**Statistical Investigation of Organizational Responses on Twitter Online Firestorm**

**Jocelyn Tan**

## Abstract

A statistical investigation on an online firestorm of JPMorgan in 2013 was conducted in this study in order to examine the effectiveness of the response provided by the company. Sentiment analysis was utilized to extract the sentiment from tweets and sentiment difference before and after the response was tested by statistical tests. Two sentiment analysis tools were used to compare the results. After observing significant increase in negative sentiments and decrease in positive sentiment from the two different tools, the study concluded that the response was not very effective.

## Introduction

Twitter has been a great platform for word-of-mouth propagation, which makes it one of the major social media for marketing campaign and communication. However, the speed of propagation sometimes can make the brand image of a company in jeopardy by bringing huge waves of negative comments and complaints with outrage in a short period of time. In the study of Pfeffer et al. (2014)[1], it defined the sudden influx of messages with huge amount of negative word-of-mouth and complaints towards companies in social media as online firestorm. In this paper, through a study of an online firestorm case of JPMorgan, the consequences of online firestorm have been revealed and the effectiveness of the organizational response of the company has been examined using statistical tests.

## Case Study

At 2:41 PM, November 6th, 2013, the JPMorgan sent a tweet from its corporate account to announce a live Twitter Q&A about leadership and career advice hosted by one of its executives. It created a hashtag "#AskJPM" and encouraged participants to submit questions using this hashtag.


One day later, November 7th, 2013, it posted a second tweet related to #AskJPM as a reminder.

[1] Pfeffer, Zorbach, Carley. (2014). Understanding Online Firestorms: Negative word-of-mouth dynamics in social media networks Journal of Marketing Communications. Vol. 20, Nos. 1–2, 117–128, http://dx.doi.org/10.1080/13527266.2013.797778

J.P. Morgan @jpmorgan

What career advice would you ask a leading exec at a global firm? Tweet a Q using #AskJPM. On 11/14 a $JPM leader takes over @JPMorgan

4 11:36 AM - Nov 7, 2013

See J.P. Morgan's other Tweets

On November 8th, 2013, it revealed the executive who hosted the Q&A session:

J.P. Morgan @jpmorgan

He's been a part of Wall St.'s biggest deals @Dell, @GM. $JPM Vice Chairman Jimmy Lee takes over @JPMorgan on 11/14 #AskJPM

5 10:33 AM - Nov 8, 2013

See J.P. Morgan's other Tweets

One week later, it sent out a reminder tweet:

J.P. Morgan @jpmorgan Follow

$JPM VC Jimmy Lee is taking over @JPMorgan on 11/14 at 1pm ET. Tweet Qs using #AskJPM & learn more about him here: jpm.com/x/a/NFPV4Fj

10:26 AM - 13 Nov 2013

While hardly any retweet or attention on the original tweet, on November 13th, the hashtag started to be used by Twitter users frequently to post questions related to the ethic of the bank. JPMorgan had been negotiating an agreement with the U.S. over bad mortgages and two ex-employees were indicted for their attempt to cover up a huge trading loss. Some of the questions asked by Twitter users:

The Other 98% @other98 · 13 Nov 2013

Replying to @jpmorgan

Hey, @jpmorgan: What's it like knowing that you've escaped punishment for securities fraud again & again? #AskJPM other98.com/dont-negotiate…

1 42 33

1 more reply

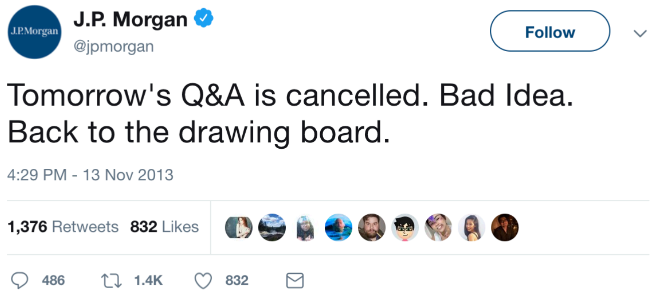
Chris Lee @Chris_Lee_Cal · 13 Nov 2013

Replying to @jpmorgan

@jpmorgan Why is it when poor people commit crimes we need more police and when rich people commit crimes we need deregulation? #askJPM

13 9

Hours later, JPMorgan realized the Q&A session had turned into an online firestorm which might be out of their control and they called it off at 4:29 PM, Nov 13th, 2013.




## Methodology

### Twitter Mining Framework

In general, hashtags are used in tweets before a keyword or phrases relevant to the topic of the user, with no space between the hashtags and the phrases, in order to categorize the content, help user keep track of the content and updates the relevant topic.[2] To identify the tweets related to this online firestorm, the hashtag #AskJPM created only for this event is used.

### *Data Collection*

Twitter data can be accessed by using Twitter API which has rate limit and limitation on fetching historical data[3]. The size of the data was anticipated to be huge based on the nature of online firestorm and tweets in 2013 were needed for this particular case.
To avoid the limitations of Twitter API and obtain the completeness of the dataset, an advanced search option in Twitter was used[4]. After carefully analysing the hashtags used for people to comment on the event, #AskJPM is chosen as the only hashtag to use during the data retrieval process to avoid collecting noisy data.
Related tweets (searched by #AskJPM) were collected from November, 6th, 2013 to February, 4th, 2014. The initial size of the collected tweets was 13,634. Some of the tweets

---

2 S. Das, et al., Extracting patterns from Twitter to promote biking, IATSS Research (2018), https://doi.org/10.1016/j.iatssr.2018.09.002

3 Twitter. Consuming streaming data. Retrived from: https://developer.twitter.com/en/docs/tutorials/consuming-streaming-data#overview

4 Historical tweet Access, https://github.com/Jefferson-Henrique/GetOldTweets-python Accessed Apr 29, 2019.

were deleted because they contained empty content after data cleaning. For the final analysis, 11,024 relevant tweets were analysed.
The response was posted by the company on 4:29 pm, November 13, 2013. There were 492 tweets before the responses and 10,532 tweets after the responses.
From figure 1, it can be observed that majority of the tweets related to the events were posted within three months of the start of the event. The high volume and quick speed confirmed the features of an online firestorm.

