

Emotional detection in Twitter: forceful police against black victims

What's this project?

After news media reports an incident of police force against the black victim (the shooting of Trayvon Martin in 2012 and that of Michael Brown in 2013), Twitter erupted with emotional reactions to the events. Our team used rule-based sentiment analysis to discover the trends of people's inner voice on these events over time. Tracking the patterns of people's emotions on Twitter about the police force against unarmed blacks in the United States is of great importance.

Why this project?

Our sentiment analysis may help reveal people's intentions in advance to some extreme passive behaviors such as suicide to some degree.

Data Overview

- 1,003,492 tweets in 2012
- 942,569 tweets in 2013
- 222,204 users in discussion
- Top hot tweet was retweeted by 37807 times
- Top active twitter user posted tweets 2169 times

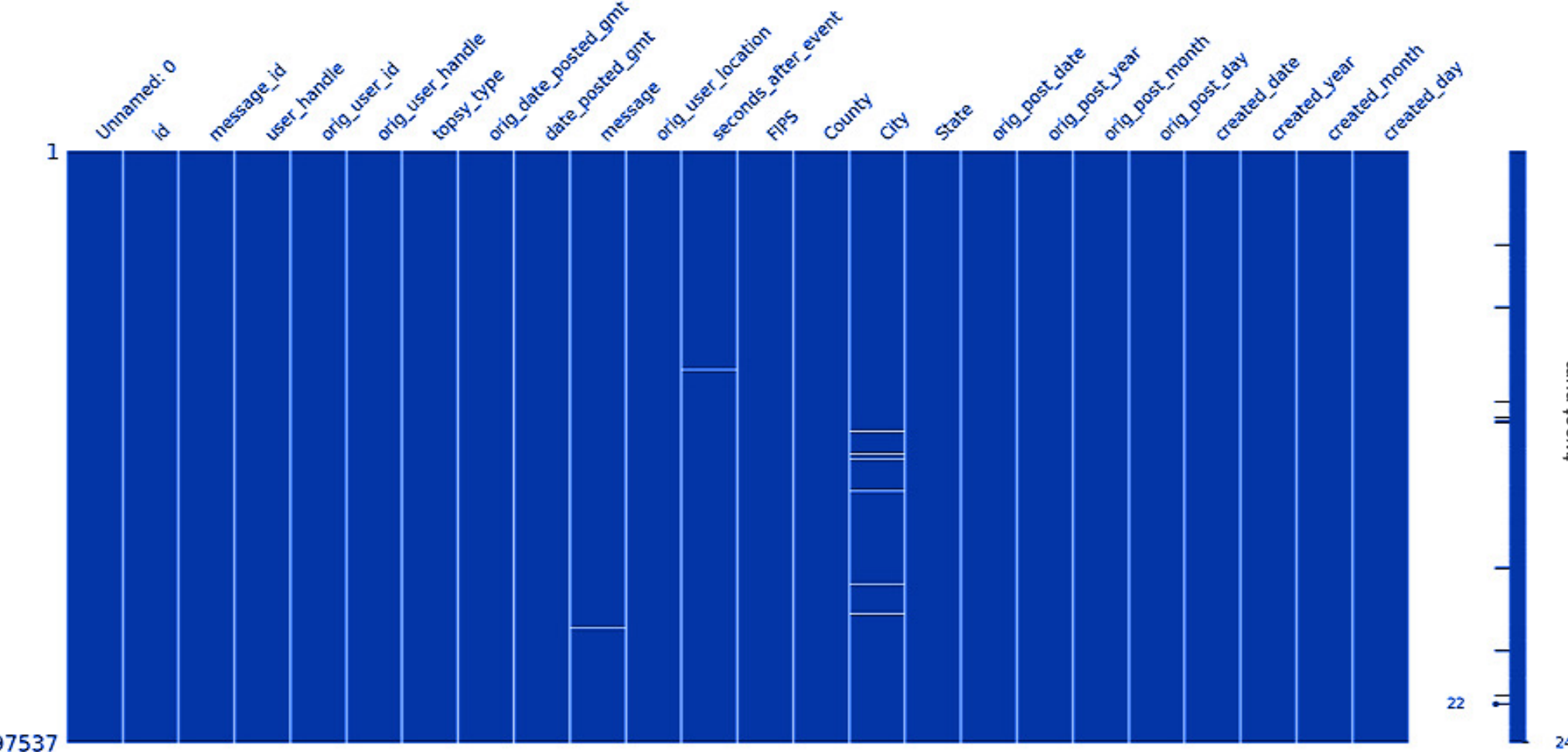


Figure 2 Missing pattern of cleaned twitter data



Figure 4 Word cloud on media data

Methods

- Detecting a subset of negative emotions like sadness, anger, anxiety to better understand how negative emotions change in tweets through time
- Correlation analysis
- Hypothesis testing on time to reach the tweets of half peak volume
- Spline analysis

