Political Ads in the 2016 US Presidential Election

Team Members

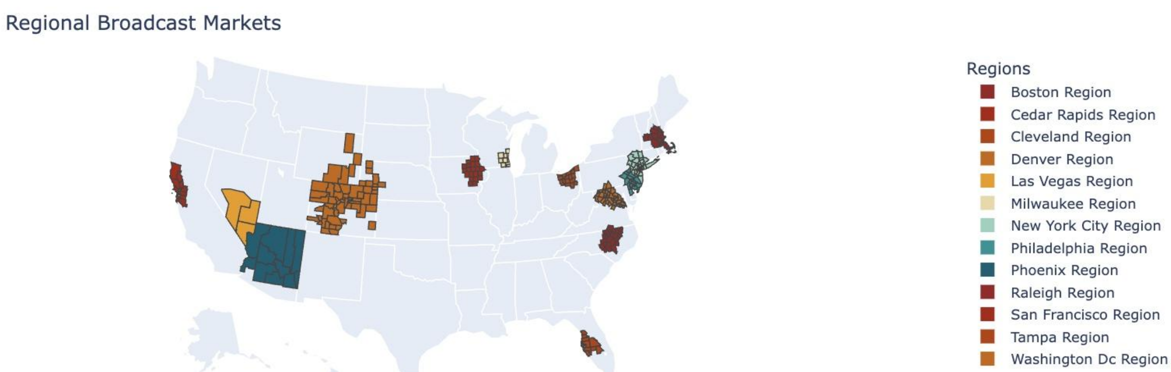
Bronte Baer, Jean-Luc Jackson, Christian Montecillo, and Richard Robbins

Project Summary

In 2016, the underdog won. Donald Trump beat Hillary Clinton 304 to 227 in Electoral College votes with a deficit of ~2.86 million popular votes.[1] How this occurred remains a topic of great debate.

For this project, we analyzed election-related television advertising data. Our working dataset included information for 85,133 advertising broadcasts. We considered four broad categories: (1) volume, (2) subject matter, (3) tone, and (4) timing. Furthermore, parts of our analysis involved comparing the number of ads to the number of votes in a region for the candidates. Our focus was on the presidential general election (excluding the primaries) and was limited to the two major-party candidates. Donald Trump became the presumptive nominee of the Republican party on May 26, 2016.[2] Hillary Clinton became the presumptive nominee of the Democratic Party on June 6, 2016.[3] Accordingly, we only considered ads that aired from June 1, 2016, through and including Election Day, November 8, 2016.

We looked at data covering thirteen regional broadcast markets in the United States. For each of the four study categories, we analyzed the data as a whole and in the respective regions. A map showing the regions appears below. As the map makes clear, and as we discuss below, the regions cover only a portion of the country and do not align with single states or cities.



Data

Sources

Our primary dataset was assembled by the Political TV Ad Archive, a project of the Internet Archive.[4][5] It covers the broadcast of presidential and senate ads for the 2016 elections in select parts of the country. We refer to this dataset as the **advertising dataset**.

The advertising dataset is specific to thirteen Nielsen Designated Market Areas (DMAs) in areas considered to be political battleground regions.[6] Each DMA is assembled from a group of counties and includes more than one city, none covers an entire state, and several include portions of more than one state. The Political TV Ad Archive provides more information about the markets it covers.[7][8] We use the term **region** instead of "market", "area", or other words for consistency and clarity. Appendix A includes a list of the regions and the number of ads aired in each covered in this report.

We supplemented the advertising dataset with an election returns dataset. The MIT Election Data and Science Lab collects and shares election data.[9] We used 2016 election return data from its County Presidential Returns 2000-2020 dataset to consider how our analysis of advertising patterns relates to the voting results in the relevant regions.[10] We refer to this dataset as the **returns dataset**. We used information on US television markets derived from Nielsen DMAs to identify the counties in each region in order to derive returns by region.[11] The regions comprise 268 counties.

A notable limitation is that the advertising dataset is not a comprehensive collection of all ads aired in the relevant regions. The Political TV Ad Archive began collecting advertisements after June 1, 2016 for some regions and stopped collecting before November 8, 2016 in other regions. A complete listing of start and stop dates per region is listed on the Political TV Ad Archive website.[7] The Internet Archive site indicates that they maintained constant TV collection points in three broadcast markets: Philadelphia, San Francisco, and the Washington, DC area, and that San Francisco is where they collected national cable television shows. For the general election, they expanded collection efforts in other markets. While they attempted to collect a vast amount of information concerning broadcast advertising, they never indicated that they attempted to capture all broadcast advertising in a market. We note that the dataset only includes data for 48 broadcast airings in the New York City region during the relevant time frame.

Cleaning and Refining the Advertising Dataset

Detailed information about the fields included in the source advertising dataset can be found on the Political TV Ad Archive data download page. [7] We reference a few of the fields below and more in the body of the report. A complete listing of the fields in the advertising dataset after we cleaned it can be found as Appendix B to this report. The filtering process is accomplished by the **filter_presidential_ads** Jupyter notebook in our project GitHub repository. We retained a few fields we did not use for the sake of completeness and to allow us to expand the scope of our work if the situation warranted.

We cleaned the advertising dataset by:

- removing data for ads related to races other than the presidential race
- removing data for ads that did not mention Hillary Clinton or Donald Trump
- removing data for airings before June 1, 2016 and after election day (November 8, 2016)
- removing unused or redundant columns of no apparent value to our work

We refined the advertising dataset by:

- replacing the **location** field values with shorter more convenient string values and renaming that field to **region_id**, for example, “San Francisco-Oakland-San Jose, CA” became “san_francisco_region”
- adding a **beneficiary** field as described in detail below
- deriving convenient date related fields from **start_time** and **end_time** that reflect local times and dates

The advertising dataset does not identify whether a particular ad was for the benefit of one candidate or the other, rather only that a candidate is mentioned in the ad. Knowing that an ad naming both candidates has a positive, negative, or mixed tone gives us no sense of whether the ad supported one candidate or the other. Slightly more than 24% of the ads subject to our review mentioned both candidates.

