Capstone Report 2:

Social Media Sequential influence around incidents of police use of force against unarmed Black victims

Andrea Navarrete, Omar Pardo, Jose Lopez, Shadi Fadaee, Shimeng Feng

Professor: Colin Wayne Leach & Courtney Cogburn

1. Introduction

In order to further understand the influence of Twitter activity around issues of police violence against black victims, our team performed more analysis on the tweets applying topic modeling, sentiment analysis, and combining both topics and emotions across time. As indicated in our last report, we defined four categories: tweets including both *Ferguson* and *Michael Brown*, only Ferguson, only *Michael Brown,* or neither. For this report, we will focus on the first category, that represents 10% of the tweets. Working on that specific category, we conclude on 12 topics. Mapping the topics with timestamps allowed us to understand the shift of topics. For sentiment analysis, our team explored beyond polarized emotions, a.k.a. positive and negative and further explored both various dimensions of emotions and the intensity of each dimension. We have limited our final approach to using Vader and LIWC by examining the lexicon and application. In the following sections, we will describe our approaches, analysis, and discoveries.

2. Data Pre-processing

Aside from the data cleaning section within our first report, our team revisited the tweets field to generate clean text data to ensure the quality of further analyses. After limiting the hashtags under our key interests (refer to report 1), we have further performed the following data cleaning process:

1) Replaced link within a post with 'URL' and created a new indicator for media links
2) Cleaned punctuation and special symbols
3) Removed 'RT' and mentioning ('@') within tweets and created a new field to indicate retweets
4) Selected tweets written in English language

The provided dataset has no indicator of the relationship between retweets and the original tweet. There are multiple ways a user can retweet or refer to the same content within Twitter. For example, a user can retweet using the retweet button on the user interface, which shows up as a tweet text starting with 'RT'. Or a user could mention the original user and refer to a certain part of the tweet with additional information. Also, a user could screenshot the same content, add it as image information. Thus, is extremely challenging to exclusively build connections between retweets and tweets due to the diverse methods of retweets. challenging to identify retweets and the original tweet due to the diverse retweets methods. As of now, the team limited retweet using method 4 listed above. We understand this is not a comprehensive method and will still require further improvement.

3. Topic Modeling

   A. Introduction

As mentioned in the previous report, one of the motivations behind this project is to understand the underlying themes behind conversations happening on Twitter around the time of the incidents. We worked in the development of a process that will help us understand what is the best way to cluster all the different tweets into groups by affinity. We pre-processed data through the creation of a pipeline that included the following steps:

1) Decomposing the text into a bag of words.
2) Dropping stopwords using the english dictionary.
3) Creating bi-grams and trigrams for common expressions.
4) Lemmatizing the text to group different inflections of a word as the same term.
5) Building a dictionary with the resulting terms, to allow mathematical manipulation of text.

Since our data is comprised of tweets that were broadcasted during August, 2014 -the month when Michael Brown was shot and killed by officer Darren Wilson in Ferguson, Missouri- we wanted to understand what were the topics discussed surrounding this event. For that reason, we ran a subset of the data (only tweets containing words related to both Michael Brown and Ferguson) through the pipeline, with the objective of understanding what aspects of the incident were being talked about by the population.

B. Models

As mentioned in the previous report, we used the Latent Dirichlet Allocation model to group tweets into themes. We found that the Mallet implementation rendered a higher coherence score and the topics made more human sense, so we used it for our pipeline.

C. Output and analysis

The metric we use to assess performance from the topic modeling algorithm is coherence. A higher coherence means that the words within each topic has a higher semantic interpretability. Applying the Mallet model on comparing different number of topics, we discovered that the coherence score reaches a maximum as 0.37 when the corpus is grouped into 12 topics.

In order to better understand this grouping, we created a word cloud with the ten most relevant terms within each of the 12 topics. This chart is shown in figure 1, and in it we can see that terms are related to themes such as family (topic 1), protests (topics 0 and 5) and Michael Brown's funeral (topic 11). One of the next steps from this analysis will be to meet with our faculty, who as domain experts can help us define names for the topics.