% action type in Twitter

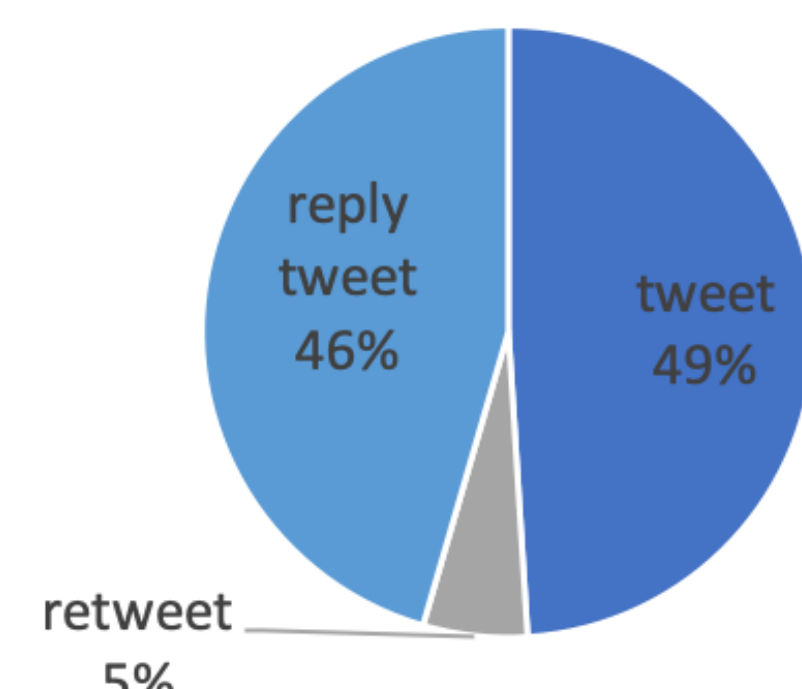


Figure 1 Action type percentage

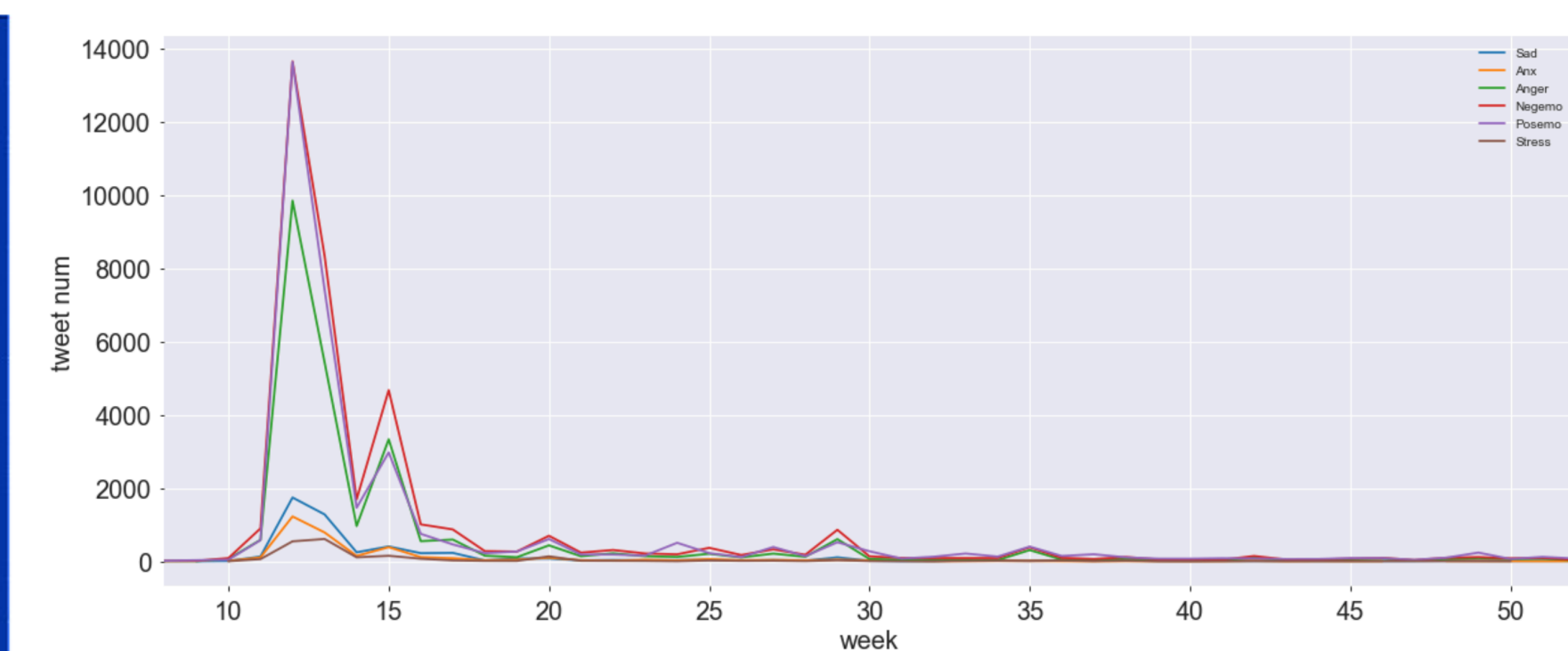


