Social Media for Giant Instant Opinion Polls:
Twitter Political Index

Rishab Aiyer Ghosh
Co-Founder & Chief Scientist

TOPSY LABS®

election.twitter.com

TOPSY LABS

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.

underlying *shift*

in *information dissemination*

TOPSY LABS

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

TOPSY LABS

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

Public conversations were *private information*

{ $\longmapsto$ *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

TOPSY LABS

{ *public* conversations are *publicly accessible* }

we can all *be*

Watching *#sandy*
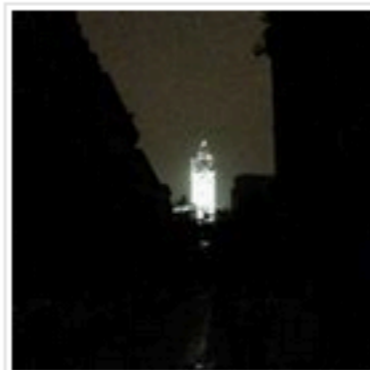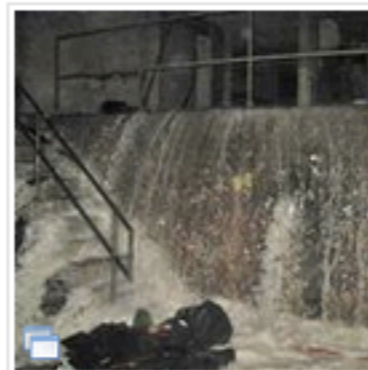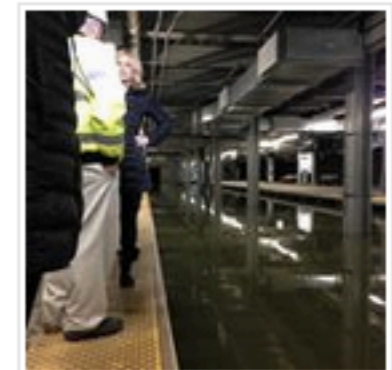
Topsy.com/s?q=sandy&type=image&window=d

{ *public* conversations are *publicly accessible* }

no reporters to intermediate

lots of NOISE

TOPSY LABS

{ 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)

TOPSYLABS

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
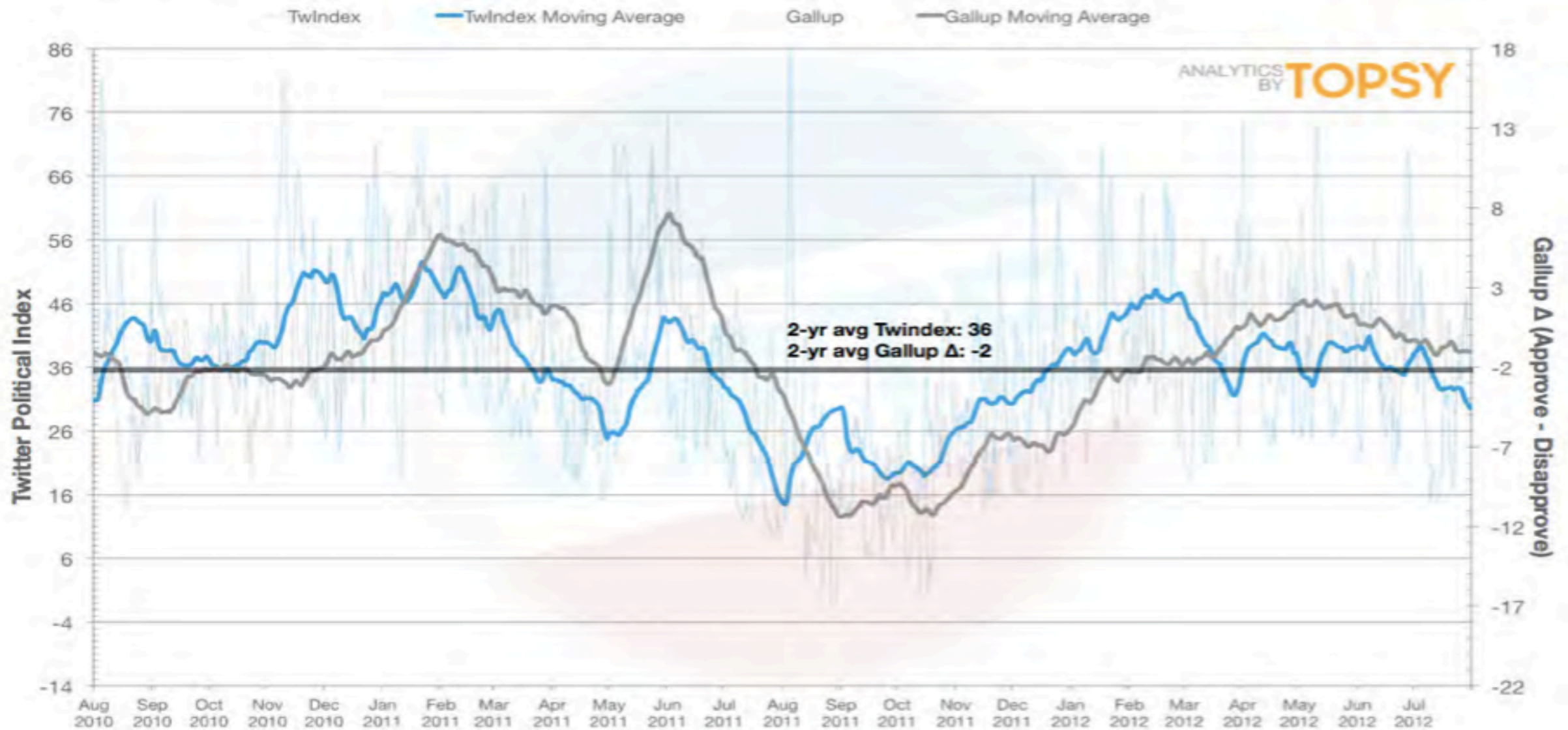 - 73% agreement when all 3 human judges agree

