

Tweet Analysis

Presidential Statistics in 140 character

Bruno Korbar, Roman Huszar, Naofumi Tomita

Goal

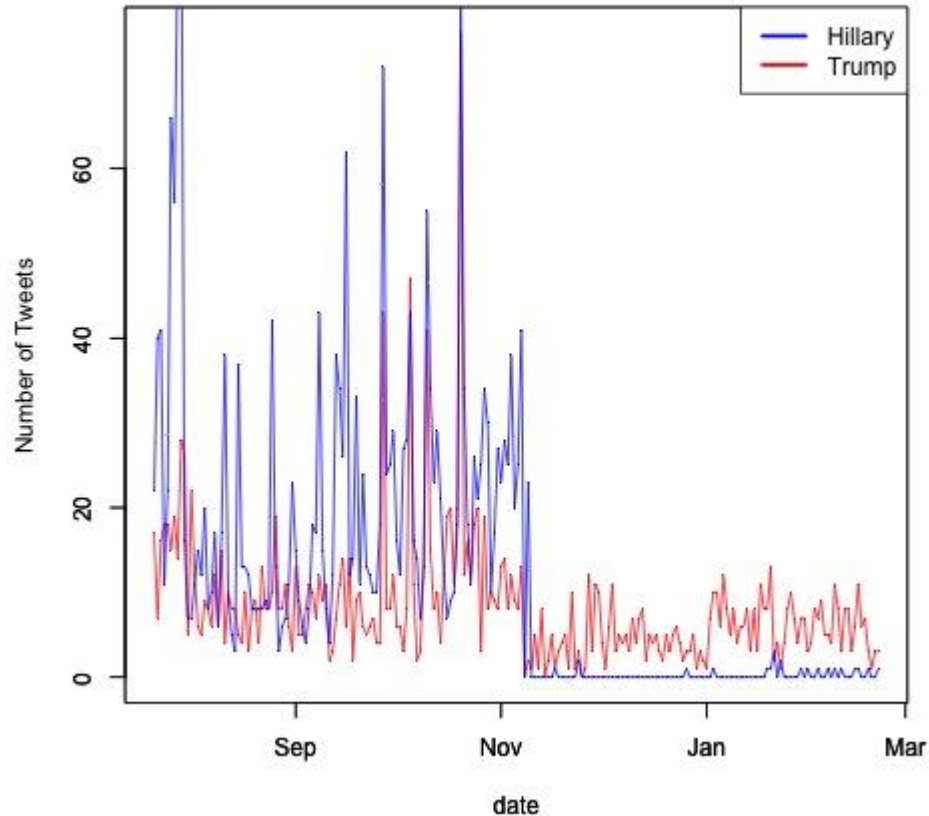
1. Understand what Twitter statistics are important
2. Answer questions about the individual based on the data

Data

- Scraped Twitter data using Twitter's API in Python
- **@realDonaldTrump**, and **@HillaryClinton**
- ~3200 tweets each

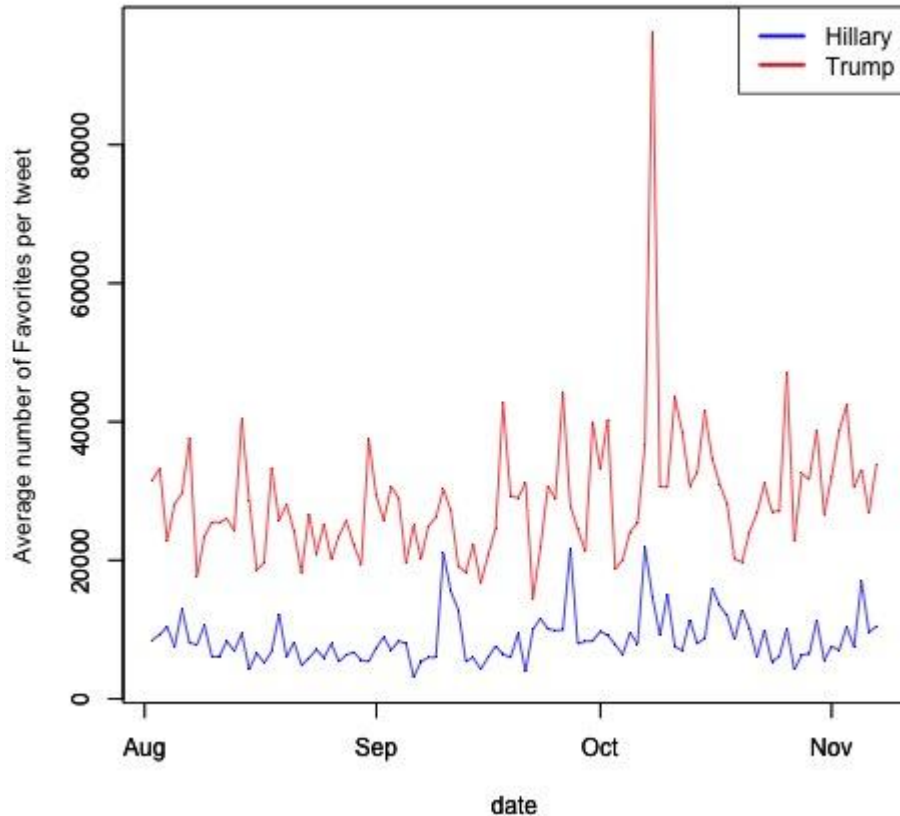
What makes tweets popular?

The number of Tweets (Chronological)