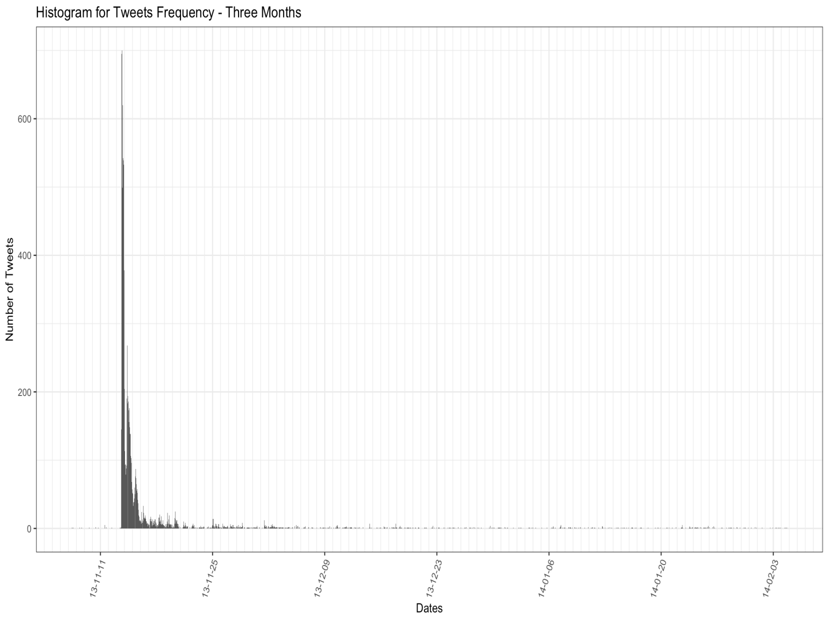



*Figure 1 Histogram for Tweets Frequency - Three Months*

From figure 2 which mainly focus on the frequency within the first two weeks (10275 tweets, 93.2% of total tweets in the final dataset). The peaks located on November 13th, 2013 which was the same day the company posted their response to the online firestorm.

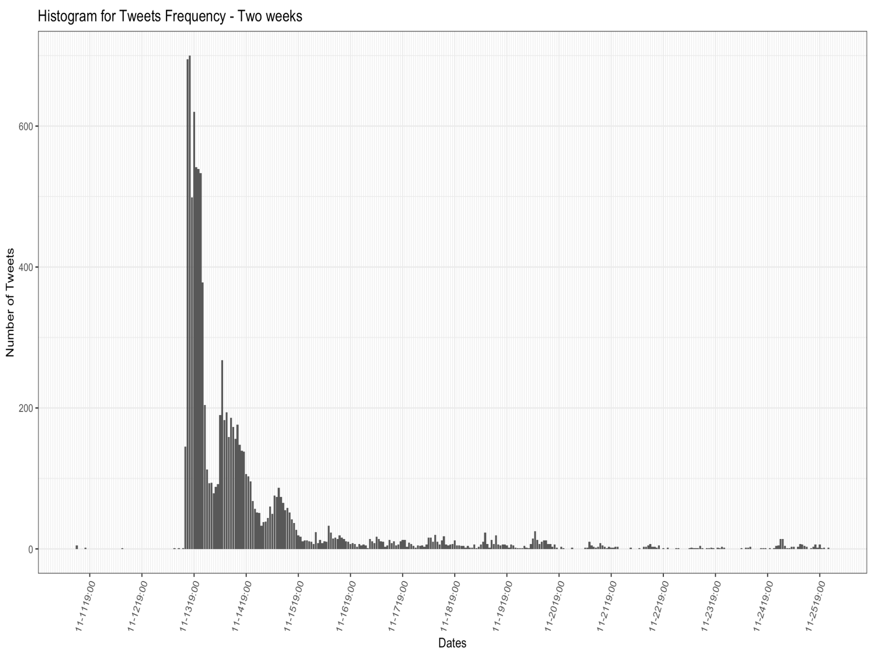



*Figure 2Histogram for Tweets Frequency - Two Weeks*

From figure 3 which mainly focus on the frequency within the 5 days of the response (9348 tweets, 84.8% of the final dataset). The peaks located between 4 pm to 5 pm on November 13th, 2013 which within 1 hour when the company posted their response to the online firestorm.

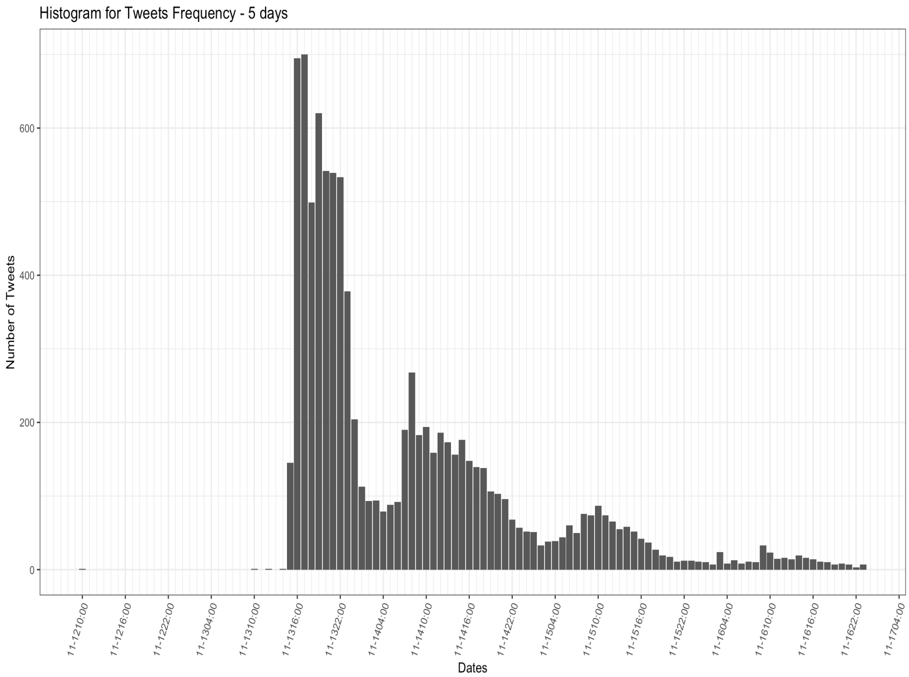



*Figure 3Histogram for Tweets Frequency - 5 days*

From figure 4 which mainly focus on the frequency within the day of the response (4654 tweets, 42.2% of the final dataset). The peaks located between 4 pm to 5 pm on November 13th, 2013 which within 1 hour when the company posted their response to the online firestorm. Interestingly, the volume was much smaller before the response was posted.

