



Mining Twitter Data to Augment NASA GPM Validation

Bill Teng^{1,2}, Arif Albayrak^{1,2}, George Huffman³, Bruce Vollmer²,
Carlee Loeser^{1,2}, Jim Acker^{1,2}

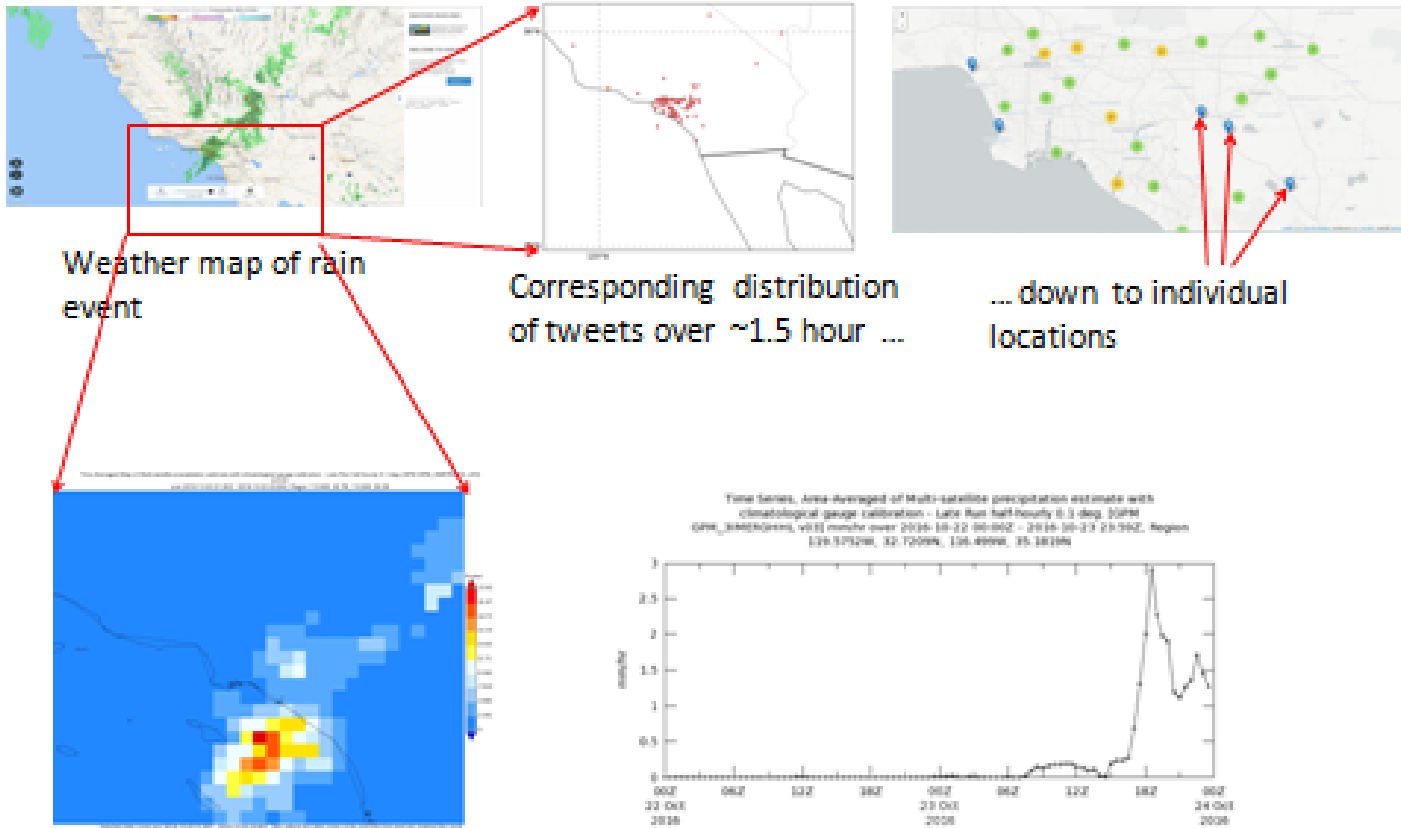
¹NASA GSFC 610.2; ²ADNET Systems, Inc.; ³NASA GSFC 612

NASA CESP (NNH16ZDA001N)

AGU 2017 Fall Meeting, Dec. 14, 2017

Motivation

Rain event in Los Angeles area



NASA Giovanni visualization of GPM map (left) and time series (right) of the LA area rain event



