

# The Effect of Facebook Ad Spending on Elections: Research on the 2020 United States Presidential Elections

**Team #19**

Jackson Duke | Saad Rasheed | Kelly Schweinsberg-Geffen | Matthew Chang

November 20, 2022

## Introduction

With the growth in social media usage, we see several ways it affects the society we live in. As more users adopt this form of technology, companies and individuals alike are looking for ways to leverage it to their advantage. One of the biggest ways we are seeing it being utilized is via ads. Social media companies, like Facebook, Twitter, Snapchat, and more, make revenue by placing advertisements throughout their platform that are paid for by companies or individuals. Ad spending on social media is projected to reach over \$173 billion in 2022 and this will account for 33% of all digital advertising spending in 2022 making it the second biggest market in digital ads. Facebook alone accounts for over 1/3 of that spend in 2022.

With us having congressional elections this year, our group aims to examine political advertisement spending on Facebook and the impact it has on elections. Whenever elections come around, voters are bombarded with political advertisements in an effort to sway opinions. Social scientists have long been interested in the effects of political mass persuasion and communication. With the advent of social media and its rapid increase in consumption, advertisement in this space has become an enormous market to reach.

## Problem Overview

With billions of dollars being spent on advertisements for political campaigns, we wanted to analyze if the money spent was worthwhile and if it is, where does it have the biggest influence. We believe this research can help impact business decisions such as: donation amounts to political campaigns from corporations and individuals, the influence of advertising on shaping societal support and views, and the usefulness of spending to put ads on Facebook. After conducting this project, we believe we will be able to give a tangible quantification of the impact of ad spending on election results which in turn demonstrates the influence of digital marketing on overall public perception and views. More broadly, companies will be able to utilize our findings to understand the overall efficacy of digital marketing. For Meta, they will be able to quantify how much influence and impact their platforms have on society.

## Literature Survey

Prior to jumping into our approach, we wanted to cover some other approaches which we researched that have influenced our path forward.

### The Effect of Social Media on Elections: Evidence from the United States

Several social media sites and applications have gained popularity in recent years. In the last decade, the social media site Twitter gained significant traction with an external event: South by Southwest festival (SXSW) as evidenced in “The Effect of Social Media on Elections: Evidence from the United States” by Fujiwara, Muller, and Schwarz (2022). In their research paper, Fujiwara, Muller, and Schwarz comment on the impact of social media on political outcomes with a two-stage least squares (2SLS) regression analysis.

Fujiwara, Muller, and Schwarz’ (2022) analysis was able to provide empirical evidence that social media can affect voter and election turnout in the United States. Fujiwara, Muller, and Schwarz first provided background evidence that the sharp rise of Twitter popularity with the SXSW festival is uncorrelated with previous trends in election results and county characteristics. From there, they outlined an analysis of the social media platform’s effect on voters, and specifically its effects on independent and moderate voters. Their method focuses on the addition of Twitter to counties after the SXSW festival compared to similar counties that had no increase

in Twitter use after SXSW. The analysis points out that counties that had increased Twitter use after the SXSW festival had lower Republican voter turnout, especially for the 2016 and 2020 elections. They point to several reasons for this including the fraction of voter ages that use Twitter as well as the number of broadcasted tweets with political content. These factors and their coefficients are all accounted for in their robust two-stage least squares regression equations that estimate (1) the logarithm of the number of Twitter users in counties and (2) election outcomes. From their analysis of recent presidential elections, their research led to the following conclusion: “We estimate that a 10% increase in a county’s number of Twitter users lowered the vote share of Republican presidential candidate Donald Trump by 0.2 percentage points (p.p.) in both the 2016 and 2020 presidential elections”. This study answered similar research questions to our proposed topic and employs a modeling structure that aligns with our present approach.

### Social Pressure on Social Media: Using Facebook Status Updates to Increase Voter Turnout

Another article that was analyzed and found to relate to our topic is one that also looks at Facebook’s influence on voter turnout. “Social Pressure on Social Media: Using Facebook Status Updates to Increase Voter Turnout” by Katherine Hanschen focuses on Facebook’s ability to be used as a communication channel to influence and increase voter turnout, effectively surpassing traditional communication styles and their influence.