For these reasons, and because we considered it essential to identify likely beneficiaries of ads, we conducted a manual review of the ad sponsors. We identified ads sponsored by an organization created to support one of the candidates as benefiting that candidate. Three sponsors accounted for 83% of the ad airings covered by our project: Hillary for America, Donald J Trump for President, and Priorities USA Action (the single largest democratic party super PAC).[12]

We reviewed the ads themselves where information about a sponsor was insufficient for us to reach a determination. The advertising dataset includes URL links to videos of the ads as aired. A sponsor typically produced a small number of ads that aired numerous times. In a few cases, the ad was sponsored by an organization supporting a third-party candidate, in which case the ad was not intended to benefit either Hillary Clinton or Donald Trump. Since our analysis was focused on differences between those two, we excluded those ads. The Internet Archive sponsored two ads (each airing several times), one of which showed Donald Trump in a very negative light and another that showed him in a favorable light. We flagged the first ad as being for the benefit of Clinton and the second as being for the benefit of Trump.

Before removing ads that aired before June 1, 2016 or after election day, and after restricting the dataset to the presidential election and eliminating ads that did not mention either Trump or Clinton, there were 41 sponsoring organizations in the advertising dataset. Of these, 19 were marked as supporting Clinton, 17 as supporting Trump, 4 that supported other people, and 1 with mixed support as described above. The **sponsor_correlation.csv** file contained in the **dataset** portion of our project repository contains the result of our review. We used that to refine the advertising dataset as described above.

At the end of this process, the advertising dataset covered 85,133 airings from June 1, 2016, through election day. That dataset is named **political_ad_pres_airing_from_june.gz** and is contained in the **dataset** portion of our project repository. For the balance of this report, when we refer to the advertising dataset, we mean the advertising dataset after being cleaned and refined as described above.

Cleaning and Refining the Returns Dataset

You can find detailed information about the fields included in the source returns dataset on the page describing the returns dataset.[10] We reference a few of the fields below and more in the body of the report. A complete listing of the fields included in the returns dataset after we cleaned it can be found in Appendix C to this report. The filtering process is accomplished by the **filter_county_returns** Jupyter notebook in our project GitHub repository.

We cleaned the returns dataset by:

- removing data for elections other than 2016

- removing data for counties not included in regions covered by the project
- removing unused or redundant columns of no apparent value to our work
- removed a column that contained obvious math errors and which was easily derived from other columns in the dataset

We refined the returns dataset by:

- adding a **region_id** field as described below
- reshaping from long to wide format in relevant part as described below

The regions used in the advertising dataset do not correspond to any publicly available source of election returns that we are aware of due to Nielsen's strict protection of access to their data. We used information on US television markets derived from Nielsen DMAs to identify the counties in each region. We believe that the counties we associate with regions are fully contained in their respective regions other than Solanco County in California and Apache County in Arizona. We did not attempt to partition those two counties and attributed them to the San Francisco and Phoenix regions, respectively. We believe that our list is a reasonable proxy for the definitive list of counties that Nielsen assigned to the regions in 2016. The **region_county_makeup.csv** file in our project repository's dataset portion contains our county to region translation.

The returns dataset included a single candidate votes column and then, for each county, three rows corresponding to votes for the democrat, republican and other candidates. This "long format" structure made our work more complex than necessary, so we reshaped the dataset to "wide format" by pivoting those three rows to columns.

The regions comprise 268 counties.

The returns dataset resulting from this process is named **enriched_county_pres_2016.gz** and is contained in the **dataset** portion of our project repository. For the balance of this report, when we refer to the returns dataset, we mean the returns dataset after being cleaned and refined as described above.

The awesome_puppies Module and Visualization Templates

We created the **awesome_puppies** module to support the project. It provides our primary project data frames and several related functions that facilitate DataFrame grouping and aggregating for visualization preparation. The module is contained in the **code** portion of our project repository. Summary documentation for **awesome_puppies** is included in Appendix D. The module was named after our team's informal name.

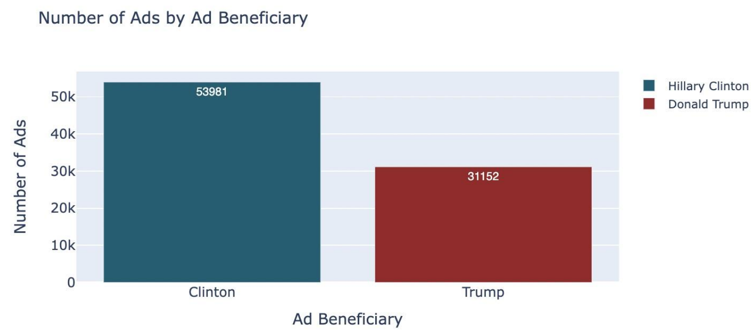
We also created various visualization templates that all team members could use, making building multiple visualizations for different data points efficient. Refer to Appendix E for a list of these visualization templates.

A listing of our project repository is included as Appendix F.