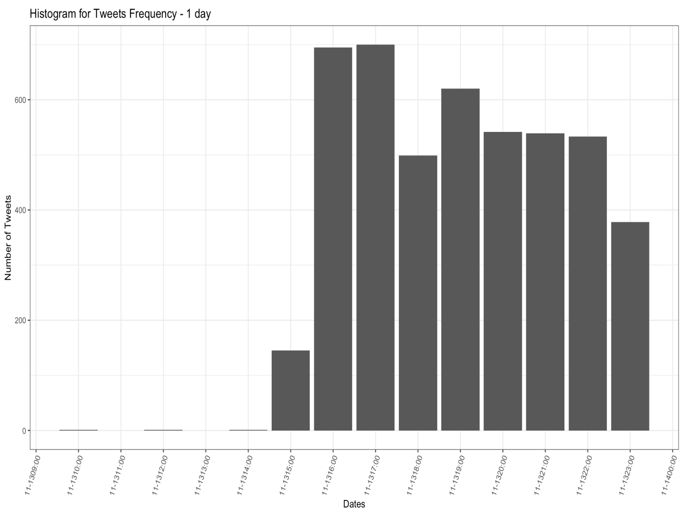



*Figure 4 Histogram for Tweets Frequency - 1 day*

Data Cleaning

The collected tweets contains some noisy data such as web links, redundant contents, non-ASCII and handles. To avoid affecting the sentiment analysis in later stage, those noisy contents were all removed using R base package. For hashtags, only the number sign "#" has been removed due to the fact that Twitter users tended to use hashtags to express their feelings and some of them wrote the whole content with "#" in front of each word.

Sentiment Analysis

Two different approaches were used to compare the results.

LIWC

LIWC (Linguistic Inquiry and Word Count) is a text analysis software that provides evaluation of emotion, cognition and structure of a given text based on the dictionary consisting of words and categories. [5] Among the results given by LIWC output, all the sentiments and outputs were analysed to compare with tidytext results.

NRC from Tidytext

NRC lexicon is used for extracting the sentiment for each tweet. The NRC Emotion Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). After utilizing the functions provided in this package, the term frequency table of 10 sentiments were generated. Total negative sentiment and total positive sentiment were calculated from those 10 sentiments to compare with LIWC results.

Statistical Test

Independent t-tests were conducted to examine whether there was significant change in sentiment before and after the response.
There were 492 tweets before the response and 10,532 tweets after the response. Thus, the sample size for tweets before the response ('before' dataset) was 492 and the sample size for after response ('after dataset') was 10,532.
First, normality of the 'before' dataset has been tested using Shapiro-Wilk test[6] for it contains much fewer tweets. For the 'after' dataset, the number of observation violated the limitation of the Shapiro-Wilk test in the stats packages in R. The limitation is used to avoid the fact that for large amounts of data even very small deviations from normality can be detected, leading to rejection of the null hypothesis even though for practical purposes the data is normal. Thus, it was assumed that it was close to normal distribution by central limit theorem. Further investigation can be conducted if needed.
F-test to compare the variance was also conducted in order to use the right t-test for different sentiment.

5 Goncelves et al. (2014). Comparing and Combining Sentiment Analysis Method. Retrievd from: https://arxiv.org/pdf/1406.0032.pdf

6 Shapiro-Wilk Test. Retrieve from: https://en.wikipedia.org/wiki/Shapiro%E2%80%93Wilk_test

Tests were conducted using the stats packages in R for Tidytext results and Python for LIWC results.

## Results and Discussion

LIWC

Among 93 sentiments generated by the LIWC software, the categories in LIWC output in the table below had significant changes before and after the response. The mean of sentiment score of 'Negative emotion', 'Anxiety', 'Affiliation', 'Risk', and 'Leisure' increased significantly after response. Most of the sentiments were negative sentiments. The dictionary words decreased significantly while the punctuation and netspeak increased. Further investigation of the sentiments for different punctuations can be conducted due to the fact that the exclamation has increased significantly after response which might be a sign of surprise or anger.

| Sentiments | Category / Examples | Diff | Sentiments | Category / Examples | Diff |
|---|---|---|---|---|---|
| **Analytic** | Analytical Thinking | Increase | **Dic** | Dictionary words | Decrease |
| **Sixltr** | Words > 6 letters | Increase | **Function** | It, to, no, very | Decrease |
| **affect** | Affective processes: happy, cried | Increase | **Pronoun** | I, them, itself | Decrease |
| **negemo** | Negative Emotion | Increase | **Ppron** | I, them, her | Decrease |
| **anx** | Anxiety | Increase | **You** | Second person | Decrease |
| **drives** | drives | Increase | **Shehe** | Third pers singular | Decrease |
| **affiliation** | Ally, friend, social | Increase | **Auxverb** | Am, will, have | Decrease |
| **risk** | risk | Increase | **Adverb** | Very, really | Decrease |
| **leisure** | leisure | Increase | **Conj** | And, but, whereas | Decrease |
| **informal** | Informal language | Increase | **Verb** | Common verbs | Decrease |
| **netspeak** | Btw, lol, thx | Increase | **Interrog** | How, when what | Decrease |
| **Colon** | | Increase | **Quant** | Few, many, much | Decrease |
| **Exclam** | | Increase | **Cogproc** | Cause, know, ought | Decrease |
| **Apostro** | | Increase | **Discrep** | Should, would | Decrease |
| **OtherP** | Other punctuation | Increase | **Tentat** | Maybe, perhaps | Decrease |
| | | | **Differ** | Hasn't, but, else | Decrease |
| | | | **focuspresent** | Today, is, now | Decrease |

*Table 5 LIWC Output Significant T Test Results*

Note: Diff (Significant mean difference (After – Before)

NRC from Tidytext

Among 10 sentiments generated by the 'nrc' lexicon and the 2 sentiments calculated representing the total negative and total positive sentiments, only 'positive' and 'trust' were significantly different before and after response. For 'positive', the mean decreased from 0.99 to 0.88. For 'trust', the mean dropped from 0.63 to 0.53. Both of them were positive sentiments and decreased significantly after response.

