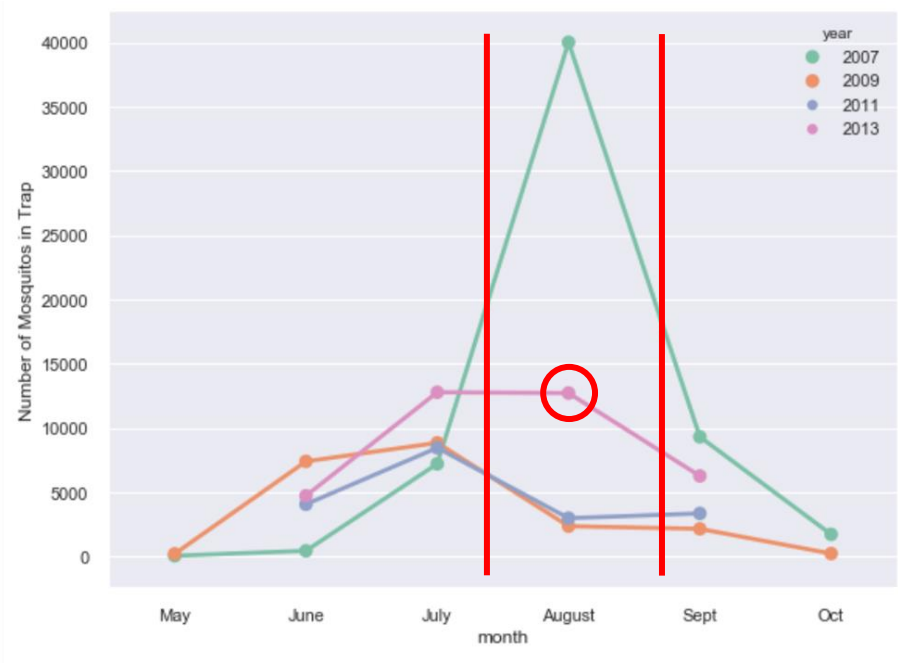
EDA - Species of Mos. w. WNV Present



EDA - Trend of WNV across year



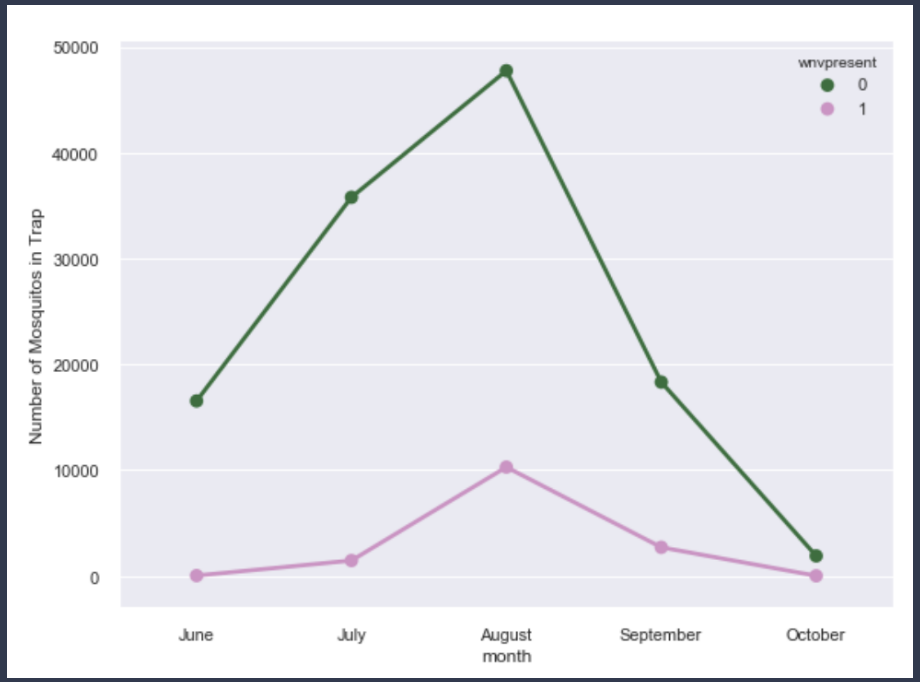
EDA - Trend of Mos Trapped



EDA - Species with WNV Across Species

EDA - Species Trapped across year

EDA – WNV vs Mos.



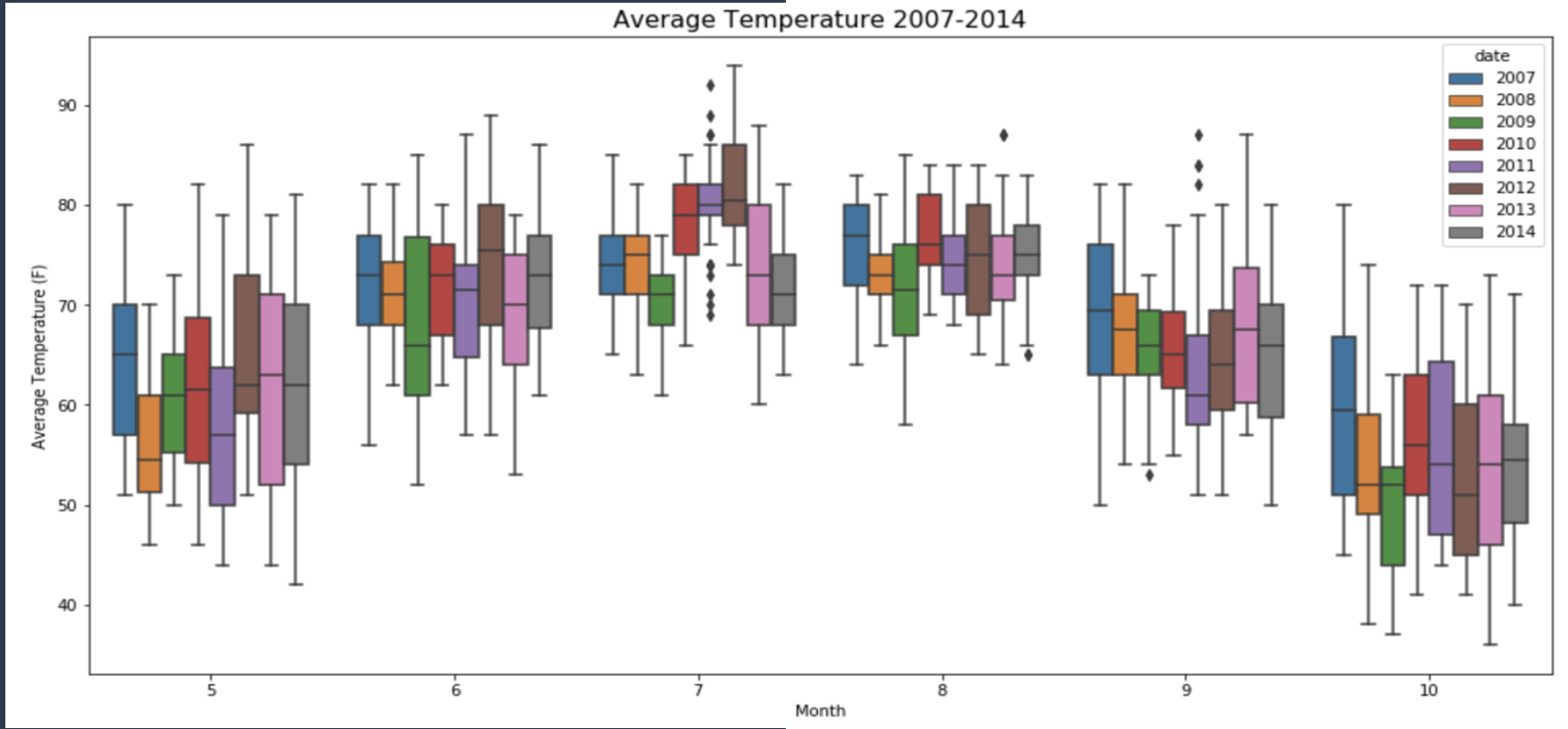
Findings so far:

- 2 Dominant species that contribute to the WNV transmission
- July - Sept seems to be the VNR ‘season’
- Peaks on August
- Generally, WNR & Mos. Trapped seems to have the same trend, but, there is a drastic change in 2013 August.

