Project 4: West Nile Virus

Members : Edward, Derrick and Tresha

Content:

- Introduction and Problem Statement
- Exploratory Data Analysis
- Modelling
- Cost-Benefit Analysis
- Executive Summary

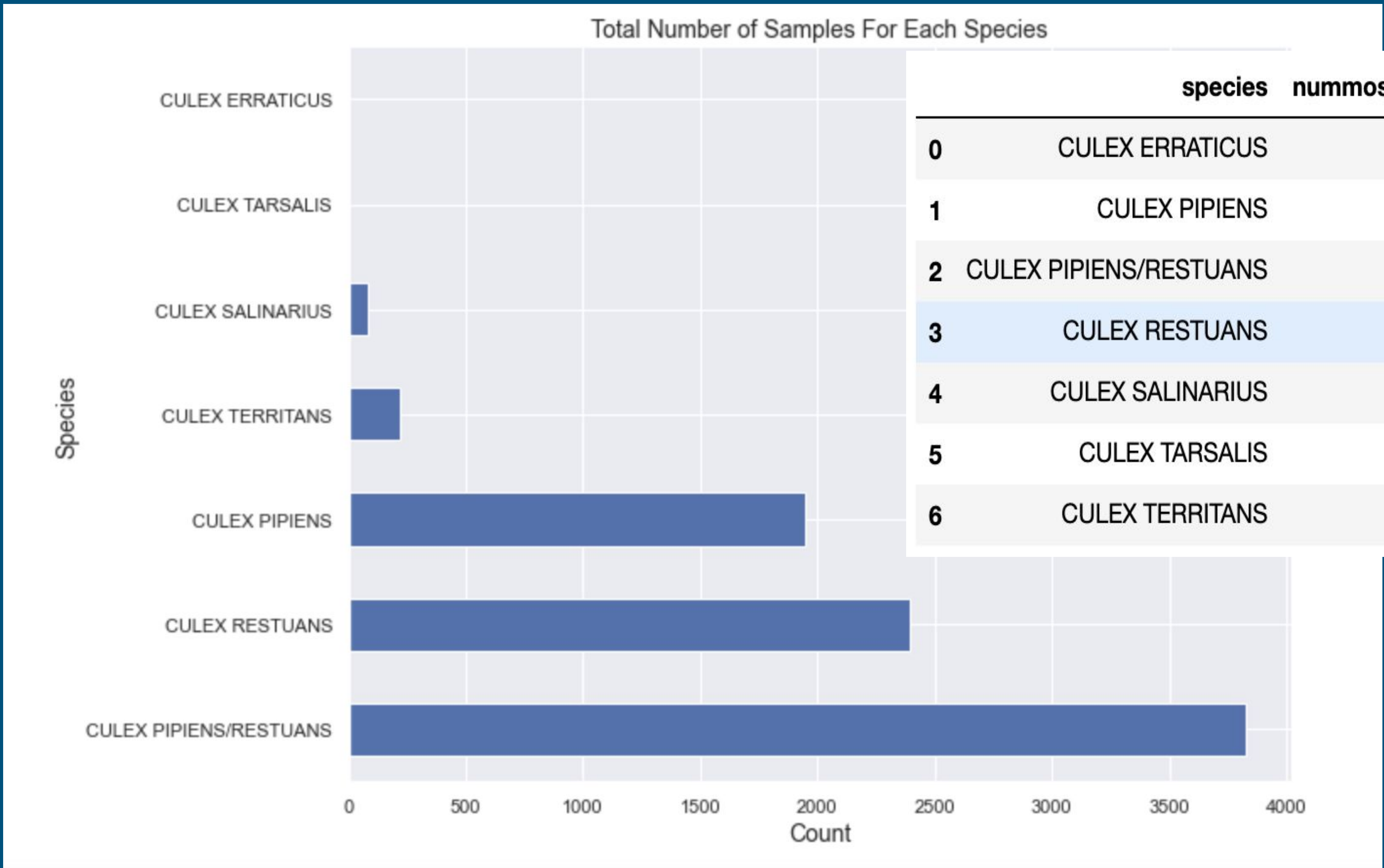
Introduction

- West Nile Virus is a leading mosquito-borne disease
- Cases occur mainly during summer and fall
- No vaccines
- Recent epidemic of West Nile Virus in the Windy City

Problem Statement

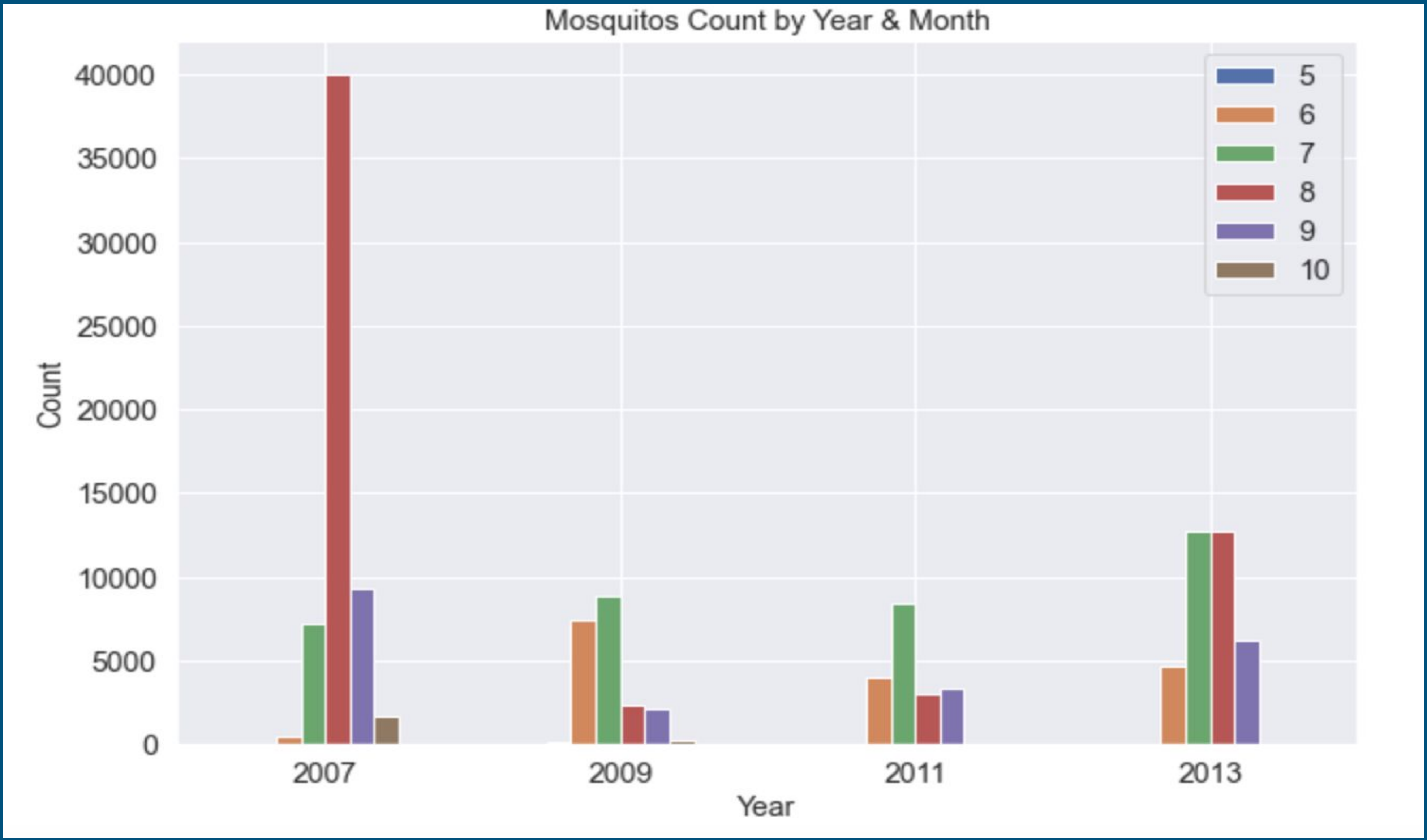
- We are from Disease and Treatment Agency
- To build a classifier model
- Deploy pesticides after a cost-benefit analysis

