West Nile Virus Prediction

Predict West Nile virus in mosquitos across the city of Chicago

$40,000 · 1,306 teams · 2 years ago

Competition Data

Edit

- mapdata_copyright_op...
- noaa_weather_qclcd_d...
- mapdata_copyright_op...
- sampleSubmission.csv...
- spray.csv.zip
- test.csv.zip
- train.csv.zip
- weather.csv.zip
- west_nile.zip

noaa_weather_qclcd_documentation.pdf 85.17 KB

Download

Data Description

Ready to explore the data? Kaggle Scripts is the most frictionless way to get familiar with the competition dataset! No data download needed to start publishing and forking code in R and Python. It's already pre-loaded with our favorite packages and ready for you to start competing!

Data Description

In this competition, you will be analyzing weather data and GIS data and predicting whether or not West Nile virus is present, for a given time, location, and species.

Every year from late-May to early-October, public health workers in Chicago setup mosquito traps scattered across the city. Every week from Monday through Wednesday, these traps collect mosquitos, and the mosquitos are tested for the presence of West Nile virus before the end of the week. The test results include the number of mosquitos, the mosquitos species, and whether or not West Nile virus is present in the cohort.

Main dataset

These test results are organized in such a way that when the number of mosquitos exceed 50, they are split into another record (another row in the dataset), such that the number of mosquitos are capped at 50.

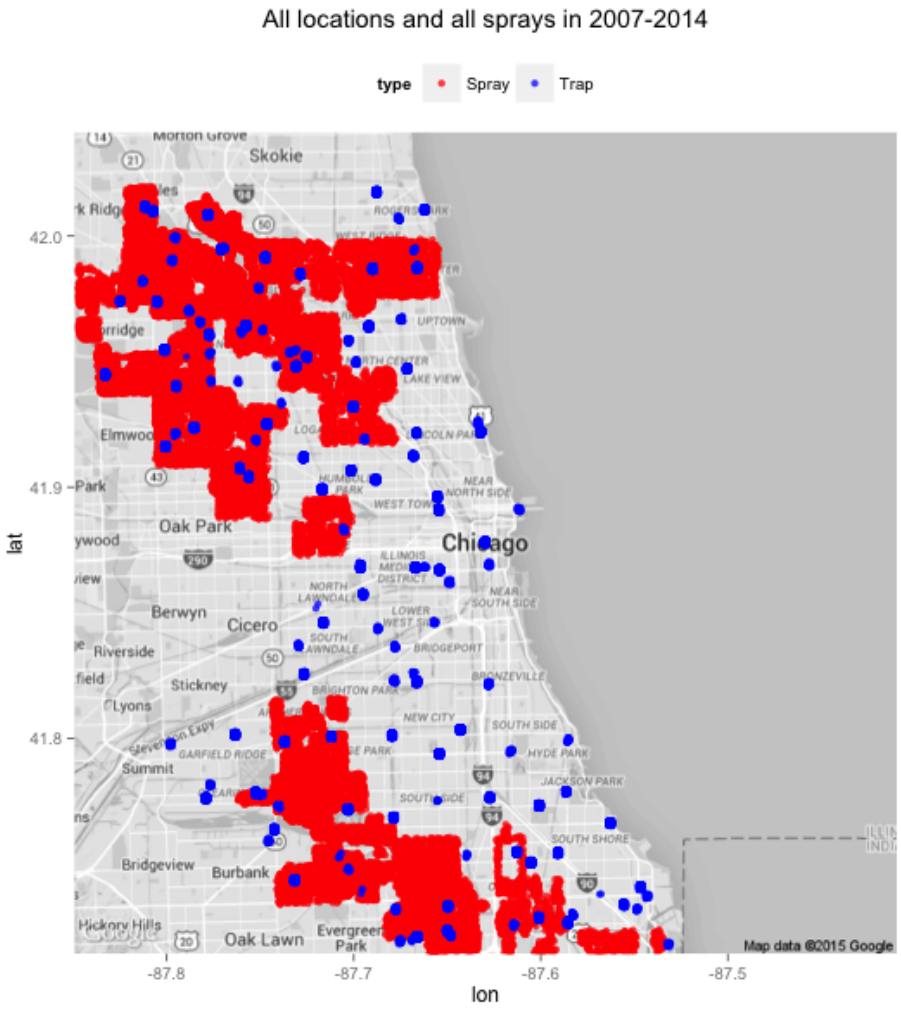
The location of the traps are described by the block number and street name. For your convenience, we have mapped these attributes into Longitude and Latitude in the dataset. Please note that these are derived locations. For example, Block=79, and Street= "W FOSTER AVE" gives us an approximate address of "7900 W FOSTER AVE, Chicago, IL", which translates to (41.974089,-87.824812) on the map.

1220.

Please note that not all the locations are tested at all times. Also, records exist only when a particular species of mosquitos is found at a certain trap at a certain time. In the test set, we ask you for all combinations/permutations of possible predictions and are only scoring the observed ones.

Spray Data

The City of Chicago also does spraying to kill mosquitos. You are given the GIS data for their spray efforts in 2011 and 2013. Spraying can reduce the number of mosquitos in the area, and therefore might eliminate the appearance of West Nile virus.



Weather Data

It is believed that hot and dry conditions are more favorable for West Nile virus than cold and wet. We provide you with the dataset from NOAA of the weather conditions of 2007 to 2014, during the months of the tests.

Station 1: CHICAGO O'HARE INTERNATIONAL AIRPORT Lat: 41.995 Lon: -87.933 Elev: 662 ft. above sea level
Station 2: CHICAGO MIDWAY INTL ARPT Lat: 41.786 Lon: -87.752 Elev: 612 ft. above sea level

Map Data

The map files **mapdata_copyright_openstreetmap_contributors.rds** and **mapdata_copyright_openstreetmap_contributors.txt** are from Open Streetmap and are primarily provided for use in visualizations (but you are allowed to use them in your models if you wish).

Here's an example using **mapdata_copyright_openstreetmap_contributors.rds**, and here's one using **mapdata_copyright_openstreetmap_contributors.txt**.

File descriptions

- **train.csv, test.csv** - the training and test set of the main dataset. The training set consists of data from 2007, 2009, 2011, and 2013, while in the test set you are requested to predict the test results for 2008, 2010, 2012, and 2014.
  - Id: the id of the record
  - Date: date that the WNV test is performed
  - Address: approximate address of the location of trap. This is used to send to the GeoCoder.
  - Species: the species of mosquitos
  - Block: block number of address
  - Street: street name
  - Trap: Id of the trap
  - AddressNumberAndStreet: approximate address returned from GeoCoder
  - Latitude, Longitude: Latitude and Longitude returned from GeoCoder
  - AddressAccuracy: accuracy returned from GeoCoder
  - NumMosquitos: number of mosquitoes caught in this trap
  - WnvPresent: whether West Nile Virus was present in these mosquitos. 1 means WNV is present, and 0 means not present.
- **spray.csv** - GIS data of spraying efforts in 2011 and 2013
  - Date, Time: the date and time of the spray
  - Latitude, Longitude: the Latitude and Longitude of the spray
- **weather.csv** - weather data from 2007 to 2014. Column descriptions in noaa_weather_qclcd_documentation.pdf.
- **sampleSubmission.csv** - a sample submission file in the correct format