Hanschen approaches her analysis of social media influence on voter turnout by analyzing the effect of Facebook statuses with “social pressure messaging”. Her research provides evidence that “tagged subjects in voting reminders on Facebook with social pressure messaging produced substantial gains in turnout, amounting to percentage point effects ranging from 15.8 to 24.3 over the control group” (Hanschen, 2020). Hanschen conducted her analysis through three studies that all utilized logistic regression modeling. The logistic regression analyses were used to determine if the treatments (exposure to various political status updates) affected participation and/or voter turnout at different points during the election cycle. She executed three unique approaches to status updating and included multiple variations of populations in her experiment. The studies focused on several styles of status structure, including peer-to-peer social pressure (in which confederates posted status updates in line with “get out the vote” mentality to participants), as well as direct post on-page and pride and shame tactics to more specific audiences. The results were evident in the studies. Hanschen used chi-square tests of independence to validate her results and concluded “These experimental results demonstrate that Facebook status updates can be used to increase voter turnout, producing gains in participation in excess of what has traditionally been found from face-to-face voter contact.” (Hanschen 2020). While our approach to the research question does not target status updates only, our procedures and goals for this project are very similar to Hanschen’s study.

## Data

To perform our study, we are using several datasets.

- Facebook Ad Library
  - We utilized the Meta Ad Library to extract political advertisement data per presidential candidate (specifically Joe Biden and Donald Trump) that reveals amount spent, audience reached, and region and demographic distribution of that audience in the time leading up to the 2020 Presidential Election.

- Harvard Dataverse – 2020 Precinct-Level Vote Counts
  - We also collected data from Harvard Dataverse for the 2020 election votes which we will aggregate up to the state or region level.
- Independent Voter Project, State Disclosures, PEW Research – Voter Party Registration
  - To provide a baseline of votes for each candidate, we decided to use the voters registered to the democratic or republican party in each state. Subtracting these voters out of the total registered population provides a pool of “undecided” voters who could go between each candidate. Of the 51 states/DC, 33 disclosed these numbers for 2020. In 18 states, registration by party is not done, so we used the estimates from the PEW Research Institute to estimate the democratic and republican populations.
- US Census Bureau
  - We utilized tables from the U.S. Census Bureau to obtain data for Voting and Registration by Age and Gender at the state level for the 2020 and 2016 election. This data was then aggregated to the region level for ease of modeling output analysis.

We will use these datasets in conjunction to answer our primary research question: Does increased political ad spending on Facebook (Meta) correlate with larger voter turnout? In addition, we aim to answer our supporting research questions:

1. Does political ad spending influence one demographic more than others?
2. Does political ad spending influence one state/region more than others?

### Missing Data

In our election votes dataset from the Harvard Dataverse, we were missing the demographics of the individual voter. Initially, we were unable to find any other robust and reliable datasets which contained this information. Through the US Census Bureau and some data massaging, we were able to find and utilize a sufficient dataset to analyze advertisement spending and its effect on voter demographics.

### Data Wrangling

The cleaning and preparation of our data have taken a considerable amount of time. Since we are working with large raw datasets, several of which are 900K+ rows in size, we had to clean and parse the data while also finding ways to aggregate at levels which translated across all datasets. This needed to be approached strategically, as there were memory and computing constraints with the compounding of operations. The different formats of the data sources in conjunction with the sheer scale of the data proved to be a challenge when wrangling the data. However, we have managed to clean and combine the necessary datasets in meaningful ways for our research questions, resulting in several tables which we use for our modeling in R.

The raw data from Meta ads required parsing of Json objects to properly classify regional statistics of audience reach. The data was flattened, transformed, and transposed to produce an output that was compatible with the Harvard Dataverse dataset. This was accomplished through Python programming, using tools such as pandas and NumPy, to produce a readable result in csv/xlsx, which would be further used in R for combining and then regression modeling.

The raw data from Harvard Dataverse required cleaning by way of aggregation. 50 states were combined into regions (states) to simplify the data and match it to the schema of the Meta data. After combining the tables via matching attributes, we successfully created a data set for Facebook ad audience reach in U.S. regions alongside voter registration and turnout in those

regions. Separately, we kept the cleaned Meta political ad data to drill down into the demographic specifics of those regions to identify any factor correlations.