Topsy Sentiment

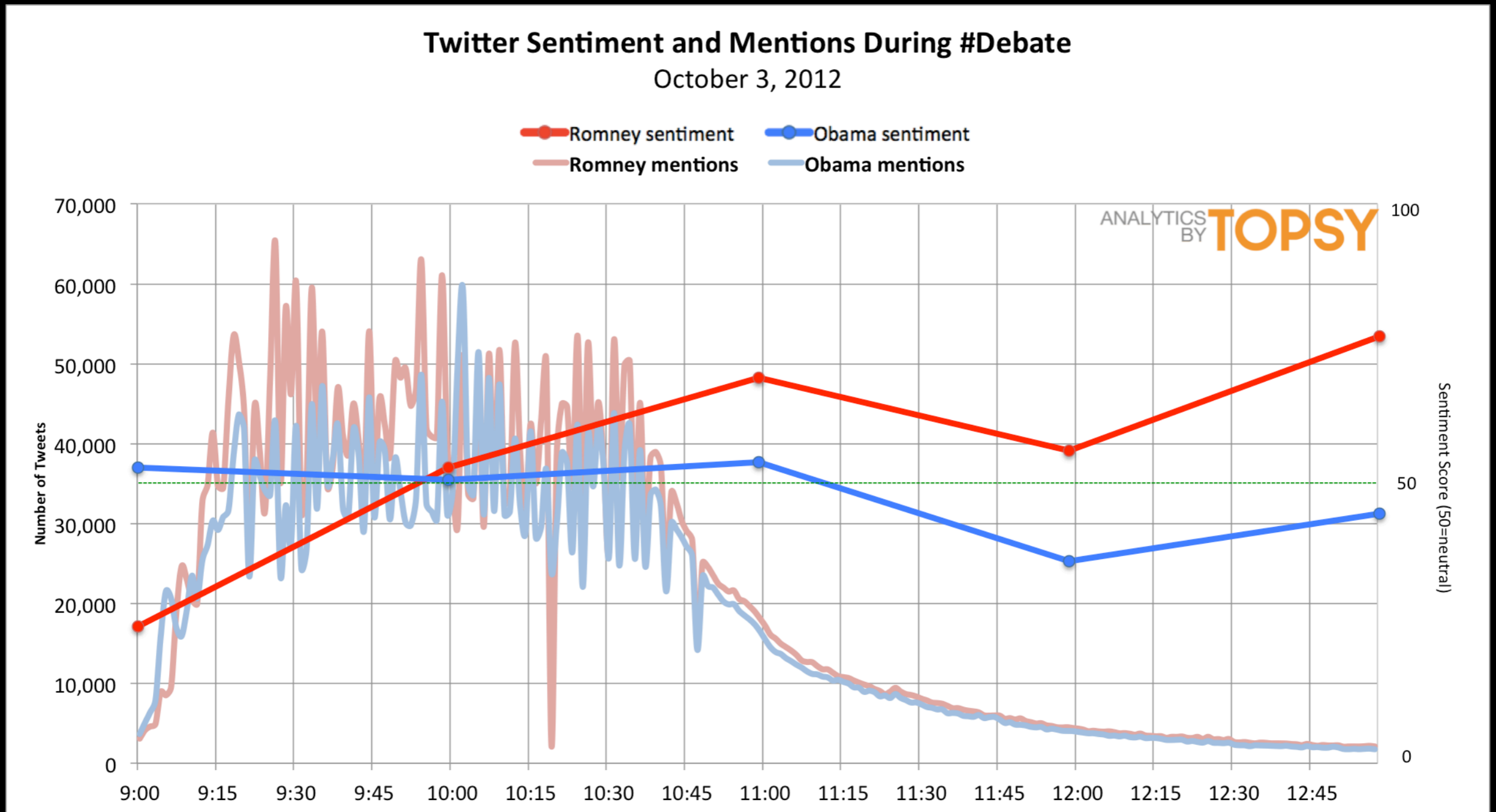Sentiment analysis technology developed in-house
- Topsy applies scale to compensate for specificity: Sentiment over 1000s or 1000000s vs polls of 100s
- Topsy Sentiment for any term (e.g. "Obama") normalized