Russian Interference in 2016 Presidential Election: A Data Visualization

Haomiao Han / hh696@cornell.edu
Kristy Chen / sc3248@cornell.edu
Jacobs Technion-Cornell Institute, Cornell Tech, Cornell University

[This report is completed as a class project for Cornell Tech's INFO 5330 *Technology, Media, and Democracy* class.]

The Dataset

https://mith.umd.edu/irads/data.zip

This dataset, compiled by Lindblad et al. from the Univ. of Maryland, contains over 3000 Facebook advertisements that the Russia's Internet Research Agency purchased. This data was extracted from the Social Media Advertisements, which is a dataset released by Facebook to the House Intelligence Committee.

For each advertisement, the dataset has captured its metadata and content, including title, subject, image, time created, clicks, interest, placement and price paid, along with other details.

Visualization 1: Ads & Targeted Groups

We have seen news reports suggesting that the IRA have created ads centering on already divisive social issues like Black Lives Matter and the Tea Party on Facebook. It also purchased advertisements for the fake accounts created, which focused on varied social/political interests.
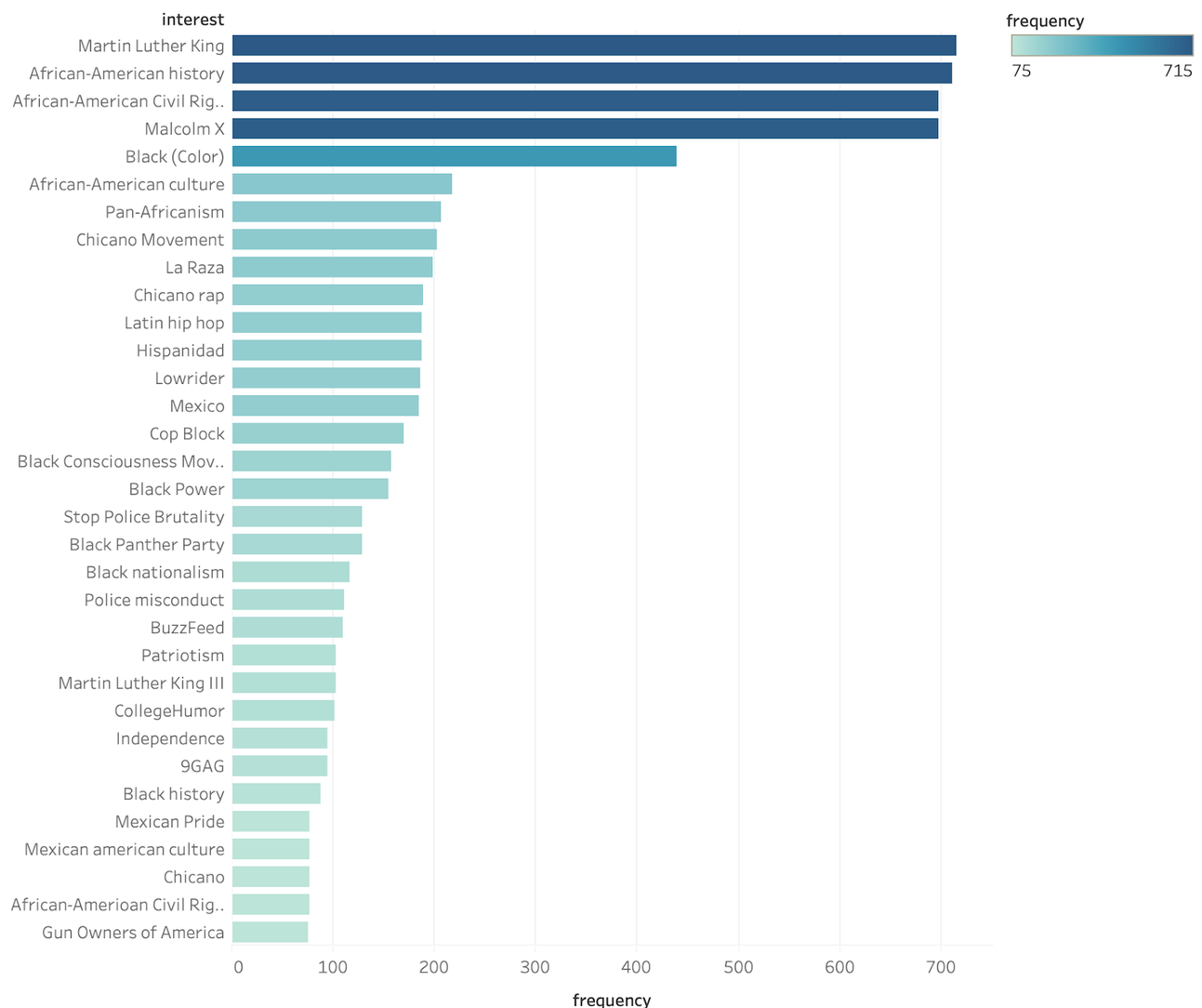
With the dataset, we counted the targeted interests for each ad - in other words, if the ad wants to target "Martin Luther King", for example, Facebook will show the ad to people who have demonstrated interested in MLK - and created this graph. Given that race is one of the prominent and more controversial issues in the U.S., a lot of

