# WORLDWIDE INFLUENZA SURVEILLANCE THROUGH TWITTER

Michael J. Paul, Johns Hopkins University
Mark Dredze, Johns Hopkins University
David A. Broniatowski, George Washington University
Nicholas Generous, Los Alamos National Laboratory

93° *All-indigo rainbow*

the ONION
America's Finest News Source

VIDEO · POLITICS · SPORTS · SCIENCE/TECH · LOCAL · ENTERTAINMENT

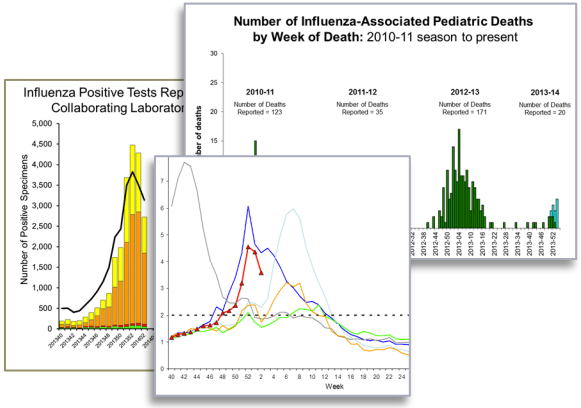
# Hip, Laid-Back Doctor Refers To Influenza As 'The Flu'

NEWS IN PHOTOS · Doctors · Local · Disease · Healthcare · ISSUE 50·45 · Nov 14, 2014

Share on Facebook 10.3K Share on Twitter 402 78

# INFLUENZA SURVEILLANCE

- Government flu monitoring is the gold standard
  - But reports have a delay of ~2 weeks
    (or longer, if the government shuts down ☺)




- Text-driven systems can produce estimates **immediately**
  - This study: **tweets**
    - advantage: huge, public, free

# INFLUENZA SURVEILLANCE

Most systems have focused on the United States

- CDC ILINet is the gold standard for US
  - Sentinel network of thousands of US providers
  - Hospitals report % of outpatients seen for influenza-like illness
  - Weekly reports of estimated ILI prevalence
- Data is lagged by 5-12 days
  - This lag can be even longer in other locations

# WORLDWIDE SURVEILLANCE

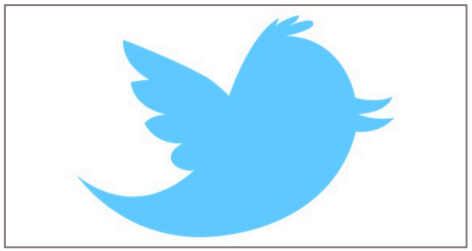
We evaluated Twitter flu surveillance for 10 English-speaking countries

- We also have ongoing work evaluating several US states/ counties/cities

| Location | Source, URL, and Description |
|---|---|
| Australia | Department of Health<br>http://www.health.gov.au/internet/main/publishing.nsf/Content/cda-surveil-ozflu-flucurr.htm<br>ILI Sentinel taken from the Australian Influenza Surveillance Report figure 6: "Weekly rate of ILI reported from GP ILI surveillance systems" with the unit described as "Rate per 1,000 consultations". Weeks begin Monday. |
| Canada | Public Health Agency of Canada<br>http://www.phac-aspc.gc.ca/fluwatch/13-14/index-eng.php<br>ILI Sentinel taken from FluWatch Report figure 5: "Influenza-like-illness (ILI) consultation rates by report week" with the unit described as "Rate per 1,000 patient visits". Weeks begin Sunday. |
| Ireland | Health Protection Surveillance Centre<br>http://www.hpsc.ie/A-Z/Respiratory/Influenza/SeasonalInfluenza/Surveillance/InfluenzaSurveillanceReports/<br>ILI Sentinel taken from Influenza Surveillance Report figure 1: "ILI sentinel GP consultation rates per 10,000 population" with the unit described as "ILI rate per 100,000 population". Weeks begin Monday. |
| New Zealand | Institute of Environmental Science and Research<br>https://surv.esr.cri.nz/virology/influenza_weekly_update.php<br>ILI Sentinel taken from Influenza Weekly Update figure 2: "Weekly consultation rates for influenza-like illness in New Zealand, 2010-2014" with the unit described as "Consultation rate (per 100,000)". The Influenza Weekly Update only reports during the influenza season in New Zealand which typically lasts between weeks 18 to 44. Weeks begin Monday. |
| South Africa | National Institute of Communicable Diseases<br>http://www.nicd.ac.za/?page=surveillance_bulletin&id=15<br>Hospital consultation data taken from National Institute of Communicable Diseases Monthly Surveillance Bulletin. The unit measured is the number of private hospital outpatient consultations with a discharge diagnosis of pneumonia and influenza. Weeks begin Sunday. |
| United Kingdom | Public Health England<br>http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/SeasonalInfluenza/EpidemiologicalData/<br>ILI Sentinel taken from the National Influenza Report in the tables from the "Weekly consultation rates in national sentinel schemes" section. Weeks begin Monday. |
| United States | Centers for Disease Control and Prevention<br>http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html<br>ILI Sentinal data from the U.S. Outpatient Influenza-like Illness Surveillance Network (ILINet). The CDC coordinates the network and publishes weekly reports showing the percentage of outpatient consultations for ILI. National rates as well as rates for the 10 HHS regions are available. Weeks begin Sunday. |

