

## **Final Report: Ad Campaign Analysis on 2016 Presidential Campaigns**

**Team Members:** Annie Friar, Phoebe Yueh, Benjamin Lee

**Github:** [https://github.com/UC-Berkeley-I-School/Project2\\_Lee\\_Yueh\\_Friar/](https://github.com/UC-Berkeley-I-School/Project2_Lee_Yueh_Friar/)

### **Introduction**

The 2016 Election Cycle involved many consequential races, most notably for President. Presidential elections have long been tightly contested<sup>1</sup>, and 2016 was no exception. The 2016 contest saw hundreds of millions of dollars spent on television advertising between the two candidates for president, although the total amount was less than what was spent in the prior two elections.<sup>2</sup> Some have argued that the decline in investment in television advertising was due to a lack of returns as the media landscape has shifted to more digital advertising, while others pointed to the fact that both candidates for president had high name recognition.<sup>3</sup>

Despite the rising importance of digital advertising, television still gets 70% of the share of total spending on political advertising. Across all races, an estimated \$2.75 billion was spent to air over 4.25 million ads in the 2015-2016 election cycle.<sup>4</sup> Using data from the Political TV Ad Archive, we explored questions around how different television advertising strategies were deployed during the 2016 election cycle.

### **Research Questions**

Our research sought to answer three main questions:

- First, how did overall advertising strategies differ among the states in the dataset?
- Second, did presidential candidates deploy different messaging strategies?
- Finally, how did candidates' advertising strategies differ in battleground states?

### **Data**

*Primary Dataset:* Political Ad Archive from 2016 Elections

Source: <http://politicaladarchive.org/data/>

Description: Dataset contains political ads from the 2016 election in select markets for key battleground states in the primaries.

Key Fields:

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"

#### *Supplemental Datasets:*

- Unique ads metadata details to understand what ads were fact checked or not by partner organizations for the Political TV Ad Archive.
  - Source: <http://politicaladarchive.org/data/>
- State election results for 2016 elections to identify how ad trends and results by state differed for candidates
  - Source: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/42MVDX>
- Data from the Federal Elections Committee (FEC) on candidate parties to add in metadata on candidate or sponsor party support
  - Source: <https://www.fec.gov/data/candidates/president/presidential-map/>
- Data from Open Secrets used to classify Political Action Committees (PACs) by the party and candidate they support
  - Data manually compiled by looking up PACs individually using: <https://www.opensecrets.org/>

## Data Cleaning & Assumptions

### *Primary Dataset*

In the primary dataset, our analysis focused mainly on advertising before the presidential election happened in 2016. For that reason, we selected only rows with race = 'PRES' (presidential), cycle = '2016', and end\_time before November 9th, 2016. We assumed that all rows where race and cycle contain null values could be dropped. As 'Message' was a critical column for us to indicate the positive and negative marketing strategies, we fixed the inconsistent naming convention so that there were only four distinct messaging types: pro, con, mixed, and unknown. We also performed data cleaning to extract only state abbreviations from the 'location' column. Using the lambda function to split the city and state by comma and a user-defined function, the column was correctly formatted with 15 unique values included.

We additionally wanted to explore differences in candidates messaging strategies in terms of their focus on different key topics. To do so, we relied on the 'subjects' field. We examined the top 20 ad subjects, and used user-defined functions to categorize them into seven broad categories or themes.

### *Supporting Datasets*

For the Unique Ads Metadata, we performed a left outer join to add columns 'reference\_count', 'air\_count', 'market\_count', and 'transcript' into the main dataset by utilizing 'archive\_id' and created a column called 'fact\_checked' to determine the reliability of a political ad. We defined an ad as being 'fact-checked' if the reference\_count column has a value greater than or equal to one.

