Political Ads in the 2016 Presidential Election

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

[CHRISTIAN]

Hello, we are Team Awesome Puppies and our project investigated Political Ads in the 2016 Presidential Election

Our group consisted of Bronte Baer, Jean-Luc Jackson, Rich Robbins, and me, Christian Montecillo

Agenda

Project Description

Cleaning and Refining the Data

Data Insights and Questions

Conclusions

[CHRISTIAN]

This is what we're going over today:

We'll talk about our project.

How we cleaned and refined our data.

The insights we gained by asking questions about the data.

And finally our conclusions.

Project Description

[CHRISTIAN]

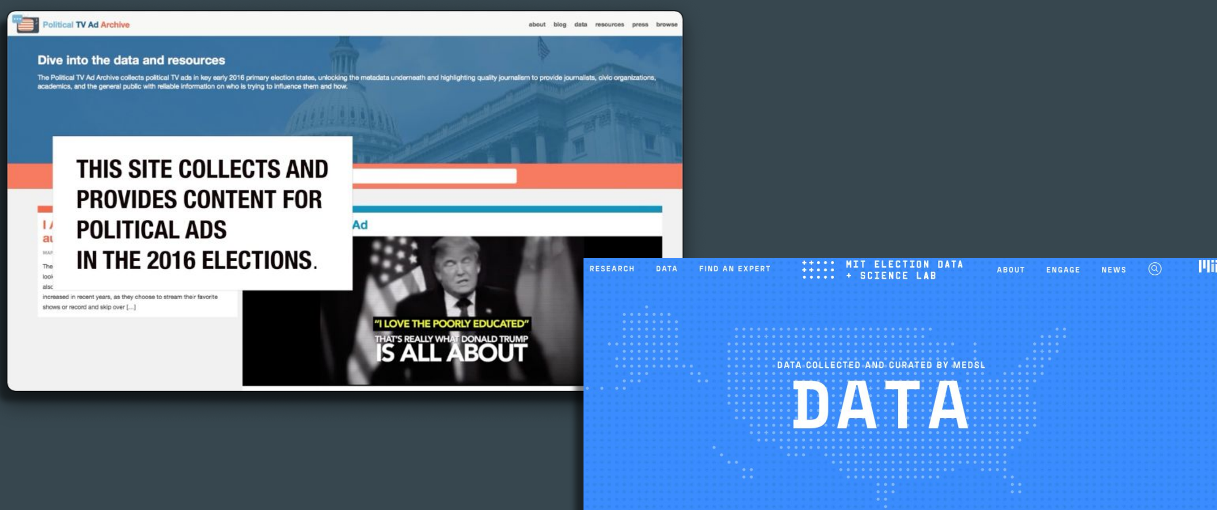
In 2016, the underdog won. Donald Trump beat Hillary Clinton 304 to 227 in Electoral College votes even with a deficit of ~2.86 million popular votes. How this occurred remains a topic of great debate.

Our project analyzed election-related TV ads, you know -- the ones that consume our airwaves in the months leading up to Election Day.

Our primary dataset included information for 85,000 advertising broadcasts in 13 regional broadcast markets in the United States as seen on this map. One thing to note as we discuss our project: the regions cover only a portion of the country and do not align with single states or cities. As an example the Denver Region covers a majority of the counties in Colorado and parts of Wyoming and Nebraska.

To focus our analysis, we looked at four broad categories which I'll introduce later and we limited our analysis to the presidential general election (which excluded the primaries) and was limited to Clinton and Trump.

