

# ANALYSIS OF THE WORDS OF THE 2016 PRESIDENTIAL CANDIDATES

---

Erica Kim

July 27, 2016

*presentation for Schireson Associates*

# Candidates' words: rich dataset

- Hillary Clinton has given ~500 formal speeches (dating back to May 1969)
- Donald Trump has given ~40 formal speeches, statements, and interviews (dating back to March 2013)
- Both candidates use twitter extensively



# Why do we care?

- Words have meaning
  - *Convey priorities and intentions*
- What issues is a candidate focusing on?
  - *Is this an issue that we care about?*
  - *Does the candidate's stance represent our view?*
- What audience is a candidate seeking to attract?
  - *Is there an alignment with us?*

# Lots of words, so little time...

- How can we streamline the process of analyzing everything the candidates say?

# Solution: Use data science to efficiently answer these questions

*For speeches:*

- Scrape speech transcripts from various websites
- Create text corpus for each candidate
- Tokenize, stem, and vectorize words
- Group speeches using unsupervised clustering algorithm

# Example of analysis of Clinton's speeches:



## Hillary's Top 10 Words

health & health care  
school  
education  
government  
children  
unite/united  
vote  
parents  
security  
issue

# Example of analysis of Clinton's speeches

(continued)

## *9 categories of speech topics:*

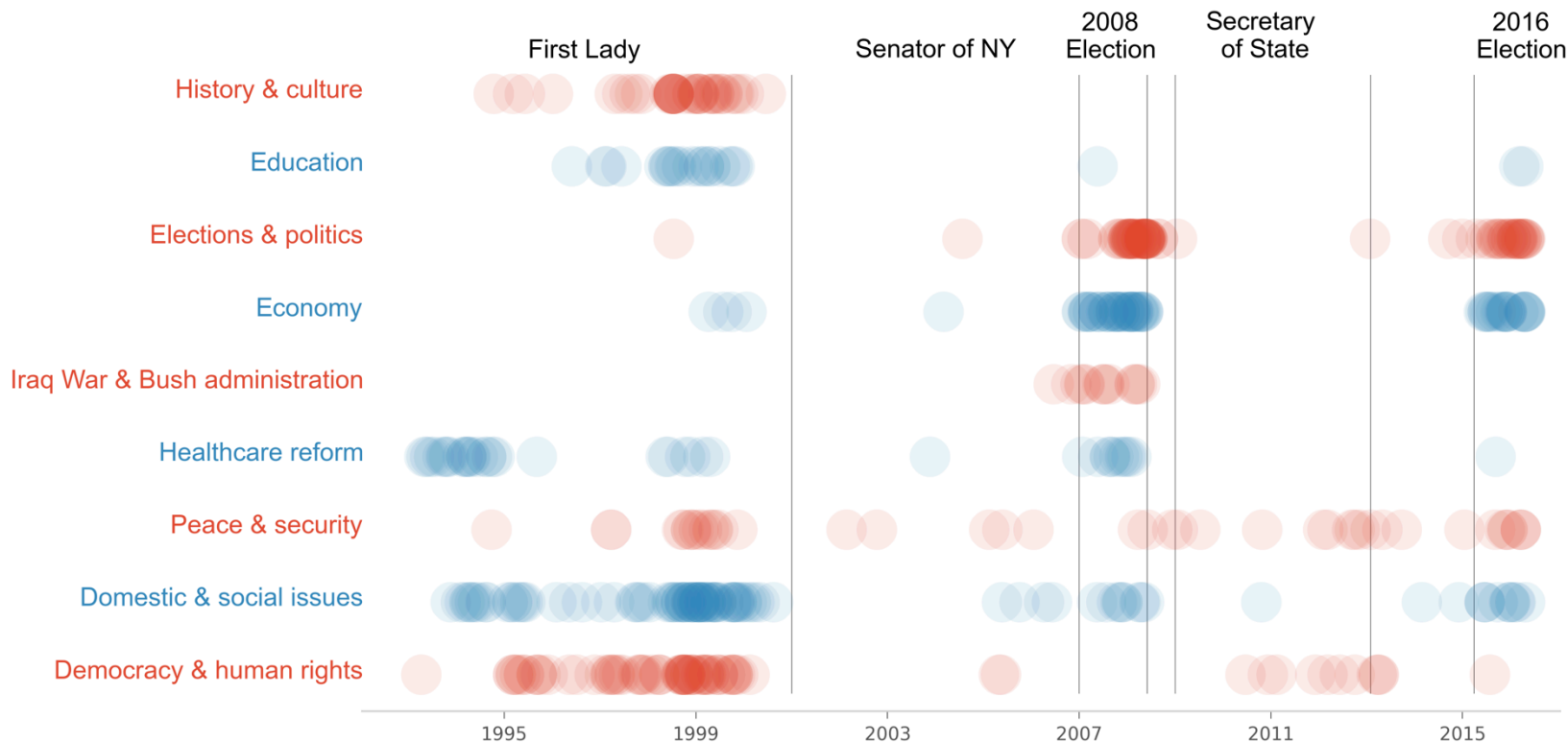
- History, culture, and honor
- Education
- Elections & politics
- Economy
- Iraq War & the Bush Administration
- Healthcare reform
- Peace & security
- Domestic & social issues
- Democracy & human rights