Questions and Insights

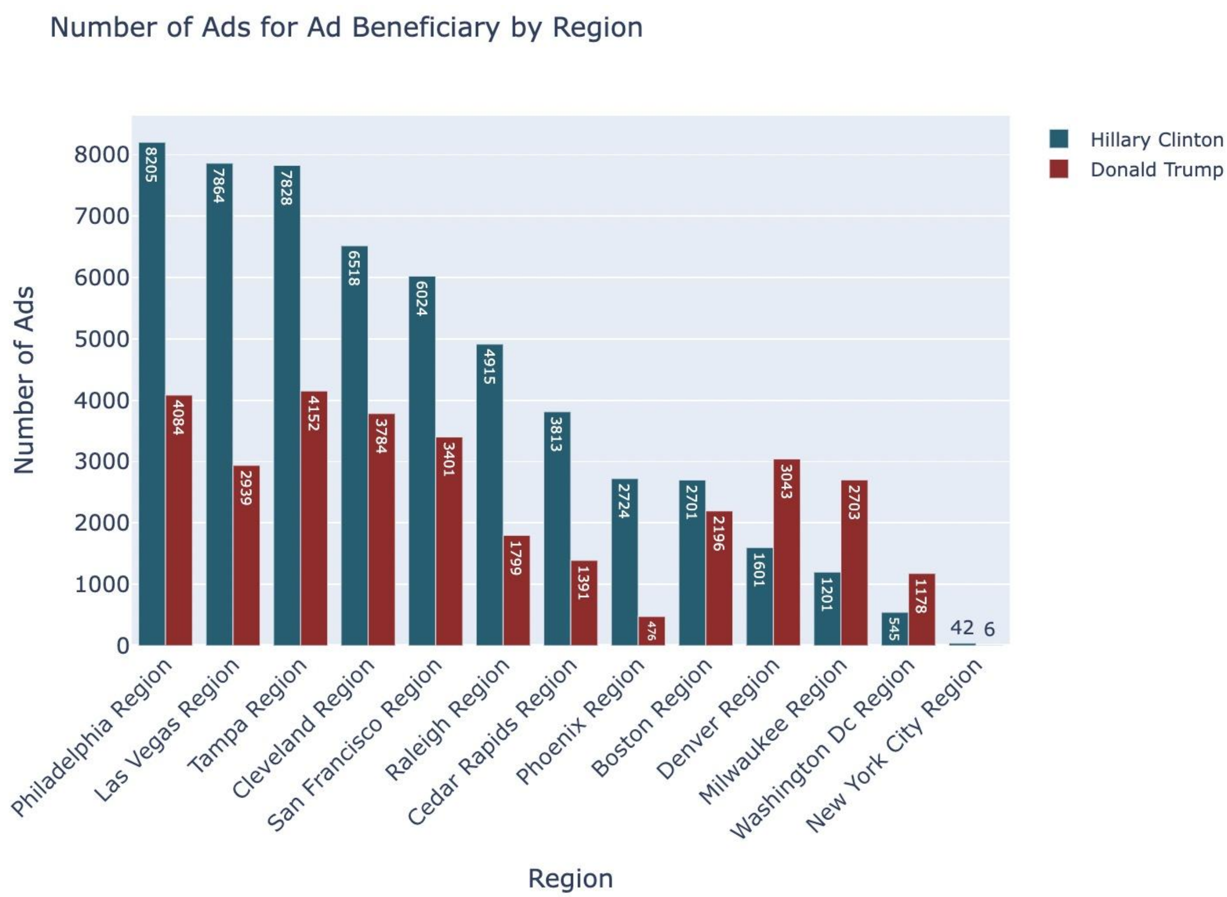
Volume

The advertising dataset covers 85,133 broadcasts, of which 63.4% were for Clinton and 36.6% for Trump.



Did one candidate air more ads than the other in the thirteen regions where the ad data was collected?

The following chart shows the regional breakdown of advertising volume for each ad beneficiary (in other words, the candidate who benefited from the ad airing). In particular, we note that Clinton advertising outpaced Trump advertising in each of the thirteen regions other than Denver, Milwaukee, and Washington DC. In the ten regions where Clinton advertising exceeded Trump advertising, the volume difference was meaningful in every region other than Boston. It is important to note that the New York City region does not have comprehensive ad data, as evident in the total of forty-eight ads included in the dataset.



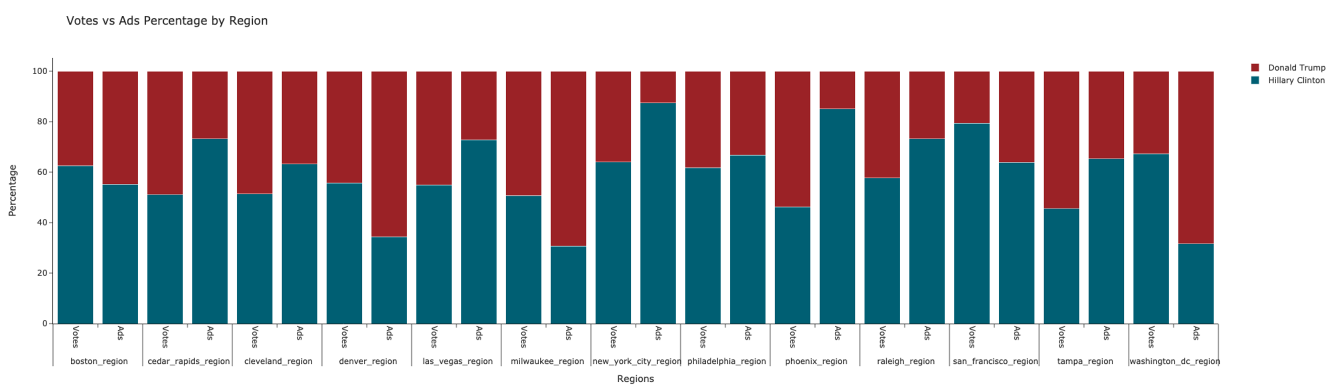
Was the candidate who aired more ads in a region the candidate who won the vote in that region?

An interesting table we created shows the discrepancy in the number of ads aired for a candidate's benefit versus the number of votes for the candidate. In some regions, we see the candidate who aired the most ads was *not* the candidate who received the most votes. These regions are highlighted in red in the figure.

Region	Most Votes	Most Ads
Boston Region	democrat	democrat
Cedar Rapids Region	democrat	democrat
Cleveland Region	democrat	democrat
Denver Region	democrat	republican
Las Vegas Region	democrat	democrat
Milwaukee Region	democrat	republican
New York City Region	democrat	democrat
Philadelphia Region	democrat	democrat
Phoenix Region	republican	democrat
Raleigh Region	democrat	democrat
San Francisco Region	democrat	democrat
Tampa Region	republican	democrat
Washington Dc Region	democrat	republican

In the Cedar Rapids region, Clinton had roughly 2.75 times more ads than Trump yet captured the vote by a margin of only 2.24%. On the other hand, in the Milwaukee region, Trump had 2.25 times more broadcast advertising than Clinton and still lost by a margin of 1.35%.