Although two different methods generated different sentiments due to the default of the software or library, the implication of the results were the same. The response did not effectively resolve the outrages of the Twitter users.

Discussion

Apart from the statistical test, the organizational strategy can be categorized as 'respond' according to the paper of Thomas et al (2012). [7] It was defined as strategy involving listening to, acknowledging, and resolving the negative feedback through social media potentially. If the response was effective, this strategic option could be used to quickly react to their clients or even convert them into loyal customers. However, one disadvantage could be the requirement of appropriate time of response. Also, the paper pointed that in situations where companies were unfairly and inaccurately attacked and the response was not well-received, using this option will not be effective. In this case, the number of comments grew rapidly after the response could be a sign of ineffective response.

## Conclusion

By investigating the #AskJPM case, the study has examined the effectiveness of this particular response provided by JPMorgan in term of sentiment difference before and after the response by using statistical tests. Both NRC lexicon and LIWC software provided similar results that the response was not very effective regarding soothing the tension and quell the negative word of mouth.

[7] Thomas, Jane B.; Peters, Cara O.; Howell, Emelia G.; and Robbins, Keith (2012) "Social Media and Negative Word of Mouth: Strategies for Handing Unexpecting Comments," Atlantic Marketing Journal: Vol. 1 : No. 2 , Article 7. Retrieved from: https://digitalcommons.kennesaw.edu/amj/vol1/iss2/7

**Appendix**

**Table 1 NRC Result**

**NRC**

| Emotion | Normal Dist | Equal Variance | Mean (before) | Mean (after) | P-Value | Result (Significant Diff) |
|---|---|---|---|---|---|---|
| Anger | No | Yes | 0.38655462 | 0.3859515 | 0.98573095 | No |
| Anticipation | No | Yes | 0.45658263 | 0.4777778 | 0.56627838 | No |
| Disgust | No | Yes | 0.28571429 | 0.2817369 | 0.88850988 | No |
| Fear | No | Yes | 0.40896359 | 0.4252874 | 0.63681392 | No |
| Joy | No | Yes | 0.40896359 | 0.3720307 | 0.25900493 | No |
| Negative | No | No | 0.78151261 | 0.8598978 | 0.08703644 | No |
| Positive | No | Yes | 0.99159664 | 0.8777778 | 0.02033461 | Yes |
| Sadness | No | Yes | 0.37815126 | 0.3661558 | 0.71231887 | No |
| Surprise | No | Yes | 0.2605042 | 0.250447 | 0.70894894 | No |
| Trust | No | Yes | 0.6302521 | 0.5320562 | 0.01380497 | Yes |
| Total Negative | No | Yes | 2.24089636 | 2.3190294 | 0.58095592 | No |
| Total Positive | No | Yes | 2.74789916 | 2.5100894 | 0.08748104 | No |

## Table 2 LIWC T Test Results

## Significant Increase/Decrease in Sentiment

| Sentiments | Before_Normality | After_Normality | Equal_Variance | Before_Mean | After_Mean | Diff | T-test_P | Significant |
|---|---|---|---|---|---|---|---|---|
| **Analytic** | No | No | Yes | 58.2855691056909 | 65.9509048613758 | Increase | 7.93810216496432E-07 | Yes |
| **Sixltr** | No | No | No | 15.8744105691057 | 18.9198604253702 | Increase | 1.7275584009635E-08 | Yes |
| **Dic** | No | No | No | 73.1655487804877 | 70.3299164451193 | Decrease | 7.62097988270589E-05 | Yes |
| **function** | No | No | No | 43.2573577235773 | 38.8381418533991 | Decrease | 1.77688230681138E-12 | Yes |
| **pronoun** | No | No | Yes | 11.6935975609756 | 9.65092100265839 | Decrease | 5.7356886011646E-07 | Yes |
| **ppron** | No | No | Yes | 7.73615853658537 | 5.96161033042153 | Decrease | 4.16872511676754E-08 | Yes |
| **you** | No | No | No | 4.71817073170732 | 3.07353304215723 | Decrease | 4.97917469054272E-09 | Yes |
| **shehe** | No | No | No | 0.431910569105691 | 0.233653627041398 | Decrease | 0.0253432288666305 | Yes |
| **auxverb** | No | No | No | 9.47101626016261 | 7.59562856057715 | Decrease | 1.73815441680772E-10 | Yes |
| **adverb** | No | No | Yes | 4.95768292682927 | 4.4115704519559 | Decrease | 0.0424746069920561 | Yes |
| **conj** | No | No | No | 4.48002032520325 | 3.59914736042534 | Decrease | 0.000728153803579778 | Yes |
| **verb** | No | No | Yes | 16.1182520325203 | 13.4534058108619 | Decrease | 1.61335102923153E-09 | Yes |
| **interrog** | No | No | No | 3.33886178861788 | 2.5532662362324 | Decrease | 0.000231016728937805 | Yes |
| **quant** | No | No | No | 1.86209349593496 | 1.50029434105584 | Decrease | 0.0393132092513737 | Yes |
| **affect** | No | No | No | 5.44532520325204 | 6.86136251424223 | Increase | 1.97239421566035E-05 | Yes |
| **negemo** | No | No | No | 2.12123983739837 | 3.14223224458789 | Increase | 6.47547290383616E-06 | Yes |
| **anx** | No | No | No | 0.142357723577236 | 0.490653247246488 | Increase | 4.93344528625302E-12 | Yes |
| **cogproc** | No | No | Yes | 11.0504471544716 | 9.44086213444723 | Decrease | 0.000107127117945445 | Yes |
| **discrep** | No | No | No | 1.81260162601626 | 1.24745442461072 | Decrease | 0.00190713251809469 | Yes |
| **tentat** | No | No | No | 3.35548780487805 | 2.08751044436003 | Decrease | 1.79665996991745E-06 | Yes |
| **differ** | No | No | No | 2.85447154471545 | 2.12118401063424 | Decrease | 0.00133049094482078 | Yes |
| **drives** | No | No | No | 7.32867886178862 | 9.45724933535872 | Increase | 4.06330727597E-09 | Yes |
| **affiliation** | No | No | No | 1.8050406504065 | 3.64926129889855 | Increase | 3.56788182288268E-23 | Yes |
| **risk** | No | No | No | 0.918252032520325 | 1.31425370300039 | Increase | 0.02687586052143 | Yes |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **focuspresent** | No | No | Yes | 10.7989430894309 | 9.69432491454592 | Decrease | 0.00379889067366329 | Yes |
| **leisure** | No | No | No | 0.68619918699187 | 2.12966388150398 | Increase | 1.15484885635569E-35 | Yes |
| **informal** | No | No | No | 3.11315040650406 | 3.93024971515377 | Increase | 0.00676969124422486 | Yes |
| **netspeak** | No | No | No | 2.09764227642276 | 2.97511203949865 | Increase | 0.000248583284067875 | Yes |
| **Colon** | No | No | No | 0.617357723577236 | 1.05760634257501 | Increase | 0.000566279941699644 | Yes |
| **Exclam** | No | No | No | 0.420040650406504 | 1.30896315989366 | Increase | 1.32037090120445E-12 | Yes |
| **Apostro** | No | No | No | 1.89020325203252 | 2.518463729586 | Increase | 0.000433813608504916 | Yes |
| **OtherP** | No | No | No | 0.998150406504065 | 1.90956703380174 | Increase | 7.05180633268258E-09 | Yes |