The raw data from US Census Bureau was cleaned by utilizing Alteryx. This was accomplished by combining columns, merging and joining datasets, summarizing by states and regions, transposing, and formatting to produce an ultimate dataset with the Meta ad data to synergize with age and demographic analysis.

## Initial Hypothesis

We expect the number of votes for the candidate to increase based on their ad spending. Some variables that we are considering are age, gender, and region. With more advertisements targeted at older voters, we expect there to be a larger impact on voter turnout in that demographic – the same can be expected of region-specific political advertisements.

## Experiment 1: Ad Spend Influence on Undecided Voters

### Modeling

We utilized linear regression as the model to analyze our Primary Research Question, along with its supporting questions. Two separate datasets were created, one for Joe Biden and one for Donald Trump. For each model, we used a combined dataset consisting of 3 tables: ad spending by candidate, votes per candidate by state and region, and overall voter registration by state and region. The combined dataset for each candidate will contain the following fields: state, region, ad dollars spent on Facebook ads in the state, and the undecided votes won to capture the increased votes obtained, calculated by subtracting the total votes by the registered voters for that party. A sample output of the Biden table is shown below.

	state	region.x	biden_ad_dol	undecided_won
	<chr>	<chr>	<dbl>	<dbl>
1	ALABAMA	South	421074.7	-402013
2	ALASKA	West	122274.7	75114
3	ARIZONA	Southwest	4979514.9	379059
4	ARKANSAS	South	286092.5	335424
5	CALIFORNIA	West	7436104.8	1418784
6	COLORADO	West	1603606.8	676332

After creating the combined table, we fit a linear regression model using the undecided votes won as the dependent variable and the ad spending and region as independent variables. The following code was used:

```
biden_merged.lm <- lm(undecided_won ~ biden_ad_dol + region.x, data = biden_merged)
trump_merged.lm <- lm(undecided_won ~ trump_ad_dol + region.x, data = trump_merged)
```

We then performed a log transformation on the ad spend by state to see if a better fit was made.

## Results

For both candidates, a linear-log model seemed to provide a better fit to the data, as shown in the R-squared and Adjusted R-squared metrics:

	lin-lin	lin-log	lin-log (no outliers)		lin-lin	lin-log	lin-log (no outliers)
--	---------	---------	--------------------------	--	---------	---------	--------------------------

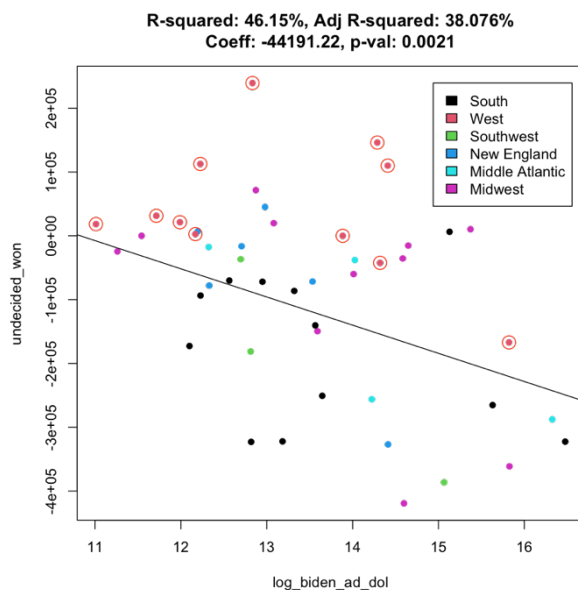
R-squared	31.71	39.42	46.15
Adj R-squared	22.18	20.72	38.08
Ad P-value	0.02	0.04	0.002
Ad Coefficient	-0.02	-38092	-44191