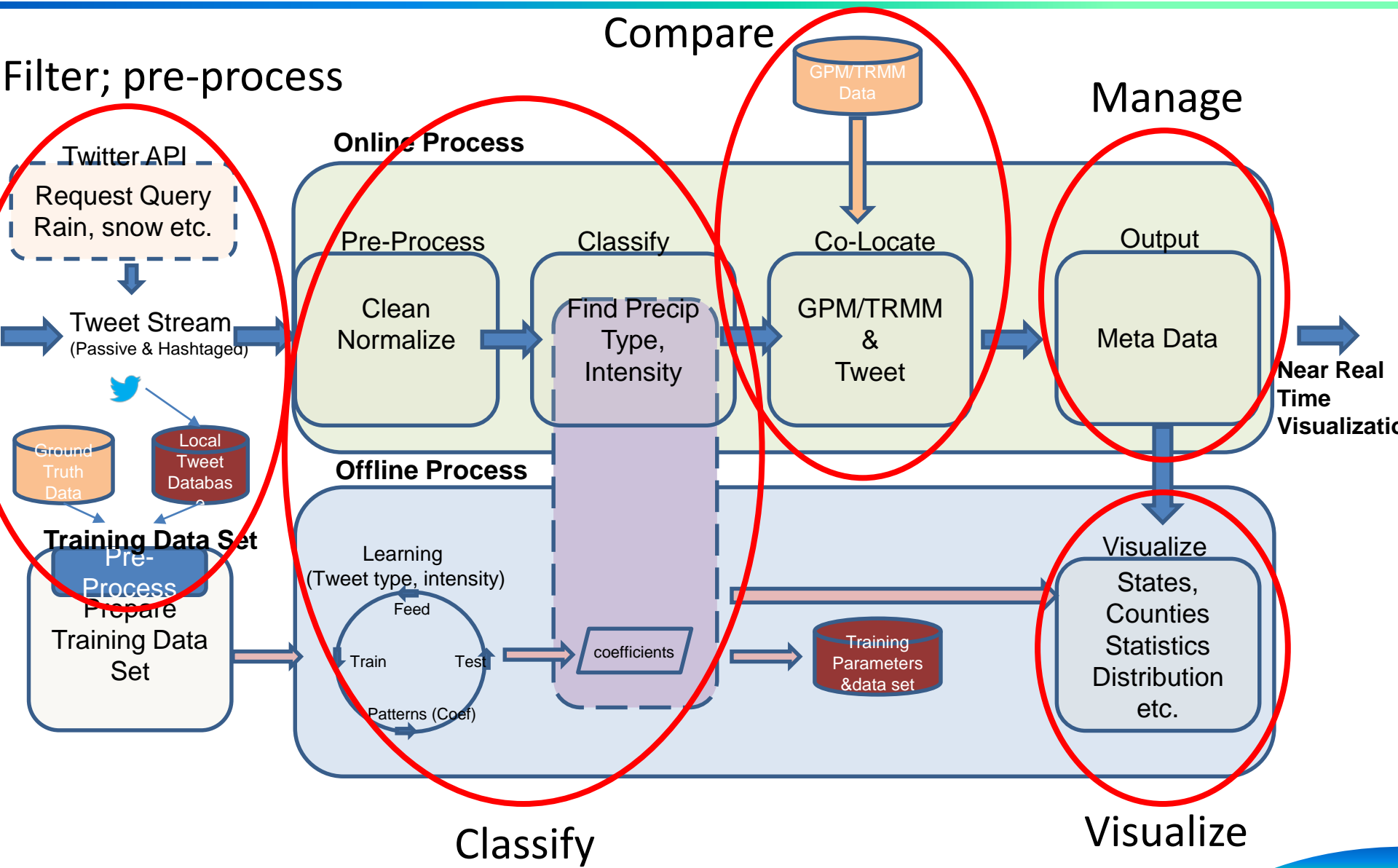
Organic network of rain gauges

Space-time-varying set of “precipitation tweets”

- Reading the “gauge measurements”
 - Develop infrastructure for processing and analyzing tweets
- Enhancing quality of tweets; engaging with “active” participants
- Applying processed tweets to satellite data validation
- Managing tweet data



Tweet processing infrastructure





Example tweets

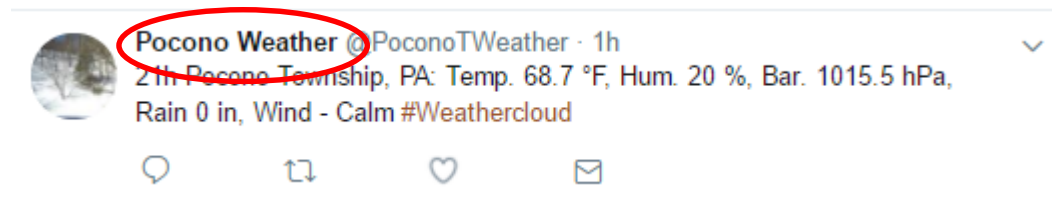


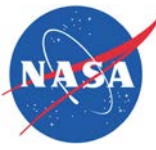
Relevant tweet

Not relevant tweet



Weather station tweet





Anatomy of a tweet (status object)

Tweet text

- `id`: The tweet's unique ID. These IDs are roughly sorted & developers should treat them as opaque (<http://bit.ly/dCkppc>).
- `text`: Text of the tweet. Consecutive duplicate tweets are rejected. 140 character max (<http://bit.ly/dCkppc>).
- `created_at`: The tweet's creation date.
- `in_reply_to_status_id`: The ID of another tweet that this tweet is in reply to. Won't be null unless the author of the referenced tweet is mentioned.
- `in_reply_to_screen_name`: The screen name & user ID of replied to tweet author.
- `in_reply_to_user_id`: The author's user ID.
- `retweeted`: Whether this tweet has been retweeted.
- `truncated`: Whether the text has been truncated to 140 characters. Only possible from SMS.
- `user`: The author's user info.

User info

- `screen_name`: The author's screen name.
- `name`: The author's user name.
- `bio`: The author's biography.
- `url`: The author's URL.
- `location`: The author's "location". This is a free-form text field, and there are no guarantees on whether it can be geocoded.
- `profile_background_color`: Rendering information for the author. Colors are encoded in hex values (RGB).
- `profile_image_url`: The author's profile image URL.
- `profile_text_color`: The creation date for this account.
- `created_at`: Whether this account has contributors enabled (<http://bit.ly/50npuu>).
- `contributors_enabled`: Whether this account has contributors enabled (<http://bit.ly/50npuu>).
- `fav_count`: Number of favorites this user has.
- `statuses_count`: Number of tweets this user has.
- `friends_count`: Number of users this user is following.
- `time_zone`: The timezone and offset (in seconds) for this user.
- `lang`: The user's selected language.
- `protected`: Whether this user is protected or not. If a user is protected, their tweets are not visible except to "friends".
- `followers_count`: Number of followers for this user.
- `verified`: Whether this user has a verified badge.
- `contributors`: The contributors' (if any) user IDs (<http://bit.ly/50npuu>).
- `geo`: The place ID.
- `coordinates`: The printable names of this place.
- `place`: The type of this place - can be a "neighborhood" or "city".
- `country`: The country this place is in.
- `bounding_box`: The bounding box for this place.
- `source`: The application used to post this tweet.

Geo-tag info

- `geo`: Whether this user has geo enabled (<http://bit.ly/4gY77>).
- `coordinates`: The place ID.
- `place`: The place ID.
- `name`: The printable names of this place.
- `type`: The type of this place - can be a "neighborhood" or "city".
- `country`: The country this place is in.
- `bounding_box`: The bounding box for this place.
- `source`: The application used to post this tweet.

Map of a Twitter Status Object
 Raffi Krikorian <raffi@twitter.com>
 18 April 2010



Approach to crowd-sourcing

- *Not* require participants to explicitly “sign up” to contribute.
- To effectively crowd-source, a large source of crowd is needed.
- Twitter is such a source.



