Russian interference in the U.S. 2016 presidential elections

Alexandra DeKinder, Ameya Karnad, Kulkanya Lekhyananda and Nitasha Nair

January 30, 2019

Background

In January 2017, major US agencies such as the Central Intelligence Agency(CIA), Federal Bureau of Investigations (FBI), and the National Security Agency(NSA) reported that there were evidences that the Russian government had conducted a campaign to influence the 2016 Presidential election. The campaign was aimed at damaging Hillary Clinton's presidential campaign, in the process, undermining the US democratic process.

According to the report, it was found that the tweets were originating from the Russian "Trolls Factory" called the Internet Research Agency. In this assignment, we analyse the 3 million trolls tweets dataset obtained by Clemson University to explore the strategy used by the trolls.

Hypothesis

A strategy utilized by the trolls was to divide the democratic party supporters by expressing support for Bernie Sanders and deriding Hillary Clinton in an effort to drive low voter turnouts overall and encouraging them to vote against Clinton. At the same time, tweets were rallying the right wing votes.

We will analyze the tweets around a specific event, the Democratic National Convention (DNC), held between July 25, 2016 - July 28, 2016. The DNC is a series of presidential nomination conventions held by the Democratic Party. The main purpose of the convention is to confirm the nomination of the president and vice president candidates. In 2016, DNC had two main candidates competed for the presidential nomination - Hillary Clinton and Bernie Sanders. At the same time, in june and july 2016, hackers released emails by DNC members suggesting that the party leadership had undermined Bernie Sanders campaign and in turn favouring the Clinton campaign.

Research Questions

1. What behaviour do the left leaning and right leaning trolls exhibit during the period of observation? Is it different from the general trend ?
2. Comparing words and hashtags used by left and right leaning trolls around the same topic
3. Can we identify certain characteristics of a troll

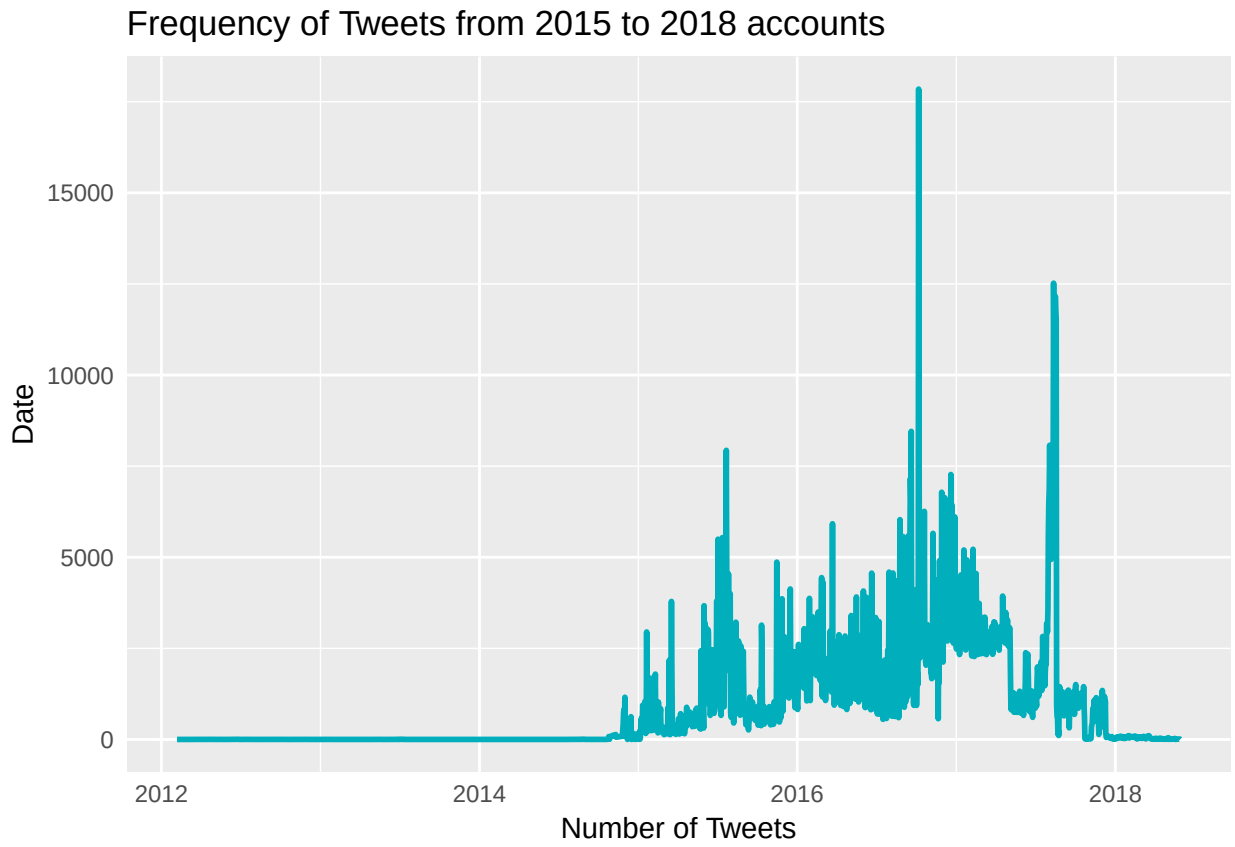
Data

We first conducted general data analysis on the whole set of tweets to draw out general trends. We further analyzed tweets from June 1 2016 to August 31 2016 to validate our hypothesis and find other insights that might arise out of the data. Due to the focus of our analysis being on how these tweets affect sentiment during the election, we have used a subset tweets that are in English