In the chart below, the numbers in the Phoenix region are especially interesting. There, Clinton broadcasts eclipsed Trump broadcasts by more than a factor of five, and she still lost the vote by over 7%. Arizona voted for the Republican candidate in the four preceding presidential elections.[13][14][15][16]



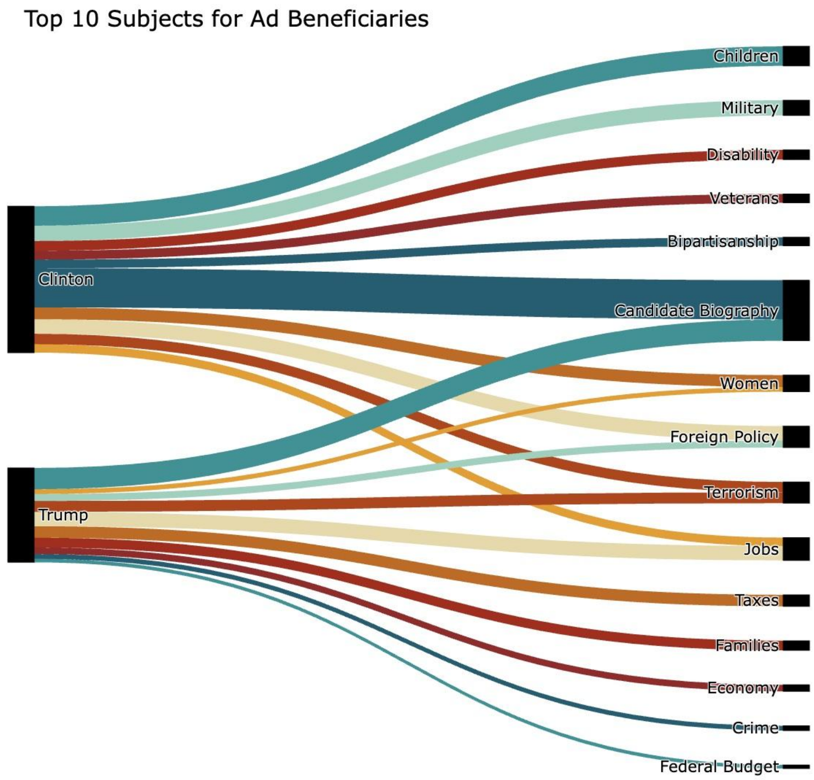
Subject Matter

Analyzing ad subject matter was more complex than we anticipated. Almost all of the ads had more than one subject listed in the ads dataset "subjects" field; therefore, when we first looked at the data, it was hard to draw any conclusions about the subjects covered in the ads. Once we separated the subjects into rows, we noticed interesting patterns.

"Candidate Biography" was always the most commonly covered subject, no matter how we sliced the data. If we looked at subjects by candidate, region, or message tone, "Candidate Biography" was always the most prevalent. This makes sense from a logical standpoint. While we do not know the exact methods used to identify subjects covered in the ads, we have all seen political ad campaigns, and we can imagine that most ads will be classified as "Candidate Biography" because, by nature, political ads are about the candidates.

Are there certain subjects each candidate covered more than others?

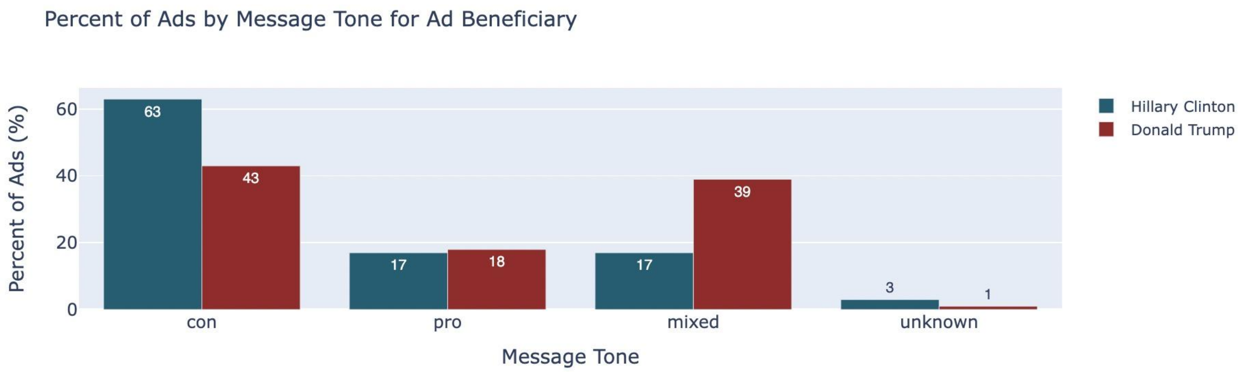
The Sankey diagram we created depicts the flow and overlap of the ad beneficiaries' top ten subjects covered in their ads. We were not surprised to see that both candidates mentioned "Jobs," "Terrorism," "Women," "Foreign Policy," and "Candidate Biography" more than other topics. Furthermore, it is interesting to see that "Military," "Disability," and "Veterans" are amongst Hillary Clinton's top ten, and "Taxes," "Economy," and "Federal Budget" are amongst Donald Trump's most covered subjects.



