

DATASCI200 Project 2

Ad Campaign Analysis on 2016 Presidential Campaigns



Annie Friar
Benjamin Lee
Phoebe Yueh

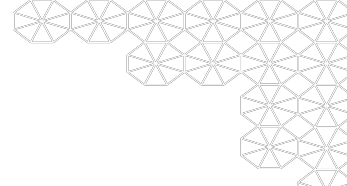
Intro & Research Question

How did advertising strategies differ during the 2016 election cycle?



Data

Primary Dataset : The Political TV Ad Archive



variable	example	definition
archive_id	435303	This is the unique identifier for a specific airing of a political ad
start_time, end_time	20:50:49, 20:51:09	Date/time ad aired, start and end
race	PRES, AZS1	The federal race the ad is targeted toward.
cycle	2016	Election cycle
sponsor_type	PAC	Candidate committee, Super PAC, 501(c), 527 etc
candidates	Bernard Sanders, Hillary Clinton	Candidate(s) named in ad; input by Internet Archive researchers. Note: if the only mention of a candidate in an ad is "I'm so-and-so and I approve this message," that candidate's name is not listed here.
message	Pro, con	Pro, con, mixed; input by Internet Archive researchers. Pro = ad mentions one or more candidates in a positive way, no negative message about any candidate (Important: this applies only to candidates running in current election and race). Con = ad mentions one or more candidates in a negative way. Mixed: Any ad that mentions more than one candidate in particular race, with significant positive content about one or more candidates and negative content about one or more candidates
type	Campaign ads	Campaign ad, issue ad, unknown, input by Internet Archive researchers. Most ads are "campaign ads"

Data

Supporting Dataset

- **Unique Ad Metadata**

(data from <http://politicaladarchive.org/data/>)

- **2016 Candidate Affiliations**

(data scraped from <https://www.fec.gov/data/candidates/president/presidential-map/>)

- **Presidential Election results (2012 & 2016)**

(data from <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/42MVDX>)

- **Categorized PACs**

Data Cleaning Process & Assumptions

Primary Dataset

- Keeps rows where
 - race = 'PRES' (presidential)
 - cycle = '2016'
 - end_time <= '2016-11-09'
- Fix inconsistent naming conventions

message_type	location
pro	Las Vegas, NV
con	Sioux City, Iowa
mixed	New York City, NY
mix	Cleveland, Ohio

Data Cleaning Process & Assumptions

Supporting Dataset

o Unique Ad Metadata

- Perform left outer join by 'archive_id'
- Create a column called 'fact_checked' if 'reference_count' > 0

column	example	definition
air_count	1900	Total number of times this particular ad has aired
market_count	6	Total number of markets this particular ad has aired
reference_count	5	Total number of fact/source checks from partner organizations
transcript	I ran for office because...	Transcript for the ad
fact_checked	0,1 (binary)	Check if the ad has reliable sources

Data Cleaning Process & Assumptions

Supporting Dataset

o 2016 Candidate Affiliations

- Use the lambda function to split the dataset into a reference table with candidate names, parties, and party abbreviations;
- Use the reference table in self-defined functions to extract key information from the 'candidates' column in our original dataset.

column	example	definition
num_cand	2	The ad is showing 2 candidates.
cand_in	[Trump, Clinton]	Trump and Clinton are shown in this ad.
maj_party	1R1D	There are 1 Republican and 1 Democrat shown in this ad.

cand_nm
Clinton [DEM]
Trump [REP]
Sanders [DEM]
Cruz [REP]
Carson [REP]
Rubio [REP]
Bush [REP]

Data Cleaning Process & Assumptions

Supporting Dataset

o Categorized PACs

- Perform left outer join by the 'sponsor' column;
- Classify if a specific PAC is supporting either Clinton or Trump
- **Assumption:** we only include PAC that showed more than 100 ads during the general election cycle (dates after June 16th, 2016)

column	example	definition
sponsor_summary	Clinton	Classify which candidate the PAC shown in the 'sponsors' column in the original dataset is supporting

unique_values	counts	support
Hillary for America	38315	Clinton
Donald J Trump For President	22617	Trump
Priorities USA Action	11250	Clinton
Rebuilding America Now	2735	Trump
Women Vote!	1102	Clinton
Future45	960	Trump
Reform America Fund	803	Trump
NRA Institute for Legislative Action	791	Trump