Biden Dataset Model Testing

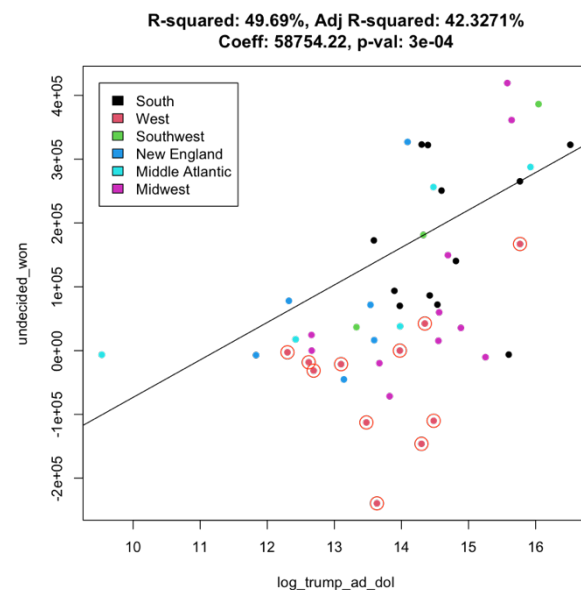
R-squared	37.29	37.39	49.69
Adj R-squared	28.74	28.85	42.33
Ad P-value	0.001	0.001	0.0003
Ad Coefficient	0.0283	66912.09	58754.22

Trump Dataset Model Testing

Both lin-log models found two variables to be statistically significant at the 95% confidence level. For Biden, Ad Spend by State (p-val = 0.002) and the West Region (p-val = 0.048) were significant. For Trump, Ad Spend by State (p-val = 0.0003) and the West Region (p-val = 0.005) were significant as well.



Biden Dataset Visualization



Trump Dataset Visualization

As seen above, there appears to be a negative relationship between Biden's ad spend and the undecided votes won, where a 1% increase in ad spend translates to 440 votes lost. However, this trend is flipped for the US West region (Alaska, California, Colorado, Hawaii, Idaho, Montana, Nevada, Oregon, Utah, Washington, Wyoming), where Biden gained 1007 votes for every 1% increase in ad spend.

These trends are reversed for Trump. As seen above, there appears to be a positive relationship between Trump's ad spend and votes won, a 1% increase in ad spend translates to 584 votes won. However, in the West region, Trump lost 1275 votes for every 1% increase in ad spend.

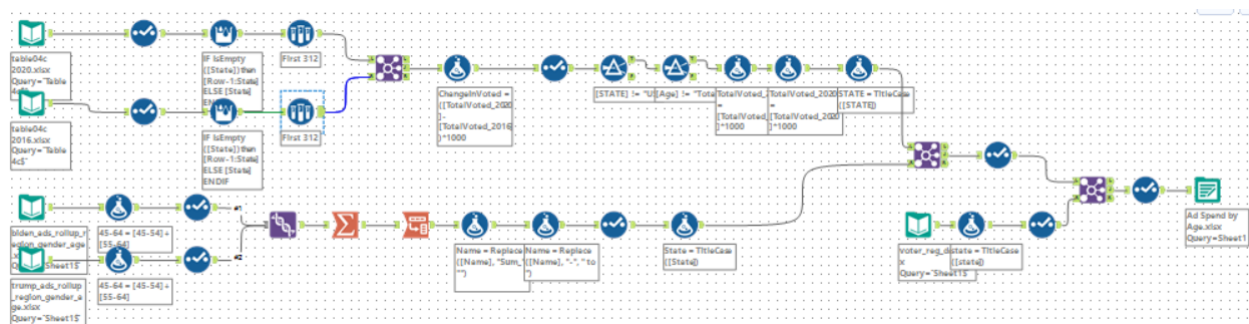
Overall, outside of the West region which seemed to reject Trump's ads in favor of Biden, Trump's ads seemed more effective at bringing out the vote for the 2020 election.

## Experiment 2: Ad Spend Influence on Age Groups

### Modeling

For this experiment, we continued with our approach of using linear regression. To answer this question, we utilized the Meta ad spend data well as census bureau data for voter registration for the 2016 and 2020 presidential election by age group. As expected with using multiple datasets, a lot of data cleaning was required to prepare for modeling. We first used python to summarize the ad spend data by region and age group. Next, we used Alteryx to combine columns, summarize data by states and regions, transpose data and format data, and

combine the two datasets. The workflow used for this along with the first few rows of the merged dataset is shown below.