ads created by IRA targeted users who are interested in topics such as Martin Luther King, African-American history, and African-American Civil Rights.

We can also see that, presumably in order to achieve maximum impact, a lot of these ads target people who show interest in non-political topics such as "Chicano rap", "Buzzfeed" and "Latin hip hop".

Interests Count
IRA Ads on Facebook, 2015-2018

interest		frequency
		75 715

Martin Luther King
African-American history
African-American Civil Rig..
Malcolm X
Black (Color)
African-American culture
Pan-Africanism
Chicano Movement
La Raza
Chicano rap
Latin hip hop
Hispanidad
Lowrider
Mexico
Cop Block
Black Consciousness Mov..
Black Power
Stop Police Brutality
Black Panther Party
Black nationalism
Police misconduct
BuzzFeed
Patriotism
Martin Luther King III
CollegeHumor
Independence
9GAG
Black history
Mexican Pride
Mexican american culture
Chicano
African-Amerioan Civil Rig..
Gun Owners of America

frequency: 0   100   200   300   400   500   600   700

Sum of frequency for each interest. Color shows sum of frequency. The view is filtered on sum of frequency, which includes values greater than or equal to 75.

Visualization 1: Ads & Targeted Groups

Methodology: We first used python (with pandas and numpy) to preprocess the data and generate a dictionary of frequencies. The graph is then created using

Tableau.

Visualization 2: Money Spent & Influence

In total, the IRA spent about $100,000 to purchase over 3500 advertisements on Facebook.

The graph shows the number of people reached and cost for each advertisement. This is based on the ~2,000 ads that we have sufficient data, as the original dataset have missing data on some ads.

Impressions, Clicks & Money Spent
IRA Ads on Facebook, 2015-2018
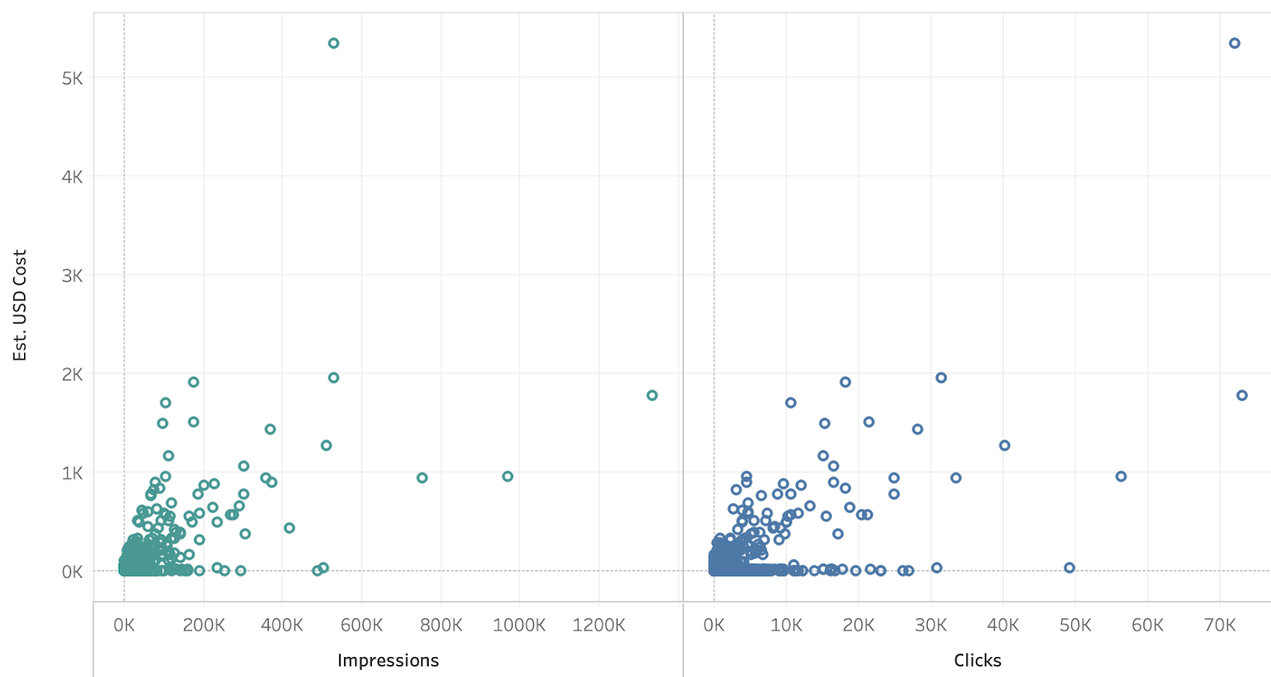
Sum of Item Type Metadata:Impressions and sum of Item Type Metadata:Clicks vs. sum of Est. USD Cost. Details are shown for pd_df_index.

