

# Social Media for Giant Instant Opinion Polls: Twitter Political Index

Rishab Aiyer Ghosh  
Co-Founder & Chief Scientist

**TOPSY** LABS<sup>®</sup>



# THE TWITTER POLITICAL INDEX

OCTOBER 29, 2012: ROMNEY +5 [LEARN MORE](#)

[Tweet](#)

[Follow @gov](#)



BARACK OBAMA

34 [↑ +1](#) SINCE YESTERDAY

[Follow @BarackObama](#) 10M+ followers



MITT ROMNEY

39 [↑ +2](#) SINCE YESTERDAY

[Follow @MittRomney](#) 1.6M followers

[election.twitter.com](http://election.twitter.com)

II. DASHBOARD

#Twindex History ▶

TOPIC AREAS

- Abortion
- The Economy
- Education
- Energy and the Environment
- Foreign Policy
- Gay Rights
- Healthcare
- Immigration
- Retirement
- Taxes
- Terrorism

HOT TOPICS

- Benghazi
- Horses and Bayonets
- Pay Equality
- PBS Funding

CANDIDATES

- Obama/Biden
- Romney/Ryan

# #TWINDEX

## HISTORY



ANALYTICS BY **TOPSY**

## NEWS

- Twitter Government** @gov 1 day  
Our growing list of gov't, political, public safety and weather accounts providing Hurricane #Sandy info: [twitter.com/gov/sandy](http://twitter.com/gov/sandy)  
View on Twitter
- Twitter Government** @gov 25 Oct  
Tweeting about the U.S. presidential election? Use hashtag #election2012 and join the global conversation! [pic.twitter.com/7yFWBq61](http://pic.twitter.com/7yFWBq61)  
View on Twitter

# election.twitter.com

all ~200 million English-language tweets

every day

minute-by-minute sentiment

every term in every tweet

scaled score for Obama and Romney

aggregates and stores this

forever.