# Example of analysis of Clinton's speeches:

(continued)

## Clinton's speech topics change with time

Each speech is plotted by its date and category, with 490 speeches plotted in total. As First Lady, the themes of her speeches were across all of the topics, except: elections & politics (she wasn't running a campaign), the economy (the Clinton years coincided with boom times), and anything relating to the Iraq War/Bush Administration (these events had not occurred yet). Clinton's speech topics during the current 2016 election campaign are generally consistent with those from the 2008 election campaign - elections & politics and the economy are the dominating themes of speech in both periods. However, while much of her focus during the 2008 election campaign was additionally on the Iraq War/Bush Administration and healthcare reform, the current 2016 election campaign is more focused on the theme of general peace & security.

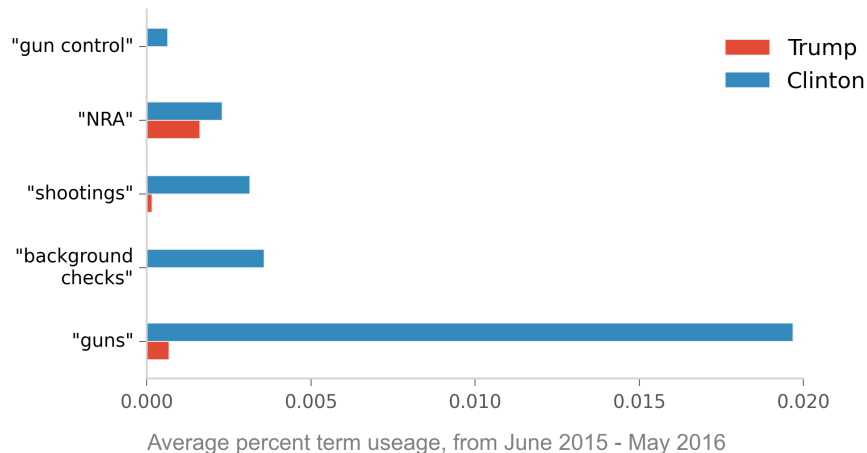




# Example of analysis of both candidates' speeches:

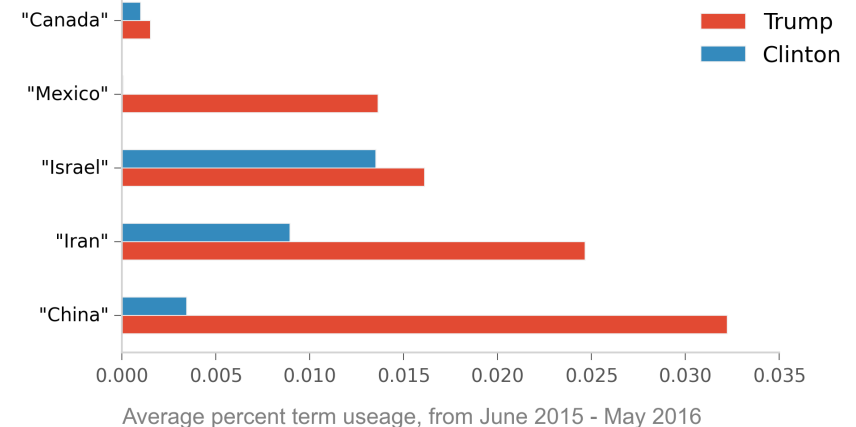
## Gun control: Proving their left-wing beliefs?

