

Make your own

### Michael J. Paul Johns Hopkins University.

©2014 by the Johns Hopkins Social Media and Health Research Group | 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



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



week ending 7/27/2014

ima

ering

city during the specified week. The cities are colorcoded by the HHS region to which they belong (see legend). User Contributed Data

Hover on a city to see the name and ILI for the week. Select a city by clicking on a bubble on the map. To select multiple cities define a region on the map (left-click and drag; release when area is selected). To clear selection click on the greyed out background of the map.

proportional to the ILI incidence reported for the

week. The size of the bubble for a city is

Selection can also be made using the tree map in the bottom left panel. Clicking on a colored rectangle(see caption below), highlights all cities in that state. To remove selection, click on white space below the heatmap as indicated.

The bottom right panel plots predictions for the selected cities (either from the map or the heatmap) during the remainder of the season. Hover on the markers to retrieve ILI estimates and standard deviations. In cases where the chart has several overlapping plots, select a marker to highlight a city's plot (select multiple cities by defining a region as described above). The width of a curve is proportional to the standard deviation of the estimate whereas brightness is inversely related i.e. a leaner, brighter plot indicates a more confident estimate.

Please use the tabs at the top to look at additional visualizations. To view forecasts for each city use tab 2. Tab 3 allows for comparison of the forecasts from different models and tab 4 to compare the current season with previous three seasons. Within each tab use the selection hoves on the

NC

TX







### Crowdbreaks

Disease surveillance, #crowdsourced. Powered by you.



### Help us make crowdbre With a single click, you can help us improve o

answering the questions below. [How does thi

### Is this tweet about Influenza

"Before I sleep

### I. Sexually transmitted diseases

change prev. day: -9.78% | change prev. week: -9.92% 14 days ago 7 days ago

### sickweather

•Man hes

Map Satellite

Hampshire

Massachusetts Bost



Live Map

Alerts

Widgets

Sign In



### 🚏 Public Health Emergency

NowTrending.HHS.gov beta Trends - by condition - by location -

### Sickweather - Sickness Forecas Mapping

Everyday thousands of people around the glob media sites like Facebook and Twitter when the close to them) get sick. Just as Doppler radar so indicators of bad weather, Sickweather scans so indicators of illness, allowing you to check for t sickness as easily as you can check for the chan See How Sickweather works

Download on the

### 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)

U.S. Department of Health & Human Services

### **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

### ad more about how we calculate this)

### **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** Home About Trends

Tracking Health Trends via Social Media

User: (user) (Logout)

### **Health Tweets**

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



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





Tracking Health Trends via Social Media

Create a plot by selecting plot options below.

Select plot resolution.	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.)

Plot

I	Disease: Influenza (Classifier)	Remove
I	Location: United States	

Add data

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 Data on this site is collected from third-party social media. Reported health information may be incomplete or inaccurate. Tracking Health Trends via Social Media

Create a plot by selecting plot options below.

Select plot resolution.	Select plot value type. O 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.)

Plot

Disease: Influenza (Classifier) Location: United States	Remove
Disease Influenza (Classifier) 💠	Remove
Location	
- Select	
United Kingdom 🗘 No Cities Available ≑	

Add data



### Select a map.



### 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 \$

Database: -Select Database-



-Select state -

÷



![](_page_24_Figure_2.jpeg)

Map Options	Date: Tuesday, January 01, 2013 Resolution: Week	Disease: Influenza (Classifier)
Resolution: Week +		
Date: 01/01/2013	Create Animation	-
Disease:	Create Animation	
Influenza (Classifier) +	Animation Options	a family and the second second
Show Animate Replay	Resolution: -Select- +	
	Transition time (ms): 100	
-Select state - \$	From: First date	3 The
	To: Last date	- C - Fraze
		2 hour Sul
	Proporto Animotion	1
	Prepare Animation	4
	Shading granularity (maximum value for re-	
	Snauing granularity (maximum value for ra	IGE). 0.048159 (100%)

![](_page_25_Picture_3.jpeg)

![](_page_26_Figure_2.jpeg)

# 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
- Tell your friends!
- Email for an account: contact@healthtweets.org