March 2017 winter storm

Number of tweets collected

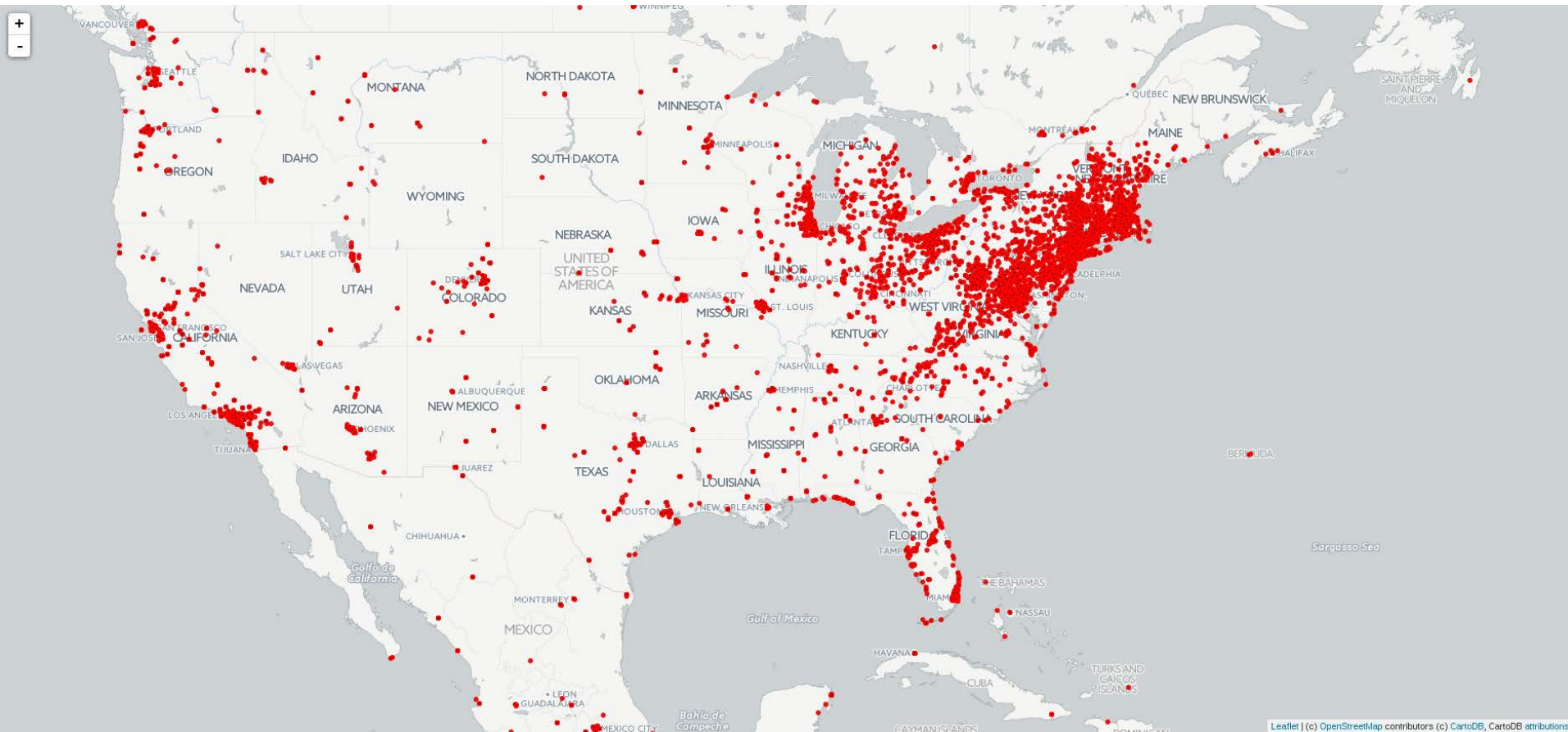
Start: 2017-03-13 22:23:12 – End: 2017-03-14 19:05:31

	Total # tweets	# tweets w/ geo-location	# tweets w/ geo-tag (place)
Global	1,227,390	22,880	34,535
U.S.		13,269	20,349