Furthermore, to get the party affiliation of each candidate, we used data from the official website of the Federal Election Commission as another supporting source. Once again, using the lambda function, we split the dataset with values all formatted as 'candidate name [PARTY]' (e.g. Trump [REP]) into a reference table with candidate names, parties, and party abbreviations as columns. As the 'candidate' column in the primary dataset refers to the candidates who are mentioned in the ads, we used the reference table and self-defined functions to extract multiple candidate names in single strings and created below three columns to enhance our analysis:

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

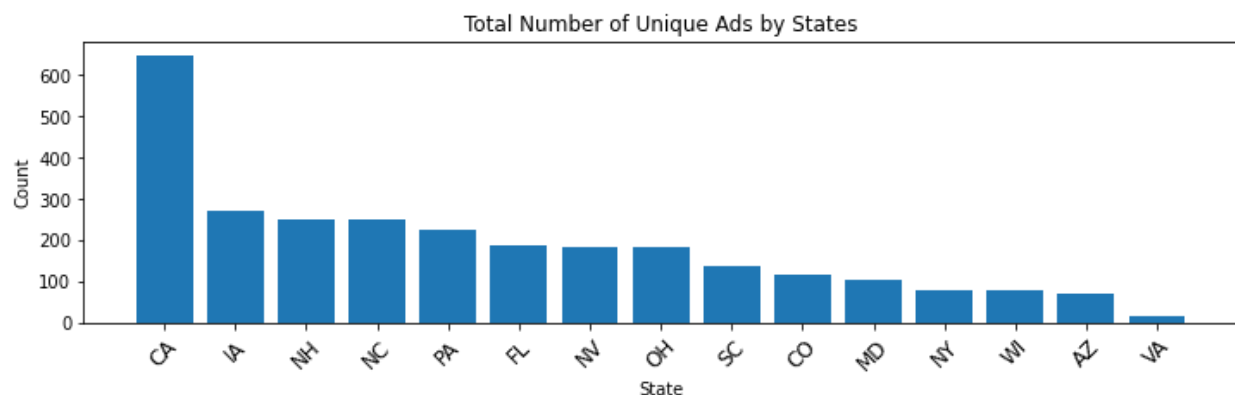
Since one of our core focuses was comparing the advertising strategies of Clinton and Trump, we created a column named ‘sponsor\_summary’. To do so, we looked at sponsors that showed more than 100 ads during the general election cycle (dates after June 16th, 2016). We then took that list of sponsors and used Open Secrets to establish if each sponsor supported Clinton or Trump as a candidate, or if it was unclear. Other PACs, including ones from Primary races and ones supporting other candidates, are labeled as ‘Other’. We then pulled this information back into our primary dataset by using a left join. When analyzing data by sponsor, we excluded records where the sponsor’s allegiance was unclear or where they sponsored a primary candidate who did not participate in the general election.

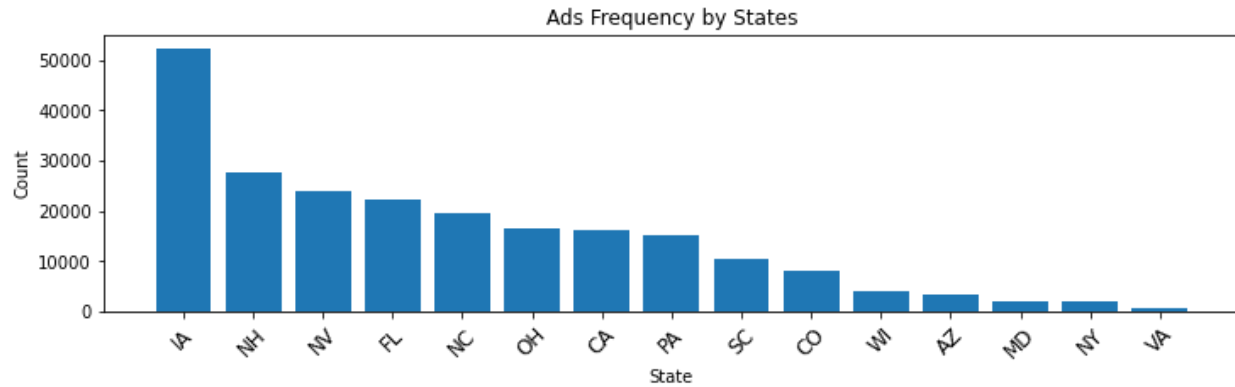
Presidential Election results for each state comparing 2012 and 2016 are also added to our primary dataset. Since the supporting data we gathered from Harvard Dataverse ranged from 1976 to 2020, only rows with ‘year’ between 2010 and 2020 are kept. We further removed rows where states are not shown in the main dataset and parties are neither ‘DEMOCRAT’ nor ‘REPUBLICAN’. The below columns were then created and used as insights to 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