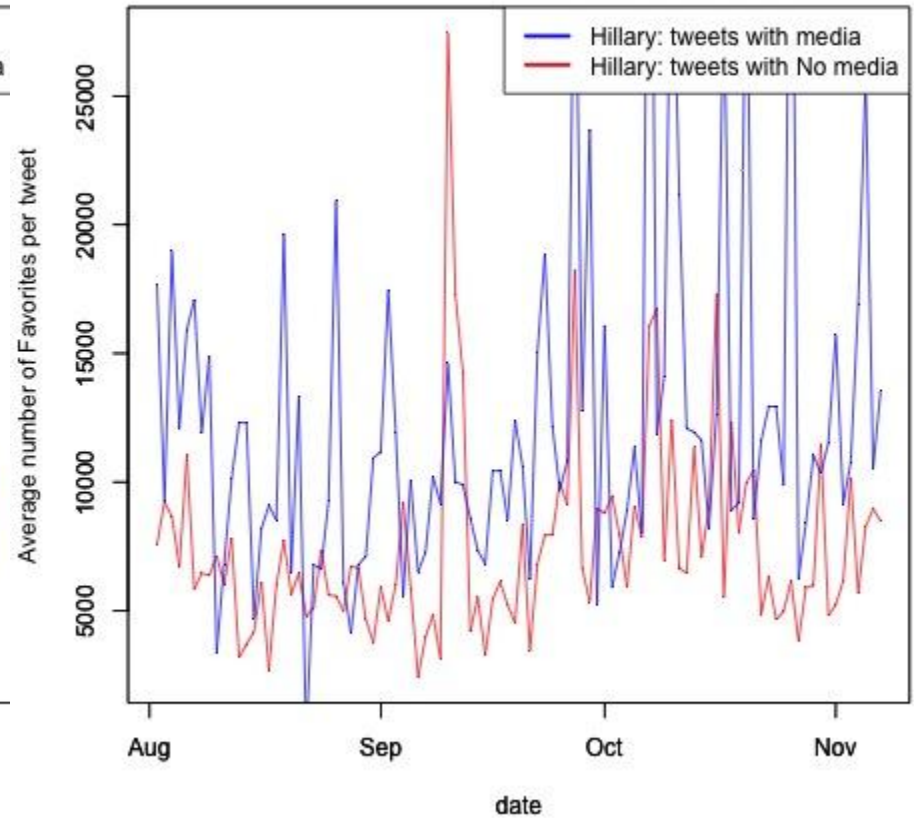
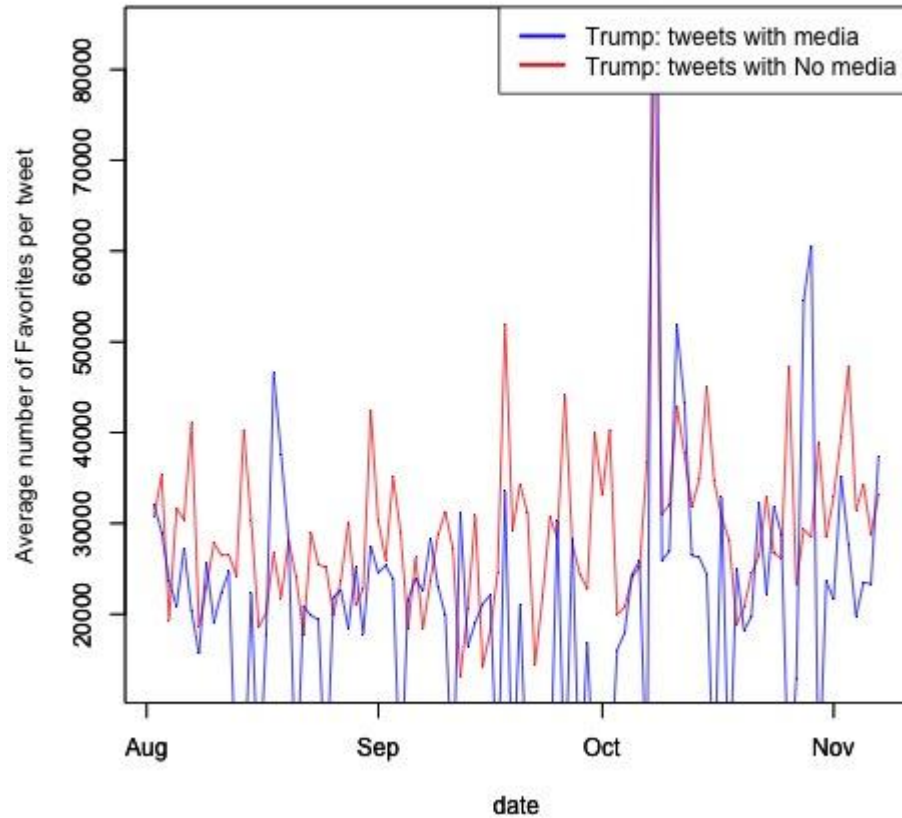
- The number of tweets **drops** immediately **after the election day** (Nov 8th)
- Before Nov 8th, Mr. Trump constantly tweets ($SD=9.946908$), while the number of tweets by Ms. Clinton fluctuates ($SD=14.31795$)

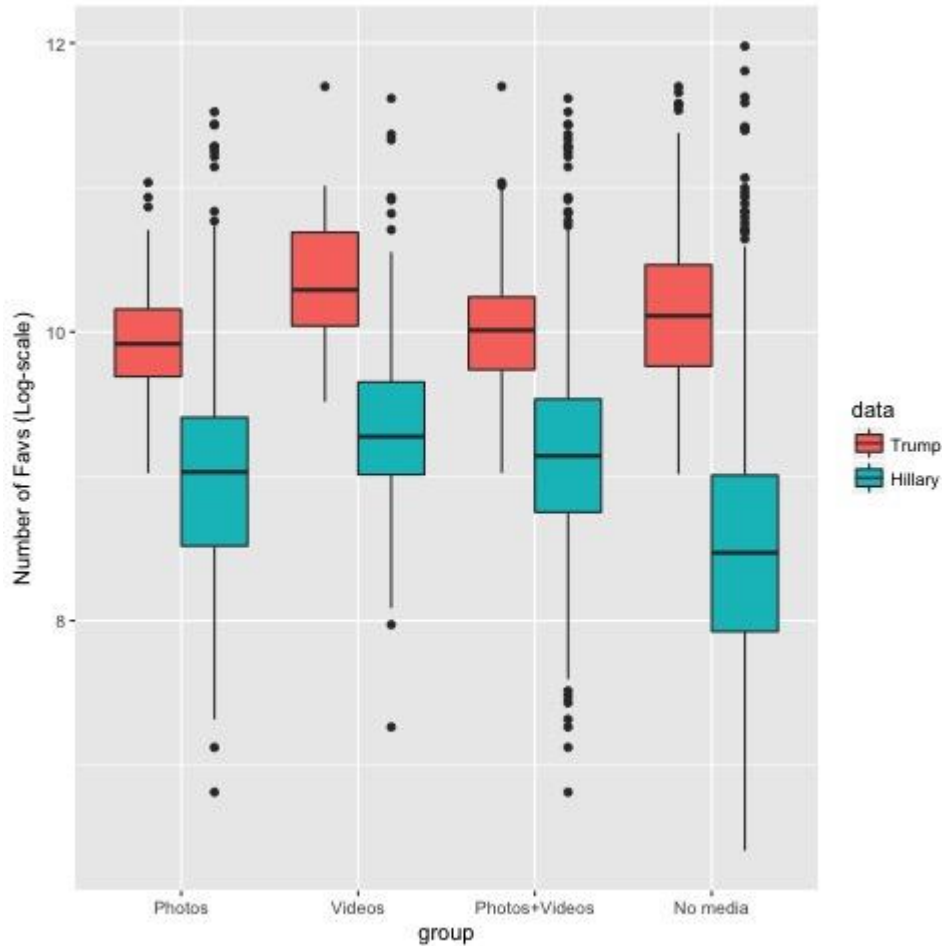
Average number of Favorites earns (Chronological)



Mr. Trump earns 3.3 times more favorites than Ms. Clinton

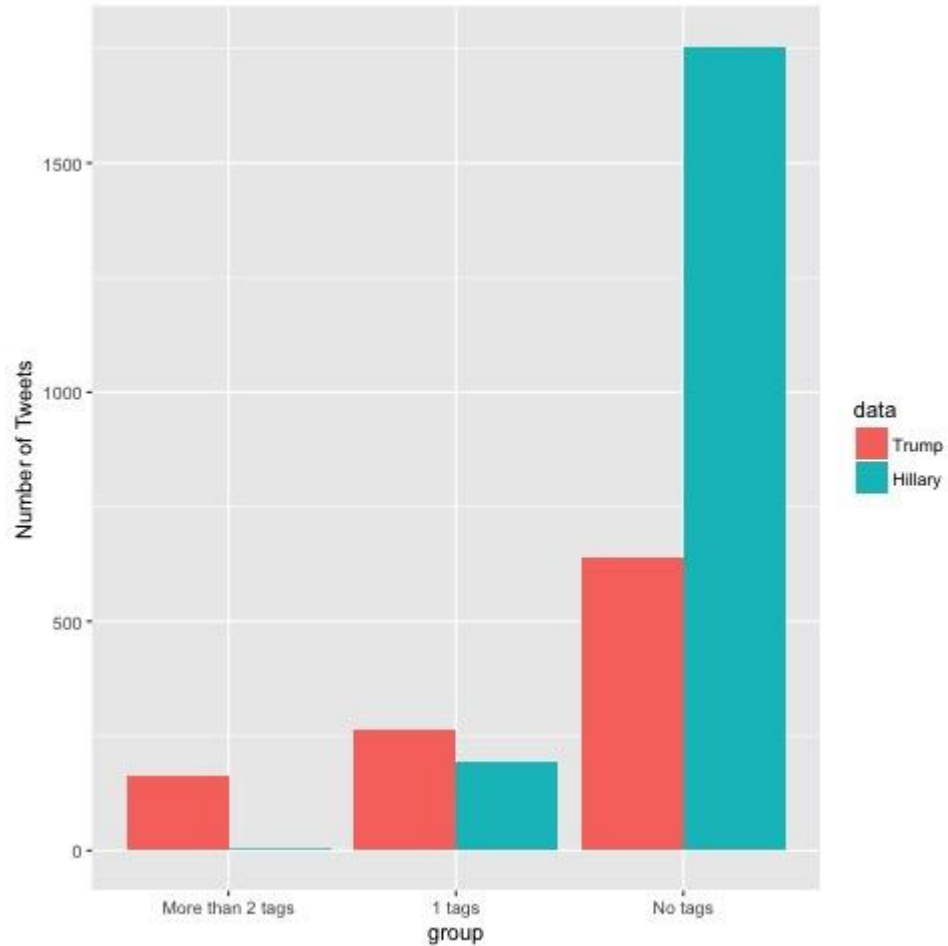
Is tweeting with photos/videos Effective?