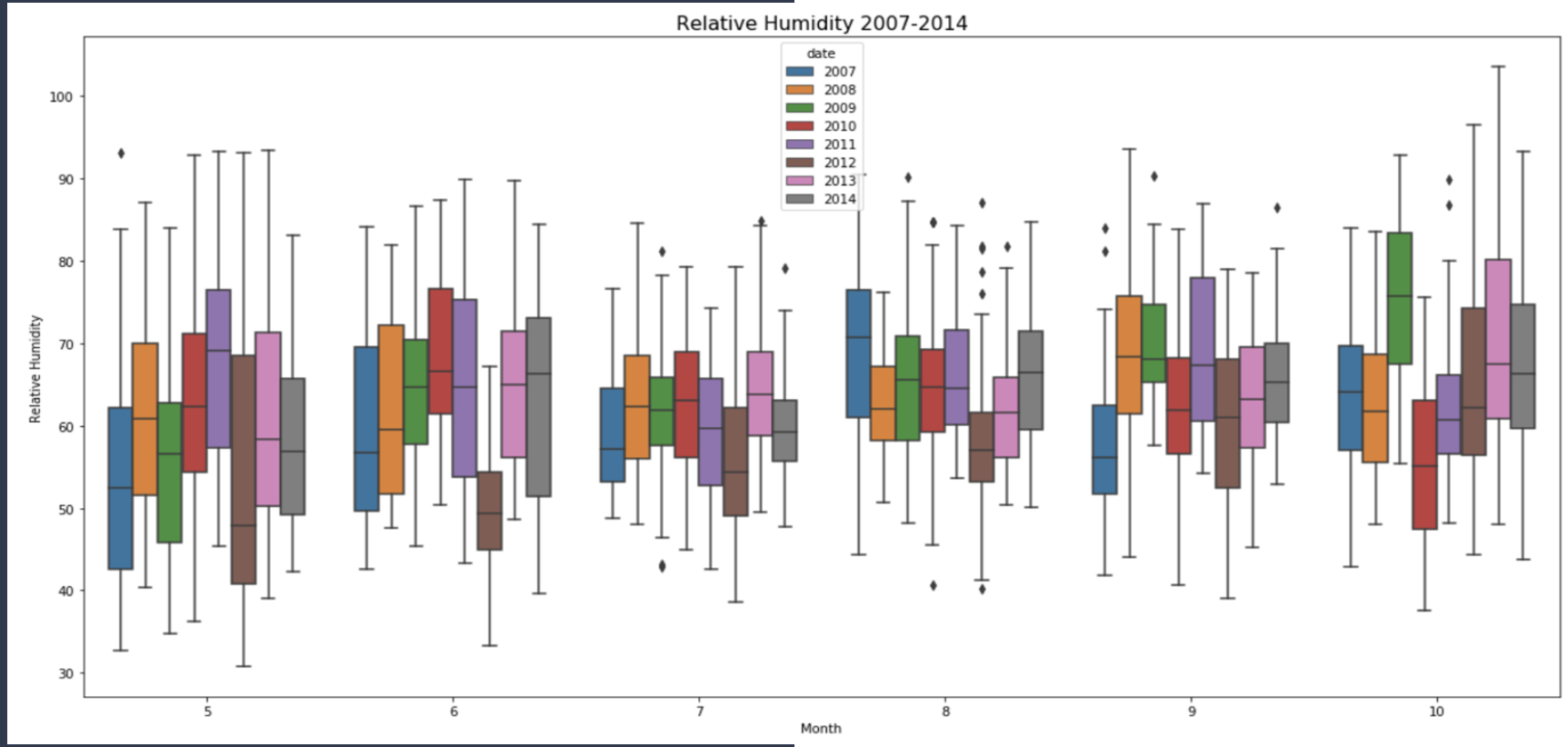
- Pipiens seems to be more of a species among all, that contribute to WNR transmission
- Pipiens/Restuans seem to have a opposite relationship during 2011

- Not all Mosquitos are transmitter of WNV

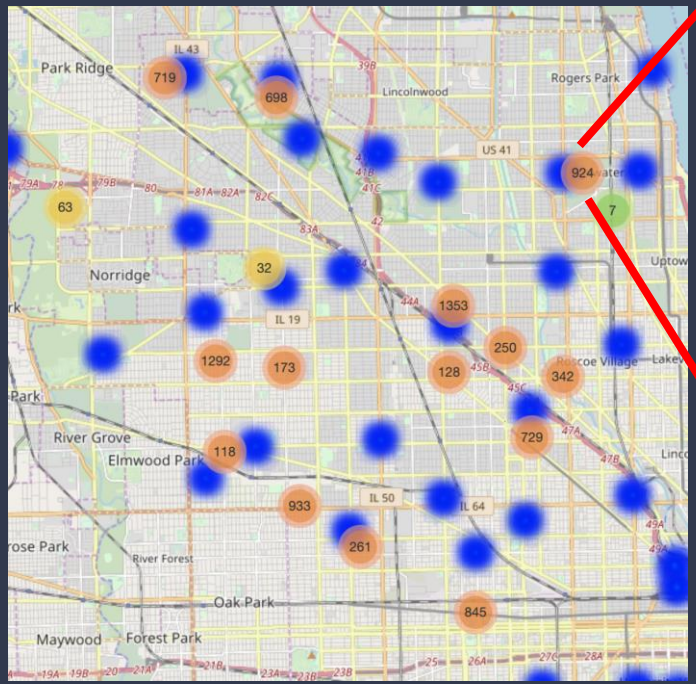
EDA – Trend of Tavg

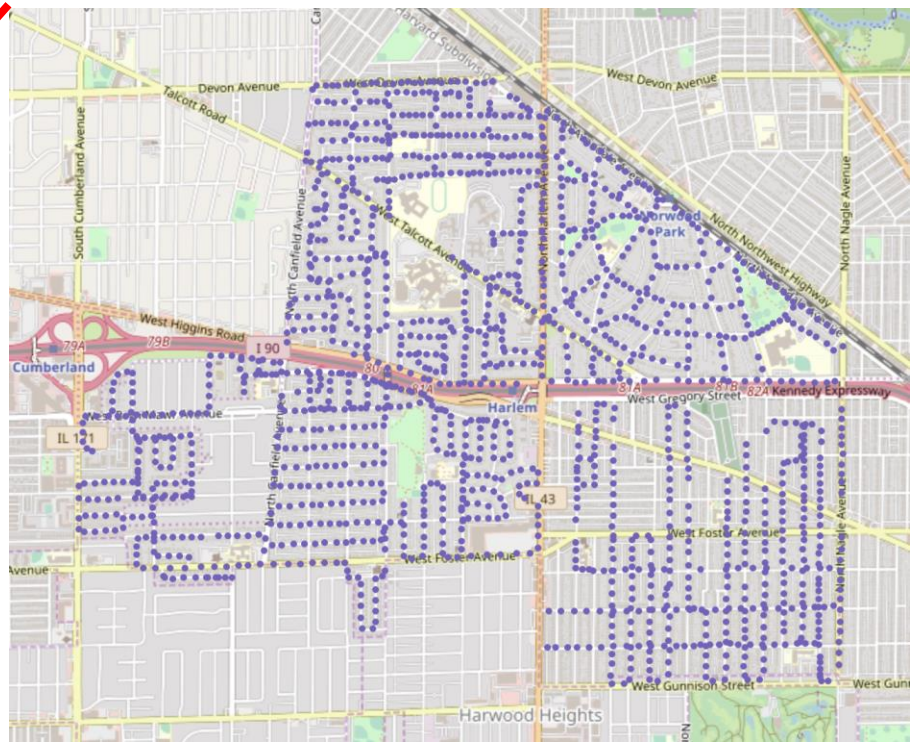


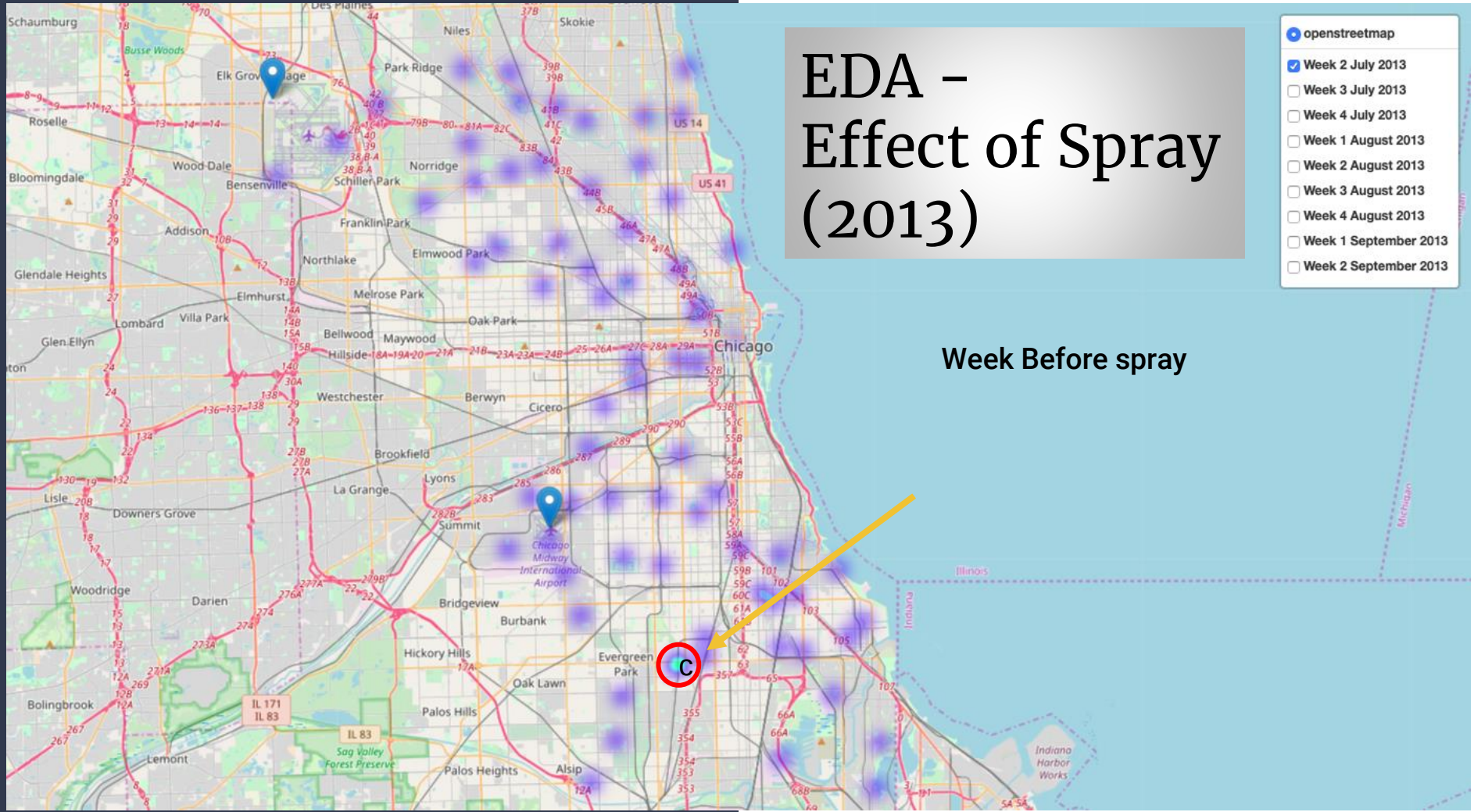
EDA – Trend of R.H.



