



POLITICAL RHETORIC ANALYSIS

Emma Tarcson



PROJECT PLAN PT.I

- Wanted to focus on rhetorical aspects
- Changed -> focus on modes of persuasion with Republican and Democratic speech/text data
 - “Hot Topic” emphasis
 - Use of descriptive words

Why so political?

BACKGROUND



VICE PRESIDENTIAL DEBATE



2020



THE UNIVERSITY OF UTAH



CHARLOTTE

RNC2020

2020 DEMOCRATIC
NATIONAL
CONVENTION

MILWAUKEE, WISCONSIN

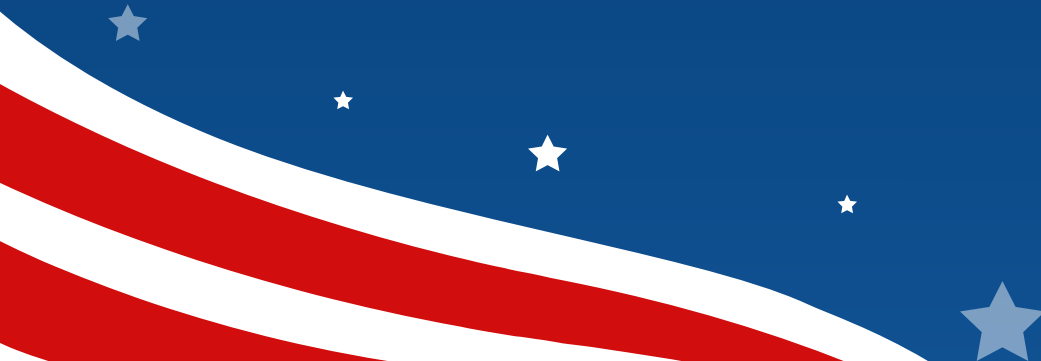
2020 PRESIDENTIAL
DEBATE

★★★★ BELMONT UNIVERSITY

WHAT IS POLITICAL RHETORIC?

1.) How politicians use language features to further persuasion

2.) Synonymous with “Agenda” -> What is the Politician trying to push? What issues does the politician care about?





DATA COLLECTION

1. RNC and DNC speeches
2. Presidential debate and VP debate
3. Party Platforms

2020 Democratic Convention Speeches
Top speeches from the 2020 Democratic Convention (NLP)

Christian Lilliefeld • updated 8 months ago (Version 1)

Data Tasks Code Discussion Activity Metadata Download (108 KB) New Notebook

Usability 8.8 License CC0: Public Domain Tags social science, politics

Description

Context

The 2020 Democratic National Convention was a presidential nominating convention that was held from August 17 to 20, 2020, at the Wisconsin Center in Milwaukee, Wisconsin, and virtually across the United States. At the convention, delegates of the United States Democratic Party formally chose former Vice president Joe Biden and Senator Kamala Harris of California as the party's nominees for president and vice president, respectively, in the 2020 United States presidential election.

2020 Republican Convention Speeches
Top speeches from the 2020 Republican Convention (NLP)

Christian Lilliefeld • updated 8 months ago (Version 4)

Data Tasks Code Discussion Activity Metadata Download (247 KB) New Notebook

Usability 10.0 License CC0: Public Domain Tags social science, politics, languages

Description

Context

The 2020 Republican National Convention was a presidential nominating convention in which delegates of the United States Republican Party selected the party's nominees for president and vice president in the 2020 United States presidential election, held from August 24–27, 2020.

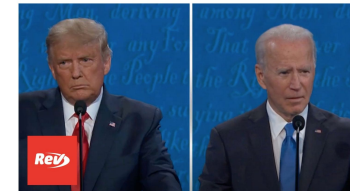
Kamala Harris & Mike Pence 2020 Vice Presidential Debate Transcript



Revy • Blog • Transcripts • Debate Transcripts • Kamala Harris & Mike Pence 2020 Vice Presidential Debate Transcript

Vice President Mike Pence and Senator Kamala Harris participated in the 2020 vice presidential debate on October 7, moderated by Susan Page of USA Today in Salt Lake City. Read the transcript of the full debate here.

Donald Trump & Joe Biden Final Presidential Debate Transcript 2020



Revy • Blog • Transcripts • 2020 Election Transcripts • Donald Trump & Joe Biden Final Presidential Debate Transcript 2020

Full transcript of the second and final 2020 presidential debate between Donald Trump and Joe Biden. The debate was moderated by Kristen Welker of NBC News on October 22, 2020 in Nashville, Tennessee.