Data Cleaning Process & Assumptions

Supporting Dataset

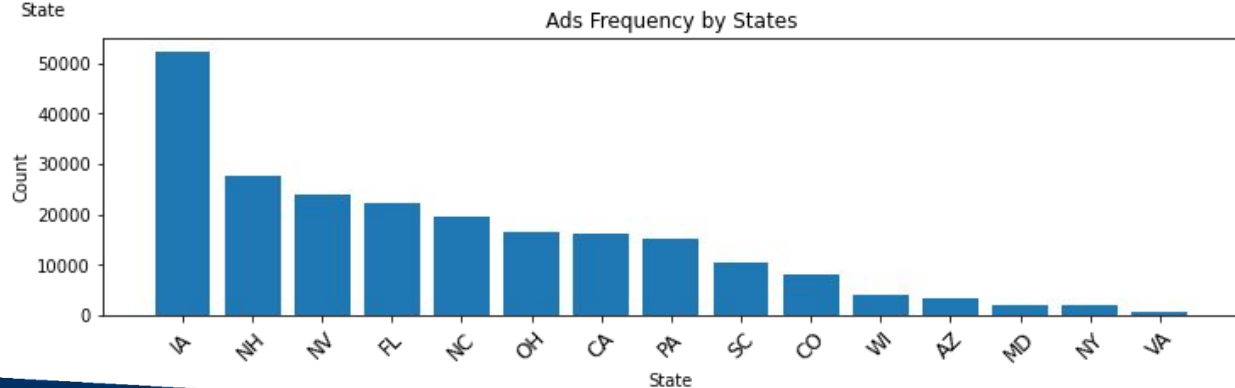
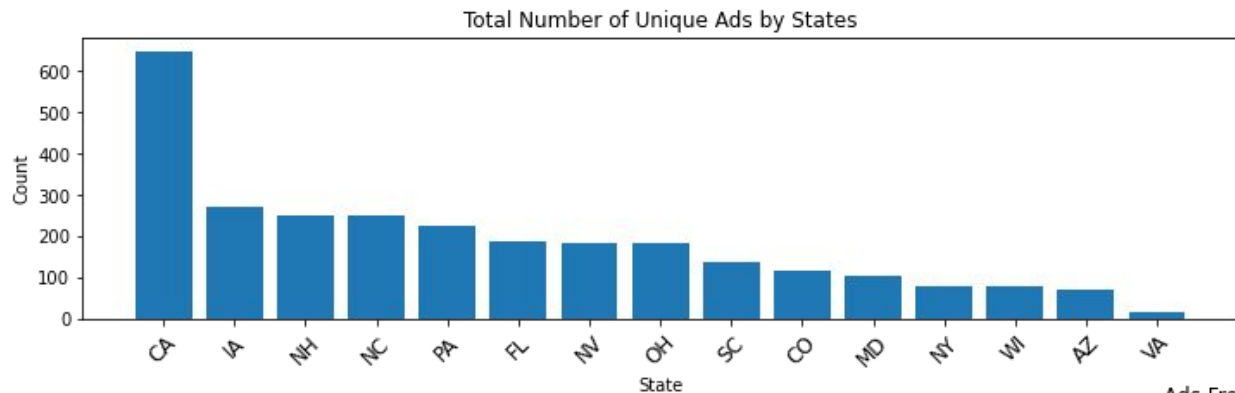
○ Presidential Election Results 2012 & 2016

- Keep rows where
- 'year' > 2010 and < 2020
- 'party_detailed' = DEMOCRAT and 'party_detailed' = REPUBLICAN
- States are shown in the original dataset
- Determine the effectiveness of political ads in our analysis

variable	example	definition
winner_name	CLINTON, HILARY	winner in the state
winner_party	DEMOCRAT	winner's respective party
winner_votes_pct	0.479178	# of votes for the winner / total # of votes
result_change_summary	DEMOCRAT, MARGIN TIGHTENED	vote margin shift from 2012 to 2016