{ *shift* }

underlying *shift*

in *information dissemination*

*{ previously }*

Previously public sources, were *published* sources: news in *publications*

*{ previously }*

Previously public sources, were *published* sources: news in *publications*

Public conversations were *private information*

{  $\longleftarrow$  *shift* }

Previously public sources, were *published* sources: news in *publications*

Public conversations were *private information*

Public conversations are *public information*

Publicly accessible



*{ public conversations are publicly accessible }*

*we can all be*

*in ohio with voters*

*{ public conversations are publicly accessible }*

*we can all be*

*Watching #sandy*

🕒 Latest Results

Past 1 Hour **4,879**

**Past 1 Day 161K**

Past 7 Day **357K**

Past 30 Day **363K**

All Time **394K**

Search

- Everything
- Links
- Tweets
- ▶ **Photos**
- Videos
- Experts

Network

- Google Plus
- ▶ **Twitter**

Sort

- ▶ **Relevance**
- Date

Language

- ▶ **All Languages**
- English
- 中文
- 日本語
- 한국어
- Русский



nycshopguide  
2 hours ago **15**



sunherald  
13 minutes ago **49**



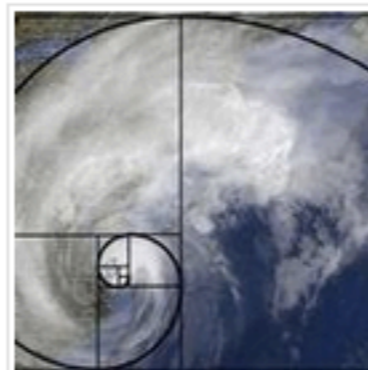
lohud  
31 minutes ago **41**



boingboing  
3 hours ago **2,765**



wsj  
11 minutes ago **26**



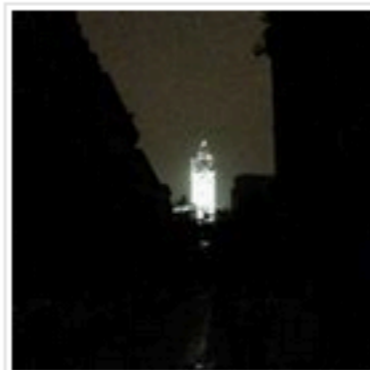
megan  
3 hours ago **42**



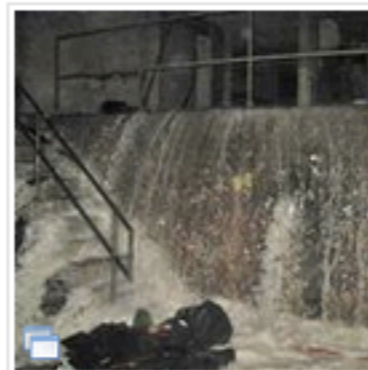
mpf2011  
2 minutes ago **10**



ojezap  
31 minutes ago **30**



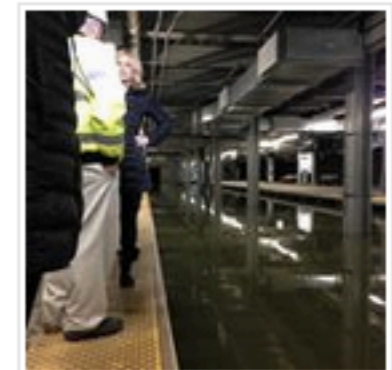
danabrunetti



storvfuloro



alvssahertzia



cecilvtvnan

[Topsy.com/s?q=sandy&type=image&window=d](https://Topsy.com/s?q=sandy&type=image&window=d)

*{ public conversations are publicly accessible }*

*no reporters to intermediate*

*lots of NOISE*

*{ public conversations are publicly accessible }*

*technology to extract signal*

*best signal requires*

*ALL THE DATA*

*{ public conversations are publicly accessible }*

*technology to extract signal*

*best signal requires*

*ALL THE DATA*

(Topsy indexes 200 billion tweets & 12 billion links  
from >100 million authors)

# Machine Sentiment :- (

- Computer-based sentiment analysis is mostly designed for well-written texts, such as newspaper articles
- Never the best quality, especially on 140 characters

# Machine Sentiment :- (

- Computer-based sentiment analysis is mostly designed for well-written texts, such as newspaper articles
- Never the best quality, especially on 140 characters
- Very little context in the Tweet itself





# Topsy Sentiment

Sentiment analysis technology developed in-house

- Human curated dictionaries with simple grammar
- Tested on several samples
- On 10k sample with 3 human judges per tweet, Topsy had:
  - 90% agreement with random human judgement
  - 73% agreement when all 3 human judges agree

