

Charlie Jarvis  
Geraldine Moriba  
Theodora Boulouta  
Flor Coelho

Data Mining  
Cable News  
Transcripts

The goal of this project is to analyze cable television news to identify patterns and trends in content, bias and coverage.



# Methodology

Lexicon  
generation:  
**which words** tell  
us that the  
topic is being  
discussed?



Transcript  
segmentation:  
**where** is the  
topic being  
discussed?

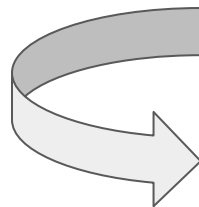


Analysis:  
**how** is the topic  
being addressed?

Determine sentiment, bias,  
and frequency of discussion

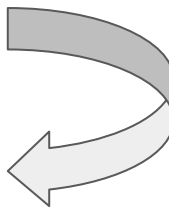
seemingly slick phoenix illegal outline postpone  
ban hawk stance anti curb  
visa reform immigration comprehensive ice lawful fiery plan  
trade flopping soft topic pillar crop  
proposal construction manpower

Iterative Process



code refinement

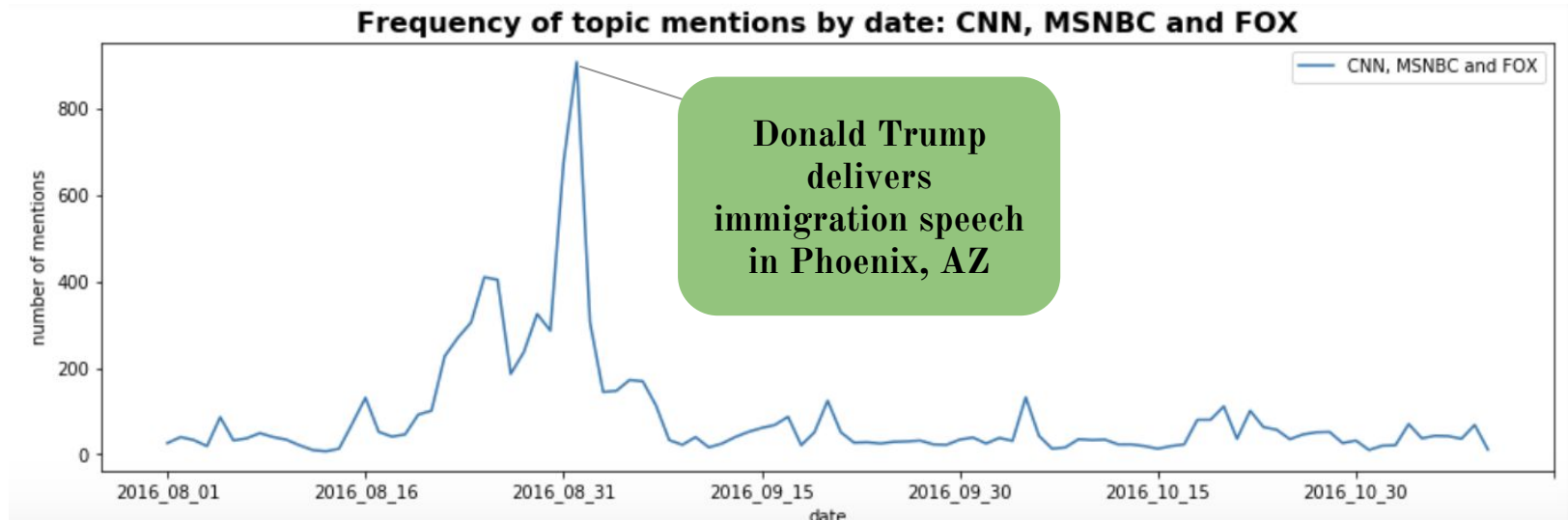
manual verification



# Coverage Frequency

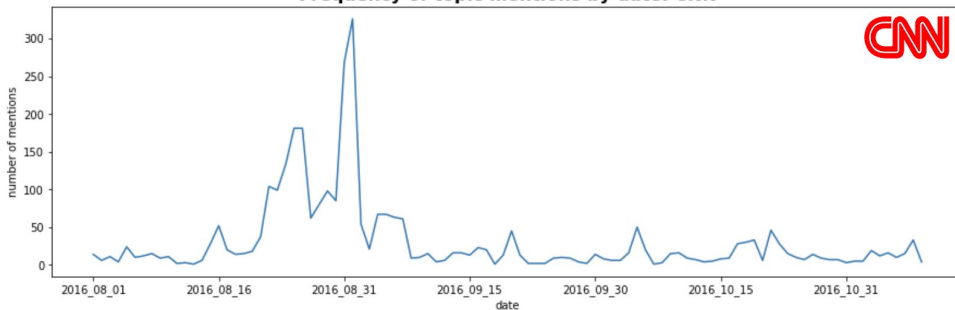
When did immigration coverage  
spike in 2016?

