Changes of Donald Trump before and after knowing campaigning against Hillary Clinton

Fei Shi

12/18/2017

Introduction

Donald John Trump is the 45th and current President of the United States, in office since January 20, 2017. As it is well known that before entering politics, he was a businessman and television personality. And after becoming the presumptive Republican nominee, Donald John Trump shifted his focus to the general election. After June 6, 2016, Hillary Clinton became the presumptive Democratic nominee, and after that Trump began campaigning against Hillary Clinton, as we know Hillary Clinton is most likely to be President at that time, we want to know what changes of Donald Trump before and after knowing campaigning against Hillary Clinton , is he some feared or disappointed? In this project, we will analysis this topic using texting mining techniques.

Material and Methods

Data

The data used in this project are tweets from Donald John Trump mainly in year 2016, the range is from 2015-12-14 to 2016-08-08, the source of the data comes from tweets website and can be easily downloaded from 'http://varianceexplained.org/files/trump_tweets_df.rda'. There are 1512 with 16 variables in the original data, but for safety, variables such as location information are dropped, we only use variables 'id', 'created', 'text' in our analysis, 'id' is the uniqute identifier for each tweet, 'created' is the time line while 'text' is the content of the tweet.

Tools

We will use the following R packages and make reference from the text book https://www.tidytextmining.com/ as well as github https://github.com/dgrtwo/dgrtwo.github.com/blob/master/_R/.