Findings

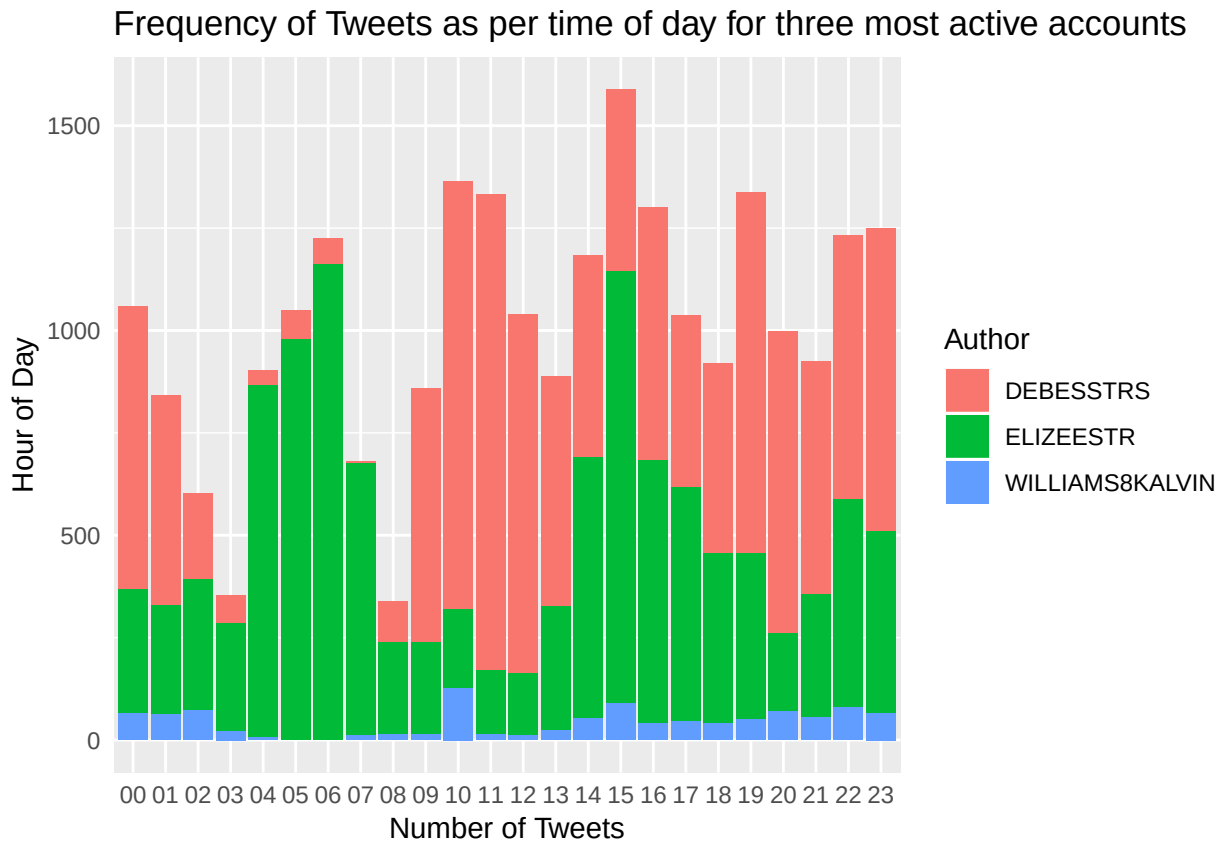
1 - Frequency of tweets

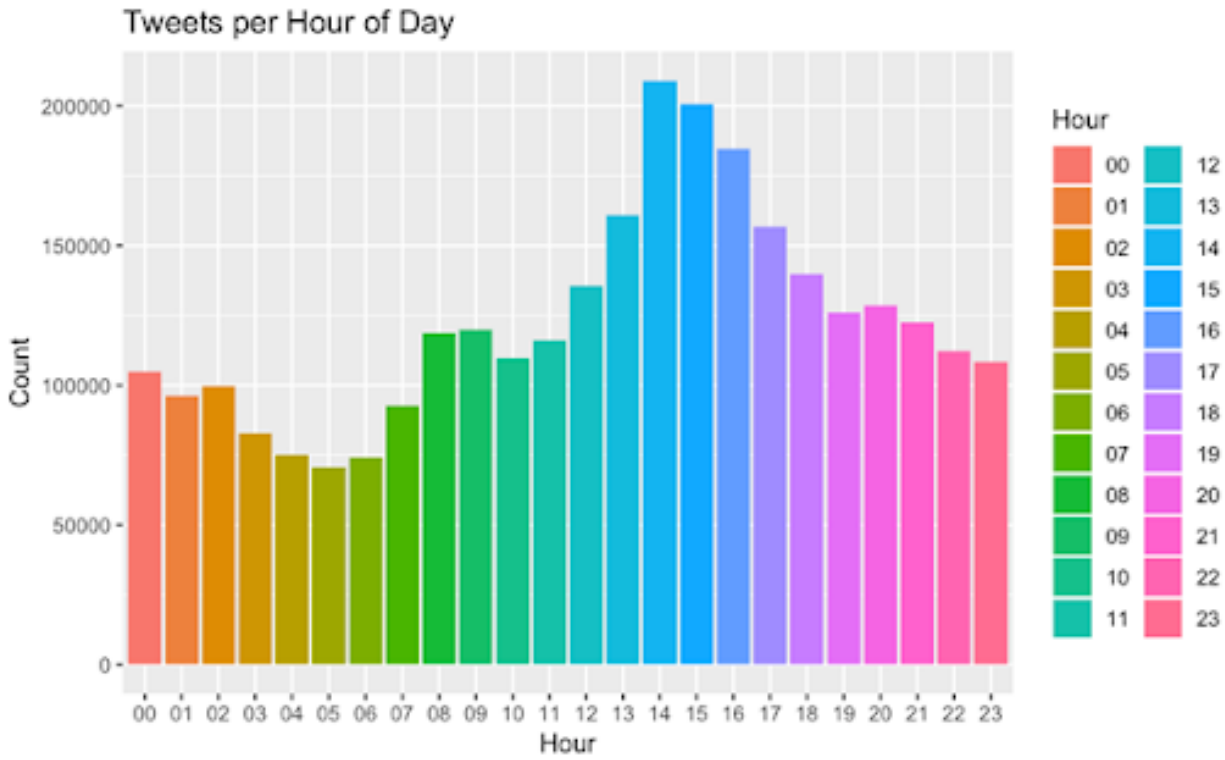
When the whole set is considered, we see that the frequency of tweets is higher between 2015 and 2018, with it peaking just before the end of 2016 around the elections. Interestingly, the tweets don't subside post elections, but continue till the end of 2017.