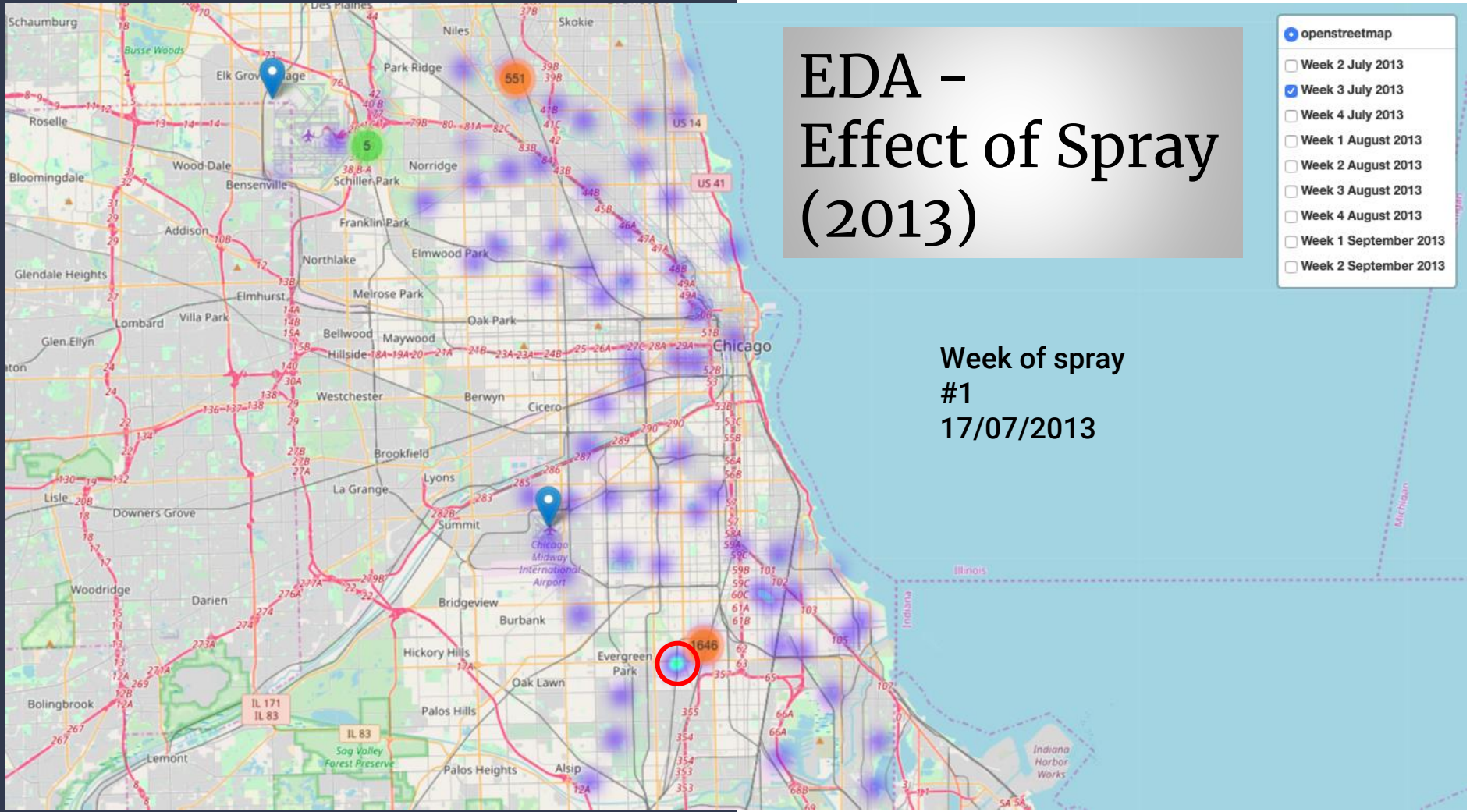
Spray cluster representation:

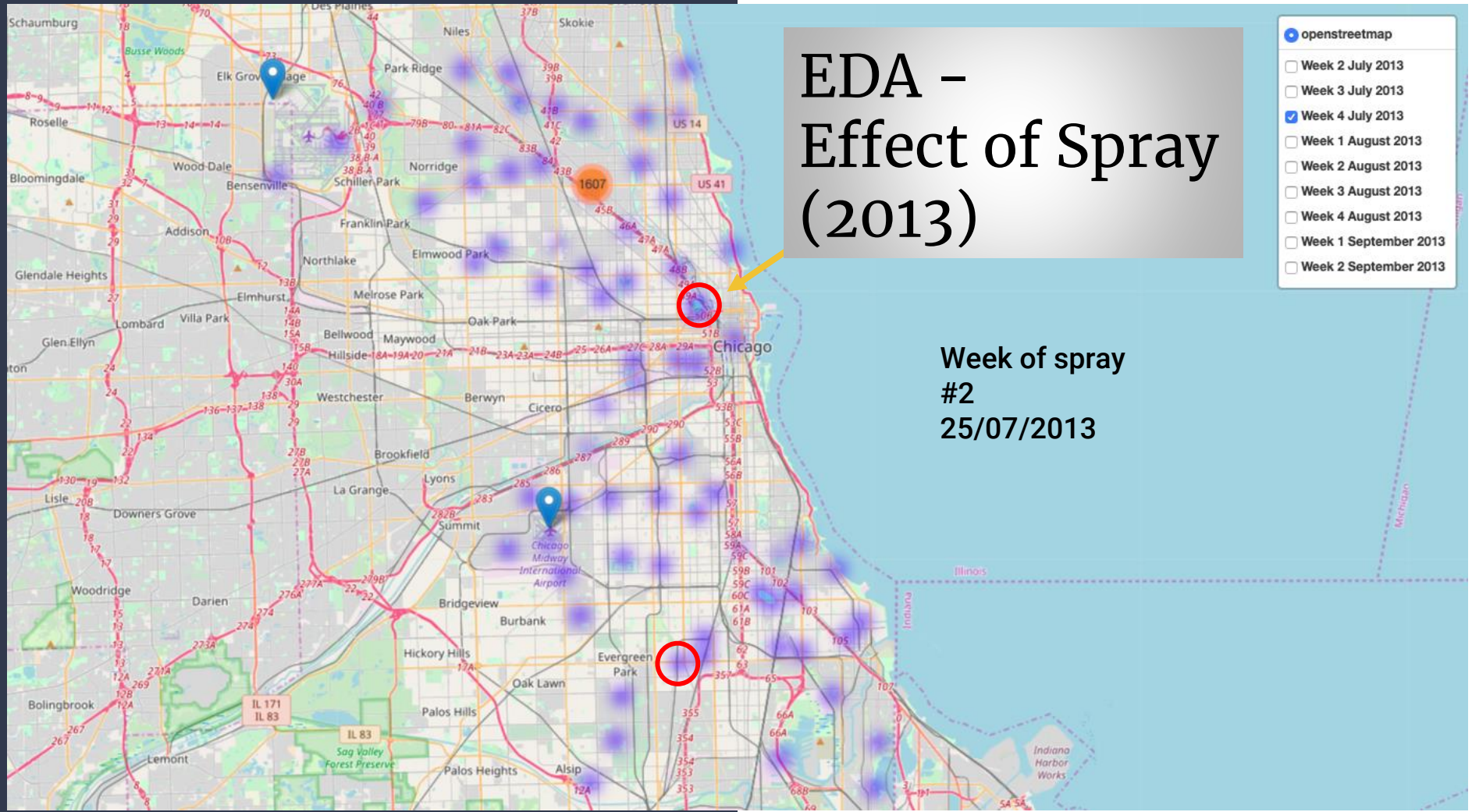


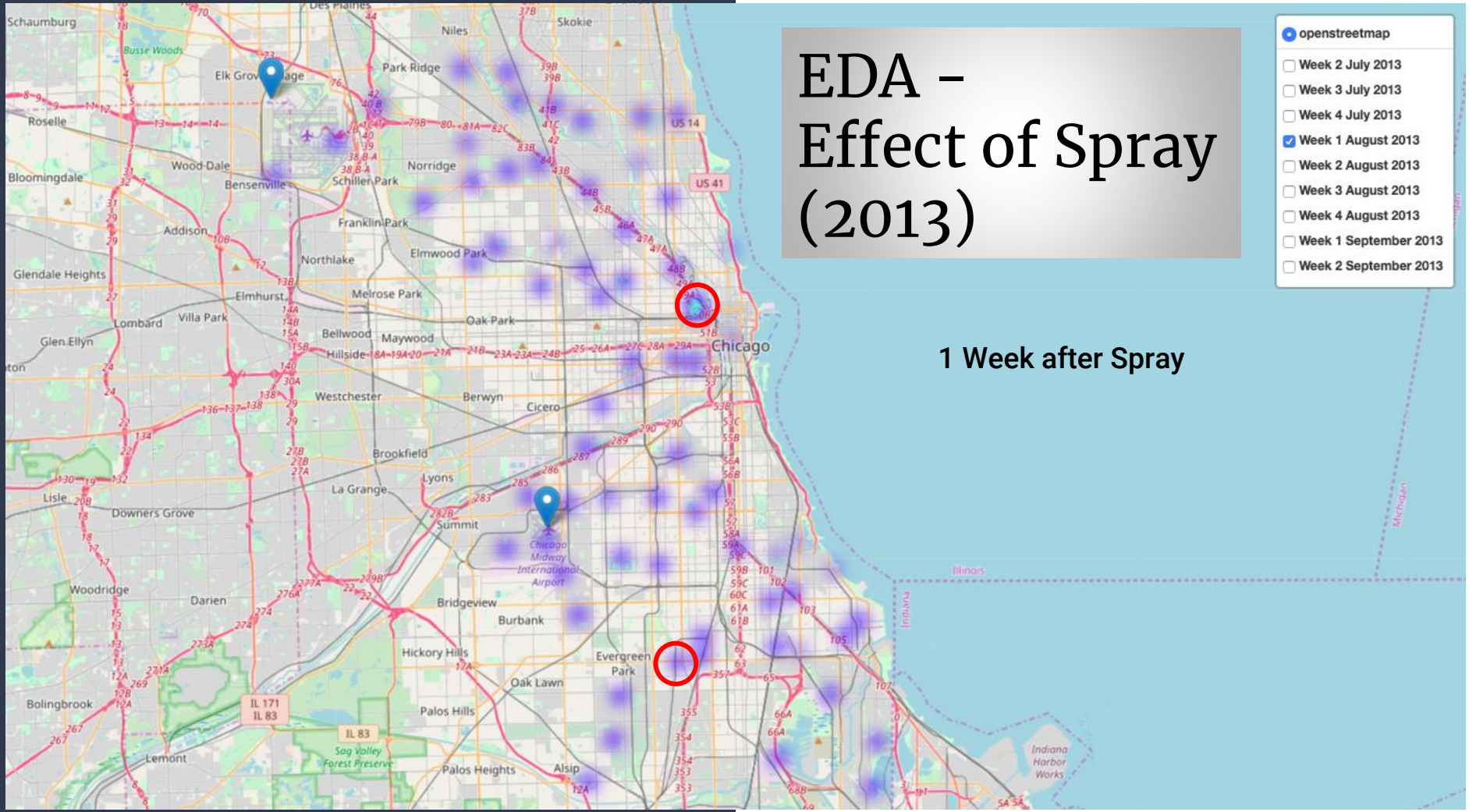


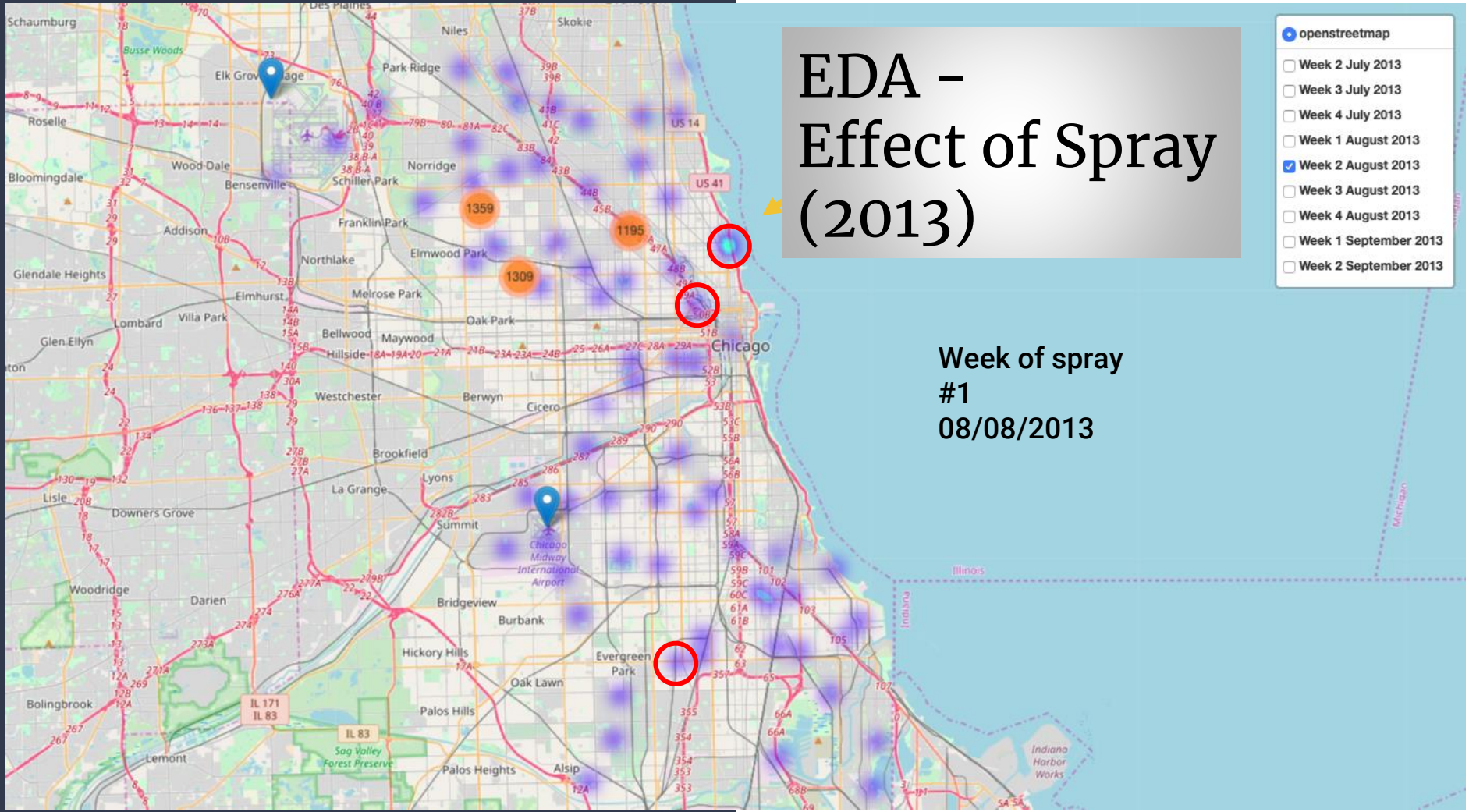

EDA –
Effect of Spray
(2013)
openstreetmap
Week 2 July 2013
Week 3 July 2013
Week 4 July 2013
Week 1 August 2013
Week 2 August 2013
Week 3 August 2013
Week 4 August 2013
Week 1 September 2013
Week 2 September 2013
Week Before spray
C
Chicago
Schaumburg
Niles
Skokie
Park Ridge
Roselle
Bloomingdale
Wood Dale
Bensenville
Norridge
Schiller Park
Addison
Franklin Park
Elmwood Park
Northlake
Glendale Heights
Elmhurst
Melrose Park
Lombard
Villa Park
Oak Park
Bellwood
Maywood
Hillside
Glen Ellyn
Westchester
Berwyn
Cicero
Brookfield
La Grange
Lyons
Lisle
Downers Grove
Summit
Chicago Midway International Airport
Woodridge
Darien
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Bolingbrook
Palos Hills
Lemont
Palos Heights
Alsip
IL 171
IL 83
US 14
US 41
Illinois
Indiana
Michigan
Indiana Harbor Works

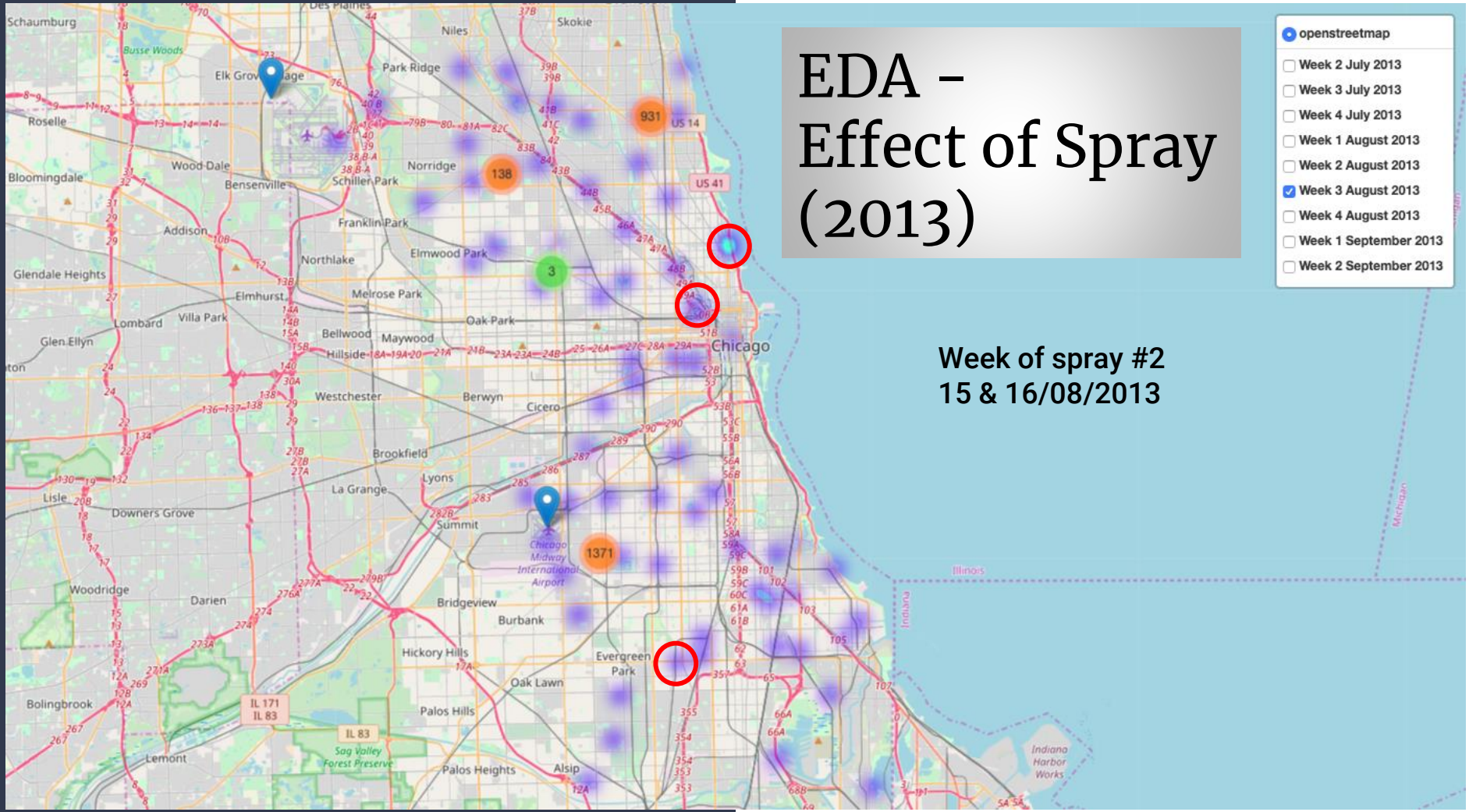
EDA – Effect of Spray (2013)