Tone

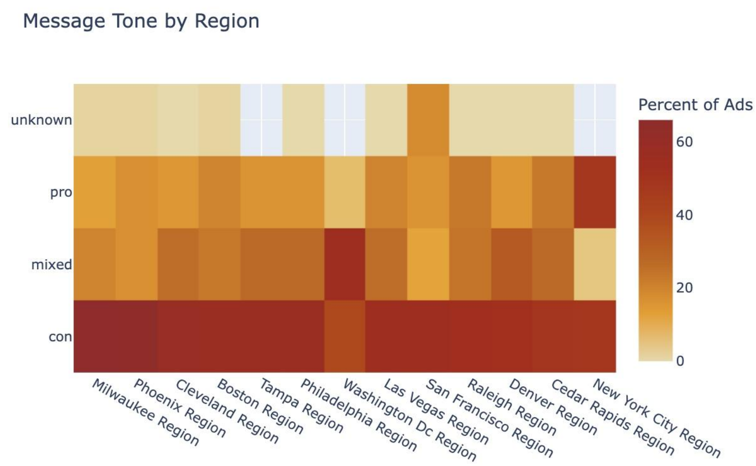
What was the ratio of negative and positive ads for each candidate?

Both candidates aired a higher percentage of ads with negative messages for their benefit. At first glance, it looks like Hillary Clinton aired a higher percentage of negative ads, 63%, than Donald Trump, 43%; however, upon examining the chart further, we noticed that 39% of Donald Trump's ads were classified as "mixed" messages, versus Hillary Clinton's 17% "mixed." This adds a layer of complexity. When we consider the "negative" messages added to the "mixed" messages, we can conclude that both candidates had negative tones in around 80% of their messages. Since Clinton aired more ads than Trump, we chose to look at the message tone by percent of ads rather than the number of ads. This gave us a better understanding of which candidate favored negative messaging rather than which candidate had more negative messages since we already know Clinton had more ads (as shown in the figure above indicating the number of ads by ad beneficiary).



Did candidates' use of positive and negative advertising vary by region?

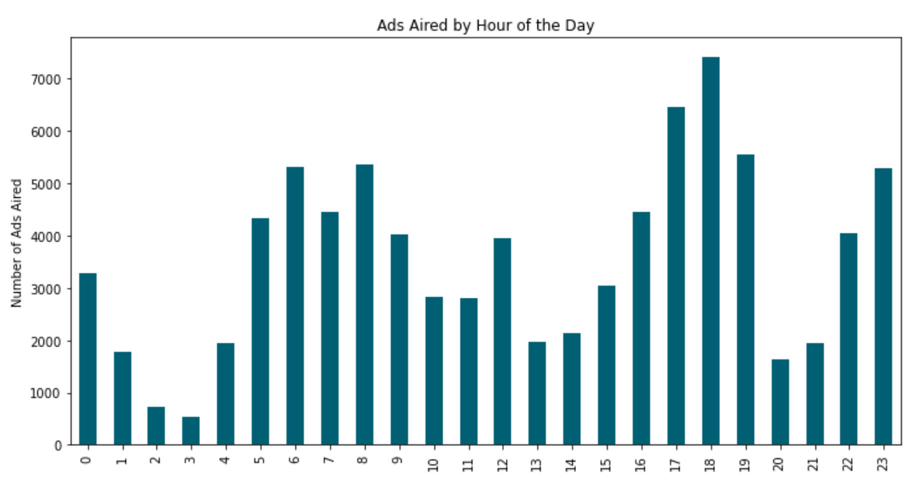
For analyzing the message tone across regions, we decided to look at the percent of ads rather than the number of ads. If we looked at the number of ads, we would not clearly understand the message tone by region since some regions, like New York, had notably fewer ads recorded in the Political TV Ad Archive dataset. In the heatmap below, we can see that negative messaging was common in all thirteen regions.



Timing

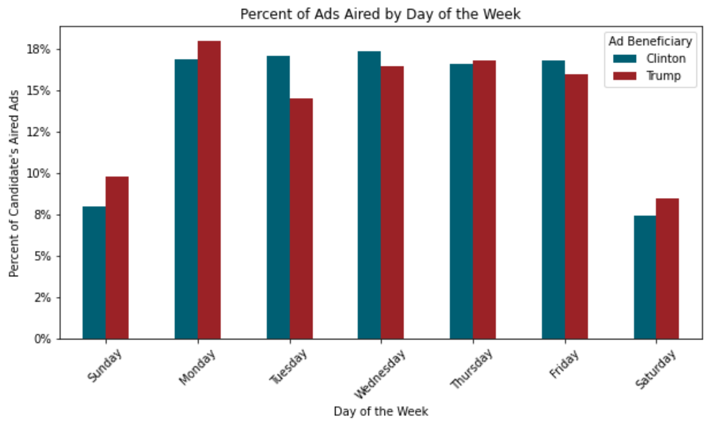
Did the candidates air more ads during a specific time of day?

We started with a broad look at when the ads were typically aired. We first had to convert each ad's air time from UTC to its region's local timezone before grouping it by hour of the day.



As seen in the plot above, the broadcasting of ads peaked around 6 AM and 6 PM and dropped at 3 AM and 1 PM. The story for this "rollercoaster" can be reasonably explained given a traditional workday: people usually watch TV before and after work around 5 AM and 5 PM, tune in again before bedtime, and are usually asleep by midnight (varying between weekdays and weekends).

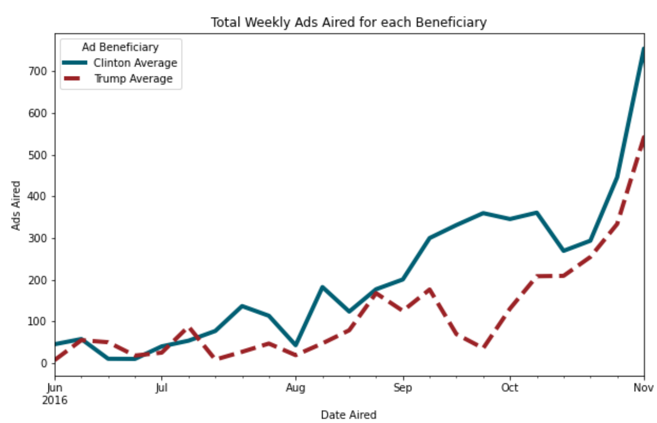
Were more ads shown on weekends or weekdays?