2 Timing of tweets

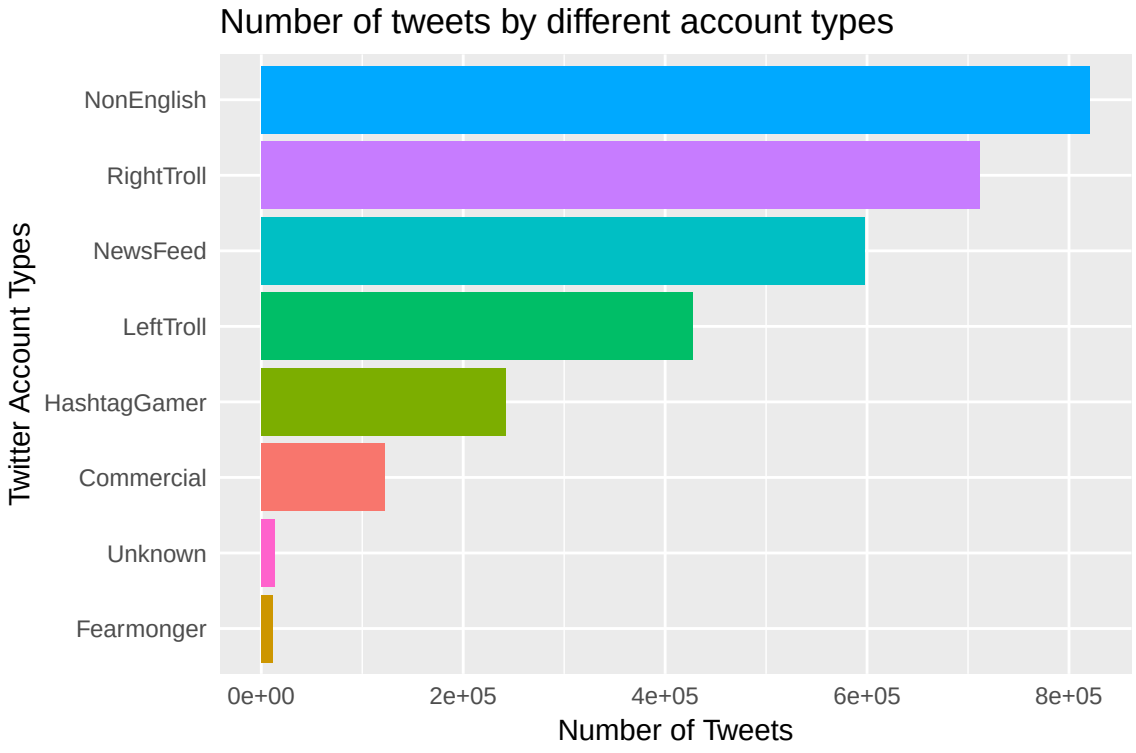
As the figures above show, there is more activity in the afternoon and early-evening hours.The exact cause of the heightened activity in the afternoon/evening would need further analysis; however, it is clear that these times are more active.