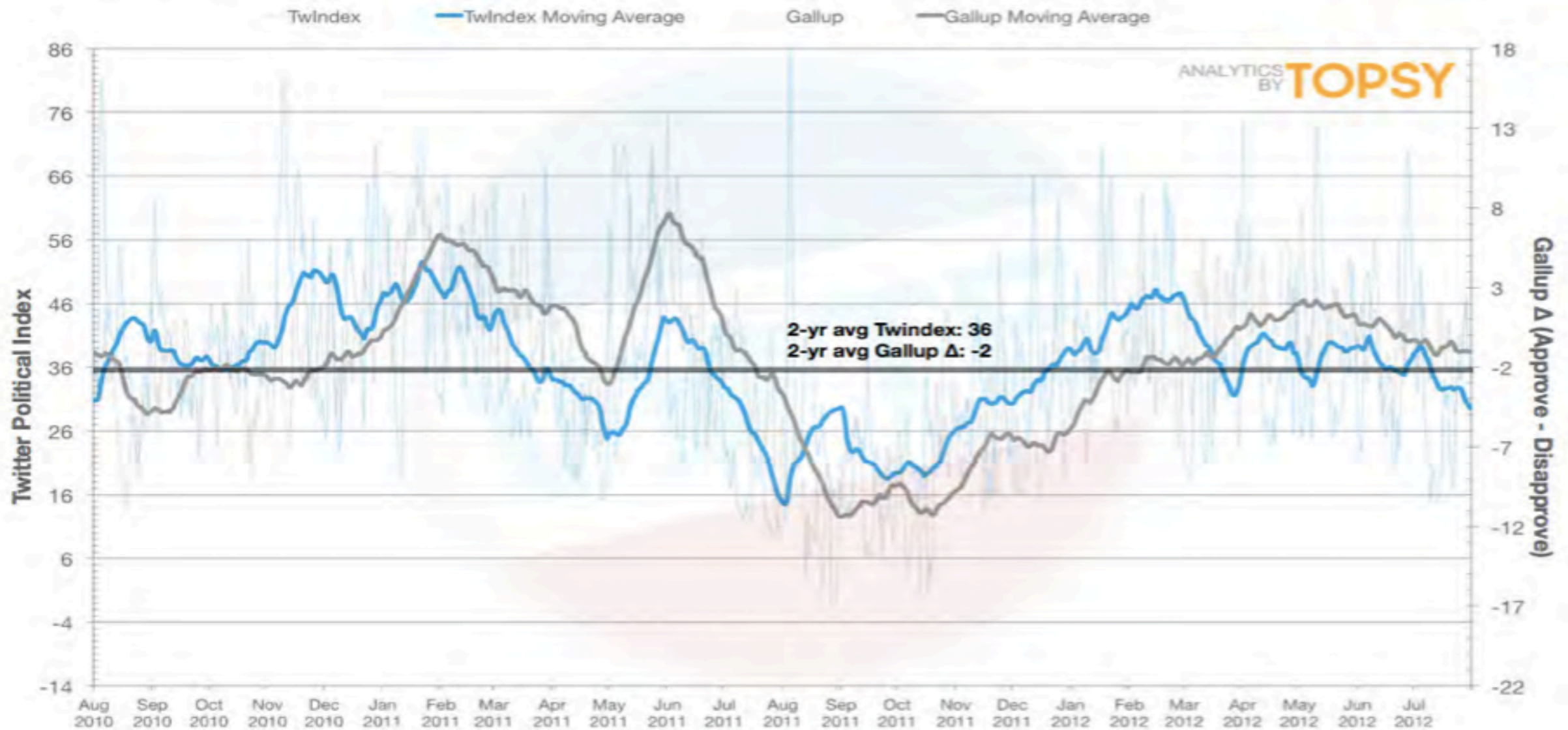
# Topsy Sentiment

Sentiment analysis technology developed in-house

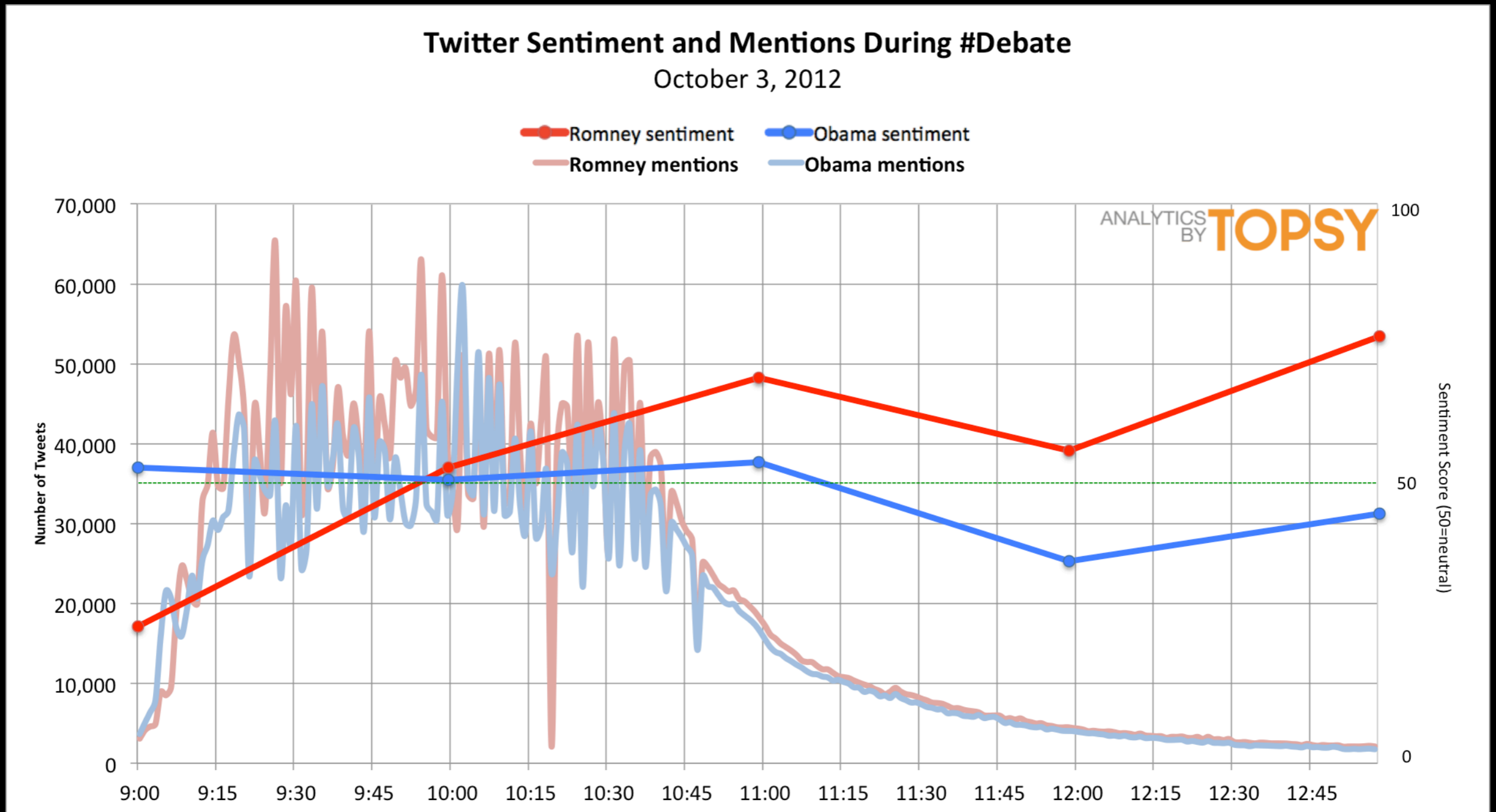
- Topsy applies scale to compensate for specificity:  
Sentiment over 1000s or 1000000s vs polls of 100s
- Topsy Sentiment for any term (e.g. “Obama”) normalized against sentiment for *all* terms, as sentiment is computed for *all* terms, phrases in *all English-language tweets (~200m)* to compute scaled percentile score, 0-100 (50 is neutral)
- Proof point: Twitter Political Index, correlation with Gallup approvals for President Obama for 2 years

# Topsy Sentiment for Obama: 2 years

Twitter Political Index: A Comparison to Gallup  
with 30-day moving averages — August 1, 2010 - July 31, 2012  
*Axis range adjusted to center respective data set averages.*