The American Presidency Project

DOCUMENTS STATISTICS MEDIA ARCHIVE PRESIDENTS ANALYSES

Documents

ARCHIVE GUIDEBOOK CATEGORIES ATTIBUTES

Categories

PRESIDENTIAL (25668) ▼
PRESIDENTIAL (1976) ▼
ELECTIONS AND TRANSITIONS (2026) ▼
MISCELLANEOUS (64) ▼
CONGRESSIONAL (3)

2020 Democratic Party Platform

August 17, 2020

DEMOCRATIC NATIONAL CONVENTION LAND ACKNOWLEDGEMENT

The Democratic National Convention wishes to acknowledge that we gather together in our own values on land that have been stewarded through many centuries by the ancestors and descendants of Chief Nations who have been here since time immemorial.

We honor the communities native to this country, and recognize that our country was built on Indigenous homelands.

We pay our respects to the millions of Indigenous people throughout history who have protected our lands, waters, and animals.

2016 Republican Party Platform

July 18, 2016

We declare this platform with veneration and gratitude

To all who stand strong in the face of danger
So that the American people may be protected against it —
The men and women of our military,
of our law enforcement, and the true responders
of every community in our land —
And to their families.



Note: Removed certain words for platform



DATA COLLECTION PT.2

RNC and DNC speeches:

- downloaded from kaggle
- Platforms:
- downloaded from The American Presidency Project
- Pres and VP Debates:
- scraped from REV -> CSV files

```
banned = ['Republican', 'Democrat', 'Republicans', 'Democrats']  
f = lambda x: ' '.join([item for item in x.split() if item not in banned])  
platform["sentences2"] = platform["sentences"].apply(f)
```



Out[24]:

	transcript	Aff
Speakers		
Kamala Harris	Thank you, Susan. Well, the American people ha...	D
Kamala Harris	Can you imagine if you knew on January 28th, a...	D
Kamala Harris	... right to reelection based on this.	D
Mike Pence	Susan, thank you. And I want to thank the Comm...	R
Mike Pence	And I believe it saved hundreds of thousands o...	R
...
Mike Pence	I look at the relationship between Justice Rut...	R
Kamala Harris	First of all, I love hearing from our young le...	D
Kamala Harris	And brings me to Joe, Joe Biden. One of the re...	D
Kamala Harris	Joe has a longstanding reputation of working a...	D
Kamala Harris	Brecklin, when you think about the future, I d...	D

207 rows x 2 columns

- Created DataFrames and added political affiliations for all 4 sources

```
import scrapy  
  
class Presdebate20Spider(scrapy.Spider):  
    name = 'presdebate20'  
    #allowed_domains = ['https://www.rev.com/blog/transcripts/donald-trump-joe-biden-final-presidential-debate-transcript-2020']  
    start_urls = ['https://www.rev.com/blog/transcripts/donald-trump-joe-biden-final-presidential-debate-transcript-2020']  
  
    def parse(self, response):  
        #extracting using CSS selectors ?  
        text1 = response.xpath("//div[@class='fl-callout-text']/p").re(r'\n.\n?')  
        #text1 = response.xpath("//div[@class='fl-callout-text']/p/text()").extract()  
        time = response.xpath("//div[@class='fl-callout-text']/p/a/text()").extract()  
        speaker = response.xpath("//div[@class='fl-callout-text']/p").re(r'>[A-Z][a-z]+ [A-Z][a-z]+')  
  
        for item in zip(text1, time, speaker):  
            scraped_info = {  
                'speaker': item[2],  
                'time_stamp': item[1],  
                'transcript': item[0]  
            }  
  
            #give the info to scrapy  
            yield scraped_info
```

speaker	time_stamp	transcript
>Kristen Welker	00:18	Good evening, everyone. Good evening. Thank you so much for
>Donald Trump	07:37	How are you doing? How are you?</p>
>Kristen Welker	07:58	And I do want to say a very good evening to both of you. This de
>Kristen Welker	08:27	The goal is for you to hear each other and for the American peop
>Kristen Welker	09:03	... during this next stage of the coronavirus crisis. Two minutes u
>Donald Trump	09:04	So as you know, 2.2 million people modeled out, were expected
>Donald Trump	09:41	There was a very big spike in Texas. It's now gone. There was a v
>Donald Trump	10:04	I can tell you from personal experience, I was in the hospital. I ha
>Kristen Welker	11:06	Okay. Former Vice President Biden to you. How would you lead

PROJECT PLAN PT.2

Analysis Hypothesis:

A classifier should be able to predict D or R based on the 3 models in this order:
highest accuracy with speeches, then next the debate, then the platform.