# Immigration: 2016

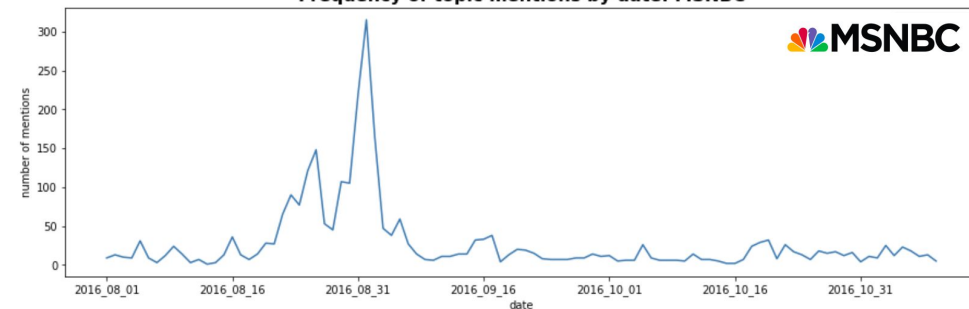


In 2016, the highest frequency of coverage occurred towards the end of August.

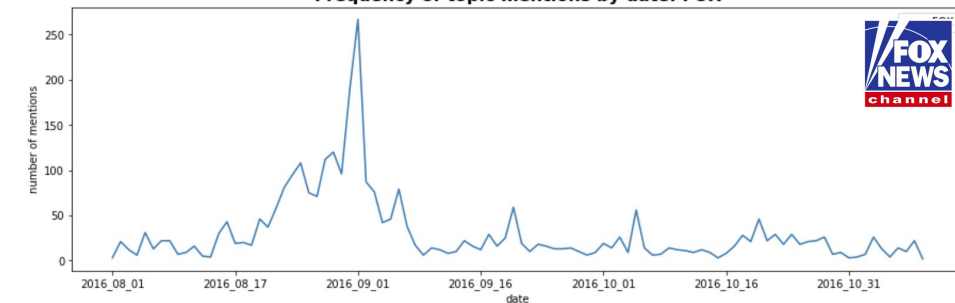
Frequency of topic mentions by date: CNN



Frequency of topic mentions by date: MSNBC



Frequency of topic mentions by date: FOX



## Top Immigration Coverage Dates (2016)

### FOX:

2016\_09\_01 : 267  
 2016\_08\_31 : 189  
 2016\_08\_29 : 120  
 2016\_08\_28 : 112  
 2016\_08\_25 : 108

### CNN:

2016\_09\_01 : 326  
 2016\_08\_31 : 268  
 2016\_08\_25 : 181  
 2016\_08\_26 : 181  
 2016\_08\_24 : 133

### MSNBC:

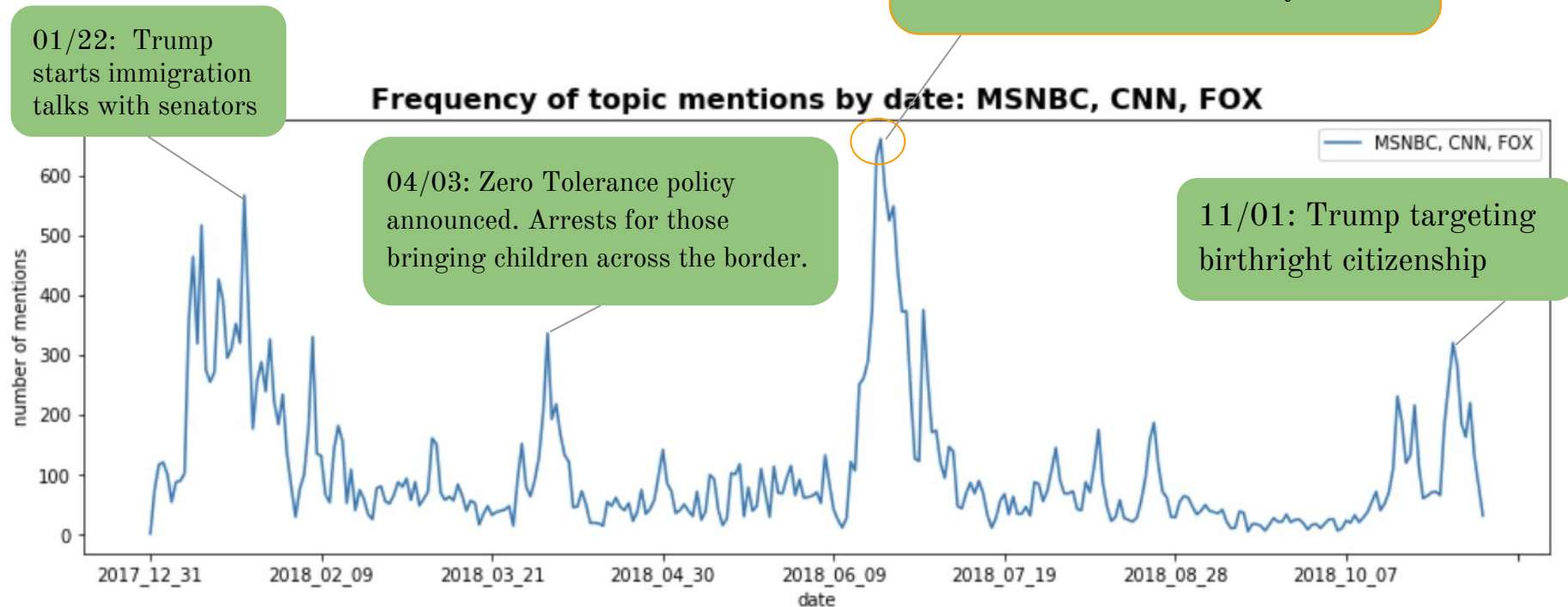
2016\_09\_01 : 315  
 2016\_08\_31 : 218  
 2016\_09\_02 : 165  
 2016\_08\_26 : 148  
 2016\_08\_25 : 121

The pattern of coverage of immigration is relatively consistent between each cable news channel.

