

# HealthTweets.org

## A Platform for Public Health

Health Tweets  
Tracking Health Trends via Social Media

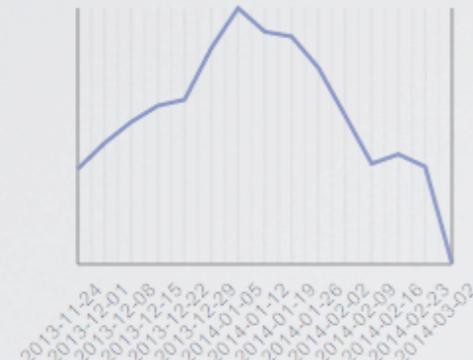
Home About Trends

User: (user) [Logout](#)

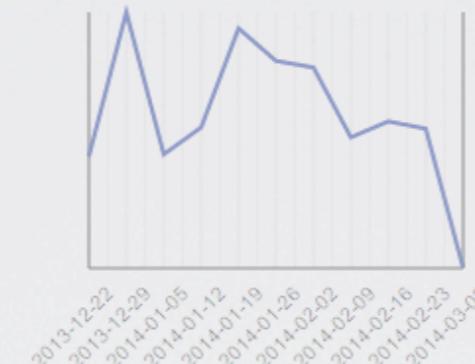
## Surveillance using Twitter

Health Tweets  
Follow health trends based on tracking Twitter messages. [\(Learn More\)](#)

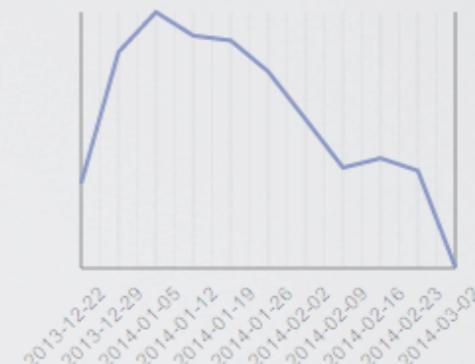
United States Flu Rate



United Kingdom Flu Rate



New York State Flu Rate



[Make your own](#)

Michael J. Paul  
Johns Hopkins University

Last updated: Feb. 24, 2014 | Processed 2,804,320,831 tweets since July 18, 2011.

©2014 by the Johns Hopkins Social Media and Health Research Group | [contact@healthtweets.org](mailto:contact@healthtweets.org)

Data on this site is collected from third-party social media. Reported health information may be incomplete or inaccurate.

Renyuan Cheng, Mark Dredze, David Broniatowski

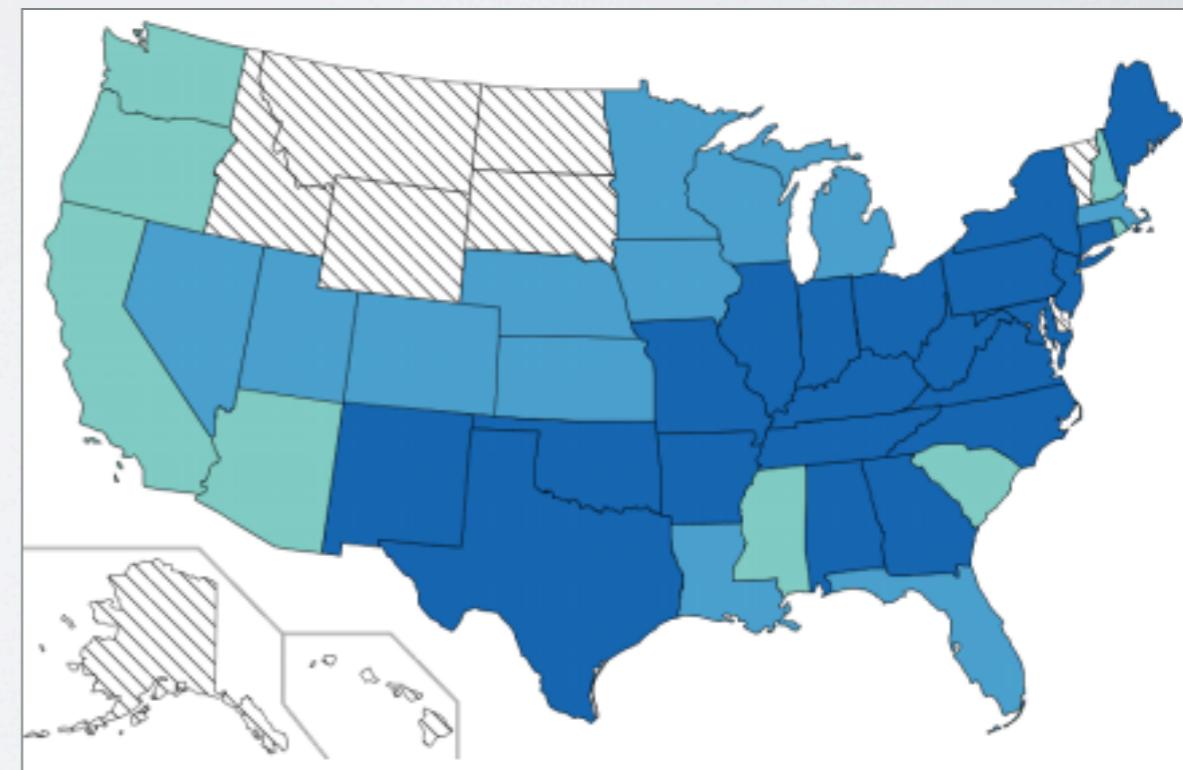
# Social Media for Public Health

- Mental health
- Drug use
- Tobacco use
- Emergency preparedness and response
- Vaccination sentiment
- ...