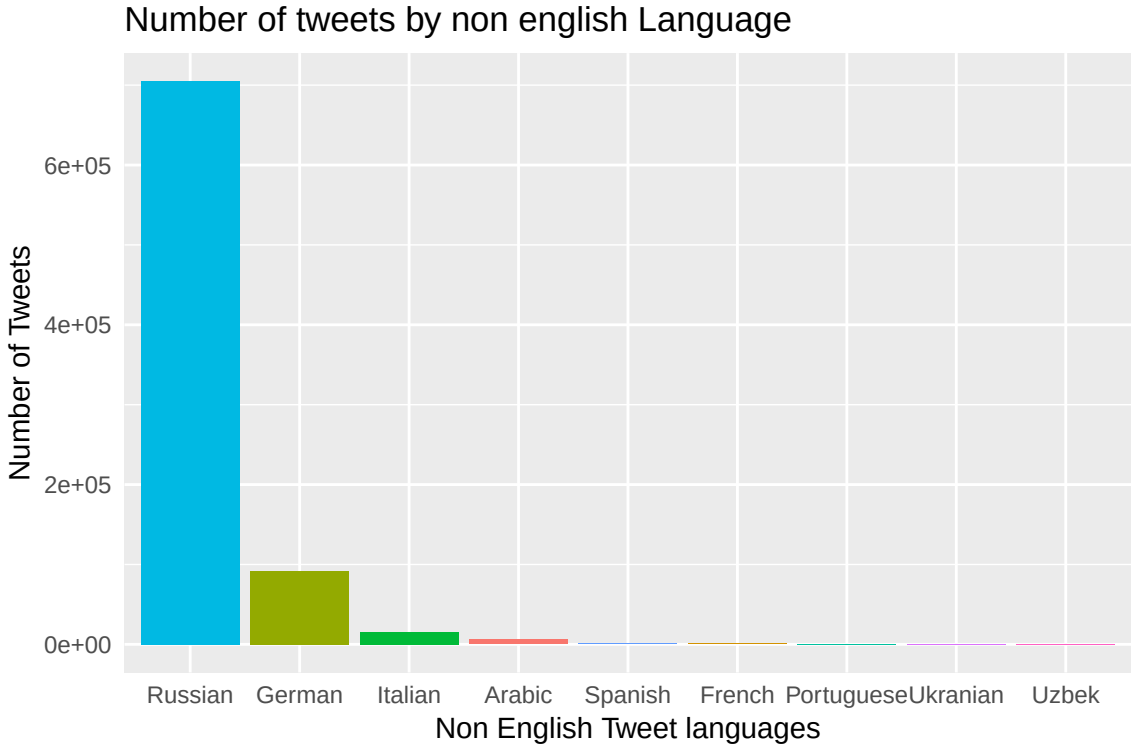

Frequency of Tweets as per time of day for three most active accounts
Hour of Day
1500
1000
500
0
00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Number of Tweets
Author
DEBESSTRS
ELIZEESTR
WILLIAMS8KALVIN


Tweets per Hour of Day
Count
200000
150000
100000
50000
0
00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour
Hour
00 01 02 03 04 05 06 07 08 09 10 11
12 13 14 15 16 17 18 19 20 21 22 23

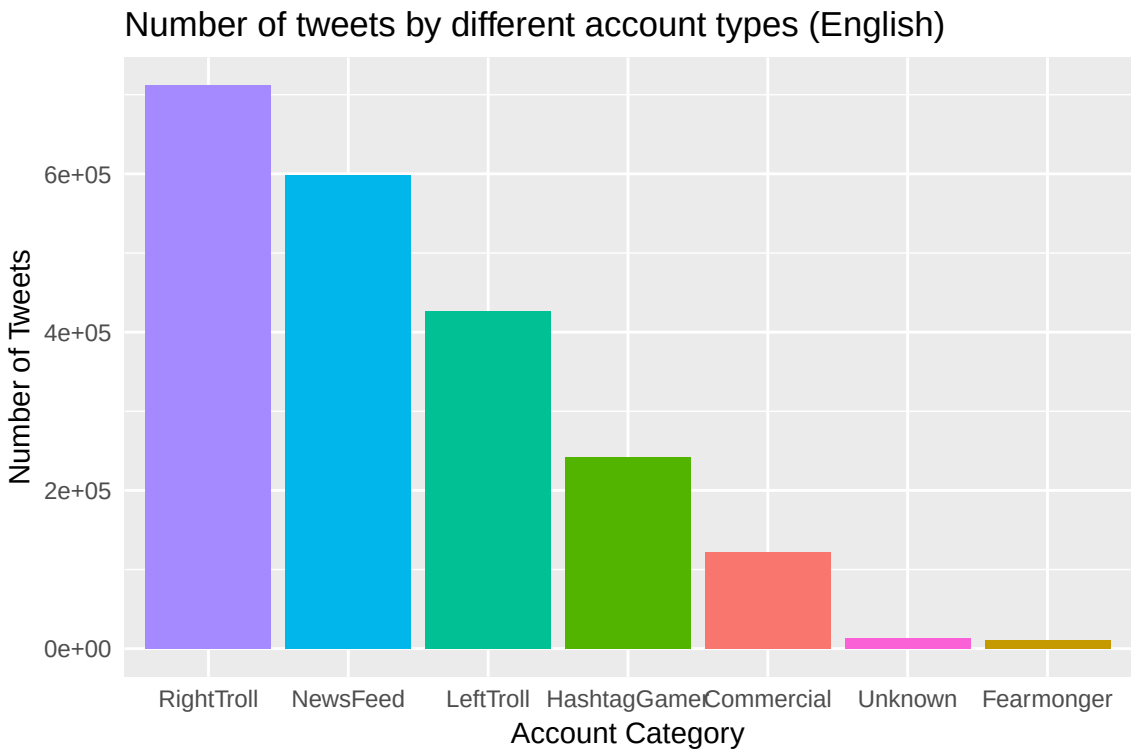
3 Tweets according to account types

The tweets were in 56 different languages, out of which 71.9% of tweets were in English and the tweets came from 37 different regions. The number of non-english tweets was substantially higher than english tweets, with more right leaning trolls than left leaning trolls.





If we analyze the non-english tweets further, we see that russian tweets dominate followed by German, Italian, Arabic and Spanish.

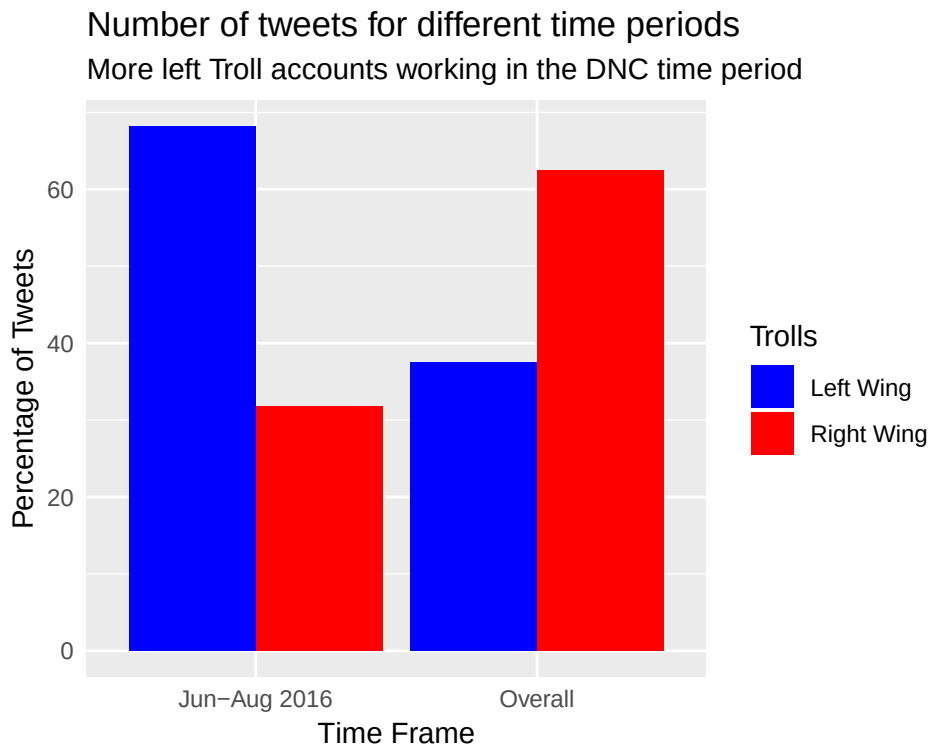


Most of the Tweets are Right Wing/ Left wing trolls, or are news feeds.

4. Analysis During DNC (tweets from the period 1st of June 2016 to 31st of August 2016):

4.1. Frequency of tweets:

We look at right and left wing trolls and concentrate around the period when DNC emails were leaked and the mood of right and left wing trolls. We hypothesize that the Russian trolls were trying to divide the democratic party by cashing in on the email leaks and trying to display negative sentiment on Hillary Clinton, while trying to play up Bernie Sanders Supporters by encouraging them to vote against Hillary.

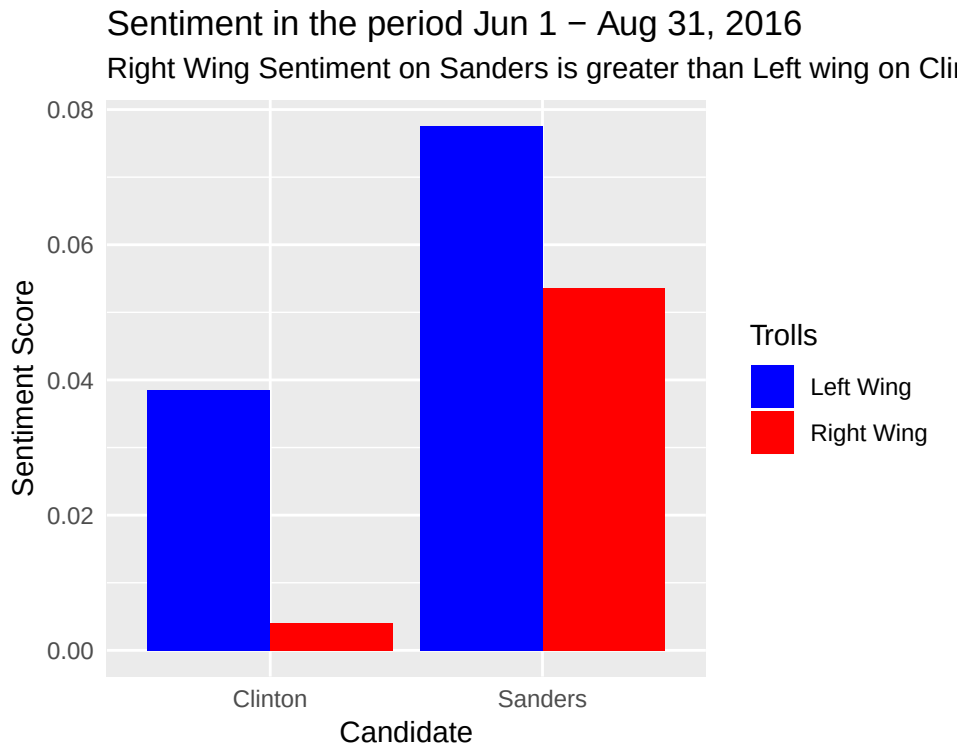


An interesting insight is that though the overall data shows that the Right Troll tweets were about twice of Left Wing trolls, the left troll accounts were more active in the period of June- August 2016