Alteryx Workflow used to clean, format, and combine datasets

State	Region	Age	TotalVoted_2016	TotalVoted_2020	ChangeInVoted	Combined_Ad_Spend
Alabama	South	18 to 24	212000	194000	-18000	120613.2838
Alabama	South	25 to 34	310000	315000	5000	300670.1395
Alabama	South	35 to 44	306000	365000	59000	370706.8172
Alabama	South	45 to 64	761000	824000	63000	1097736.003
Alabama	South	65+	507000	549000	42000	587242.9301
Alaska	West	18 to 24	26000	29000	3000	19838.9226
Alaska	West	25 to 34	58000	41000	-17000	54444.02014

Combined Dataset

After creating the combined table, we fit a linear regression model using the change in votes from the 2016 election to the 2020 election as the dependent variable and the ad spending, age, and region as independent variables. The following code was used:

```
mod <- lm(ChangeInVoted~Combined_Ad_Spend+Age+Region, data=data)
```

In an attempt to better fit the data, we also attempted a log transformation of the combined ad spend variable which resulted in a lin-log model. We were unable to transform the dependent variable, change in voter registration, as it had negative values and we are not able to take log of negative values. The following code was used for the lin-log model:

```
data$log_Combined_Ad_Spend <- log(data$Combined_Ad_Spend)
mod2 <- lm(ChangeInVoted~log_Combined_Ad_Spend+Age+Region, data=data)
```

## Results

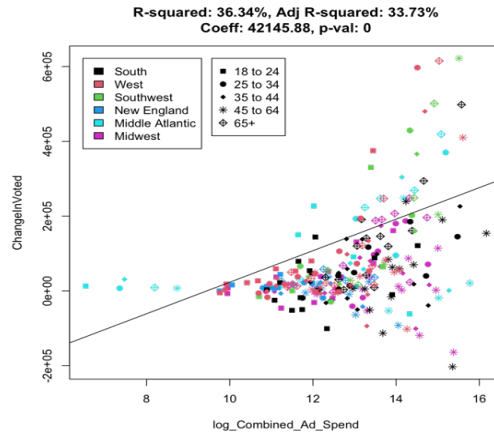
Our results indicate that the lin-log model is a better fit as indicated by the higher  $R^2$  and adjusted  $R^2$  values. For this model the following variables were deemed to be statistically significant: Combined Ad Spend, Ages 45-64, Region.Midwest, Region.New England, and Region.Southwest. The comparison of the lin-lin and lin-log models are shown below.

	lin-lin	lin-log
R-squared	35.09	36.34
Adj R-squared	32.43	33.73
Ad P-val	6e-15	5e-16
Ad Coeff	0.04	42145.88

Lin-Lin and Lin-Log Model Comparison

Based on the modeling output, we can say that ad spend does impact voter turnout. More specifically, the coefficient can be interpreted as a 1% increase in combined ad spend translates to a 420 increase in voter turnout from 2016 to 2020. The visual for this relationship is shown below.





Relationship between log\_Combined\_Ad\_Spend and ChangeInVoted

Contrary to the relationship between ad spend and change in voted relationship, for voters between ages 45-64, a 1% increase in combined ad spend translates to 464 decrease in voter turnout. This is contrary to what we hypothesized initially where we expected there to be a positive impact on voter turnout for older voters with more advertisements targeted at that demographic. Similarly, a 1% increase in combined ad spend for the Midwest, New England, and South regions translates to ~382 decrease in voter turnout based on modeling outputs.

## Experiment 3: Ad Spend Influence on Gender Groups

### Modeling

Like the previous experiment focusing on age groups, we also conducted a study on the effects of Facebook ad spending toward genders as well as regions. For this experiment, we utilized the Meta ad data to determine money spent overall by gender and region for the 2020 election as well as US Census data from both the 2016 and 2020 elections to determine voter turnout by gender and region.

The voter data from the US Census Bureau was already divided by gender but the relevant information had to be separated into its own table to be usable in R programming. Using the original .csv file downloaded from the US Census website, only relevant gender data was effectively extracted using Excel built-in functions to isolate the critical information. The result were two tables from both 2016 and 2020 elections that contained the pertinent information needed to conduct a regression analysis on gender and region.