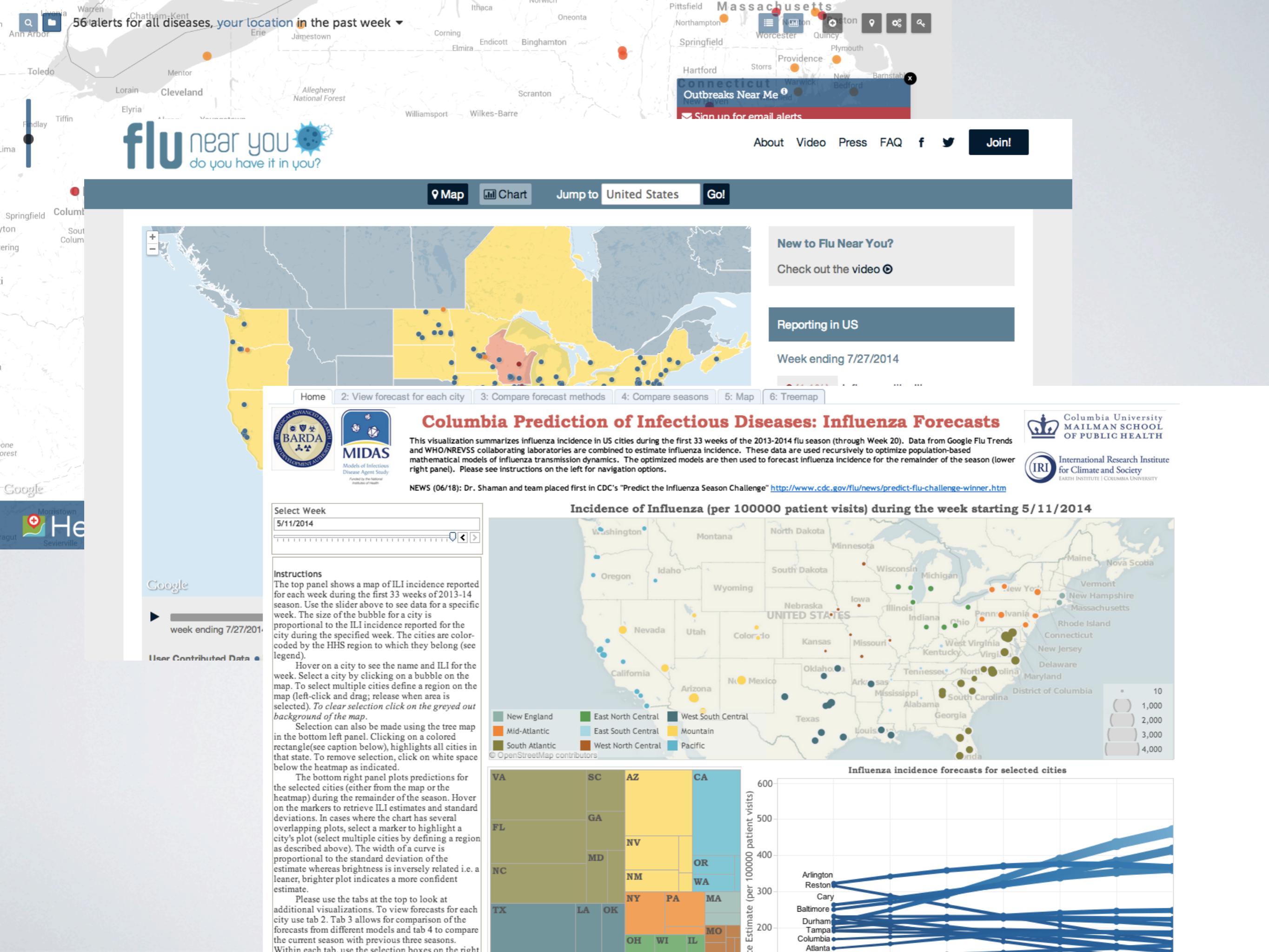
# Influenza Surveillance

- Use Web data to track current influenza rates
  - Twitter messages
  - Google Search Queries
  - Wikipedia searches
  - ...



