**Introduction**

Political scientists often cite the 1960 Presidential debate between John F. Kennedy and Richard Nixon as the first example of television playing a role in a political election. Televisions broadcasted images of the youthful and energetic Kennedy alongside the ill and uncomfortable Nixon into the homes of millions of Americans. While scholars debate the true impact of these images on the 1960 election, it is undeniable that the way political candidates are portrayed on television has been important to electoral success ever since. Television advertisements are one of the primary means a campaign has to get out its message out and portray its candidate in a positive light or denigrate his or her competitors. In this paper, we will examine the television advertisements broadcasted across the country in the 2016 U.S. presidential primary for both parties. We will address a series of research questions that examine the use, impact, and scope of the time-honored tradition of political advertisements in this 21st century election.

Before we could dive into the research questions, we set out to better understand our data. For this analysis, we started with the Television Ad Airings dataset from the Political TV Ad Archive. This gave us an observational dataset of more than 180,000 rows representing an individual airing of an advertisement. Additionally, we acquired a dataset with the total votes by county for each presidential candidate. The dataset did not contain all states that were able to vote in the primary possibly due to lack of publish results. The next step was to clean the datasets and prepare the datasets for analysis. For the TV Ad Archive dataset, we first removed any advertisements that were in regards to House or Senate races and as well as any which were aired after the California primary as these are likely to be general election-focused ads. Second, we chose to associate each ad with the candidate who was supported by the ad’s sponsor, rather than the candidate the ad was actually talking about. For positive ads, these two are the same candidate, but this is not the case for contrast or negative ads. Our logic was that the fundamental goal of buying television ad time is to promote a particular candidate, party, or ideology and we should gauge the success or failure of an ad in that context, rather than the ads impact on its subject. We also had no way of knowing what the impact of a mixed or negative ad would be and therefore chose to remove these variables from the dataset after an analysis was completed. This also led us to drop observations where the affiliation of the sponsor was either “opposes” or “none”, again not knowing what these categories meant for an ad we would not want the variables to shew our analysis. Ads classified as “none” were largely oriented towards issues rather than candidates. Study of these ads could lead to other interesting findings, but these research questions are out of the scope of this paper. Finally we removed any observations where the sponsor affiliation was unknown. This left us with a dataset of roughly 115,000 for our analysis for the TV Ad Archive dataset. The total votes dataset contained county, fips, and fraction votes that were removed due to limit of scope. Variations of merges between the two datasets were preformed throughout the analysis and can be viewed in corresponding documentation.

**What is the positive/mixed/negative ad message distribution for each candidate and how does it change by state and over time?**

Our first research objective was to take a closer look at the message variable for each of the advertisements in our dataset. Each advertisement is classified as positive (mentions one or more candidates in a positive way with no negative message about any candidate), negative (mentions more than one candidate in a negative way), and mixed (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). By looking at the distribution of these different categories, we can gain some insight into each candidate’s strategic approach and better understand the nature of these two races overall. The results from this analysis are summarized in Table 1 below and ordered by percent positive.

**Table 1:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Party** | **Candidate** | **% Positive** | **% Mixed** | **% Negative** |
| Democrat | Bernie Sanders | 99.9% | <0.1% | 0% |
| Republican | Carly Fiorina | 99.2% | <0.1% | 0% |
| Democrat | Hillary Clinton | 91.5% | 4.7% | 3.6% |
| Republican | Donald Trump | 72.2% | 27.3% | <0.1% |
| Republican | Chris Christie | 68.8% | 24.6% | 6.6% |
| Republican | John Kasich | 68.2% | 25.8% | 1.6% |
| Republican | Marco Rubio | 45.5% | 35.8% | 17.3% |
| Republican | Mike Huckabee | 34.8% | 65.2% | 0% |
| Republican | Ted Cruz | 31.2% | 37.1% | 25.4% |
| Republican | Jeb Bush | 24.7% | 35.9% | 39.3% |

Note: A small number of ads were classified as unknown for the message. These ads will not show up in the above table, but they are included in the total percentage of ads run. This means that some rows add up to less than 100%.

Republican candidates were much less likely than Democrats to run advertisements that were completely positive in nature, with the exception of Carly Fiorina who was not in the race for very long. This makes sense as most pundits would agree that the GOP race was more adversarial and combative than the Democratic contest. Some Republicans like Mike Huckabee tended to focus on creating a contrast between candidates with a mixed message, while others like Ted Cruz and Jeb Bush utilized both mixed and negative advertisements to bring down support for their opposition.

