Predicting West Nile Virus Outbreaks

Kihoon Sohn, Awab Idris, James Linek, and Leticia Beeck

Outline

1. Define the Problem
2. What data did we analyze?
3. Our Methods and Models
4. Cost-Benefit Analysis
5. Key Takeaways, Recommendations and Next Steps

Defining the Problem

- When and where will different species of mosquitoes test positive for West Nile Virus?

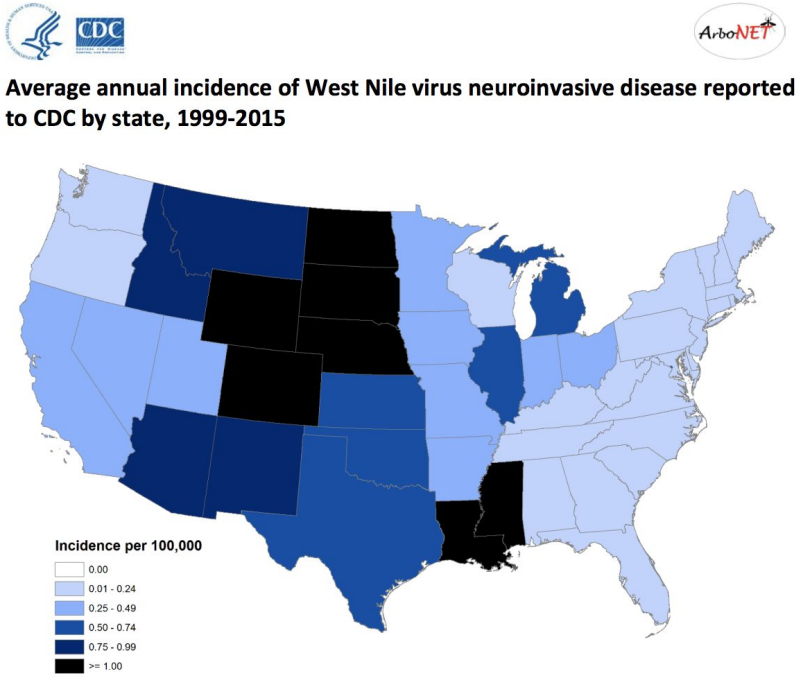
What data did we analyze

- The city of Chicago's GIS data for spray efforts in 2011 and 2013
- Chicago's weather conditions from 2007 to 2014 during the months of testing
- The training data from the main dataset.
    - The training data provided us with from 2007, 2009, 2011, and 2013 which would help us predict the test results for 2008, 2010, 2012, and 2014

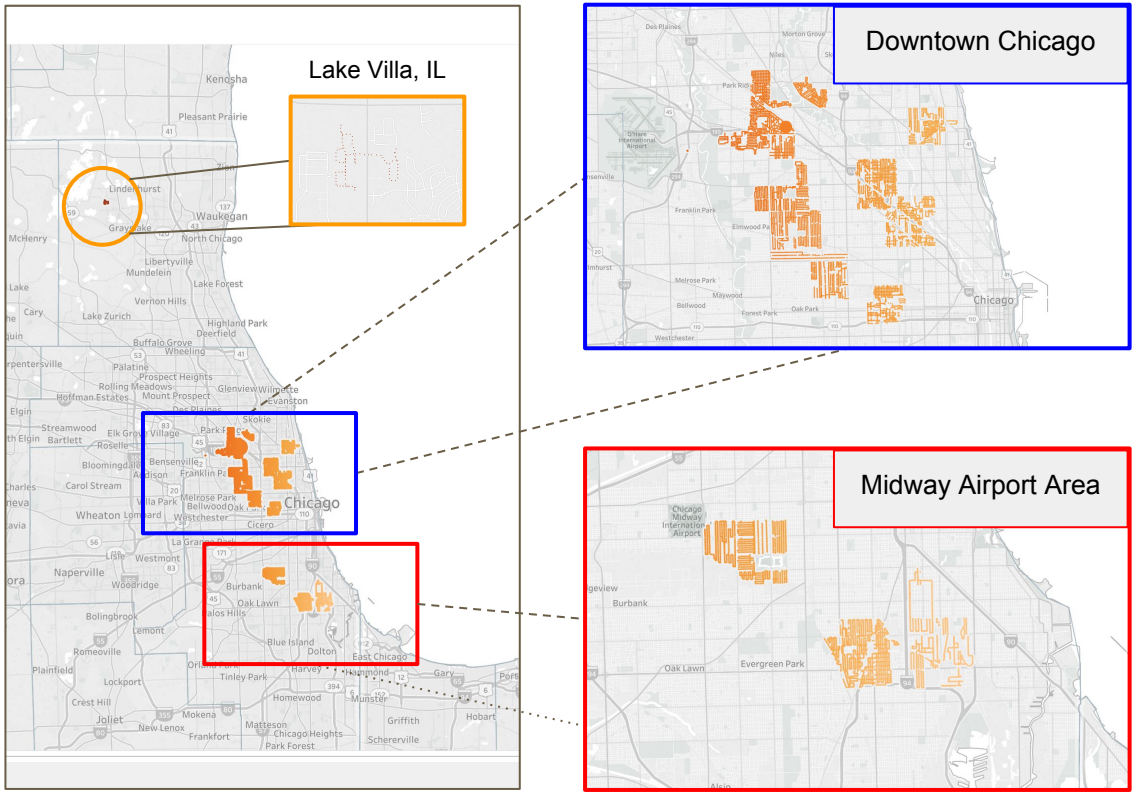
West Nile Virus

- Arrived to the U.S. in 1999
- First occurrence in NYC affected at least 62 people
- **At least 37,000 people were infected and 1,500 in deaths**