# TWITTER FLU DETECTION

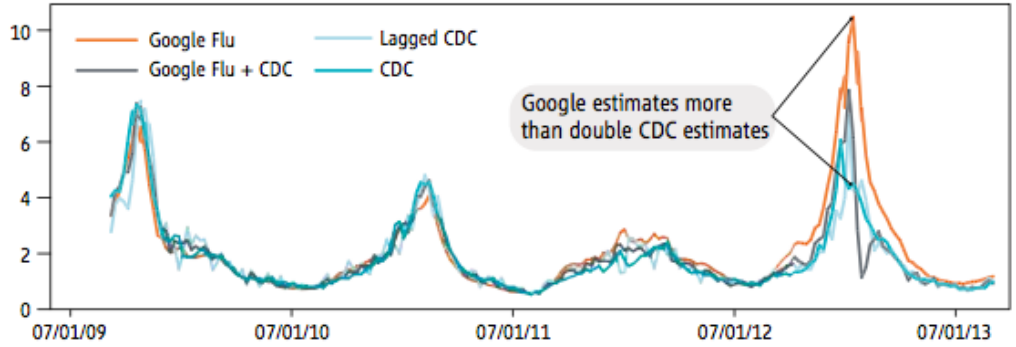
- We used our state-of-the-art Twitter system
  - Lamb et al (2013) and Broniatowski et al (2013)
- Two streams downloading data since Nov 2011
  - 1% sample and stream filtered for health keywords
  - About 4 million per day
- Cascade of tweet classifiers:
  - Relevant to health
  - Relevant to flu
  - Indicates flu **infection** (vs general awareness)
- Can produce daily or weekly prevalence estimates



$$\frac{\text{\# of tweets classified as flu infection}}{\text{\# of tweets from full sample}}$$

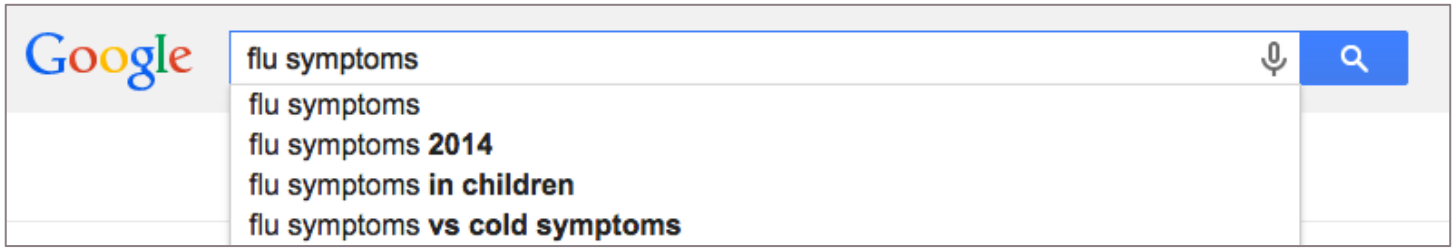
# TWITTER FLU DETECTION

**The infection vs awareness distinction matters!**




From Lazer et al., *Science*, 2014