against sentiment for *all* terms, as sentiment is computed for *all* terms, phrases in *all 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
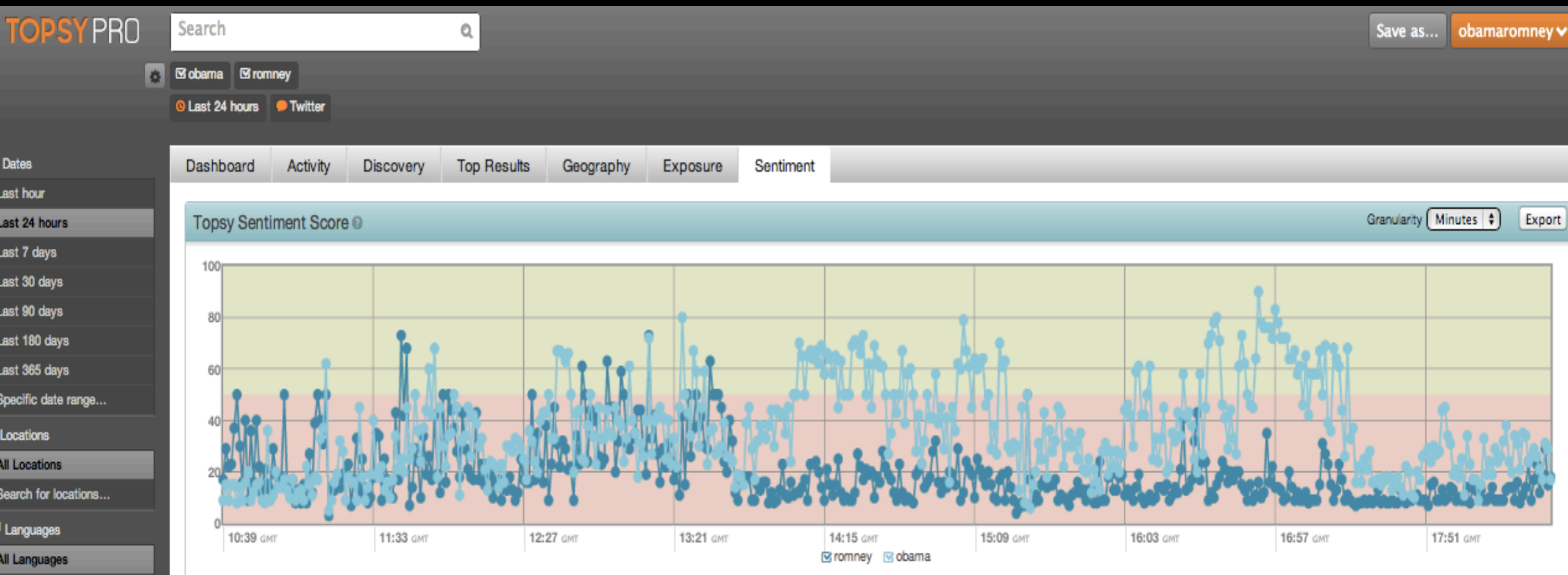Axis range adjusted to center respective data set averages.