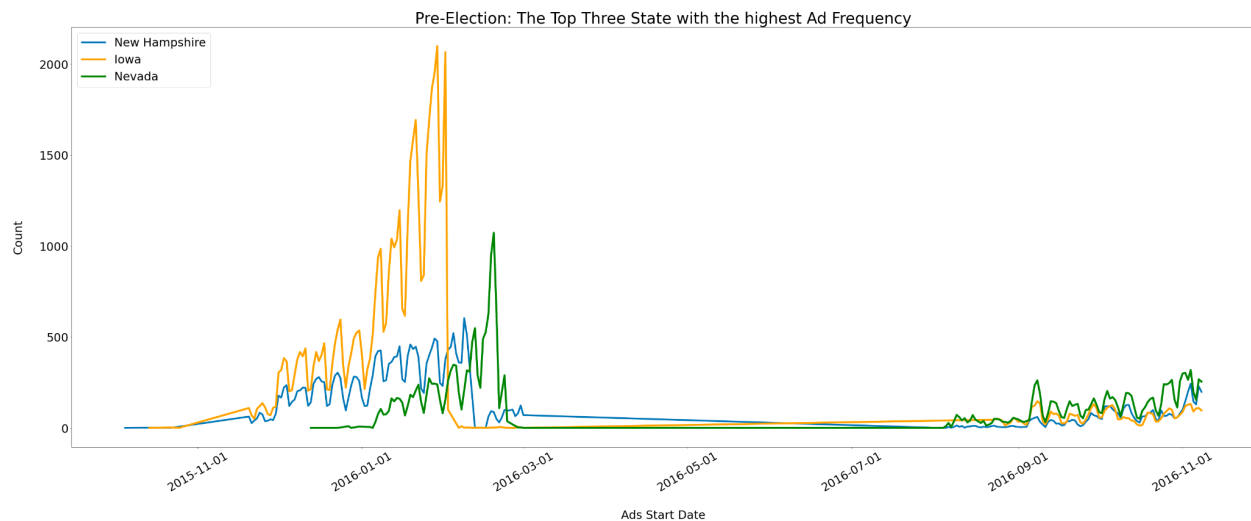
## Charts & Insights

### How do overall advertising strategies differ among states?





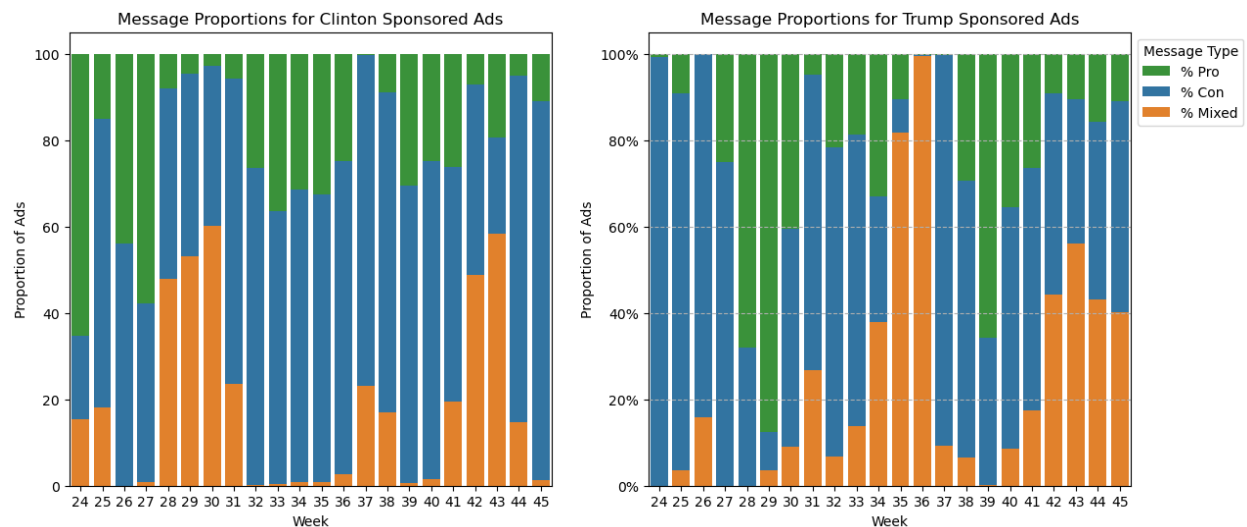
From the first bar chart, we observed that California had the highest number of unique ads aired on television, whereas the second chart implied that Iowa had the highest campaign ad frequency in the pre-election period (in terms of the total number of times ads were aired). Because California is the state with the largest population in the country and Bay Area donors are generally considered to be a wealthy group, we found candidates were proactive in getting exposure and potentially more fundraising opportunities in California. On the other hand, the rationale behind Iowa being the state with the highest aired counts may be that it's long been a swing state, it has a relatively small size, and its media market is much more affordable so the average cost of advertising is lower.