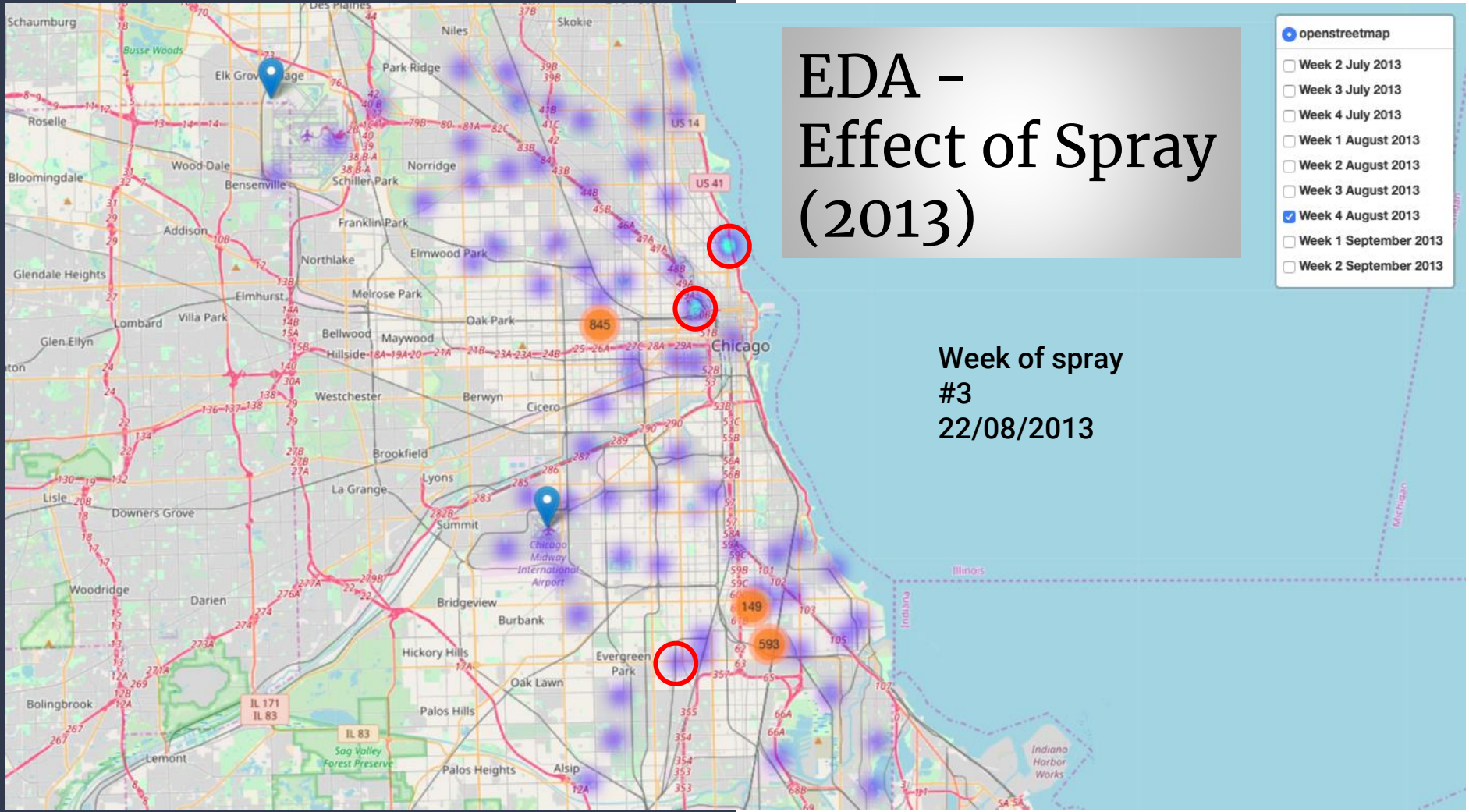


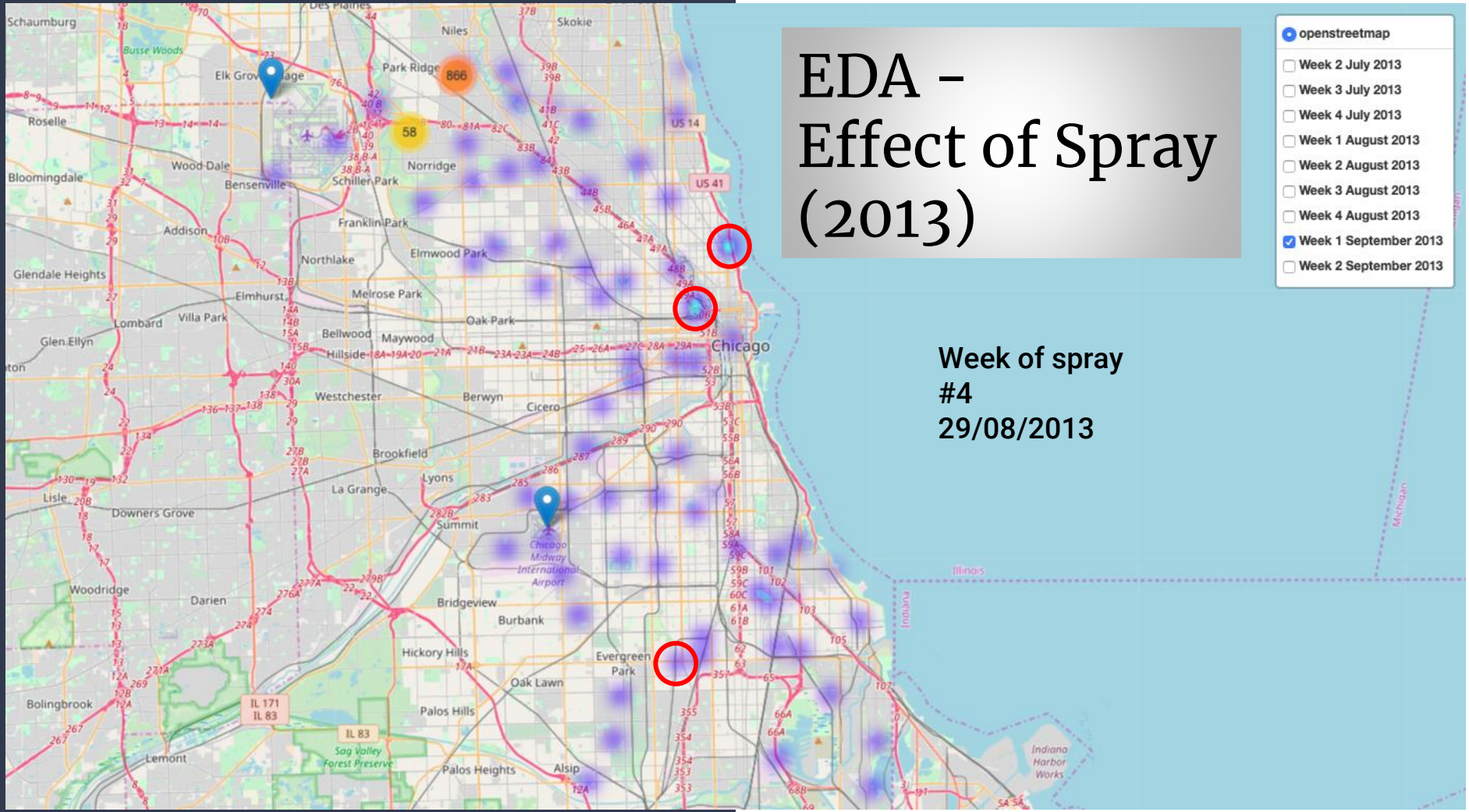

EDA –
Effect of Spray
(2013)
openstreetmap
Week 2 July 2013
Week 3 July 2013
Week 4 July 2013
Week 1 August 2013
Week 2 August 2013
Week 3 August 2013
Week 4 August 2013
Week 1 September 2013
Week 2 September 2013
Week of spray
#2
25/07/2013
Chicago
Oak Park
Cicero
Berwyn
Evergreen Park
Chicago Midway International Airport
Schaumburg
Elk Grove Village
Park Ridge
Niles
Skokie
Roselle
Bloomingdale
Wood Dale
Bensenville
Schiller Park
Norridge
Addison
Franklin Park
Elmwood Park
Northlake
Glendale Heights
Elmhurst
Melrose Park
Lombard
Villa Park
Glen Ellyn
Bellwood
Maywood
Hillside
Westchester
Brookfield
La Grange
Lyons
Summit
Lisle
Downers Grove
Woodridge
Darien
Bridgeview
Burbank
Hickory Hills
Oak Lawn
Palos Hills
Palos Heights
Alsip
Bolingbrook
Lemont
Sag Valley Forest Preserve
US 14
US 41
IL 171
IL 83
1607
Illinois
Indiana
Michigan
Indiana Harbor Works


