Russian Twitter Trolls Analysis

Steven Cotton, Seth Posner, Yifan Huang, Kimi Wang

Background - Dataset

- In February 2018, NBC News published its database of more than 200,000 tweets that Twitter has tied to "malicious activity" from Russia-linked accounts during the 2016 U.S. presidential election.
- Russian troll accounts have been accused of influencing American elections by impersonating real user accounts
- NBC News open-sourced the dataset to help fight against the threat Russian trolls and bots pose to U.S. elections



Proposal - Can You Spot a Russian Troll

- We are interested in finding common features (e.g. frequency of words, phrases, misspellings) of these Russian Twitter Troll's tweets and identifying patterns of behavior. This could help people identify fake news and prevent them from affecting elections.
- Alternatively, we could discover that Russian twitter trolls are so advanced that an average twitter user could not discriminate between a troll's tweet and a user's tweet
- Goal: Is it possible to tell a Russian twitter troll from a typical user



Dataset

- NBC News 200,000 known Russian twitter bot/troll tweets
- A second dataset was needed to compare with the troll tweets
- Chris Albon, a political and data scientist, scraped twitter for 6.5 million tweet ids from the day of the election
- Ed King, a linguistics PhD student at Stanford, took over 300,000 of these tweets ids and made the full tweet data available on kaggle.com



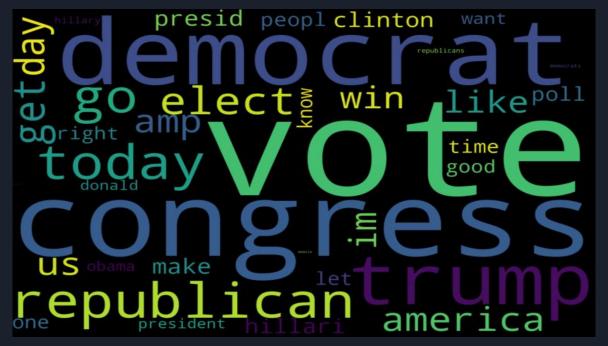
Data Cleaning

- The following were removed from consideration in our analysis
 - Emojis
 - Spaces

- The following were considered "Important Words" and were kept in the analysis
 - o hillary, clinton, donald, trump, obama, USA, vote
 - elect, president, democrat, republican, democrats
 - o republicans, crooked, emails