## Insignificant Increase/Decrease in Sentiment

| Sentiments | Before_Normality | After_Normality | Equal_Variance | Before_Mean | After_Mean | Diff | T-test_P | Significant |
|---|---|---|---|---|---|---|---|---|
| **WC** | No | No | Yes | 14.9268292682927 | 14.5498480820357 | Decrease | 0.181852266334321 | No |
| **Clout** | No | No | No | 69.2467479674796 | 69.4573442840871 | Increase | 0.884893873741208 | No |
| **Authentic** | No | No | Yes | 31.4142479674797 | 29.6601243828337 | Decrease | 0.267041388214001 | No |
| **Tone** | No | No | Yes | 42.3743699186994 | 39.9524762628202 | Decrease | 0.157114524459071 | No |
| **WPS** | No | No | Yes | 8.58579268292683 | 9.03374952525637 | Increase | 0.0517927127866307 | No |
| **i** | No | No | Yes | 1.53065040650407 | 1.47644037219902 | Decrease | 0.763245234224388 | No |
| **we** | No | No | Yes | 0.461544715447154 | 0.613982149639197 | Increase | 0.15998401272452 | No |
| **they** | No | No | Yes | 0.593943089430894 | 0.564483478921386 | Decrease | 0.763671336533796 | No |
| **ipron** | No | No | Yes | 3.92638211382114 | 3.68354063805541 | Decrease | 0.328449300426041 | No |
| **article** | No | No | Yes | 5.29002032520325 | 5.34277250284841 | Increase | 0.842971929595447 | No |
| **prep** | No | No | Yes | 9.93548780487806 | 10.247086023547 | Increase | 0.38470262384687 | No |
| **negate** | No | No | Yes | 1.07855691056911 | 1.25870584884163 | Increase | 0.208587262434085 | No |
| **adj** | No | No | Yes | 3.79473577235772 | 3.96532187618682 | Increase | 0.527344335347187 | No |
| **compare** | No | No | Yes | 2.21443089430894 | 1.88166919863274 | Decrease | 0.0937584191124655 | No |
| **number** | No | No | Yes | 1.39436991869919 | 1.18745632358527 | Decrease | 0.227073507427783 | No |
| **posemo** | No | No | Yes | 3.23256097560975 | 3.68633497911125 | Increase | 0.119709689015342 | No |
| **anger** | No | No | Yes | 0.754491869918699 | 0.950004747436388 | Increase | 0.159060250016324 | No |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **sad** | No | No | Yes | 0.603516260162602 | 0.706215343714399 | Increase | 0.40578041007003 | No |
| **social** | No | No | Yes | 10.9010365853659 | 10.697939612609 | Decrease | 0.622387172322361 | No |
| **family** | No | No | Yes | 0.112418699186992 | 0.178838777060387 | Increase | 0.277694302228667 | No |
| **friend** | No | No | Yes | 0.273983739837398 | 0.233872958602355 | Decrease | 0.534690180445451 | No |
| **female** | No | No | Yes | 0.14010162601626 | 0.0978627041397645 | Decrease | 0.317074836444582 | No |
| **male** | No | No | Yes | 0.614979674796748 | 0.43842290163312 | Decrease | 0.0610861551897677 | No |
| **insight** | No | No | Yes | 2.42689024390244 | 2.25957462969995 | Decrease | 0.418414648455222 | No |
| **cause** | No | No | Yes | 2.34810975609756 | 2.06707368021268 | Decrease | 0.13382365785199 | No |
| **certain** | No | No | Yes | 1.04587398373984 | 1.16981105203191 | Increase | 0.389869155597423 | No |
| **percept** | No | No | Yes | 1.98888211382114 | 1.81604063805545 | Decrease | 0.354823610318167 | No |
| **see** | No | No | Yes | 0.780264227642276 | 0.801869540448163 | Increase | 0.857191773803658 | No |
| **hear** | No | No | Yes | 0.616382113821138 | 0.528518799848084 | Decrease | 0.398005809974584 | No |
| **feel** | No | No | Yes | 0.489268292682927 | 0.374803456133689 | Decrease | 0.18230525657236 | No |
| **bio** | No | No | Yes | 1.44599593495935 | 1.35602734523358 | Decrease | 0.608937844415694 | No |
| **body** | No | No | Yes | 0.447479674796748 | 0.473587162932019 | Increase | 0.790334303760869 | No |
| **health** | No | No | Yes | 0.406483739837398 | 0.273208317508546 | Decrease | 0.0784048460008932 | No |
| **sexual** | No | No | Yes | 0.211666666666667 | 0.228588112419294 | Increase | 0.827547113796424 | No |
| **ingest** | No | No | Yes | 0.412256097560975 | 0.354343904291683 | Decrease | 0.523338067219453 | No |
| **achieve** | No | No | Yes | 1.54540650406504 | 1.60031143182682 | Increase | 0.765826914261544 | No |
| **power** | No | No | Yes | 2.84778455284553 | 3.11892992783894 | Increase | 0.253894818880014 | No |
| **reward** | No | No | Yes | 1.59475609756097 | 1.66603778959362 | Increase | 0.693498877923222 | No |
| **focuspast** | No | No | Yes | 2.7735569105691 | 2.46093619445498 | Decrease | 0.135636384510475 | No |
| **focusfuture** | No | No | Yes | 1.23184959349593 | 1.09353778959362 | Decrease | 0.33288575228071 | No |
| **relativ** | No | No | Yes | 10.3491260162602 | 10.1513074439801 | Decrease | 0.640755415758571 | No |
| **motion** | No | No | Yes | 1.43758130081301 | 1.50923091530574 | Increase | 0.654198810562757 | No |
| **space** | No | No | Yes | 5.33361788617887 | 5.08347037599692 | Decrease | 0.382769428888127 | No |
| **time** | No | No | Yes | 3.54894308943089 | 3.70995822255979 | Increase | 0.53547960791185 | No |
| **work** | No | No | Yes | 3.62926829268293 | 3.63754557538926 | Increase | 0.974521081709406 | No |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **home** | No | No | Yes | 0.355386178861789 | 0.29057349031523 | Decrease | 0.367702510338057 | No |
| **money** | No | No | Yes | 2.5700406504065 | 2.49095233573869 | Decrease | 0.714417915641713 | No |
| **relig** | No | No | Yes | 0.19270325203252 | 0.233393467527535 | Increase | 0.569097616972014 | No |
| **death** | No | No | Yes | 0.219593495934959 | 0.192353778959362 | Decrease | 0.701870456347191 | No |
| **swear** | No | No | Yes | 0.54150406504065 | 0.441621724268896 | Decrease | 0.362855153340719 | No |
| **assent** | No | No | Yes | 0.830284552845529 | 0.687382263577673 | Decrease | 0.207593389113874 | No |
| **nonflu** | No | No | No | 0.402073170731707 | 0.236315039878466 | Decrease | 0.143293014087421 | No |
| **filler** | Yes | No | Yes | 0 | 0.013929927838967 | Increase | 0.397434552726709 | No |
| **AllPunc** | No | No | Yes | 19.0785162601626 | 25.4801072920623 | Increase | 0.367863492105563 | No |
| **Period** | No | No | Yes | 5.23156504065041 | 5.79783327003417 | Increase | 0.25817210044647 | No |
| **Comma** | No | No | Yes | 2.24071138211382 | 2.09134827193315 | Decrease | 0.484346465879738 | No |
| **SemiC** | No | No | Yes | 0.0508130081300813 | 0.117939612609191 | Increase | 0.33303369775482 | No |
| **QMark** | No | No | Yes | 6.65717479674798 | 9.19116027345223 | Increase | 0.713656300656557 | No |
| **Dash** | No | No | Yes | 0.609329268292683 | 1.06663311811622 | Increase | 0.175571779465134 | No |
| **Quote** | Yes | No | Yes | 0 | 0.00539783516900874 | Increase | 0.562418670915346 | No |
| **Parenth** | No | No | Yes | 0.363373983739837 | 0.415315229775921 | Increase | 0.689083697548909 | No |