Figure 3 Tweet num vs week in 2012

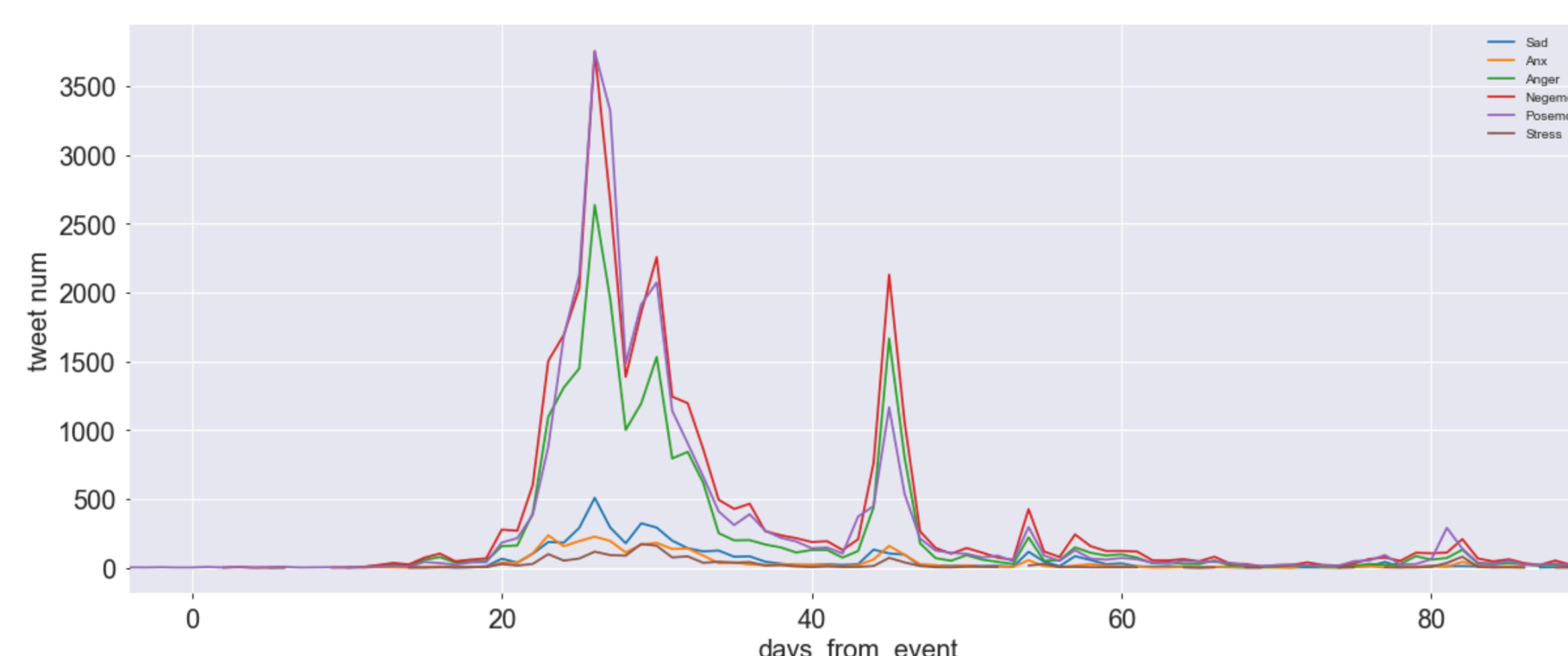


Figure 5 General tweet num vs days from the event

Results