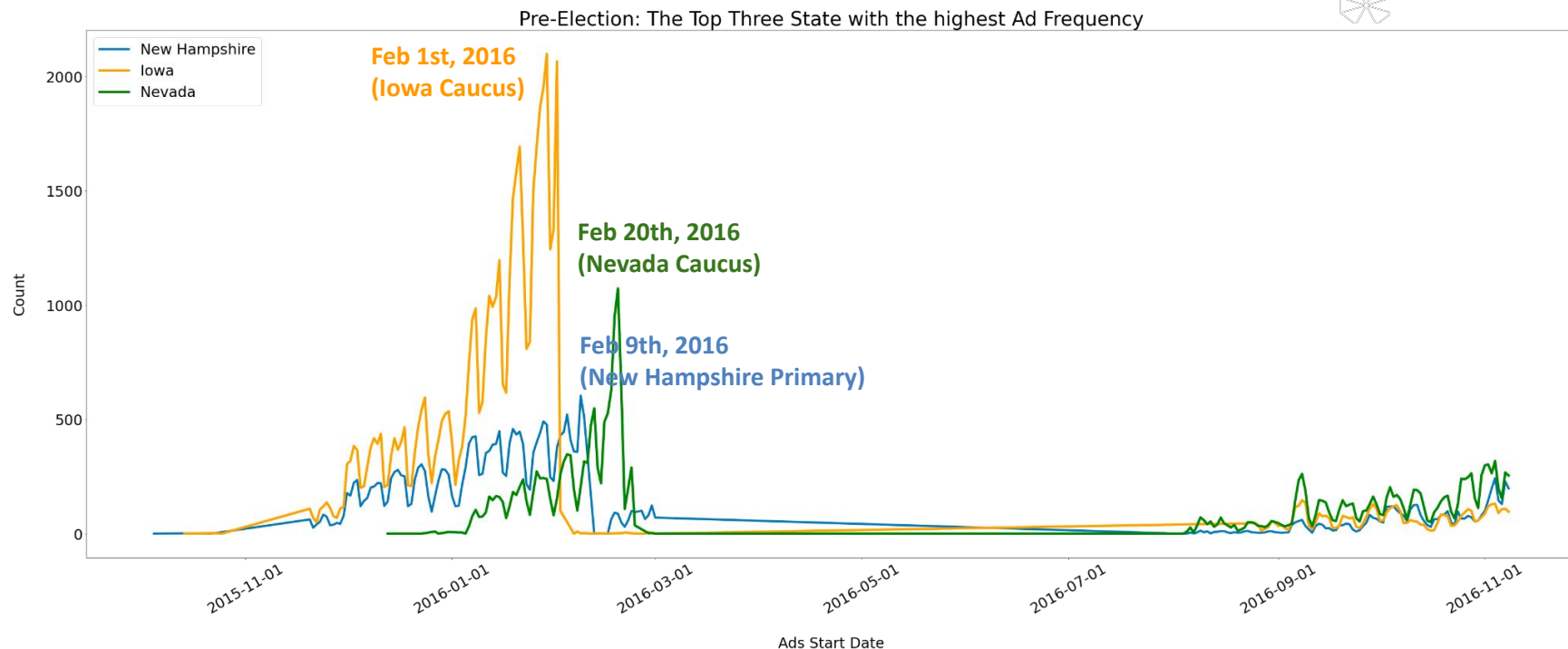
Question 1:

How did advertising strategies differ **among** states?



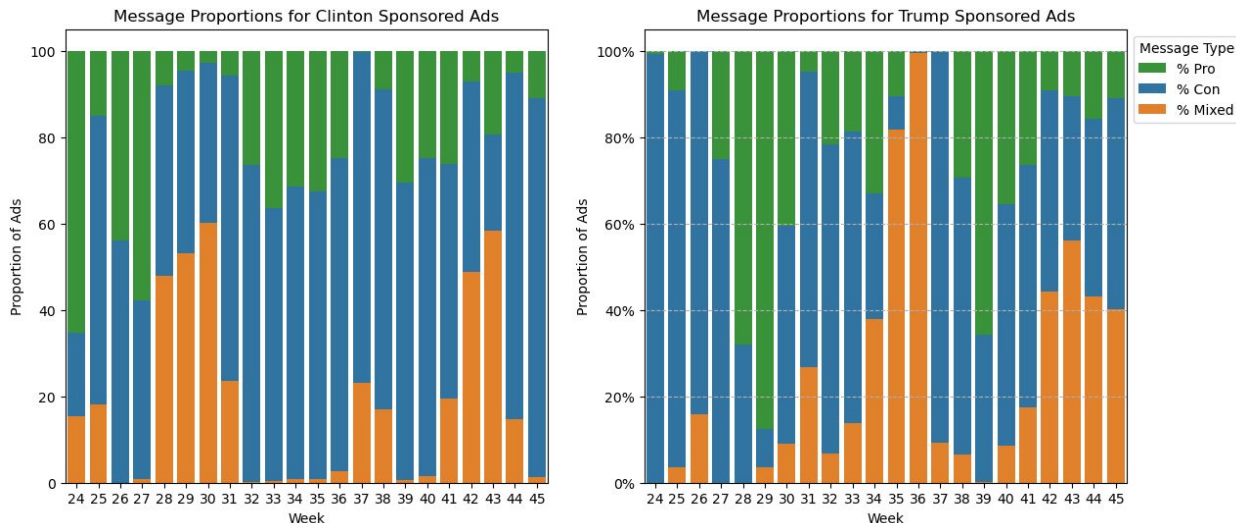
Question 1:

How did advertising strategies differ among states?