We then plotted the top three states that have the highest ad frequency into a time-series line graph. From there, we observed that the three maximum points where ads are aired the most number of times all happened at the date of caucuses or primaries for each state (e.g. Feb 1st for Iowa [orange], Feb 9th for New Hampshire [blue], and Feb 20th for Nevada [green]). Thus, we can infer that candidates value early caucuses/primaries as a critical time to start building momentum in the election, so they advertise more to get media attention.

## Did presidential candidates deploy different messaging strategies?

To understand how messaging types differed for each presidential candidate, we plotted a time series chart of the proportion of ads aired by messaging type. To understand how each presidential candidate's messaging strategy differed, we relied on the sponsor of the advertising. We grouped together both advertising by the candidate's official campaign, and by super PACs on the candidate's behalf. We focused specifically on the general election, meaning advertisements that were placed after both candidates had secured the nomination (June 16, 2016 or week 24).

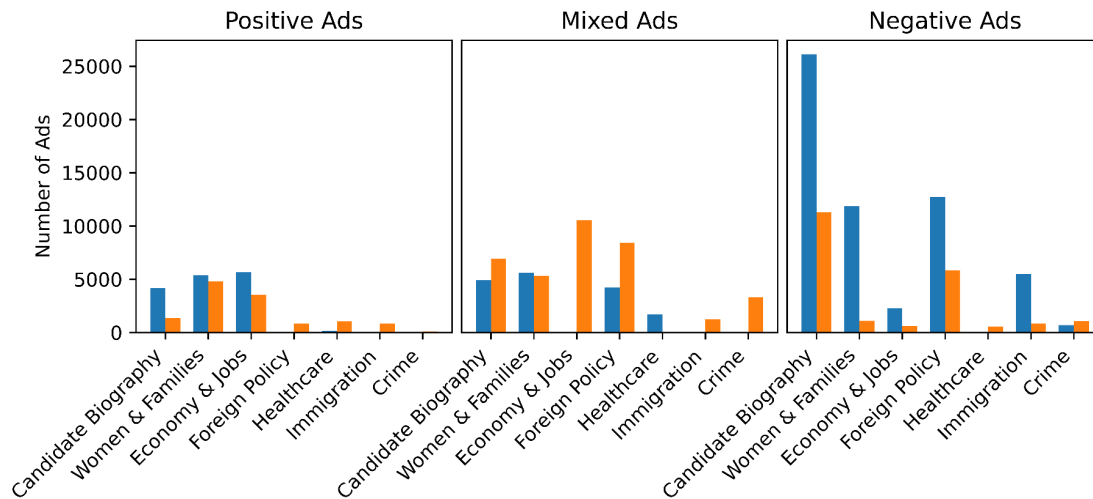


In the above chart, we can see ad proportions by messaging type for Clinton (left) and Trump (right). The main takeaway from these charts is that both candidates deployed a significant amount of negative advertising. Negative advertising made up 60-70% of Clinton's television advertising in the weeks leading up to the election. In contrast, Trump's strategy seemed to focus more on a mix of all messaging types, where he employed a much higher percentage of mixed advertisements than Clinton, especially later in the campaign.

To investigate further, we explored key topics candidates focused on in their advertising during the 2016 campaign. We relied on the category mapping work we did during our data cleaning. We extracted seven major themes from candidate messaging, which were: Candidate Biography, Women & Families, Economy & Jobs, Foreign Policy, Healthcare, Immigration, and Crime.

When exploring this data, we wanted to differentiate between positive, mixed, and negative advertising to better understand the intention behind the messaging. For example, advertising on Candidate Biography could be a candidate promoting themselves, or attacking the other candidate's experience or character.