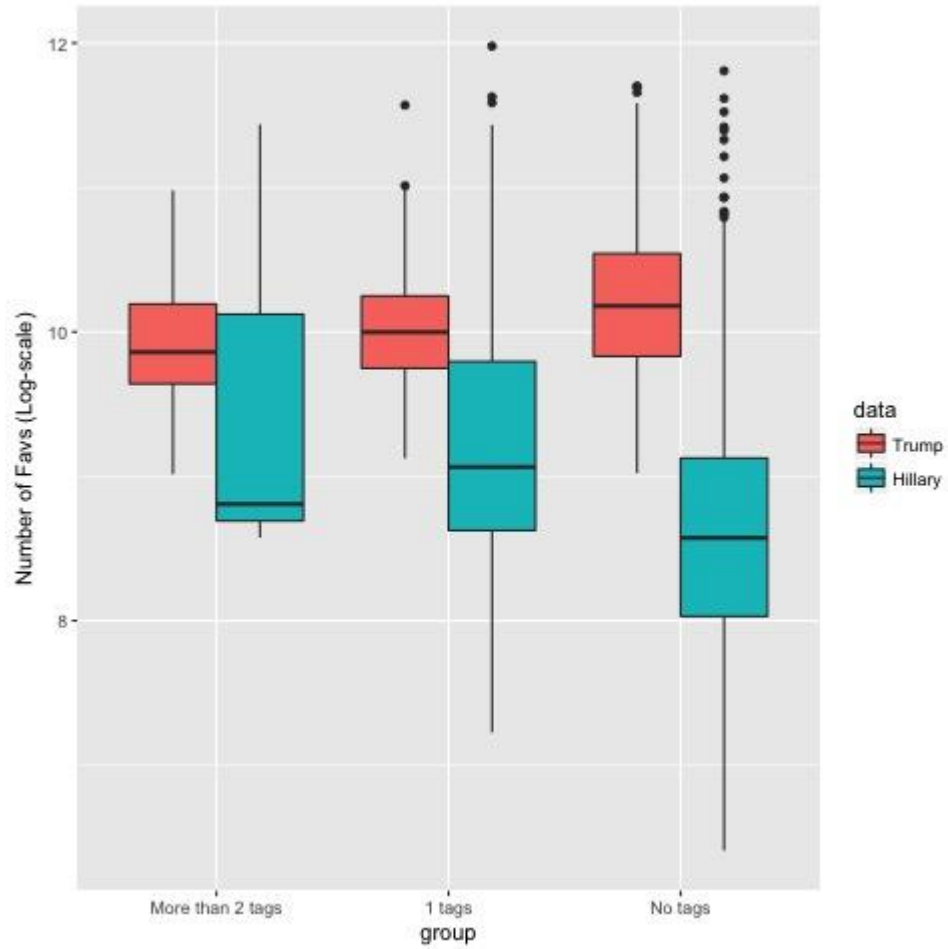


Mr. Trump is likely to receive more favorites on **plain tweets**, while Ms. Clinton receives more on tweets **with photos/videos**

How Hashtags are Effective?



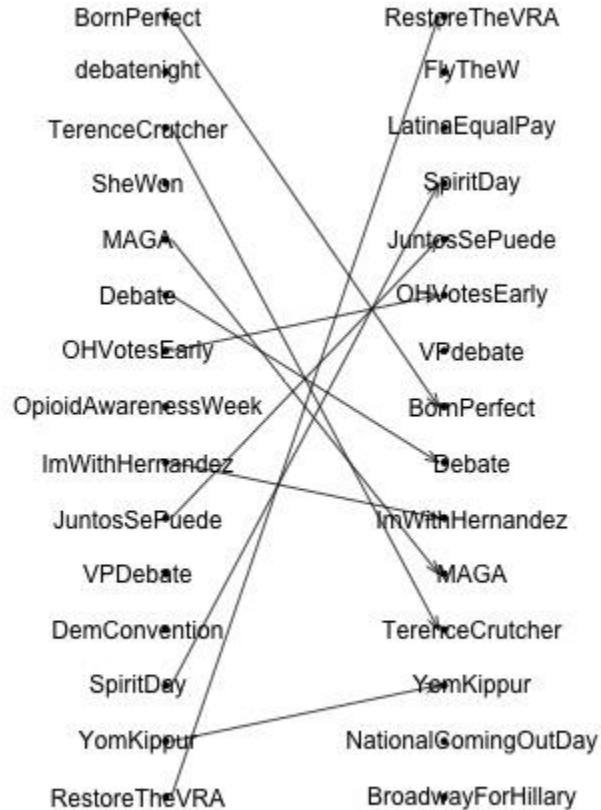
Mr. Trump favors hashtagging



But Mr. Trump earns less favs on tweets with hashtags...

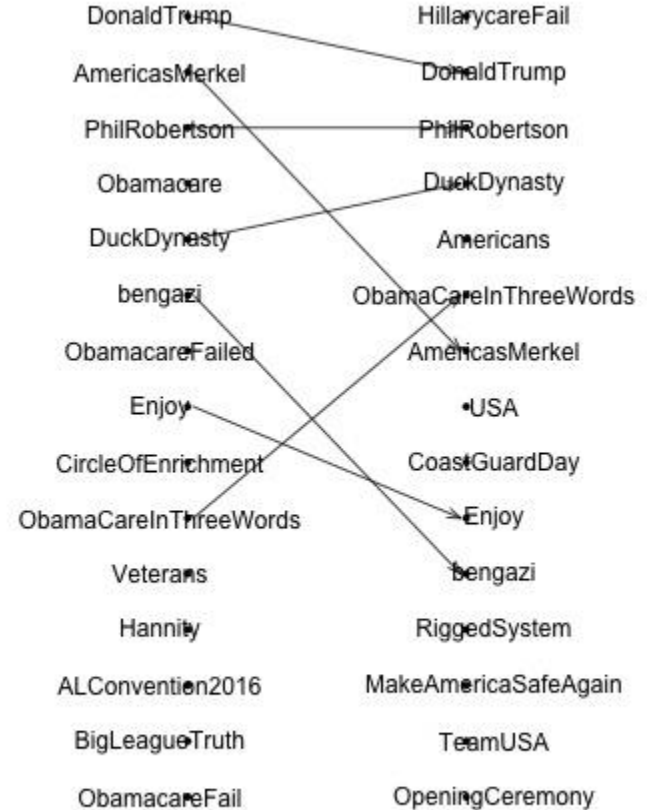
Better to repeat same hashtags?