Twitter Sentiment Analysis

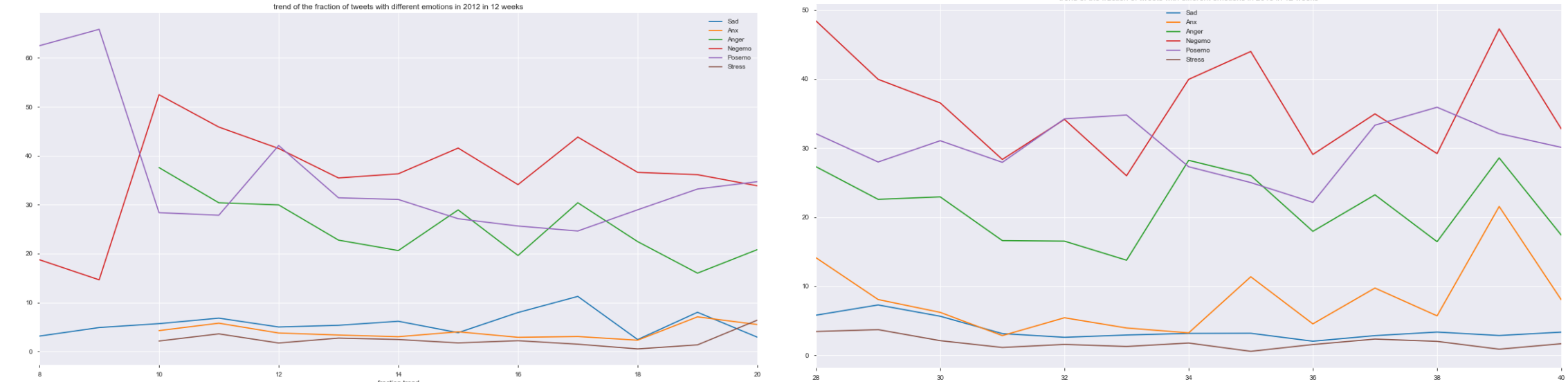


Figure 6 Trends of the fraction of tweets with different emotions in 2012 and 2013