Google concluded that media attention was a primary cause of their huge overestimate in 2012-2013




"flu symptoms" – not an experiential query

# TWITTER FLU DETECTION

## Features:

- Stylometry
  - Retweets, user mentions, URLs, emoticons
- 8 manually created word classes

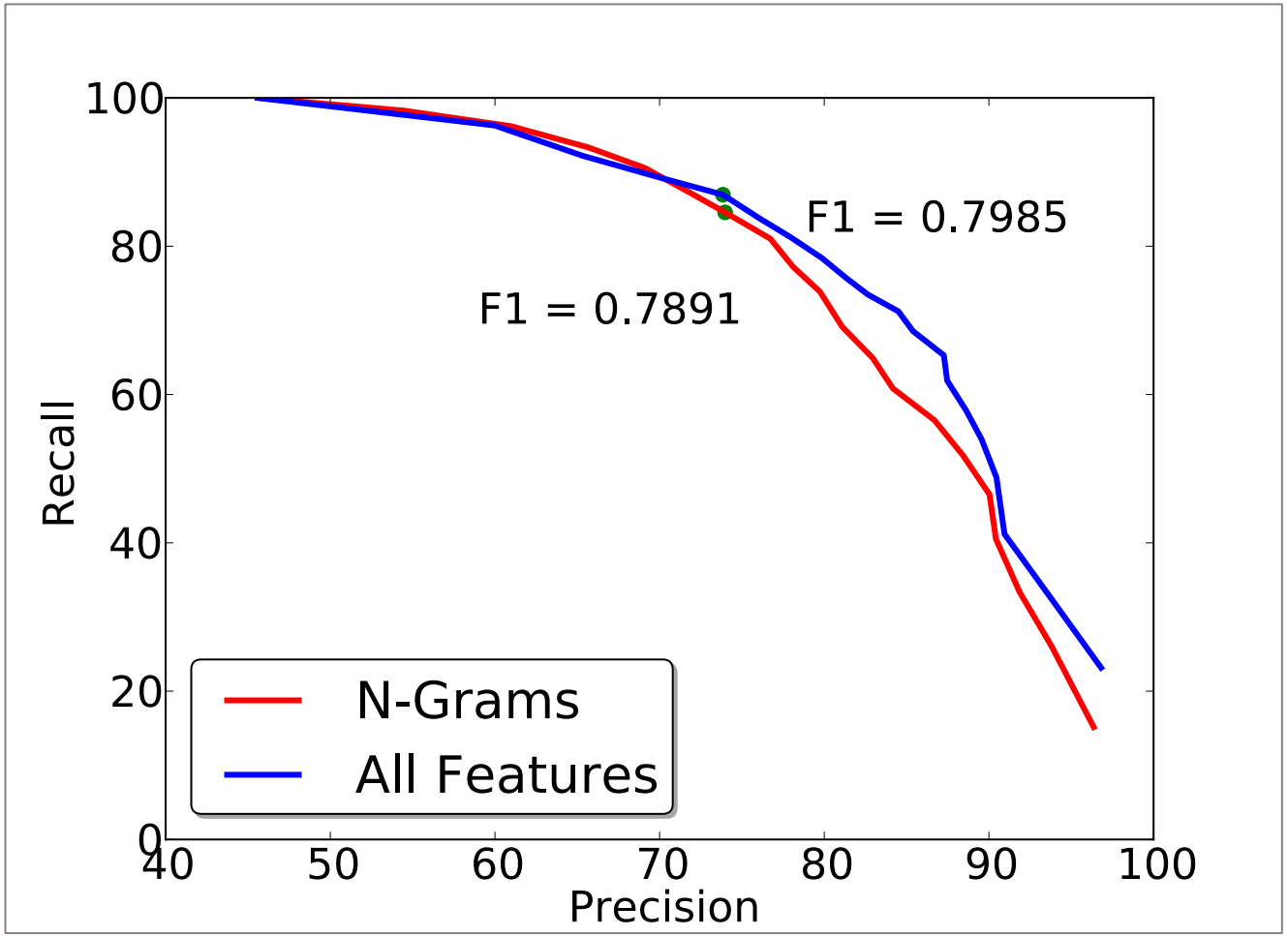
| | |
|---|---|
| Infection | getting, got, recovered, have, having, had, has, catching, catch, cured, infected |
| Disease | bird, flu, sick, epidemic |
| Concern | afraid, worried, scared, fear, worry, nervous, dread, dreaded, terrified |
| Treatment/ Prevention | vaccine, vaccines, shot, shots, mist, tamiflu, jab, nasal spray |
| … | … |

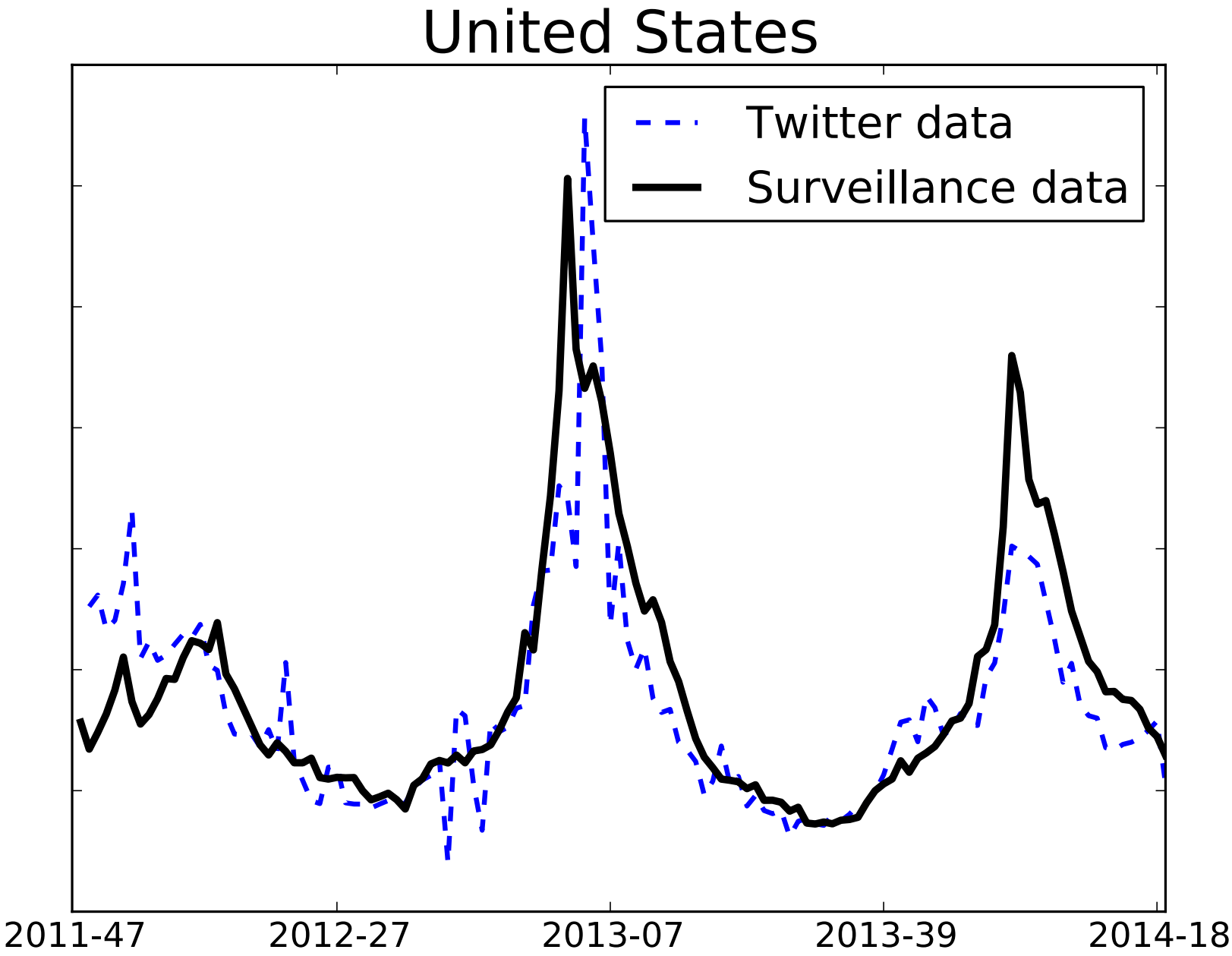
# TWITTER FLU DETECTION

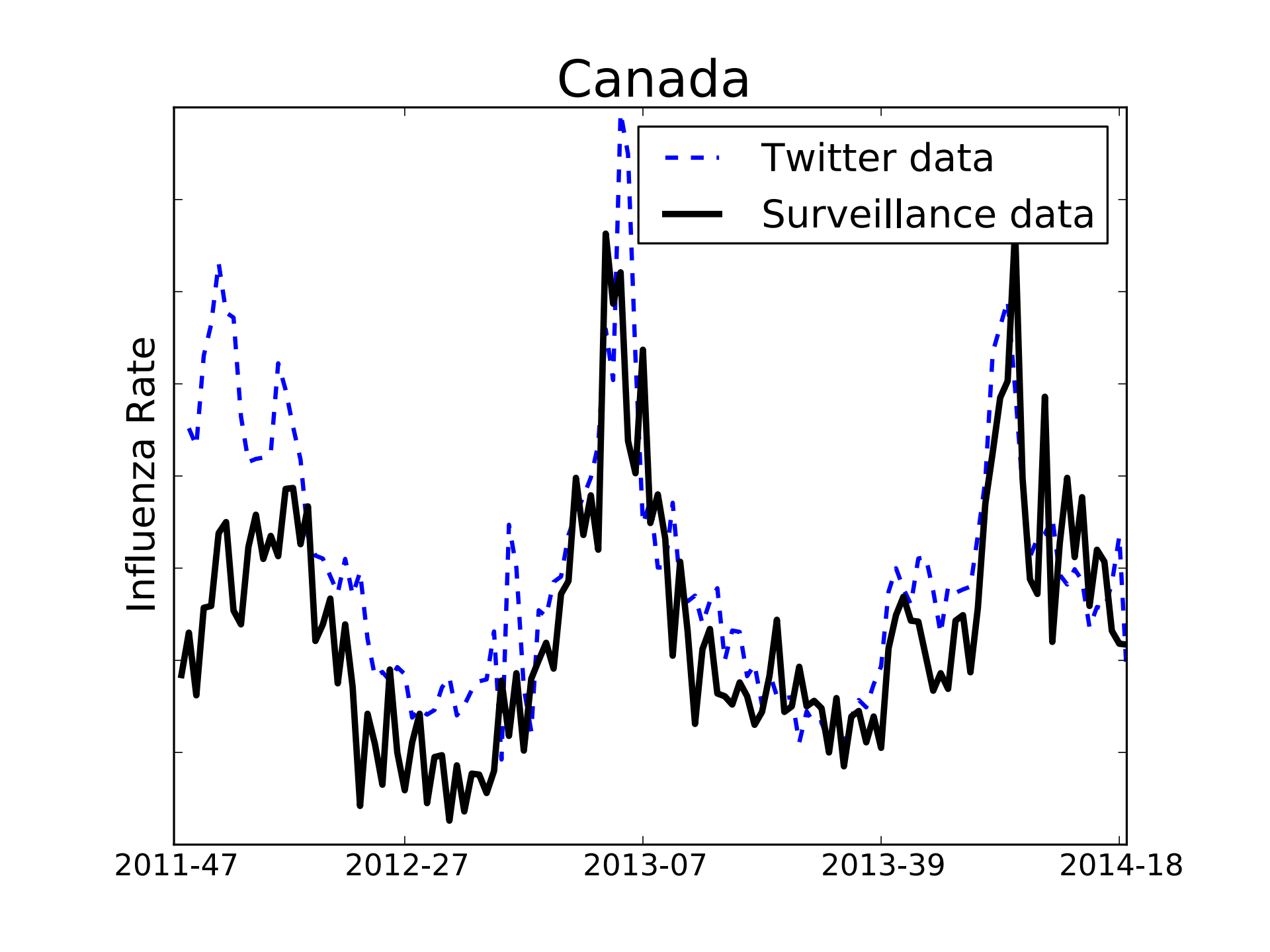
**Features:**

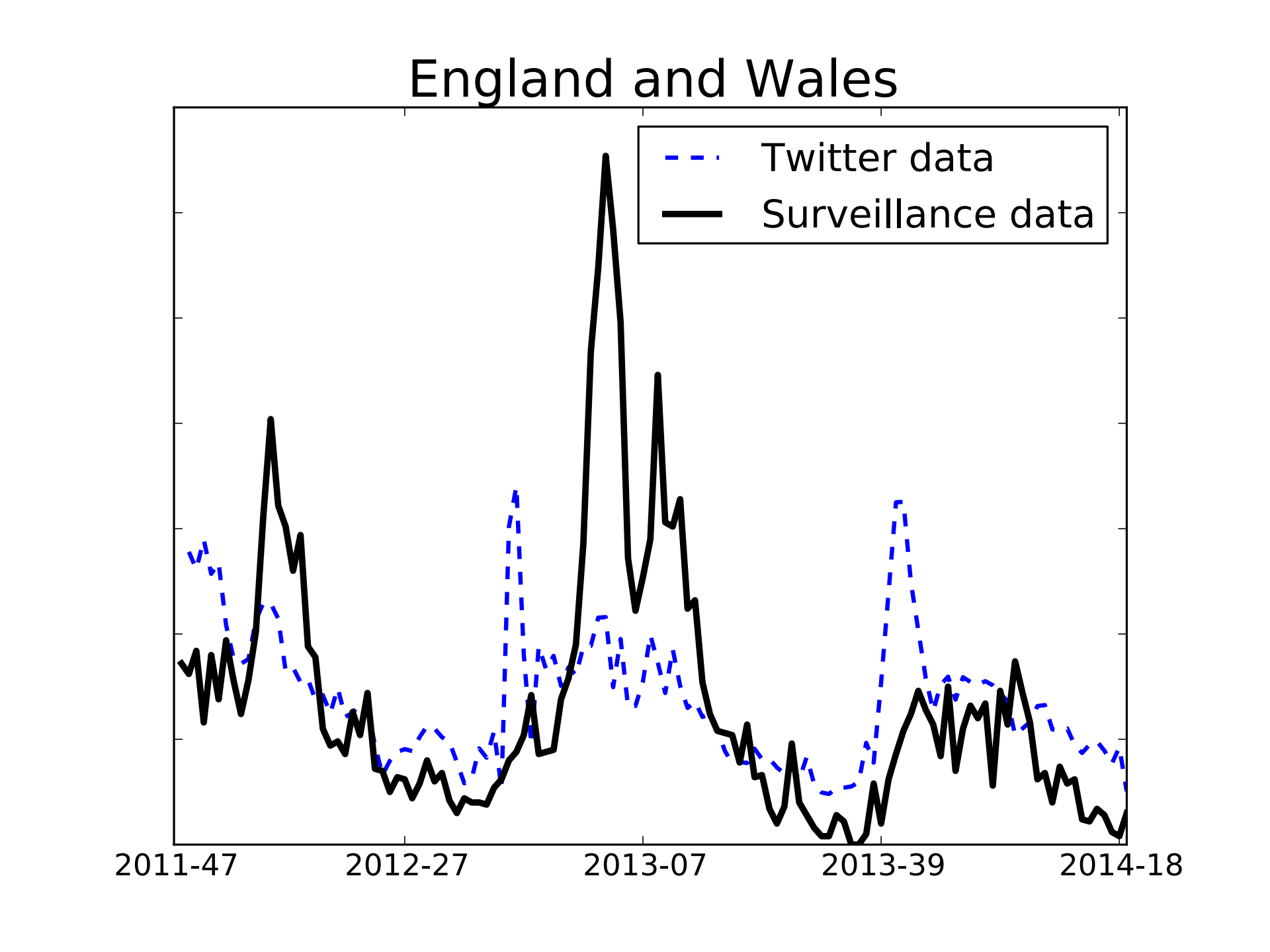
- Part of speech templates
  - (subject,verb,object) tuples
    - always a good feature, IMO
  - numeric references
    - "100 more cases of swine flu"
  - whether "flu" is a noun or adjective
    - "tired of the flu" vs "tired of the flu hype"
  - whether "flu" is the subject or object
    - "I have the flu" vs "the flu is going around"
  - … and others

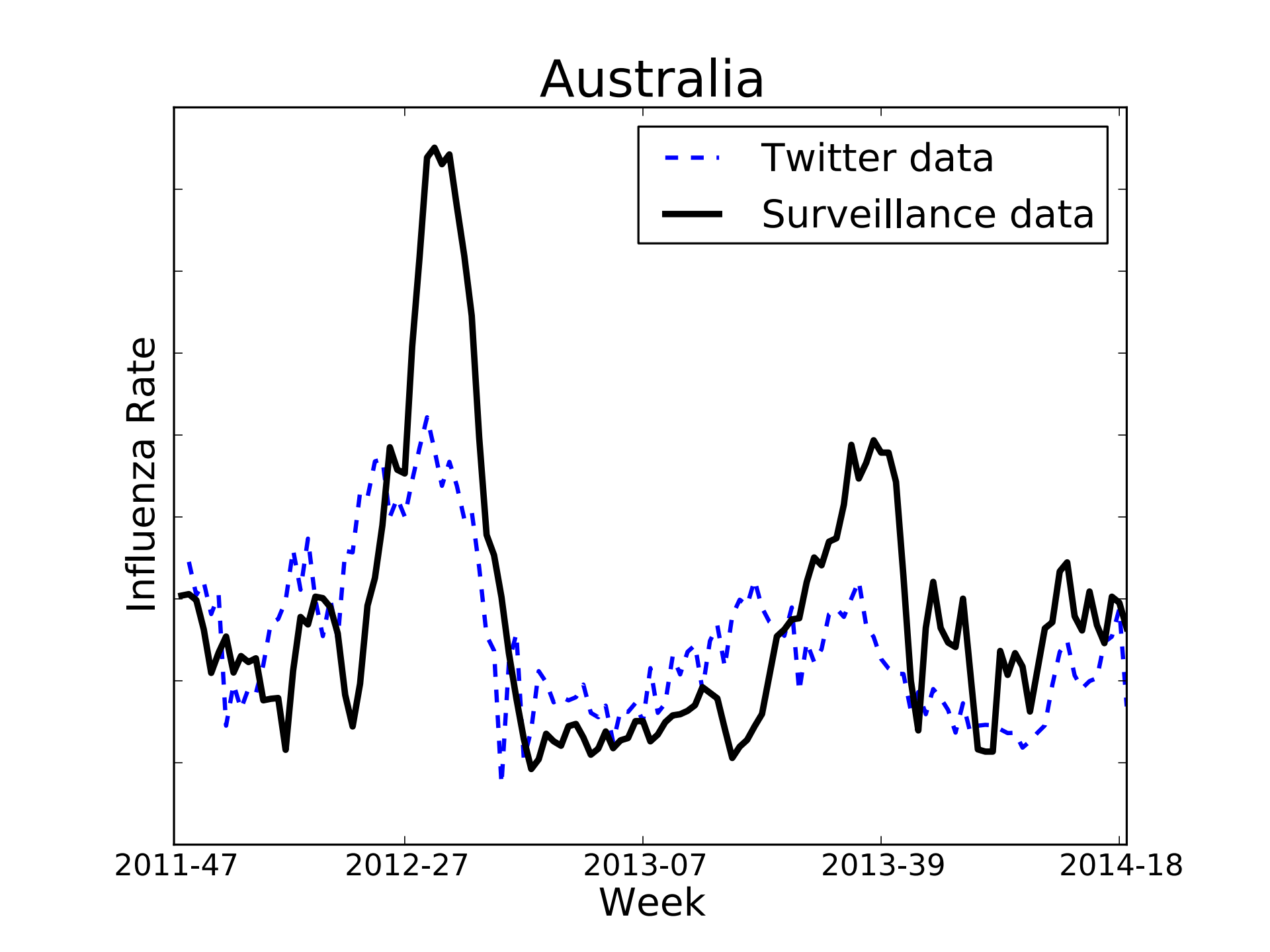
# TWITTER FLU DETECTION

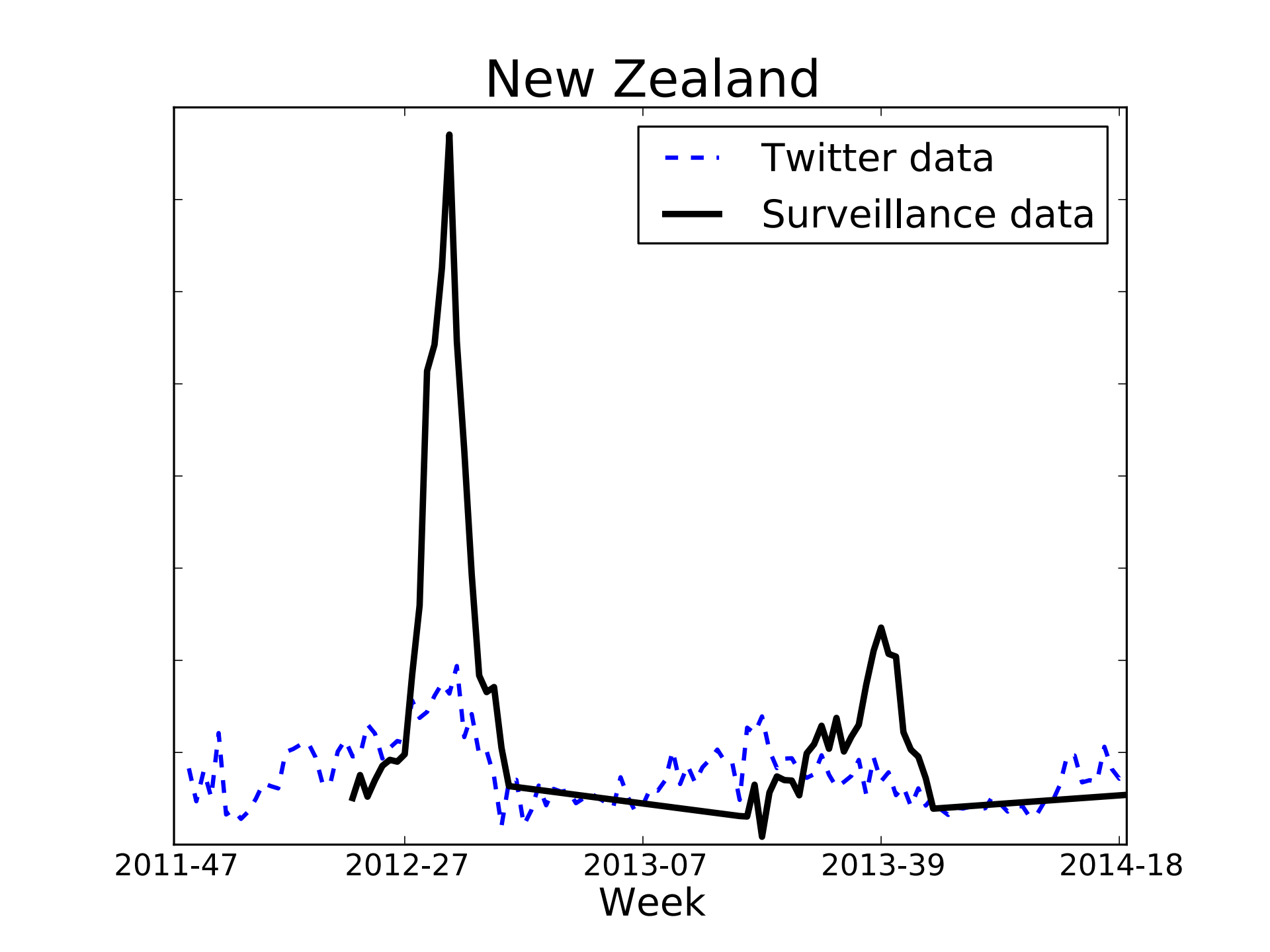


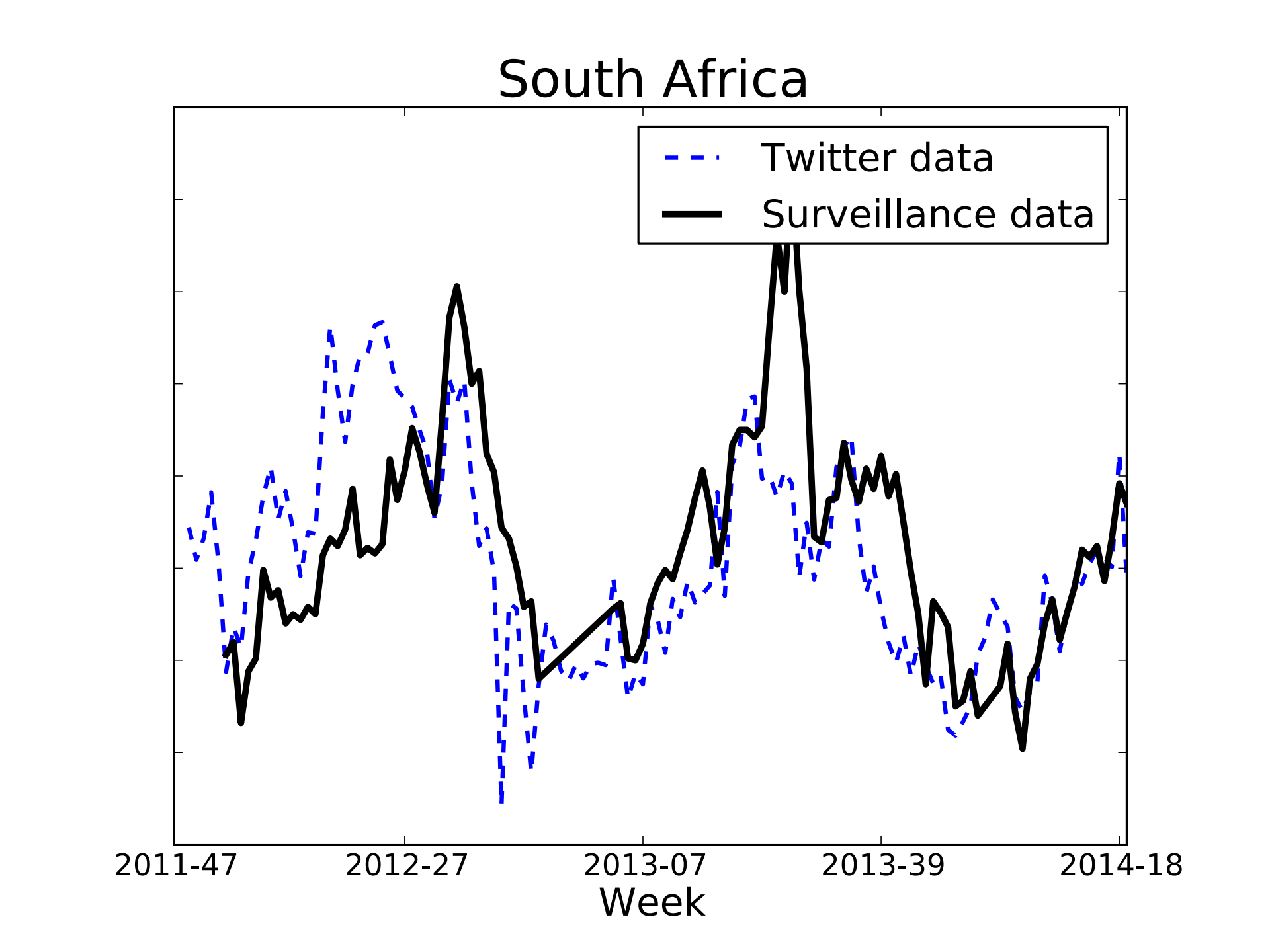

United States
Twitter data
Surveillance data
2011-47
2012-27
2013-07
2013-39
2014-18


Canada
Twitter data
Surveillance data
Influenza Rate
2011-47
2012-27
2013-07
2013-39
2014-18


England and Wales
Twitter data
Surveillance data
2011-47
2012-27
2013-07
2013-39
2014-18


Australia
Twitter data
Surveillance data
Influenza Rate
2011-47
2012-27
2013-07
2013-39
2014-18
Week


New Zealand
Twitter data
Surveillance data
2011-47
2012-27
2013-07
2013-39
2014-18
Week


South Africa
Twitter data
Surveillance data
2011-47
2012-27
2013-07
2013-39
2014-18
Week

# NOWCASTING MODEL

- Linear autoregressive model:

$$\hat{y}_w = \alpha_0 y_{w-\ell} + \alpha_1 y_{w-\ell-1}$$

- This is a strong baseline

# NOWCASTING MODEL

- Linear autoregressive exogenous model:

$$\hat{y}_w = \gamma z_w + \alpha_0 y_{w-\ell} + \alpha_1 y_{w-\ell-1}$$

Weekly estimate from
our Twitter system

# NOWCASTING EVALUATION

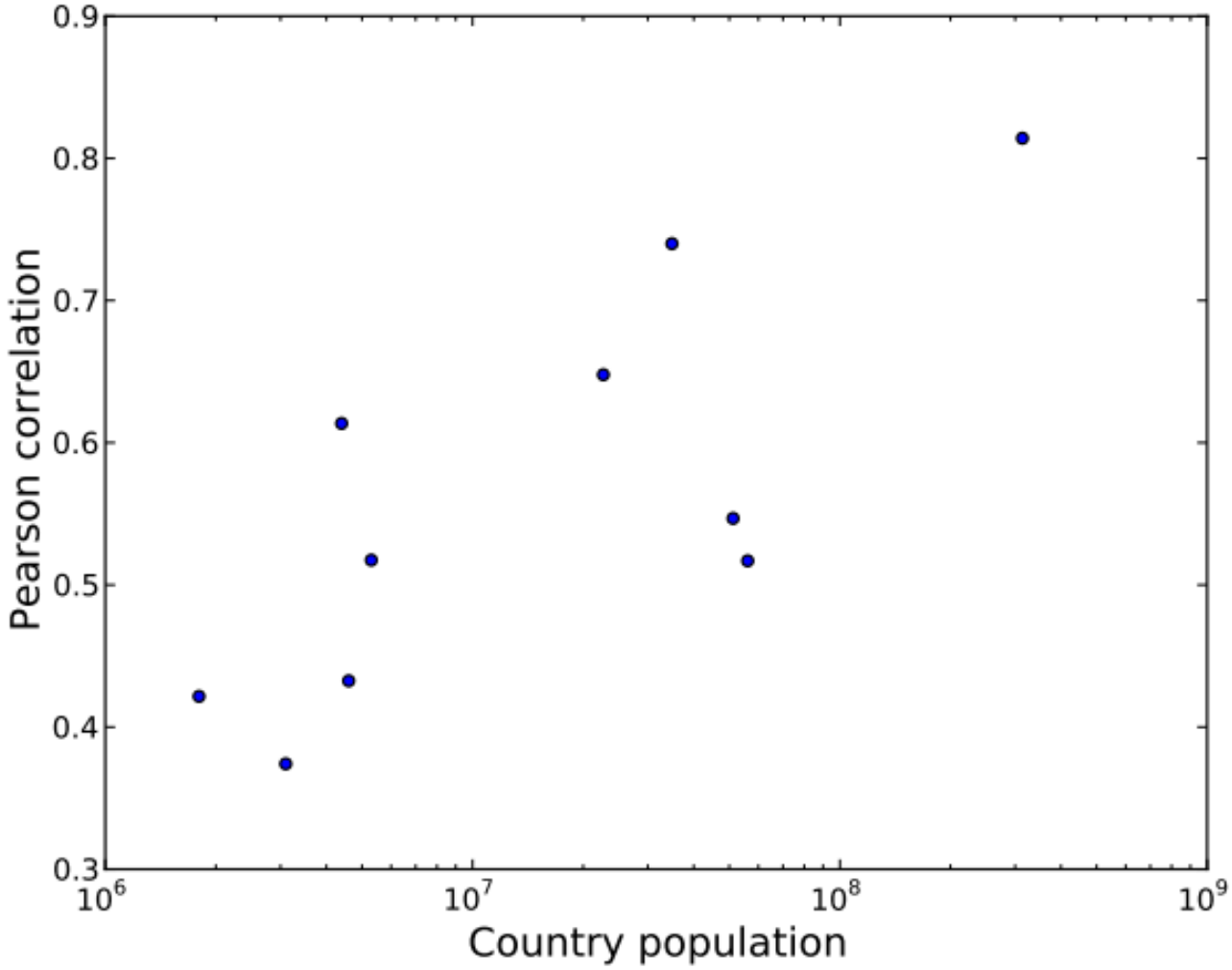
| Location | Pop. | # Tweets / week | | Delay | $n$ | Time Range |
|---|---|---|---|---|---|---|
| | | All | Flu | | | |
| Australia | 22.7m | 60,332 | 304 | 12–26 | 125 | 201149 – 201418 |
| Canada | 34.9m | 137,608 | 544 | 6–20 | 125 | 201149 – 201418 |
| England+Wales | 56.1m | 339,387 | 1,935 | 4–11 | 122 | 201149 – 201418 |
| Ireland | 4.6m | 30,180 | 211 | 4–11 | 113 | 201149 – 201418 |
| New Zealand | 4.4m | 9,003 | 46 | 3–10 | 45 | 201220 – 201344 |
| Northern Ireland | 1.8m | 6,415 | 46 | 4–11 | 122 | 201149 – 201418 |
| Scotland | 5.3m | 32,212 | 184 | 4–11 | 122 | 201149 – 201418 |
| South Africa | 51.2m | 33,095 | 495 | >30 | 105 | 201203 – 201418 |
| United States | 314.0m | 2.1m | 5,846 | 6–13 | 125 | 201149 – 201418 |
| Wales | 3.1m | 14,169 | 96 | 4–11 | 122 | 201149 – 201418 |

# NOWCASTING EVALUATION

| Location | $r$ | MSE Red. (%) | | |
|---|---|---|---|---|
| | | $\ell$=1 | $\ell$=2 | $\ell$=3 |
| Australia | 0.648* | 10.5 | 19.5* | 29.2* |
| Canada | 0.740* | 7.7 | 24.1* | 37.2* |
| England+Wales | 0.517* | -0.6 | 7.8 | 9.5 |
| Ireland | 0.433* | 1.6 | 5.4 | 7.4 |
| New Zealand | 0.614* | 18.0 | 37.0 | 59.2 |
| Northern Ireland | 0.422* | 5.3 | 6.1 | 8.7 |
| Scotland | 0.517* | -3.2 | -0.5 | 4.2 |
| South Africa | 0.547* | 5.6 | 17.2 | 25.3 |
| United States | 0.814* | 15.3 | 17.7 | 33.6* |
| Wales | 0.374* | 2.7 | 6.8 | 3.3 |

Results from 5-fold cross validation

# NOWCASTING EVALUATION



# NOWCASTING EVALUATION

- Our results are assume that the gold data is actually available, but this isn't usually true
  - For example, the CDC data is always revised in later weeks.
  - Our experiments used the revised data rather than the data that would have actually be available at the time
  - Therefore the nowcasts have unrealistically low error
    - In an earlier study, Twitter reduced error by 30% if we used the correct data, butonly  6% if we used the revised data

# CONCLUSION

- Our system works well in several countries
  - Tweet quality varies by country
  - But so does the quality of sentinel surveillance
- Limitation: only works in English-speaking locations
  - Want to train models for additional languages in future

# THANK YOU