Top 15 frequently used tags VS Top 15 Effective tags



Hillary Clinton

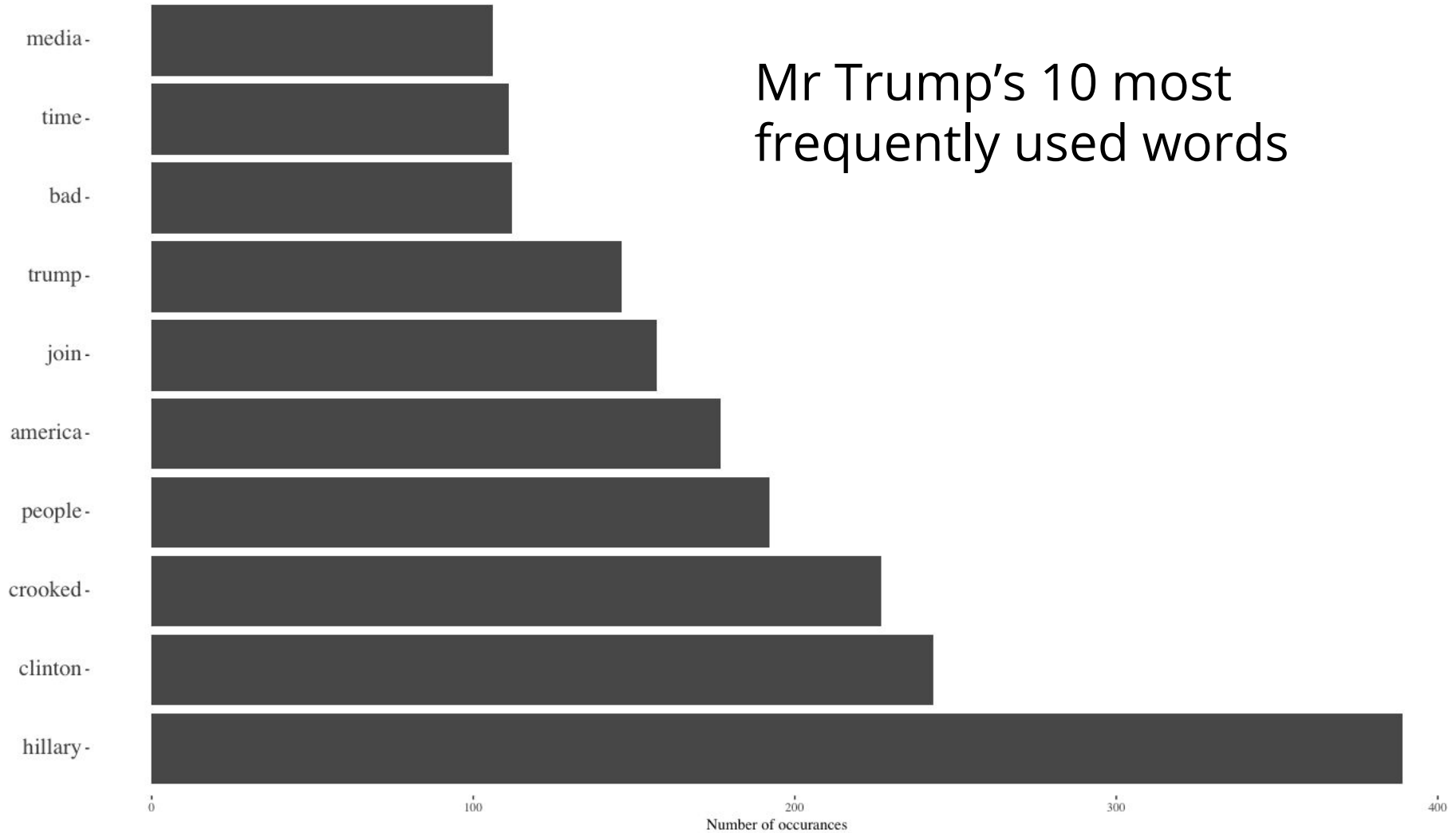
Top 15 frequently used tags VS Top 15 Effective tags



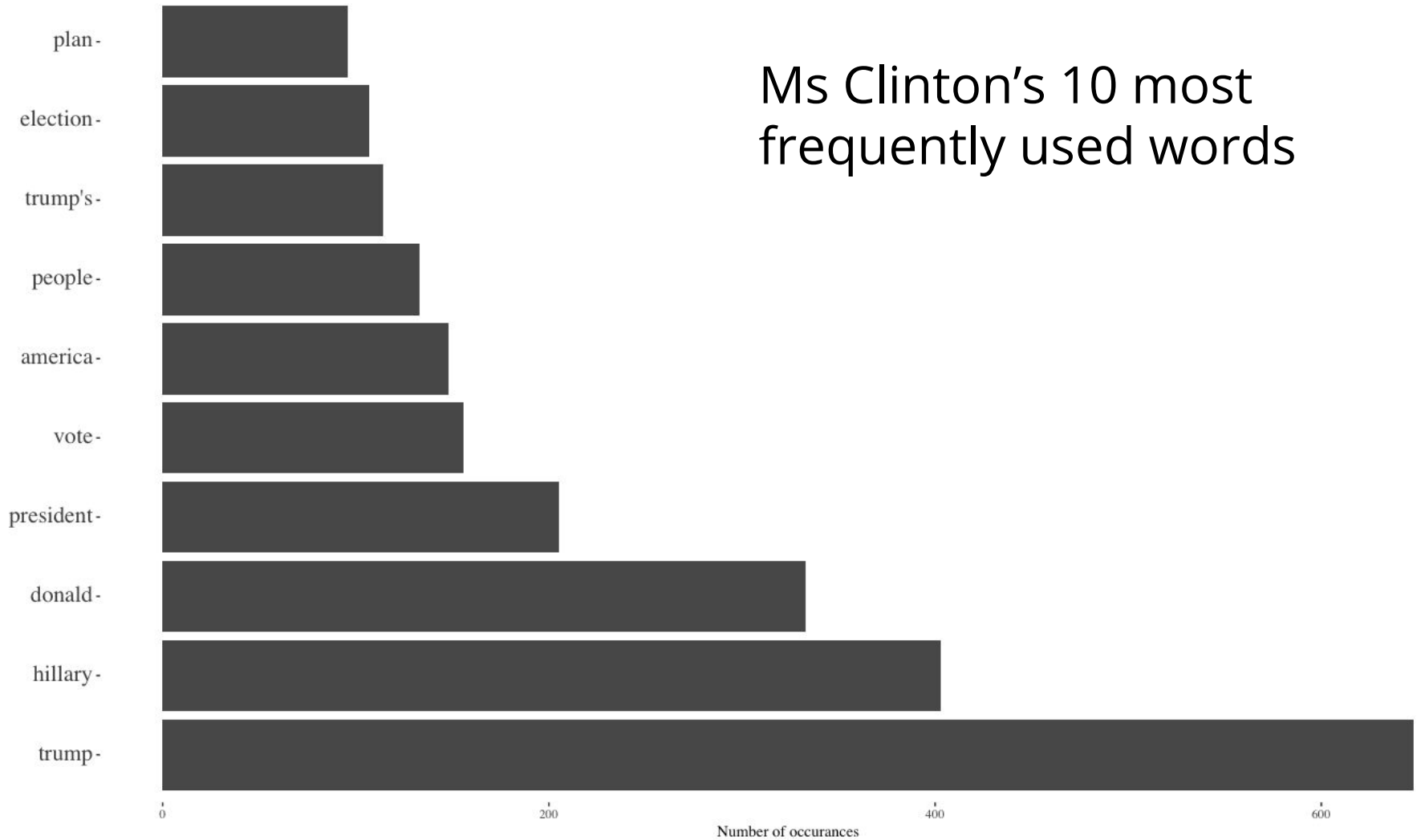
Donald Trump

What are they tweeting?