# What about 2018?

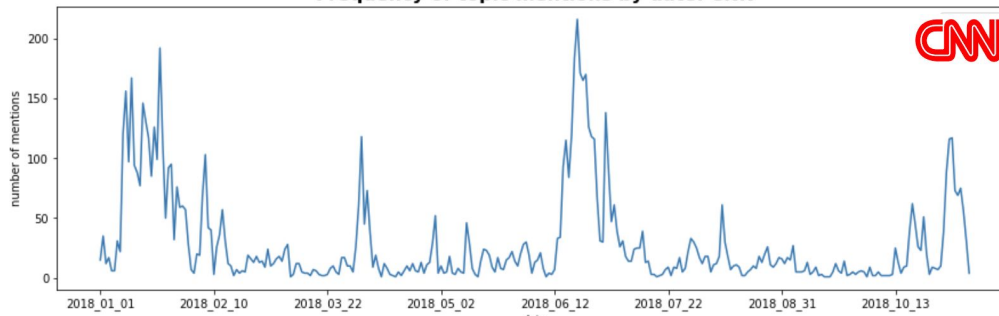
When did immigration receive the most coverage?

# Immigration: 2018

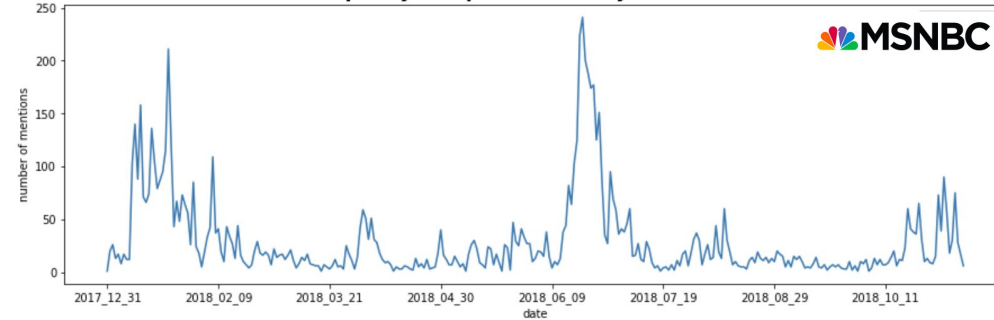




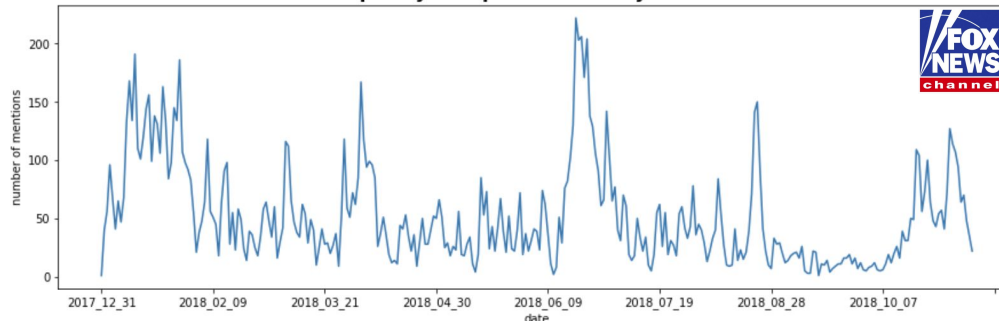
Frequency of topic mentions by date: CNN



Frequency of topic mentions by date: MSNBC



Frequency of topic mentions by date: FOX



## Top Immigration Coverage Dates (2018)

FOX top scores:

2018\_06\_19 : 222

2018\_06\_21 : 206

2018\_06\_23 : 204

2018\_06\_20 : 203

2018\_01\_12 : 191

CNN top scores:

2018\_06\_20 : 216

2018\_01\_22 : 192

2018\_06\_19 : 184

2018\_06\_21 : 171

2018\_06\_23 : 170

MSNBC top scores:

2018\_06\_20 : 241

2018\_06\_19 : 224

2018\_01\_22 : 211

2018\_06\_21 : 200

2018\_06\_22 : 188

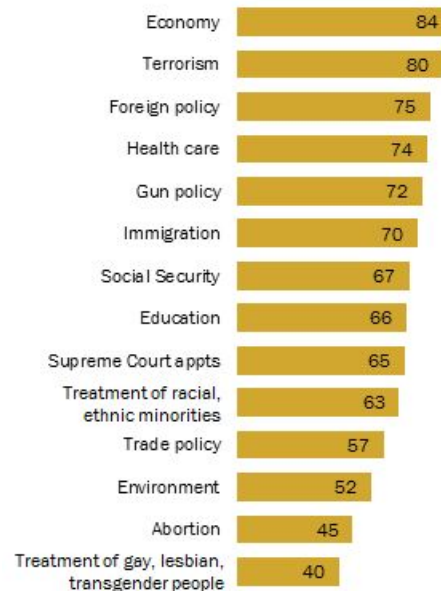
Peak coverage dates of immigration are similar across all cable news channels, but **FOX News** has the largest pattern of coverage.