**R Scripts**
**Histogram**

```
#time series graph
library(ggplot2)
library(scales)
library(lubridate)
setwd("~/Desktop/Research/Sentiment Analysis")
tw<-read.csv("AskJPM_cleaned.csv")
tw$date.2<-with(tw,ymd_h(paste(year,month,day,hour,sep="-")))
tw$ymd<- with(tw,ymd(paste(year,month,day,sep='-')))
tw_11<-
subset(tw,tw$month=="11"&(tw$day=="11"|tw$day=="12"|tw$day=="13"|tw$day=="
14"|tw$day=="15"|tw$day=="16"|tw$day=="17"|tw$day=="18"|tw$day=="19"|tw$day
=="20"|tw$day=="21"|tw$day=="22"|tw$day=="23"|tw$day=="24"|tw$day=="25")&t
w$year=="2013")
tw_12<-
subset(tw,tw$month=="11"&(tw$day=="12"|tw$day=="13"|tw$day=="14"|tw$day=="
15"|tw$day=="16")&tw$year=="2013")
tw_13<-subset(tw,tw$month=="11"&(tw$day=="13")&tw$year=="2013")
gra1<-ggplot(tw,aes(tw$date.2))+
  geom_histogram(stat="count")+
  scale_x_datetime(breaks=date_breaks("2 weeks"),minor_breaks=date_breaks("1
day"),labels=date_format("%y-%m-%d"))+
  theme_bw()+
  theme(axis.text.x = element_text(angle = 70, hjust = 1))
gra1+
  ggtitle("Histogram for Tweets Frequency - Three Months")+
  labs(y='Number of Tweets',x="Dates")
gra2<-ggplot(tw_11,aes(tw_11$date.2))+
      geom_histogram(stat="count")+
      scale_x_datetime(breaks=date_breaks("24 hour"),minor_breaks=date_breaks("1
hour"),labels=date_format("%m-%d%H:%M"))+
      theme_bw()+
      theme(axis.text.x = element_text(angle = 70, hjust = 1))
gra2+
  ggtitle("Histogram for Tweets Frequency - Two weeks")+
  labs(y='Number of Tweets',x="Dates")
gra3 <- ggplot(tw_12,aes(tw_12$date.2))+
  geom_histogram(stat="count")+
  scale_x_datetime(breaks=date_breaks("6 hour"),minor_breaks=date_breaks("1
hour"),labels=date_format("%m-%d%H:%M"))+
  theme_bw()+
  theme(axis.text.x = element_text(angle = 70, hjust = 1))
gra3+
  ggtitle("Histogram for Tweets Frequency - 5 days")+
  labs(y='Number of Tweets',x="Dates")
gra4 <- ggplot(tw_13,aes(tw_13$date.2))+
  geom_histogram(stat="count")+
```

````
 scale_x_datetime(breaks=date_breaks("1 hour"),minor_breaks=date_breaks("1
hour"),labels=date_format("%m-%d%H:%M"))+
 theme_bw()+
 theme(axis.text.x = element_text(angle = 70, hjust = 1))

gra4+
 ggtitle("Histogram for Tweets Frequency - 1 day")+
 labs(y='Number of Tweets',x="Dates")
````

