Predicting West-Nile in Chicago

Carlton Mezetin, Neil Liberman, Nathan Hall

West Nile Virus

- First US detection 1999 in a set of birds
- Most commonly spread mosquitoes
- Major Reporting Issues
  - 720 vs 54000
    - Most cases go undiagnosed
    - Mild cases are often detected in blood donations
- No Vaccine
  - Mosquito prevention

Why Chicago?

- Chicago Population

  -2.7 million residents

- West Nile in Chicago
  - 2001 first case
  - 2002 First Human case
  - 2004 West Nile Surveillance system

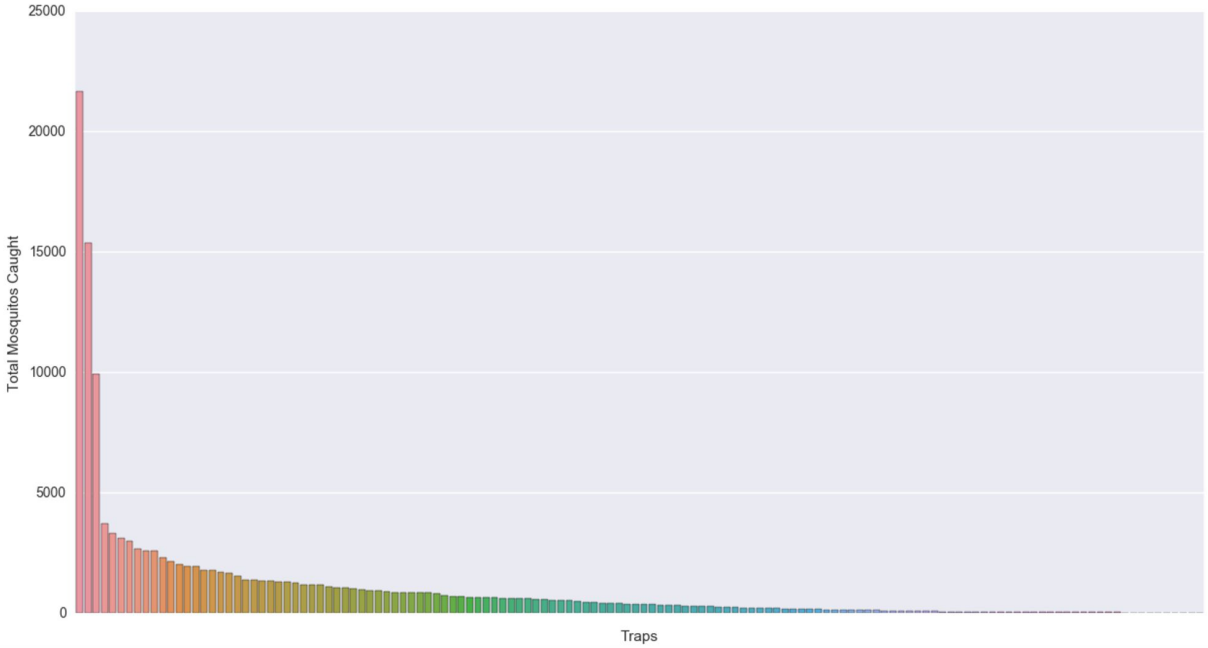
- **Challenge Overview**
- *Our Strategy*
- *Our Assumptions*
- *How important is climate to mosquitos?*
- *Can we make this better with more data?*
- *Are there any problems between the test and train data sets?*
- *So what did we find?*
- *Can't we just weight the future if we know it?*
- *How to solve the REAL problem.*
- *How well did that work?*
- *Where to go next?*

Kaggle Competition

- The kaggle competition called us to predict where and when West Nile would be present in Chicago
- They provided us with Trap data
- Each entry represents a mosquito species from a trap.
  - Species
  - Trap #
  - Location information
  - If West Nile was present
  - Etc.

- **Challenge Overview**
- *Our Strategy*
- *Our Assumptions*
- *How important is climate to mosquitos?*
- *Can we make this better with more data?*
- *Are there any problems between the test and train data sets?*
- *So what did we find?*
- *Can't we just weight the future if we know it?*
- *How to solve the REAL problem.*
- *How well did that work?*
- *Where to go next?*

What This Tells Us About Controlling West Nile...



Our Strategy

West Nile Virus is highly correlated with mosquito density, therefore, we sought to find variables associated with a thriving mosquito population.

Much time was therefore spent familiarizing ourselves with the climate factors affecting mosquito populations as well as the behavior of mosquitoes.

Having zeroed in on potential indicators, we accumulated as much relevant data as we could find.