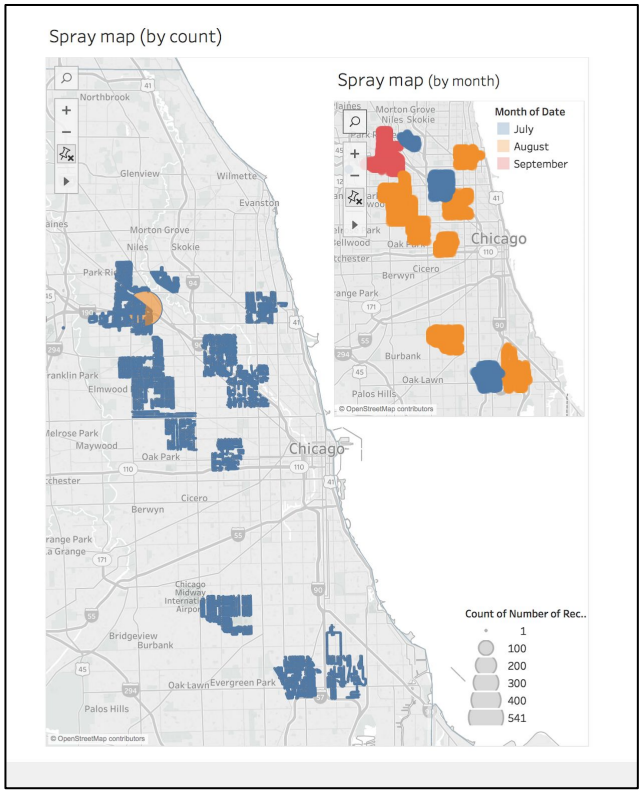
CDC Report(1999~2012)



Spray map in detail (2011, 2013 in spray dataset)



Spray map in detail (2011, 2013 in spray dataset)



Modeling and EDA

- Main features we focused on:
  - Temperature (min, max, and average)
  - Date (Day, month, year)
  - Precipitation
- We dropped certain features we did not think were as relevant to the model, primarily ones relating to addresses since we already had longitude and latitude coordinates
- We also dropped Snowfall and Water 1 (empty column)

Our Models

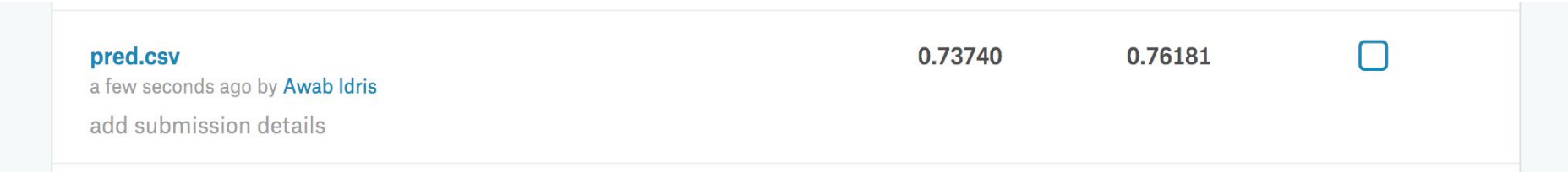
- *By GridSearchCV, we tried multiple parameters for tuning our modeling.*
- *These models are listed below:*

Machine Learning Model with GridSearchCV. Results from the test set:

- ***Logistic Regression .9474***
- *AdaBoost Classifier .9440*
- *Random Forest .9371*

Results

- **Here is our final Kaggle Submission Score :**



Cost Thus Far

- Up until 2014, the CDC has estimated that the virus has cost the U.S. about **$778 million**

Monetary Costs

- **Education** is crucial to help people understand the gravity of WNV and thus take precautionary measures
  - In order to improve this, schools will have to have more resources available as well as host talks that address the issue
- The National Center for Biotechnology Information estimated that the cost of an uncomplicated case of someone infected by WNV was approximately **$1,000 per case**,
  - ~**$165 per day** for a 5 day period due to lost productivity
  - + ~**$175 in medical costs** which includes 1 ambulatory care visit, medications, and diagnostic tests.
  - In 2017, the CDC recorded only 68 neuroinvasive cases, 19 non-neuroinvasive cases, and 5 deaths for the state of Illinois
  - Therefore, about **$19,000** (1,000*19) would be spent a year for non-neuroinvasive cases
- Neuroinvasive WNV illness **full recovery costs**:
  - **~$27,500** per case *This cost was estimated by an economics study done during the 2002 WNV epidemic in Louisiana
  - Therefore, about **$1,870,000** (27,500*68) would be spent a year for neuroinvasive cases
- **Pesticide Spray**
  - An average of **$5,179.40** per square mile