The plot above shows ad airing by day of the week with ad counts normalized to show percentages per day relative to each beneficiary's total ad count. This plot shows that weekdays were more targeted than weekends for ads in our advertising dataset. Both beneficiaries are included to demonstrate that similar strategies were used in daily advertising.

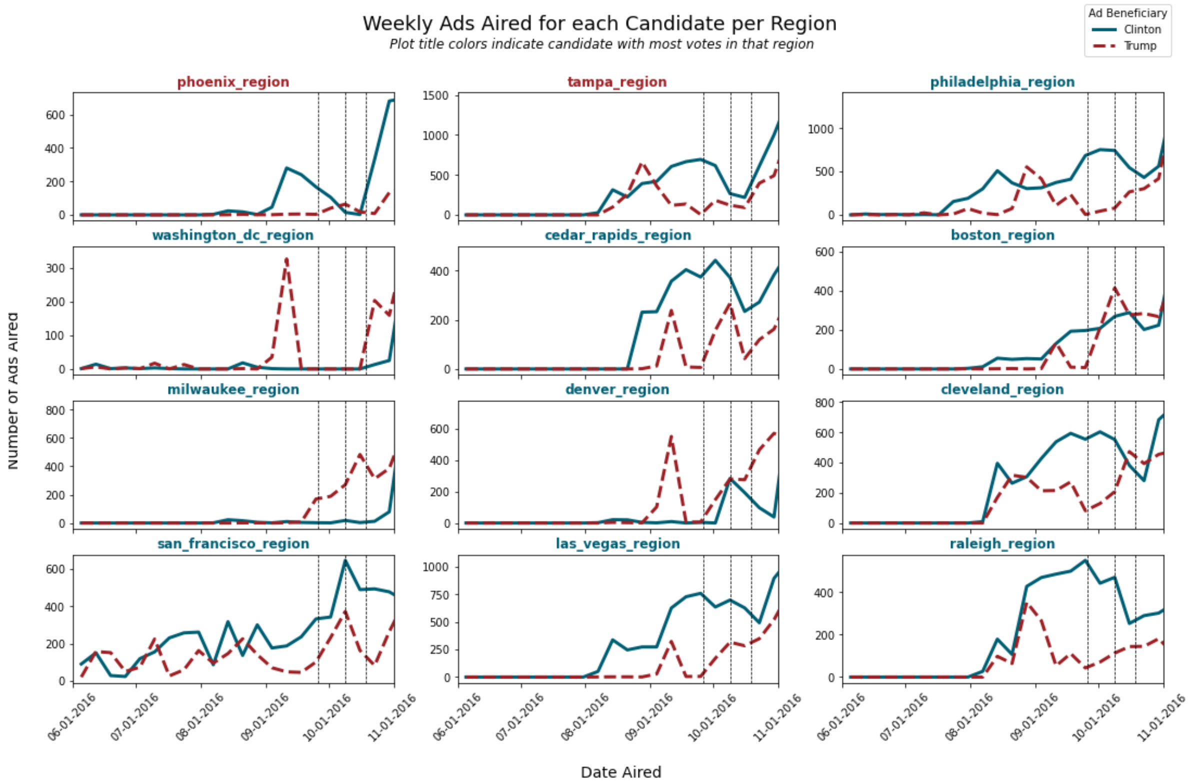
How did advertising frequency change as the election drew nearer?

The first plot below shows each candidate's total ads aired per week over the general election period we considered. Both the Clinton and Trump camps aired a slow and steady stream of ads throughout the summer, then ramped up in the last few weeks leading up to Election Day. We see one deviation starting in September: this is when Clinton's camp started its ramp-up in its broadcasts until Election Day. Trump's, in contrast, had a retraction period in the middle of September before increasing in October until Election Day.



Does the advertising frequency vary across regions? How did the number of ads aired compare to the candidate with the most votes in the region?

The plot below shows the total ads aired per week for each region. The color of each subplot's title indicates which candidate received the most votes on Election Day, an indicator built using the returns dataset. Three vertical lines are included to mark the three presidential debates on September 26, October 9, and October 19 of 2016.



The plots for the regions above tell a story of how candidate campaigns chose to air ads over time. As demonstrated in the Volume section above, the candidate that advertised the most in a given region was not necessarily the candidate that received the most votes. For example, in the Phoenix region we can see that Trump's campaign barely invested in airing ads compared to Clinton's campaign — in the end, the Phoenix region had more votes for Trump. Meanwhile in the Boston region, both candidates increased advertising steadily until a final peak just before the election (Clinton received the most votes). A general observation across all the regions plotted above is that there were many sudden spikes surrounding presidential debates.

Conclusion

After our data exploration and analysis of the Political Ad Archive and MIT Election Data and Science Lab datasets, we successfully answered the questions we had for each of the four categories. However, due to the advertising dataset being non-comprehensive, we can not conclusively say that there is a strong correlation between the number of ads a candidate broadcasts and the number of votes received.

During our discussion about Volume, we saw disparities where the candidate with the most ads did not receive the most votes. We also learned from our Subject Matter analysis that both candidates wanted you to get to know them, and a large part of their messages discussed their bios. The story from our Tones section was similar – both candidates had overwhelmingly negative messages about the other and did not focus much on any of the positive traits about themselves. And finally, from our Timings investigation, we saw that candidates had very similar strategies to when they played ads.

While we have shown you where the two candidate advertising campaigns varied, overall, this data suggests that the candidates appeared to adopt similar advertising strategies.