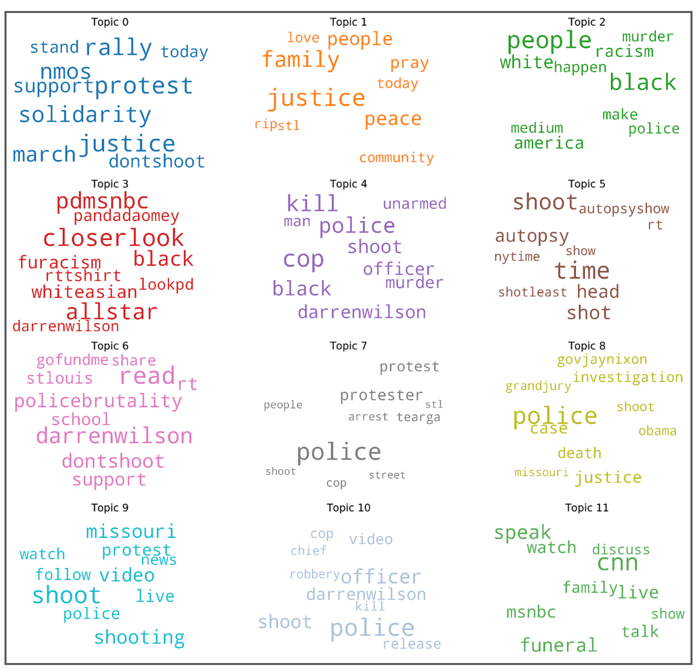

Figure 1. Word Clouds per topic

Next, we decided to analyze how different topics had a varying share of the tweets during the month of August. The topic modeling algorithm allowed us to determine the dominant topic for each tweet, as well as a measure. This enabled us to count how many tweets could be categorized into each topic, as well as to measure how strongly each tweet was related to its dominant topic. Finally, we wanted to measure the behavior of the topics that drove conversation, either by seeing which topic triggered the most retweets and which topic sparked the largest amount of original tweets.

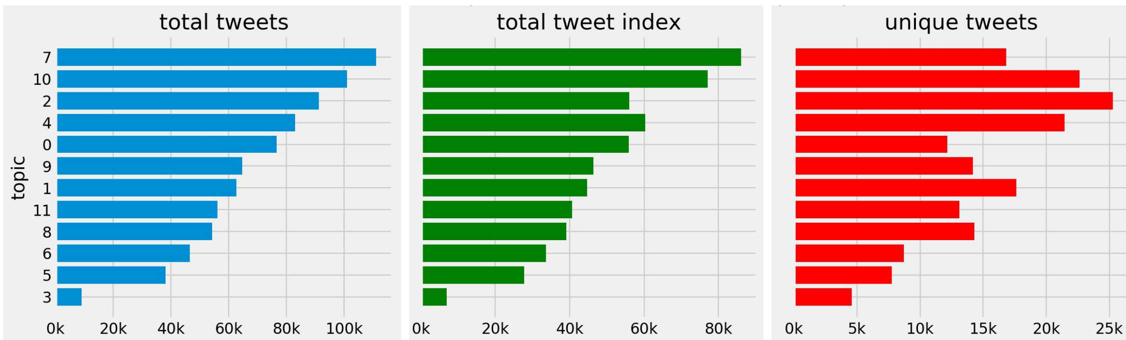

Figure 2. Distribution of tweets, contributions and unique tweets per topic

Figure 2 shows these patterns. It is interesting to see how topic 7 (related to riots and protests) had the most tweets attributed during the period, as well as the most tweets where it was the most dominant topic. However, in terms of unique tweets, it came in 5th place, which might be due to the fact that people were trying to stay informed about protests. Similarly, although topic 2 (dealing with words related to race, police and murder) had the most unique tweets, it had the 3rd largest number of total tweets, and in terms of dominance measure it ranked 4th.

Since one of the main objectives of this project is to understand how the events influence Twitter conversations through time, we created plots that show this evolution. Figure 3 contains a time series of number of tweets by dominant topic, as well as the share of the conversation had on each day that could be attributed to a theme.