- *Challenge Overview*
- *Our Strategy*
- **Our Assumptions**
- *How important is climate to mosquitos?*
- *Can we make this better with more data?*
- *Are there any problems between the test and train data sets?*
- *So what did we find?*
- *Can't we just weight the future if we know it?*
- *How to solve the REAL problem.*
- *How well did that work?*
- *Where to go next?*

Conducive Conditions for Mosquitoes

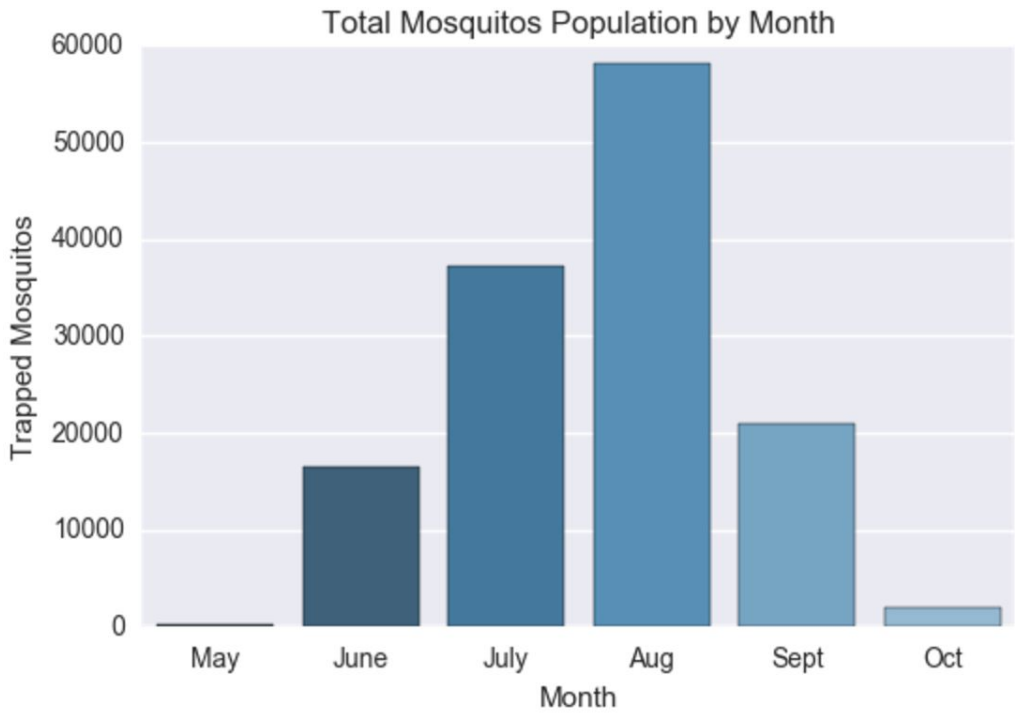
- Standing Water
- High Temperatures
- High Humidity

Other Factors That May Attract Mosquitoes

- Carbon Dioxide
- Lactic Acid
- Movement
- Grassy Areas

- *Challenge Overview*
- *Our Strategy*
- *Our Assumptions*
- **How important is climate to mosquitos?**
- *Can we make this better with more data?*
- *Are there any problems between the test and train data sets?*
- *So what did we find?*
- *Can't we just weight the future if we know it?*
- *How to solve the REAL problem.*
- *How well did that work?*
- *Where to go next?*

Data Supporting Climate Conditions



- *Challenge Overview*
- *Our Strategy*
- *Our Assumptions*
- *How important is climate to mosquitos?*
- **Can we make this better with more data?**
- *Are there any problems between the test and train data sets?*
- *So what did we find?*
- *Can't we just weight the future if we know it?*
- *How to solve the REAL problem.*
- *How well did that work?*
- *Where to go next?*

Finding More Data

We now had an idea of what factors could signal a high mosquito population. What data was available?

Weather:

- Temperature
- Dew Point
- Humidity
- Pressure
- Wind
- Precipitation

- *Challenge Overview*
- *Our Strategy*
- *Our Assumptions*
- *How important is climate to mosquitos?*
- **Can we make this better with more data?**
- *Are there any problems between the test and train data sets?*
- *So what did we find?*
- *Can't we just weight the future if we know it?*
- *How to solve the REAL problem.*
- *How well did that work?*
- *Where to go next?*

Finding Data (continued)

Public Parks

Mountain Bike Trails

Bodies of water (> 0.001 sq. miles)

Local Farms

CO2 Building Emissions

- *Challenge Overview*
- *Our Strategy*
- *Our Assumptions*
- *How important is climate to mosquitos?*
- *Can we make this better with more data?*
- **Are there any problems between the test and train data sets?**
- *So what did we find?*
- *Can't we just weight the future if we know it?*
- *How to solve the REAL problem.*
- *How well did that work?*
- *Where to go next?*

One big problem...