References

[1] Wikipedia. (2016). *2016 United States presidential election*. 2016 United States presidential election. Retrieved November 18, 2021, from https://en.wikipedia.org/wiki/2016_United_States_presidential_election

[2] NPR. (2016). Amita Kelly, *Donald Trump Clinches GOP Nomination* https://www.npr.org/2016/05/26/479588197/donald-trump-clinches-gop-nomination

[3] AP News. (2016). Hope Yen, Stephen Ohlemacher, Lisa Lerer, and Catherine Lucey, *AP count: Clinton has delegates to win Democratic nomination* https://apnews.com/article/4c9c850385c84b12ad5b85fda49743f9

[4] Political TV Ad Archive http://politicaladarchive.org/

[5] Internet Archive https://archive.org/

[6] Nielsen DMA Regions https://www.nielsen.com/us/en/contact-us/intl-campaigns/dma-maps/

[7] Political TV Ad Archive Data Download http://politicaladarchive.org/data/

[8] Political TV Ad Archive Data Blog http://politicaladarchive.org/revamped-political-tv-ad-archive-website-offers-new-features-political-ad-tracking-in-key-states-for-general-election/

[9] MIT Election Data + Science Lab https://electionlab.mit.edu/

[10] Harvard Dataverse, County Presidential Election Returns 2000-2020 https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VOQCHQ

[11] List of United States television markets https://en.wikipedia.org/wiki/List_of_United_States_television_markets

[12] Priorities USA Action https://en.wikipedia.org/wiki/Priorities_USA_Action

[13] 2000 United States presidential election in Arizona https://en.wikipedia.org/wiki/2000_United_States_presidential_election_in_Arizona

[14] 2004 United States presidential election in Arizona
https://en.wikipedia.org/wiki/2004_United_States_presidential_election_in_Arizona

[15] 2008 United States presidential election in Arizona
https://en.wikipedia.org/wiki/2008_United_States_presidential_election_in_Arizona

[16] 2012 United States presidential election in Arizona
https://en.wikipedia.org/wiki/2012_United_States_presidential_election_in_Arizona

Appendix Reference

Appendix A: Regions and Ad Counts

Philadelphia Region	12,289
Tampa Region	11,980
Las Vegas Region	10,803
Cleveland Region	10,302
San Francisco Region	9,425
Raleigh Region	6,714
Cedar Rapids Region	5,204
Boston Region	4,897
Denver Region	4,644
Milwaukee Region	3,904
Phoenix Region	3,200
Washington DC Region	1,723
New York City Region	48

Appendix B: Advertising Dataset Columns

network	str: TV channel on which the ad aired
region_id	str: name of the region in which the ad aired
program	str: name of the TV program in which the air aired
program_type	str: “news” or “not news”
start_time	datetime: when the aid aired, in local time
end_time	datetime: when the airing of the ad ended, in local time
archive_id	str: a unique alphanumeric id for each ad identified, corresponding with id used on PoliticalAdArchive.org.
embed_url	str: url for the ad
sponsors	str: organization sponsoring the political ad, as it appears in the ad
sponsor_types	str: candidate committee, Super PAC, 501(c), 527 etc
subjects	str: subjects covered in ad; subject index from PolitiFact, input by Internet Archive researchers
candidates	str: 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	str: see the discussion below
beneficiary	str: the name of the candidate who benefited from the ad
date	str: the calendar date on which the ad was aired
day_of_week	str: the day of the week on which the ad was aired
air_time	str: the time of day on which the ad was aired
duration	timedelta: the length of time of the ad

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 then running. Con = ad mentions one or more candidates in a negative way. Mixed: Any ad that mentions more than one candidate in a particular race, with positive content about one or more candidates and negative content about one or more candidates. Important: if the only mention of a candidate in the ad is "I'm so-and-so and I approve this message," and the rest of the ad is an attack on the opponent, that ad will be categorized as "con" and the only candidate listed will be the opponent.

Appendix C: Returns Dataset Columns

state	str
county_name	str
county_fips	str
region_id	str
democrat	int64
republican	int64
other	int64
totalvotes	int64

A FIPS code is a five-digit Federal Information Processing Standards code that uniquely identifies counties and county equivalents in the United States. While it appears to be a number it is more commonly accessed as a string with leading zeros. For more information see.

The votes cast in the election are recorded as democrat, republic or other with their sum being reflected as totalvotes.

Appendix D: The awesome_puppies Module

The awesome_puppies module supplies project dataframes and functions.

- Dataframes: ads_df and votes_df (see Appendix C and D, respectively)
- Region functions: regions, region_fips, region_string
- Vote functions: votes_by_region, votes_by_party
- Ad functions: ad_counts_by_region_and_beneficiary
- Other functions: groupby_agg

`regions()`
: Returns a list of all regions

`region_fips()`
: Returns a series of fips (as a numpy array) by region

`region_string(region_id)`
: Return the display string for a region_id string

`votes_by_region()`
: Returns votes by region dataframe.
The columns are the votes, by party, for each region

`votes_by_party()`
: Returns votes by party dataframe.
The columns are regions

`ad_counts_by_region_and_beneficiary()`
: Returns a DataFrame with ad counts by beneficiary. The result has region_ids as its indices and [Clinton, Trump] as its columns.
Access a region's ad counts by df.loc["<region_id>"]
Access a beneficiary's ad counts by df.loc[<region>].Clinton or .Trump

`groupby_agg(df, groupby_list=[], func='size', fill_value=None)`
: This function generalizes the application of groupby() and aggregate() functions to the passed DataFrame "df". If multiple labels are provided in the "groupby_list", this function unstacks the resulting MultiIndex DF by the number of labels - 1. Aggregate functions are passed as strings to the "func" argument (such as "size", "sum", "mean", etc.) and "size" is applied by default.

```
df.groupby(<groupby_list>).agg(<func>).unstack(<unstack_level>)
```

Appendix E: Visualization Templates

We created templates for these visualizations:

- Bar charts
- Line plots
- Sankey diagrams
- Heat maps
- Tree maps
- Choropleth maps

The various visualization templates can be found throughout the final_report_code/jupyter_notebooks folder in our GitHub repository.