March 2017 snow event

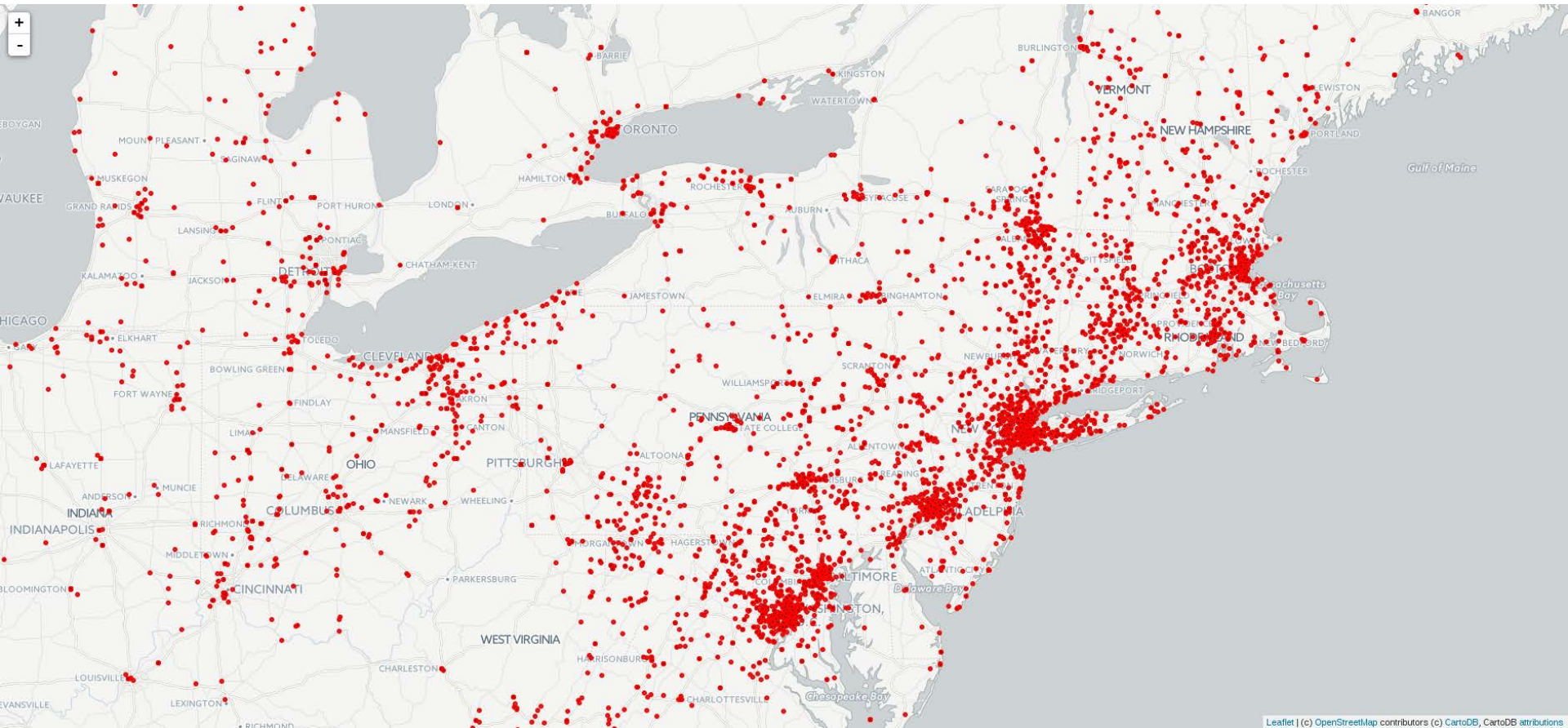
Distribution of tweets





March 2017 snow event

Distribution of tweets

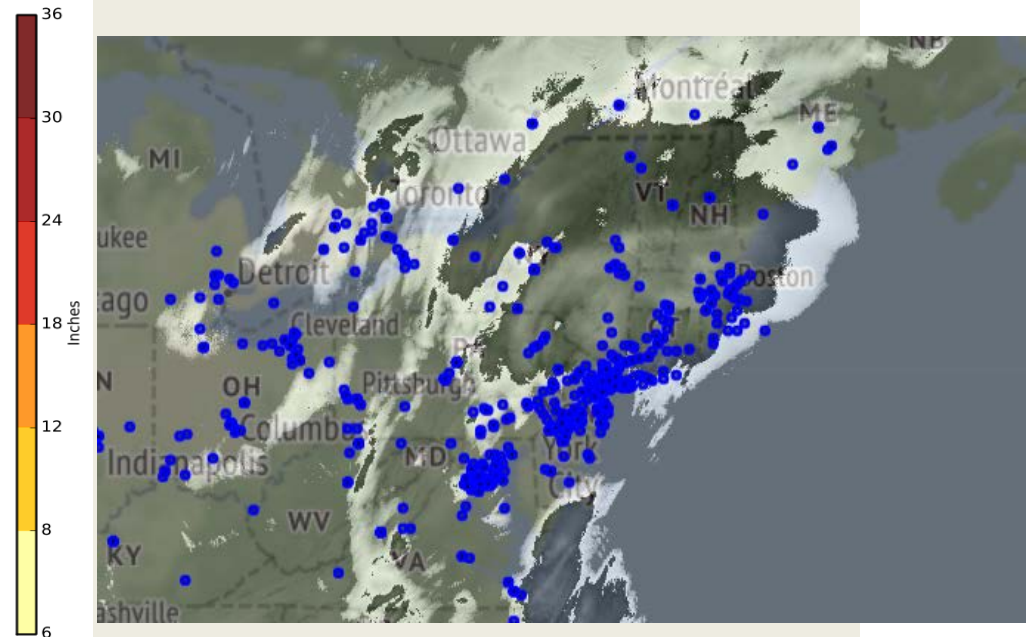
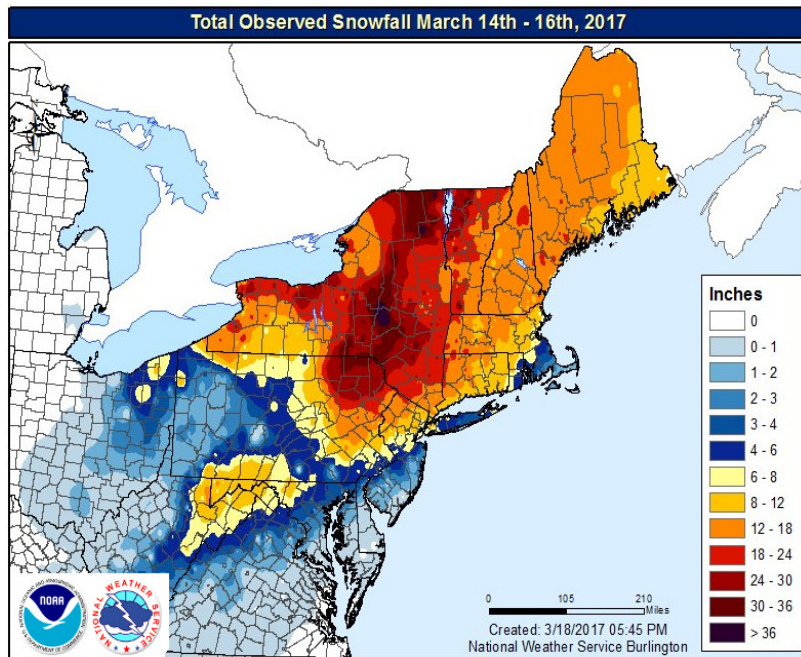


Leaflet | (c) OpenStreetMap contributors (c) CartoDB, CartoDB attributions



March 2017 winter storm

Observed snowfall & MRMS*-tweet map



*Multi-Radar/Multi-Sensor System (NOAA NSSL)



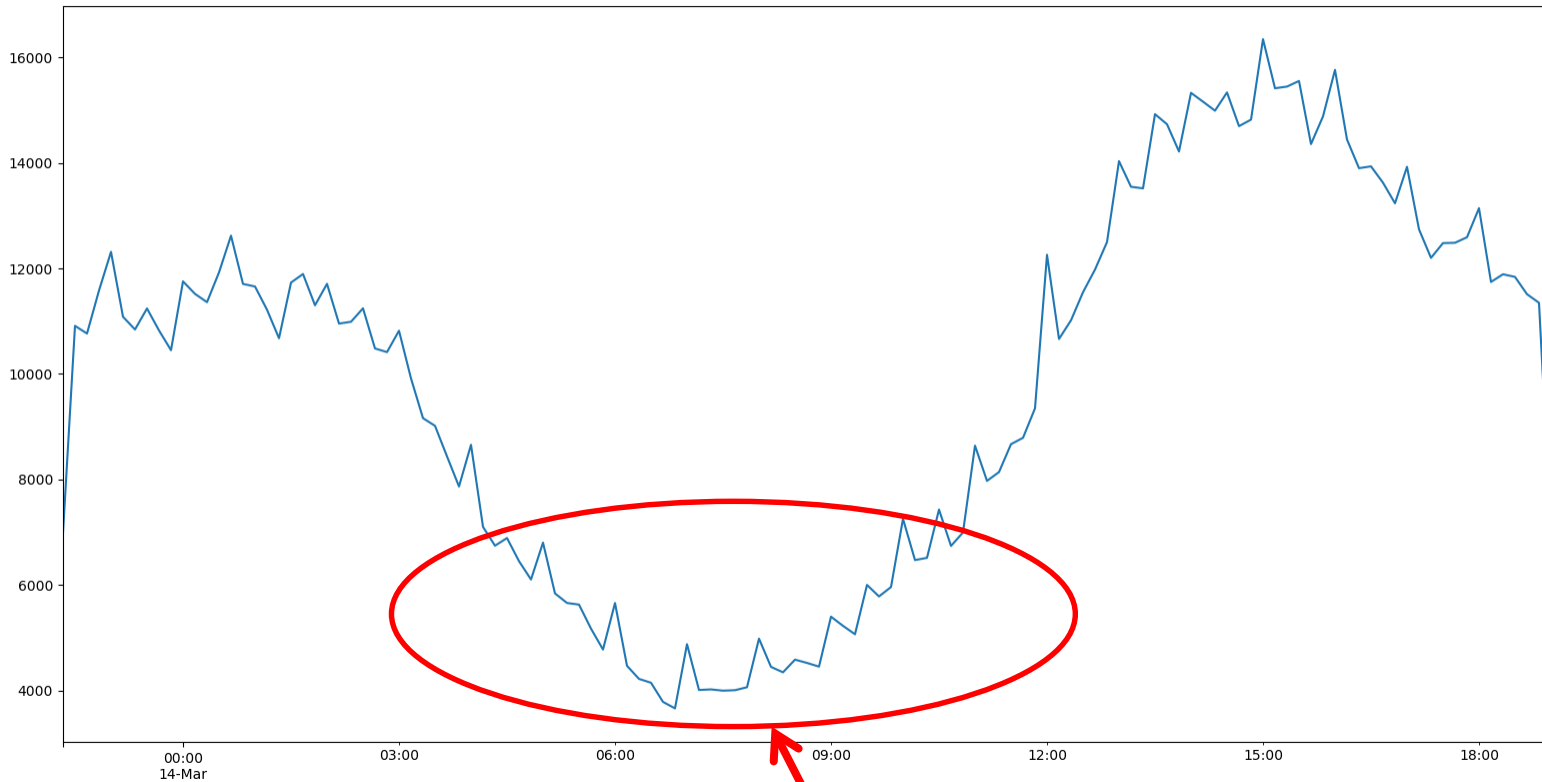
Comparing tweets with MRMS

- Analysis ongoing; thus far, majority of MRMS-tweet pairs agree (i.e., m1/t1, m0/t0).
- Many “passive” tweets are “regular” (e.g., from amateur weather stations), e.g.,
 - Wind 0.0 km/h N. Barometer 1006.7 mb, Falling. Temperature 8.3 °C. Rain today 8.8mm. Humidity 93%
- Implication for quality of “passive” tweets.