Cleaning and Refining the Data

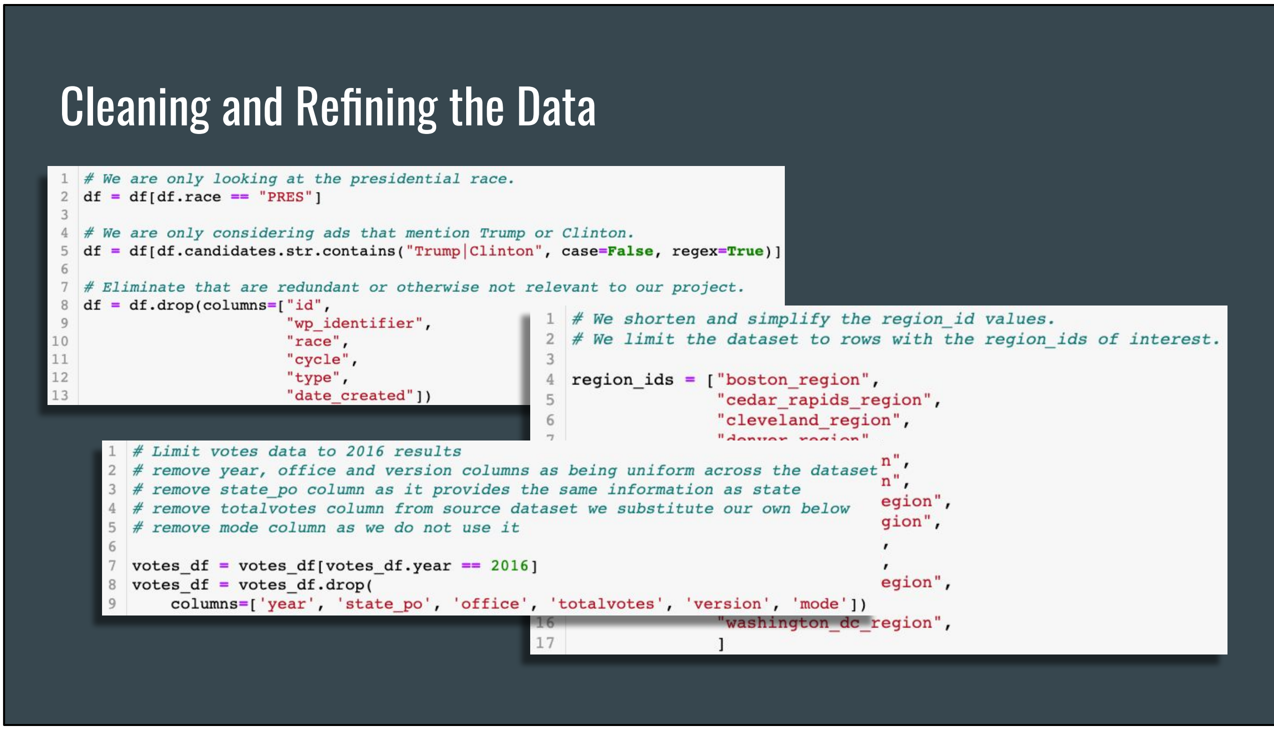


[RICH]

Our primary dataset was assembled by the Political TV Ad Archive covering the broadcast of presidential and senate ads for the 2016 elections in select parts of the country.

[CLICK]

We supplemented the advertising dataset with an election returns dataset from The MIT Election Data and Science Lab. We used 2016 election return data from a broader dataset.



[RICH]

I will highlight the three most important things we did when we cleaned the data, our written report has much more detail about how we removed data that was redundant or not relevant to our work.

First:

The advertising dataset does not indicate which candidate benefits from an ad but indicates who sponsored the ad.

We used public information about sponsors to identify the intended beneficiary of each ad.

This derived data field was critical to our work.

Second:

All time data was presented to us in UTC format, in order for us to do important time-related work, we localized all times.

Third:

As Christian mentioned, the regions in the dataset do not correspond to particular states or cities, rather, they are a collection of 268 counties.

So, we used public data to identify the counties in each of the regions in order to determine the voting results by region.

This derived data field was critical to our work.

Data Insights
Four Main Topics

1. Volume
2. Subject Matter
3. Tone
4. Timing

[CHRISTIAN]

For insights, we covered four broad topics:

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

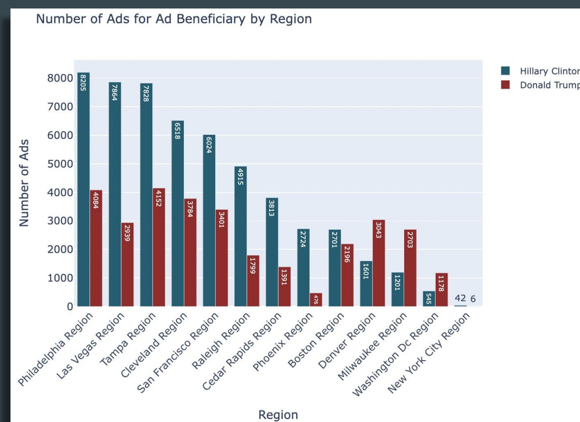
Subject Matter
Are there certain subjects each candidate covered more than others?