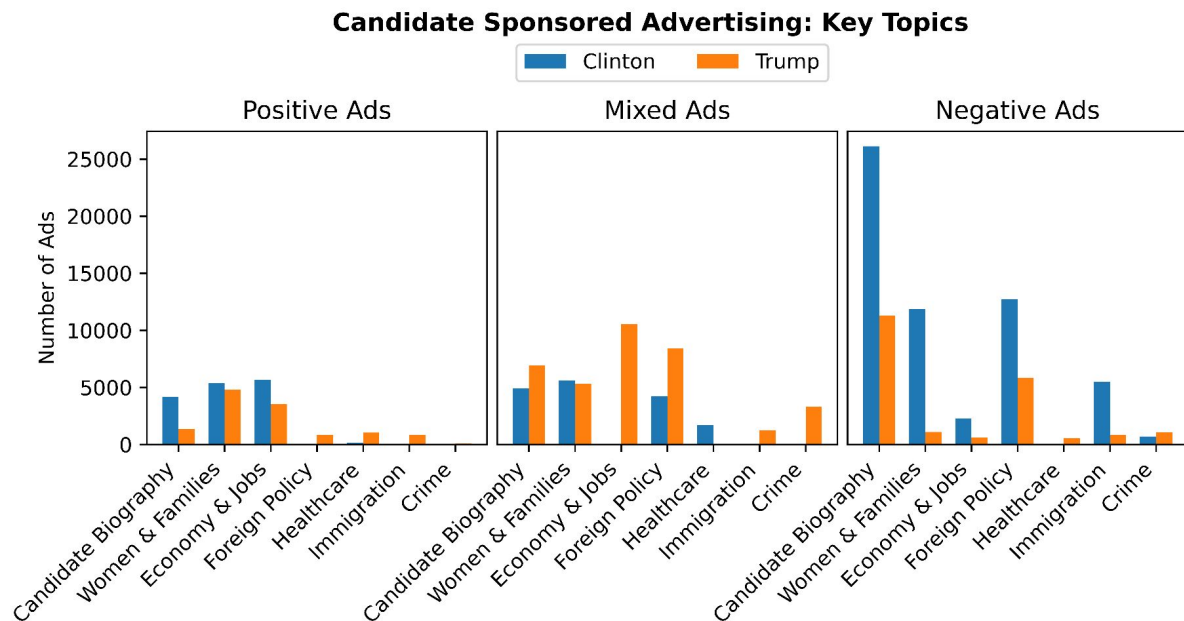
Question 2:

How did messaging strategies differ between candidates?



Question 2:

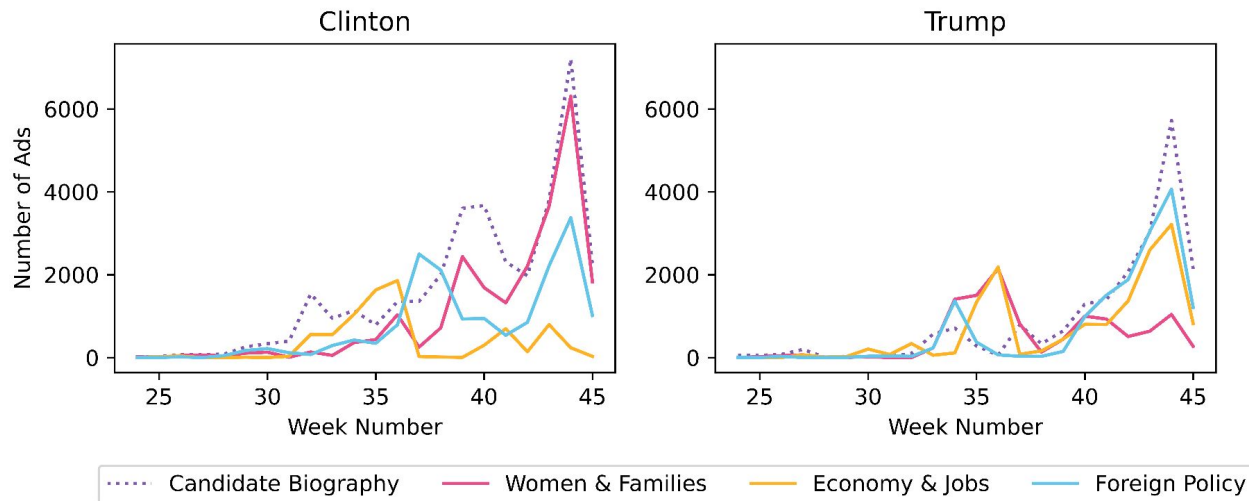
How did messaging strategies differ between candidates?