2-yr avg TwIndex: 36
2-yr avg Gallup Δ: -2

Obama vs Romney, first debate

Twitter Sentiment and Mentions During #Debate
October 3, 2012

ANALYTICS BY **TOPSY**

Legend: Romney sentiment — Obama sentiment — Romney mentions — Obama mentions
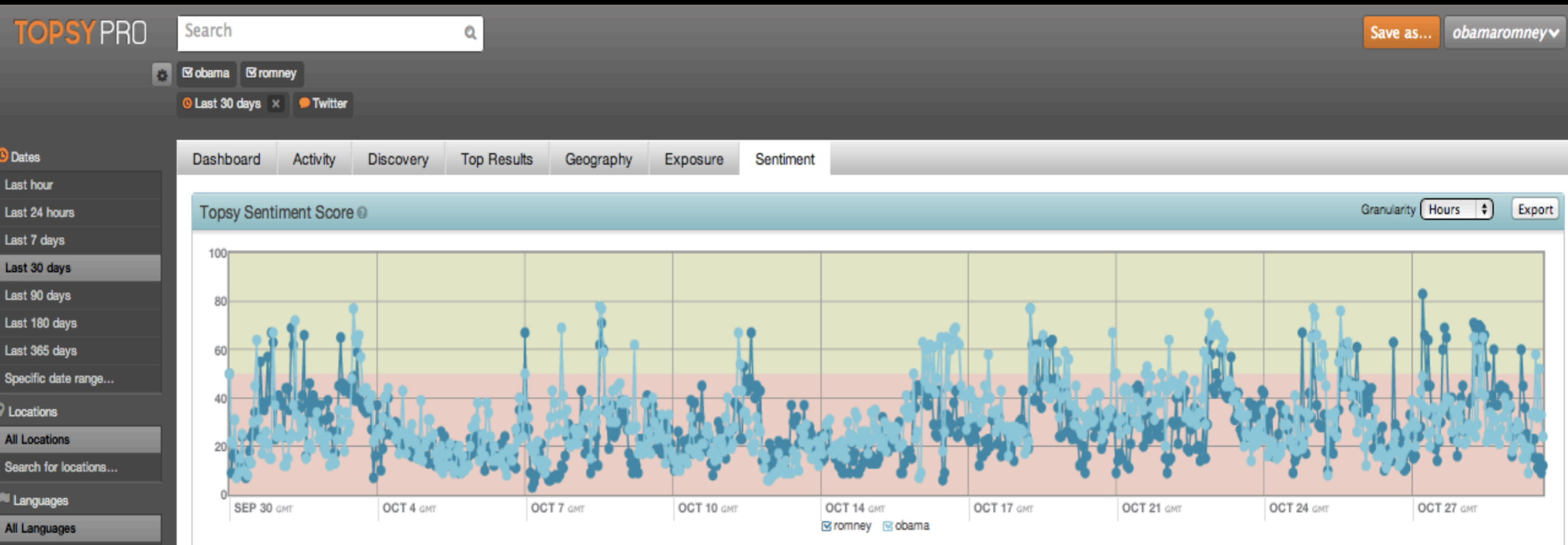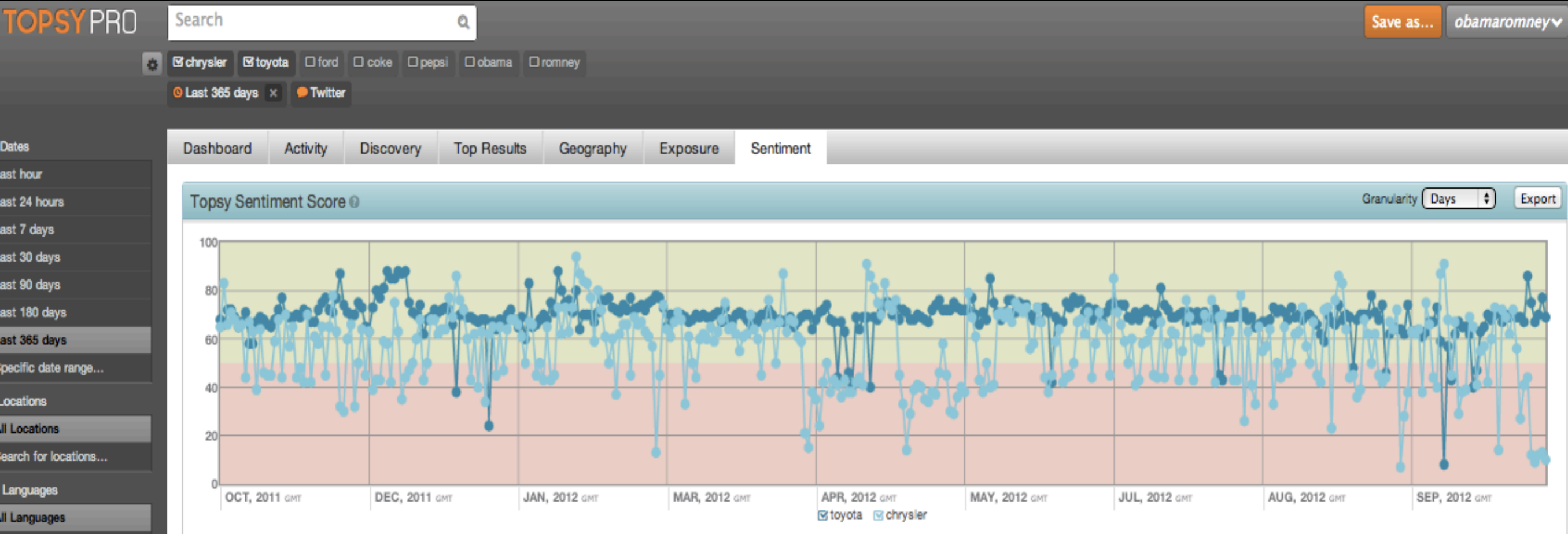
TOPSY LABS