# What about the Coverage Frequency of Other Election Topics?

The economy is one of the top ten election topics determined by PEW Research Institute voter polling results.

## Economy and terrorism are top issues for voters in 2016

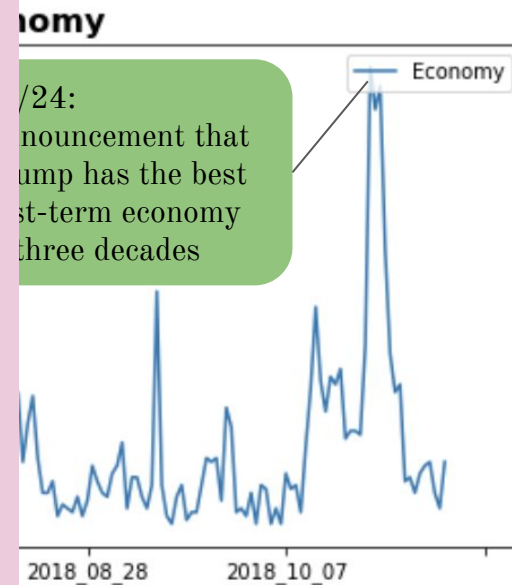
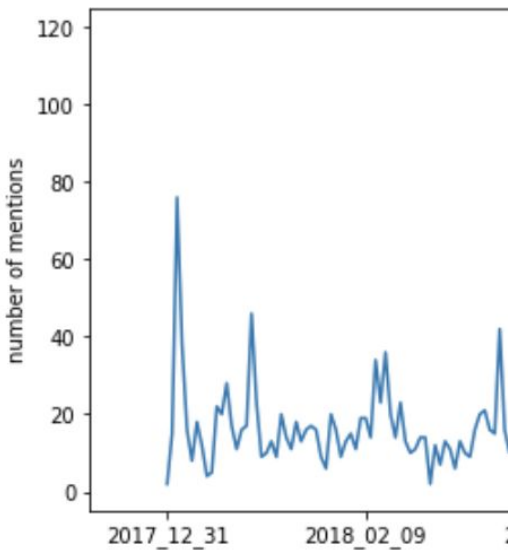
*% of registered voters saying each is "very important" to their vote in 2016 ...*



Note: Based on registered voters. Q40.  
Source: Survey conducted June 15-26, 2016.

PEW RESEARCH CENTER

# The Economy

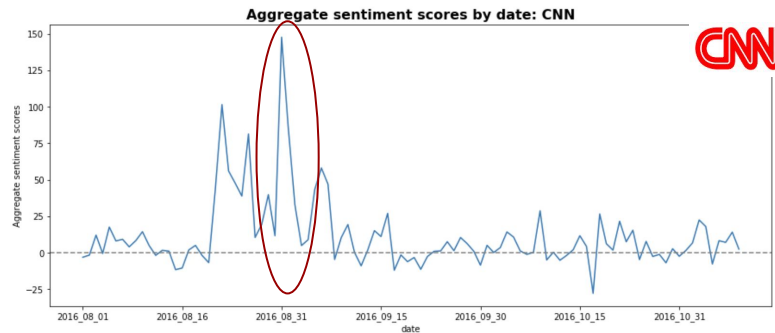
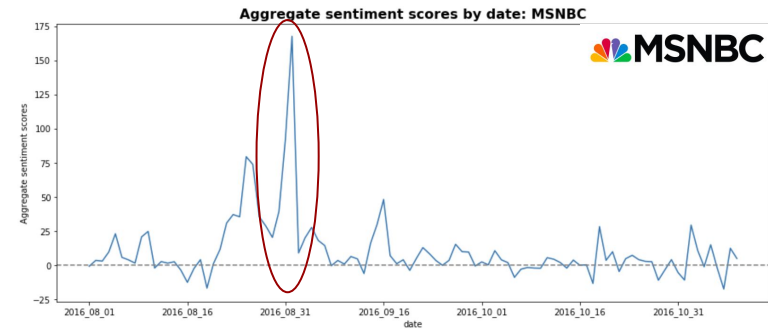
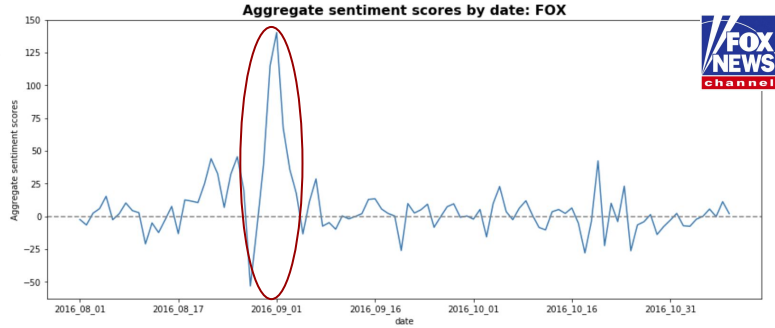