State	region	male_registered_16	female_registered_16	male_voted_16	female_voted_16
ALABAMA	South	1155	1371	944	1151
ALASKA	West	181	177	152	156
ARIZONA	Southwest	1485	1660	1273	1496
ARKANSAS	South	679	777	574	667
CALIFORNIA	West	7663	8433	6833	7583

Example of Cleaned Gender Data, 2016

After the datasets were cleaned a linear regression analysis was performed in R using gender and region as categorical independent variables as well as total ad money spent for the election. The dependent variable, like the age group study, was the difference in voter turnout between the 2016 and 2020 elections. The following code was used:



```
gender.lm <-lm(Difference_2016_2020 ~ comb_ad_spend + Gender + region, data =
merged_voted_gender_2016_2020)
```

In order to instill confidence in our modeling approach and follow the same method as the age group study, we also attempted a log transformation of the combined ad spend independent variable as well as the difference in vote turnout dependent variable for linear-log, log linear, and log-log regression analyses. The transformed regression analyses did not provide improved model fit, however, and the linear regression model proved to be the most effective way to model the ad spend influence on gender groups data.

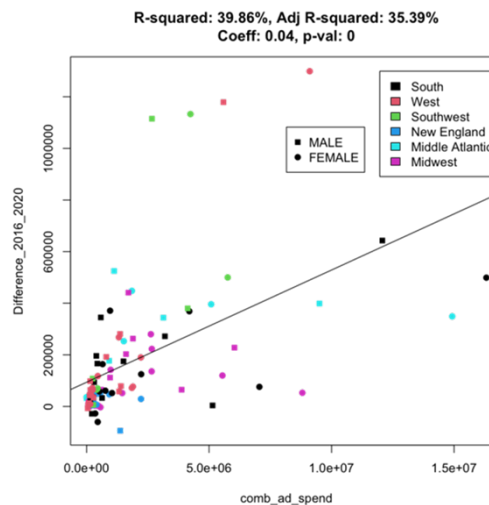
	lin-lin	lin-log	log-lin	log-log
R-squared	39.86	32.89	25.71	31.72
Adj R-squared	35.39	27.90	21.75	26.29
Ad P-val	1e-08	2e-06	0.00014	1.4e-05
Ad Coeff	0.04	47490	2.36e-07	0.34

Lin-Lin, Lin-Log, Log-Lin and Log-Log Model Comparison

## Results

As stated above, our results confirmed that the linear regression model was the most effective way to model the difference in voter turnout from the 2016 to 2020 elections by gender and region. We used the  $R^2$  and adjusted  $R^2$  value to draw this conclusion. For this model, the significant variables were: RegionSouthwest ( $2.103e+05$ ) and comb\_ad\_spend ( $4.348e-02$ ). From these results, we can conclude that ad money spent does indeed impact vote turnout, which supports our initial hypothesis. Gender does not necessarily explain the variation though. Ad money spent and Region Southwest can better explain the variation in vote turnout.

The model results can be summarized as: For every \$1 spent on Meta ads, there is a 0.04 increase in votes from 2016 to 2020. However, since the gender categorical variable is not significant, this means that although female voter turnout average is a large, positive number (93,920) and male voter turnout is even higher in comparison, this does not explain the variation in voter turnout. Additionally, since the RegionSouthwest variable is significant, we can conclude that states the southwestern region had a more significant impact on voter turnout than other regions. The visualization for these results is shown below:



Relationship between Comb\_Ad\_Spend and Difference\_2016\_2020

## Conclusion

Based on our analysis and experiments conducted above, we can say ad spending does have a positive impact on and correlates to a larger voter turnout answering our primary research question. Based on the first model, we can see that Trump's ads seemed more effective than Biden's ads at bringing out the vote for the 2020 election, apart from the West region which seemed to be more influenced by Biden's ads. Based on the second model, we can say that a 1% increase in ad spend leads to a 420 increase in voter turnout.

As far as regions, we can see that New England, Midwest and South had a decrease in voter turnout as ad spend increased but the Southwest region was positively impacted. These are the only regions we can speak to as far as impact as they were the only ones that were statistically significant.