Demo

Demo

(sentiment per minute, over past several hours)

Demo

(sentiment per hour, over past month)
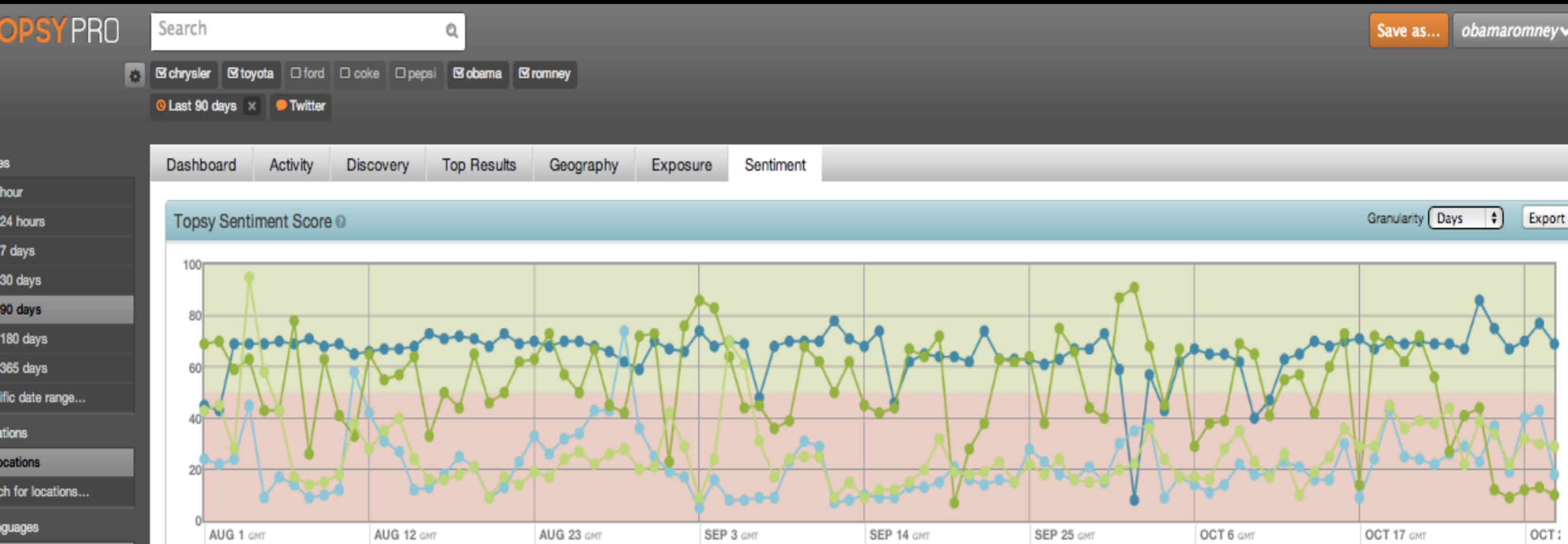
Demo

(sentiment per day, over past year)

Demo

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

TOPSY LABS

Questions?

TOPSY LABS®

Contact

Rishab Ghosh
Co-Founder & Chief Scientist
rishab@topsy.com

Eddie Smith
CRO
eddie@topsy.com