# Analysis of Immigration as a News Topic

*How did cable news cover this topic?*

# Sentiment Analysis

2016

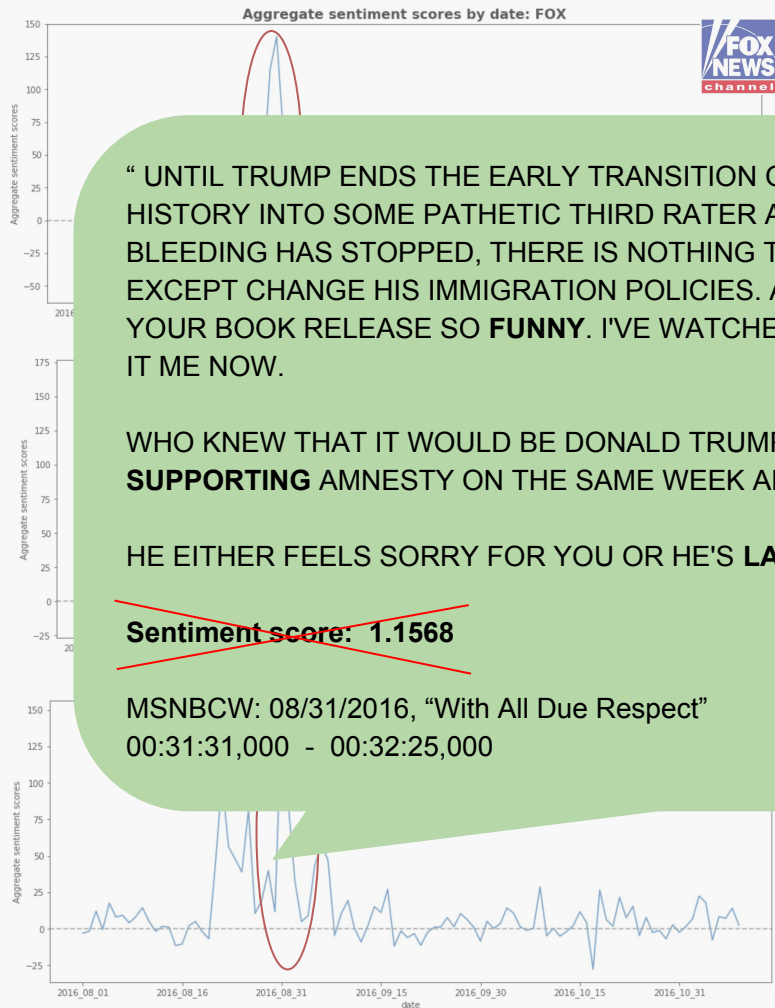


Fox has the most negative sentiments in relation to immigration.

All three cable networks share the highest sentiments scores on the same date. What happened on that day?

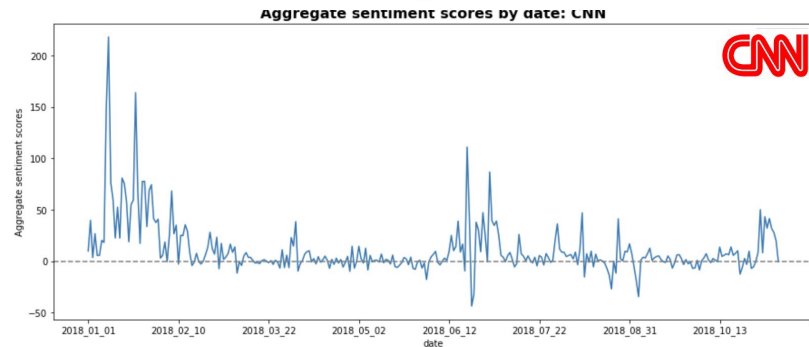
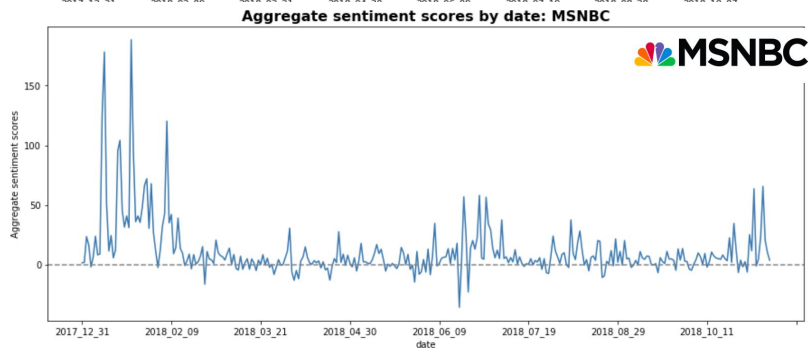
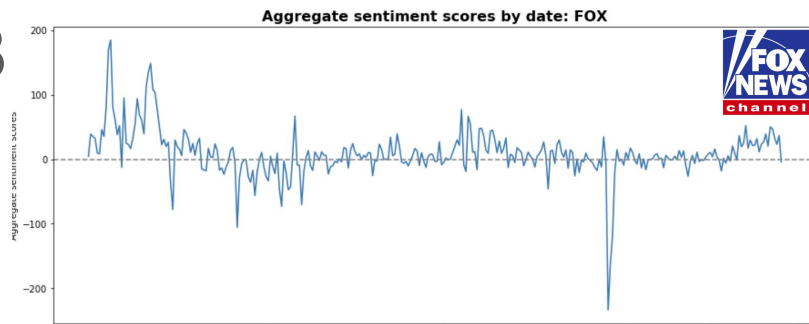
*'Donald Trump's Immigration Speech*

2016



## *Donald Trump's Immigration Speech*

# 2018



Sentiment spikes (positive or negative) = high emotion moments

Most positive sentiment scores in all channels are on **January 9 and 10th**.  
What happened on these days?