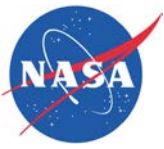


March 2017 snow event

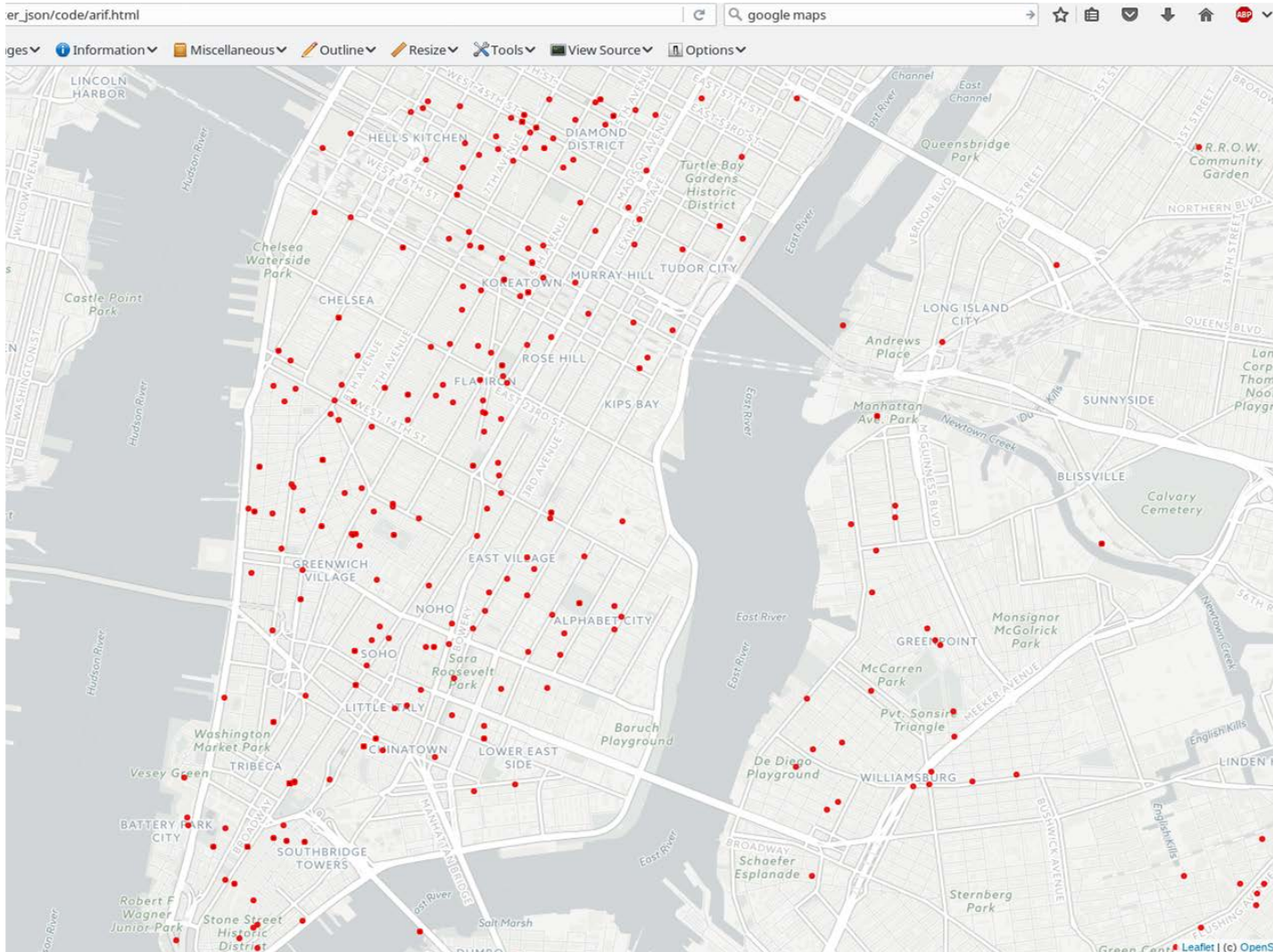
Time-varying # of tweets



~Midnight – 6 am, U.S. east coast



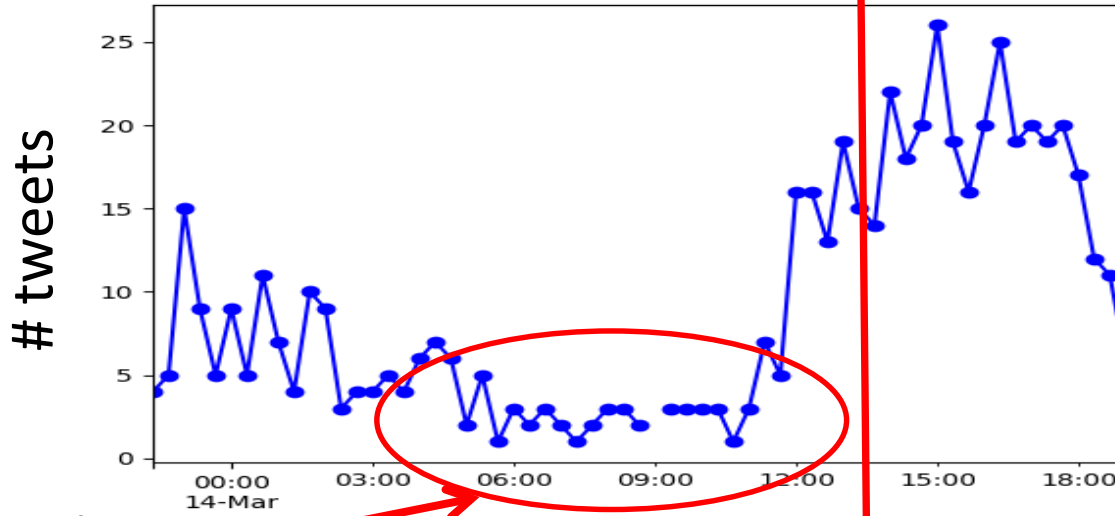
March 2017 snow event Manhattan



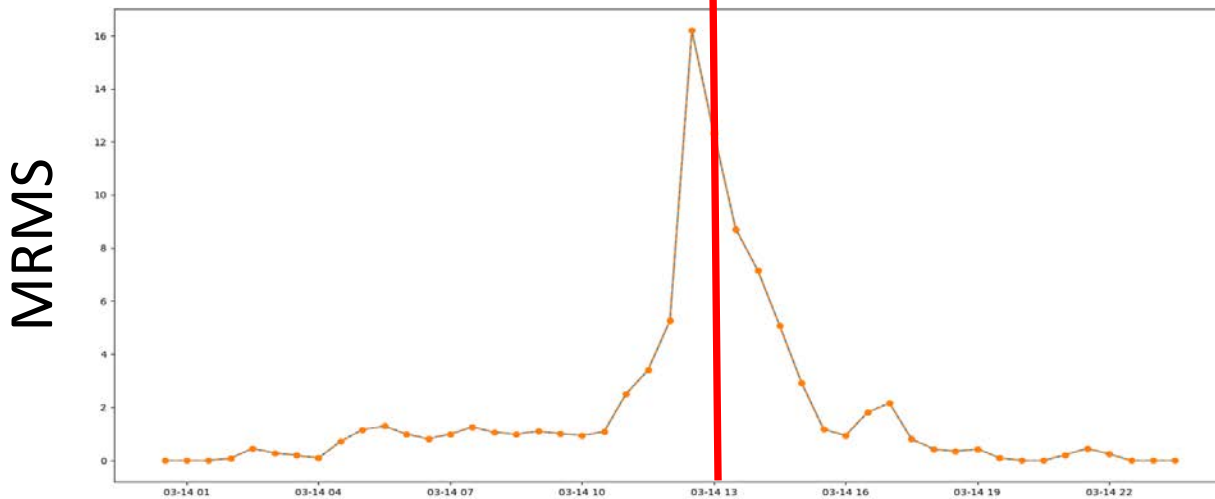


March 2017 snow event

Manhattan



~Midnight - 6 am

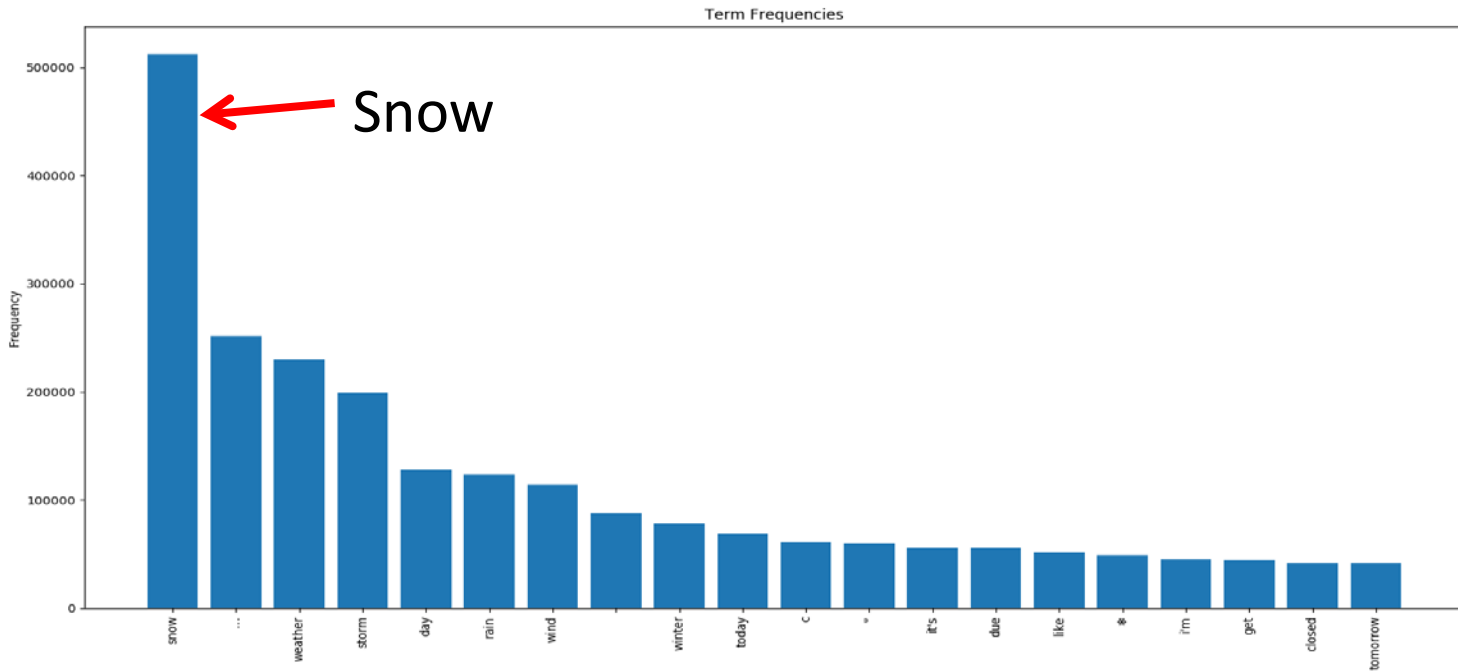