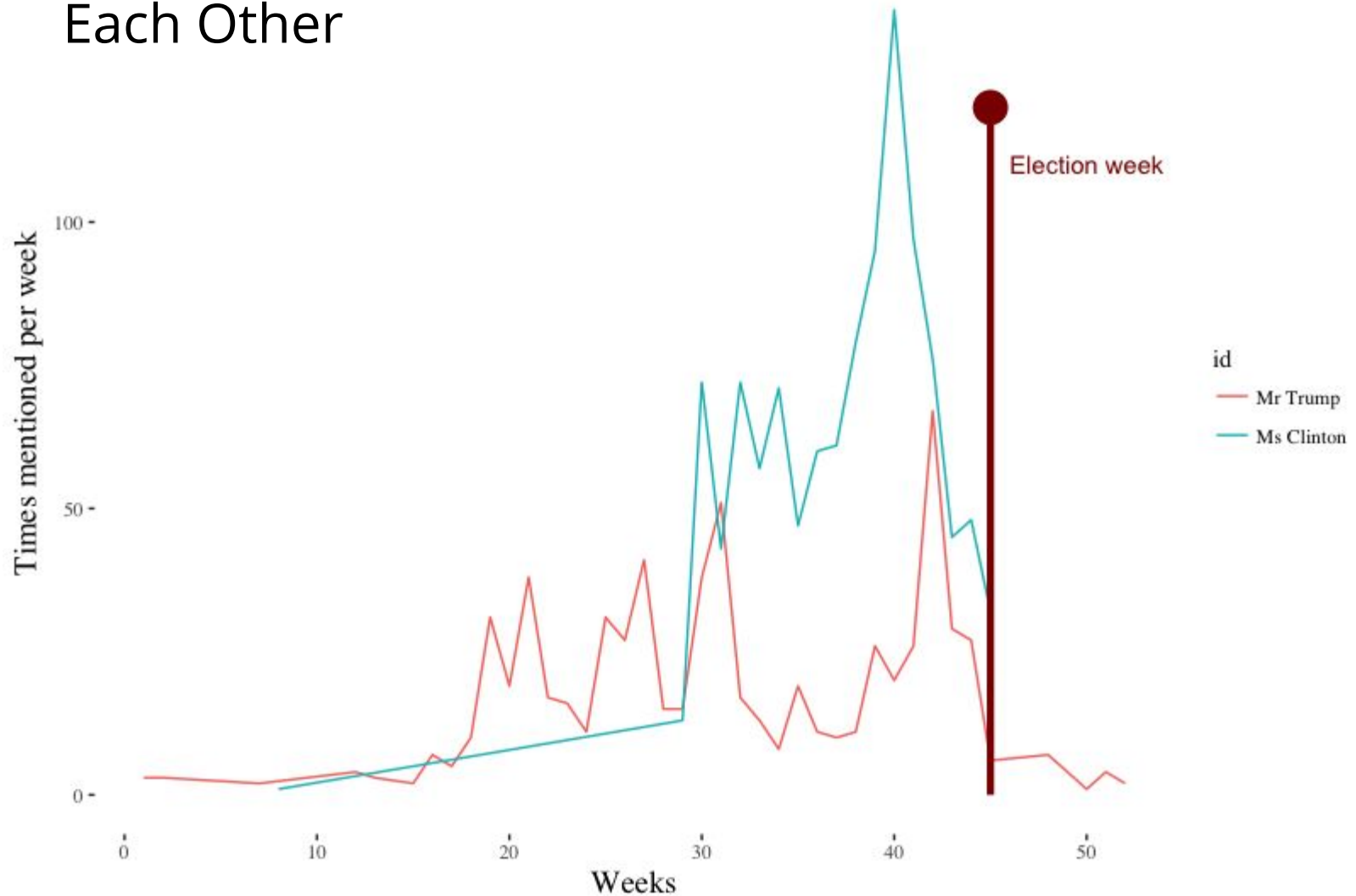
Mr Trump's 10 most frequently used words



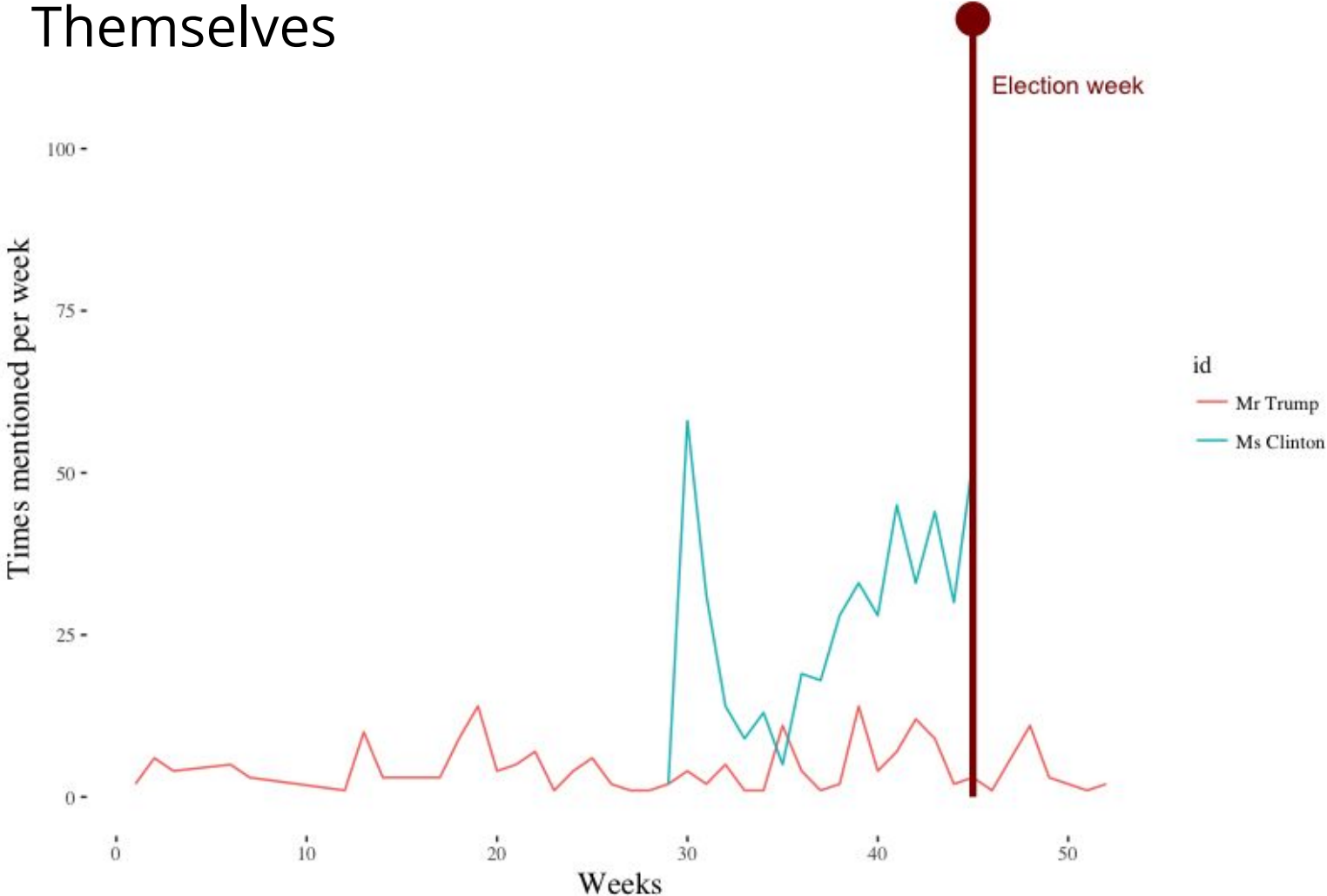
Ms Clinton's 10 most frequently used words