Visualization 2: Money Spent & Influence

From the graph, we can see that the IRA have spend less than $500 on most of the ads, which in turn received around 0-200K impressions (views) and 0-10K clicks.

Nonetheless, we can also see some interesting peculiarities: for example, the two dots, close to each other, on the "Impressions" graph have around 500K impressions each, but the moeny spent to create these ads are minimal. Both ads are created

under the Facebook page "Blacktivist", and both feature a video - this one has a lighthearted, "viral" video, whereas the other ad has a seemingly more sensantionalist one. It is intriguing to see how so little money spent on propaganda (approximately $7 for the first one and $20 for the second one) can create a large impact.

Methodology: We first used python (with pandas and numpy) to preprocess the data: as most of the original payment currency is in Russian Rubles (!), we converted them into US Dollars based on historical exchange rates. The graph is then generated using Tableau.

Visualization 3: Words Frequency in Ads

We then turn to a more NLP focused persepective and created this visualization to show the number of times each word appeared in the ads.

We can see that most ads aimed to target on socially divisive issues with the words used in the subjects such as *black, white,* and *racism.* The word *2nd* most likely refers to the discussion on the 2nd Amendment, while *police, shot* and *brutality* likely refer to the controversy surrounding police brutality. The frequent apperance of these words corroborates with our first graph, which shows that a lot of these ads exploit racial issues in the U.S.
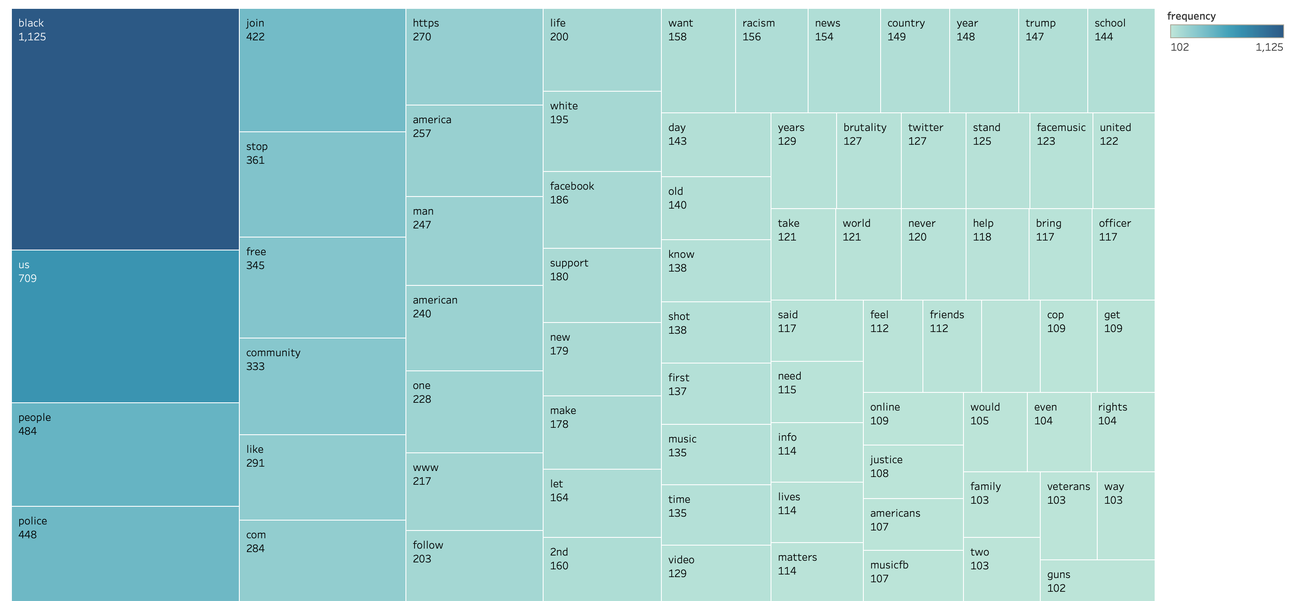
Word and sum of frequency. Color shows sum of frequency. Size shows sum of frequency. The marks are labeled by word and sum of frequency. The view is filtered on sum of frequency, which includes values greater than or equal to 100.