March 2017 snow event

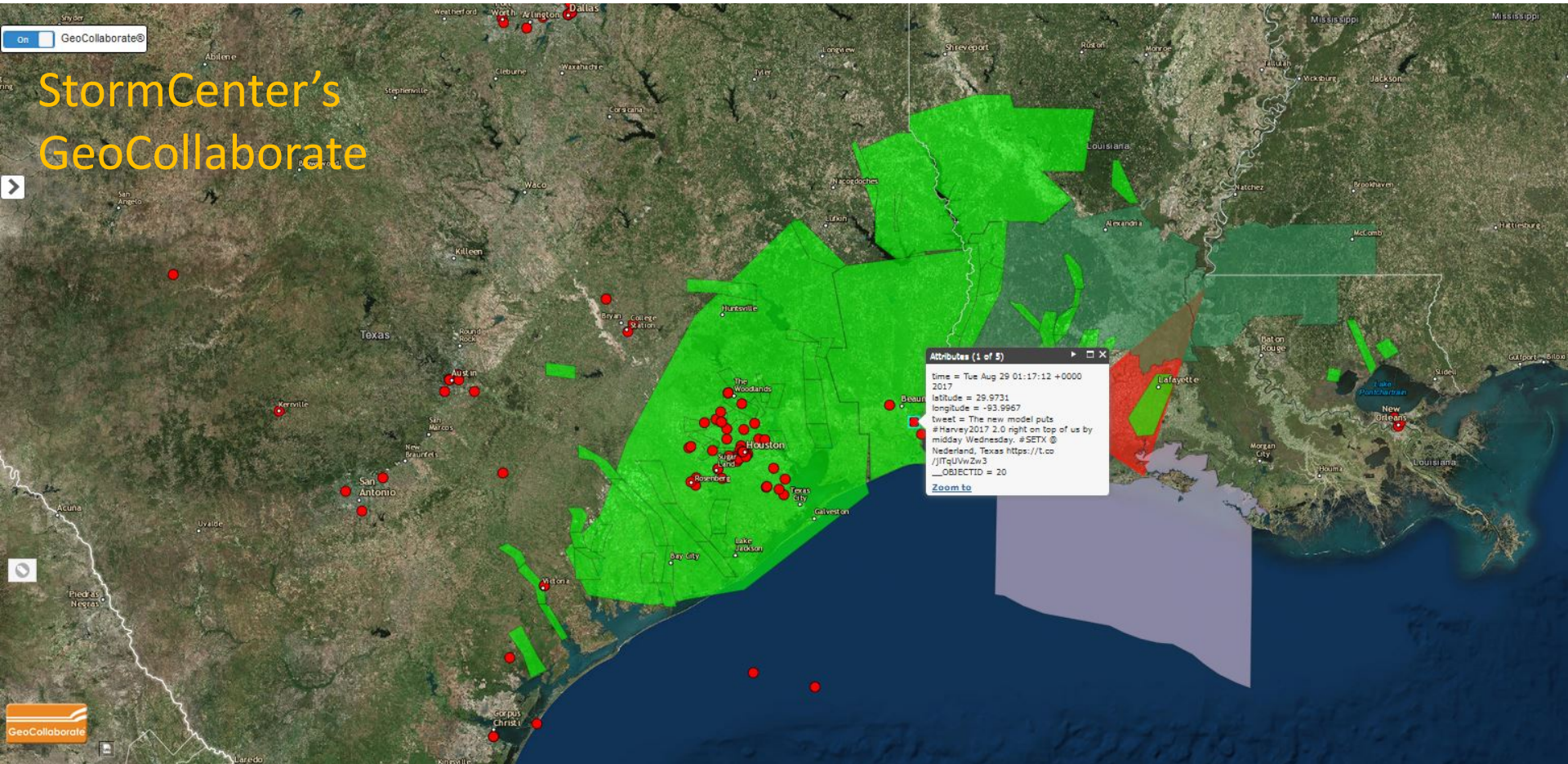
Frequency of words in tweets

	Snow	Weather	Storm	Rain	Wind	❄️	☃️
Global	512,414	229,836	198,798	123,773	113,915	48,789	
U.S. point location only	14,094	4,362	1,850	4,644	9,778	1,686	250



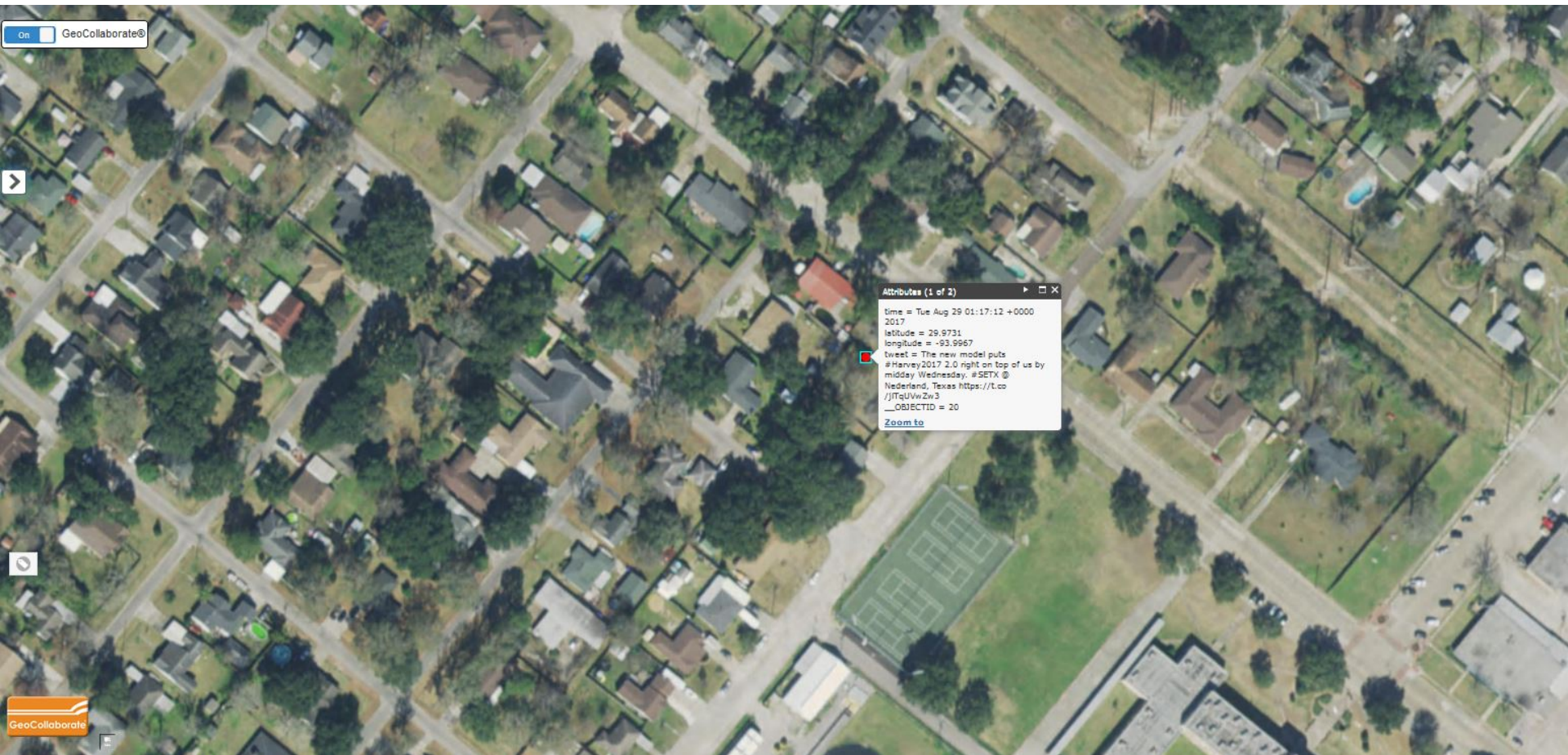


August 2017 Hurricane Harvey Tweets in GeoCollaborate





August 2017 Hurricane Harvey Tweets in GeoCollaborate





Summary

- Infrastructure is generic, i.e., not specific to a given measurement, social medium, or satellite mission.
- Twitter data have potential for earth science applications.



Questions?



Extras



Engagement w/ Active Participants

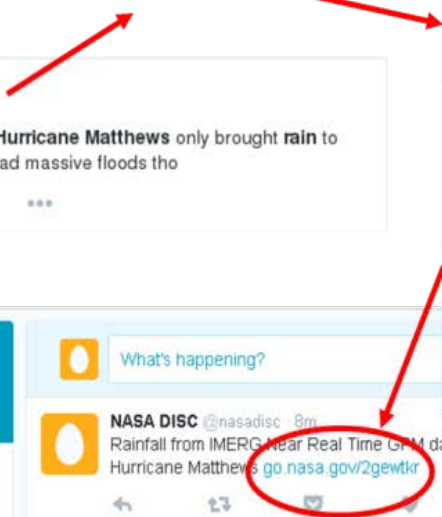
Filter and extract “rain” tweet (Oct. 7, 2016)
about Hurricane Matthew.