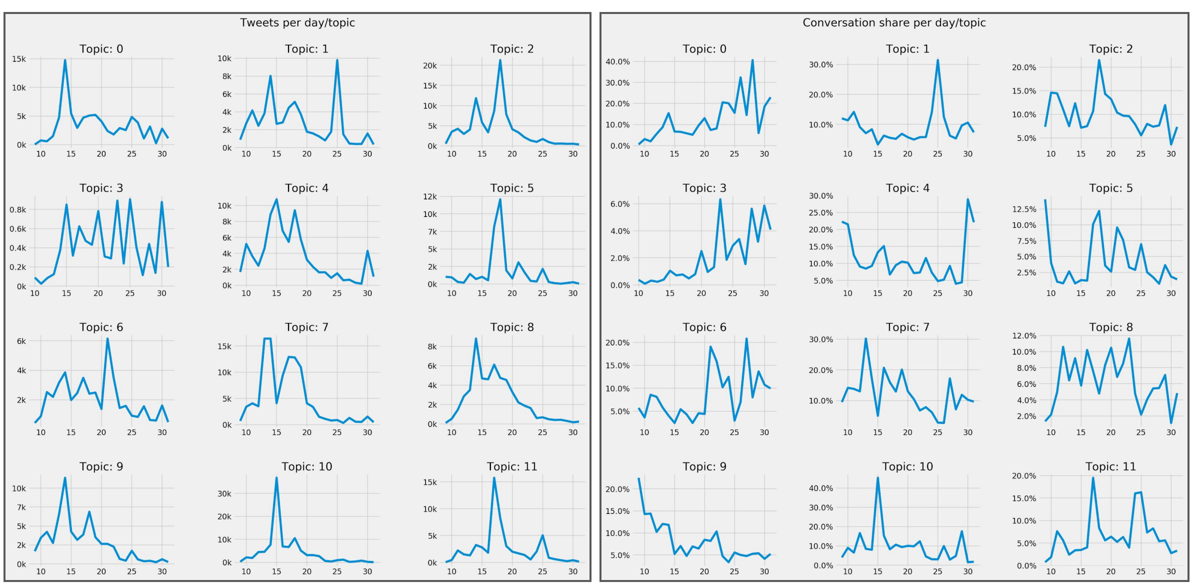

Figure 3. Number of tweets (left) and share of tweets (right) per day of August for each topic

The charts on figure 3 help us draw several insights from the data set, as it is now possible to understand and relate topics with events that took place over the month of August, 2014. Take for instance the behavior of topic 10: It spiked on August 15th, right on the day when officer Darren Wilson's name was released to the public. It also drove the conversation, as over 40% of the tweets in our sample are related to that topic.

Topic 0 allows for interesting insights to be drawn. The number of tweets related to it increased to a maximum just a few days after the murder of Michael Brown. This topic contains terms related to protests like *#dontshoot*, that took place on August 12th, hence the spike. Although the number of tweets drops towards the last days of the month, it became one of the topics that were on most tweets of our sample, specially after the funeral on August 25th.

Topic 1 (related to family, solidarity, justice and prayers) reached its maximum number of tweets around times of distress, like when the Missouri Highway patrol took over security in Ferguson, when the National Guard was ordered into the area and on the day of the funeral. It will be interesting to see if there is a relevant connection between a specific emotion and this topic.

4. Sentiment analysis
    A. Introduction

The following are some corpus / lexicon that we have identified. We have successfully applied both Vader and LIWC on the tweets and generated lists of words for each emotion. Then we explored how emotions change throughout different time periods.

LIWC (Linguistic Inquiry and Word Count): The default LIWC2015 Dictionary is composed of almost 6,400 words, word stems, and select emoticons. It contains 41 LIWC keys around psychological constructs (e.g., affect, cognition, biological processes, drives). The psychological processes section of the categories include positive and negative emotion, anxiety, anger and sadness.

For this analysis, we iterate for each tweet to find the list of words (word stems) of every LIWC key in each of the word categories. In this way for a particular tweet we have a list of words for every key category, a dummy variable that indicates the presence or not of a word form that key category on the tweet and the ratio calculated as the total number of matched words divided by the total length of tweet. Further analysis will use the presence or not of a word in each key LIWC category in order to analyse the percentage of tweets that contain a specific word category over time.