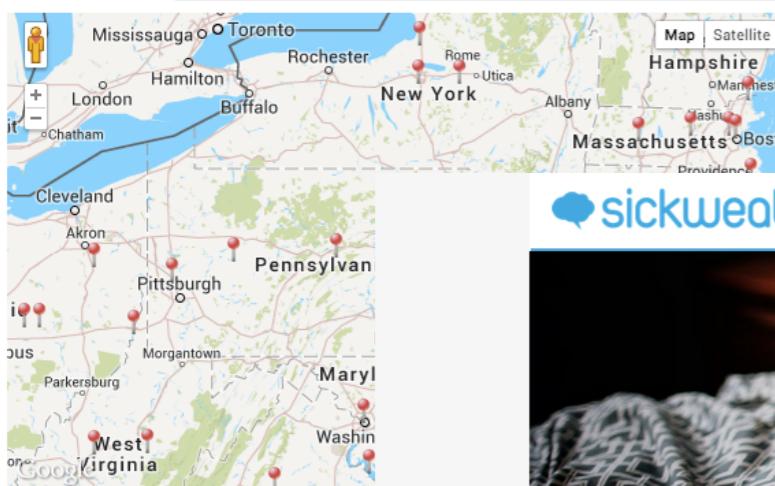
# Research to Practice

- Over two dozen papers demonstrating ability to track influenza with Twitter
- How can a public health official make use of these results?
- Gap between research code and timely, ready to use information



Select the geographic area of interest, and find out which diseases are trending.

[Geographic Map](#) [Select US state\(s\)](#)



### Help us make crowdbreaks better

With a single click, you can help us improve our system by answering the questions below. [\[How does this work?\]](#)

Is this tweet about Influenza?

"Before I sleep

# crowdbreaks

Disease surveillance, #crowdsourced. Powered by you.

S

▼ 1. Sexually transmitted diseases

change prev. day: -9.78% | change prev. week: -9.92%

14 days ago      7 days ago      past 24 hours

[Live Map](#) [Alerts](#) [Widgets](#) [Sign In](#)

Protect your loved ones from illness with Sickweather.

U.S. Department of Health & Human Services

### Sickweather - Sickness Forecasting and Mapping

Everyday thousands of people around the globe share their sickness stories on social media sites like Facebook and Twitter when they (or someone close to them) get sick. Just as Doppler radar scans the sky for indicators of bad weather, Sickweather scans social media for indicators of illness, allowing you to check for the chance of sickness as easily as you can check for the chance of rain. See [How Sickweather works](#)

[Download on the App Store](#)

### Public Health Emergency

[NowTrending.HHS.gov](#) beta Trends ▾ by condition ▾ by location ▾

### Welcome!

We are tracking disease trends, 140 characters at a time

In March 2012, the Assistant Secretary for Preparedness and Response at the Department of Health and Human Services launched a challenge competition titled Now Trending: #Health in My Community. This contest challenged entrants to create a web-based application that searched open source Twitter data for health topics and delivered analyses of that data for both a specified geographic area and the national level. This website is a result of that contest. The information available below and throughout the website is a tool intended for health departments and other health entities to use in multiple ways such as serving as an indicator of potential health issues emerging in the population, building a baseline of trend data, engaging the public on trending health topics, or cross-referencing other data sources.

The data and metrics on this site represent data for up to the last two weeks. Full historical data is being maintained but is not publicly available at this time.

**4,476,036**  
tweets gathered from Twitter's Streaming API. All of them match at least one of the 234 condition terms currently tracked across 27 conditions.

**122,165 (2%)**  
tweets with a sensor-based location  
(read more about how we calculate this)

**2,721,258 (60%)**  
tweets with a popular user profile location  
(read more about how we calculate this)

#### Conditions by Tweet Count

std natural disaster acute respiratory illness gastroenteritis common cold influenza tuberculosis pertussis malaria meningitis rabies anthrax polio dengue pneumonia tick borne disease varicella cholera mosquito borne disease smallpox measles tetanus yellow fever mumps typhoid chagas enterovirus legionnaires disease diphtheria

#### Top 20 Tweet Locations

national capital region, republic of the philippines, california, usa, texas, usa, orlando, fl, alaska, usa, new york, ny, florida, usa, calabarzon, republic of the philippines, los angeles, ca, georgia, usa, ohio, usa, pennsylvania, usa, melbourne, new york, usa, north carolina, usa, new jersey, usa, central luzon, republic of the philippines, arizona, usa, nigeria, virginia, usa

#### Top 20 User Locations

us london philippines usa #215love #lcm new york uk indonesia worldwide canada california united states singapore texas australia earth india manila, philippines los angeles nigeria

# Our Goal

- Create a platform for research collaboration with public health officials
  - Deliver timely health trends from social media
  - Receive detailed feedback to facilitate new research

# Requirements

- Goal: show users health trends in Twitter data
- Requires
  - Trend identification
  - Tweet geolocation
  - Historical data



# Trend Identification

- System uses a variety of approaches to trend identification
  - Keyword based
  - Machine learning classifiers
- Example: Influenza
  - A series of statistical machine learning classifiers
  - Lamb et al, 2013. Broniatowski et al, 2013.