Trump Calls For 'Bill Of Love' Allowing DACA Recipients To Remain

January 9, 2018 · 3:52 PM ET

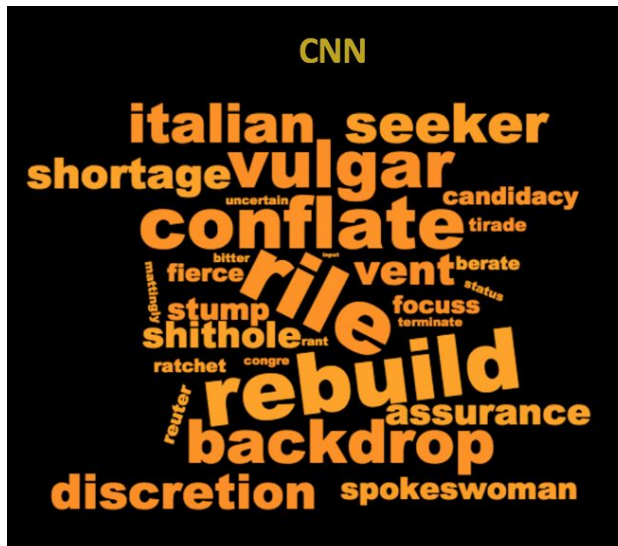
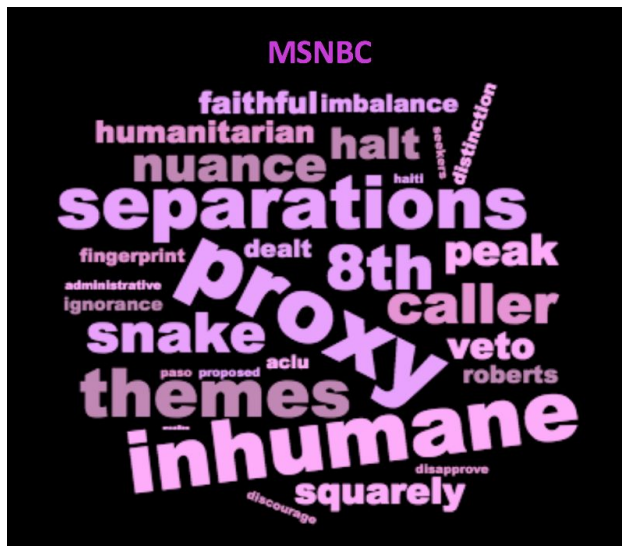


# Political Bias Analysis

A more nuanced approach

# A Sample Of AI Generated Immigration Related Words Frequently Used By Each Cable News Network Exclusively (2018)

Formula: Only MSNBC = MSNBC frequently used words - CNN frequently used words - FOX News frequently used words