For LIWC, we have successfully determined the following word categories fields from the lexicon as important keywords to analyse over time:

- Emotions: Sad, Anger, Risk, Death, Anxiety
- Pronouns: I, We, You, SheHe, They
- Social: Family, Friend
- Other: Health, Drives, Affiliation, Swear, Percept, Sexual, Negate, Achieve, Power, Reward

Valence Aware Dictionary and Sentiment Reasoner (Vader): Sentiment is a rule-based sentiment model for social media/microblog-like context (Hutto, C.J. & Gilbert, E.E. (2014)). Vader is validated by multiple independent human judges and it incorporates a "gold-standard" sentiment lexicon. The VADER sentiment lexicon is sensitive to the polarity and the intensity of sentiments expressed in social media (Hutto, C.J. & Gilbert, E.E. (2014)).

The compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). This metric gives a unidimensional measure of sentiment for a given sentence, creating a normalized, weighted compound score (Hutto, C.J. & Gilbert, E.E. (2014)).

The positive, negative and neutral scores are ratios for proportions of text that fall in each category, such that they all add up to 1. If a multidimensional measure of sentiment is needed for a sentence, these are the metrics to use (Hutto, C.J. & Gilbert, E.E. (2014)).

VADER Sentiment Analysis on Twitter:

We applied Vader's scoring process to the tweets to discover the emotions and the intensity score of emotions for each tweet. We ran the Vader sentiment analyzer on each tweet to get the positive, negative, neutral and compound scores for each tweet. Then for each tweet we obtain the list of words that have corresponded to each category to understand the contribution of the words in the tweet in the scoring output. Whenever we wanted to find the intensity score for a single word we obtained the compound score of that word. We decided to focus mostly on the compound score of each tweet as it gives a normalized measure of sentiment for that specific tweet. The VADER sentiment analysis takes into account word_order, negation, words specific to twitter and emojis amongst other details specified in their paper.

The next step is to combine the LIWC dictionary with the intensity scoring from Vader. For each LIWC category, we run the Vader sentiment analyzer on the words of that category to find the intensity score of the words in each category. Then to understand the sentiment of the categories as a whole, we looked at the distribution of the intensity scores of words of a category, shown below.

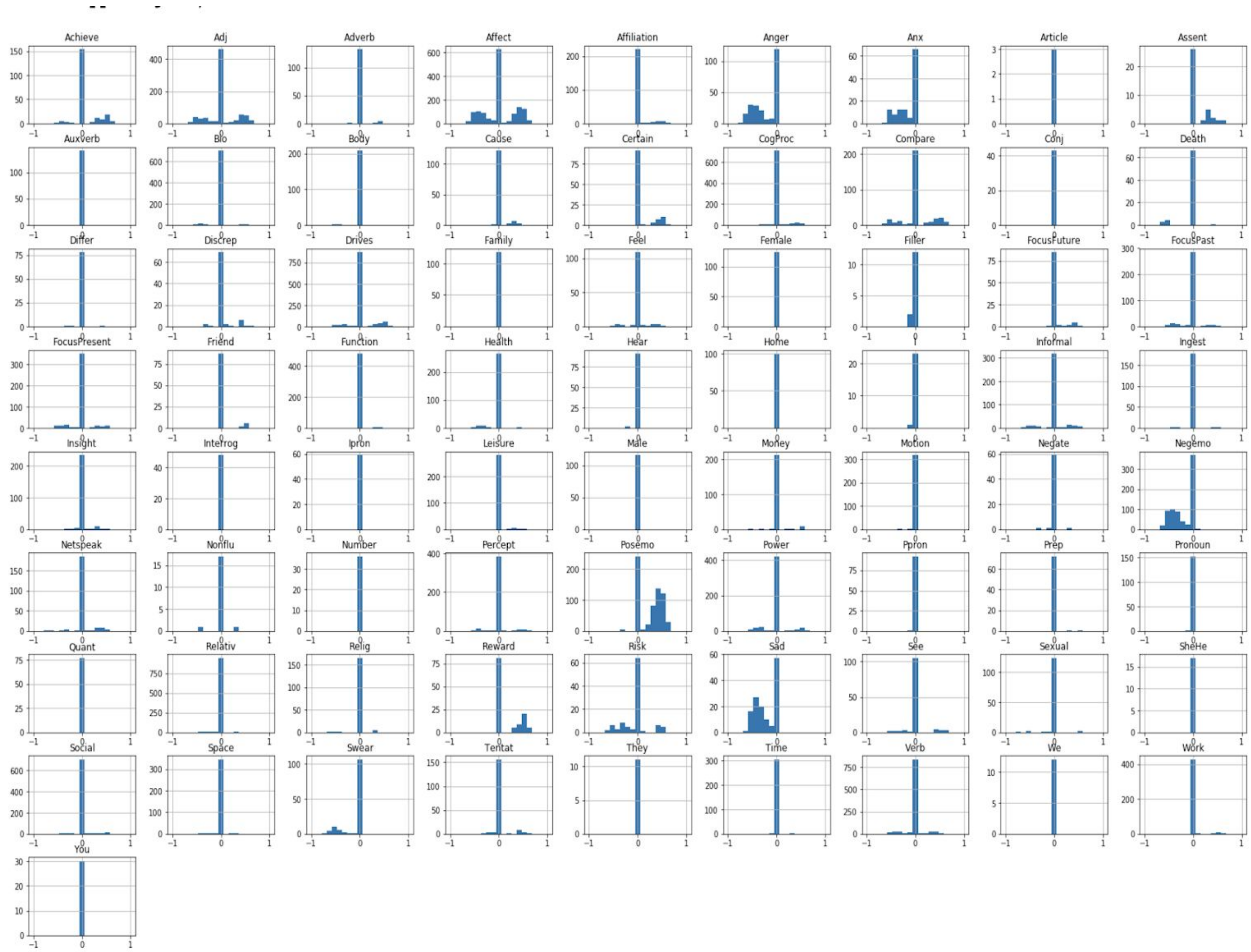

Figure 5. Distributions of categories in LIWC dictionary with Vader score

As we can see in the distributions, most of the categories' words have no sentiment associated with them and all categories have a majority of neutral words. However, some categories like anger, anxiety, Risk, Sad and Negemo indicate negative scoring words. Categories including affiliation, assent, certain, reward, Posemo show words that have positive scores. Adjective and affect show symmetric behavior around the mean, meaning that amongst the non-neutral words, there is a similar number of positive and negative words.
We then run the LIWC categories along with VADER intensity scores on each tweet, obtaining the scores for each tweet, the words of the tweets that match the LIWC categories, and the intensity of the words of the tweet corresponding to a LIWC category. This gives a holistic understanding of the sentiment intensity of a tweet, the sentiment of individual words in the tweet, and the list of the words of the tweet that corresponds to each LIWC category. However, as we already obtain the intensity scores using VADER and as we are analyzing the LIWC categories individually, we found that the combination of LIWC and VADER scores was not necessary for our analysis.

In the final step, we combined our sentiment analysis with the topics we found. We grouped our tweets by each topic and looked at the intensity scores over time. More details of this step will be discussed in the next report.

NRC Word-Emotion Association Lexicon: The lexicon includes two sentiments and 8 emotions such as anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The lexicon can be applied to both English and over 100 other languages. Lexicon can be downloaded directly for non-commercial usage. Further research on NRC has revealed that the lexicon under each emotion category has included broader content than the emotion definition derived from Emotion Knowledge: Further Exploration of Prototype Approach (Shaver et al.) Thus, the output could introduce noise and misunderstanding of the emotion changes within crowds during the events. Additionally, the lexicon was generated using Mechanical Turks through labeling emotions based on various context. This method could potentially introduce a certain level of bias and affect reliability. Through consideration and discussion with the professor. We decided not to proceed with the NRC emotion dictionary. In addition to the emotion lexicon, further research shows NRC also has an emotion/affect intensity lexicon to show intensities of emotions, and a valence, arousal, and Dominance(VAD) Lexicon. As indicated, valence is the positive-negative or pleasure--displeasure dimension, arousal is the excited--calm or active-passive dimension; and dominance is the powerful-weak or 'have full control'-'have no control' dimension. Even though both lexica have adjusted the rating score consistency through Best-Worst Scaling, we decided not to proceed because of the broad words included in the final dictionary.

B. Further understanding of emotions

There are many different methods of evaluates emotion, and considering the purpose of this research, we have focused on the Hierarchical circumplex model of emotion. There are two major dimensions, activation/ deactivation and unpleasant/pleasant. For the unpleasant and pleasant dimension, an example is that condemn is associated with a greater degree of anger than irritated. This is in sync with the intensity model provided by VADER. It is important to clarify that the activation emotion does not lead to action readiness, but potentially more stress and mental influence. And an example of the activation and deactivation dimension is alert vs. calm. Thus our analysis of emotion through various dimensions could unveil physiological arousal, motor expression, and psychological emotion impact.

In addition, we are interested in the Typology of group relevant emotion, which can be identified through words conveys Group or individual, ingroup or outgroup, or subject or object of emotion. These words help us understand the flow of emotion through different events and how individuals' relationship with the world through social media.

C. Output and analysis

Using the LIWC dictionary, we are able to see the trends for the different keywords based on the percentage of tweets that use a word related to that specific keyword. By looking at the different trends it is possible to notice different patterns:

- The percentage of tweets related to risk, negata, sexual are decreasing, while perception and power increases over time.
- There are important peaks:
    - On the Funeral day (25) especifically for Sad, Death, Affiliation and Family (with a noticeable changed), while Reward, Power, Anger decreased.
    - On August 20th we see a spike in Swear. This might be due to the heavy protests on 18, 19 and 20 when national guard was ordered there due to the riots.
    - August 28th Protest for Affiliation, Percept, Anger, Risk, Reward, as well as more We and You pronouns used.

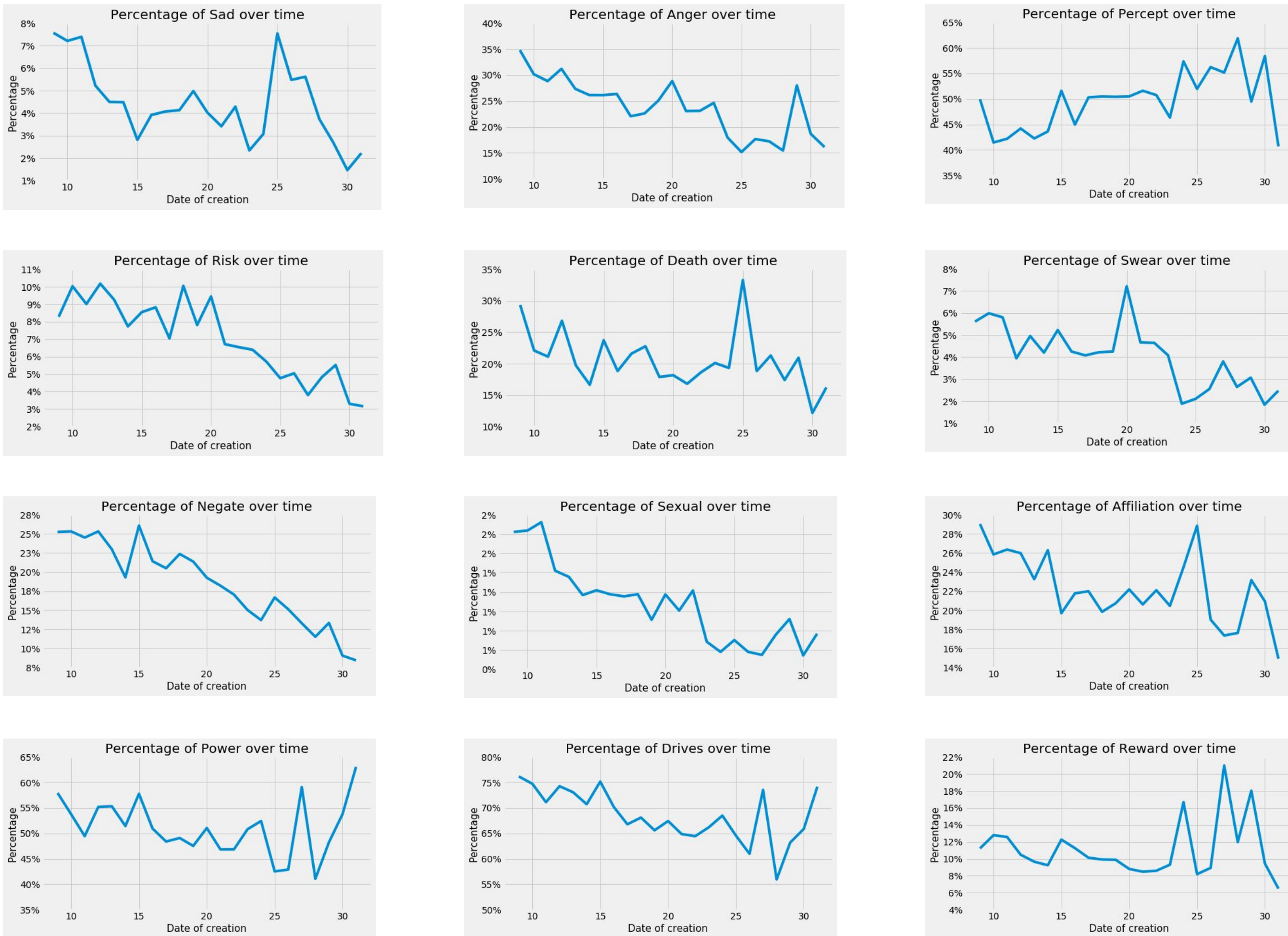

Figure 6. Percentage of words per day for August, 2014

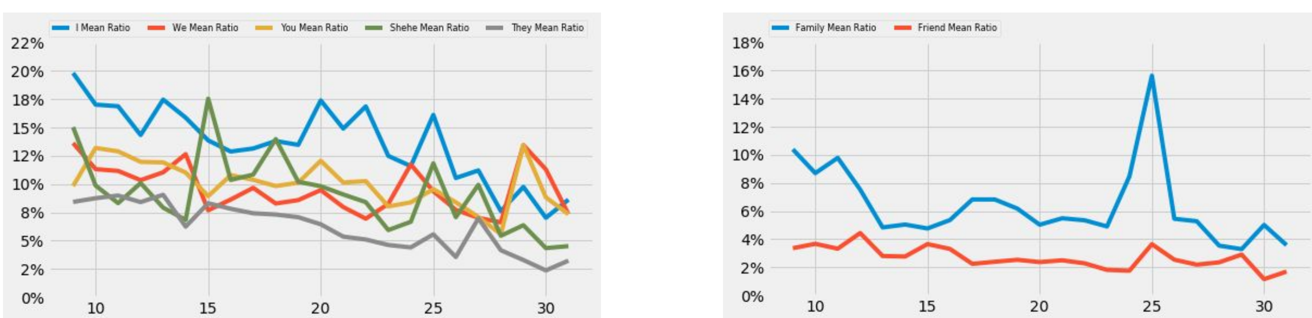

Figure 7. Ratio of pronoun use per day

5. Additional - Image Analysis:

Since a picture is worth a thousand words, we have further explored a sample of tweet images through clustering to discover the similarities for further exploration. Here are some examples:

