

# REPORT PROBLEM # 5

*Prepared By:*

*Anuj Rastogi*

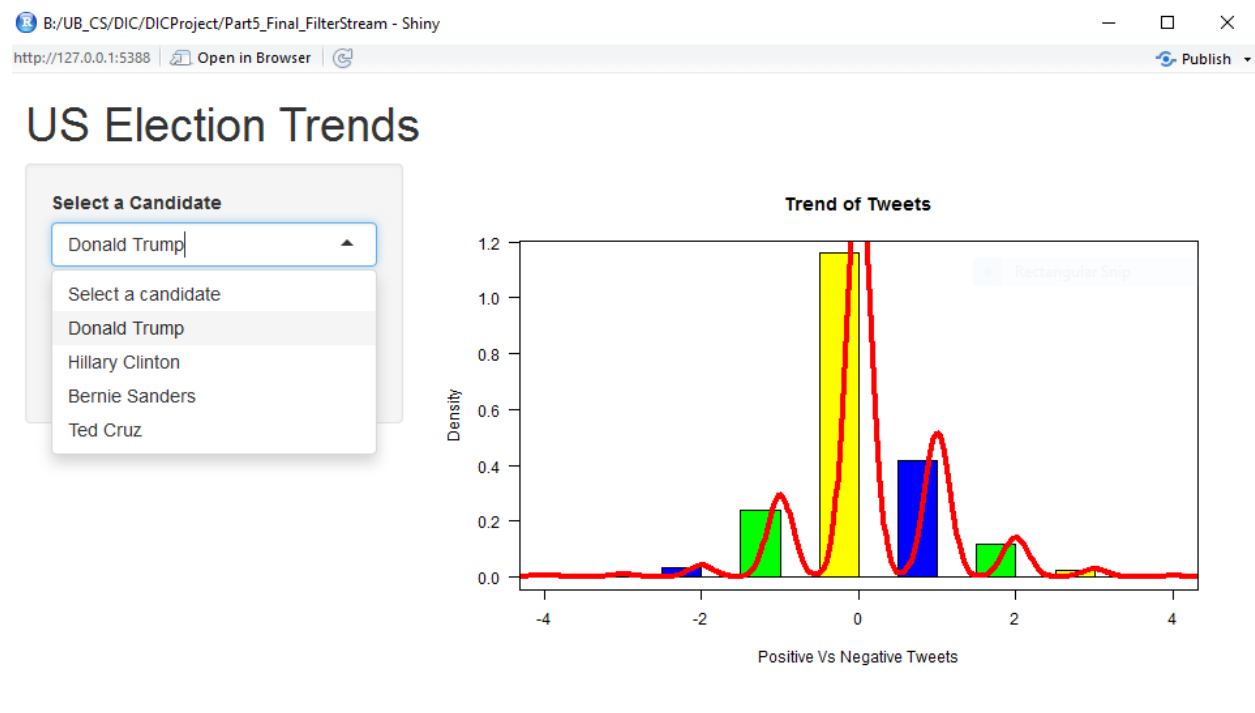
[anujrast@buffalo.edu](mailto:anujrast@buffalo.edu)

*PersonNumber# 50134324*

## **ANALYSIS OF TWITTER DATA FOR THE WEEK**

In this part of the project we have made a UI that based upon the election candidates in USA displays the trends of their popularity among people. For this we had to do sentiment analysis of the tweets and their nature if they were positive or negative.

The following is an image of the website developed:



To analyze this trend I used a 'Sentiment\_new.R' that, I downloaded from '<https://github.com/mjhea0/twitter-sentiment-analysis>'. Now this file basically compares the tweets against a list of positive and negative words. Based on that, this separates the tweets into positive and negative tweets. The following R script is used to figure out the trend of tweets. This also forms the basis of our server side coding.

```

if(candidateName == 'Donald Trump' & radioSelect == 'Weekly Trend'){

  a = readLines('DonaldTrump.json',warn = FALSE)
  x = jsonlite::fromJSON(a)
  pos = scan('positive-words.txt', what = 'character', comment.char = ';')
  neg = scan('negative-words.txt', what = 'character', comment.char = ';')

  #This file sentiment_new.R along with the list of positive and negative words has been taken from url: https://github.com/mihea0/twitter-sentiment-source('sentiment_new.R')
  y = x$text
  cleaned = iconv(y, 'UTF-8','ASCII')
  analysis = score.sentiment(cleaned, pos.words = pos, neg.words = neg)
  hist(analysis$score, xlab = 'Positive Vs Negative Tweets', las = 1, main = 'Trend of Tweets' , col = c('blue','yellow','green'), probability = T)
  lines(density(analysis$score), col = 2, lwd = 4)
  box()

}

```

```

if(candidateName == 'Donald Trump' & radioSelect == 'Now'){
  load('my_oauth.Rdata')
  #serching Twitter
  filterStream(file.name = "Donald.json", # Save tweets in a json file
               track = c("Donald Trump","president","election"), # Collect tweets mentioning either Affordable Care Act, ACA, or Obamacare
               language = "en",
               timeout = 10, # Keep connection alive for 60 seconds
               oauth = my_oauth) # Use my_oauth file as the OAuth credentials

  tweets_df <- parseTweets('Donald.json', simplify = FALSE)
  pos <- scan('positive-words.txt', what = 'character', comment.char = ';')
  neg <- scan('negative-words.txt', what = 'character', comment.char = ';')
  y = tweets_df$text
  cleaned = iconv(y,'UTF-8','ASCII')
  analysis = score.sentiment(cleaned, pos.words = pos, neg.words = neg)

  #This file sentiment_new.R along with the list of positive and negative words has been taken from url: https://github.com/mihea0/twitter-sentiment-source('sentiment_new.R')
  hist(analysis$score, xlab = 'Positive Vs Negative Tweets', las = 1, main = 'Trend of Tweets' , col = c('blue','yellow','green'), probability = T)
  lines(density(analysis$score), col = 2, lwd = 4)
  box()
}

```

The above is the server side code. For weekly analysis we already have tweets that will be used to identify who is getting the most negative feedback. But for current trend 'streamR' package of twitter will be used to pull in the latest data from twitter Stream API. This has been done above by the use of '**filterStream()**' API.

This filterStream() API is used to fetch the current tweets from twitter. After parsing them into JSON format the text field can be extracted from them and hence, they can be used for analyzing if they are positive or negative in nature.

From there a graph is plotted that will show the count and the probability of tweets as positive or negative. Also, the above code is repeated for all the other four candidates.

### EXPLANATION OF THE CURVE:

The curve is a histogram of probability. It has in X-axis, two divisions based on the position of 0. If the bar is on the right side of 0 that means that the bar represents a set of positive tweets with the score of tweets representing the value at which the bar itself stands. The more the score of tweets the more positive they are.

Similarly, the lesser the score of tweets from zero the more negative they are. If the tweet bar stands at 0, that means that people related to those tweets are either indifferent or neutral. Hence, they will not be counted towards winning or losing trend.

Also, there is a line that is the probability curve. The more the curve bulges towards the positive side the better it is for the candidate. Hence, after all considerations we can say that the trend will be calculated as

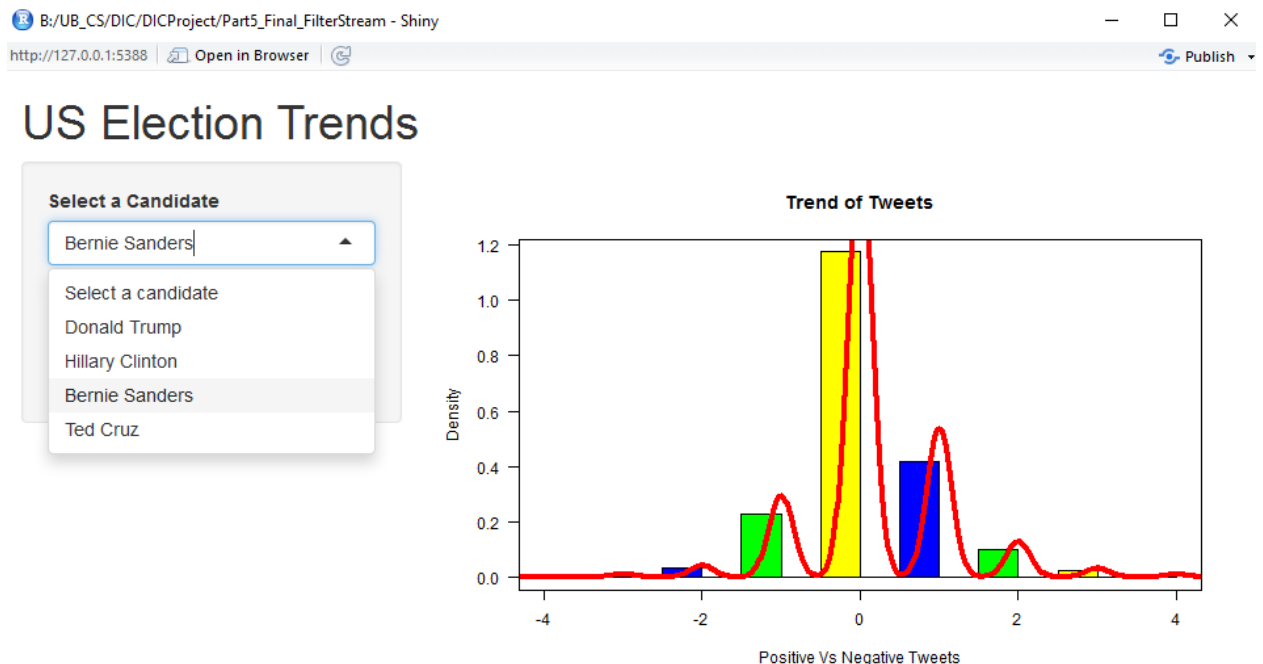
$$\text{TREND} = \text{Number of positive tweets} - \text{Number of negative tweets} \quad (1)$$

The more the value of trend the better it is for candidates. Now let us analyze the real weekly data obtained for our four election candidates.

### CLOSE ANALYSIS OF WEEKLY TRENDS

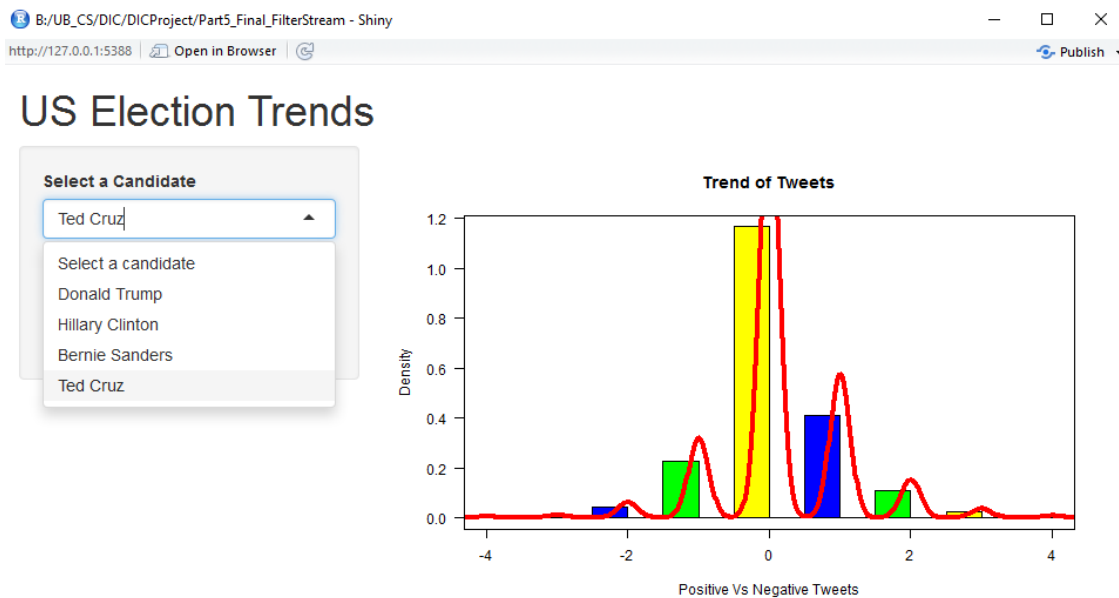
#### Weekly Trends Plots:

- **Bernie Sanders:**



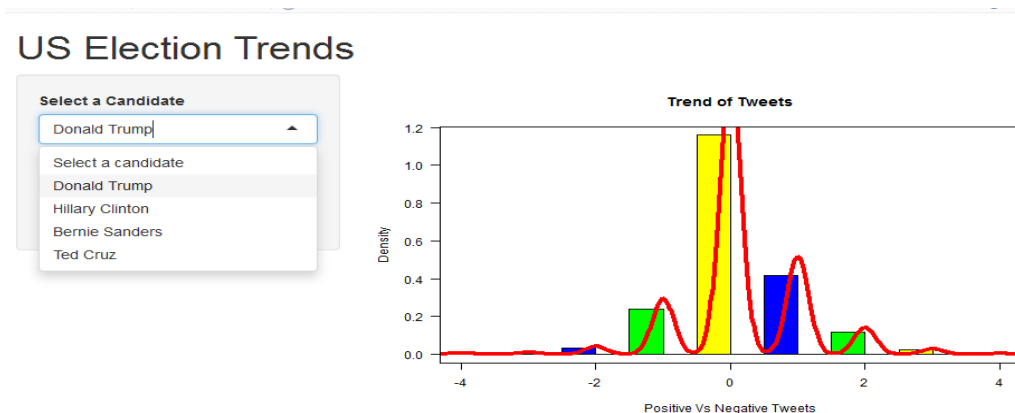
The graph has a lot of people who are confused about his candidacy, thus a huge graph at zero represents their tweets. Now overall, the blue graph at right side of 0 seems taller than the negative graph on left side of 0. Hence, for Bernie Sanders, the trend seems positive at least for the last week.

- **Ted Cruz :**



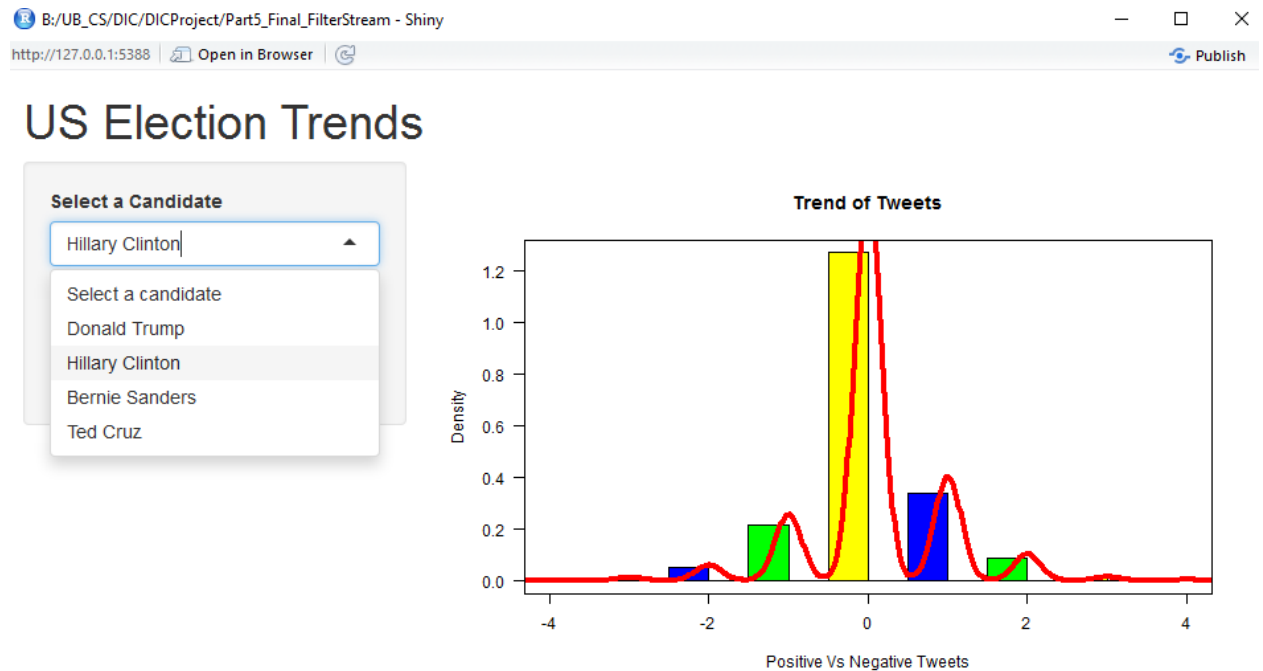
For Ted Cruz, as well the trend for the last week represents a positive vote. Although according to (1) the larger the value of trends the better it is for the candidate. Hence, between Bernie Sanders and Ted Cruz we can from there decide who has an upper hand.

- **Donald Trump**



The same is the case for Donald Trump. However, the number of neutral tweets for Donald Trump remains lesser than the above two.

- **Hillary Clinton**



For Hillary Clinton, as well the same situation persists. Thus, from the above graphs we have at least got an idea on how to analyze the trends based on the understanding of the graph.

### **SUMMARY**

*From the analysis of the week almost all the candidates for now have an equal trend for winning. The graph not only divides the tweets based upon strong and weak sentiments but also gives a chance to know how many among the followers of a candidate are strong and how many are weak supporters. Thus, it provides an extended analysis of public sentiments and also, provides us with a probability curve whose peak will mold in the direction of trend.*