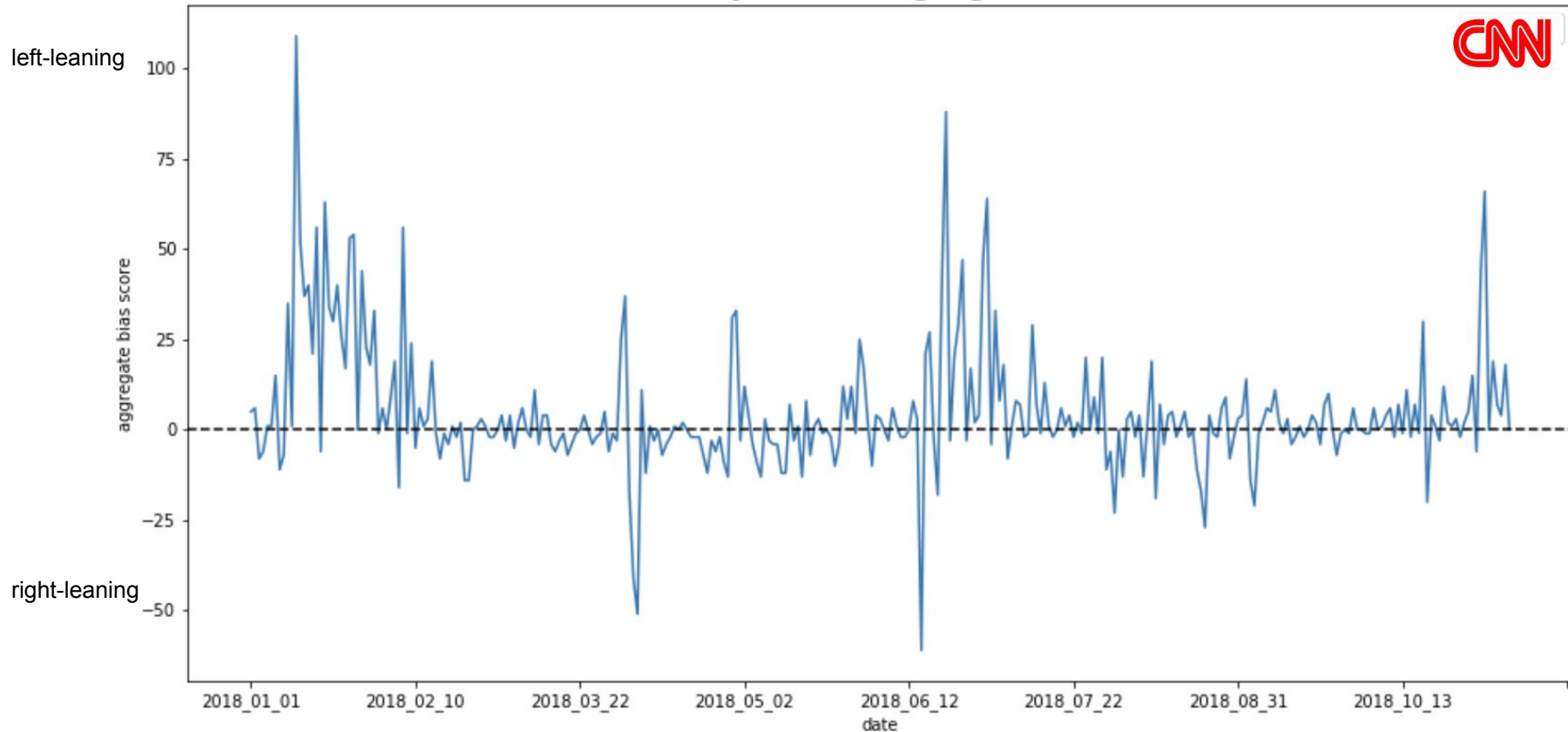


# High Inference Language Lexicons (Sample)

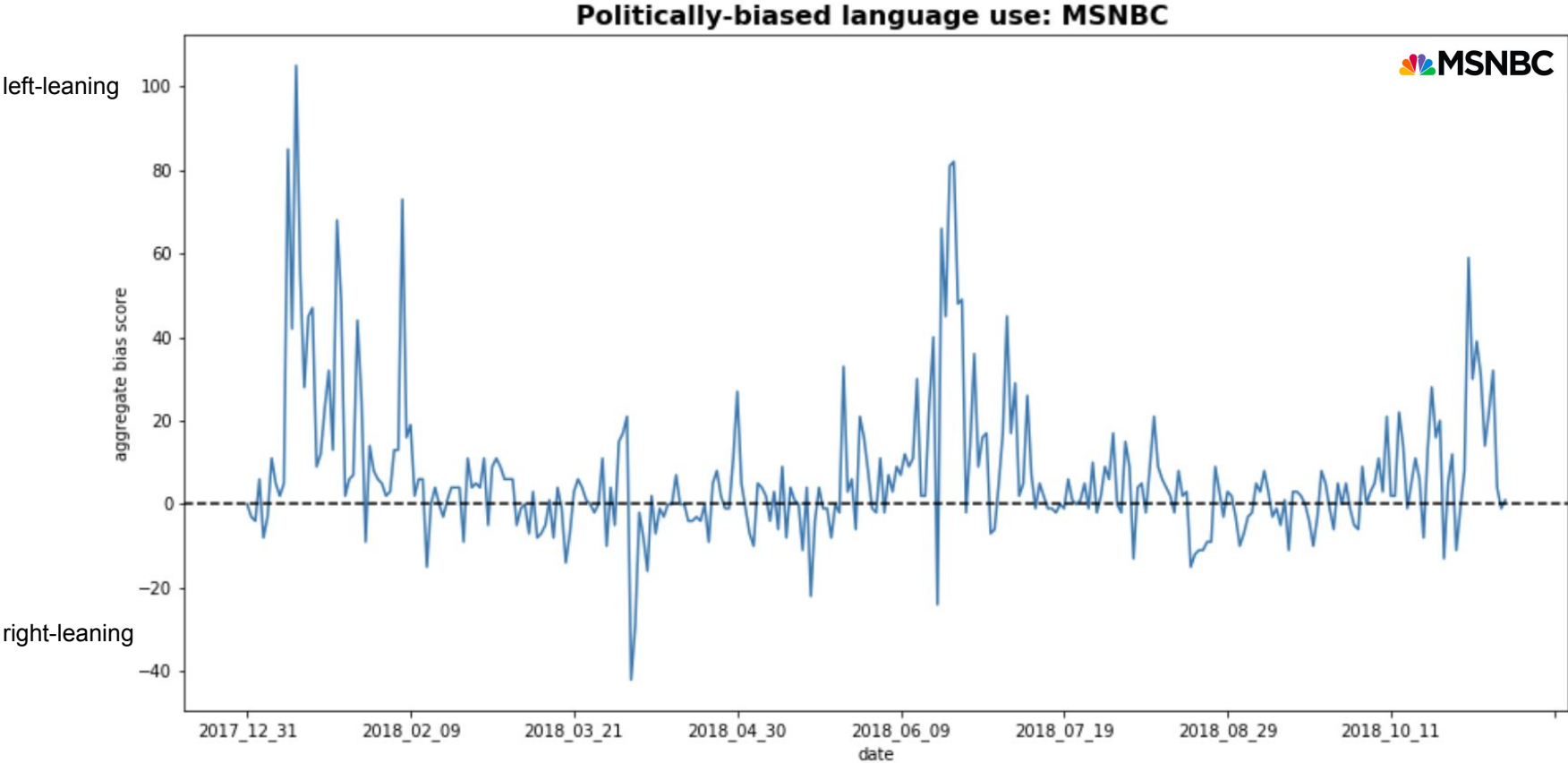
Right biased wording when used with immigration	Left biased wording when used with immigration
Abused	Abiding
Acquitted	Accountability
Aggressive	Activate
Alien	Advisors
Allows	Advocacy
America first	Aggressively
Winning again	Alienate
America safe	Alt-right
American lives	Amnesty
American worker	Anthem
Anchor baby	Anti-Hispanic
Apprehend	Anti-immigrant
Apprehended	Anti-Latino
Architect	Anxiety
Arrest	Approach
Assaulting	Assurance
Avalanche	Asylum seekers