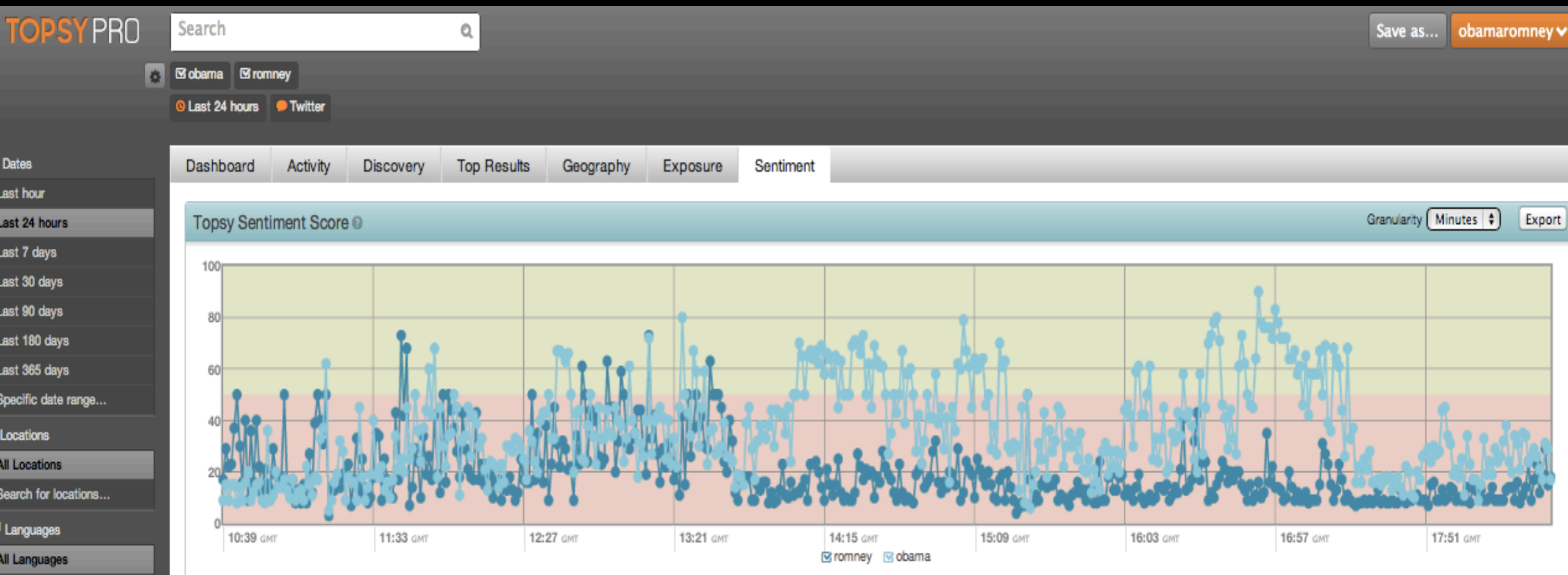


# Obama vs Romney, first debate



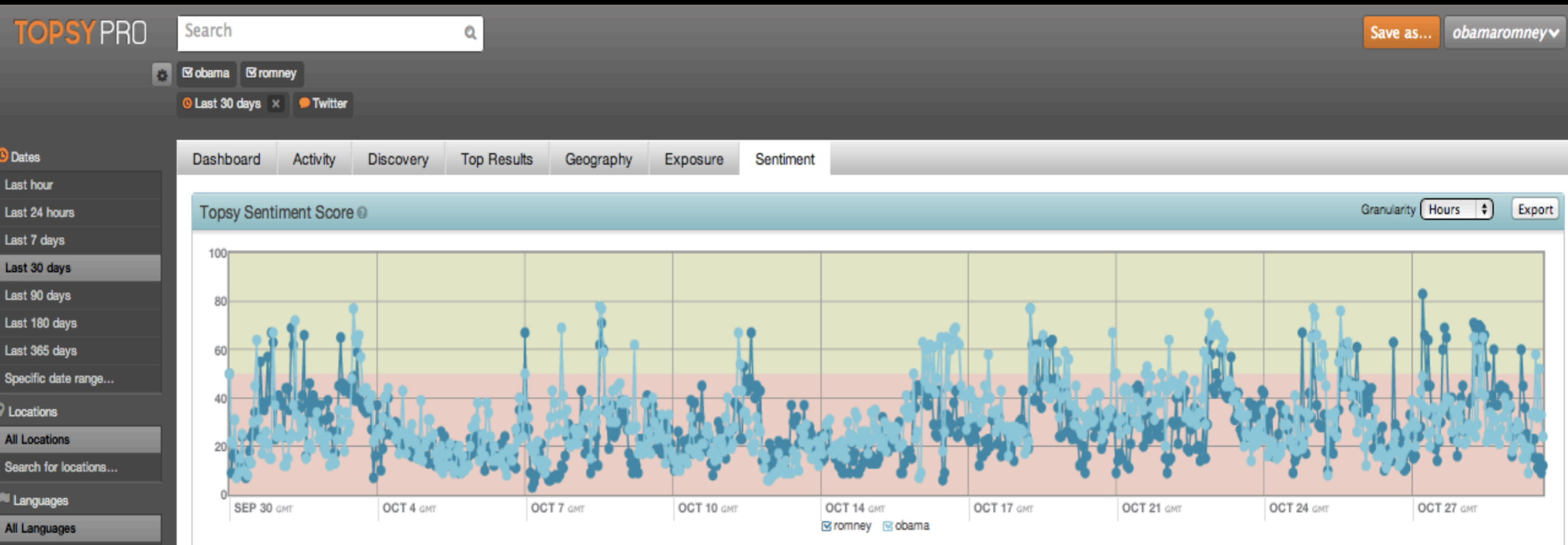