Results

Time series plot

First, we create a new variable called 'Status' which contains two categories, the first one is 'Before' which means the tweet created time before 2016-06-06 the day Hillary Clinton became the presumptive Democratic nominee, and the Status 'After' the time after that day, and then we plot the number of tweets vs time hours seperately for the two status:
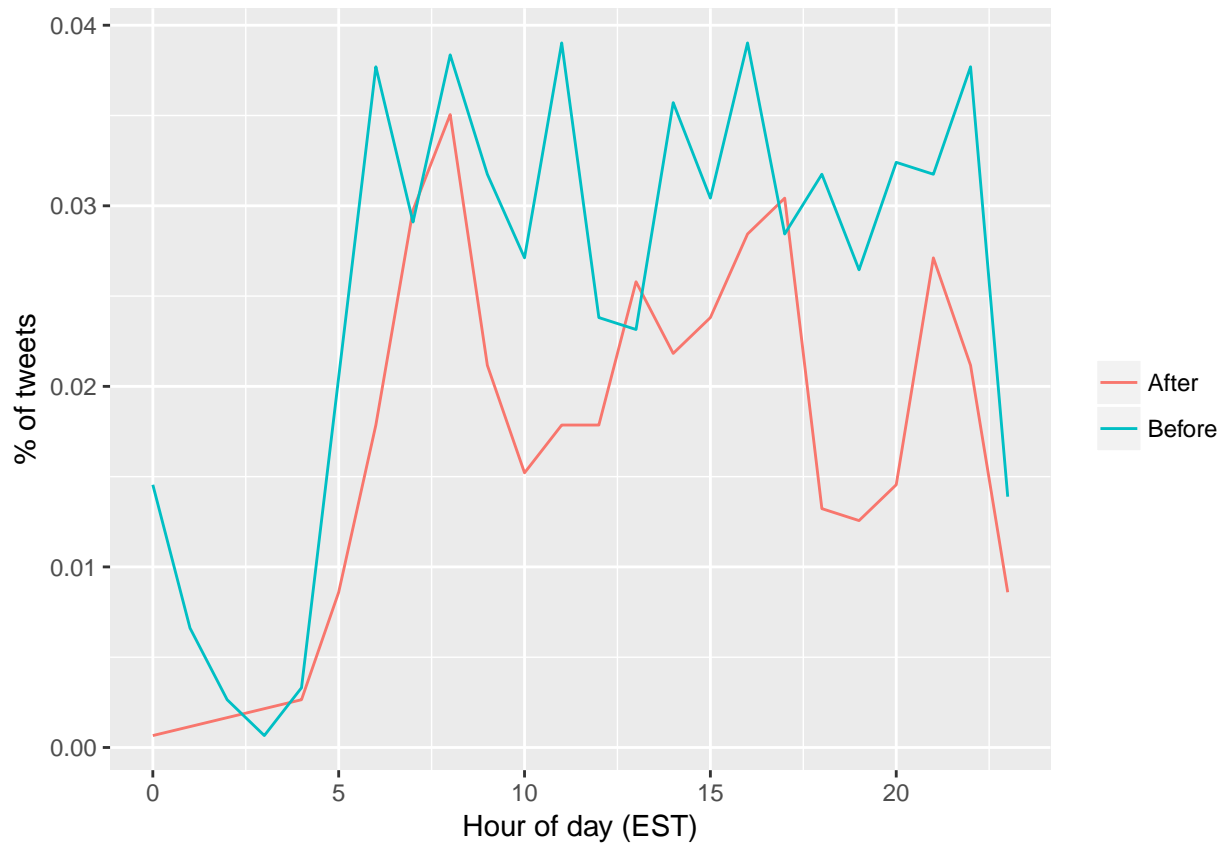
So it can be found that after Trump knowing campaigning against Hillary Clinton, it appears that he becomes much more busy while having much less time to post tweets, the numbers of tweets are obviously lower than before knowing campaigning against Hillary Clinton, it seems like Trump become serious about the campaign.
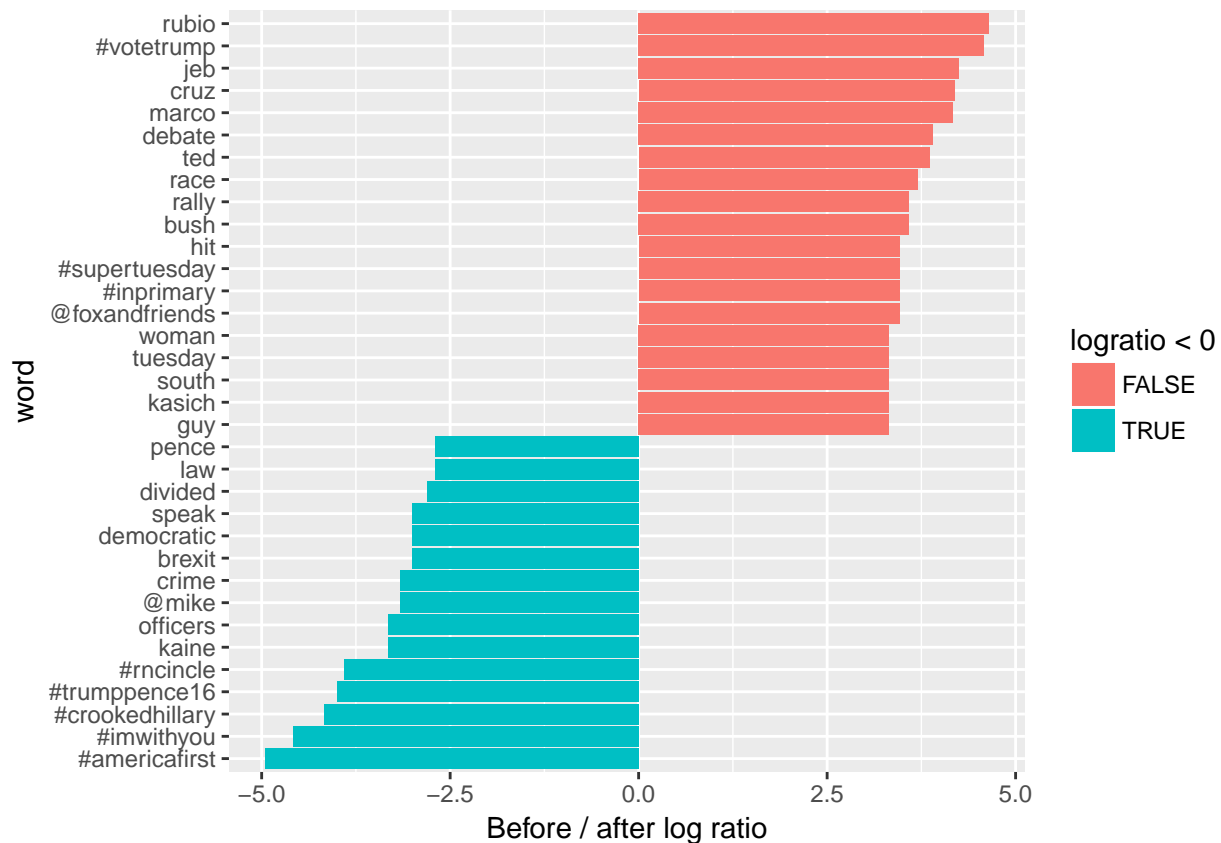
Words used in the two periods

Like the reference did, we remove tweets which are likely not be posted by Trump himself and show most common words used:
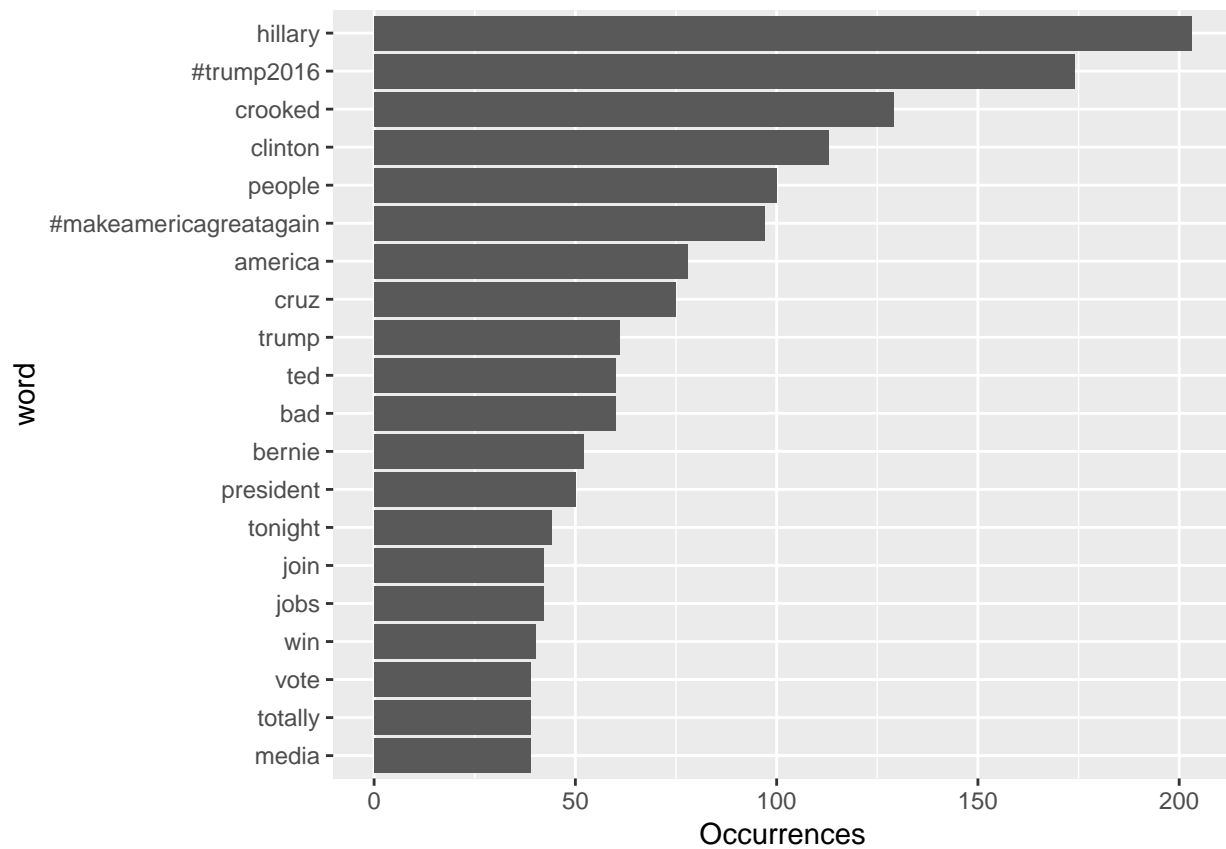
Obviously, he care more about hillary cliton than himself, also, he cares about jobs, people, media as well as vote and president and etc.

Next, we show the words most different from two status:

The blue one stands for before campaigning against Hillary Clinton, it appears hard to understand the words, but the red one for after campaigning against Hillary Clinton, it can be seen that he cares about 'vote trump' and 'debate' as well as 'woman'. To see more clearly, we use wordclouds to show the maps of the words in the two periods:

```
# A tibble: 6 x 3
            word         After       Before
           <chr>         <dbl>        <dbl>
1       @karlrove 0.0002446782 0.000180538
2         @reince 0.0002446782 0.000180538
3 @trumpturnberry 0.0002446782 0.000180538
4        @usatoday 0.0002446782 0.000180538
5          accept 0.0002446782 0.000180538
6          acting 0.0002446782 0.000180538
```

We can find that before, Trump is selfish, the world Trump is very large, but the two words hillary, clinton are larger after 2016-06-06, so Trump really talks more about hillary clinton than the period before. And next, we use the following plot to show the shared words and different words across the two periods:

As the words in the line are the common words in the two periods, we can find whether or not before and

after compaigning against hillary clinton, Trump cares about job and people also voters as the three words are in the line, also, Trump always like joke, so joke is shown in the line too, but we can find something different, on the top of the tail of the line, we can find hillary clinton is in the side of after, it means Trump indeed talks about hillary clinton more ofthen.

Sentiment analysis

Finally, in this section, we will use Sentiment analysis to compare the sentiments including anger, disgust, fear, joy, negative, sadness and etc as follows:

```
# A tibble: 6 x 2
       word sentiment
      <chr>     <chr>
1    abacus     trust
2   abandon      fear
3   abandon  negative
4   abandon   sadness
5 abandoned     anger
6 abandoned      fear

# A tibble: 6 x 4
  Status     sentiment total_words words
   <chr>         <chr>       <int> <dbl>
1  After         anger        4087   237
2  After  anticipation        4087   178
3  After       disgust        4087   154
4  After          fear        4087   228
5  After           joy        4087   145
6  After      negative        4087   397
```

After obtained the sentiments, we use Exact Poisson tests to test whether the ratio between two rate parameters are equal or not as well as compute confidence intervals:

```
# A tibble: 10 x 9
# Groups:    sentiment [10]
     sentiment  estimate statistic        p.value parameter   conf.low
         <chr>     <dbl>     <dbl>          <dbl>     <dbl>      <dbl>
 1        anger 1.0462530       237 0.6030053621  230.9711 0.8793936
 2 anticipation 0.8494316       178 0.0902722724  196.1556 0.7002924
 3      disgust 1.1467693       154 0.2246865697  142.6586 0.9192195
 4         fear 1.3794741       228 0.0007125726  191.9098 1.1421298
 5          joy 0.8470455       145 0.1183340839  160.0664 0.6835166
 6     negative 1.0447443       397 0.5245329927  387.2163 0.9142375
 7     positive 1.0446415       454 0.4908421750  442.8362 0.9222420
 8      sadness 1.0344132       216 0.7172772078  211.8651 0.8624902
 9     surprise 0.7850214       106 0.0494955052  122.7034 0.6120579
10        trust 1.1464875       313 0.0751209393  289.9876 0.9831102
# ... with 3 more variables: conf.high <dbl>, method <fctr>,
#   alternative <fctr>
```
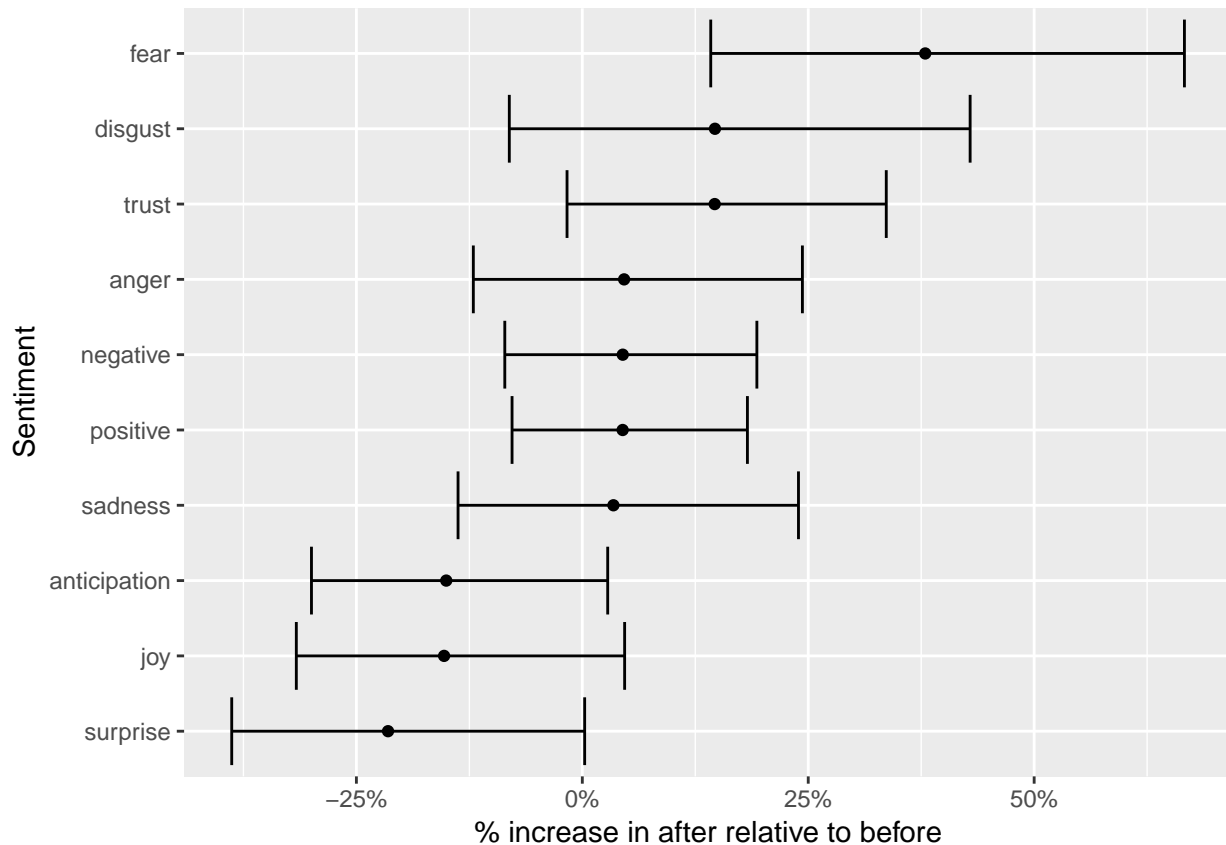
From the plots, it can be found fear, anger, negative, disgust, sadness are the main sentiments after campaigning against Hillary Clinton as there are different levels percentages increase in after status than before, actually, these results reflect the sentiment of Trump changed from 'joy' before campaigning against Hillary Clinton, on one hand, Trump might fear failture, on the the hand, he defame Hillary Clinton. However, he sometimes feel anger and sadness about rumors from Hillary Clinton. After all, it is USA president election, Trump really change before and after knowing campaigning against Hillary Clinton, besides, in future analysis, with more tweets data, we can compare the changes between before and after Trump becomes USA president.