Candidates Mentioning Each Other



Candidates Mentioning Themselves



Interesting Statistics

@realDonaldTrump

21.65% of tweets mention Clinton

6.47% of tweets mention himself

@HillaryClinton

34.35% of tweets mention Trump

14.43% of tweets mention herself

Interesting Statistics

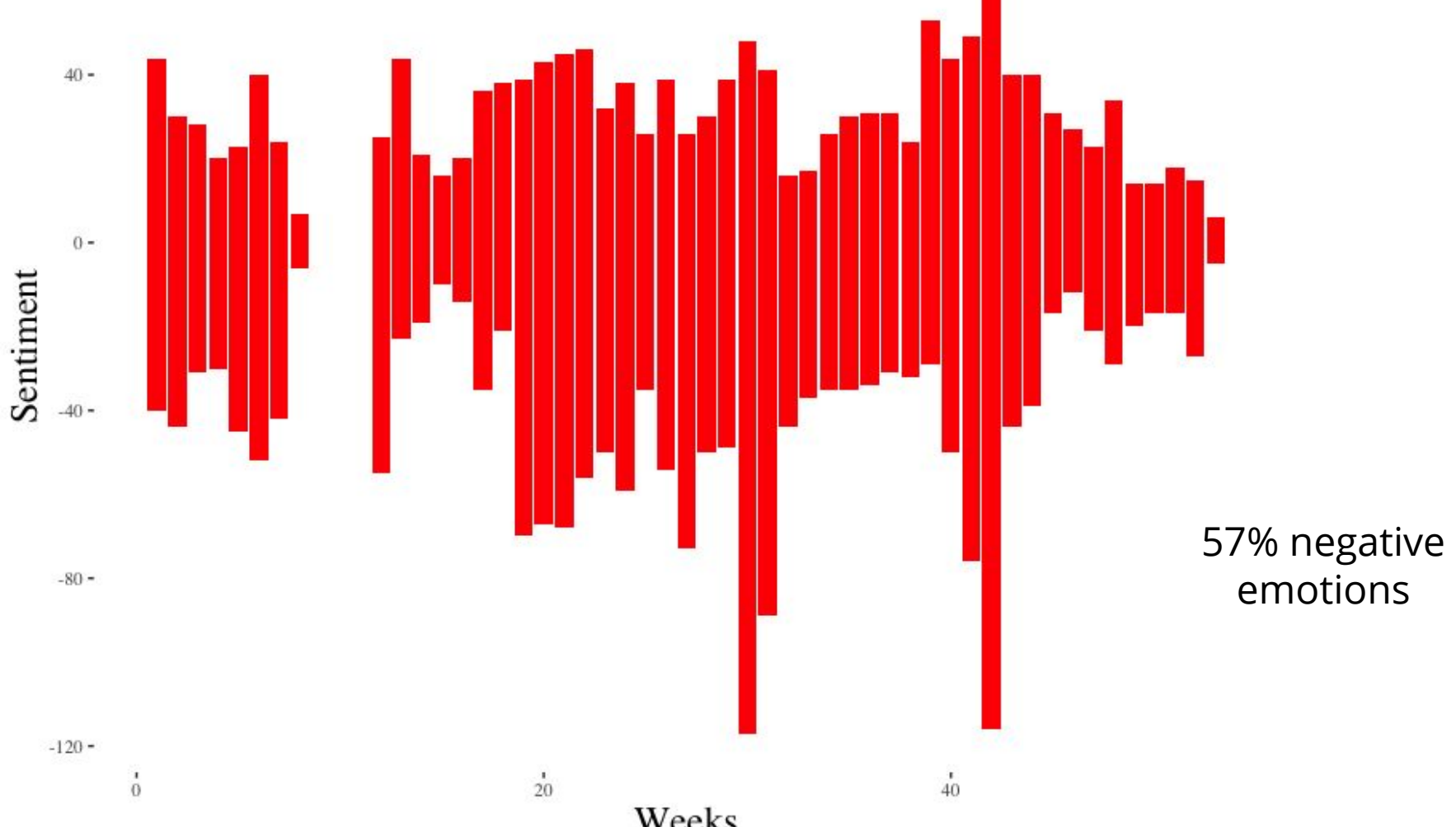
When @HillaryClinton mentioned @realDonaldTrump she got

6,525 favourites per tweet

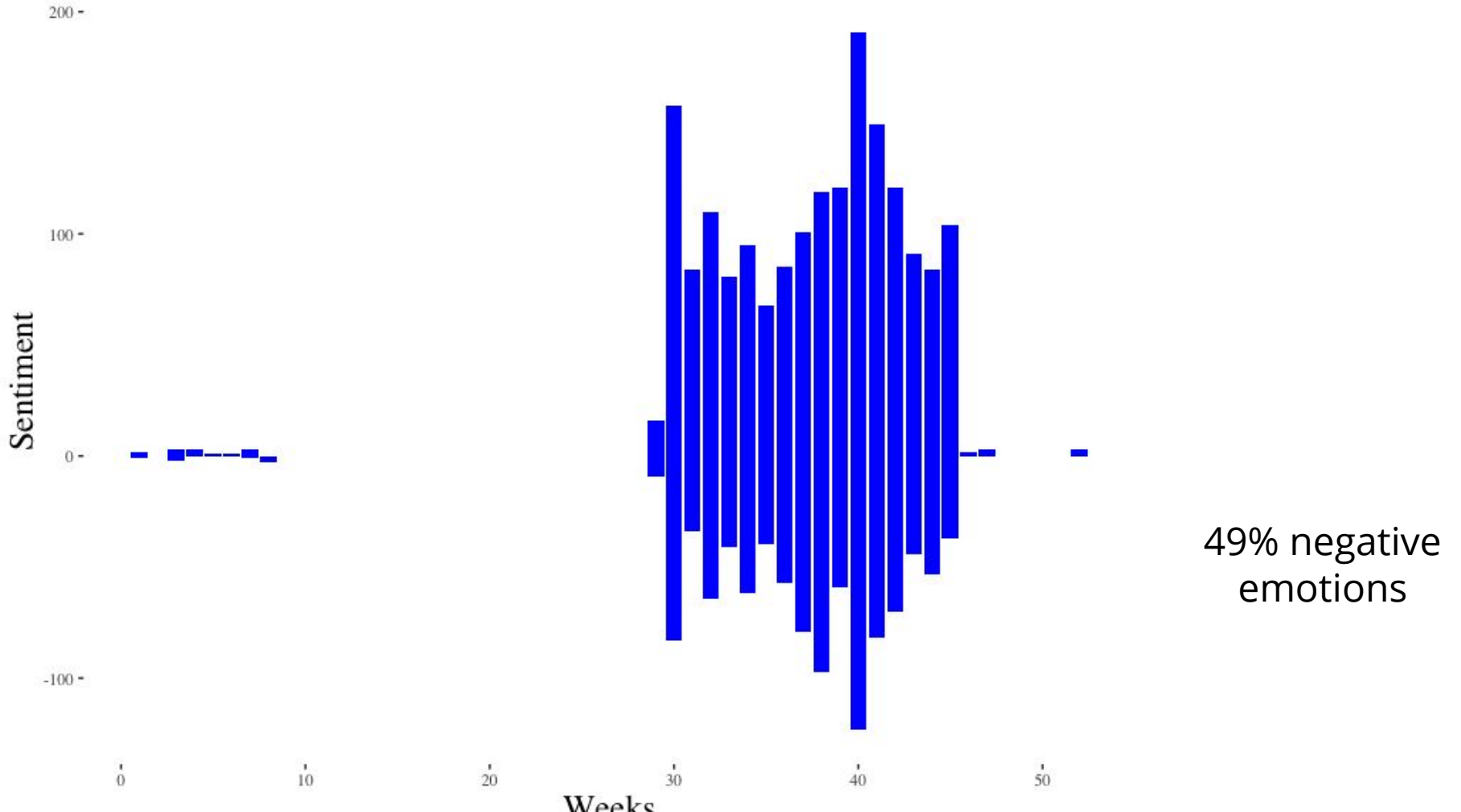
When @realDonaldTrump mentioned @HillaryClinton he got

29,258 favourites per tweet

Emotions - Mr. Trump



Emotions - Ms. Clinton



Typo Hunting

- On average people will have 4 mistakes per 100 words [1,2]
- Assuming binomial distribution (each word can have a typo) with

$$P = 0.04,$$

Expected number of typos in 3200 tweets is 128 ($sd = 11.09$)

- How often would presidential candidates make spelling errors?