Visualization 3: Words Frequency in Ads

Methodology: We first used python to preprocess the data by converting all sentences into lower case, removing punctuation and tokenizing the ad contents, using NLTK as a NLP toolkit. We then produced a dictionary of word frequencies and generated the graph with Tableau.

Visualization 4: Sentiment toward Hillary Clinton vs. Donald Trump

Continuing with our NLP-focused analysis, we created this pair of histograms showing the distribution of sentiments of the ad contents.

We can see that despite the mean sentiments are somewhat similar for both candidates (0.012 for Clinton and 0.119 for Trump), there are a considerable amount of positive ads about Trump, whereas we don't see that many positive ads for Clinton. This graph can support the news reporting that the IRA's main focus is to create discord within the American society and publish negative ads against both sides. The graph also shows that how the IRA has also swayed the vote in favor of Trump.

Sentiment Analysis of Ad Content
(-1.0 - most negative, 1.0 - most positive)
IRA Ads on Facebook, 2015-2018
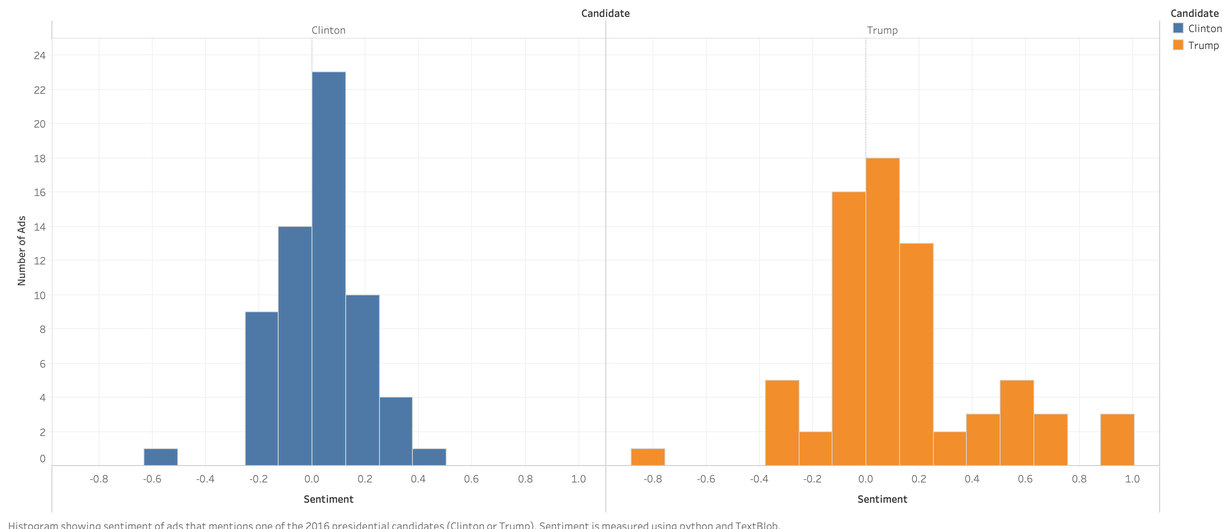
Histogram showing sentiment of ads that mentions one of the 2016 presidential candidates (Clinton or Trump). Sentiment is measured using python and TextBlob.

Visualization 4: Sentiment toward Hillary Clinton vs. Donald Trump

Methodology: We first used python to preprocess the data by identifying the ads that mention the keywords "Hillary", "Clinton" or "Trump", and categorized these ads into Clinton-related ads and Trump-related ads. We then used TextBlob as a NLP toolkit to estimate the sentiment of each ad. A graph is then generated with Tableau. It should be noted that most sentiment analysis toolkits available have their own limitations in estimating the sentiments of tweets, which is not an easy task. Therefore, this graph should be used in an experimental/reserach context only.

That's it!

Thanks for taking time to read our report. Don't hesitate to let us know if you have any further questions - we can be reached at hh696@cornell.edu and sc3248@cornell.edu.