Train Data Set

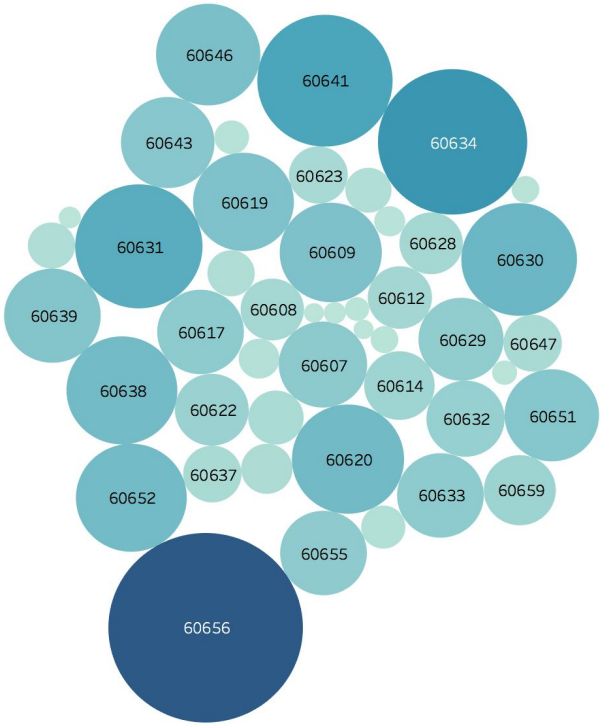
	NumMosquitos	WnvPresent
Species		
CULEX ERRATICUS	7	0
CULEX PIPIENS	44671	240
CULEX PIPIENS/RESTUANS	66268	262
CULEX RESTUANS	23431	49
CULEX SALINARIUS	145	0
CULEX TARSALIS	7	0
CULEX TERRITANS	510	0

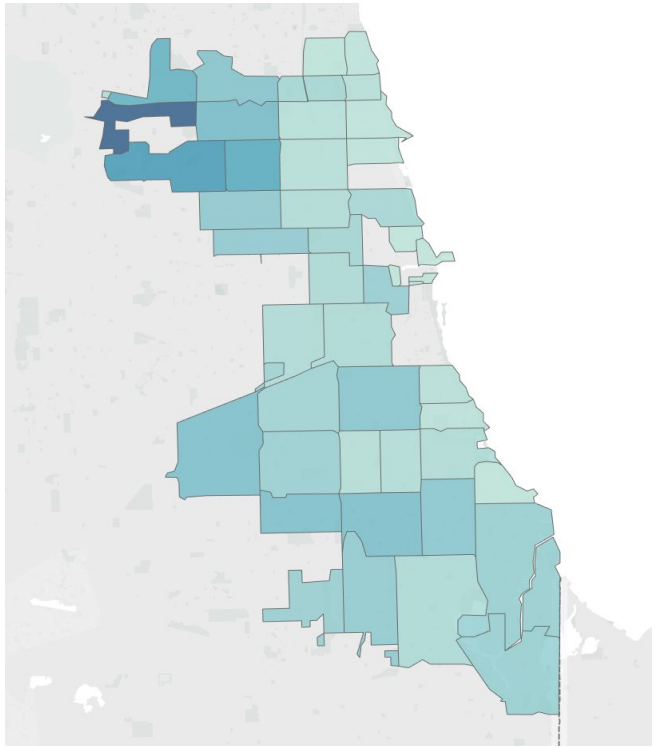
Test Data Set

	Record Count
Species	
CULEX ERRATICUS	14345
CULEX PIPIENS	14521
CULEX PIPIENS/RESTUANS	15359
CULEX RESTUANS	14670
CULEX SALINARIUS	14355
CULEX TARSALIS	14347
CULEX TERRITANS	14351
UNSPECIFIED CULEX	14345

- *Challenge Overview*
- *Our Strategy*
- *Our Assumptions*
- *How important is climate to mosquitos?*
- *Can we make this better with more data?*
- *Are there any problems between the test and train data sets?*
- **So what did we find?**
- *Can't we just weight the future if we know it?*
- *How to solve the REAL problem.*
- *How well did that work?*
- *Where to go next?*