Appendix F: GitHub Repo Structure — Tree Output

```
# This tree output has been edited to summarize our repo structure
# 12/5/21
.
├── Project2_Proposal.pdf
├── Project_2_description.ipynb
├── README.md
├── code
│   ├── Jupyter
│   │   ├── BB
│   │   │   └── # Bronte's Jupyter Notebooks
│   │   ├── CM
│   │   │   └── # Christian's Jupyter Notebooks
│   │   ├── JJ
│   │   │   └── # Jean-Lucas Jupyter Notebooks
│   │   ├── RR
│   │   │   └── # Richard's Jupyter Notebooks
│   │   ├── filter_county_returns.ipynb
│   │   ├── filter_presidential_ads.ipynb
│   │   ├── load_enriched_county_results.ipynb
│   │   └── load_presidential_ads.ipynb
│   └── Python
│       ├── BB
│       │   └── # Bronte's Python scripts
│       ├── CM
│       │   └── # Christian's Python scripts
│       ├── JJ
│       │   └── # Jean-Luc's Python scripts
│       ├── RR
│       │   └── # Richard's Python scripts
│       ├── awesome_puppies.py
│       └── test_pups.py
├── datasets
│   ├── MIT_Election_Lab
│   │   ├── 2016-precinct-president.csv
│   │   ├── enriched_county_pres_2016.gz
│   │   └── region_county_makeup.csv
│   │   └── url_link.txt
│   ├── Political_TV_Ad_Archive
│   │   ├── political_ad_entire_dataset.csv
│   │   ├── political_ad_pres_airing_from_june.gz
│   │   ├── political_ad_unique_entire_dataset.csv
│   │   ├── sponsor_correlation.csv
│   │   └── url_link.txt
│   └── README.md
├── final_report_code
│   ├── jupyter_notebooks
│   │   └── final_report_figures.ipynb
│   └── python_files
│       ├── awesome_puppies.py
│       └── maps.py
├── output
│   ├── Proposal
│   │   └── # Proposal Visualizations
│   └── create_subfolders_as_you_need.txt
├── report
│   ├── BB_figs
│   │   └── # Bronte's Figures for Final Report
│   ├── JL_figs
│   │   └── # Jean-Luc's Figures for Final
Report
│   └── figs
│       └── # Christian's Figures for Final Report
└── tree.txt

26 directories, 116 files
```

Appendix G: Final Project Plan

ID	% Complete	Task Name	Duration	Start	Finish	Predecessors	Responsible	Support
0	**100%**	**w200 Project 2**	**29 days**	**Tue 11/9/21**	**Tue 12/7/21**			
1	**100%**	**Project Proposal**	**7 days**	**Tue 11/9/21**	**Mon 11/15/21**			
2	100%	Draft proposal	4 days	Tue 11/9/21	Fri 11/12/21		B	C, J, R
3	100%	Team edits (Proposal)	2 days	Sat 11/13/21	Sun 11/14/21	2	B, C, J, R	
4	100%	Complete Proposal	1 day	Mon 11/15/21	Mon 11/15/21	3	B	C, J, R
5	100%	Milestone: 10% Turn in Final proposal	0 days	Mon 11/15/21	Mon 11/15/21	4	B	C, J, R
6	**100%**	**Written Report**	**22 days**	**Sat 11/13/21**	**Sat 12/4/21**			
7	**100%**	**Initial Data Exploration**	**3 days**	**Sat 11/13/21**	**Mon 11/15/21**			
8	100%	10% Finalize/adjust questions asked (if needed)	1 day	Sat 11/13/21	Sat 11/13/21	2	B, C, J, R	
9	100%	ID key assumptions	1 day	Sat 11/13/21	Sat 11/13/21	2	B, C, J, R	
10	100%	ID key argument (what is our argument?)	1 day	Sat 11/13/21	Sat 11/13/21	2	B, C, J, R	
11	100%	Create intial Report outline	1 day	Sun 11/14/21	Sun 11/14/21	9,10,8	B	C, J, R
12	100%	Each person signs up for section to focus on	1 day	Mon 11/15/21	Mon 11/15/21	11	B, C, J, R	
13	**100%**	**EDA**	**13 days**	**Sun 11/14/21**	**Fri 11/26/21**			
14	100%	20% Data cleaning/sanity checks	13 days	Sun 11/14/21	Fri 11/26/21	10	R	B, C, J
15	100%	20% Gather compelling figures from data using code	13 days	Sun 11/14/21	Fri 11/26/21	10	B, C, J, R	
16	100%	20% Gather compelling text and data stories	13 days	Sun 11/14/21	Fri 11/26/21	10	C	B, J, R
17	**100%**	**Report**	**8 days**	**Sat 11/27/21**	**Sat 12/4/21**			
18	100%	Draft report (each person writes their sections of reports)	5 days	Sat 11/27/21	Wed 12/1/21	16,8,14,15	B, C, J, R	
19	100%	Team edits (Written Report)	2 days	Thu 12/2/21	Fri 12/3/21	18	B, C, J, R	
20	100%	Final draft of Report (read it once through and approve)	1 day	Sat 12/4/21	Sat 12/4/21	19	B, C, J, R	
21	**100%**	**PowerPoint Presentation**	**8 days**	**Tue 11/30/21**	**Tue 12/7/21**			
22	100%	Create initial presentation layout	1 day	Tue 11/30/21	Tue 11/30/21	18	B	
23	100%	Draft presentation	3 days	Thu 12/2/21	Sat 12/4/21	22	C	B, J, R
24	100%	Team edits (Presentation)	1 day	Sun 12/5/21	Sun 12/5/21	23	B, C, J, R	
25	**100%**	**Presentation Delivery**	**2 days**	**Mon 12/6/21**	**Tue 12/7/21**			
26	100%	Dry run	1 day	Mon 12/6/21	Mon 12/6/21	24	B, C, J, R	
27	100%	Deliver Final Presentation in Class	1 day	Tue 12/7/21	Tue 12/7/21	26	C	B, J, R
28	100%	Milestone: 20% In-Class Presentation	0 days	Tue 12/7/21	Tue 12/7/21	27	B, C, J, R	
29	100%	Milestone: Push final report	0 days	Tue 12/7/21	Tue 12/7/21	28	B	C, J, R