Typical Twitter User



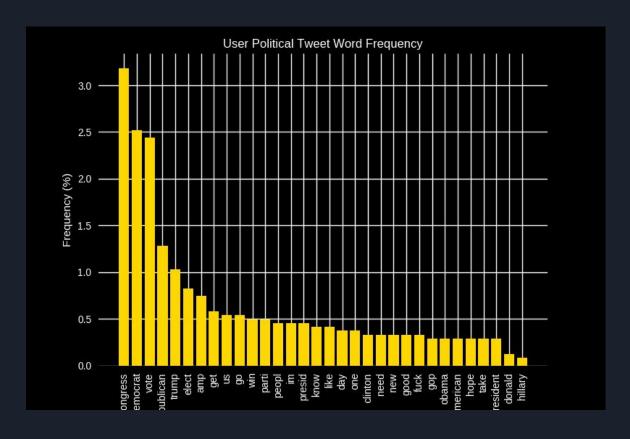


Typical User - Most Frequent Words

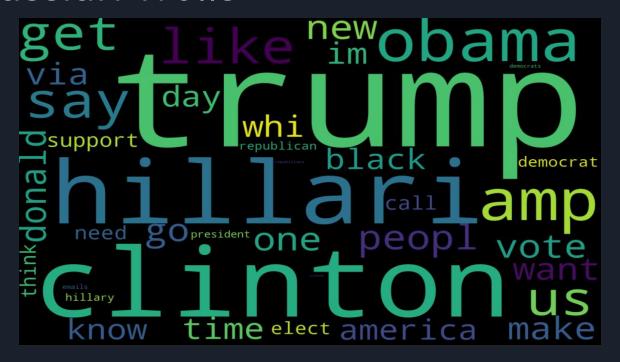
Top 5 Words

- 1. Congress
- 2. Democrat
- 3. Vote
- 4. Republican
- 5. Trump





Russian Trolls



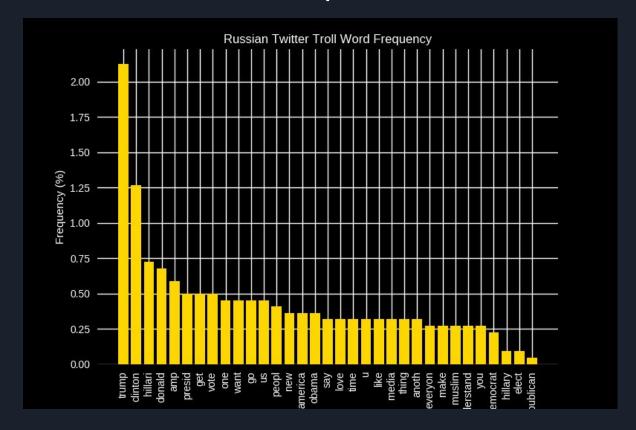


Russian Trolls - Most Frequent Words

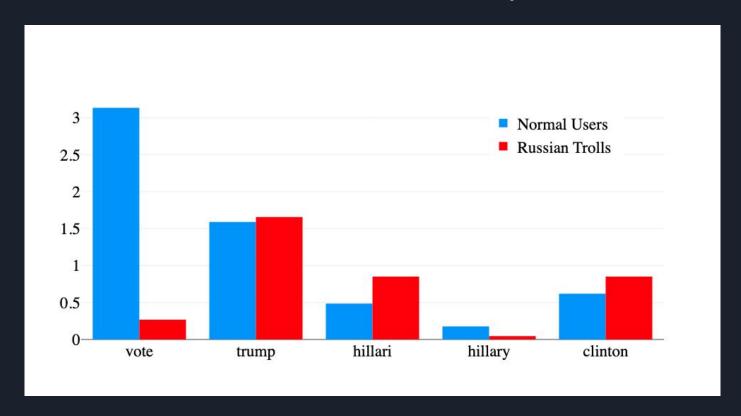
Top 5 Words

- 1. Trump
- 2. Clinton
- 3. Hillari
- 4. Donald
- 5. &



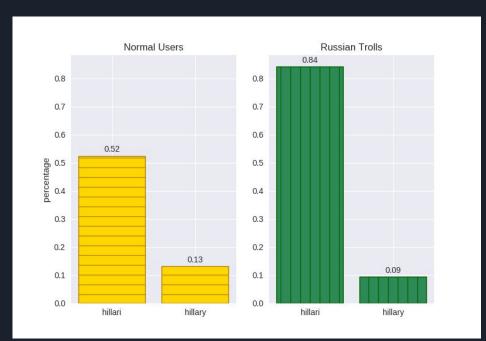


Notable Words Direct Comparison



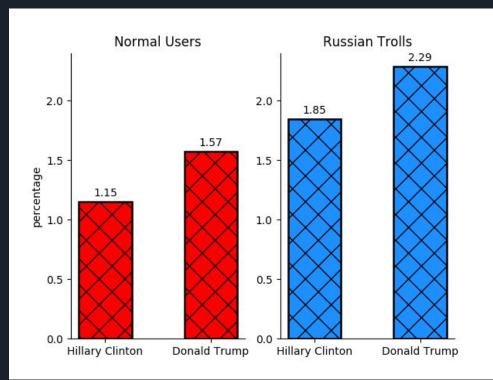
Hillari or Hillary?

- Upon discovering that Russian twitter trolls used the misspelled "Hillari", we needed to compare the frequency of Hillari and Hillary for each.
- We found that Hillari is far more common in Russian troll tweets than the typical user
 - But the typical user also misspells Hillary as Hillari more often than spelling it right!





Compiled Variations on Donald Trump vs Hillary Clinton



Variations Included

Donald Trump

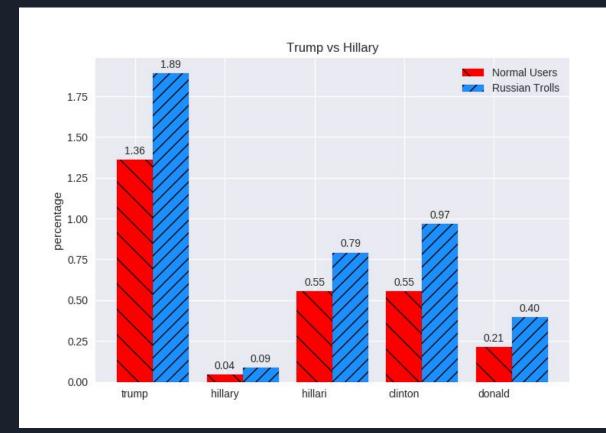
- donald
- trump

Hillary Clinton

- hillary
- hillari
- clinton



Variations on Candidate Names





Conclusions: Can You Spot a Troll?

- Russian Twitter trolls do not use the word "vote", instead preferring to single out a candidate
- Everyone frequently spells "Hillary" as "Hillari", though it appears to be a far more common misspelling for the Russian trolls
- Russian Trolls focus more on candidates than on parties
- Trolls use the ampersand (&) far more than the typical user
- While Russian Twitter trolls mention Trump far more than Clinton, the average user mentions Trump more than Clinton too, indicating Trump was just more often talked about on social media in general
- While these results will not allow you to identify a Russian troll tweet with perfect accuracy, they do give insight on when to be cautious

Possible "Identifying" Words

Russian Trolls

- Hillari
- 8

Typical User

- Vote
- Elect
- Republican/Democrat
- Congress

Non-Identifying Words

- Trump (unless used very frequently by same account, then likely Russian)
- Clinton
- Hillary
- America