In order to dive deeper into this message question, we will look at how a majority positive candidate (Hillary Clinton) and a majority negative candidate (Ted Cruz) distributed their message across states. These were both candidates who stayed in the race through all of the contests in our dataset, but made much different strategic messaging choices. Table 2 shows each candidate’s messaging percentages by state and is ordered chronologically by when the primary was held.

**Table 2:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **State** | **HC**  **% Positive** | **TC**  **% Positive** | **HC**  **% Mixed** | **TC**  **% Mixed** | **HC % Negative** | **TC % Negative** |
| IA | 92.6% | 37.4% | 6.4% | 32.5% | 0.6% | 23.8% |
| NH | 90.5% | 57.6% | 8.6% | 40.6% | 0.7% | 0% |
| SC | 99.1% | 26.1% | 0.2% | 30.3% | 0.6% | 38.0% |
| NV | 99.4% | 37.8% | <0.1% | 27.8% | 0.2% | 0% |
| VA | 86.2% | 38.5% | 1.5% | 30.8% | 12.3% | 38.5% |
| MA | 100% | 6.9% | 0% | 93.1% | 0% | 0% |
| OH | 99.7% | 18.2% | 0.2% | 81.8% | <0.1% | 0% |
| NC | 99.9% | 1.3% | <0.1% | 98.7% | 0% | 0% |
| FL | 92.6% | 2.0% | 1.0% | 57.1% | 0.2% | 40.8% |
| NY | 71.2% | 9.1% | 26.0% | 54.5% | 2.9% | 18.2% |
| PA | 79.2% | 1.1% | 5.1% | 97.3% | 15.7% | 0.5% |
| CA | 34.8% | 26.3% | 3.8% | 60.7% | 59.7% | 11.2% |
| **Tot.** | **91.5%** | **31.2%** | **4.7%** | **37.1%** | **3.6%** | **25.4%** |

Note: A small number of ads were classified as unknown for the message. These ads will not show up in the above table, but they are included in the total percentage of ads run. This means that some rows add up to less than 100%.

Starting with Hillary Clinton, we can see in the table above that the vast majority of her ads were positive. The positive percentage is consistently above 85% until the last three contests where it dips into the 70-79 percent range for New York and Pennsylvania and then drops all the way down to 35% in California. This suggests that Clinton’s team and supporters decided to go more negative with the ads as the contest drew to a close in order to ensure victory and push Bernie Sanders out of the race.

Ted Cruz on the other hand went mixed and negative from the start, seeking to draw a strong contrast between him and the other candidates. His mixed vs. negative balance was done strategically with some states such as New Hampshire, Nevada, and Ohio receiving only mixed, while South Carolina and Virginia residents also saw a steady stream of solely negative advertising. Despite all this negative campaigning, Cruz and his supports kept at least a quarter of his ads positive in an effort to initially present him well to the voters. However, as the race dragged on and he became more of a known quantity, his ads shifted towards only mixed and negative with his positive percentage never getting above 20% again until the final contest.

Please be aware that due to lack of insight into the what qualifies as a negative ad or mixed ad no further analysis was done on pro, negative and mixed ads. All further analysis in this report will be done on pro ads.

**For which states and candidates do the share of advertisements correlate the strongest with the share of the vote received?**

The ultimate goal of for campaigns and Super PACs when they pay for television advertisements is to increase their candidate’s share of the vote, whether by convincing more voters to support their candidate or discouraging voters from supporting other candidates. If this is true, we should see those candidates who have a larger share of the television time in an election receive a larger share of the vote totals. However, there are a lot of other factors that go into determining who the voters support such as ideology, party, earned media, and the persuasiveness of the advertisements, so the relationship between ads and votes is never exact. By aggregating advertisement percentages by party, candidate, and state and then comparing those rates to a separate voting results dataset, we can determine the correlation between these two variables. It is important to note however that due to prescience of so many other factors and the simplicity of correlation coefficients, these values represent simply associations and not necessarily causal relationships. Table 3 shows the results of this analysis.