Mosquitos



	species	nummosquitos	wnvpresent
0	CULEX ERRATICUS	7	0
1	CULEX PIPIENS	44671	184
2	CULEX PIPIENS/RESTUANS	66268	225
3	CULEX RESTUANS	23431	48
4	CULEX SALINARIUS	145	0
5	CULEX TARSALIS	7	0
6	CULEX TERRITANS	510	0

Mosquitos



Mosquitos Count by Year & Month
40000
35000
30000
25000
20000
15000
10000
5000
0
Count
2007
2009
2011
2013
Year
5
6
7
8
9
10

Modelling

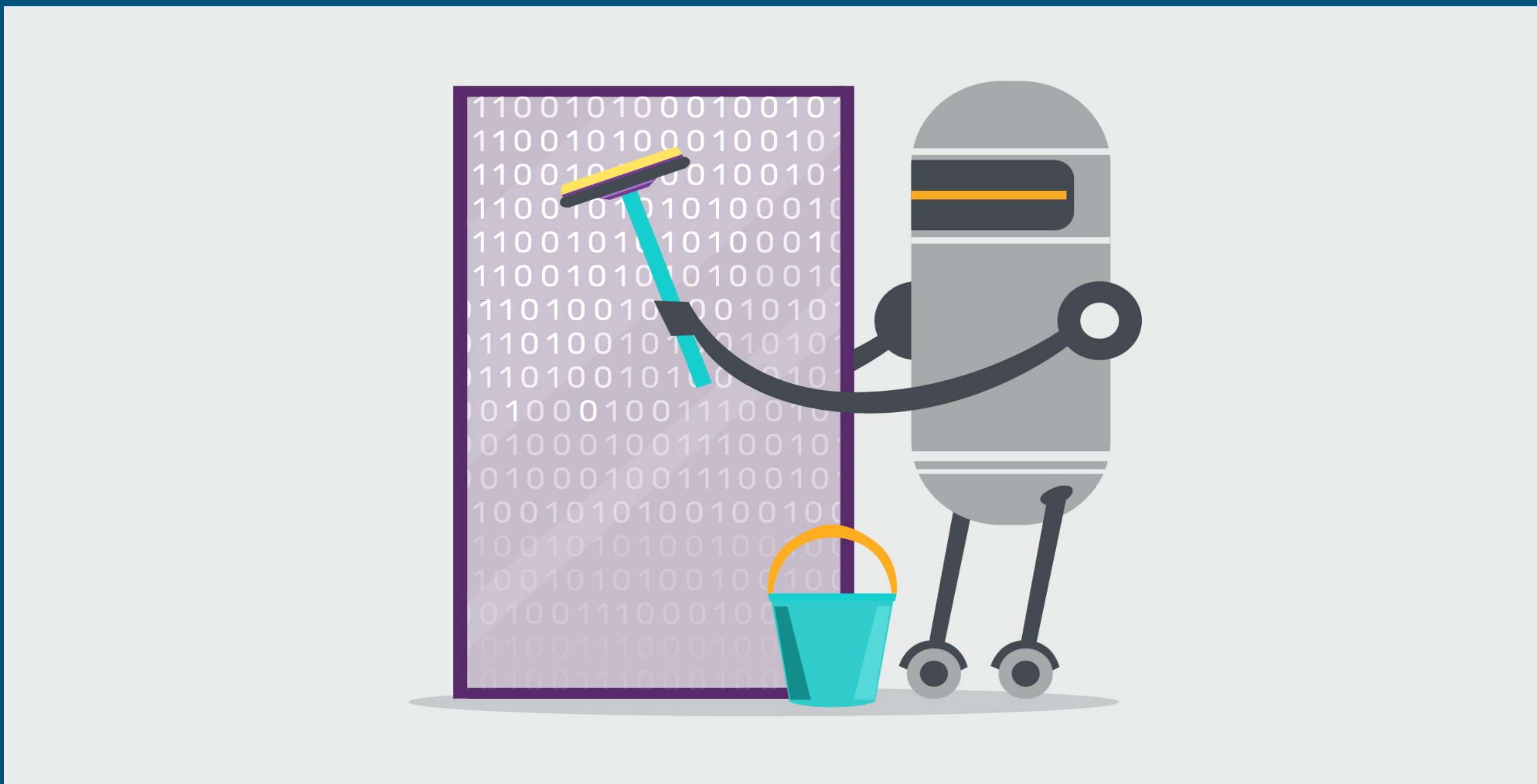
- K Nearest Neighbours
- Support Vector Machines
- Logistic Regression
- Random Forests
- Extra Trees