SPEECHES

Written with persuasive intent
(Most biased?)



DEBATES

“Spontaneous” persuasion
(Keep both sides in mind?)

PARTY PLATFORMS

Informative persuasion
(most neutral?)



ANALYSIS PROCESS:

- Tried logistic regression with the more meta features: token len, type len, ttr, sent len, char amt, etc.
- Accuracy stayed around 45% for the speeches and 50% for the debate.
 - Probably would have been lower for the platform



- these features were not good at deciphering between the two affiliations
- there are not big rhetorical stylistic differences between parties.

```
# building logistic regression
from sklearn.linear_model import LogisticRegression
X = conv_speeches[['Token_count', 'Type_count', 'TTR', 'SentAmt', 'AVGSENTLEN']]
y = conv_speeches['Aff']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0, stratify=y)
```

```
lr_model = LogisticRegression(C = 1, class_weight= None, penalty='l2') # default setting
lr_model.fit(X_train, y_train)
y_pred = lr_model.predict(X_test)
accuracy_score(y_test, y_pred)
```

0.45454545454545453

SVC ANALYSIS PROCESS:

Debate:

- 1.) Added both debate csvs together
- 2.) Gridsearch for best parameters in SVC
- 3.) Ran classifier
- 4.) Most important features (unigram, bigram)

Speech:

- 1.) Gridsearch, but the accuracy was high just from basic parameters
- 2.) Most important features

Platform:

- 1.) Gridsearch
- 2.) Ran classifier
- 3.) Most informative features

```
from sklearn.pipeline import Pipeline
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import SVC
from sklearn.model_selection import GridSearchCV

x = conv_speeches.Speech2[:1000]
y = conv_speeches.Aff[:1000]

tfidf_model = TfidfVectorizer()
svc_model = SVC()

pipe = Pipeline(steps=[('tfidf', tfidf_model), ('svc', svc_model)])

param_grid = {
    'tfidf__max_features': [1000, 2000, 3000],
    'tfidf__stop_words': ['english', None],
    'svc__C': [1E5],
    'svc__kernel': ['linear'],
    'svc__gamma': [1]
}

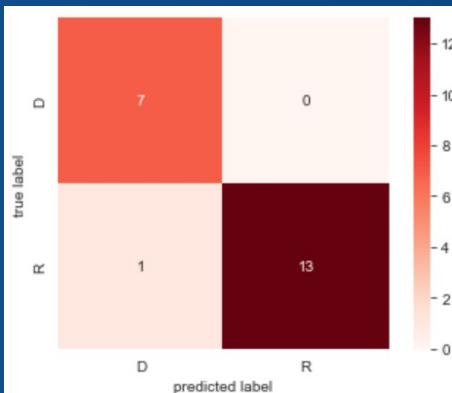
search = GridSearchCV(pipe, param_grid, n_jobs=3, cv=5)
search.fit(x, y)
```

RESULTS OF SVM CLASSIFICATION:

95.2%



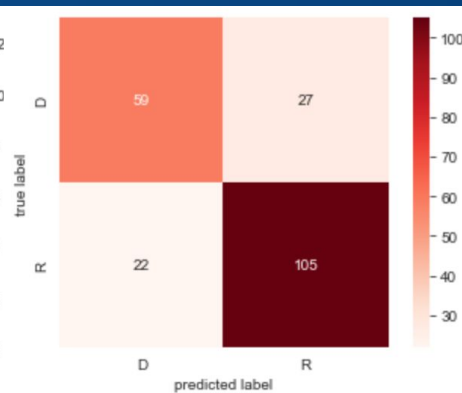
SPEECHES



77.6%



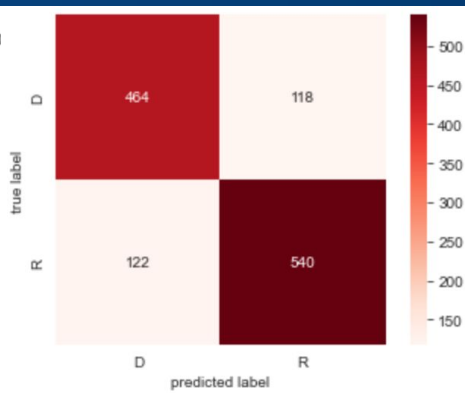
DEBATES



80.7%



PARTY PLATFORMS



★ MOST SIGNIFICANT FEATURES:

Speeches:

D:		R:	
1.)	Covid	1.)	China
2.)	Healthcare	2.)	Freedom
3.)	Virus	3.)	Heroes
4.)	Mail	4.)	Opportunity
5.)	Poverty	5.)	Dream
6.)	Rights	6.)	Democrats
7.)	Truth	7.)	Police
8.)	Women	8.)	Children
9.)	Movement	9.)	Families
10.)	Pandemic	10.)	Media
11.)	Injustice	11.)	Bless
12.)	Empathy	12.)	States
13.)	Immigrants	13.)	Law
14.)	Justice	14.)	Pandemic

Platform:

D:		R:	
1.)	Health	1.)	Government
2.)	Believe	2.)	Tax
3.)	Support	3.)	Economic
4.)	Care	4.)	Freedom
5.)	Workers	5.)	Private
6.)	Public	6.)	Religious
7.)	Rights	7.)	Military
8.)	Access	8.)	Human
9.)	Communities	9.)	Economy
10.)	Pandemic	10.)	Amendment
11.)	Students	11.)	Security
12.)	Programs	12.)	Constitutional
13.)	Affordable	13.)	Business
14.)	Climate	14.)	Defense

Debates:

D:		R:	
1.)	People	1.)	Excuse
2.)	Respond	2.)	People
3.)	Jobs	3.)	Statement
4.)	Parents	4.)	Tax
5.)	Thank	5.)	Country
6.)	Make	6.)	Know
7.)	True	7.)	Work

Apparently, there were less “hot topic” words used in the debates than I thought. -> Could also be because debates have prompts.



CONTINUED DEBATES:

Bigrams:

D:

red-states, problem-going, net-zero, commander-chief, asking-people, president-obama, abraham-lincoln, make-china, need-help, god-going, business-china, public-option, zero-emissions, future-bright, making-sure, oil-industry, let-talk, make-sure

R:

american-energy, young-people, american-jobs, close-oil, trump-tax, really-like, criminal-justice, trillion-dollars, built-cages, cutting-taxes, russia-russia, left-mess, millions-dollars, socialized-medicine, small-businesses, abraham-lincoln, big-statement, tax-cuts, supreme-court, history-country, green-new, new-deal, oil-industry, african-americans



PROJECT PLAN PT.3

ADJ hypothesis

Republicans would use more adjectives, but it would not with good accuracy.

Order of most adjectives: debates, speech, then platform

REPUBLICANS

Donald Trump -> “best”, “great”, calling candidates names, etc.

DEBATES

“Spontaneous” persuasion, sell idea

ANALYSIS PROCESS:

- Used TextBlob
- Pulled adjectives from each row of text
- Found the amount of adjectives in each row
- Compared R and D
- Compared Speech, Debate, Platforms

Extra:

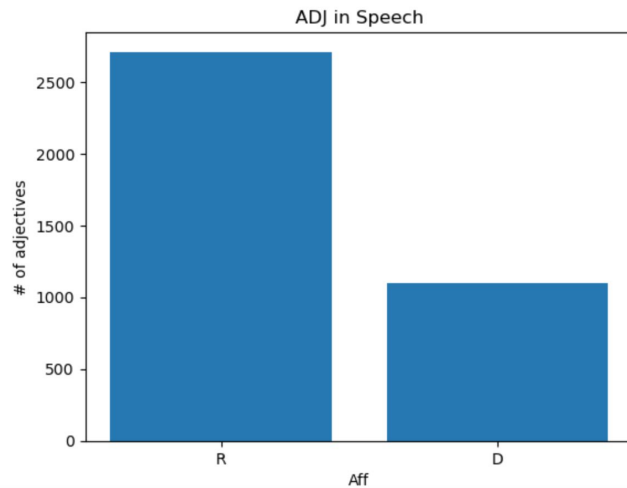
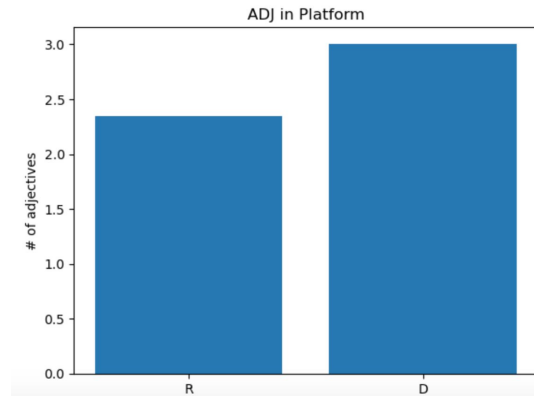
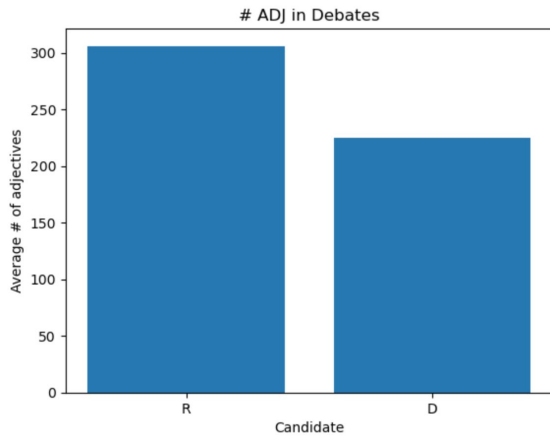
- Donald Trump vs. Joe Biden: use of comparative and superlative adjectives
- Most frequent adjectives in the debates

```
import pandas as pd
from textblob import TextBlob

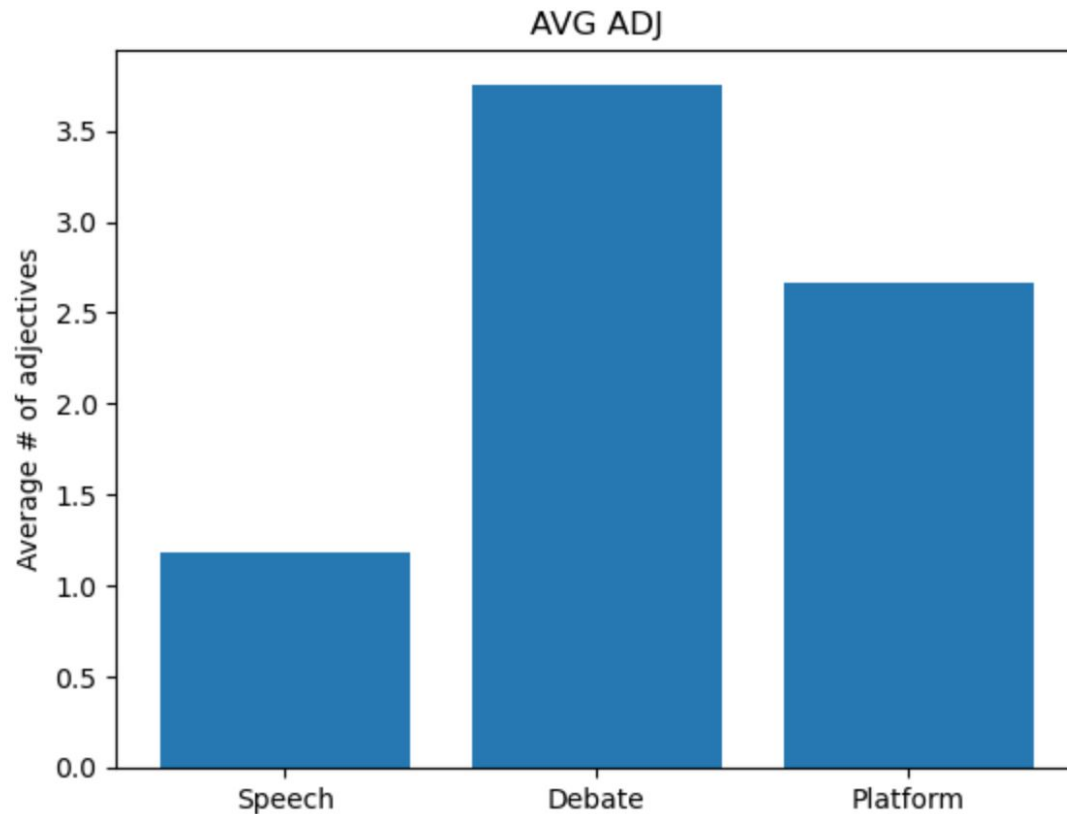
def get_adjectives(text):
    blob = TextBlob(text)
    return ' '.join(word for (word, tag) in blob.tags if tag.startswith("JJ"))

allpresdebate['adjectives'] = allpresdebate['transcript'].apply(get_adjectives)
adj_count = [len(nltk.word_tokenize(t)) for t in allpresdebate.adjectives]
allpresdebate['adj_count'] = adj_count
```