Gun control has been a hot debate topic for the Democratic primary race, as Clinton is a staunch supporter of tighter national gun control laws, while Bernie Sanders prefers a more hands-off approach. Because this is a decisive issue between the two Democratic candidates, Clinton frequently speaks about it.



## Candidates focus on different countries

Trump references other countries much more than Clinton; however, while Trump mentions Iran and China the most (as military and economic adversaries), Clinton is most focused on the U.S.'s special relationship with Israel.



# Twitter



**Hillary Clinton** ✓

@HillaryClinton

Wife, mom, grandma, women+kids advocate, FLOTUS, Senator, SecState, hair icon, pantsuit aficionado, 2016 presidential candidate. Tweets from Hillary signed -H

📍 New York, NY

🔗 [hrc.io/signup](http://hrc.io/signup)



**Donald J. Trump** ✓

@realDonaldTrump

📍 New York, NY

🔗 [DonaldJTrump.com](http://DonaldJTrump.com)

📅 Joined March 2009

✉ **Tweet to Donald J. Trump**

# Tweets: different data, different tools

## *Example tweets:*

- “I like Michael Douglas!”
- “Wow, big lines in Kansas”
- “Enough. –H <https://t.co/3her06i1Pr>”
- “President Trump?”
- “Delete your account.”

## *Processing pipeline:*

- Collect tweets using Tweepy (python library based on Twitter api)
- Create custom lexicon for specific issues

# Example of analysis of tweets

## Most frequent bigrams:

*Donald Trump:*

- “Ted Cruz”
- “Hillary Clinton”
- **“Crooked Hillary”**
- “Donald Trump”
- “South Carolina”
- “New Hampshire”
- “New York”
- “Marco Rubio”
- **“beat Hillary”**