# Tweet Geolocation

- Identify the location of tweets
- Two sources of geolocation information
  - Geotagged tweets (~3% of US data)
  - Automatically inferred locations based on profile information
    - Carmen (Dredze et al, 2013)

# Twitter Geolocation

- User profiles
  - Self-reported locations
  - 56% of users fill this in
- Tweet content
  - Language analysis
  - More involved – we don't do this here

# Carmen

- Returns structured object for each tweet
  - City, County, State, Country
- Fast and simple
  - 27,000 tweets / sec
- Code available on Github
  - Python and Java versions



# Carmen

- Uses GPS data when available
  - Get location information from Yahoo Map API
  - Mapping of user profile strings to places
    - e.g. “NYC”, “New York” -> {city: New York, state: NY, ...}
    - Manually curated
    - Automatically added aliases using location clusters created from social network structure
      - Bergsma et al, 2013
      - 4,811 unique places in our mapping

# Evaluation

- Treated GPS locations as ground truth
- Evaluated geolocation from user profiles against the ground truth
  - Accuracy (precision)
  - Coverage (recall)
- Test set: 56,000 tweets (plus 10,000 dev)

# Evaluation

- Accuracy:
  - Country: 91%
  - State+Country: 65%
  - Within 250 miles: 75%
  - Within 25 miles: 55%
- Coverage:

# Historical Data

- Trends require a temporal analysis
  - Helpful to have historical data
- Two streaming API data collections
  - Health keyword based collection
  - Public sample
- Normalization: compute per capita rates based on public sample normalization



# Demo

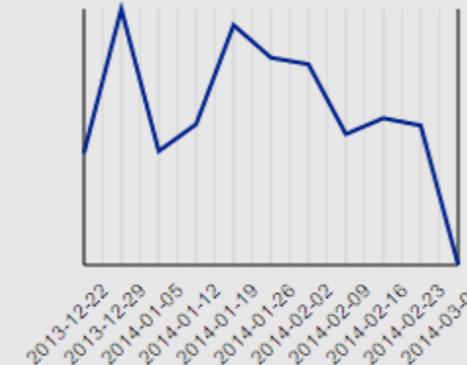
## Health Tweets

Follow health trends based on tracking Twitter messages. ([Learn More](#))

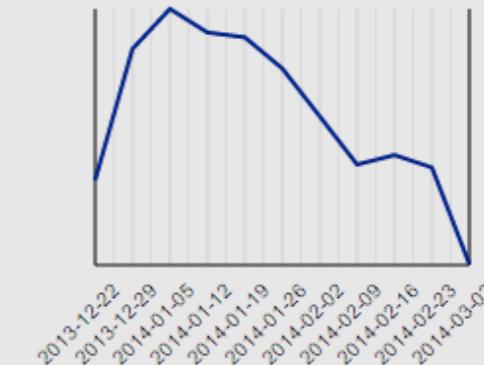
United States Flu Rate



United Kingdom Flu Rate



New York State Flu Rate

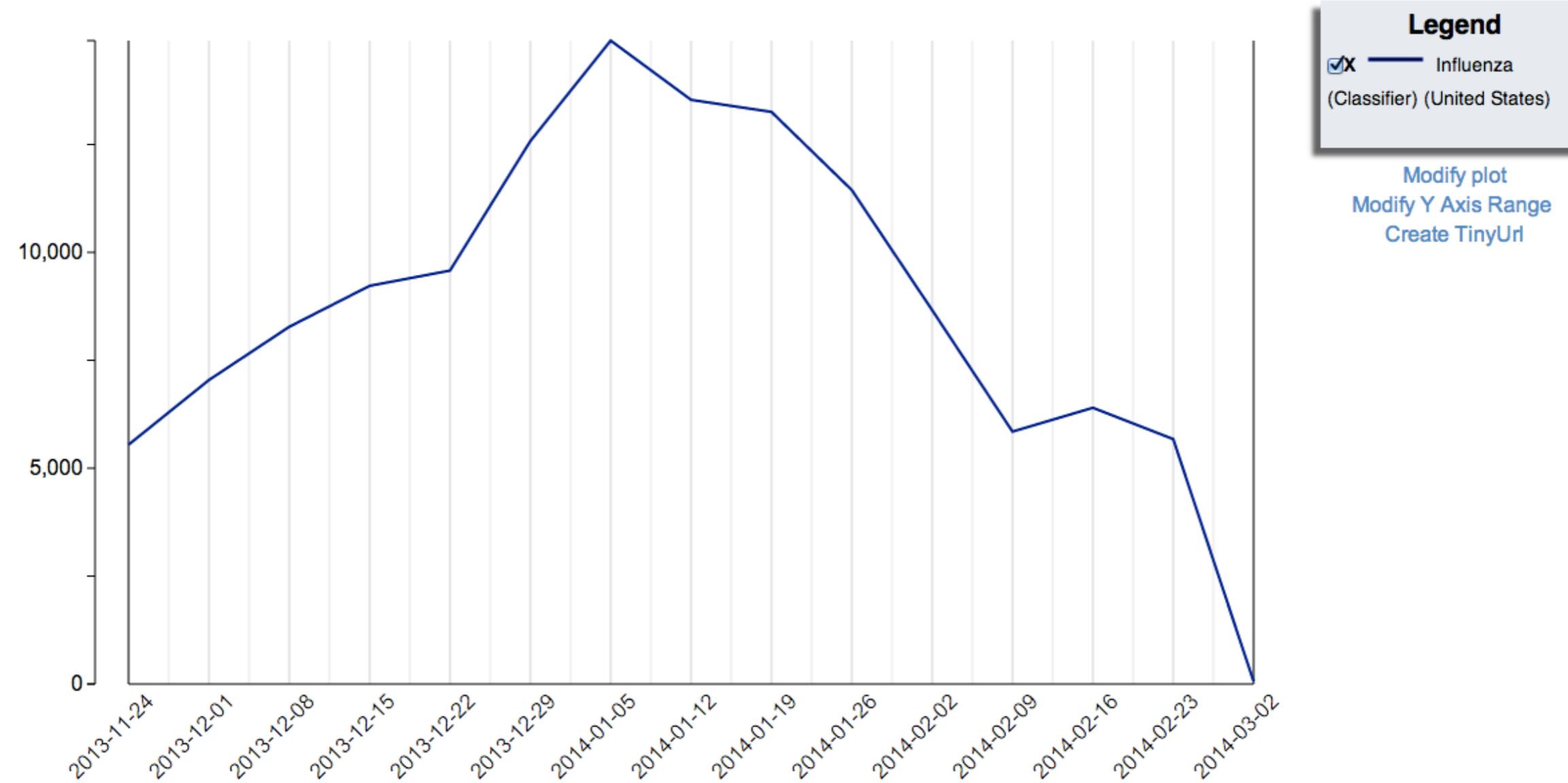


[Make your own](#)

Last updated: Feb. 24, 2014 | Processed 2,804,320,831 tweets since July 18, 2011.

©2014 by the [Johns Hopkins Social Media and Health Research Group](#) | [contact@healthtweets.org](mailto:contact@healthtweets.org)

Data on this site is collected from third-party social media. Reported health information may be incomplete or inaccurate.

**Legend**

**Influenza**  
(Classifier) (United States)

[Modify plot](#)  
[Modify Y Axis Range](#)  
[Create TinyUrl](#)

Create a plot by selecting plot options below.

Select plot resolution.

Week ▾

Select plot value type.

- Normalized Counts  
 Raw Counts

Select plot date range.

From: 11/26/2013

[First date](#)

To: 2/24/2014

[Last date](#)

Data to plot (maximum 4 plot lines.)

Disease: Influenza (Classifier)

[Remove](#)

Location: United States

[Add data](#)

[Plot](#)

Last updated: Feb. 24, 2014 | Processed 2,804,320,831 tweets since July 18, 2011.

©2014 by the [Johns Hopkins Social Media and Health Research Group](#) | [contact@healthtweets.org](mailto:contact@healthtweets.org)

Data on this site is collected from third-party social media. Reported health information may be incomplete or inaccurate.

Create a plot by selecting plot options below.

Select plot resolution.

Week ▾

Select plot value type.

- Normalized Counts  
 Raw Counts

Select plot date range.

From: 11/26/2013

[First date](#)

To: 2/24/2014

[Last date](#)

Data to plot (maximum 4 plot lines.)

Disease: Influenza (Classifier)

[Remove](#)

Location: United States

Disease

Influenza (Classifier) ▾

[Remove](#)

Location

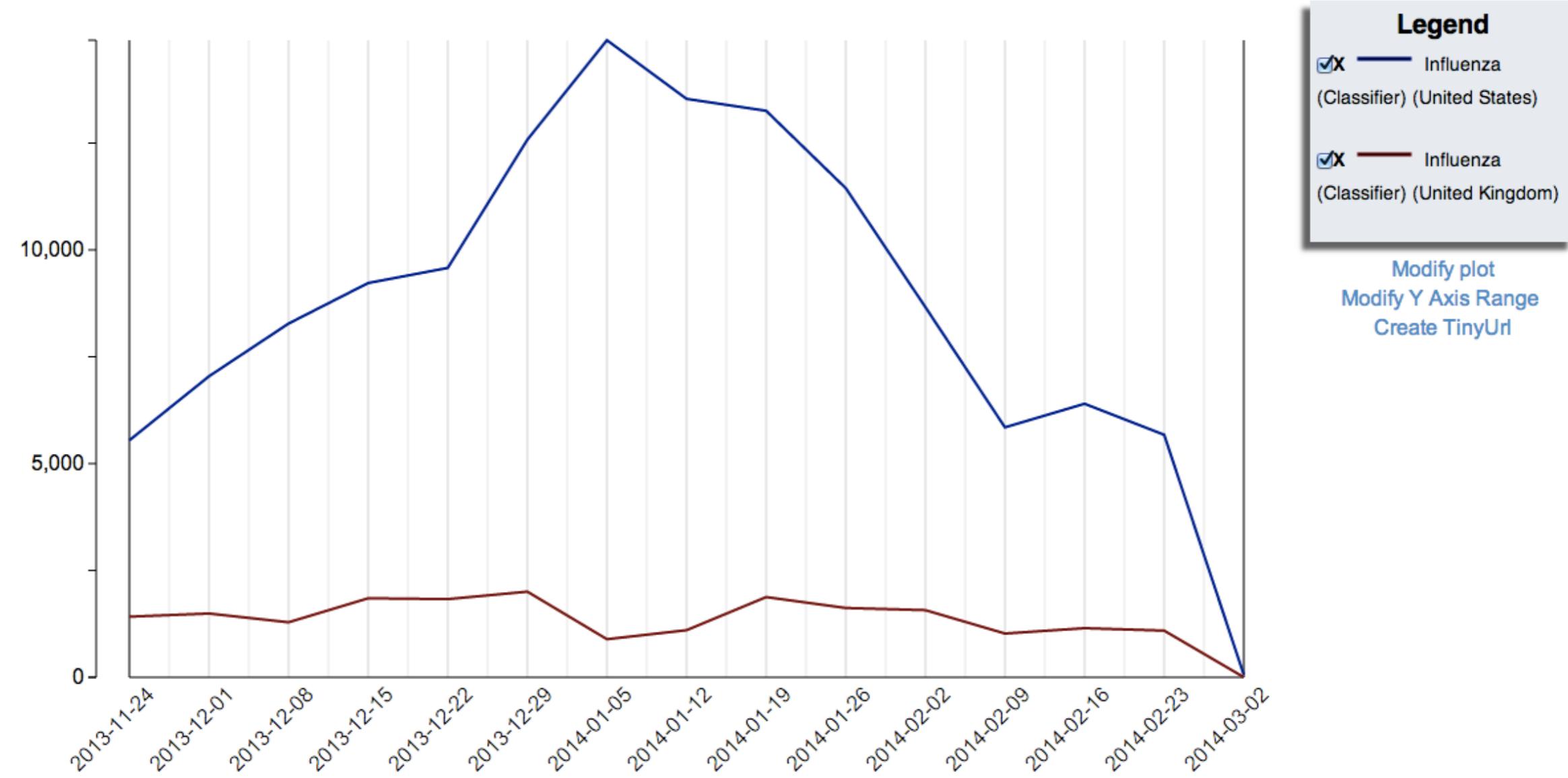
▼ Select

United Kingdom ▾

No Cities Available ▾

[Add data](#)

[Plot](#)

**Legend**

— Influenza  
(Classifier) (United States)

— Influenza  
(Classifier) (United Kingdom)

[Modify plot](#)  
[Modify Y Axis Range](#)  
[Create TinyUrl](#)

Select a map.



**United States**



**Europe**



**World**

Last updated: 01/11/2013 | Processed 4,176,681,828 tweets since 05/25/2009.

©2013 by Mark Dredze.

Data on this site is collected from third-party social media. Reported health information may be incomplete or inaccurate.  
[contact@healthtweets.org](mailto:contact@healthtweets.org)

Database:

**Map Options**

Resolution: Week

Date: 01/01/2013

Disease:

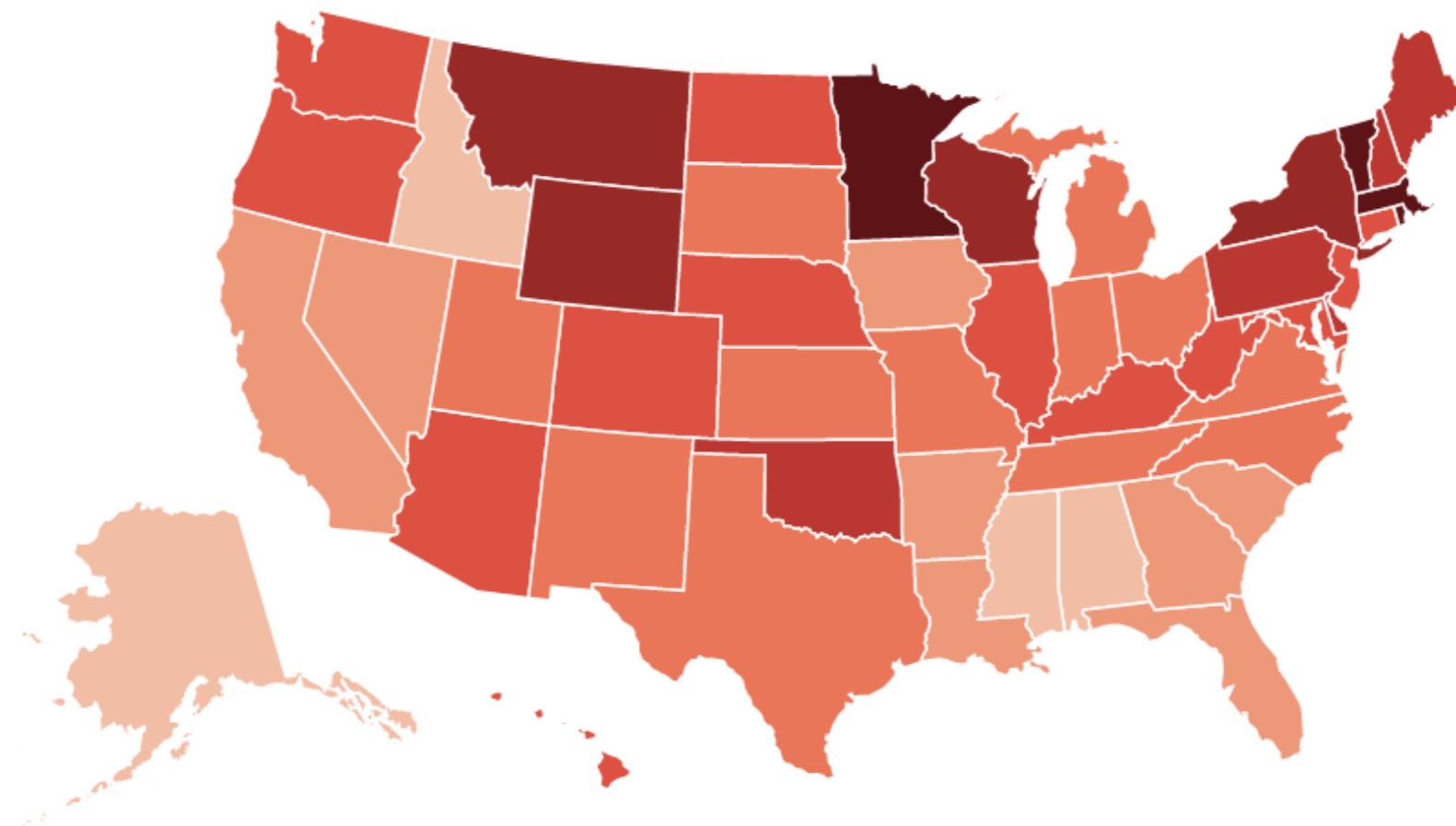
Influenza (Classifier)