**Table 3:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Breakout** | **Correlation Coefficient** | **Breakout** | **Correlation Coefficient** |
| Overall | 0.6725 | Iowa | 0.9219 |
|  |  | New Hampshire | 0.7011 |
| Democrats | 0.5502 | South Carolina | 0.7085 |
| Republicans | 0.4202 | Nevada | 0.8947 |
|  |  | Virginia | 0.8027 |
| Bernie Sanders | 0.4677 | Massachusetts | 0.8451 |
| Hillary Clinton | 0.4696 | Ohio | 0.9499 |
| Donald Trump | -0.1491 | North Carolina | 0.8103 |
| John Kasich | 0.8994 | Florida | 0.3720 |
| Ted Cruz | 0.5368 | New York | 0.3198 |
| Marco Rubio | 0.7830 | Pennsylvania | 0.8794 |
| Jeb Bush | 0.9996 | California | 0.3054 |

The moderately strong nature of the overall coefficient supports our hypothesis that share of the advertisements is an important factor in determining vote percentage, but it is still one of many factors.

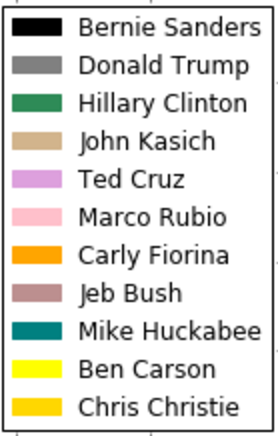
When we break this correlation down by party and candidate we see that while Democrats overall have a stronger correlation than Republicans, Bernie Sanders and Hillary Clinton have the second and third smallest coefficient among all candidates. This is because Donald Trump’s coefficient is so different from the rest of the group that it brings down the overall GOP relationship. Donald Trump’s vote percentage had basically nothing to do with how many advertisements were aired on his behalf. This result also makes sense given our understanding of the election as the Trump campaign spent much less money than his competitors and instead relied heavily on earned media and social networks to gather support.

At first glance, the state breakout does not yield much in the way of interesting insight. States that are similar in geographic size, population, and demographics can either give us very similar results in the case of Massachusetts and Pennsylvania or very different results in the case of Florida and South Carolina. It is only when we listed the correlations in chronological order of when the election was held that a pattern did emerge. From Iowa to North Carolina, the coefficients of elections held in the first month and an a half of the race never dipped below 0.70, meaning that there was a consistently strong relationship between the share of advertisements and the share of the vote. This all changed with Florida as three of the next four elections in our dataset all had coefficients under 0.40. At this point in the race, Donald Trump was winning more and more contests despite the fact that Super PACs friendly to his opponents were pouring money into advertisements aimed at defeating him. His momentum towards the GOP nomination could not be halted by the ad buys and thus the correlations in most of the next few states we substantially lower than in the previous few states. This is also partially why Ted Cruz has the second lowest coefficient among GOP candidates. Since he became the sole challenger to Trump as the race wore on, most of the “Never Trump” spending was by definition in support of him.

**Did the number of total pro TV ads aired for a presidential candidate by state and the total number of votes received for that candidate by state show any unique trends? Did any of the candidates receive a considerably high number of votes, but a low number of ads aired?**

To look more closely at the sponsor TV ads aired for single candidate and their relationship to votes received we created a bubble chart to display the candidate’s total number of votes received by state as the x axis and the state as the y axis. The states were also ordered in sequence of the date of the state’s primary election. The bubble radius is determined by the total ads aired within the state sponsoring that candidate. This graph is meant to show the overall distribution of votes by state for each candidate and display the difference in the total number of ads for each candidate by the size of the radius of each bubble. The dataset was made by joining the primary election dataset and TV Ad Archive dataset.

**Figure 1:**

****

The graph details that while many of the sponsored ads for republican presidential candidates such as Ben Carson, Chris Christie, Carly Fiorina, Jeb Bush and Mike Huckabee have low numbers of ads aired or no ads aired, most likely due to these candidates dropping out the election, total sponsored ads by state and total votes received by candidate for that state do not seem to have a direct visual trend. This may be due to, as stated above, Donald Trump’s total votes not being dependent upon TV ads aired and the large number of republican presidential candidates either dropping out of the election early or having little no sponsored candidate TV ads. The graph does detail the first state to vote within the primary and last states to vote within the primary were the most heavily advertised to. The last states to vote having a larger amount of advertisements could be due to the increased amount of time in which ads could be aired or due to sponsors increasing ads in these states. Further analysis would need to be completed to give further insight as to why more sponsored single candidate ads aired.

A small caveat for the above graph is that you cannot easily decipher if candidates followed a similar trend in total votes received and total ads aired. To look more closely at the two democratic presidential candidates within this dataset we created two bubble graphs Hillary Clinton and Bernie Sander. Upon further analysis we found that Hilary Clinton and Bernie Sanders sponsored TV ads per state and total votes per were similar. Both Hilary Clinton and Bernie Sanders had a low number of total sponsored TV ads aired in California, but each received the largest number of votes. This could be due to the large number of votes allotted to California, that California has a history of being blue state and that a majority of the republican candidates had dropped out by the time of the California primary.