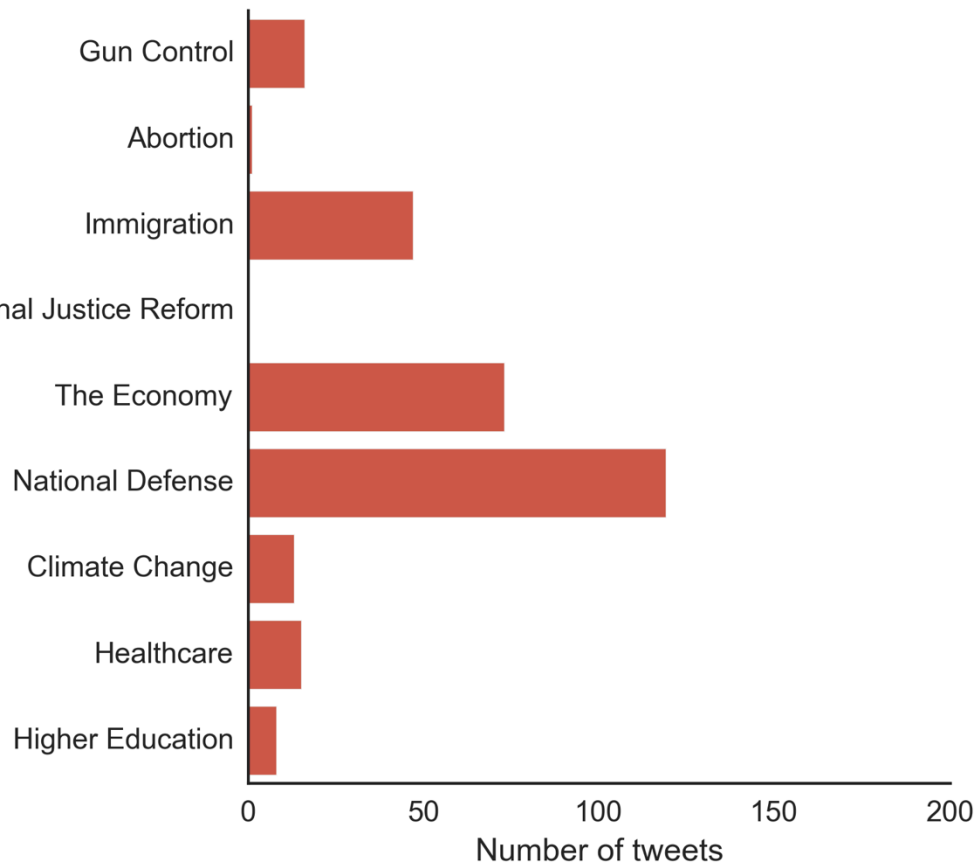
*Hillary Clinton:*

- “Donald Trump”
- “Hillary Clinton”
- **“gun violence”**
- **“health care”**
- “White House”
- “New Hampshire”
- “New York”
- **“Supreme Court”**
- **“foreign policy”**

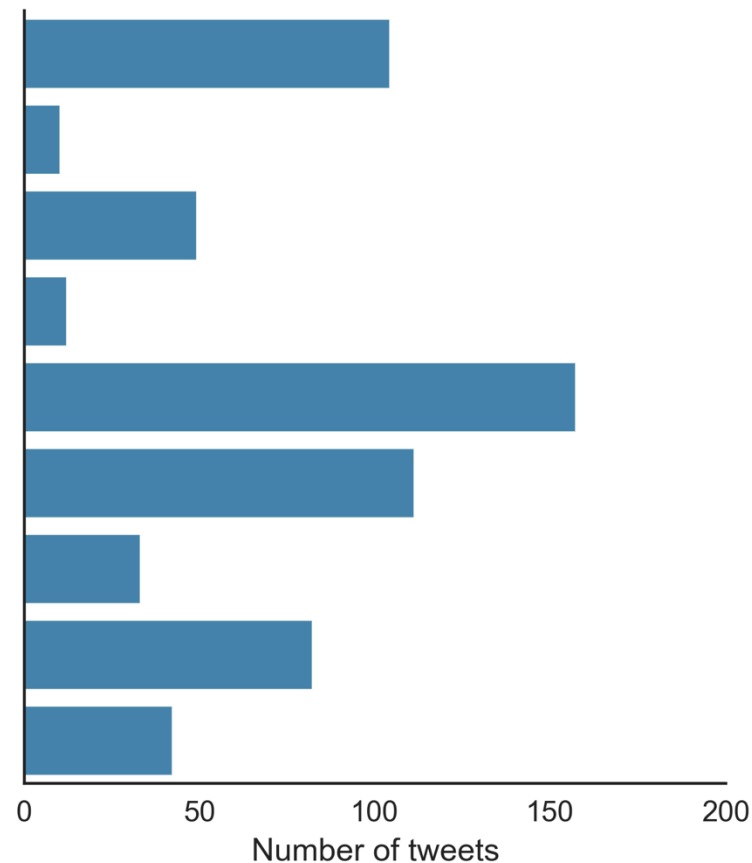
# What issues are the candidates tweeting about?

How often do candidates tweet about election issues?