When it comes to demographics, we couldn't glean as much in terms of insights as the variables were not statistically significant with and without transformations. When it came to age, the age group of 45-64 was the only demographic segment we can say was impacted by ad spend and they were impacted negatively as a 1% increase in combined ad spend translates to a 464 decrease in voter turnout from 2016 to 2020 for ages 45-64. For gender, the categorical variable was not significant, and this means that although the female voter turnout average is 93,920 and the male voter turnout is higher in comparison, this does not explain the variation in voter turnout.

## Business Application

We believe our research and modeling efforts can go beyond just analyzing the 2016 and 2020 presidential election with numerous applications and benefits. We were able to give a tangible quantification of the impact of ad spending on election results which also demonstrated the influence of digital marketing on overall public perception and views. From an application standpoint, political parties can use the region-based analysis to better strategize political campaigns in the future. Political donors can use our insights to determine how much to donate to drive a more impactful change in voter turnout. Another insight that we can clearly see is the influence of ad spending on society. Corporations could also leverage this to see which regions are more receiving of ads and target those markets for a better ROI on their advertisement spend. From Meta's standpoint, they can see the power they hold in shaping elections and could leverage this for further financial gains if they choose to do so.

## Future Research

Given more time and additional resources, there is plenty more research that we would have liked to do. We would have liked to bring in additional x variables to our datasets and models to get a more complete picture of what influenced voters. This could be datasets like TV ad spend, Twitter interactions, Facebook group interactions, or other social media activity. Also, performing similar analysis for additional elections could have provided insight on how voters behave for secondary elections like the senate, house, and governor. Doing this analysis in other countries would also have been interesting to see how Facebook's influence on elections differs abroad. Lastly, earlier this year Meta introduced new policies to limit how "sensitive" ads reach someone's feed. It would be interesting to see if this has any impact on future elections.

## Works Cited

Fujiwara, Thomas, et al. "Princeton University." *The Effect of Social Media on Elections: Evidence from the United States*, 16 Sept. 2022, <https://www.princeton.edu/~fujiwara/papers/SocialMediaAndElections.pdf>.

Haenschen, Katherine. "Haenschen Using FB Status Updates to Increase Voter Turnout." *Social Pressure on Social Media: Using Facebook Status Updates to Increase Voter Turnout*, Journal of Communication, 2020, <https://seriouspod.com/wp-content/uploads/2020/10/Haenschen-using-fb-status-updates-to-increase-voter-turnout.pdf>.

## Datasets

*Ad Library*. Facebook. Retrieved October 20, 2022, from [https://www.facebook.com/ads/library/?active\\_status=all&ad\\_type=political\\_and\\_issue\\_ads&country=US&media\\_type=all](https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=US&media_type=all)

MIT Election Lab. (2022, March 17). *MIT Election Data + Science Lab*. Retrieved October 29, 2022, from <https://electionlab.mit.edu/data>

*Party affiliation by state*. Pew Research Center. (2022, June 13). Retrieved October 20, 2022, from <https://www.pewresearch.org/religion/religious-landscape-study/compare/party-affiliation/by/state/>

Independent Voter Project. (n.d.). *Primary Elections State-by-State*. Primary Map. Retrieved October 29, 2022, from <https://independentvoterproject.org/map>

*Elections*. Arkansas Secretary of State. (n.d.). Retrieved October 29, 2022, from <https://www.sos.arkansas.gov/elections/research>

Colorado. (n.d.). *2020 voter registration statistics*. Retrieved October 29, 2022, from <https://www.sos.state.co.us/pubs/elections/VoterRegNumbers/2020VoterRegNumbers.html>

Ohio Secretary of State. (n.d.). *Secretary of State provides update on party affiliation data*. Retrieved October 29, 2022, from <https://www.ohiosos.gov/media-center/press-releases/2021/2021-10-01-a/>

Utah.gov. (n.d.). *Utah Voter Information*. Retrieved October 29, 2022, from <https://voteinfo.utah.gov/current-voter-registration-statistics/>

Wyoming Secretary of State. (n.d.). *Voter Registration Stats*. Retrieved October 29, 2022, from <https://sos.wyo.gov/Elections/VRStats.aspx>