Figure 8. Police / Military

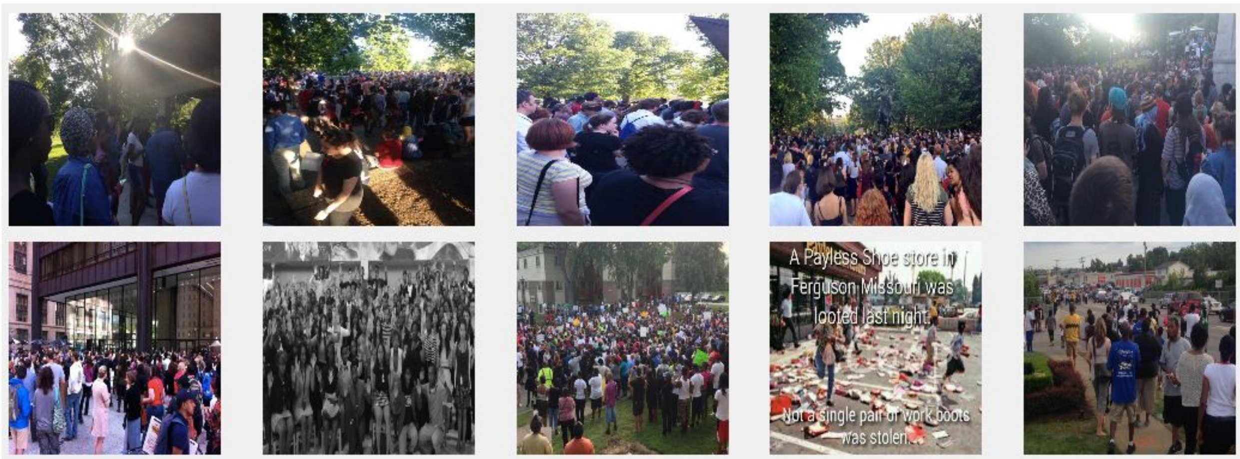

Figure 9. Crowds

Figure 10. Screens / additional text

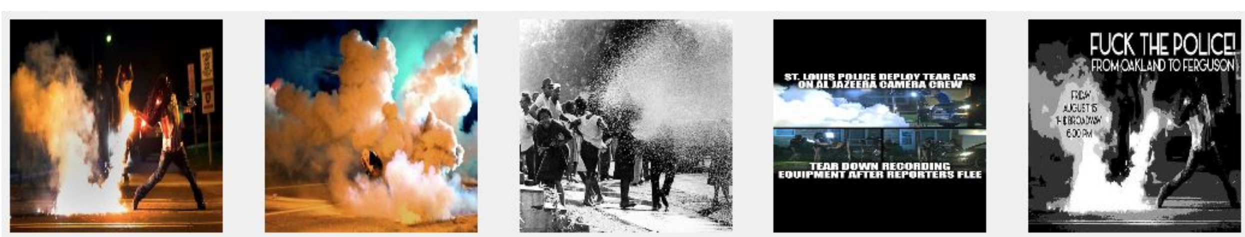

Figure 11. Violence

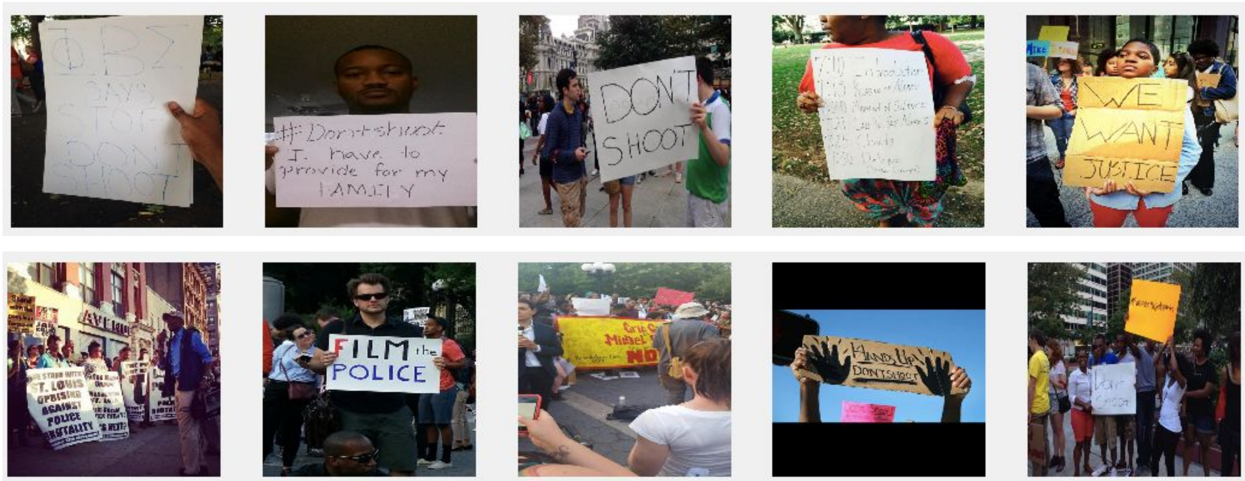

Figure 12. Protest - protester

There are other categories that we can further explore, but from the above we can further explore: text within an image and emotion conveyed through the image, and enrich the topic modeling and sentiment analysis discussed above.

6. Contribution

As in the previous progress report, each team member has made valuable contributions to the current status of the project. These contributions included further cleaning and making sense of the dataset to guide analyses for the project. Specific efforts include Andrea, Shadi and Shimeng working on sentiment analysis, Omar and Jose working on the topic modeling pipeline. Everyone contributed to the presentations made for checkpoint meetings with the faculty, and to the construction of this report.