**Data Cleaning and Sentiment Analysis in R (Using NRC lexicon)**

````
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```
## import library
```{r}
library(tidyverse)     # data manipulation & plotting
library(stringr)       # text cleaning and regular expressions
library(tidytext)      # provides additional text mining functions
library(lubridate)
library(psych)
library(dplyr)
library(textclean)
```
```{r read file}
jpm <- read_csv("AskJPM.csv")
typeof(jpm$date)
jpm$date<-as.character(jpm$date)
jpm <- jpm %>% mutate(Date=as.POSIXct(date, format = "%m/%e/%Y %R"))
jpm$response<- ifelse(jpm$Date <= as.POSIXct("2013-11-13 16:29:00"),0,1)
jpm_text<-as.data.frame(jpm$text)

```
```{r Data Clean-up}

# clean the text by removuing the hashtag
jpm_text$text_clean <- gsub("#", "", jpm_text$`jpm$text`)

jpm_text$date=jpm$Date
jpm_text$year=jpm$Year
jpm_text$month=jpm$Month
jpm_text$day=jpm$Day
jpm_text$hour=jpm$Hour
jpm_text$minutes=jpm$Minutes
jpm_text$response<- ifelse(jpm_text$date <= as.POSIXct("2013-11-13 16:29:00"),0,1)
#jpm_text<-jpm_text[,c(3,1,2,4)]

#removing the @ all together
jpm_text$text_clean <- gsub("@ ", "@", jpm_text$text_clean)
jpm_text$text_clean <- gsub('@\\S+', '', jpm_text$text_clean) # Remove Handles
````
````

```
# remove the url
jpm_text$text_clean <- gsub('http\\S+\\s*', '', jpm_text$text_clean) # Remove URLs
jpm_text$text_clean<-gsub("pic.twitter..*","",jpm_text$text_clean)

# remove non-ascii

#s<-jpm_text[4,2]
#s
#Encoding(s)<-"latin1"
#s<-iconv(s,"latin1","ASCII",sub="")
#s

library(dplyr)
jpm_text <- jpm_text %>% mutate(text_clean = iconv(text_clean, from = "latin1", to =
"ASCII")) %>% filter(!is.na(text_clean))

# remove whitespaces

jpm_text$text_clean <- gsub("^[[:space:]]*","", jpm_text$text_clean) ## Remove leading
whitespaces
jpm_text$text_clean <- gsub("[[:space:]]*$","", jpm_text$text_clean) ## Remove trailing
whitespaces

write_csv(jpm_text,"AskJPM_cleaned.csv")
```

```{r}

colnames(jpm_text)[2] <- "text"
max(which(jpm_text$response==1))
nrow(jpm_text)
before_tidy_data<- jpm_text[c(10533:11024),] %>%
  group_by(date) %>%
  unnest_tokens(word,text_clean)%>%
  ungroup()
after_tidy_data<-jpm_text[c(1:10532),] %>%
  group_by(date) %>%
  unnest_tokens(word,text_clean)%>%
  ungroup()
sentiment_before <- before_tidy_data %>%
  inner_join(get_sentiments("nrc")) %>%
  count(date,text,sentiment)%>%
  spread(sentiment, n, fill = 0)
sentiment_after<-after_tidy_data %>%
  inner_join(get_sentiments("nrc")) %>%
  count(date,text,sentiment)%>%
  spread(sentiment, n, fill = 0)
```

````
```

```{r t test}
s<-shapiro.test(sentiment_before$anger)
sentiment_before$anger
s$p.value
# data is not normal
# Mann-Whitney U test
# provided the sample size is not too small, we should not be overly concerned if the data appear to violate the normal assumption
v<-var.test(sentiment_before$anger,sentiment_after$anger)
# equality of two variances
v$p.value
res_anger<-t.test(sentiment_before$anger,sentiment_after$anger,var.equal=TRUE)
res_anger
res_anger$p.value
# no difference
e_bf<-sentiment_before[,3]
e_bf
emotion.1<-colnames(sentiment_before[3])
typeof(sentiment_before[3])
sentiment_before$anger
as.numeric(unlist(sentiment_before[3]))
```

```{r}
sentiment_before$total_negative <- rowSums(sentiment_before[,c(3,5,6,8,10)])
sentiment_before$total_positive <- rowSums(sentiment_before[,c(4,7,9,11,12)])
sentiment_after$total_negative <- rowSums(sentiment_after[,c(3,5,6,8,10)])
sentiment_after$total_positive <- rowSums(sentiment_after[,c(4,7,9,11,12)])
```

```{r for loop}
library(magicfor)
magic_for(print, silent = TRUE)

x<-c(3:14)
for (val in x) {
  emotion<-colnames(sentiment_before[val])
  e_bf<-as.numeric(unlist(sentiment_before[val]))
  e_af<-as.numeric(unlist(sentiment_after[val]))
  cat("\nThe Emotion:",emotion,"\n")
  s.before<-shapiro.test(e_bf)
  p_normal<-s.before$p
  #s.after<-shapiro.test(e_af)
  #s.after
  if (s.before$p.value < 0.05){
    cat("The distribution for",emotion,"is not normal\n")}
  v<-var.test(e_bf,e_af)
````

~~~
  p_var<-v$p.value
  if (v$p.value>0.05) {
    cat ("The variance for before/after response of",emotion,"is equal\n")
    res<-t.test(e_bf,e_af,var.equal=TRUE)
  }else{
    cat ("The variance for before/after response of",emotion,"is not equal\n")
    res<-t.test(e_bf,e_af,var.equal=FALSE)
  }
  mean_est<-res$estimate
  p_test<-res$p.value
  if (res$p.value <0.05){
    cat("The average",emotion,"before response is significantly different from after
response\n")
  }else{
    cat("The average",emotion,"before response is NOT significantly different from after
response\n")
  }

  #put(emotion,s.before$p.value,v$p.value,res$estimate,res$p.value)
  put(emotion,p_normal,p_var,mean_est,p_test)
}


