# Data Science @ Gnip

#AnalyzeBoulder Meetup

Josh Montague
Data Scientist, Gnip Inc.
@jrmontag

June 4, 2013

### Gnip

Where → Here! (Boulder, CO)

Who  $\rightarrow$   $\approx$  60 people\*

What  $\rightarrow$  social data provider

\*and hiring! http://gnip.com/careers

### sources & volumes

Publisher <sup>†</sup>	Daily Activities
Twitter	500 M
Tumblr	90 M
Foursquare	5 M
Disqus	1.7 M
Wordpress <sup>a</sup> Comments	1.7 M
Wordpress <sup>a</sup> Posts	1 M
'Engagement'b	2.4 M

a.com + .org

### 4 B activities delivered daily

<sup>&</sup>lt;sup>b</sup> likes, votes (Disqus, WP, Tumblr, IntenseDebate)

<sup>&</sup>lt;sup>†</sup> Add'l Premium Publishers: StockTwits, Newsgator; EDC Streams: Facebook, YouTube, Google+, flickr, bitly, Dailymotion, Delicious, Identi.ca, Instagram. Metacafe, Panoramio, Photobucket, Plurk, Reddit, StackOverflow, Vimeo

#### our team

3.5 data scientists (academic backgrounds)

'marketing'  $\rightarrow$  3x roles: marketing + sales + product

ask Qs, answer Qs (about our data)

## social data (v. media)

```
"body": "Preparation and Mindset is everything",
"retweetCount": 0.
   "link": "http://twitter.com".
   "displayName": "web"
                                                                                                                                                  > Follow
                                                                                                         Areeb Soomro
                                                                                                        @AreebSoomro
   "klout score": 38.
   "language": {
                                                                                                 Preparation and Mindset is everything
      "value": "en"
                                                                                                 Reply 13 Retweet * Favorite ** More
"object": {
   "postedTime": "2013-03-05T22:12:11.000Z".
                                                                                                3:12 PM - 5 Mar 13
   "summary": "Preparation and Mindset is everything",
   "link": "http://twitter.com/AreebSoomro/statuses/309063818313617408".
   "id": "object:search.twitter.com,2005:309063818313617408",
  "objectType": "note"
                                                                                                 Reply to @AreebSoomro
"actor": {
   "preferredUsername": "AreebSoomro".
   "displayName": "Areeb Soomro",
   "links": [
         "href": null.
         "rel": "me"
   "twitterTimeZone": "Mountain Time (US & Canada)".
   "image": "http://a0.twimg.com/profile_images/3342894166/cc21a568cf6a186e3b2341155c5d34d2_normal.jpeg",
   "verified": false.
   "location": {
      "displayName": "Cypress, Texas",
      "objectType": "place"
   "statusesCount": 22628,
   "summary": "ill hug ur eye lids, rubberbands last much longer when refrigerated, make sure to save the jackalopes because they are becoming extinct, i make no sense #Hashtag".
   "languages": [
      "en"
   "utcOffset": "-25200".
   "link": "http://www.twitter.com/AreebSoomro",
   "followersCount": 379.
   "friendsCount": 262,
   "listedCount": 1.
   "postedTime": "2011-06-24T12:02:24.000Z",
   "id": "id:twitter.com:323185180",
   "objectType": "person"
"twitter_entities": {
   "user_mentions": [],
   "hashtags": [].
"verb": "post",
"link": "http://twitter.com/AreebSoomro/statuses/309063818313617408".
   "link": "http://www.twitter.com",
   "displayName": "Twitter",
   "objectType": "service"
"postedTime": "2013-03-05T22:12:11.000Z"
"id": "tag:search.twitter.com, 2005:309063818313617408",
```



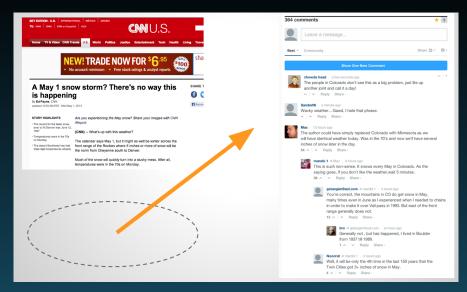
"objectType": "activity"

20

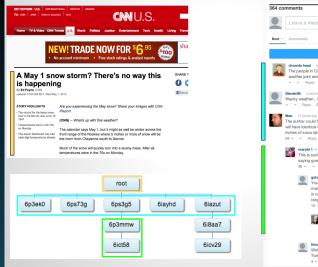
48

49

## ex: Disqus comment threads

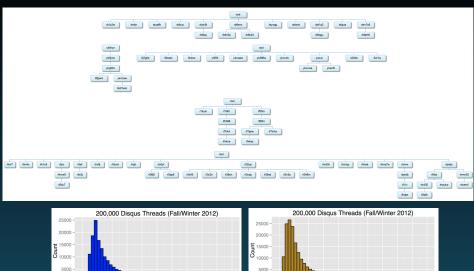


## ex: Disqus comment threads





## ex: Disqus comment threads



20 25 30 Replies To Article

Unique Comments

## thanks!

source/slides: http://bit.ly/ABMeetup

Don't forget, we're hiring! http://gnip.com/careers