Textual and Statistical Analysis of Russian IRA Facebook Posts

*The paper is written in the scope of a student-faculty collaborative summer research with professor Richard K. Merritt.

David Oniani
Luther College
oniada01@luther.edu

Month Day, Year

Abstract

The 2016 United States Presidential Election was targeted by an unprecedented intelligence and influence campaign. Arising out of Russian so-called Internet Research Agency (IRA), it sought to sow discord and attack the fissures of the United States with the ultimate goal of swaying the election results. [1] [2] Recently, some of the IRA-backed Facebook advertisements were released by The United States House Permanent Select Committee on Intelligence. All of the advertisements are in the PDF format. We have scraped the PDF files and present the results obtained by textual and statistical analysis of the above-mentioned data. Authorship attribution and sentiment analysis tests were also performed. [1] [3] We have also made the data publicly available for other researchers and/or interested people in a much nicer and easier-to-manipulate CSV format.

[1]Please note that this paper does not discuss neither social, nor political implications of these events, but attempts to explore the methods of persuasion that were employed in this influence campaign.

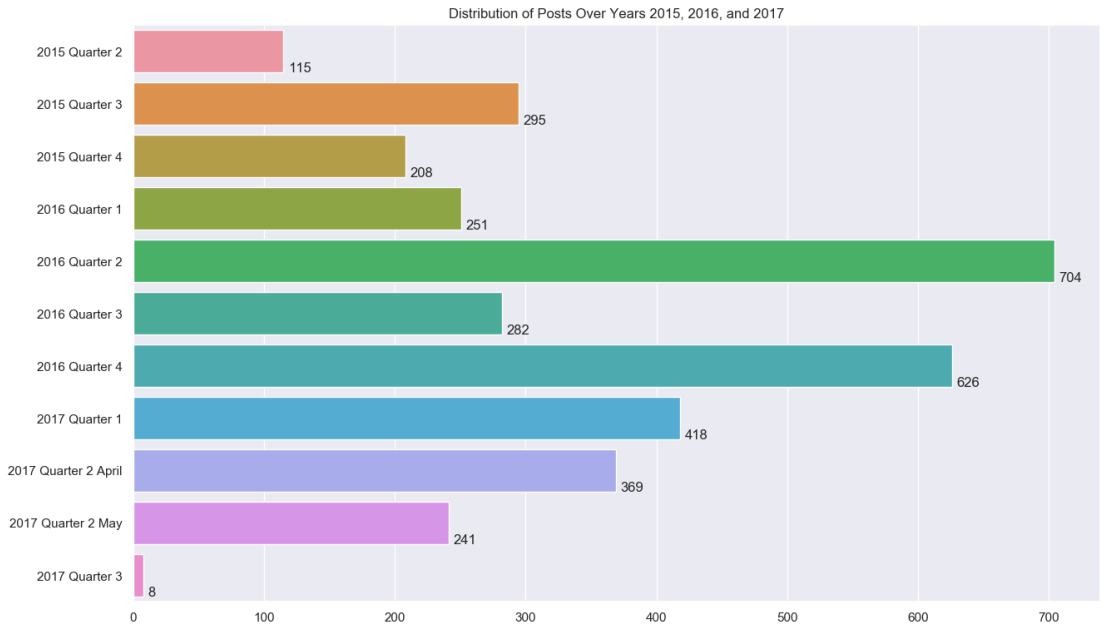
Table of Contents

Data Preparation

The data was harvested from [2] more than 3500 Russian IRA Facebook posts made publicly available in the PDF format by the House Intelligence Committee. We used the free and open-source Python library [4] `pdftotext` to scrape the data. Many CSV files were formatted in a way that it was hard for `pdftotext` to scrape it correctly. Because of this, we have manually reviewed most of the CSV files for validity. [3] All the CSV files have been made publicly available.

General Statistics

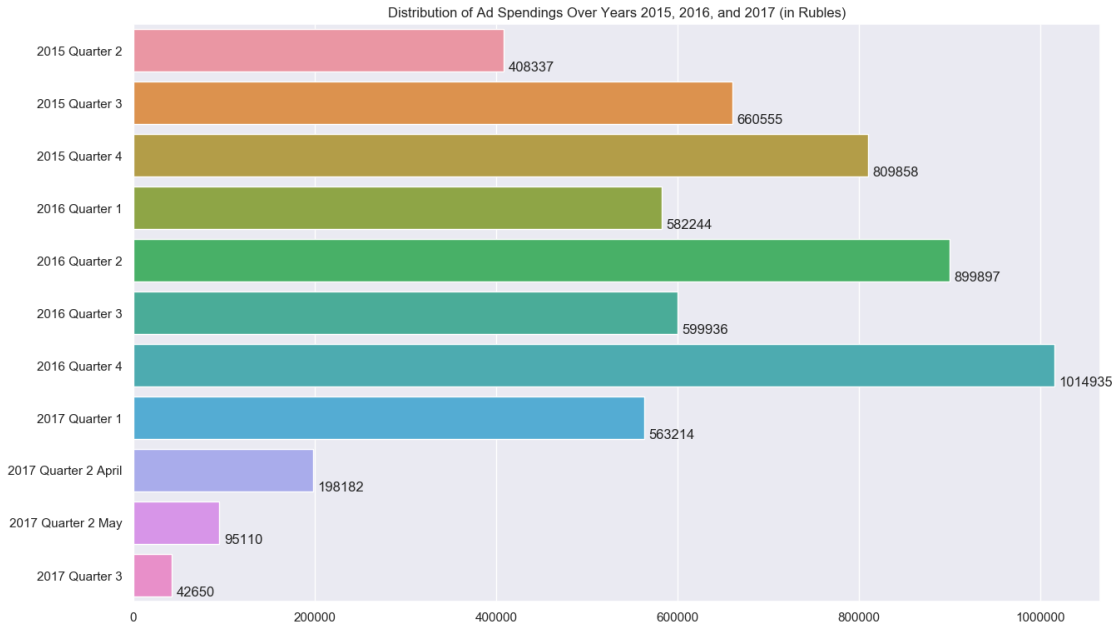
The distribution of posts over all three years shows us a bimodal distribution with two peaks in 2016 Quarter 2 and 2016 Quarter 4. Given the fact that the US Presidential Elections were held in 2016 Quarter 4, it is surprising that the second quarter had the most of the Facebook ads in it.



Distribution of Posts Over Years 2015, 2016, and 2017.

When it comes the money spent on ads, however, the fourth quarter of 2016 is, by far, the one on which IRA spent the most money on. It is also interesting that most of of the money was payed in Russian Rubles with two exceptions in 2016 quarter 3 and 2017 quarter 1 where they have spent

$74.000 and $35.330 respectively.



Distribution of Ad Spendings Over Years 2015, 2016, and 2017 (in Rubles).

Textual Analysis

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean porta purus et sem gravida rutrum. Maecenas blandit nulla ac luctus tempus. Nam finibus posuere ante, et lacinia massa vestibulum sit amet. Nulla velit arcu, efficitur quis turpis nec, sollicitudin lobortis nisi. Vivamus ut diam ut eros faucibus fringilla. Suspendisse pellentesque magna nec velit tristique sollicitudin. Morbi ultrices nec augue et molestie. Nam sapien ante, ullamcorper elementum convallis id, faucibus in lectus. Fusce pellentesque mollis velit efficitur porta. Sed finibus ligula quam, et lacinia velit posuere auctor. Donec ligula lorem, dictum nec lectus in, vehicula tincidunt massa. In hac habitasse platea dictumst.

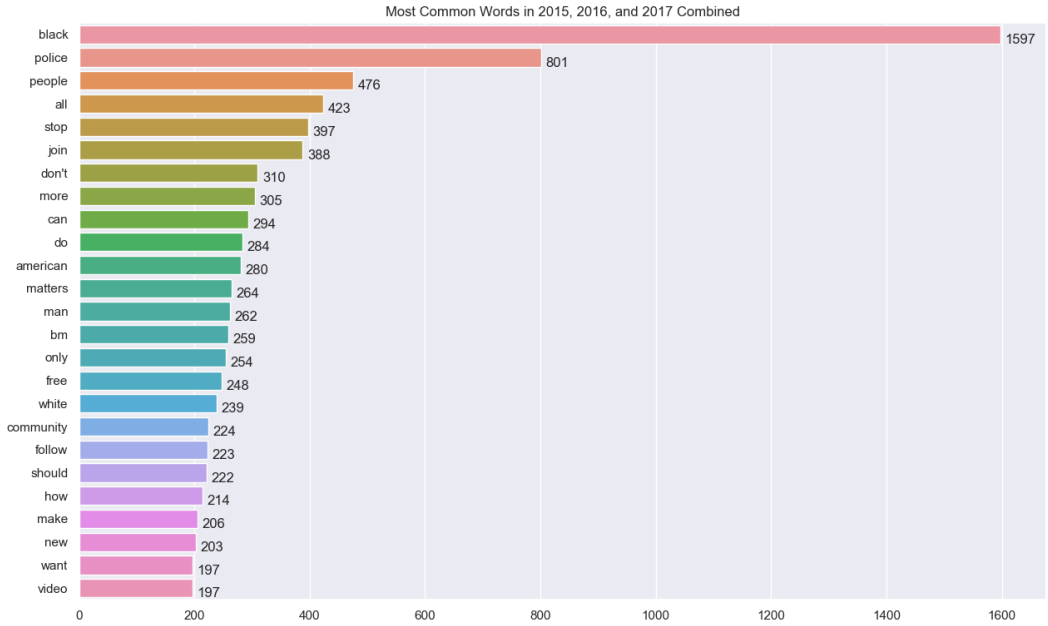


Figure.

Sentiment Analysis

For the sentiment analysis purposes, we used `python`'s `TextBlob` library. Suprisingly, out of all Facebook posts with `Ad Text`, 1643 were positive, 900 negative, and 933 neutral leaving the overall tone of the posts positive.

Analyzing Negativity			
Subjectivity Level	Year 2015	Year 2016	Year 2017
1.0	154	562	176
0.75	150	547	169
0.5	123	405	120
0.25	15	81	22
0.15	5	33	10
0.1	2	11	6
0.05	1	5	2

References

[1] The United States House Permanent Select Committee on Intelligence. *Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisements*. 2018. URL: https://intelligence.house.gov/social-media-content/ (visited on 08/27/2019).

[2] The United States House Permanent Select Committee on Intelligence. *Social Media Advertisements*. 2018. URL: https://intelligence.house.gov/social-media-content/social-media-advertisements.htm (visited on 08/27/2019).

[3] David Oniani and Richard Merritt. *CSV data scraped from the PDF files of IRA Facebook posts*. 2019. URL: https://github.com/oniani/ira-analysis/tree/master/data/csv (visited on 08/27/2019).

[4] Jason Alan Palmer. *pdftotext*. 2018. URL: https://pypi.org/project/pdftotext/ (visited on 08/27/2019).