Question 2:

How did messaging strategies differ between candidates?

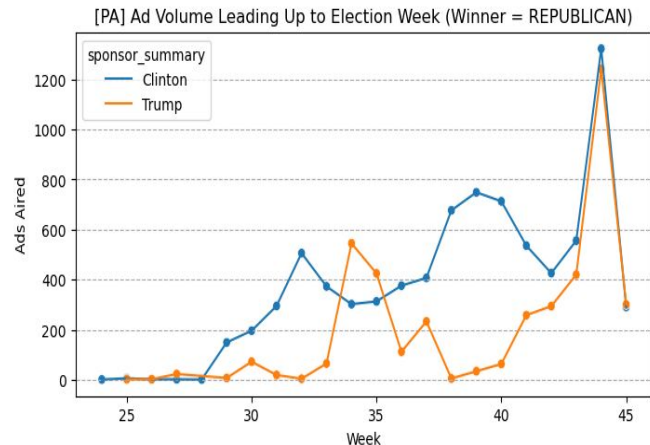
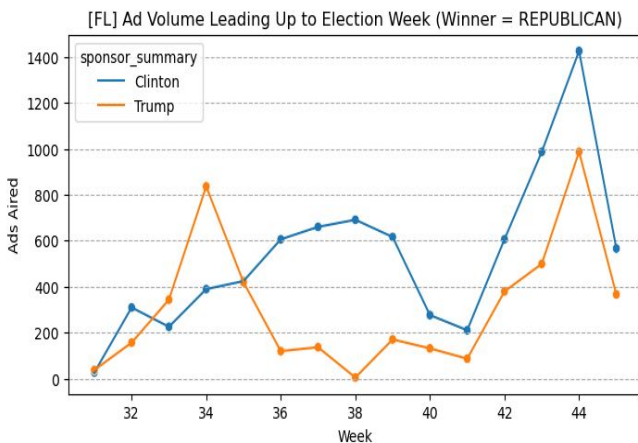
Candidate Sponsored Advertising: Select Topics by Week



Question 3:

How did advertising strategies differ between candidates by state?

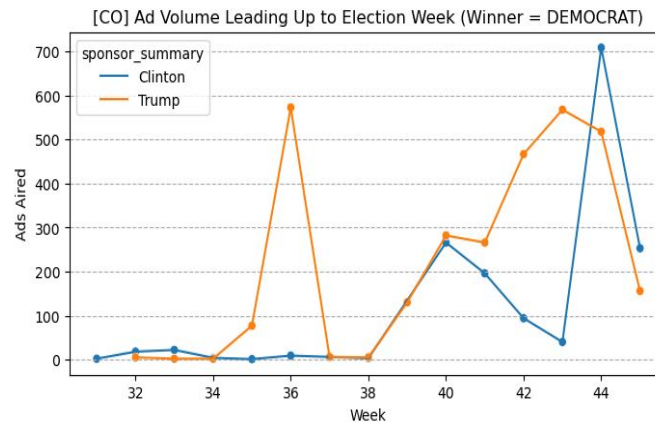
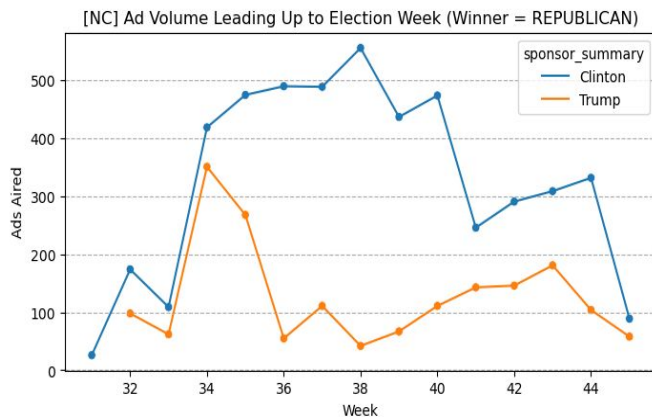
- Competitive States



Question 3:

How did advertising strategies differ between candidates by state?

- Non-Competitive States



Key Takeaways



- Television advertising still relevant
- Advertising by state reflects candidates' national strategy
- Heavy use of negative messaging