[1] Matias, E., MacKenzie, I. S., & Buxton, W. (1996)

[2] <https://analytics.twitter.com>

Caveat: how to find spelling errors?

1. Remove all *URLs*, *hashtags* and *usernames* from the tweets
2. Acquire large dictionary (we used list of 30,000 words on Google's "Project Gutenberg")
3. Find Levenshtein distance between tweet t , and each dictionary word s :
measure of the similarity between two strings
number of deletions, insertions, or substitutions required to transform t into s

Null Hypothesis:

Each user will have between 95-161 typos in 3200 tweets
(i.e. # of typos will be within 3SD of $\mu=128$)

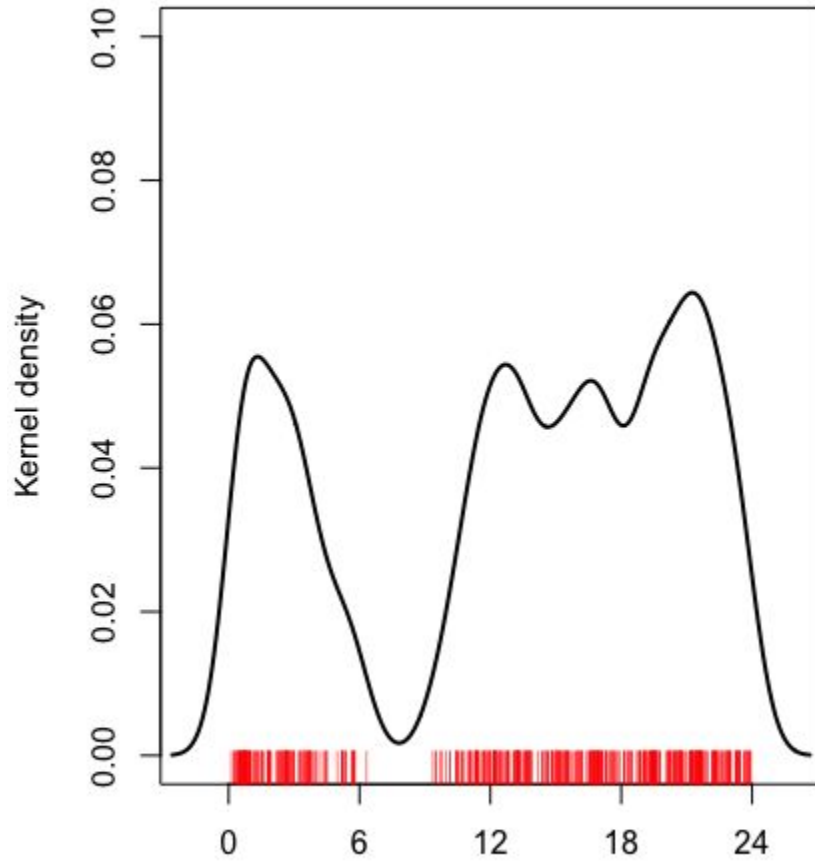
Reality

Neither Ms. Clinton, nor Mr. Trump had any typos
($p = 0$)

→ we reject the null hypothesis

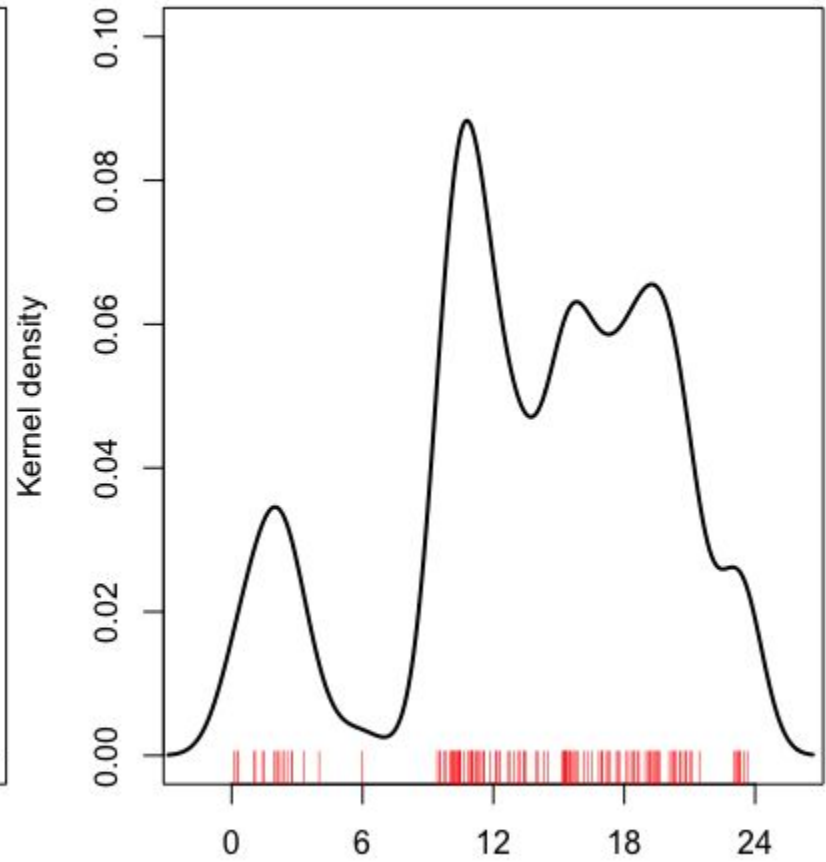
When is Trump tweeting?