Modelling Summary

Model	Train Accuracy	Test Accuracy	Sensitivity	Baseline acc	True Pos	False Pos	False Neg	True Pos
KNearest Neighbours	1.0	0.9476	0.0563	0.0539	5284	8	285	17
Support Vector Machine	0.9471	0.9458	0.0033	0.0539	5290	8	301	1
Logistic Regression	0.9488	0.9480	0.0762	0.0539	5280	12	279	23
Random Forest	0.9957	0.9544	0.4172	0.0539	5213	79	176	126
Extra Trees	0.9521	0.9097	0.6126	0.0539	4904	388	117	185

Random Forests was chosen as our model

Data Cleaning



Data Cleaning

1) Dropping Unnecessary columns
2) Changing columns to DateTime format
3) Merging cells with hit max cap for mosquito count
4) Imputing missing values
5) Changing variables to appropriate data types

Exploratory Data Analysis

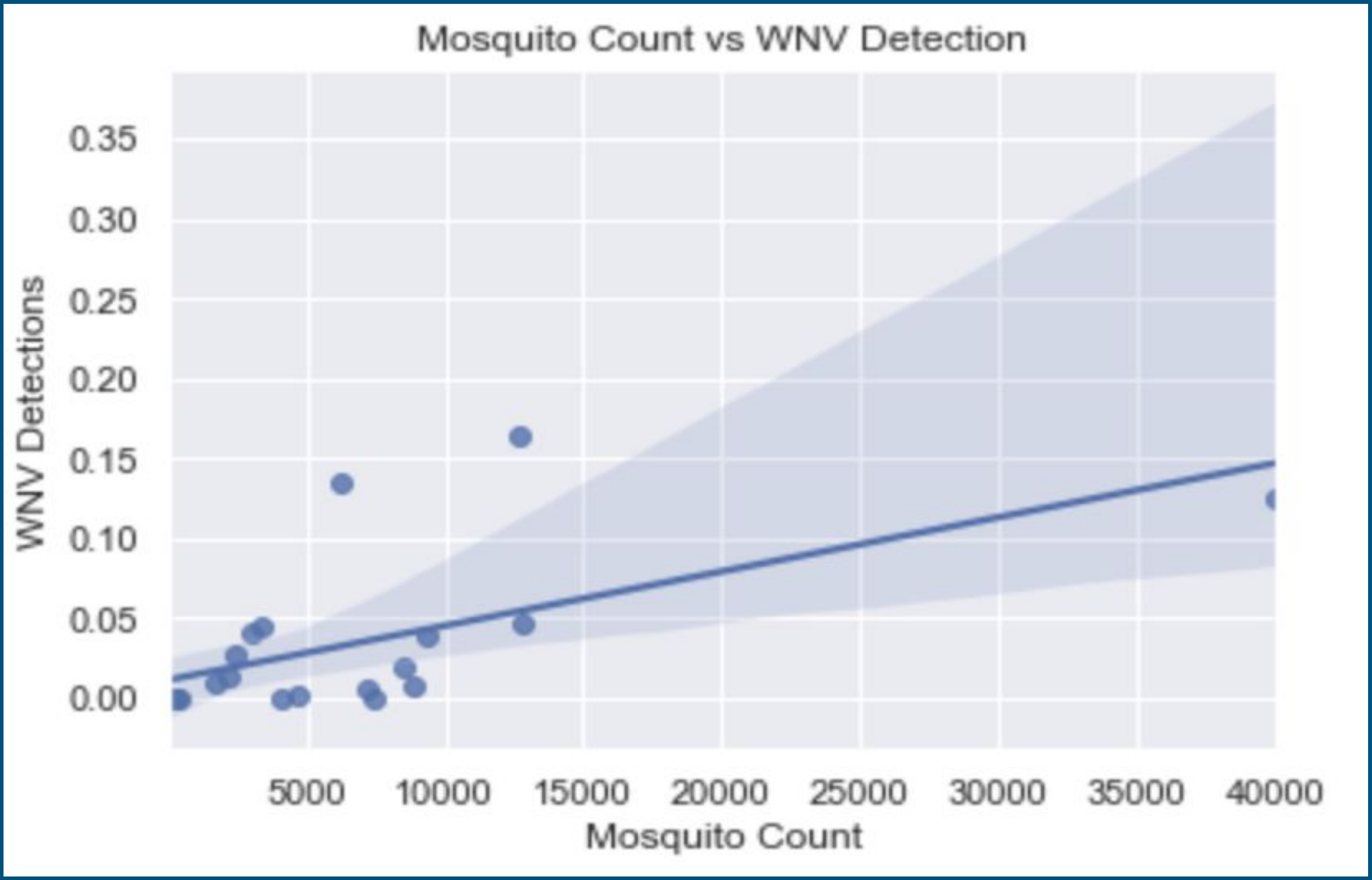
Traps

Satellite Traps

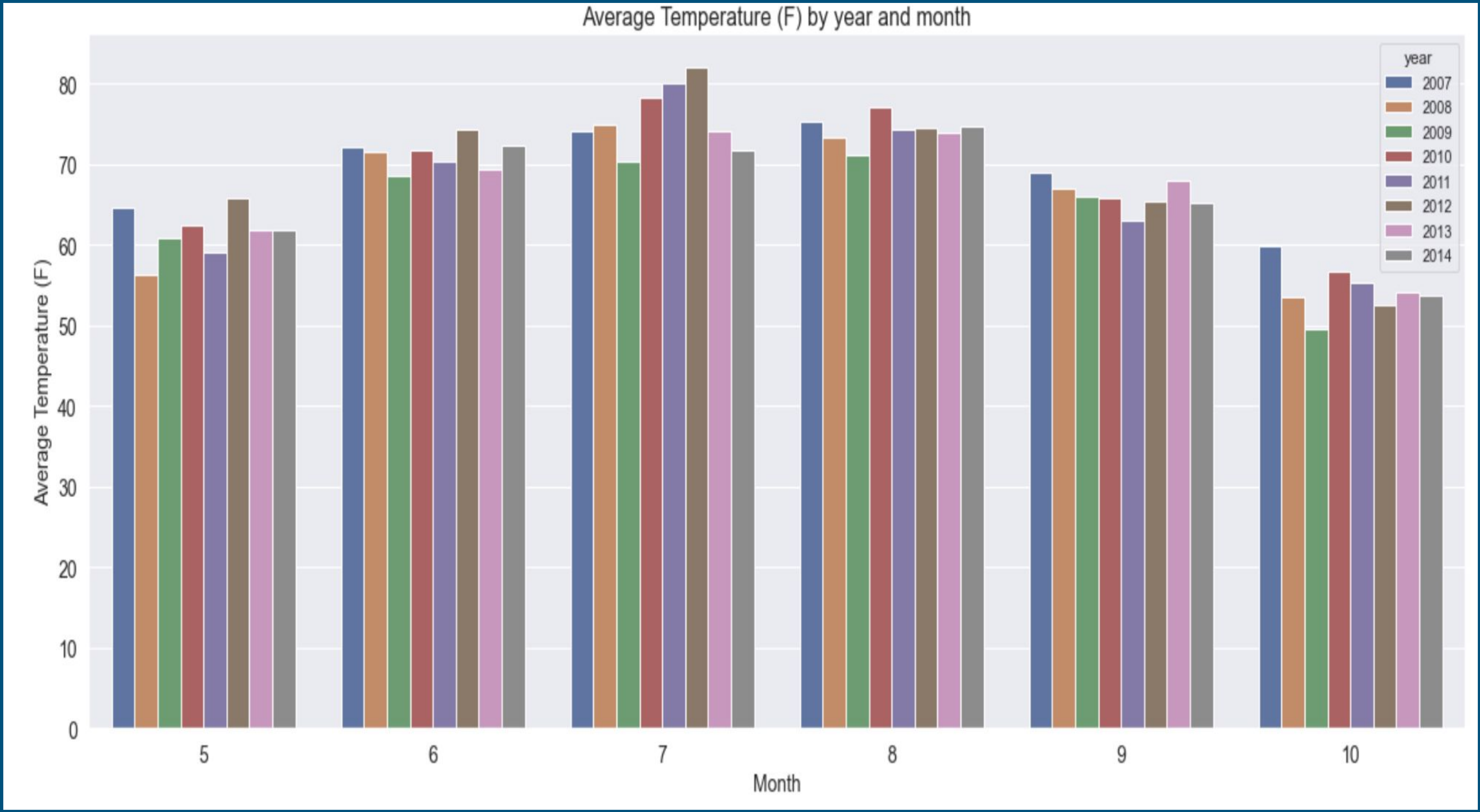