EDA –
Effect of Spray
(2013)
openstreetmap
Week 2 July 2013
Week 3 July 2013
Week 4 July 2013
Week 1 August 2013
Week 2 August 2013
Week 3 August 2013
Week 4 August 2013
Week 1 September 2013
Week 2 September 2013
1 Week after Spray
Chicago
Chicago Midway International Airport
Schaumburg
Niles
Skokie
Park Ridge
Roselle
Bloomingdale
Wood Dale
Bensenville
Schiller Park
Norridge
Addison
Franklin Park
Northlake
Elmwood Park
Glendale Heights
Elmhurst
Melrose Park
Lombard
Villa Park
Oak Park
Glen Ellyn
Bellwood
Maywood
Hillside
Westchester
Berwyn
Cicero
Brookfield
La Grange
Lyons
Lisle
Downers Grove
Summit
Woodridge
Darien
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Bolingbrook
Palos Hills
Lemont
Palos Heights
Alsip
US 14
US 41
IL 171
IL 83
Illinois
Indiana
Michigan
Indiana Harbor Works

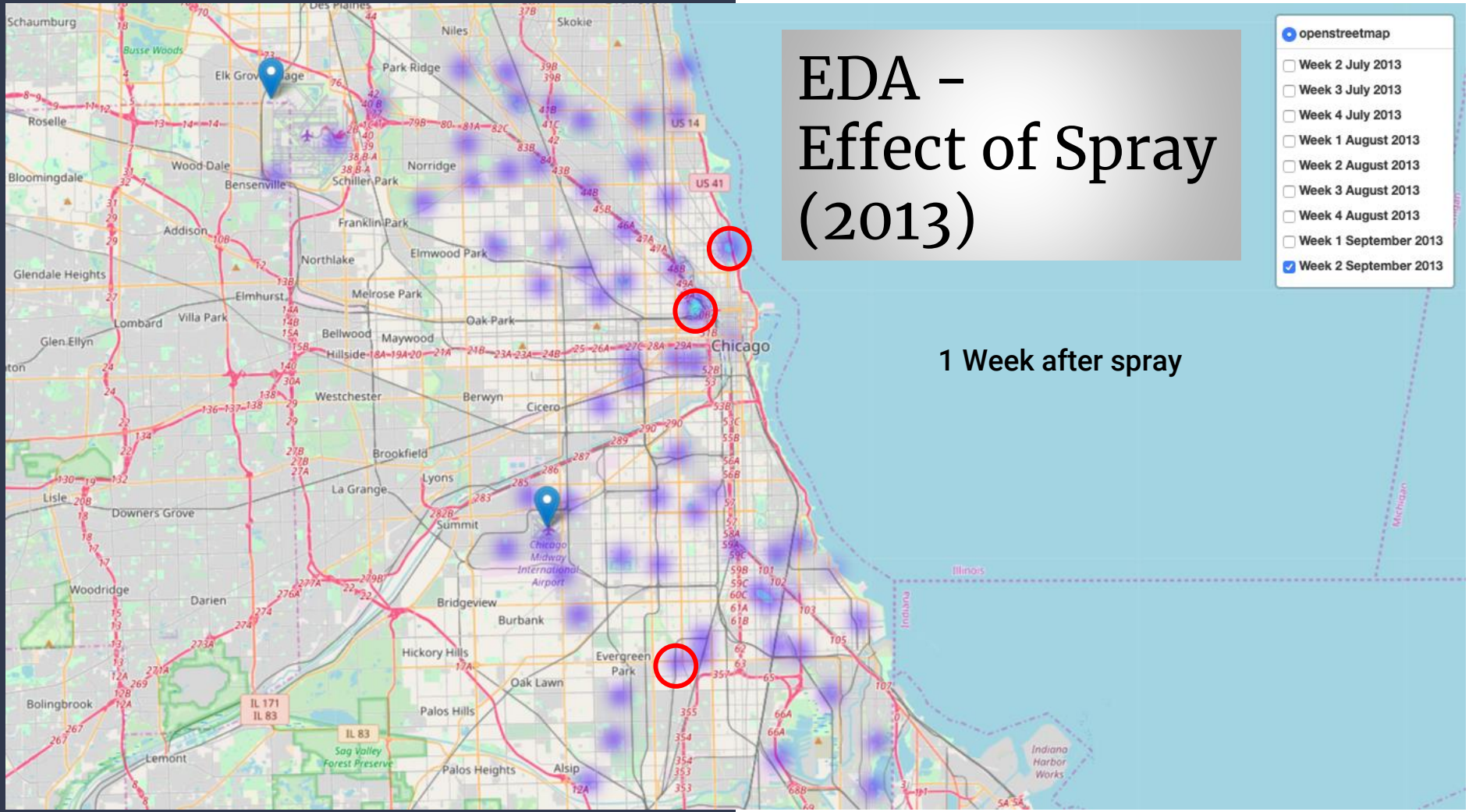

EDA –
Effect of Spray
(2013)
openstreetmap
Week 2 July 2013
Week 3 July 2013
Week 4 July 2013
Week 1 August 2013
Week 2 August 2013
Week 3 August 2013
Week 4 August 2013
Week 1 September 2013
Week 2 September 2013
Week of spray
#1
08/08/2013
1359
1195
1309
Chicago
Schaumburg
Niles
Skokie
Park Ridge
Roselle
Wood Dale
Bloomingdale
Bensenville
Schiller Park
Norridge
Addison
Franklin Park
Northlake
Elmwood Park
Glendale Heights
Elmhurst
Melrose Park
Lombard
Villa Park
Oak Park
Glen Ellyn
Bellwood
Maywood
Hillside
Westchester
Berwyn
Cicero
Brookfield
La Grange
Lyons
Lisle
Downers Grove
Summit
Chicago Midway International Airport
Woodridge
Darien
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Bolingbrook
Palos Hills
Lemont
Palos Heights
Alsip
US 14
US 41
IL 171
IL 83
Illinois
Indiana
Michigan
Indiana Harbor Works


EDA –
Effect of Spray
(2013)
openstreetmap
Week 2 July 2013
Week 3 July 2013
Week 4 July 2013
Week 1 August 2013
Week 2 August 2013
Week 3 August 2013
Week 4 August 2013
Week 1 September 2013
Week 2 September 2013
Week of spray #2
15 & 16/08/2013
Chicago
Chicago Midway International Airport
931
138
3
1371
US 14
US 41
IL 171
IL 83
Schaumburg
Niles
Skokie
Park Ridge
Roselle
Bloomingdale
Wood Dale
Bensenville
Schiller Park
Norridge
Addison
Franklin Park
Northlake
Elmwood Park
Glendale Heights
Elmhurst
Melrose Park
Lombard
Villa Park
Oak Park
Glen Ellyn
Bellwood
Maywood
Hillside
Westchester
Berwyn
Cicero
Brookfield
Lyons
La Grange
Lisle
Downers Grove
Summit
Woodridge
Darien
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Bolingbrook
Palos Hills
Lemont
Palos Heights
Alsip
Illinois
Indiana
Michigan
Indiana Harbor Works


EDA –
Effect of Spray
(2013)
openstreetmap
Week 2 July 2013
Week 3 July 2013
Week 4 July 2013
Week 1 August 2013
Week 2 August 2013
Week 3 August 2013
Week 4 August 2013
Week 1 September 2013
Week 2 September 2013
Week of spray
#3
22/08/2013
845
149
593
Chicago
Schaumburg
Niles
Skokie
Park Ridge
Roselle
Bloomingdale
Wood Dale
Bensenville
Schiller Park
Norridge
Addison
Franklin Park
Northlake
Elmwood Park
Glendale Heights
Elmhurst
Melrose Park
Lombard
Villa Park
Oak Park
Glen Ellyn
Bellwood
Maywood
Hillside
Westchester
Berwyn
Cicero
Brookfield
La Grange
Lyons
Lisle
Downers Grove
Summit
Chicago Midway International Airport
Woodridge
Darien
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Bolingbrook
Palos Hills
Lemont
Palos Heights
Alsip
IL 171
IL 83
US 14
US 41
Illinois
Indiana
Michigan
Indiana Harbor Works


EDA –
Effect of Spray
(2013)
openstreetmap
Week 2 July 2013
Week 3 July 2013
Week 4 July 2013
Week 1 August 2013
Week 2 August 2013
Week 3 August 2013
Week 4 August 2013
Week 1 September 2013
Week 2 September 2013
Week of spray
#4
29/08/2013
Chicago
Schaumburg
Niles
Skokie
Park Ridge
Elk Grove Village
Roselle
Bloomingdale
Wood Dale
Bensenville
Schiller Park
Norridge
Addison
Franklin Park
Northlake
Elmwood Park
Glendale Heights
Elmhurst
Melrose Park
Lombard
Villa Park
Oak Park
Glen Ellyn
Bellwood
Maywood
Hillside
Westchester
Berwyn
Cicero
Brookfield
La Grange
Lyons
Lisle
Downers Grove
Summit
Chicago Midway International Airport
Woodridge
Darien
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Bolingbrook
Palos Hills
Lemont
Palos Heights
Alsip
Indiana Harbor Works
Illinois
Indiana
Michigan
US 14
US 41
IL 171
IL 83
866
58

EDA – Effect of Spray (2013)