Active approach to citizen science

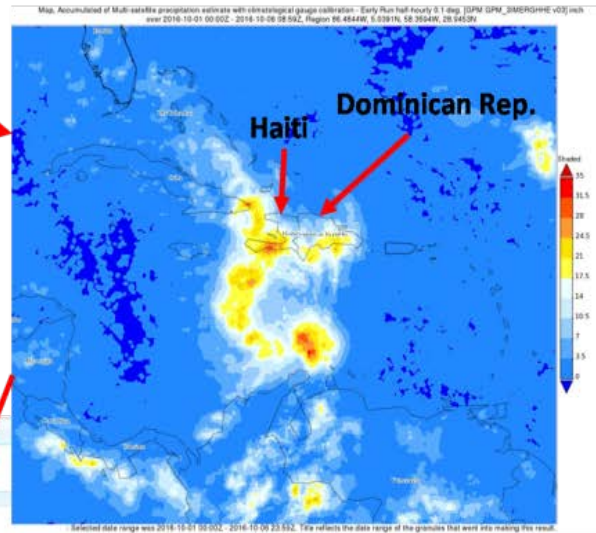
“Rain” tweet



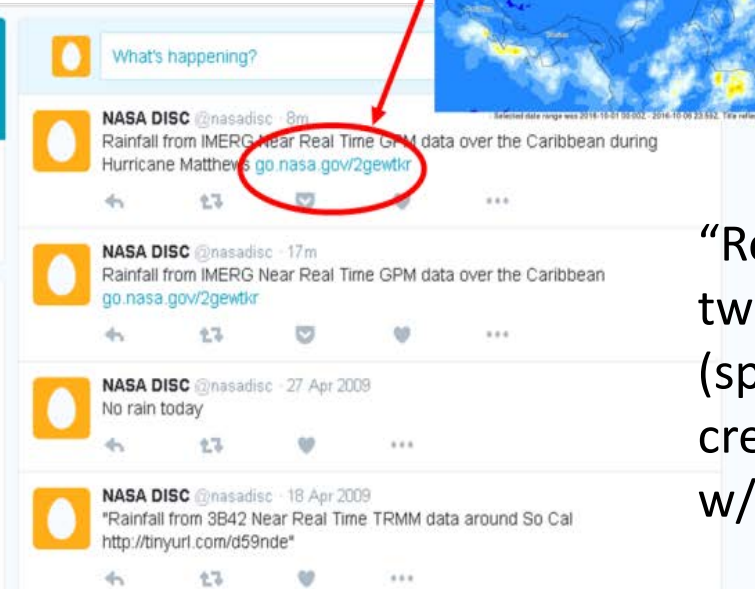
Geolocation



Giovanni rain map



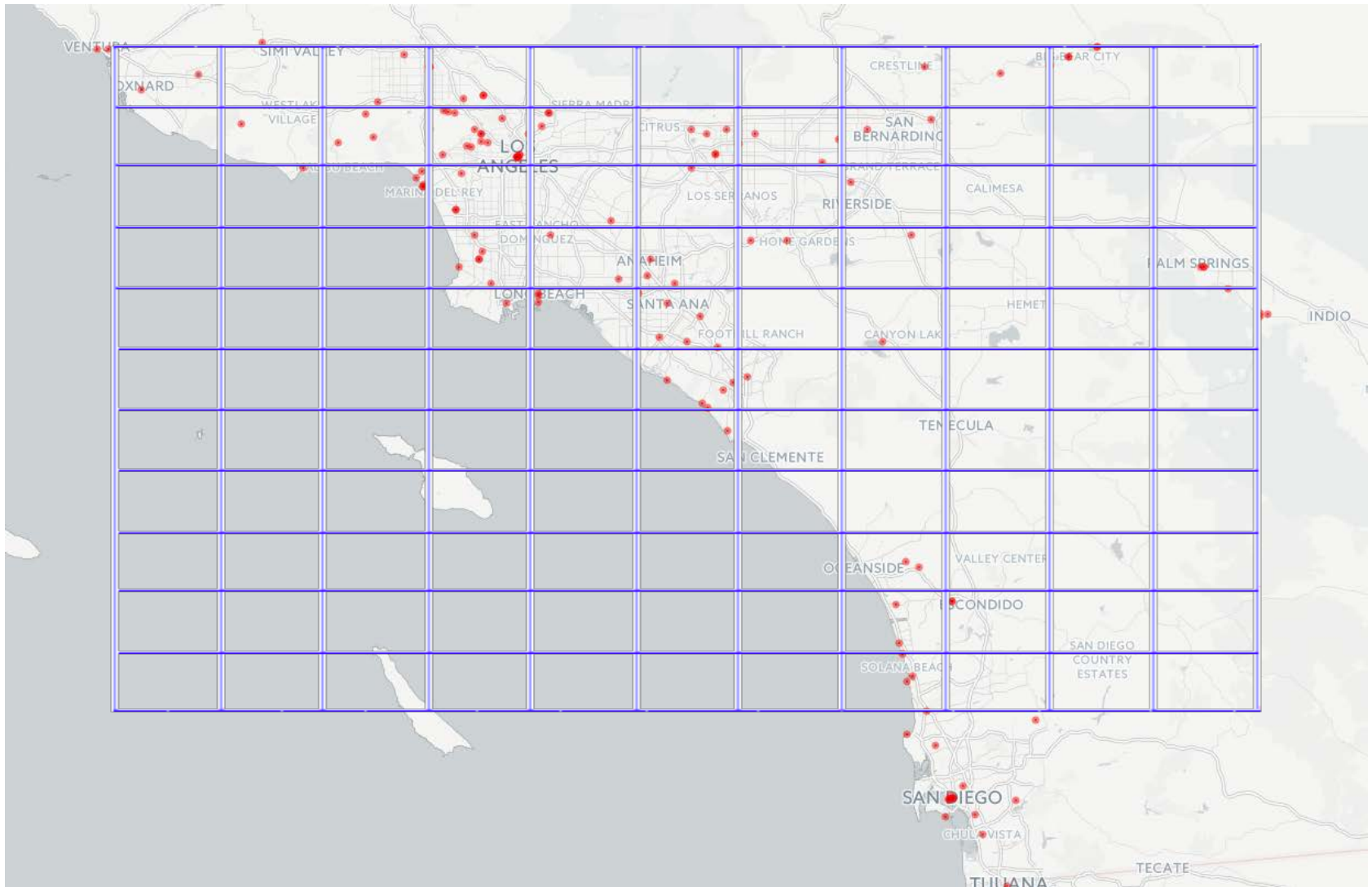
Generate rain map from
GPM, using
NASA Giovanni.



“Reply” to “rain” tweet by
tweeting to @NASADISC
(special Twitter account
created for the experiment)
w/ link to rain map.



Managing tweet data





August 2017 Hurricane Harvey

Number of tweets collected

Start: 2017-08-29 00:33:21 – End: 2017-08-29 09:36:02

	Total # tweets	# tweets w/ geo-location	# tweets w/ geo-tag (place)
Global	1,571,234	6,497	84,103
U.S.		4,010	78,906



August 2017 Hurricane Harvey

Distribution of tweets