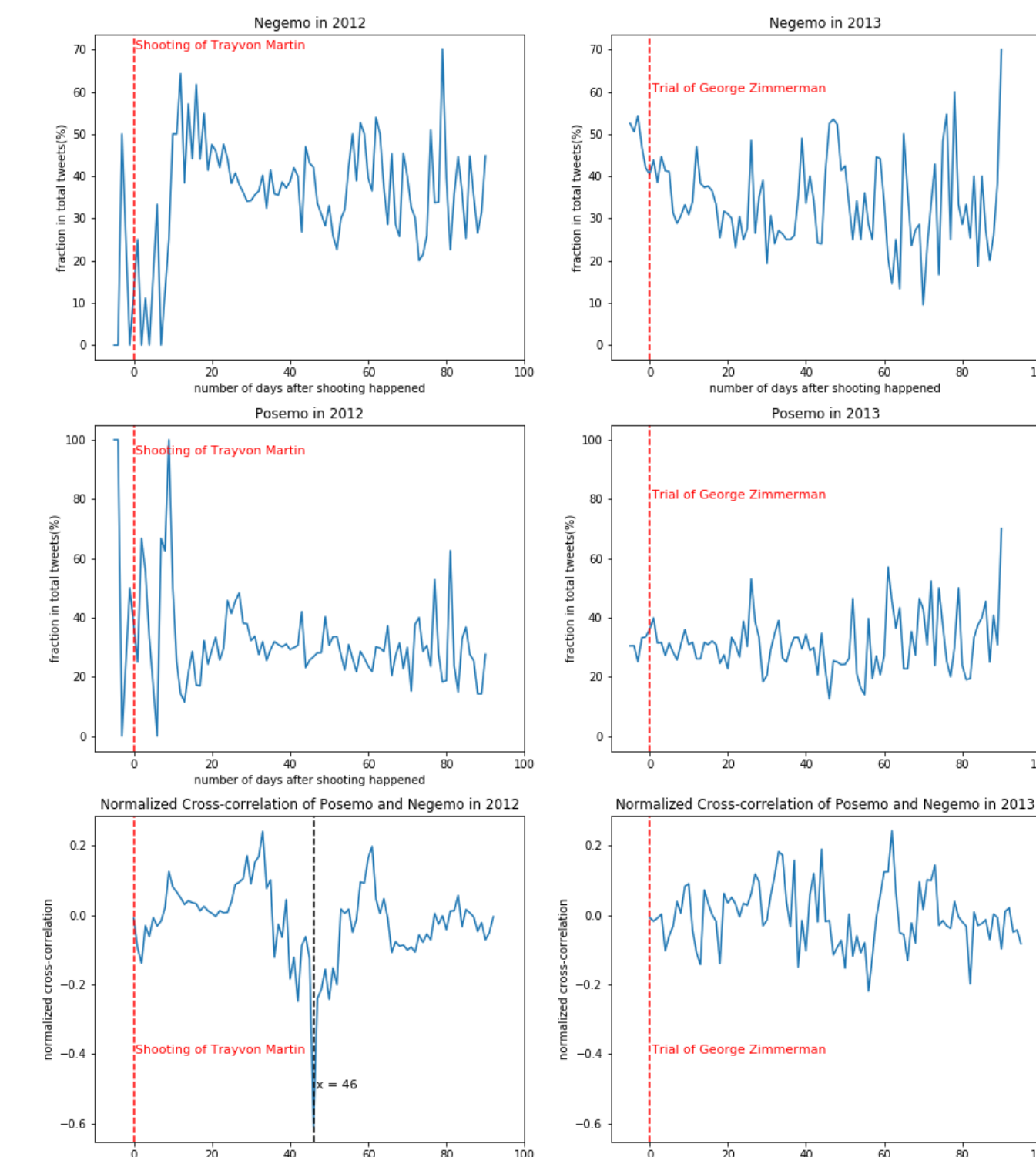


Figure 7 Cross-correlation of posemo and negemo in 2012 and 2013

Correlation Analysis

Changes in the normalized cross-correlation resonates with interrupted events.

Hypothesis testing

- The coefficient of variation in anxiety level is similar to that of negemo.
- The time needed for the number of tweets to drop to half peak volume in 2012 is at least 1 week while pattern for 2013 is hard to capture.

Conclusions and recommendations

Although analyzing Twitter data can be complicated, we have identified interesting features that can either confirm people's intuitions or counter them. The emotion pattern is more fluctuating in 2013 than in 2012, there might be interrupted events happened. This project could be improved further by gathering more detailed data, for instance, data with emoji information included, and by increasing the accuracy of emotion detection.

Acknowledgments

The authors gratefully acknowledge the support of the Columbia University Data Science Institute faculty and the LIWC dictionary, twitter dataset provided by Professor Colin Wayne Leach.

References

- text preprocessing: <https://github.com/Deffro/text-preprocessing-techniques>.
- #Bigbirds Never Die: Understanding Social Dynamics of Emergent Hashtags

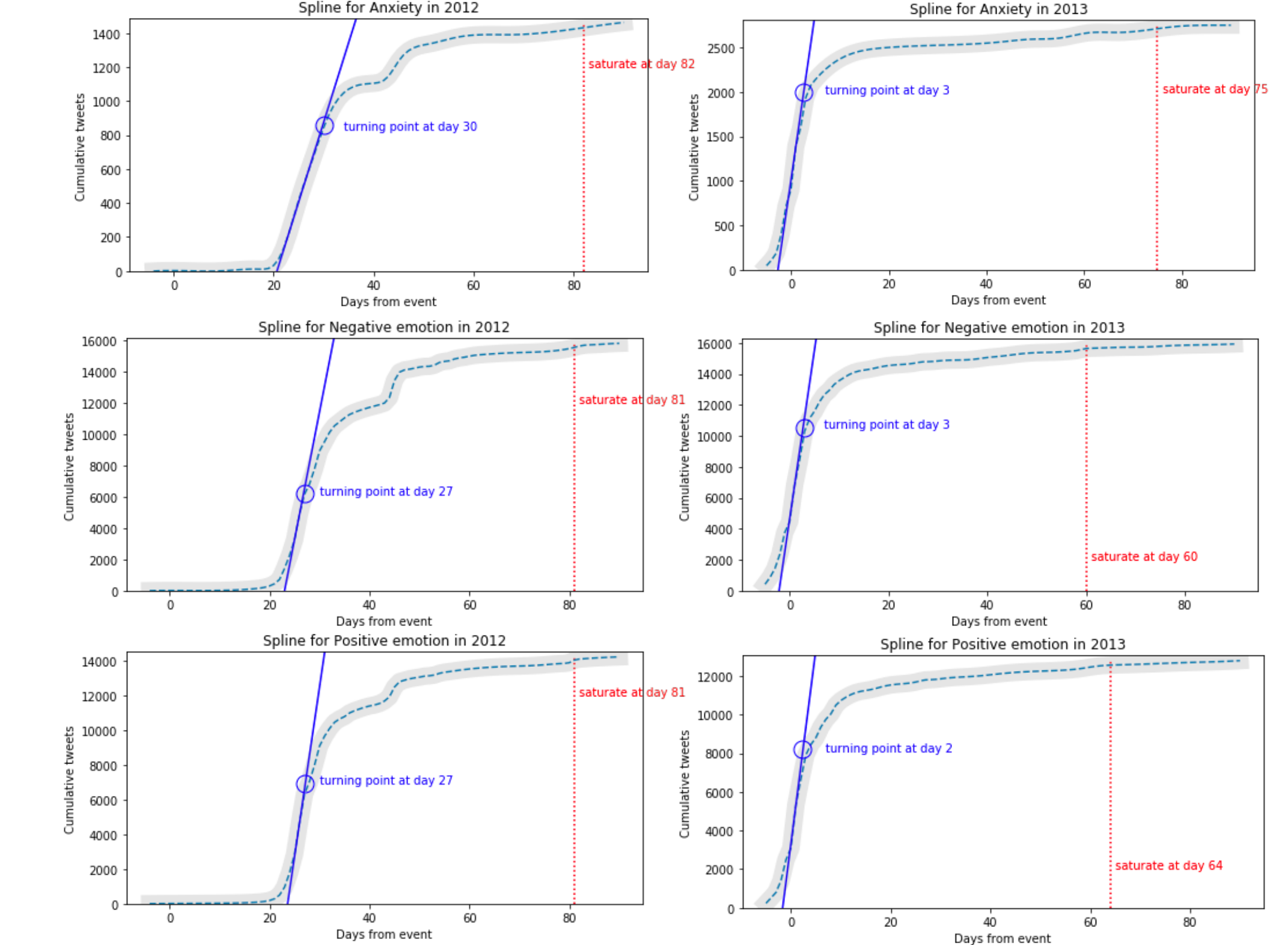


Figure 8 Spline analysis for emotions in 2012 and 2013