	species	block	trap	latitude	longitude	nummosquitos	wnvpresent	year	month	day
11	CULEX PIPIENS/RESTUANS	22	T054	41.921965	-87.632085	2	0	2007	5	29
12	CULEX RESTUANS	22	T054	41.921965	-87.632085	3	0	2007	5	29
40	CULEX RESTUANS	22	T054	41.921965	-87.632085	3	0	2007	6	5
159	CULEX PIPIENS/RESTUANS	22	T054	41.921965	-87.632085	3	0	2007	6	29
211	CULEX PIPIENS/RESTUANS	22	T054	41.921965	-87.632085	2	0	2007	7	2

	species	block	trap	latitude	longitude	nummosquitos	wnvpresent	year	month	day
9552	CULEX PIPIENS/RESTUANS	21	T054C	41.925652	-87.63359	47	0	2013	8	8
9709	CULEX PIPIENS	21	T054C	41.925652	-87.63359	45	1	2013	8	15
9708	CULEX PIPIENS/RESTUANS	21	T054C	41.925652	-87.63359	15	0	2013	8	15
9876	CULEX PIPIENS	21	T054C	41.925652	-87.63359	19	0	2013	8	22
9875	CULEX PIPIENS/RESTUANS	21	T054C	41.925652	-87.63359	11	0	2013	8	22