4.2. Sentiment Analysis:

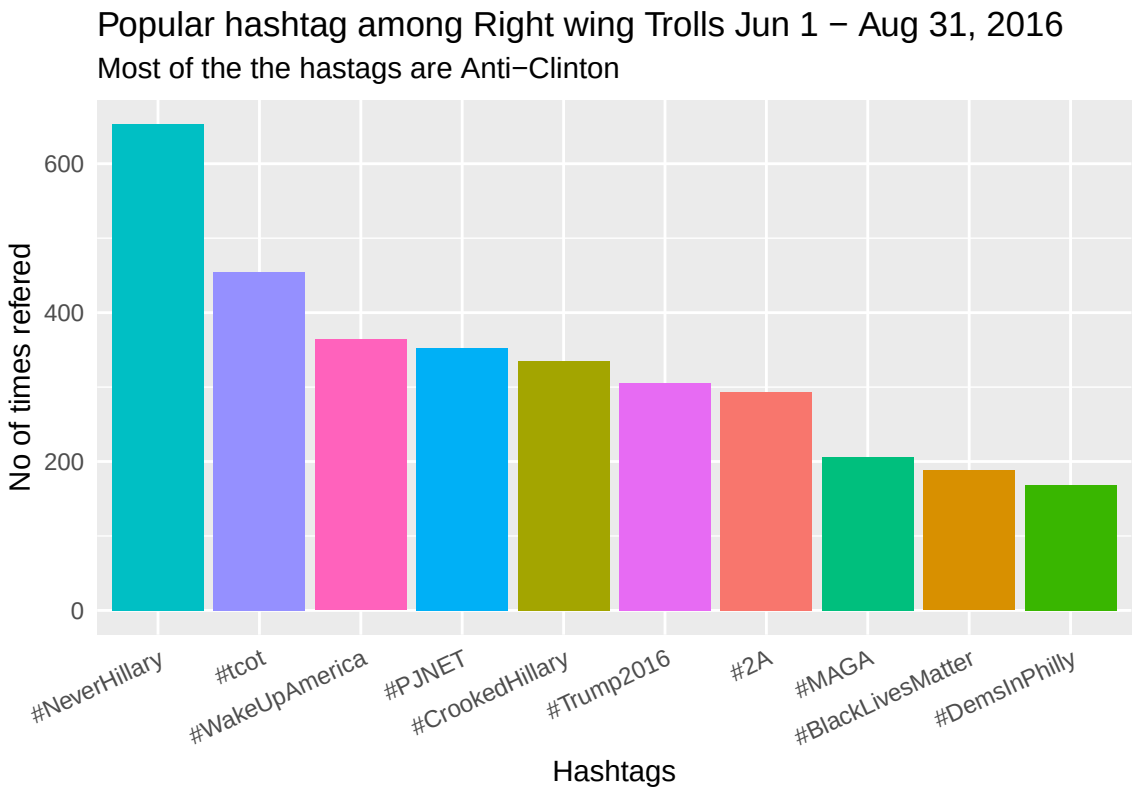
We further use the R 'SentimentAnalysis' package to calculate sentiment of the Tweeters (Both Left and Right Trolls) for Bernie Sanders and Hillary Clinton during the DNC. We will be using SentimentGI calculator of the R's 'SentimentAnalysis' , which is based on sentiment on the words of Harvard-IV Dictionary.

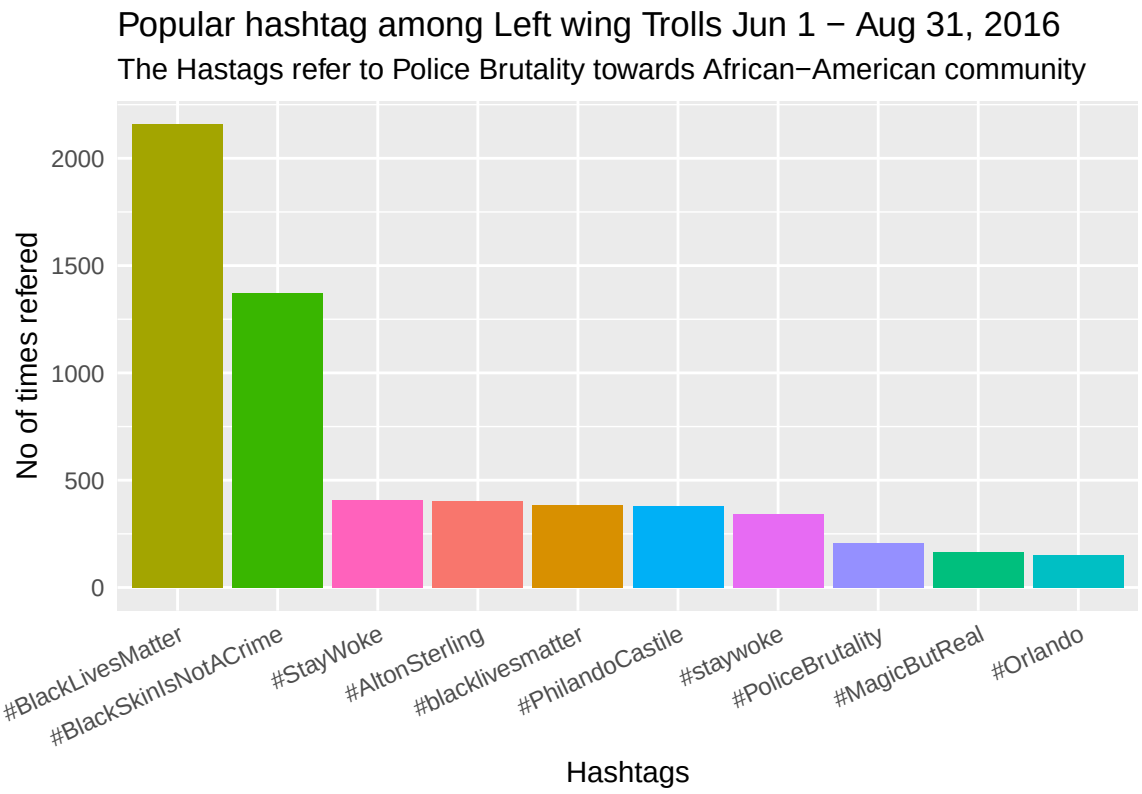
We subset the tweets where Keywords 'Bernie' and 'Sanders'(referencing Bernie Sanders) and 'Hillary' and 'Clinton' (referencing Hillary Clinton) were used. This is done seperately for Right and left wing Trolls. Each Tweet receives a score from -1 to +1 depending on the sentiment expressed in the tweet.(Negetive to Positive). The Average Sentiment Score is calculated by taking the average of the tweet sentient score for the particular group.