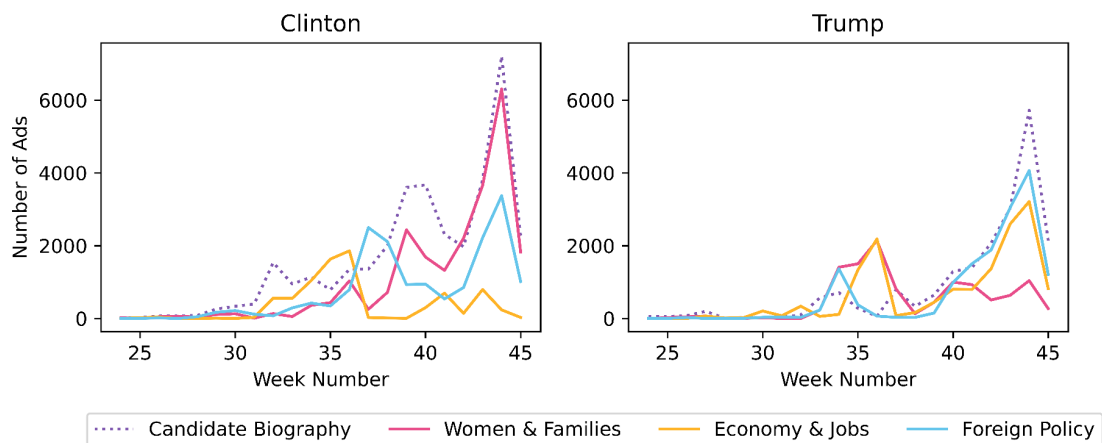
Clinton Trump



As shown in the chart above, much of the advertising during the presidential campaign was negative. We noticed that Clinton's negative advertisements focused on Trump's biography, as well as on Women & Families and Foreign Policy. Meanwhile, Trump attacked Clinton's biography, and also advertised against her on Foreign Policy. Trump deployed more mixed advertising, focusing on Economy & Jobs most of all.

Next, we were interested in exploring how these differences in messaging strategies varied over the course of the general election campaign.

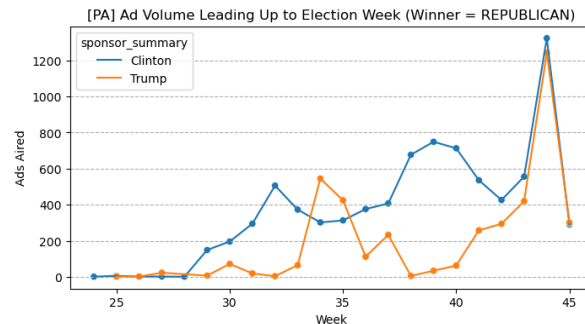
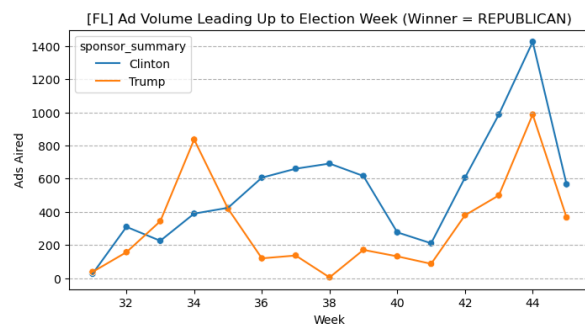
### Candidate Sponsored Advertising: Select Topics by Week



In the chart above we display the four topic categories that candidates advertised about the most. In terms of messaging strategy, both candidates focused on Candidate Biography most of all. It is interesting to note, however, that their focus on subsequent topics differed, especially in the weeks leading up to the election. Hillary Clinton's campaign and PAC supporters ran advertisements that focused on Women & Families and Foreign Policy. Clinton's campaign out-advertised the Trump campaign on most topics by virtue of their higher spending overall, but Trump advertised more on the Economy & Jobs in the final weeks than Clinton did.

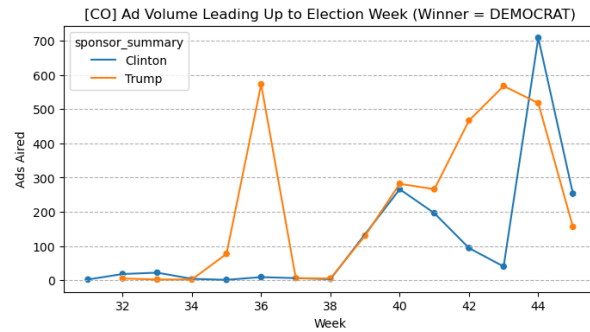
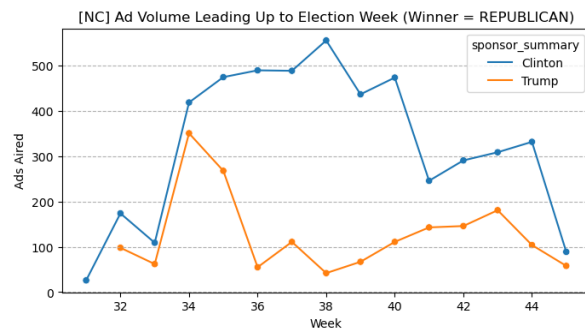
### How did candidates' advertising strategies differ in battleground states?