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

Timing
How did advertising intensity change as the election drew nearer?

Data Insights: Volume

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



[CHRISTIAN]

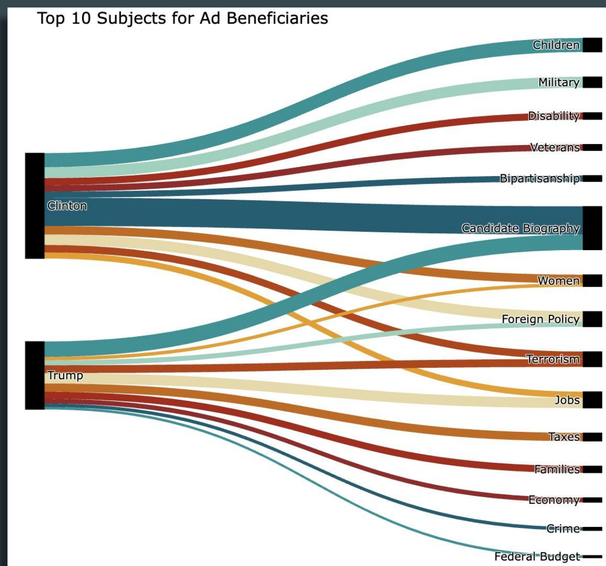
Let's talk about the Volume of ads!

Highlighted in red in this table, we see that in some regions, the candidate who aired the most ads was *not* the candidate who received the most votes.

[CLICK]

Take a look at Phoenix. Clinton aired five times more ads yet still lost by over 7%. Arizona had voted for the Republican candidate in the four previous presidential elections. Clinton's advertising focus in the region suggests that her team thought it could reverse that pattern, but they failed. Trump's camp wisely focused its energy elsewhere.

Data Insights: Subject Matter



[BRONTE]

On to Subject Matter:

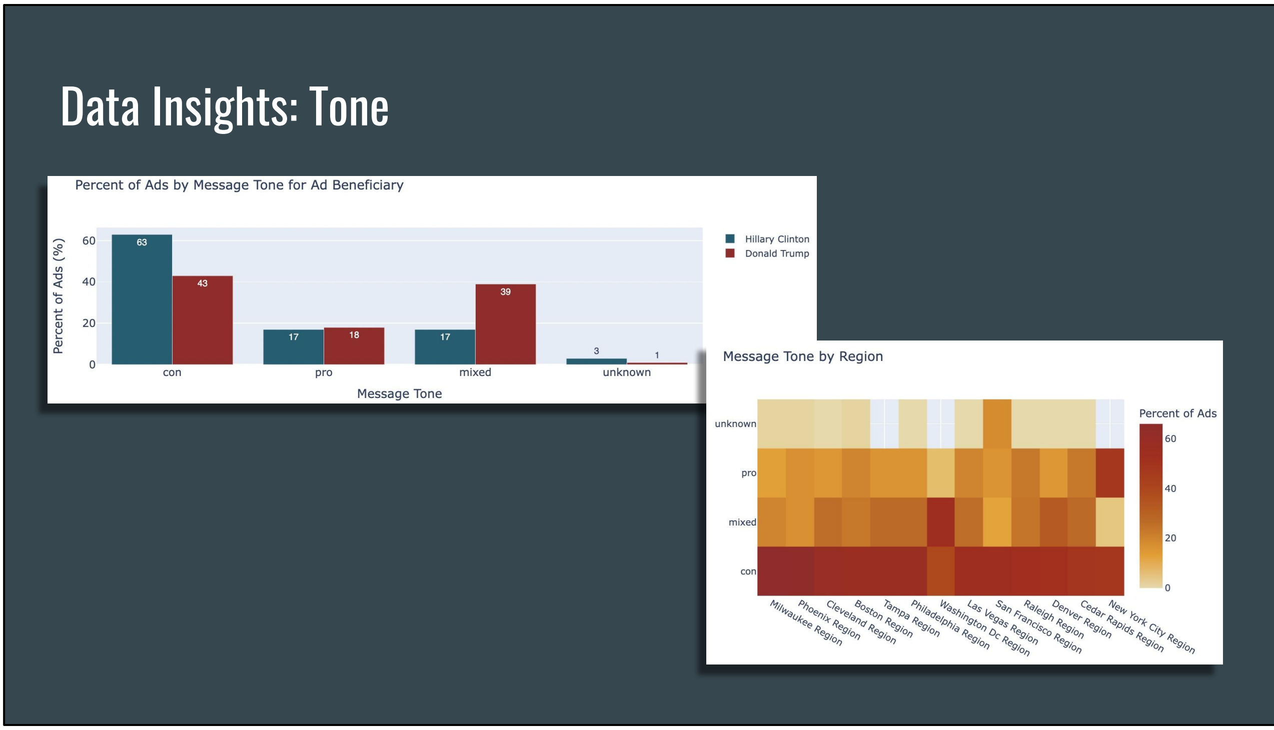
Which candidate covered which subjects? Was there overlap?