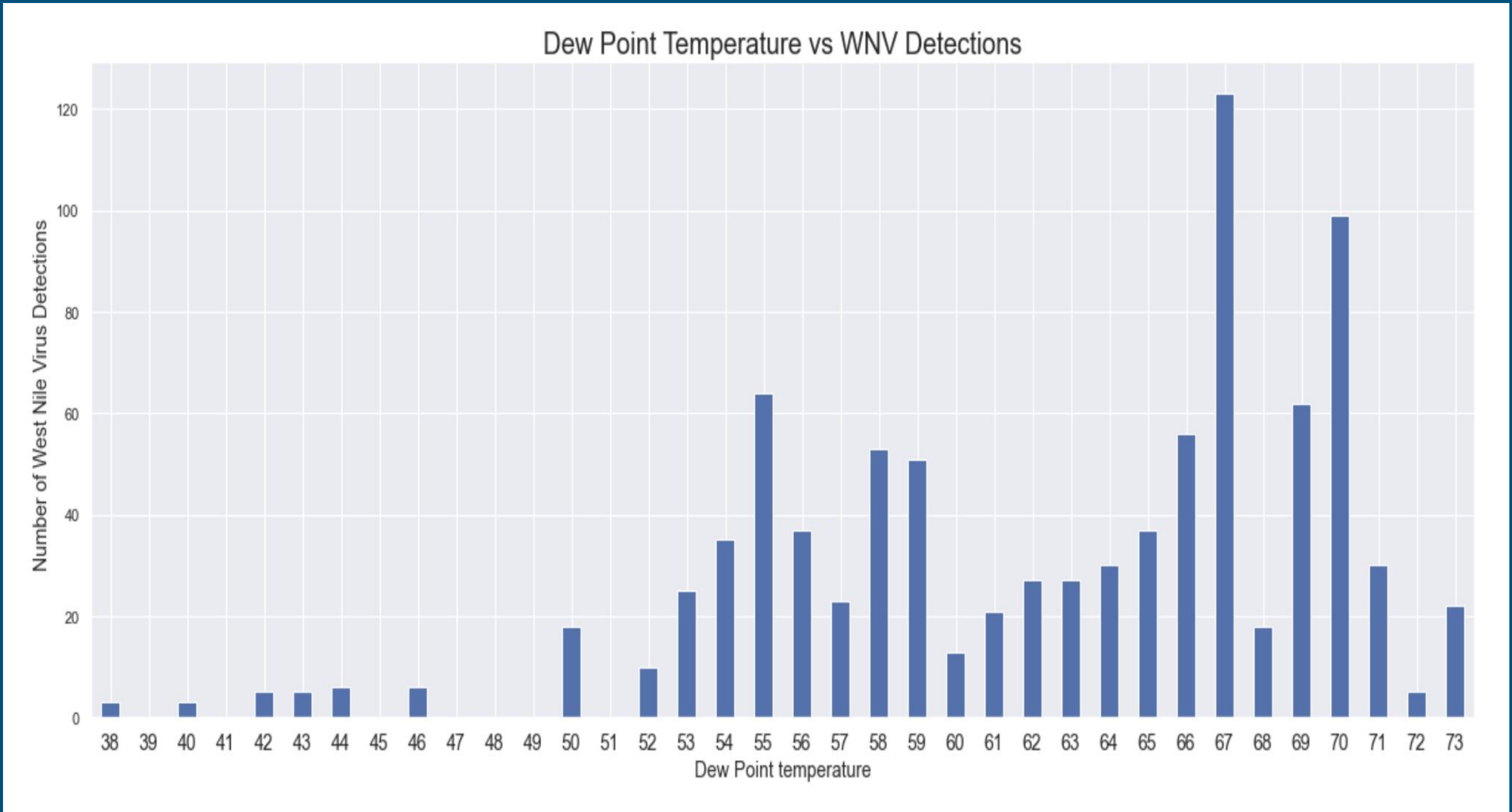
Mosquito Count vs WNV Detection



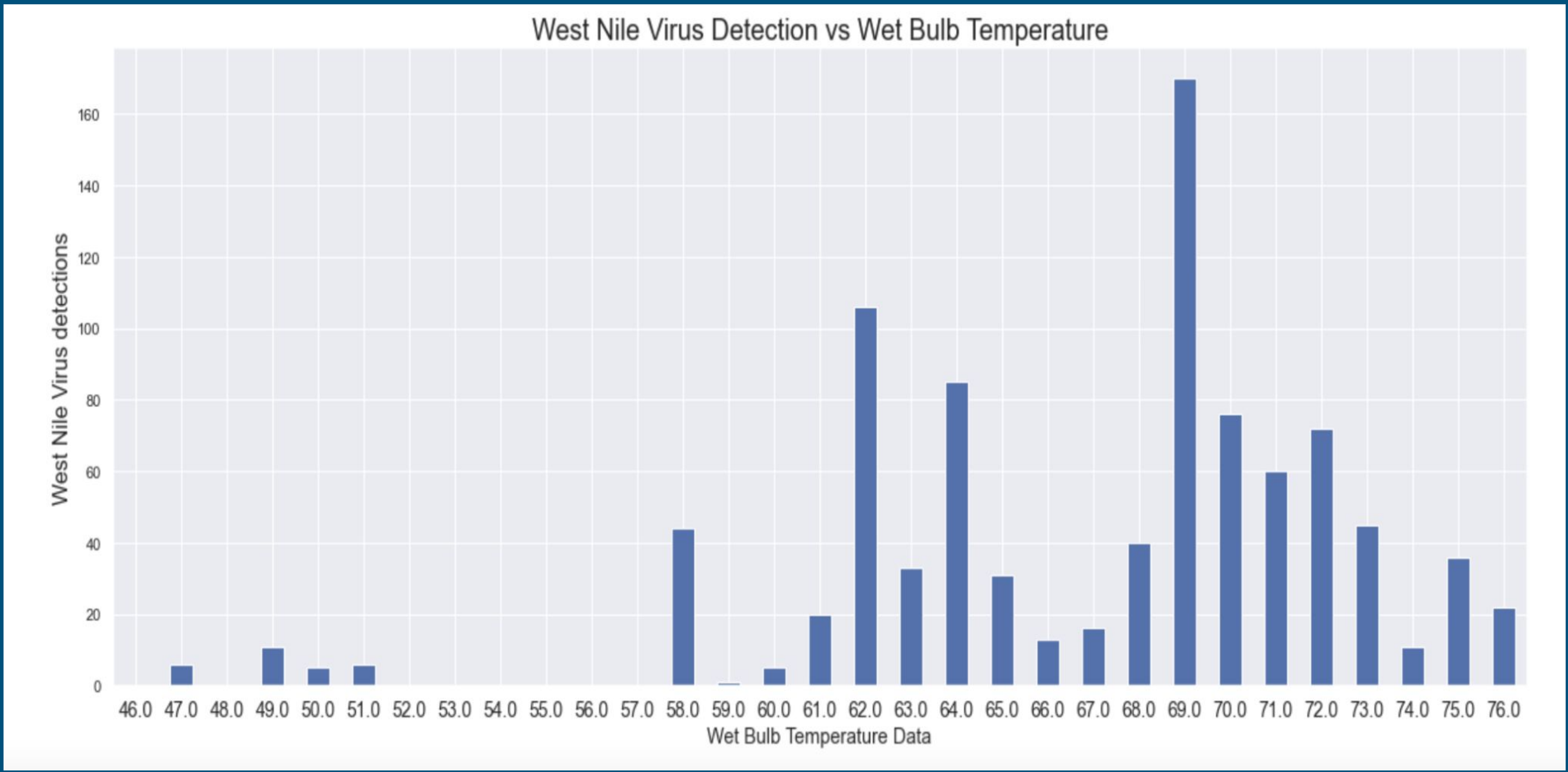
Temperatures



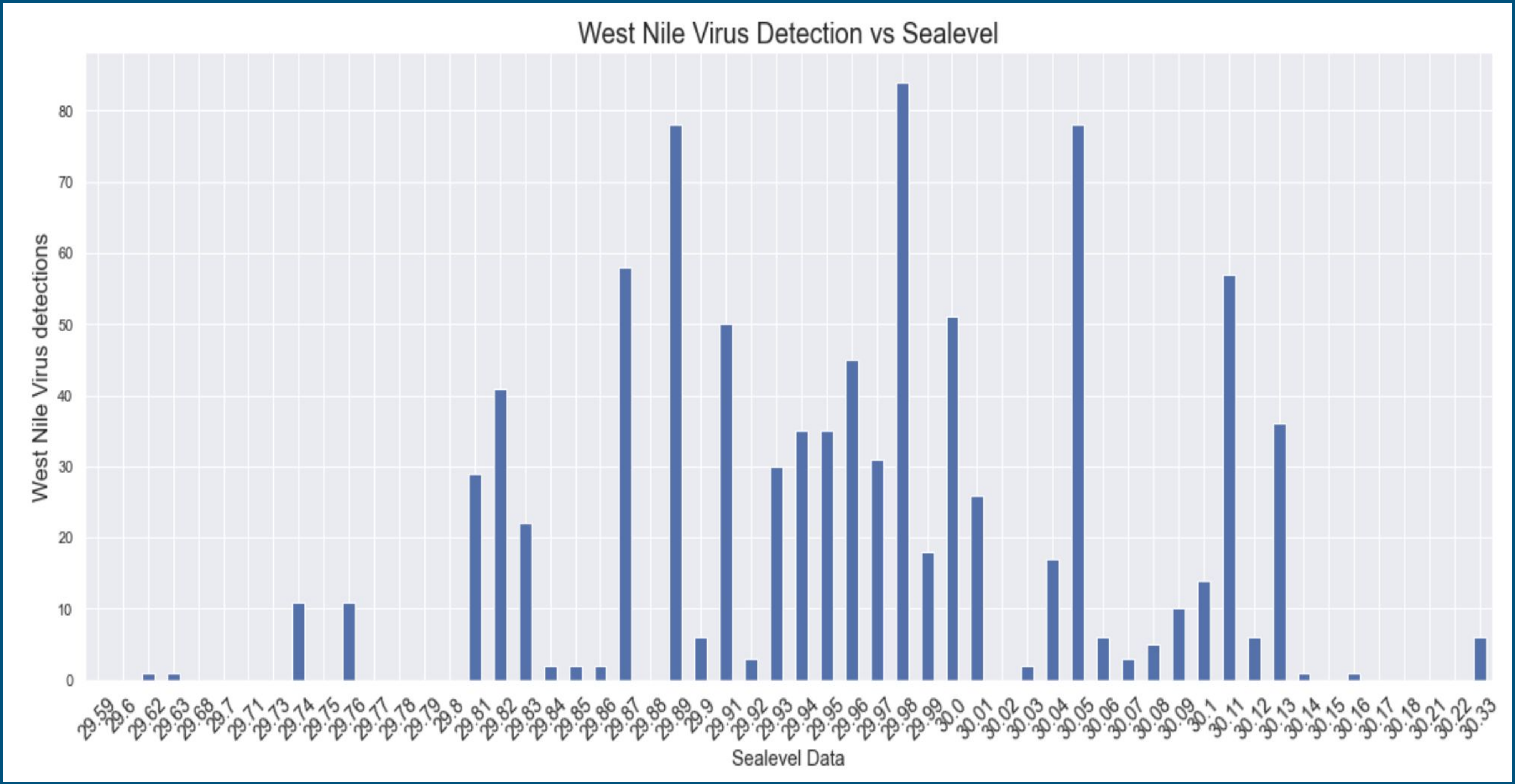
Dewpoint Temperature



Wetbulb Temperature



Sealevel

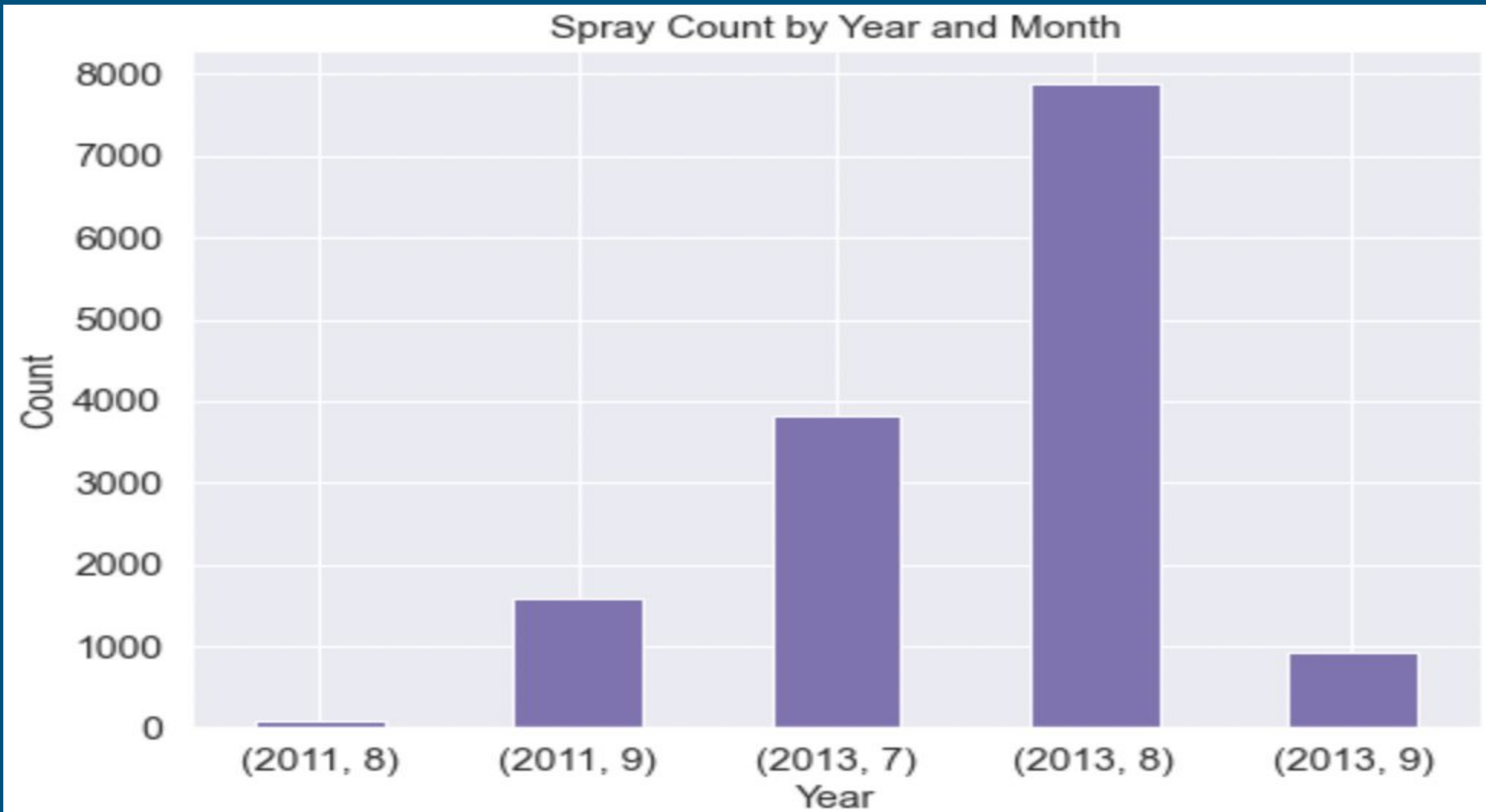


West Nile Virus Detection vs Sealevel
West Nile Virus detections
Sealevel Data