Show  Animate  Replay

Date: Tuesday, January 01, 2013

Resolution: Week

Disease: Influenza (Classifier)



-Select state -

Shading granularity (maximum value for range): 0.03275 (68%)

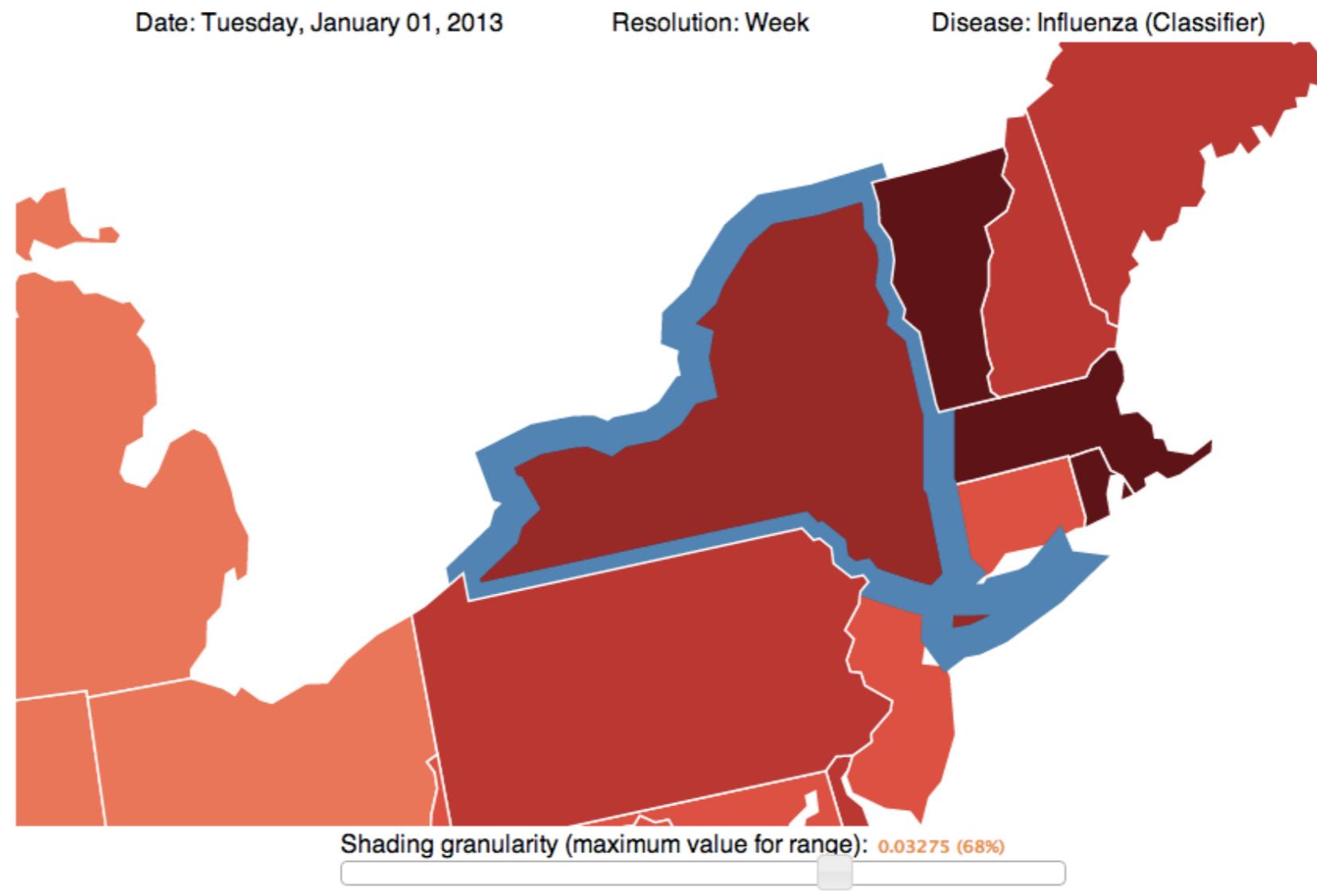
**Map Options**

Resolution: Week

Date: 01/01/2013

Disease: Influenza (Classifier)