# Bias Scores

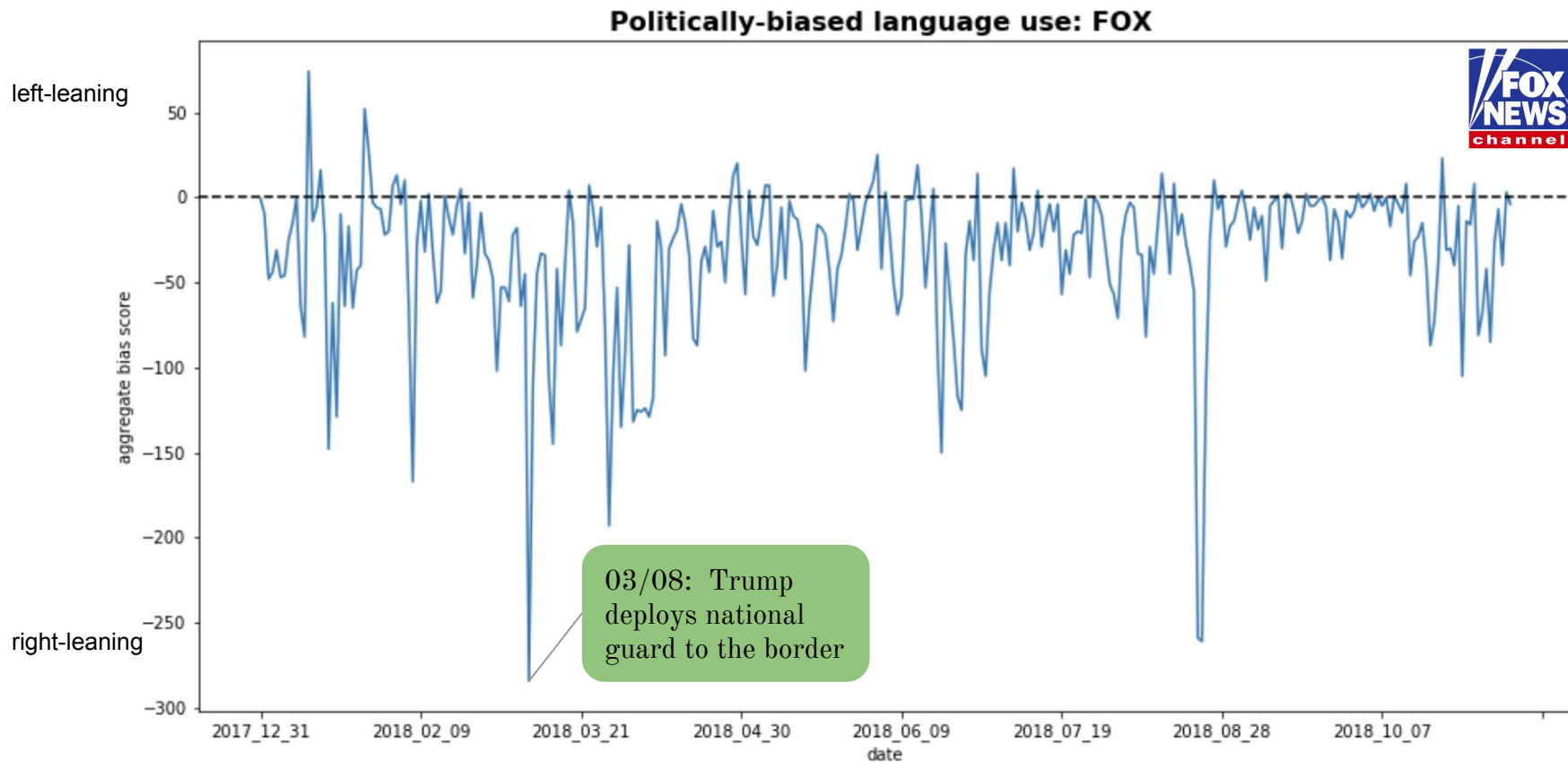
## Politically-biased language use: CNN



# Bias Scores



# Bias Scores



# What We Learned

## What works

- Extracting information / identifying topic segments based on word frequencies
- Identifying high emotion moments using sentiment analysis
- Using bias analysis to determine trends in political-leanings
- Computer-aided comparisons of coverage patterns of different topics is possible

## The Challenges

- Sentiment analysis has limitations
- Bias lexicons are difficult to create
  - Both sides often use the same words
  - Manually-generated lexicons necessary for good results
- Manual lexicons are influenced by the biases of the human programmer
- Lemmatization makes the process slow

A 3D rendering of a complex maze made of white walls. The maze features numerous doors in various colors: green, blue, orange, brown, red, and purple. Some doors have windows or circular portholes. A person in a purple and white striped shirt is visible in one of the maze's openings, looking at a blue door. The scene is brightly lit, casting shadows on the floor.

End.

Special Thanks:  
Will Crichton  
CS206 Teaching Team