We looked at candidates advertising strategies in all battleground states, and focused on a few examples of competitive vs. non competitive campaigning between the candidates. Competitive here refers to if the candidates deployed similar amounts of advertisements or not.



Taking a look at the above charts in Florida (FL) and Pennsylvania (PA), we see both Trump and Clinton compete for ad views pretty consistently following the primaries and leading up until the election. Trump seems to have a more efficient ad strategy with a few peaks in advertising on certain dates and lower average ad spending leading up to the election. Meanwhile, Clinton had more consistent spending. We notice the peak in advertisements for Trump in Week 34 in Florida, which coincided with his Immigration speech in Miami. Perhaps this type of promotion alongside key events helped Trump gain more influence in the state. We see a similar event in Pennsylvania for Clinton's campaign near Week 32 where the peak in ads coincides with the Town Hall in Harrisburg.





In the non competitive states of North Carolina (NC) and Colorado (CO), we see the opposite campaign strategies employed by both candidates. These ad strategies reflect how candidates campaign in states that lean more towards their opponent's parties. Overall Colorado leans Democratic and North Carolina tends to lean Republican. Clinton attempted a consistent ad campaign in NC following the primaries, but her spending decreased closer to the election date. This signals that Clinton's earlier campaigning was not enough to sway the results in her favor and she consequently directed her spending elsewhere. Meanwhile, Trump attempted a similar strategy in CO with a focus on running the most ads during rallies he holds in the state and close to the election date, but it was not enough to ultimately sway the voters in his favor.

## Conclusion

Our analysis on the Political TV Ad Archive helped broadly answer our research questions about how television advertising strategies differed across candidates and in different states. We were able to explore the key messaging strategies employed by each candidate for president in this analysis. In particular, we found that there was a large shroud of negativity surrounding both candidates' campaigns. After the primaries, both candidates shifted towards a large percentage of negative advertising, which was mainly focused on the opposing candidate's "Candidate Biography". Looking at these ads from a voter's perspective, it seemed the 2016 presidential campaign was more of a decision of which candidate is not as bad as the other. As one article put it, "if the campaign of 2008 was known as one of 'hope' and 'change,' the campaign of 2016 may well go down in history as one of 'fear' and 'anger.'"

We also gained valuable insights from our analysis of advertising strategy by state. It revealed how candidates decide which states to spend their advertising dollars in and how they deploy different strategies to try and win votes. Clinton's strategy in key battleground states focused on consistent messaging leading up to the election, while Trump's strategy focused on increasing television advertising during key rallies or events he attended in the state.

## References

1. DeSilver, Drew. 2020. "It's Not Just 2020: U.S. Presidential Elections Have Long Featured Close State Races." Pew Research Center. December 4, 2020.  
<https://www.pewresearch.org/short-reads/2020/12/04/its-not-just-2020-u-s-presidential-elections-have-long-featured-close-state-races/>.
2. Hunter, Lauren, Morgan Hess, Prashanth Bhat, and Shawn Parry-Giles. 2016. "2016 Presidential Advertising Focused on Character Attacks." *The Conversation*. November 17, 2016.  
<https://theconversation.com/2016-presidential-advertising-focused-on-character-attacks-68642>.
3. Pearce, Adam. 2016. "Trump Has Spent a Fraction of What Clinton Has on Ads." *The New York Times*, October 21, 2016, sec. U.S.  
<https://www.nytimes.com/interactive/2016/10/21/us/elections/television-ads.html>.
4. SIDES, JOHN, LYNN VAVRECK, and CHRISTOPHER WARSHAW. 2021. "The Effect of Television Advertising in United States Elections." *American Political Science Review* 116 (2): 1–17. <https://doi.org/10.1017/s000305542100112x>.