4.3. Popular Hashtags during the period

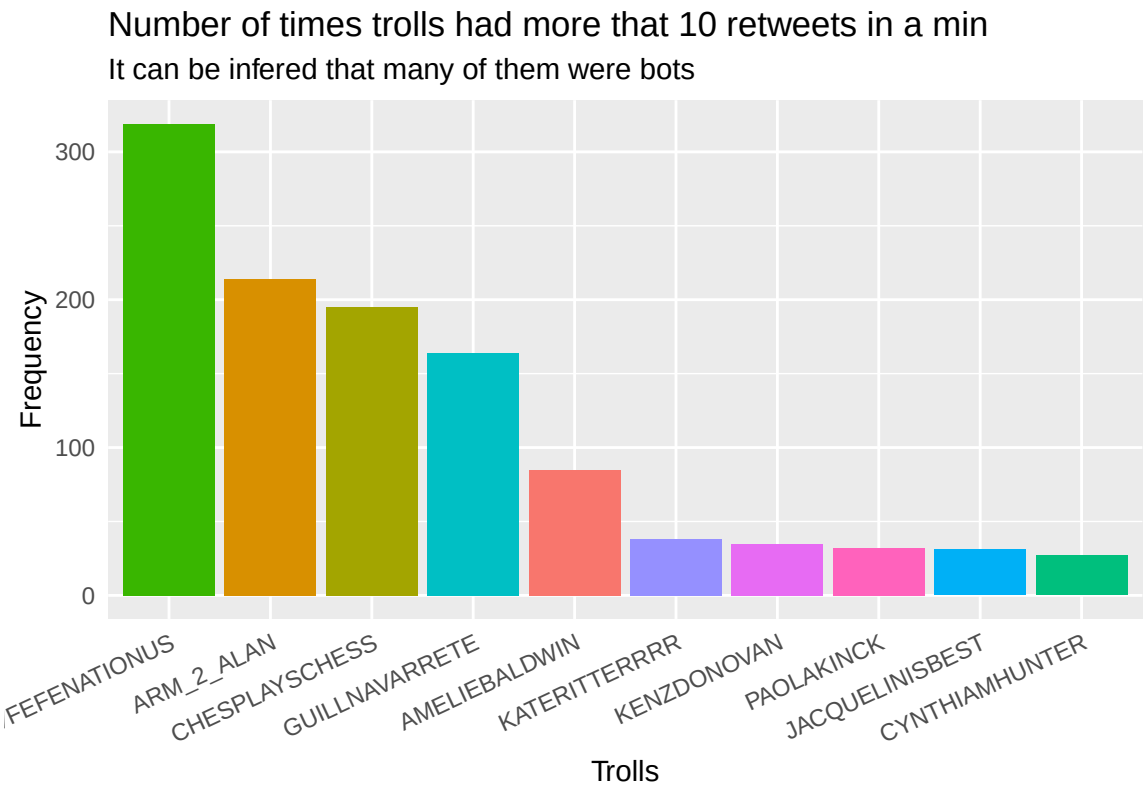
The Most Popular Hashtags referenced by the Trolls during that time period were



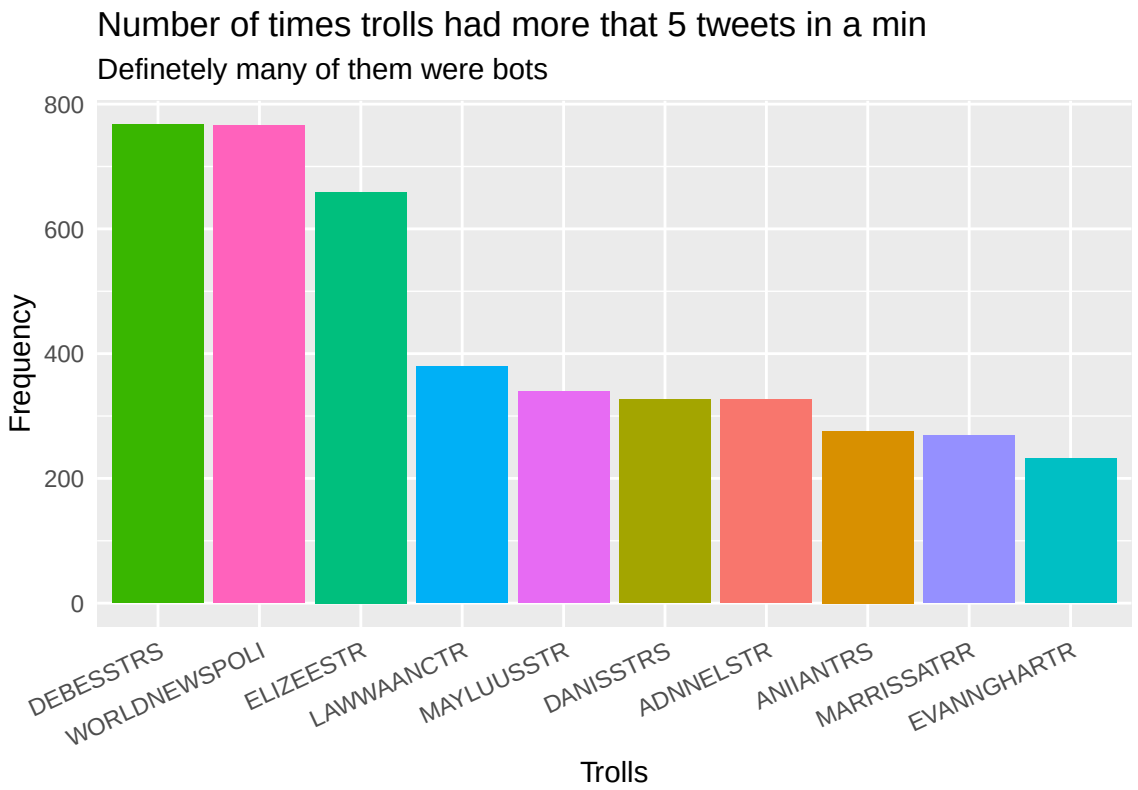

Popular hashtag among Left wing Trolls Jun 1 − Aug 31, 2016
The Hastags refer to Police Brutality towards African−American community
No of times refered
2000
1500
1000
500
0
#BlackLivesMatter
#BlackSkinIsNotACrime
#StayWoke
#AltonSterling
#blacklivesmatter
#PhilandoCastile
#staywoke
#PoliceBrutality
#MagicButReal
#Orlando
Hashtags

5 Speed of Tweets by a user

As the tweets had a time stamp to the minute, we ran analysis of the hypothesis that some of the accounts might by bots. Lets look at the Accounts that retweeted more than 10 times in a minute.



Now lets look at the Accounts that tweeted more than 5 times in a minute (which may be impossible for a human.



Top 10 instances where accounts had high number of tweets in a minute

```
##         publish_date          author count
##  1: 4/19/2017 10:56 WILLIAMS8KALVIN   116
##  2:  8/4/2017 13:36       ELIZEESTR    26
##  3:  8/16/2017 1:31       DEBESSTRS    25
##  4:  8/13/2017 1:11     MARRISSATRR    24
##  5: 8/12/2017 18:52      ALANISSTRS    23
##  6: 8/16/2017 14:28       ANIIANTRS    22
##  7: 8/13/2017 16:38       EISSYT56T    22
##  8: 10/2/2017 16:07       DANISSTRS    21
##  9:  8/3/2017 11:53       DEBESSTRS    21
## 10:  8/5/2017 23:53       DEBESSTRS    21
```

```
## Time count was greater than 5 per min  12065
```

12065 times has more that 5 'Orignal Tweets' been posted per min. WILLIAMS8KALVIN has 116 tweets per min, which makes a case for some of the accounts being bots and not human accounts"

Policy Recommendations & Future Research Questions

- The burst of activity on twitter around certain events like the DNC convention bring to light the strategy used on twitter by russian trolls more sharply. Looking at tweets around the event also help focus the analysis.

- Identifying bots based on certain characteristics of a tweet - includes the text (like specific words used and patterns of capitalizing letters), how many hashtags it includes and how often it is retweeted.

Contributions

- The final approach was decided collaboratively through group discussion and initial analysis of the data.

- Alexandra DeKinder and Ameya Karnad worked on data analysis using R

- Kulkanya Lekhyananda and Nitasha Nair worked on the presentation and policy report

References

1 - Jonathan Masters, "Russia, Trump, and the 2016 U.S. Election" (Council on Foreign Relations, 2018), https://www.cfr.org/backgrounder/russia-trump-and-2016-us-election.

2 - "Why We're Sharing 3 Million Russian Troll Tweets," FiveThiryEight, 2018, https://fivethirtyeight.com/features/why-were-sharing-3-million-russian-troll-tweets/.

3 - "Why We're Sharing 3 Million Russian Troll Tweets," FiveThiryEight, 2018, https://fivethirtyeight.com/features/why-were-sharing-3-million-russian-troll-tweets/.