Donald Trump's tweets



Hillary Clinton's tweets

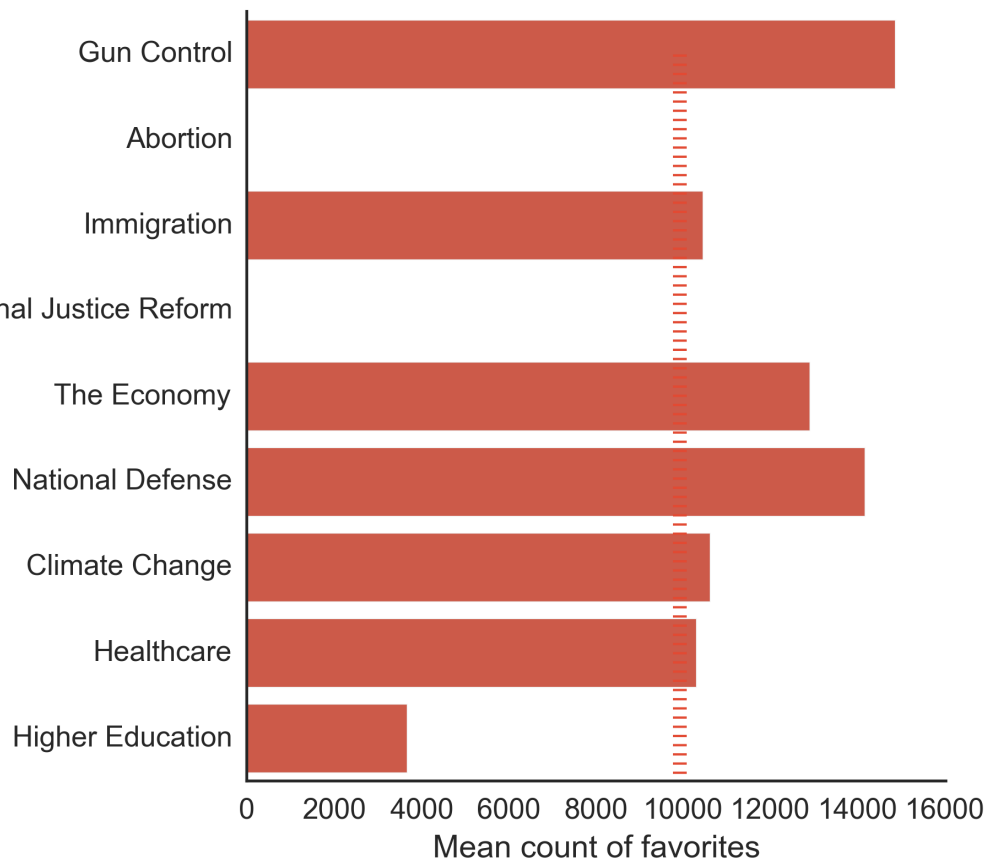


# And how are twitter followers responding?

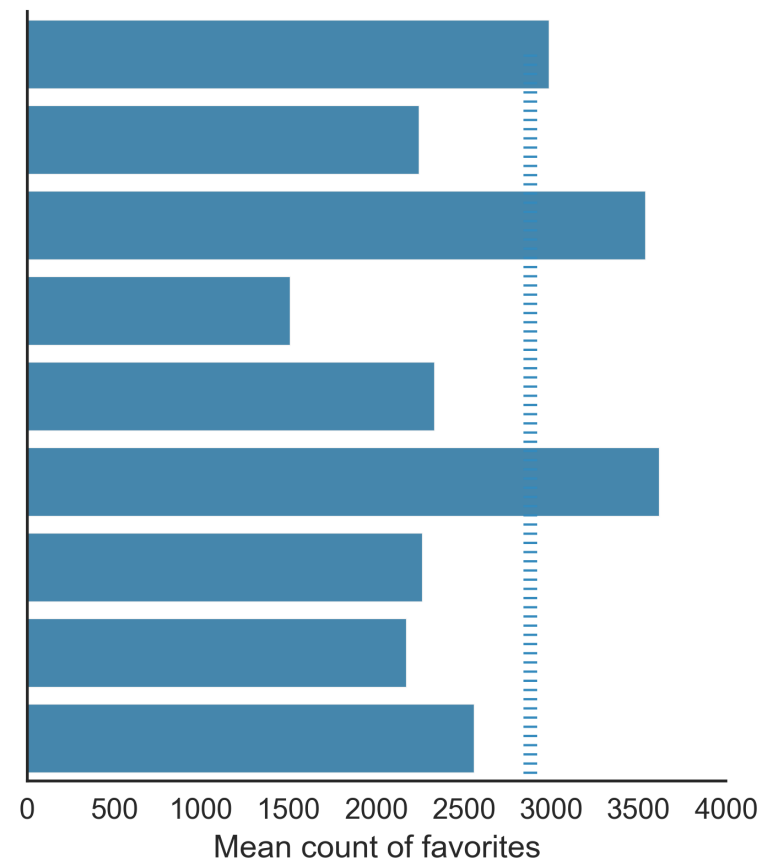
## *User engagement based on tweet topic:*

How do twitter users respond to the issues?

Donald Trump's tweets



Hillary Clinton's tweets



# Conclusion

- Quickly determine:
  - What issues the candidates are talking about
  - How often they are talking about these issues
  - How people are responding to these issues
- **...without having to read a single word**
- Networks doing political coverage
  - Streamline process of analyzing news sources
- Lobbies, interest groups
  - *Is my donation worth it?*
  - *What can I expect, come January?*