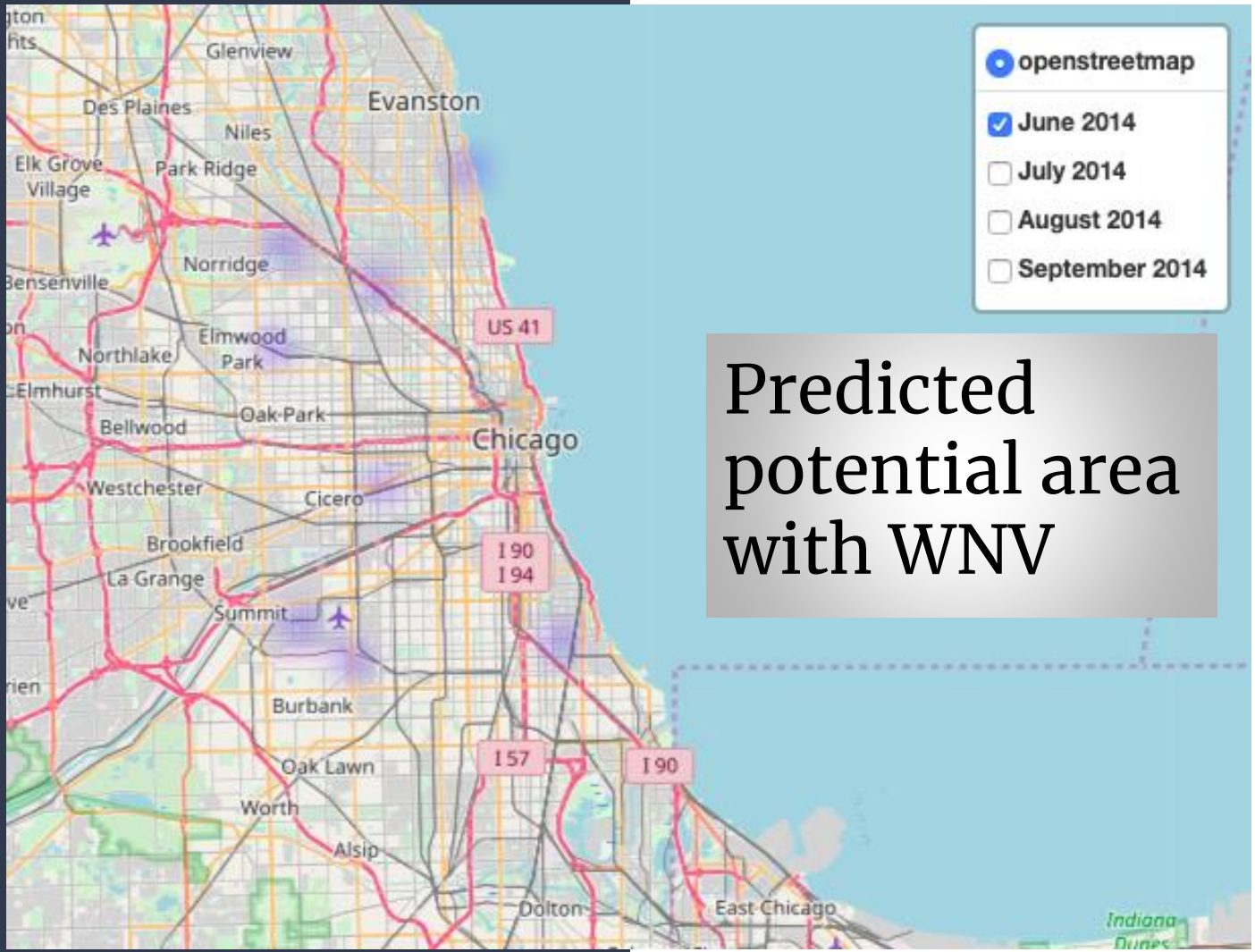


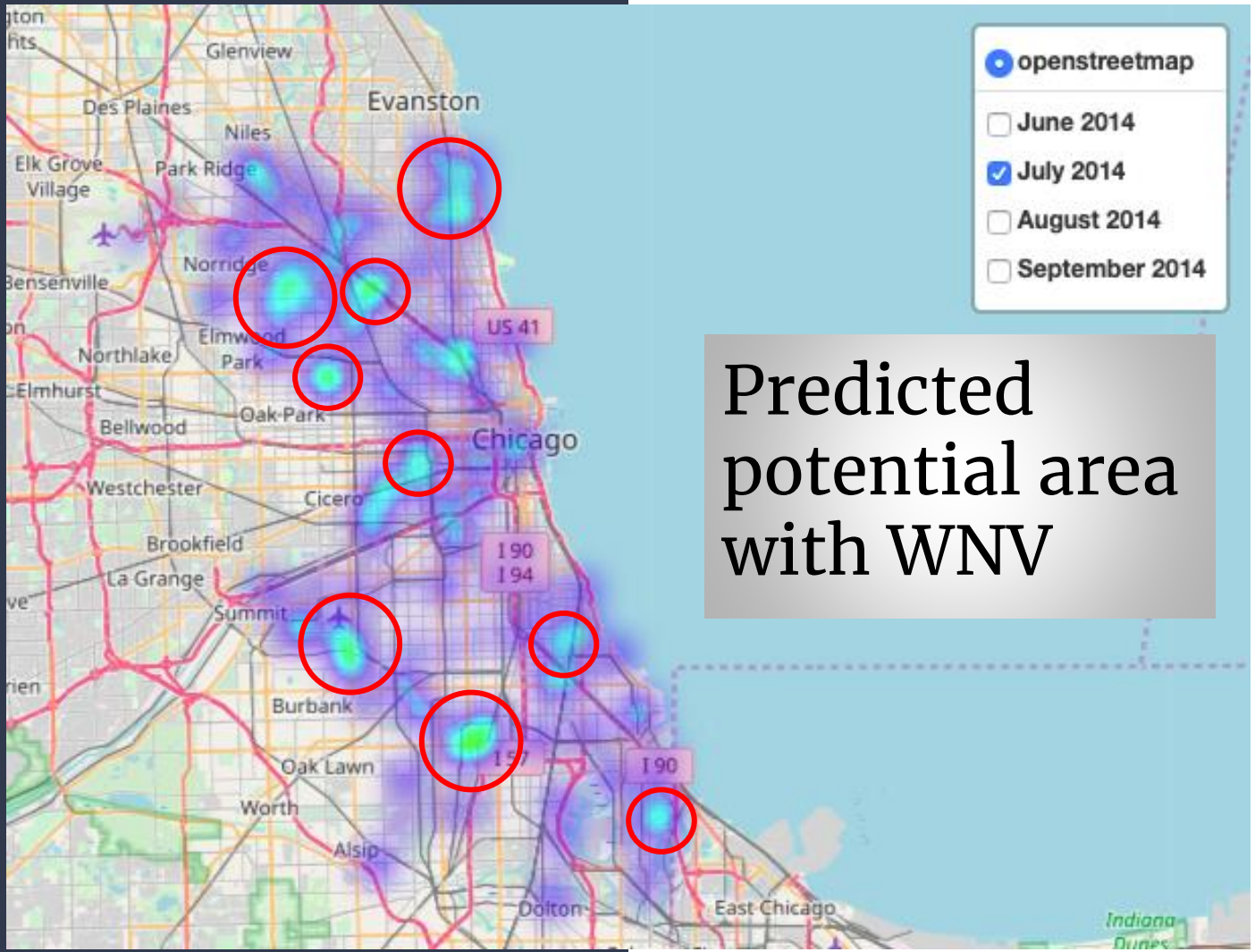
Model Selection & Evaluation

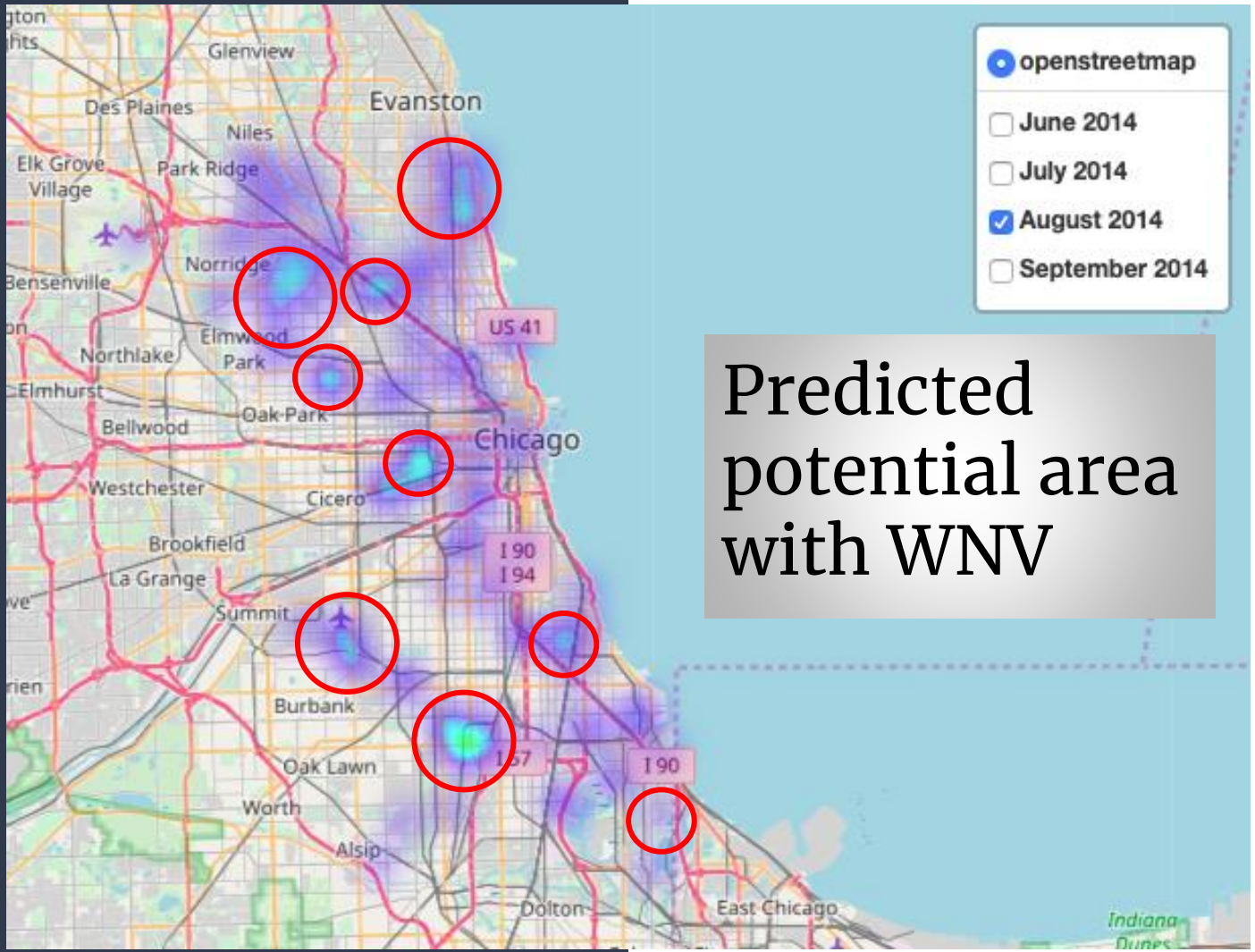
```
0     8077
1      457
Name: wnvpresent
```