30 days prior to election



Tweet timestamps, 511 in total

30 days after election

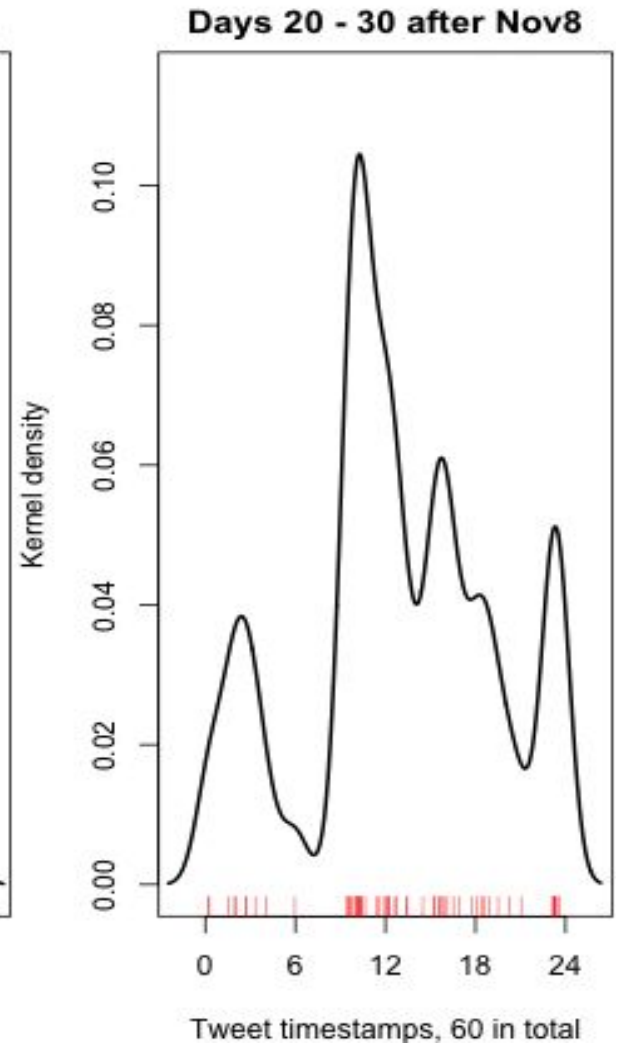
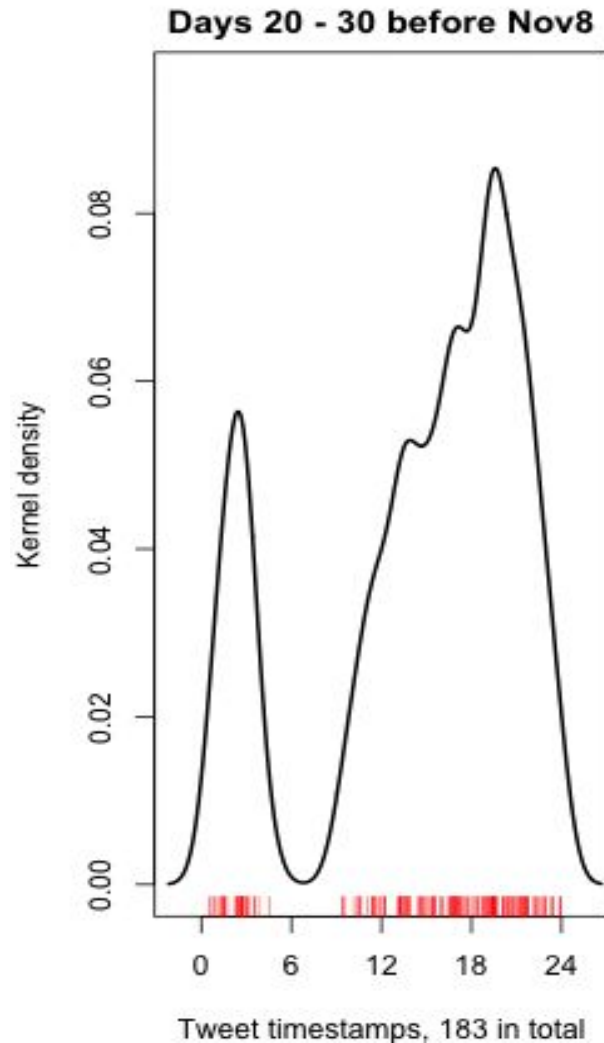


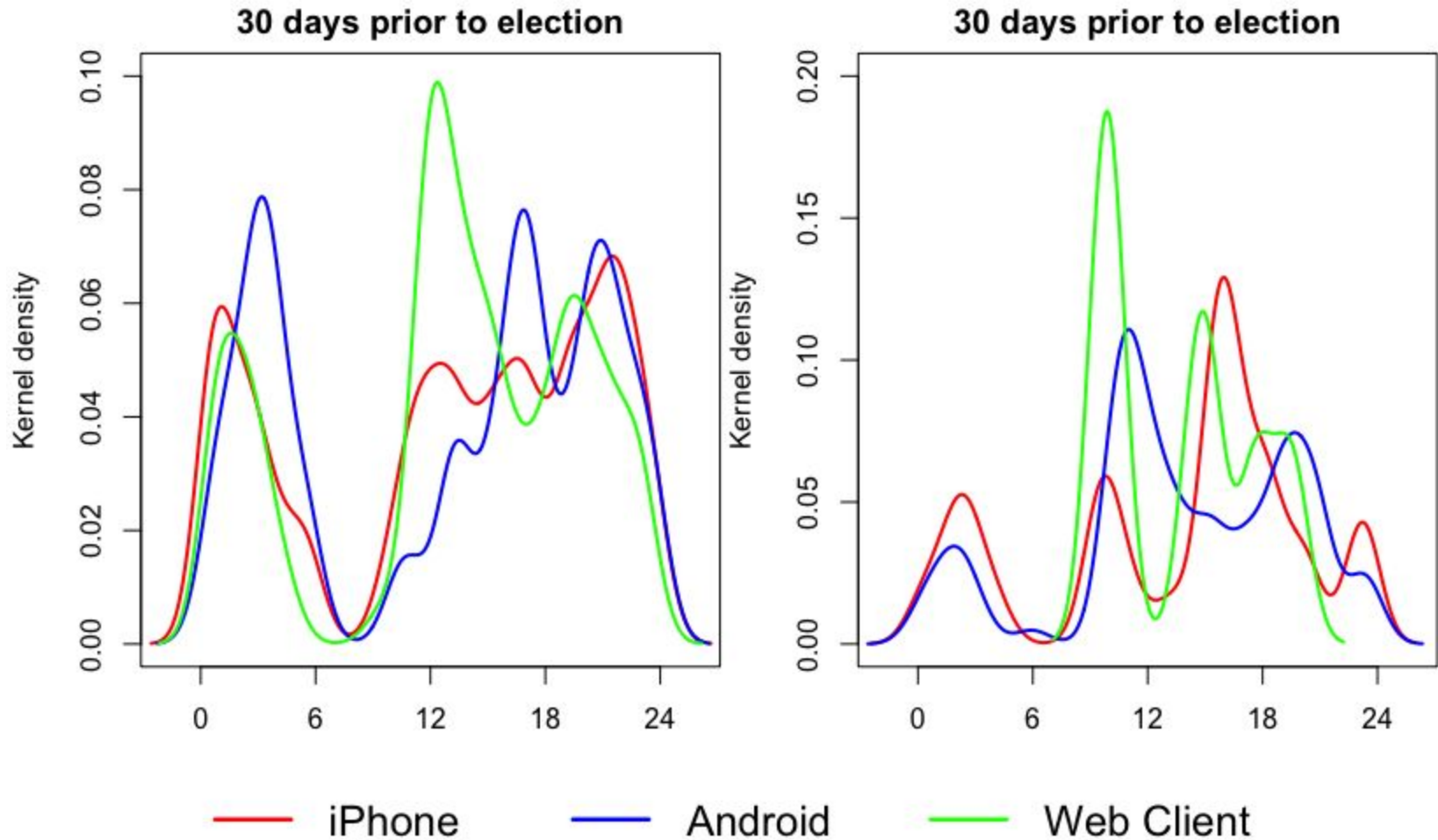
Tweet timestamps, 132 in total

Densities of tweet timestamps before and after election - over 30 day window.

Densities of tweet timestamps before and after election - computed over a 10 day sliding window, moved 5 days at a time.

Tweeting occur between noon and 9pm after the election... also less late-night tweeting.





**Densities of tweet timestamps before and after election
- over 30 day window for 3 most frequently used sources.**

Conclusions

- Harder to interpret than we thought
 - Small amount of tweets post-election
 - Tweeting behaviour might not be as predictable as we thought
- What we would do next?
 - Compare word distribution from different sources
 - Can we use them to predict authorship?
 - Can we find whether the candidate's PA is tweeting for them?
 - Can we use sentiment analysis to predict important events?