0
10
20
30
40
50
60
70
80
29.59
29.6
29.62
29.63
29.68
29.7
29.71
29.73
29.74
29.75
29.76
29.77
29.78
29.79
29.8
29.81
29.82
29.83
29.84
29.85
29.86
29.87
29.88
29.89
29.9
29.91
29.92
29.93
29.94
29.95
29.96
29.97
29.98
29.99
30.0
30.01
30.02
30.03
30.04
30.05
30.06
30.07
30.08
30.09
30.1
30.11
30.12
30.13
30.14
30.15
30.16
30.17
30.18
30.21
30.22
30.33

Optimal Conditions

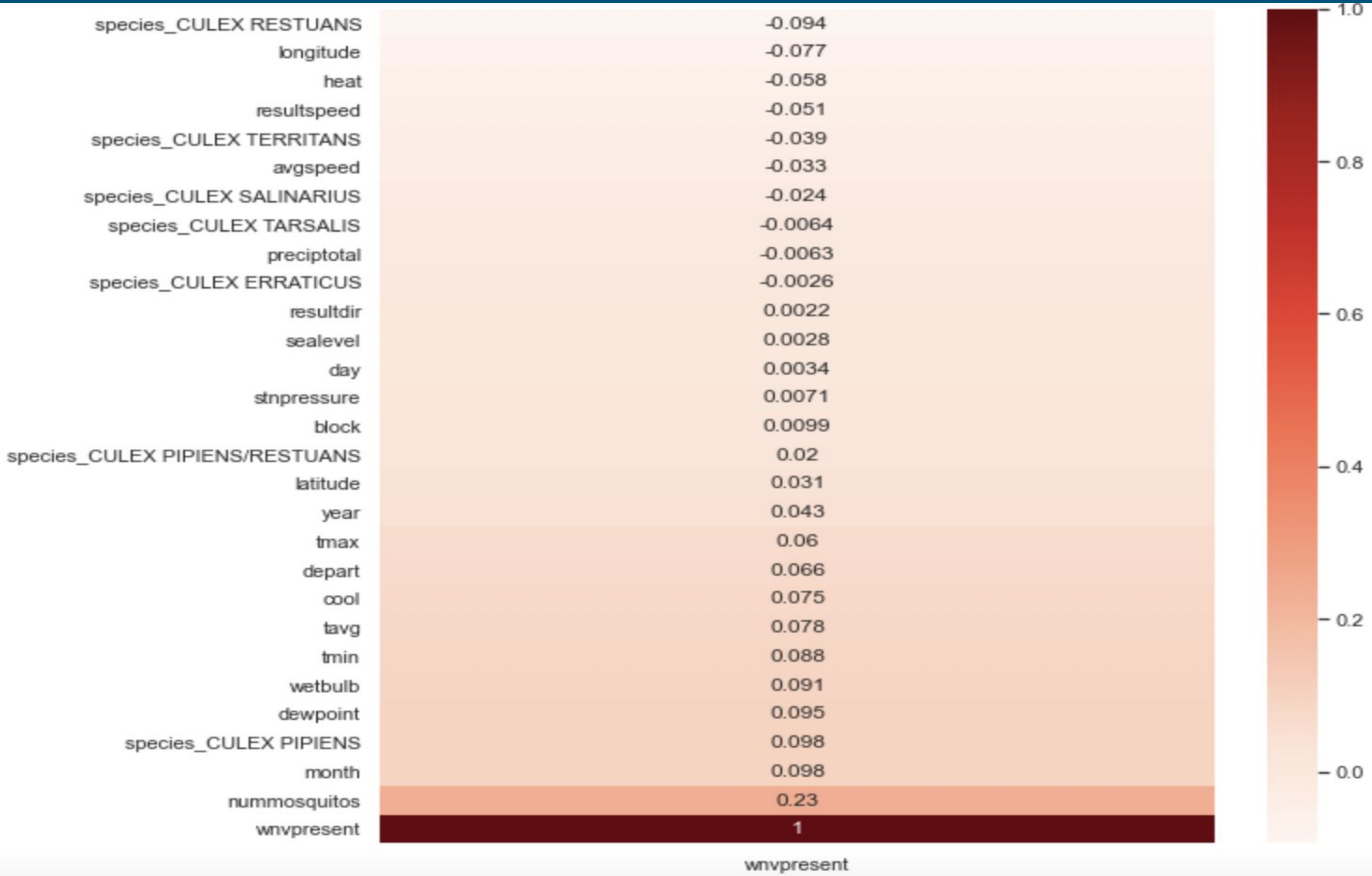
	Optimal Conditions		
tmax (°F)	78.00	cool (°F)	7.00
tmin (°F)	69.00	preciptotal (inches)	0.00
tavg (°F)	72.00	stnpressure (inches)	29.29
depart	4.00	sealevel (inches)	29.98
dewpoint (°F)	67.00	resultspeed (mph)	3.50
wetbulb (°F)	69.00	resultdir	24.00
heat (°F)	0.00	avgspeed (mph)	6.00