```
```{r}

write_csv(sentiment_before,"before_term_freq.csv")
write_csv(sentiment_after,"after_term_freq.csv")

```

```{r}

colnames(jpm_text)[2] <- "text"

#tokenization of words into tidy dataframe
#group by id,each text is split into words in new colomn 'word'

tidy_data<- jpm_text %>%
  group_by(date) %>%
  unnest_tokens(word,text_clean)%>%
  ungroup()
#write_csv(tidy_data,'/Users/xiaotonghe/Documents/research/tw_data/tidy_data.csv')




#nrc dict
lexi<- get_sentiments('nrc')%>%filter(sentiment %in% c("positive","negative"))
~~~

````
#get sentiments for each word
abc_nrc<-tidy_data_stop%>%
  inner_join(get_sentiments("nrc"),by='word')%>%
  ungroup()

#sentiments counts
sentiments_count<-abc_nrc%>%
  filter(sentiment %in% c("positive","negative"))%>%
  group_by(sentiment)%>%
  count(sentiment)

sentiment_nrc <- tidy_data_stop %>%
  inner_join(get_sentiments("nrc")) %>%
  count(date,text,sentiment)%>%
  spread(sentiment, n, fill = 0)

```
#observation 13085 (inclusive) after are after response tweets
```{r}
#after
after<-jpm_text[c(1:13084),]
write_csv(after,"JPMafter.csv")
#before
before<-jpm_text[c(13085:13634),]
write.csv(before,"JPMbefore.csv")

```


```{r dataset with dummy variable}
jpm_text$rowID<-1:nrow(jpm_text)
jpm_text$response<-ifelse(jpm_text$rowID<=13084,1,0)
jpm_text$rowID<-NULL
write_csv(jpm_text,"AskJPM_Jocelyn.csv")
```
````

**Statistical test on LIWC**

````
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```


```{r}
library(tidyverse)     # data manipulation & plotting
library(stringr)       # text cleaning and regular expressions
library(tidytext)      # provides additional text mining functions
library(lubridate)
library(psych)
library(dplyr)
````

````
library(textclean)
```
```{r}
liwc <- read_csv("LIWC.csv")
before <- subset(liwc,liwc$C == 0)
after <- subset(liwc,liwc$C == 1)
describe(liwc)
```
```{r}
library(magicfor)
magic_for(print, silent = TRUE)
x<-c(34:38)
for (val in x) {
  emotion<-colnames(before[val])
  e_bf<-as.numeric(unlist(before[val]))
  e_af<-as.numeric(unlist(after[val]))
  cat("\nThe Emotion:",emotion,"\n")
  s.before<-shapiro.test(e_bf)
  p_normal<-s.before$p
  #s.after<-shapiro.test(e_af)
  #s.after
  if (s.before$p.value < 0.05){
    cat("The distribution for",emotion,"before the res is not normal\n")}
  v<-var.test(e_bf,e_af)
  p_var<-v$p.value
  if (v$p.value>0.05) {
    cat ("The variance for before/after response of",emotion,"is equal\n")
    res<-t.test(e_bf,e_af,var.equal=TRUE)
  }else{
    cat ("The variance for before/after response of",emotion,"is not equal\n")
    res<-t.test(e_bf,e_af,var.equal=FALSE)
  }
  mean_est<-res$estimate
  p_test<-res$p.value
  if (res$p.value <0.05){
    cat("The average",emotion,"before response is significantly different from after
response\n")
  }else{
    cat("The average",emotion,"before response is NOT significantly different from after
response\n")
  }
  #put(emotion,s.before$p.value,v$p.value,res$estimate,res$p.value)
  put(emotion,p_normal,p_var,mean_est,p_test)
}

```
````

Python codes for LIWC Results

```
import pandas as pd
```
```

```
from scipy import stats

liwc = pd.read_csv('LIWC.csv',sep=',')

liwc.describe()

pd.set_option('display.max_columns', None)

liwc.head()

after= liwc[liwc.C==1]

after.shape

before= liwc[liwc.C==0]

before.shape

before_emo = before.iloc[:,3:]

before_emo.head()

after_emo = after.iloc[:,3:]

testresult=[]

columnnames =
['Sentiments','Before_Normality','After_Normality','Equal_Variance','Before_Mean','After
_Mean','Diff','T-test_P','Significant']

for col in before_emo: # for each emotion

  tempresult=[col] # get the name of the emo

  # perform normality test for before

  before_nol_p = stats.shapiro(before[col])[1]

  if before_nol_p  <= 0.05:

    tempresult.append('No')

  else:

    tempresult.append('Yes')
```

```
# perform normality test for after
after_nol_p  = stats.shapiro(after[col])[1]
if after_nol_p <= 0.05:
    tempresult.append('No')
else:
    tempresult.append('Yes')

# perform variance test
var_test_p = stats.levene(before[col],after[col]).pvalue
if var_test_p > 0.05: # can not reject the null
    tempresult.append('Yes')
else:
    tempresult.append('No')
bef_mean = before[col].mean()
aft_mean = after[col].mean()
tempresult.append(bef_mean)
tempresult.append(aft_mean)

if bef_mean > aft_mean:
    tempresult.append('Decrease')
elif bef_mean == aft_mean:
    tempresult.append('Same')
else:
    tempresult.append('Increase')
```

```
    # perform t test
    if var_test_p > 0.05: # equal variance
        t_test_p = stats.ttest_ind(before[col],after[col])[1] # get the p value
        tempresult.append(t_test_p)
        if t_test_p <= 0.05: # significantly different
            tempresult.append('Yes')
        else:
            tempresult.append('No')
    else: # not equal variance
        t_test_p = stats.ttest_ind(before[col],after[col],equal_var=False)[1] # get the p value
        tempresult.append(t_test_p)
        if t_test_p <= 0.05: # significantly different
            tempresult.append('Yes')
        else:
            tempresult.append('No')
    testresult.append(tempresult)
    # perform the equal variance test
df = pd.DataFrame(testresult, columns = columnnames)
df
df.to_csv("LIWC_t_test.csv",index=False)
```