Visualizing our Target Spray Locations

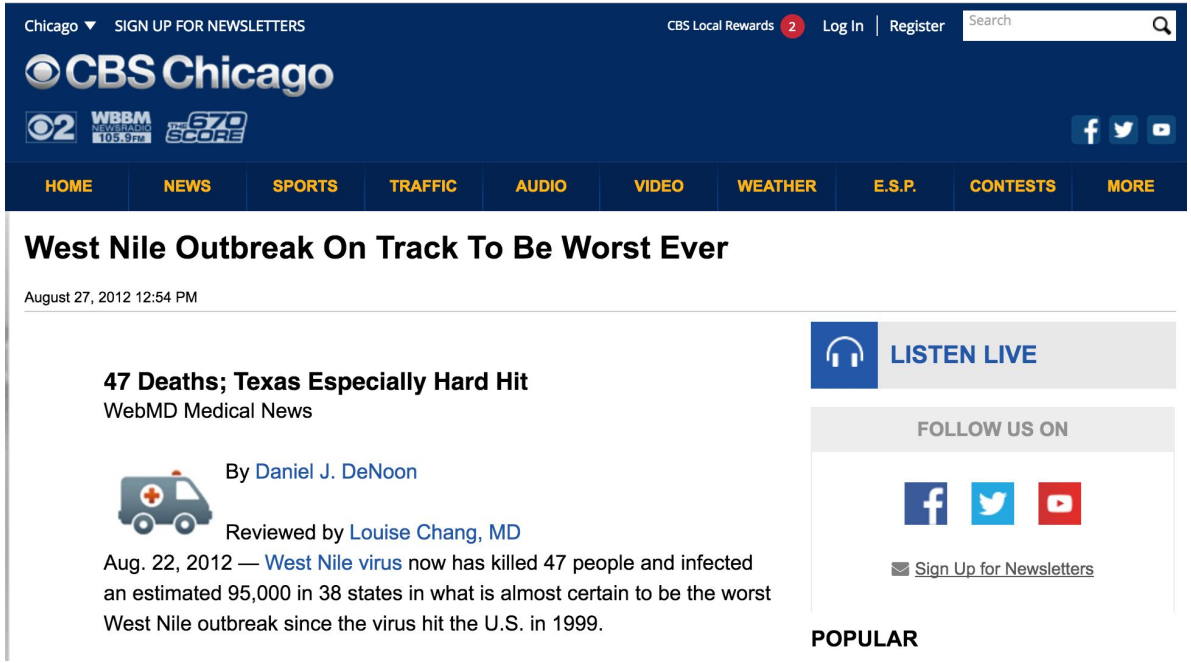




- *Challenge Overview*
- *Our Strategy*
- *Our Assumptions*
- *How important is climate to mosquitos?*
- *Can we make this better with more data?*
- *Are there any problems between the test and train data sets?*
- *So what did we find?*
- **Can't we just weight the future if we know it?**
- *How to solve the REAL problem.*
- *How well did that work?*
- *Where to go next?*

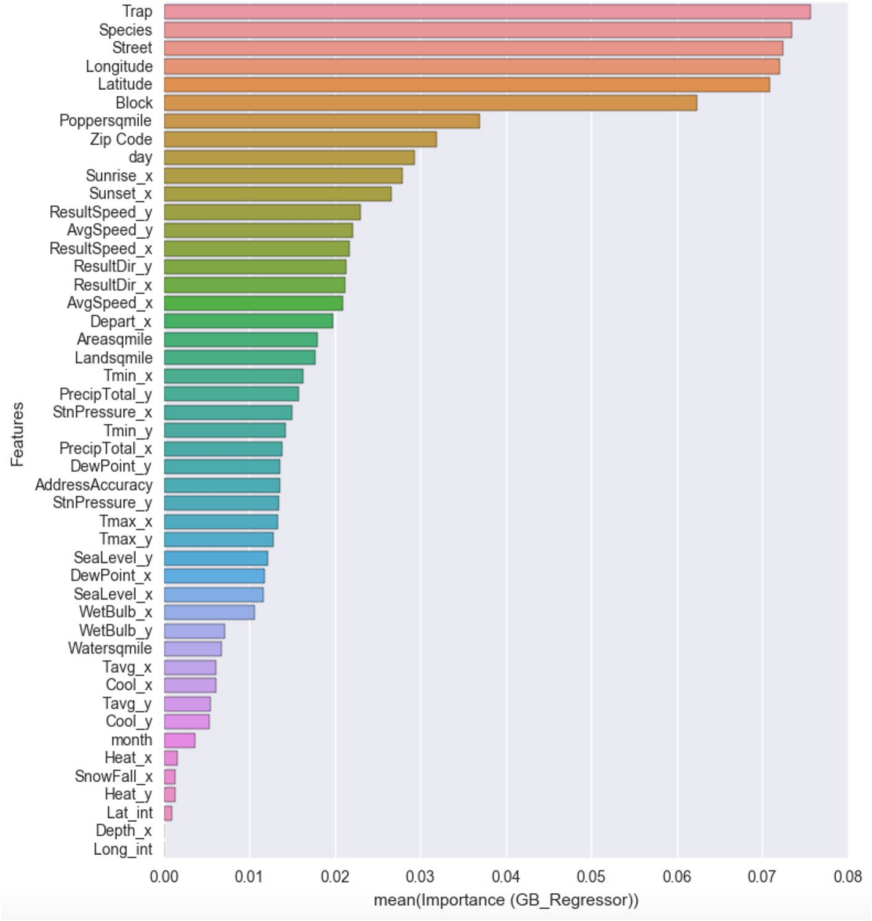
Well wouldn't that be nice.