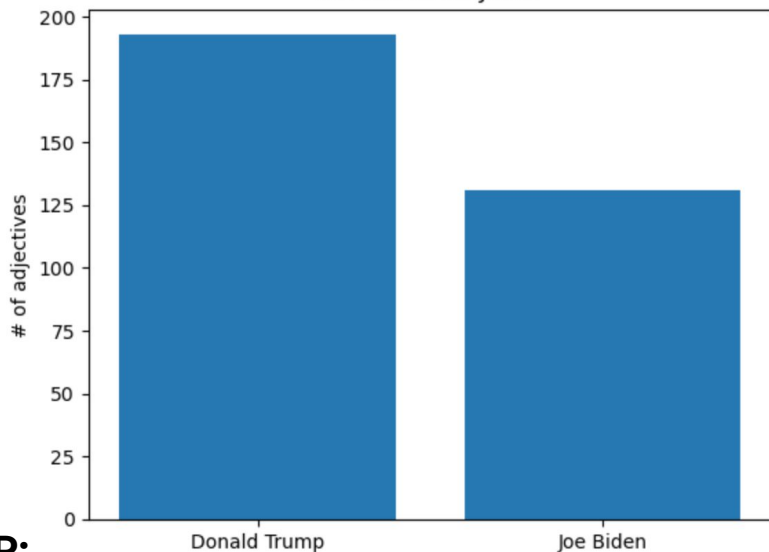
★ ADJS: D VS. R



★ AVG ADJS: SPEECH VS DEBATE VS PLATFORM



CM and SP ADJ AMT



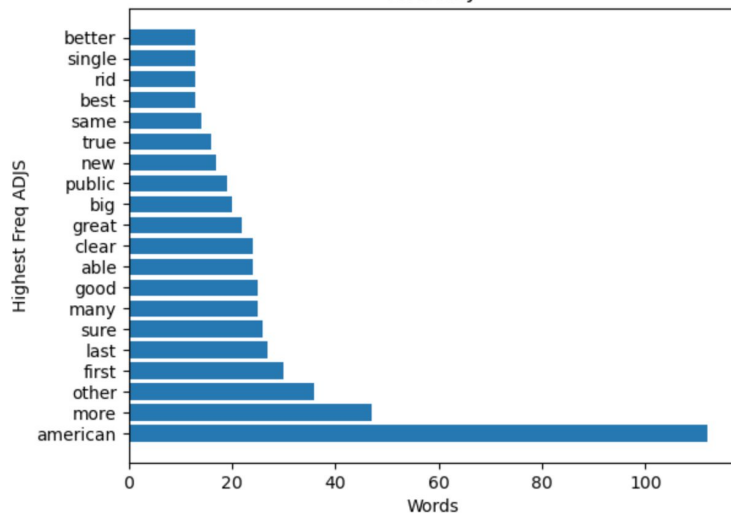
R:

greatest, real, long, natural, challenging, responsible, historic, lowest, tiny, clean, likely, nuclear, xenophobic, able, massive, common, better, cleanest, unleashed, young, treated, vice, president, armed, terrible, socialized, single, green, fair, right, true, short, rid, new, different, second, worse, vibrant, average, clear, liberal, tremendous, important, military, small, little, economic, criminal, horrible, bad, strong, good, great, public, biggest, best, big, american

D:

lower, worried, minimum, significant, second, different, white, bipartisan, look, right, ukrainian, average, black, high, social, transparent, clean, profit, new, affordable, military, low, national, common, little, economic, russian, red, real, nuclear, private, great, renewable, better, preexisting, free, young, bright, foreign, let, complete, best, republican, wrong, public, big, single, safe, able, criminal, rid, important, true, good, responsible, clear, sure, american

AVG ADJ



CONCLUSION

- Hypothesis #1: Debates and Platforms switched
- Hypothesis #2: Platforms had second most avg adjective amount



- Debates: less use of “Hot Topic” words/words to further political rhetoric, but more adjectives
- Speeches: most dividing uses of “Hot Topic” words, but with the least avg # of adjectives
- Platform: most important features were also more polarized than I thought, second highest use of adjectives



- Republicans generally used more adjectives, and words pertaining to the economy, laws, taxes, freedom, etc.
- Democrats generally used less adjectives with words pertaining to public health, social justice, civil rights, etc.

A graphic of the American flag is positioned on the left side of the slide. It features red and white wavy stripes and a blue field with white stars. A large white star with a blue outline is prominent in the center-left, with a smaller white star below it. Several small grey stars are scattered in the background.

THANKS

Do you have any question?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution