

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 Facebook advertisements, which were published and promulgated by IRA, 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 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.

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

Table of Contents

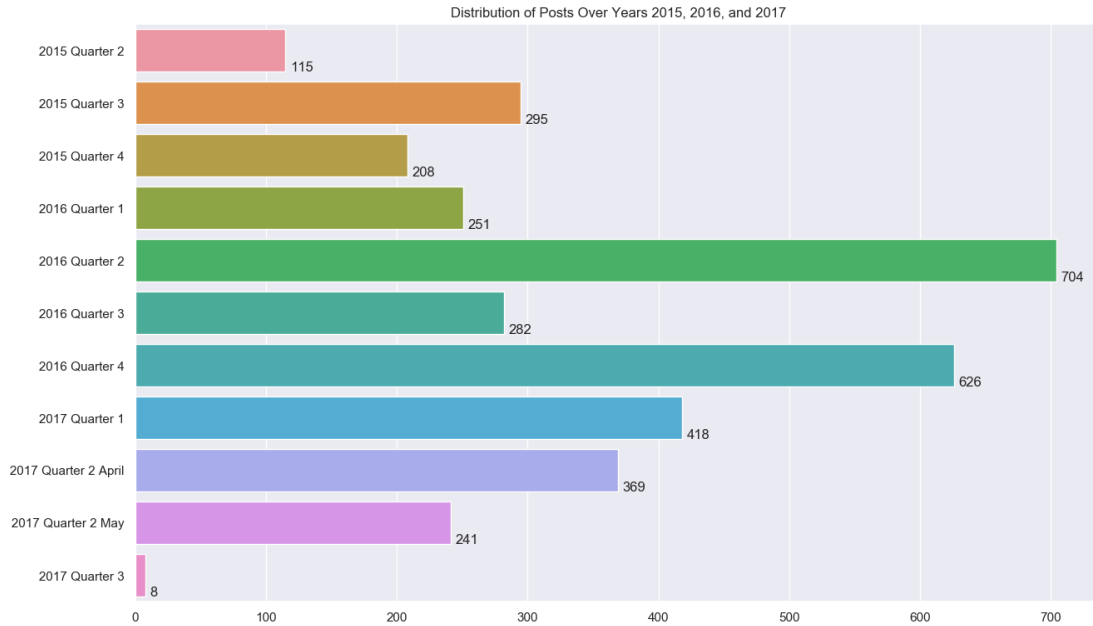
Abstract	1
Data and Preparation	3
General Statistics	3
General Statistics	4
References	6

Data and 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 scraped the data. Many CSV files were formatted in a way that it was hard to scraped the data correctly. Because of this, we have reviewed and 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.

[illegible]

Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah
Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah
Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah
Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah Blah
Blah Blah Blah Blah

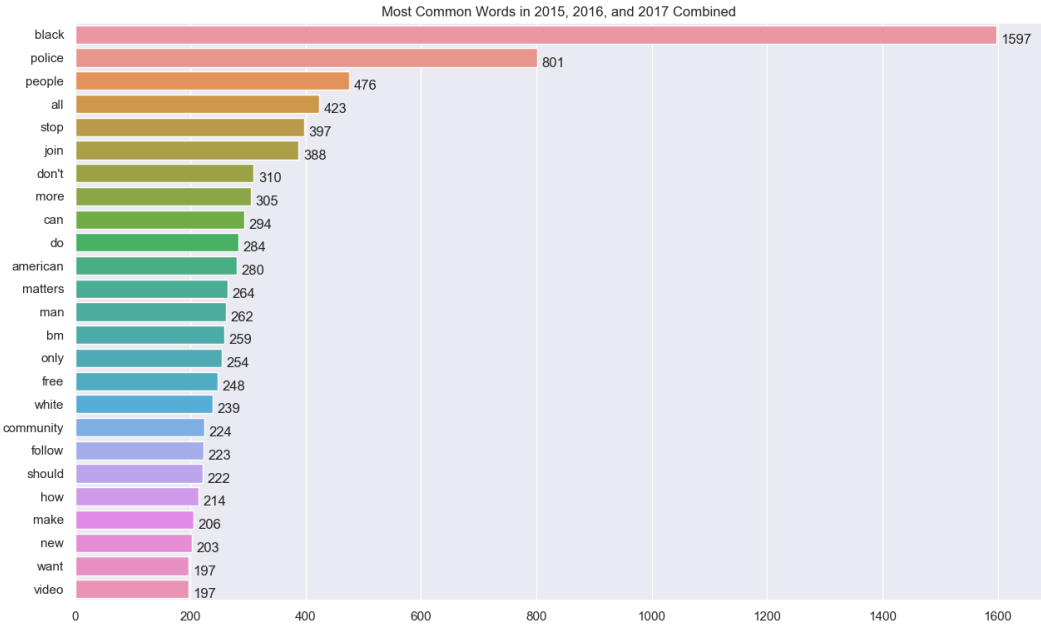


Figure.

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).