Spray



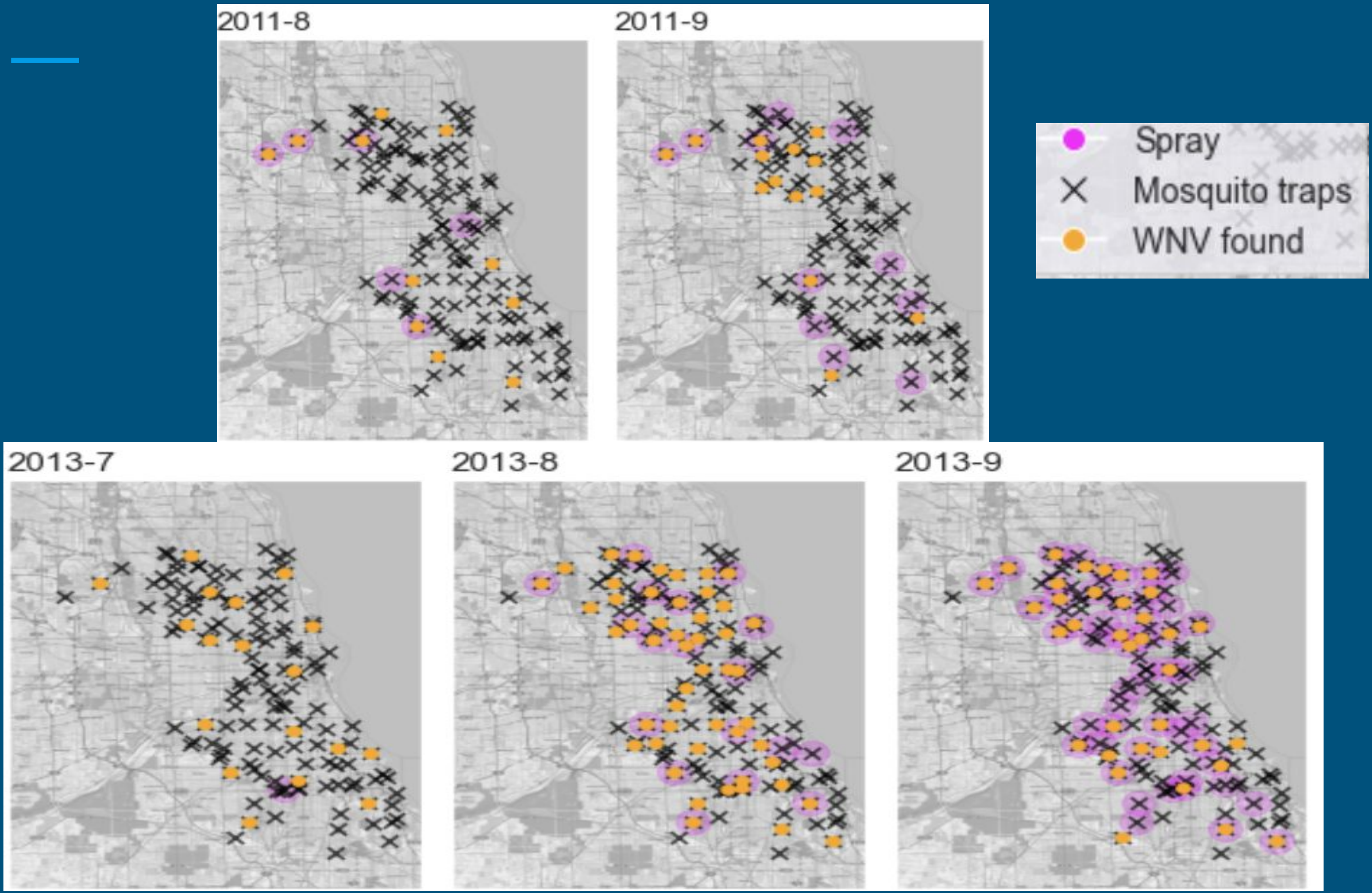
Spray Count by Year and Month
8000
7000
6000
5000
4000
3000
2000
1000
0
Count
(2011, 8)
(2011, 9)
(2013, 7)
(2013, 8)
(2013, 9)
Year

Correlation Heatmap



Cost-Benefit Analysis

Spray



2011-8
2011-9
2013-7
2013-8
2013-9
Spray
Mosquito traps
WNV found

Cost

Spray Rate : 1.5 ounces per acre

Cost per Gallon (128 ounces) : USD $300 per Gallon

Chicago Area size = 149800 acres

Spray amount in Gallons: (149,800 / 1.5) / 128 = $780.21

Total cost to spray once every fortnight in Chicago for 6 months (from May to Oct; including summer):

($780.21 x 300) x 12 = $2,808,756

Benefits

Summary of Hospital costs that could possibly be incurred if infected with the West Nile Virus:

Cost category	Mean	95% CI	Median	Range	Chicago Mean cost	Chicago cost range estimated
Total acute medical care	252,115,100	158,022,000–458,998,400	230,879,300	115,644,400–2,822,846,000	2,068,705	948,908-23,162,580
Total acute lost productivity†	22,081,260	9,550,370–63,069,700	16,144,050	7,070,480–2,643,251,000	181,185	58,016-21,688,931
Total long-term medical care	27,570,280	11,566,780–56,221,870	25,468,510	6,087,800–118,883,900	226,225	49,953-975,489
Total long-term lost productivity	26,866,800	13,526,800–48,279,320	25,416,720	7,790,800–85,567,700	220,452	63,927-702,117
Total lifetime lost productivity caused by deaths‡	449,464,800	(NA)	449,464,800	(NA)	3,688,038	NA

Analysis

The total estimated hospitalisation costs are high and this do not include outpatient visits and discharged patients follow up costs.

Nonetheless, the important takeaway from our external research is that the benefits of fogging the entire City of Chicago greatly outweights the cost. Specifically, there are huge savings (eg. hospitalisation bills) that we can reap if we manage to eradicate the West Nile Virus via fogging of entire Chicago city once every fortnightly for 6 months.

Conclusion

As we are concerned about correctly identifying the presence of the West Nile Virus, we have predominantly used the test accuracy score to choose the best performing model. Notwithstanding this, we also looked at the sensitivity score so as to ensure that we keep our false positives to a minimum. Taking these two evaluation metrics into account, we observed that our Random Forest model is the most ideal (i.e. test accuracy of 0.9544 and sensitivity of 0.4172).

Random Forest	0.9957	0.9544

Conclusion

Re cost-benefit analysis, our earlier paragraph shows that the cost of spraying the whole of Chicago fortnightly for 6 months (including the summer months) will cost about $2.8mil. This is far lower that the total estimated hospitalisation costs (see figures under "Chicago Mean Cost" from table above) from West Nile Virus and demonstrates a clear impetus for the city of Chicago to spray insecticide city-wide to reduce the spread of the virus.