# Demo

# Demo



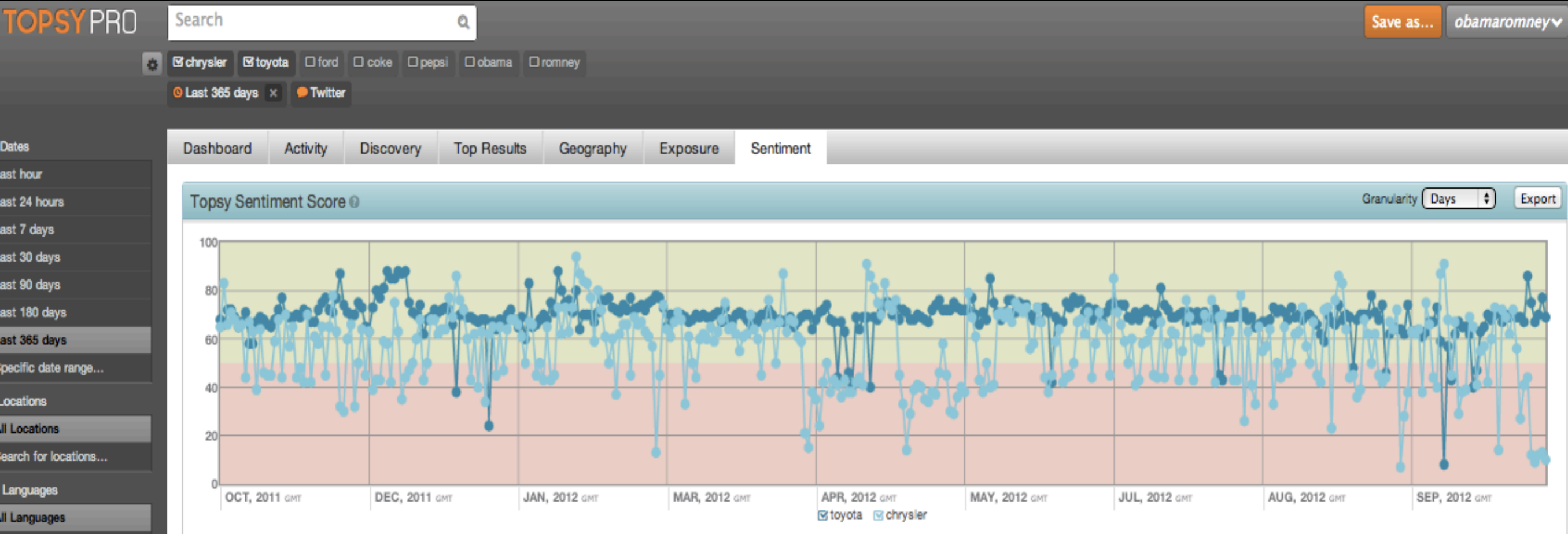
(sentiment per minute, over past several hours)