Additional graphs were made to show the distinction between the two.

Figure 2, 3, 4:





The final bubble graph was made to show the very similar nature of votes received to number of ads aired by state between the two candidates and acts as a visual representation of how close the primary election was between Sanders and Clinton.

**How does the distribution of outside group vs. candidate committee support (in the form of ads run) change as Election Day approaches? Does this vary by party?**

In our cleaned dataset, advertisements are paid for by one of two entities: the candidate’s campaign committee or a Super PAC. Campaign finance laws cap the amount of money that committees can spend on advertisements while Super PACs do not have these limits. Both kinds of groups seek to use the funds they have available as strategically as possible to influence the voters. One key aspect of voter persuasion is the amount of time between when a voter sees an ad for your candidate and when they go to the polls. We looked into how each entity uses its resources to persuade these voters in the lead up to the election by calculating the percentage of ads paid for by each group over different time periods. The overall results are shown below in Table 4.

**Table 4:**

|  |  |  |
| --- | --- | --- |
| **Time to Election** | **Super PAC Ad %** | **Candidate Committee Ad %** |
| More than 2 Months | 55.2% | 44.8% |
| 1-2 Months | 30.9% | 69.1% |
| 14-31 Days | 32.5% | 67.5% |
| 7-14 Days | 30.6% | 69.5% |
| 1-7 Days | 26.1% | 73.9% |
| Election Day | 26.9% | 73.1% |

We see in the table above a very consistent relationship between the two percentages. When the election is still a long way off, candidate committees seem content to conserve their resources and allowing Super PACs to fill the airwaves with advertisements. These ads may support the committee’s candidate to a degree, but they are not shaping their content. However, as the election draws closer, Super PACs seem to take a backseat to candidate committees as the specific messaging becomes more important. Table 5 below breaks out Table 4 by party.

**Table 5:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Time to Election** | **GOP Super PAC Ad %** | **DEM Super PAC Ad %** | **GOP Cand. Comm. Ad %** | **DEM Cand. Com. Ad %** |
| More than 2 Months | 78.8% | 23.1% | 21.2% | 76.9% |
| 1-2 Months | 72.3% | 1.0% | 27.7% | 99.0% |
| 14-31 Days | 63.6% | 0.1% | 36.4% | 99.9% |
| 7-14 Days | 63.6% | 0.7% | 36.4% | 99.2% |
| 1-7 Days | 65.7% | 0.1% | 34.3% | 99.9% |
| Election Day | 57.7% | 0.1% | 42.3% | 99.9% |

These results are even more illuminating. Overall, Democrats in the 2016 presidential primary relied on Super PACs for television advertisements much less than their GOP counterparts. Part of that is due to the fact that the Democratic runner-up Bernie Sanders publicly denounced Super PACs and refused their support, but these percentages mean that even Hillary Clinton had much less Super PAC support than any of the Republican Candidates. Even on Election Day itself, Republicans were relying on outside groups for more than half of their advertisements. Overall, these results speak to the growing rift in the perception of Super PACs by our two political parties.

**What day of the week were the most ads run and does this vary by sponsor ads in support of only one candidate?**

With Table 5 displaying the relationship between Super Pac Ads and Candidate Ads we wanted to look at the total Candidate ads aired for a candidate and see if a certain day happened to air more ads in comparison to the other days of the week. For this analysis we further broke down ad sponsorship from the TV Ad Archive dataset to include only pro sponsored ads for a single candidate. The x axis represents total candidate ads aired, while the y axis represents the candidate and is labeled by the candidates’ initials. It is important to note that the dataset used to create this graph only used sponsored ads associated with a single candidate, within the dataset you will see sponsored ads supported groups and pairs of candidates.

**Figure 5:**

****

As shown in this multi-bar graph, Bernie Sanders had the most sponsored TV ads at 5600 ads in total aired on Friday during the primary election, and the most ads aired in total. Another interesting finding was that Friday happened to be the day that the most ads were aired for the majority of the candidates with a total of 20115 being aired of the primary election. It is also interesting to see that Donald Trump lacked sponsored TV ads in comparison to Marco Rubio and had the most ads aired on Tuesday. The republican candidates in general have a smaller number of ads aired in comparison to the democratic candidates.

Please visit our GitHub repo for additional insights about the Political Ad Archive dataset and the analysis completed.