Show  Animate  Replay



**Map Options**

Resolution: Week

Date: 01/01/2013

Disease: Influenza (Classifier)

Show  Animate  Replay

-Select state -

Date: Tuesday, January 01, 2013

Resolution: Week

Disease: Influenza (Classifier)

**Create Animation**

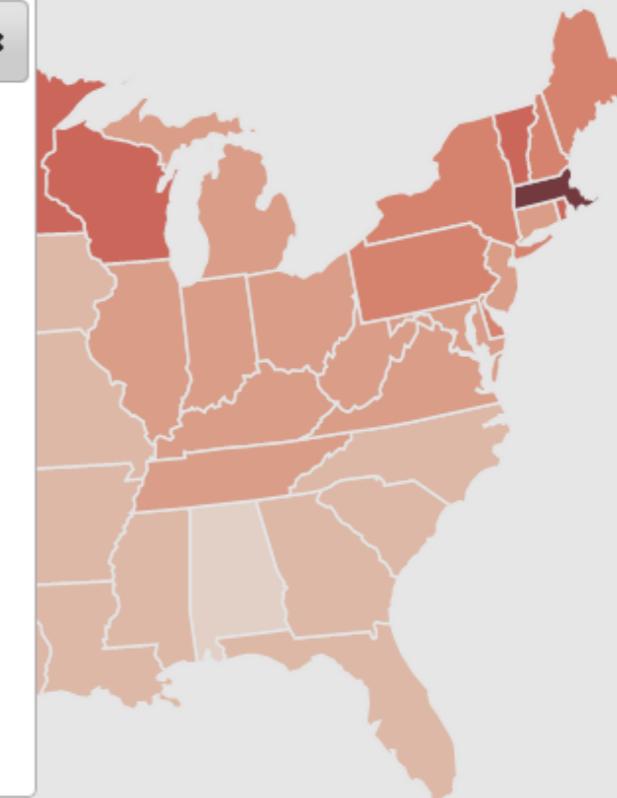
**Animation Options**

Resolution:

Transition time (ms): 100

From:  [First date](#)

To:  [Last date](#)



Shading granularity (maximum value for range): 0.048159 (100%)

United States  
Friday, January 11, 2013  
Change date:  [Most recent date](#)

[587 Known Cities](#)

### Friday, January 11, 2013

Total Tweets: 210,499  
Total Health Tweets: 192,226

Anthrax  
Rate: 0.0  
Count: 0

Bird Flu  
Rate: 0.000180523422914  
Count: 38

Cholera  
Rate: 1.42518491774e-05  
Count: 3

Common Cold  
Rate: 0.0797153430658  
Count: 16,780

Diphtheria  
Rate: 0.0  
Count: 0

Flu Vaccine  
Rate: 0.0194062679633  
Count: 1,085

### Week of January 06, 2013

Total Tweets: 0  
Total Health Tweets: 0

Anthrax  
Rate: 0  
Count: 0

Bird Flu  
Rate: 0  
Count: 0

Cholera  
Rate: 0  
Count: 0

Common Cold  
Rate: 0  
Count: 0

Diphtheria  
Rate: 0  
Count: 0

Flu Vaccine  
Rate: 0  
Count: 0

### Month of January 2013

Total Tweets: 941,287  
Total Health Tweets: 11,733

Anthrax  
Rate: 0.0  
Count: 0

Bird Flu  
Rate: 9.56137713577e-06  
Count: 9

Cholera  
Rate: 0.0  
Count: 0

Common Cold  
Rate: 0.000566246001485  
Count: 533

Diphtheria  
Rate: 0.0  
Count: 0

Flu Vaccine  
Rate: 3.29336323566e-05  
Count: 21

### Since Monday, May 25, 2009

Total Tweets: 443,715,909  
Total Health Tweets: 96,480,606

Anthrax  
Rate: 1.57533229218e-06  
Count: 699

Bird Flu  
Rate: 5.85802750651e-05  
Count: 25,993

Cholera  
Rate: 1.00514764279e-06  
Count: 446

Common Cold  
Rate: 0.0178856985721  
Count: 7,936,169

Diphtheria  
Rate: 2.41145286499e-07  
Count: 107

Flu Vaccine  
Rate: 0.000379907496172  
Count: 169,571

# Future Directions

- Share site with public health officials
  - Deliver metrics that are of immediate practical use to decision makers
  - Receive feedback for improving these metrics

# Thank You

- [www.healthtweets.org](http://www.healthtweets.org)
- Tell your friends!
- Email for an account: contact@healthtweets.org