Negative Externalities

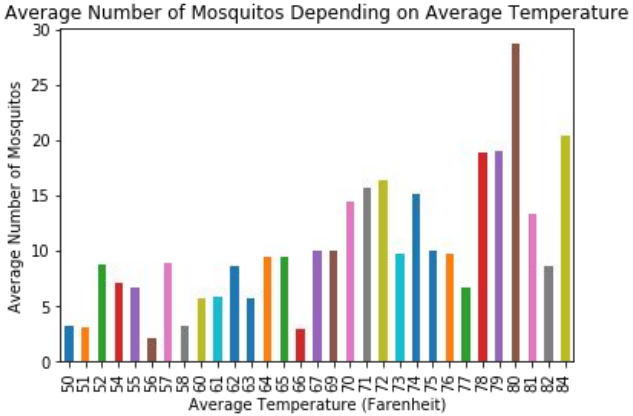
- In addition to monetary costs, there are a number of negative externalities associated with this virus:
  - People miss time at work
  - Overall health in Chicago deteriorates
  - Tourism slows down
  - Creates an unsafe environment
  - Chicago loses some of its reputation
    - No longer as attractive for conferences and big events such as concerts, sporting events, etc.

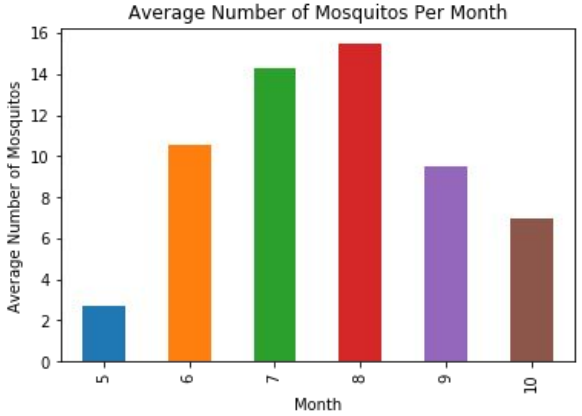
Benefits to Controlling WNV

- By implementing the model we shared, we are able to better target particular areas we believe are highly populated with virus-carrying mosquitoes
- By focusing primarily on targeted areas at certain times of the year, we will also be able to effectively reduce the number of WNV cases
- Our service and medical costs would drop significantly and thus increase overall health levels as well as Chicago's reputation and appeal

Recommendations

- The only species carrying the virus were the Culex Pipiens and Restauns, which raises another question of whether there could be a way of only targeting these kind of mosquitoes through a specific method or trap
    - This could be a key finding to continue to minimize costs and efficiently target the right species



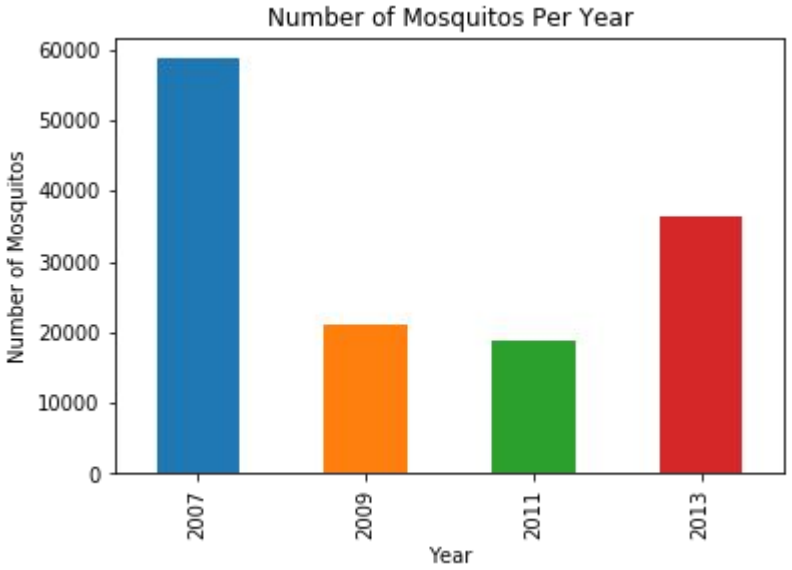


Recommendations

- Focus more on preventative measures
- Educate the citizen
  - Use insect repellent
  - Police the neighborhood to clean up sitting water
- Change Spraying Method
  - Aerial spraying: efficient way to spraying large areas
  - Mobility spraying: to cover wider area by using cars/moped/scooter to spray it
- Take closer care on the public area
  - Community center, Senior center, etc.

Next Steps

- Gather more Data
- Calculate a more accurate area of spray
- Dig deeper into why in certain months certain areas were sprayed while others weren't
- Look into why the number of mosquitos dipped after 2007, but then rose again (depicted in the graph to the right)



Let there be no more mosquito footprints left in Chicago

Thank you!

Questions?