# Demo



(sentiment per hour, over past month)

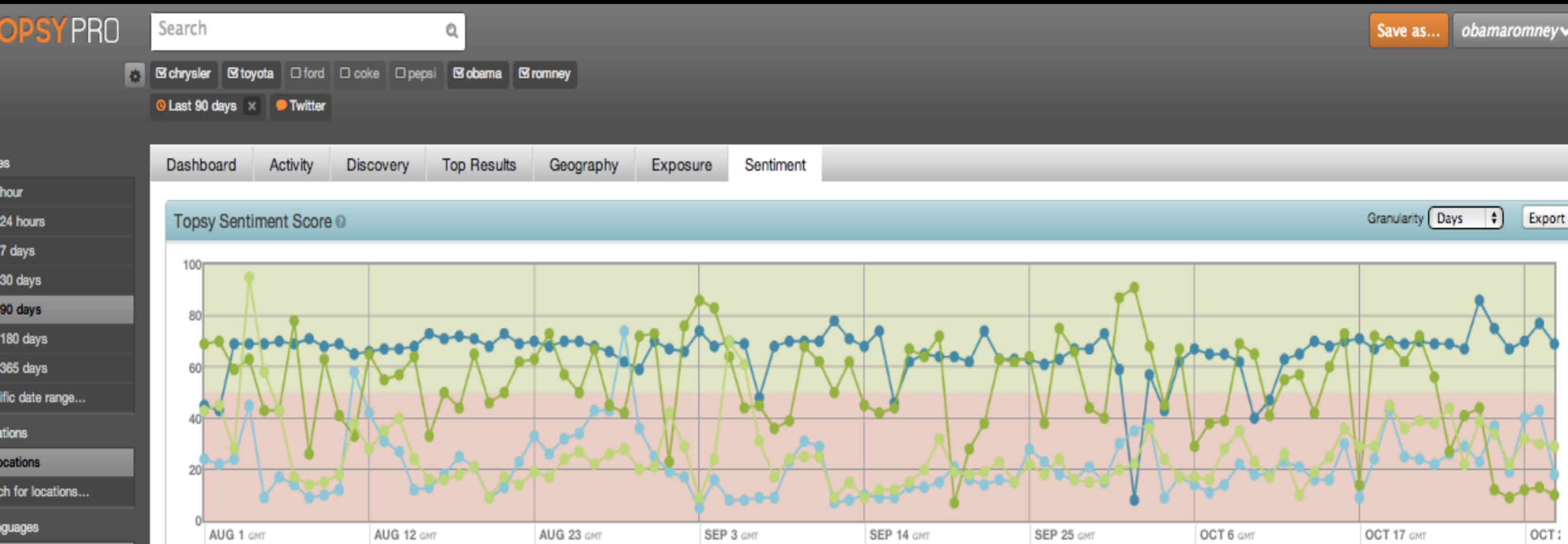
# Demo



(sentiment per day, over past year)



# Demo



(sentiment per day, past 90 days, obama, romney, toyota, chrysler)

**TOPSY** LABS<sup>®</sup>

Questions?

# TOPSY LABS<sup>®</sup>

## Contact

Rishab Ghosh  
Co-Founder & Chief Scientist  
[rishab@topsy.com](mailto:rishab@topsy.com)

Eddie Smith  
CRO  
[eddie@topsy.com](mailto:eddie@topsy.com)