- Our train set has the year 2012 which simple googling revealed...



Predicting Mosquito Population

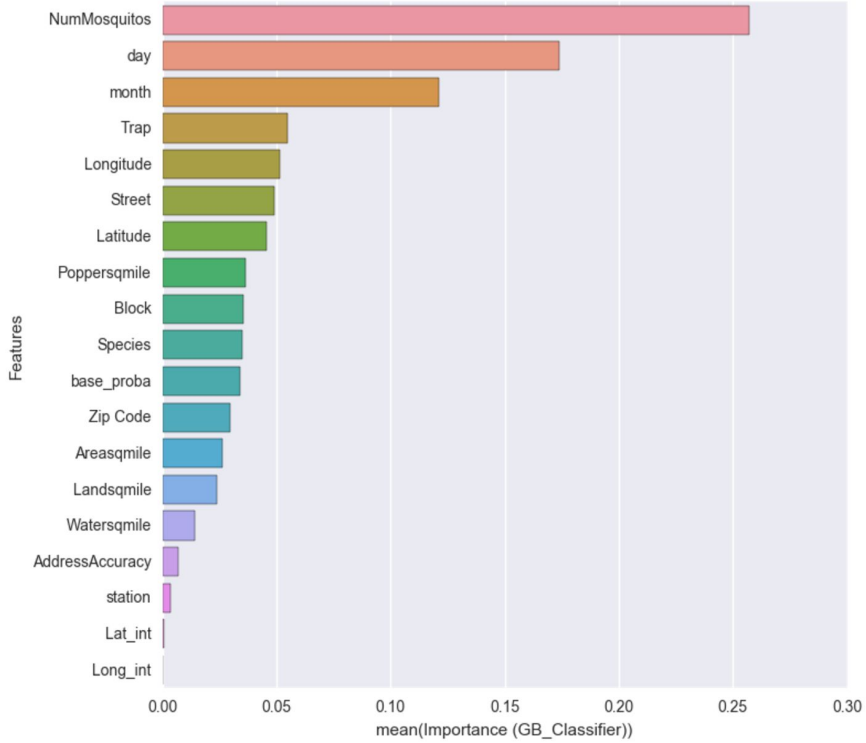
- *Challenge Overview*
- *Our Strategy*
- *Our Assumptions*
- *How important is climate to mosquitos?*
- *Can we make this better with more data?*
- *Are there any problems between the test and train data sets?*
- *So what did we find?*
- *Can't we just weight the future if we know it?*
- **How to solve the REAL problem.**
- *How well did that work?*
- *Where to go next?*



- *Challenge Overview*
- *Our Strategy*
- *Our Assumptions*
- *How important is climate to mosquitos?*
- *Can we make this better with more data?*
- *Are there any problems between the test and train data sets?*
- *So what did we find?*
- *Can't we just weight the future if we know it?*
- *How to solve the REAL problem.*
- **How well did that work?**
- *Where to go next?*

Accuracy on Kaggle

-	halluva_good_time	0.71350	-	Thu, 12 Jan 2017 15:09:15	Post-Deadline



- Challenge Overview
- Our Strategy
- Our Assumptions
- How important is climate to mosquitos?
- Can we make this better with more data?
- Are there any problems between the test and train data sets?
- So what did we find?
- Can't we just weight the future if we know it?
- How to solve the REAL problem.
- How well did that work?
- Where to go next?

Recommendations.

- Improve the predictions of west-nile outbreaks by incorporating west-nile positive case data.
- Continue to improve the predictions of mosquito population changes as that has the largest effect on west-nile outbreaks.
- Increase spraying during times of favorable breeding weather for mosquitos.