Analyzing ad subject matter was more complex than we anticipated. Almost all of the ads had more than one subject so 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.

The Sankey diagram you are looking at here shows the flow, volume, and overlap of the candidates' top ten subjects covered in their ads.

“Candidate Biography” was always the most prevalent subject matter. 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.

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.



[BRONTE]

How about the tone of the messages?

Did one candidate air more negative messages than the other? What about by region?

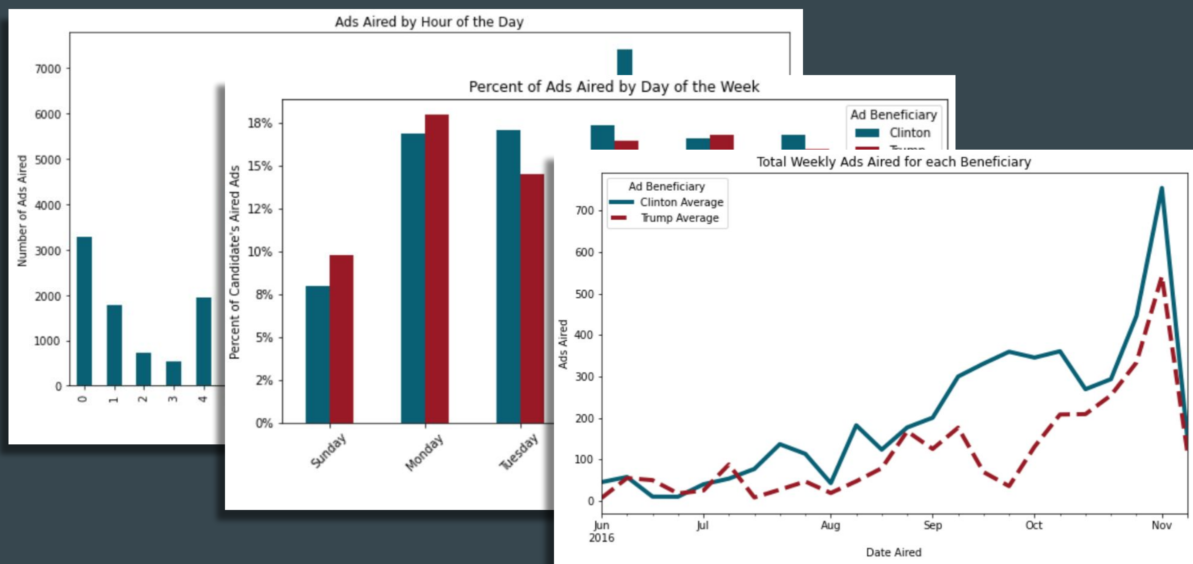
One point to note is that we chose to normalize this plot in order to give a fair comparison between the candidates. We know from the volume of ads section Christian covered, that Clinton aired more ads by count than did Trump which is why we presented this plot as percentages.

Not surprisingly, both candidates aired a higher percentage of ads with negative messages. Initially you might say that Clinton aired a higher percentage of negative ads; however, considering that “mixed” message ads also contained negative messaging, about 80% of ads by each candidate had negative tones to them.

We did the same normalization for the message tone by region. Since there were some regions that had a much smaller number of ads recorded in our dataset.

You can see that all the regions had a higher percentage of negative ads than positive ones.

Data Insights: Timing



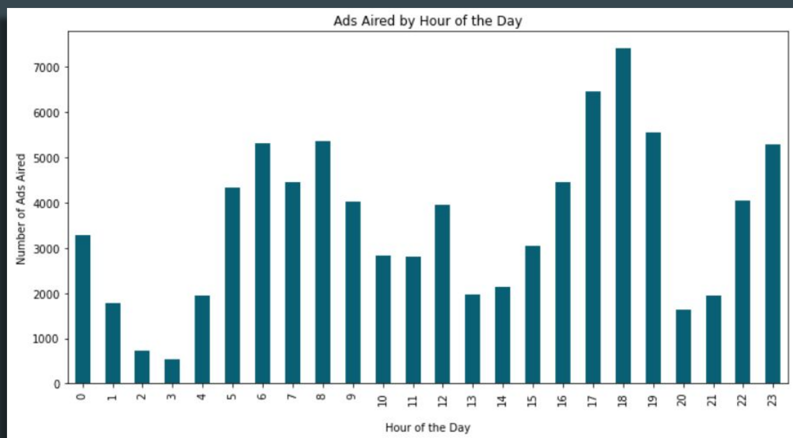
[JL]

Next we investigated the timing of ads

We looked at ads —

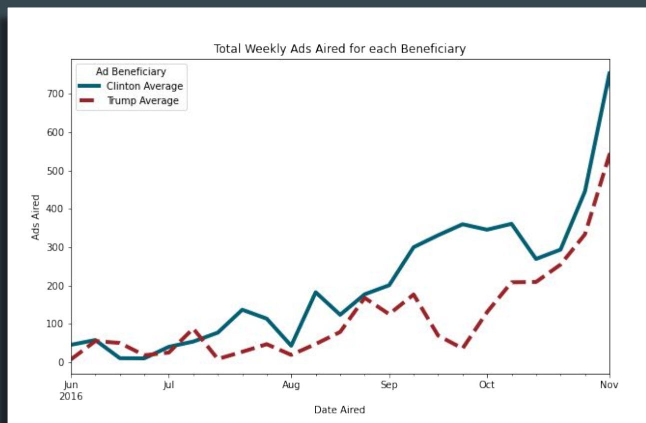
[CLICK]
By the hour [CLICK]
By the Day [CLICK]
And over time in the election season

Data Insights: Timing

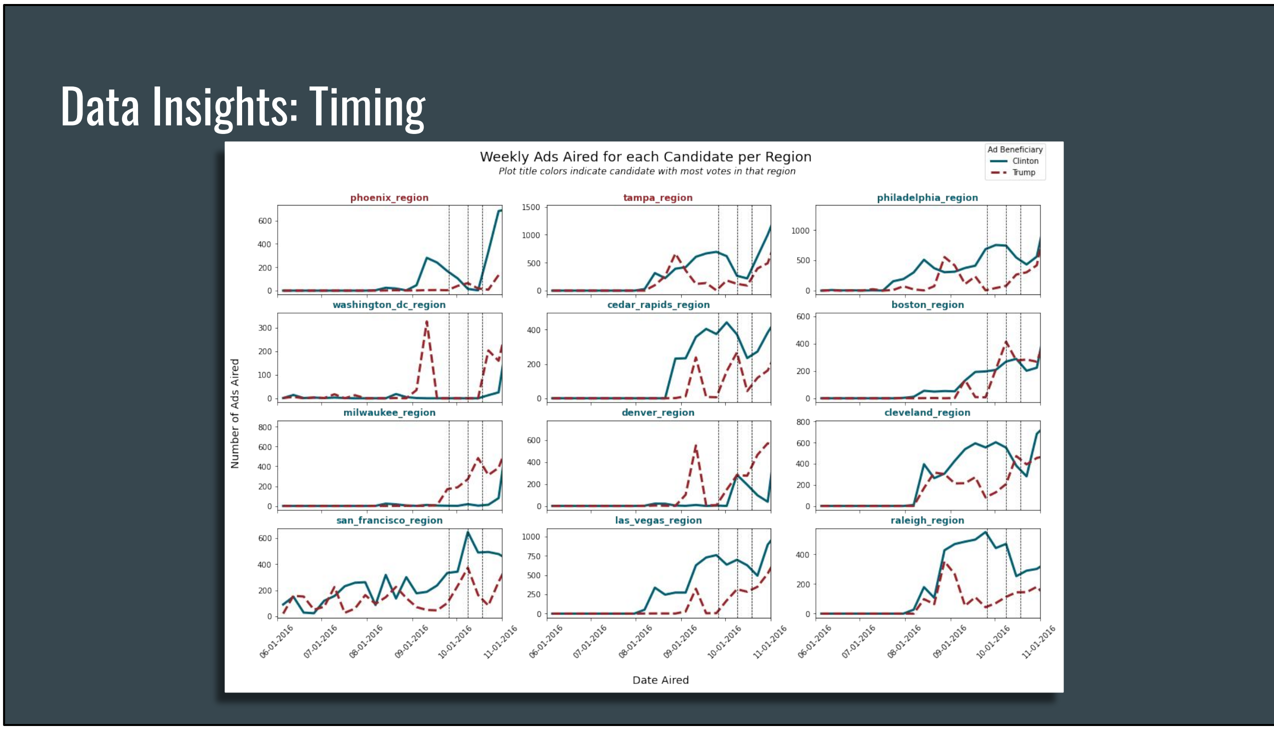


The broadcasting of ads peaked around 6 AM and 6 PM and dropped at 3 AM and 1 PM. This visual “rollercoaster” seems to follow a traditional workday

Data Insights: Timing



Here we see how advertising intensity changed as the election got closer. As you'd expect, ad volume increased significantly in the month or so prior to Election Day, crescendoing the week leading up to the election.



[JL]

This was true in all regions we inspected.

Each subplot here has a title whose color indicates which candidate received the most votes (blue = Clinton, red = Trump)

- The vertical black dotted lines show the 3 presidential debates

I wanted to highlight a couple of places.

Phoenix

- In the Phoenix region, when compared to Trump, Clinton aired many more ads.
- Trump received more votes here.

DC & Milwaukee

- The inverse happened in Washington DC and Milwaukee, where Clinton's campaign barely advertised until very close to Election Day.
- Both went to Clinton.

In general, we found it interesting how advertising peaked or valleyed based on the debates — indicating changes in advertising strategy.

- Looking at **Raleigh,** for example, Clinton's ads dropped noticeably after each debate

[CHRISTIAN]

So what did we learn?

Because of the characteristic of the original dataset we used, we can't conclusively say that there is a strong correlation between the number of ads a candidate broadcast to the number of votes that candidate received.

We saw major disparities where the candidate with the most ads did not receive the most votes.

Bronte showed you in her Subject Matter discussion that both candidates wanted you to get to know them and a large part of each of their messages discussed their bios.

Bronte then told you that the story around Ad Tone was the same. Both candidates overwhelmingly sending negative messages about the other and didn't focus much on any of the positive traits about themselves.

And finally from Jean-Luc's Timings discussion, we saw how each candidate's ad strategy shifted after each presidential debate.

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.

Thank you

Custom Module

[RICH]

After all the initial data cleaning, there was still some additional data manipulation depending on the questions being asked so….

[CLICK]

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.

This was really an integral part of our team’s ability to code our visualizations in a way that was consistent with each other because we were all working together from the same foundation. This saved us a lot of time and effort and, likely, mistakes in consistency.

And if anyone is wondering, the module was named after our team’s informal name.