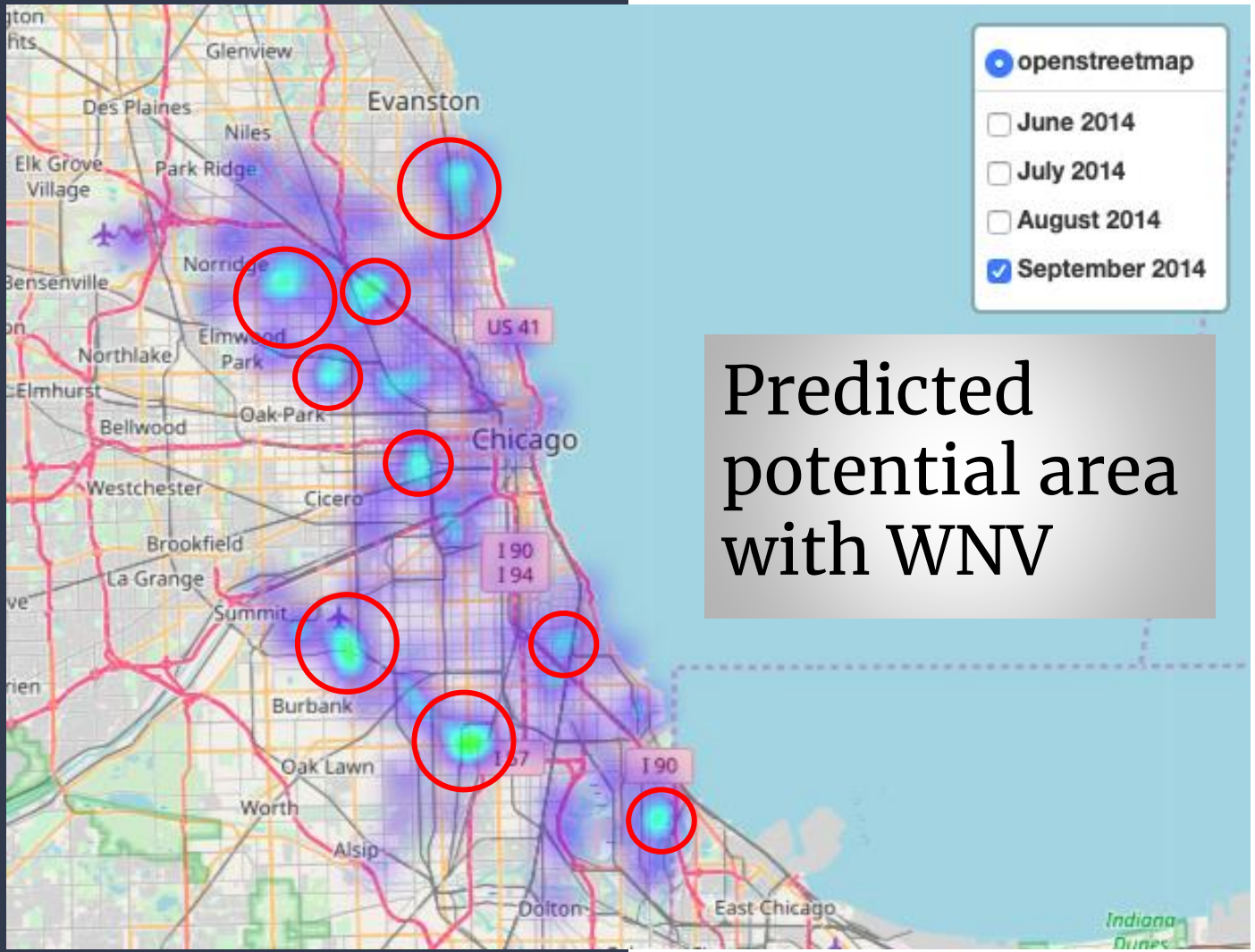
```
1     0.5
0     0.5
Name: wnvpresent
```

Model	ROC-AUC Score
Decision Tree	0.869
Random Forest Classifier	0.947
Ada Boost	0.949
Gradient Boost	0.905
Support Vector Machine	0.626


openstreetmap
June 2014
July 2014
August 2014
September 2014
Predicted potential area with WNV
Glenview
Evanston
Des Plaines
Niles
Elk Grove Village
Park Ridge
Norridge
Bensenville
Elmwood Park
Northlake
Elmhurst
Oak Park
Bellwood
Chicago
Westchester
Cicero
Brookfield
La Grange
Summit
Burbank
Oak Lawn
Worth
Alsip
Dolton
East Chicago
US 41
I 90
I 94
I 57
I 90


openstreetmap
June 2014
July 2014
August 2014
September 2014
Predicted potential area with WNV
Glenview
Evanston
Des Plaines
Niles
Elk Grove Village
Park Ridge
Norridge
Bensenville
Elmwood Park
Northlake
Elmhurst
Oak Park
Bellwood
Chicago
Westchester
Cicero
Brookfield
La Grange
Summit
Burbank
Oak Lawn
Worth
Alsip
Dolton
East Chicago
US 41
I 90
I 94
I 57
I 90


openstreetmap
June 2014
July 2014
August 2014
September 2014
Predicted potential area with WNV
Glenview
Des Plaines
Evanston
Niles
Elk Grove Village
Park Ridge
Norridge
Bensenville
Northlake
Elmwood Park
Elmhurst
Oak Park
Bellwood
Chicago
Westchester
Cicero
Brookfield
La Grange
Summit
Burbank
Oak Lawn
Worth
Alsip
Dolton
East Chicago
US 41
I 90
I 94
I 57


openstreetmap
June 2014
July 2014
August 2014
September 2014
Predicted potential area with WNV
Glenview
Evanston
Des Plaines
Niles
Elk Grove Village
Park Ridge
Norridge
Bensenville
Elmwood Park
Northlake
Elmhurst
Oak Park
Bellwood
Chicago
Westchester
Cicero
Brookfield
La Grange
Summit
Burbank
Oak Lawn
Worth
Alsip
Dolton
East Chicago
US 41
I 90
I 94
I 90
Indiana Dunes

Cost Benefit Analysis

Cost

- Cost of spray/vector control
- Indirect Costs (hospitalisations, lost productivity)

Benefit

- Economic Impact (on health, productivity)

Case Study

Table 2

Estimated inpatient and outpatient economic costs of WNND cases, Sacramento County, California, 2005*

Item	Cost per case†	No. cases to which cost applies‡	% Cases to which cost applies§	Total cost for all cases	Total cost if treatment/service were used in all cases
Inpatient treatment costs	**$33,143**	**46**	**100**	**$1,524,570**	**$1,524,570**
Outpatient costs	Cost per case¶				
Outpatient hospital treatment	$333	17	36	$5,668	$15,337
Physician visits	$450	46	100	$20,708	$20,708
Outpatient physical therapy	$909	46	100	$41,810	$41,810
Occupational therapy	$4,037	3	7	$12,111	$185,699
Speech therapy	$588	1	1	$588	$27,032
Total				$80,885	$290,586
Nursing home costs	Cost#				
Nursing home stay**	$190	2	4	$36,195	$36,195
Transportation	$65	46	100	$2,977	$2,977
Home health aides, babysitters, etc.	$1,569	7	14	$10,983	$505,211
Total				$50,154	$544,383
Total for WNND				$2,140,409	$2,844,339

In 2005, an outbreak of West Nile virus (WNV) disease occurred in Sacramento County, California.

163 Human Cases Reported

Economic Impact of the Outbreak

Vector Control Costs
$701,790 (~477m^2)

Medical Treatment & Productivity Lost
$2.28 Mil

Cost Benefit Analysis

Total economic impact: $2.98 million.
Only 15 prevented cases of West Nile neuroinvasive disease to make the emergency spray cost-effective.

Source: National Center Biotechnology Information

Conclusion & Recommendation

- Jul and Aug as the target month for vector control
- To monitor the predicted 9 clusters in 2014 as areas with a higher risk of WNV outbreak.
- The cost of vector control should also factor other trade-offs such impact on health, medical bills and loss in productivity in order to have a better sense of the overall economic cost
- Future work would include the use of medical data and the costs to have a more complete analysis.