

THE POLITICAL DATA INDUSTRY WITHIN MICROTARGETING

Ragan Carlson Jones

Political Science Department

Defended on March 18, 2020 in front of thesis advisor Nancy Billica (Political Science), Janet Donovan (Political Science), and Orly Hersh (Writing)

Chapter I. Introduction

Campaigns have consistently been tasked with appealing to voters to instigate turnout in their favor on election day. Politicians are faced with decisions about who to talk to, about what, and when to talk to them to reach the maximum amount of support possible. In the earlier days of campaigning, this might have required determining who to give a supportive lapel pin to, when it would be best to give it to them, and what to say when handing it over to them. In the digital age, the same questions are asked, but about sending an email. Who should I send it to? When should I send it? What should it say?

Specialized tactics used to answer these questions have evolved. Now, successful campaigns frequently employ the usage of data to help find insights on populations. They also will use data to precisely contact those they think optimal supportive messaging will encourage. Models of “core voters” versus “periphery voters” have been put to the test (Sigelman & Jewell 1986). Further, more advanced algorithms have been created and used to support campaign needs.

Data sources like statewide voter files and consumer data files have difficulties inherent with their usage and are a large undertaking for campaigns to handle themselves. As a result, specialized firms that compile voter data have found a niche market in which to operate. The political data industry (“the industry”) specializes in compilation and specialty interpretation of electorate data for clients. Client requested data and analysis is then handed over to campaigns and other political organizations who are tasked with creating the optimal communication with the targeted electorate.

The chapters in this paper fill in current gaps in research about the political data industry and apply appropriate analysis. Chapter II reviews the available literature and gives relevant definitions of microtargeting. Chapter III outlines the methods of analysis adopted in this research

to answer the appropriate questions at hand. Chapter IV reviews the findings from the interviews conducted and Chapter V demonstrates the processes completed by the industry and analysis firms through a novel case study of Ohio. Chapter VI discusses further connections of the findings in Chapter IV and Chapter V. Concluding Chapter VII recommends more research to be undertaken in order to better understand the chain of services that campaigns use to send messaging to the electorate.

Chapter II. Literature Review

Though the data industry has been rapidly expanding its role in U.S. campaigns and elections, there is little scholarly political science research examining either the methods or effects of this work. With little from political science to build on, there is a need to establish a basic foundation of knowledge that can help to describe and explain how political data is affecting contemporary American politics.

This literature review puts into context the importance of political data to campaigns. Specifically, it covers both the available information on the political data industry and questions raised by political theorists on the role of the voter in protecting and advancing democracy. This review covers topics of the data industry as they relate to competitive edge, discusses definitions of targeting from academics and the government alike, describes the two-step follow through of targeting, gives examples of common variables and forms of communication used by campaigns to target, and explains voter targeting's relationship to campaign money and consumer targeting. This review also considers the positive and negative implications of voter targeting. Currently, there is a lack of literature and research to describe the front end of campaign messaging, which involves voter data processing.

A. Democracy

Upholding democratic ideals is and has been at the root of American legislative and governmental norms since the establishment of the United States. Political equality and equal representation of voice, even at the lowest societal level, is demonstrated through the American electoral system setup. However, challenges to true equality within a population as large as 328 million are inevitable, especially with stratification due to racial, geographical, and economic differences (U.S. Department of Commerce 2018).

The right to vote is one of the most important tools that Americans are able to use to make their voices heard on an equal playing field. Each citizen has the right to vote in local, state, and national elections, with each vote given equal weight. But, interference in this process from outside forces has the ability to enhance systemic inequalities among voters (Sides & Karch 2008).

B. The push to mobilize

Contrary to the popularly held belief that incumbents are safer from losing an election than the first time they ran, it has been shown that their probability of losing remains the same each election cycle. They do not win elections any more easily because of incumbency, and therefore, campaign efforts have taken on all methods available to secure election success (Jacobson 1987).

There are incentives for campaigns to alter spending strategically based on impacts like geographical competitiveness and party money transfers to states. Additionally, as a first step, mobilizing candidates' core supporters is the most rational because there is no need for persuasion, but instead just mobilization. Allocating resources most efficiently would direct campaigns to spend money on mobilization of the already persuaded rather than requiring both persuasion and mobilization (Holbrook & McClurg 2005).

C. Definitions

The next step is for campaigns to find the voters they wish to mobilize and successfully reach out to them. Sending a tailored message with the intent to have a specific group "evaluate favorably," is called targeting (Holman, Schneider & Pondel 2015). Microtargeting is the process of "tailoring messages to specific groups of voters" ("Of mud and money" 2012).

This type of political targeting has been broken down into two steps, where groups must first be identified, and then the most effective mobilization tactics must then be identified (Rusch, Lee, Hornik, Jank & Zeileis 2013). More specifically, identity-based targeting is defined as using

“politicized social group” attachments to appeal to prospective voters through targeting (Holman, Schneider & Pondel 2015). The federal government has recently expressed some interest in voter targeting through the introduction of the Voter Privacy Act of 2019, which defines targeting as “target[ing] communications to an individual based on that individual’s personal information” (“Voter Privacy Act of 2019” 2019).

Some voters have made part of the process easier for campaigns by self-selecting into groups through rallies and attendance at candidate appearances (Holbrook & McClurg 2005). Similarly, the population has started to self-sort on social media platforms, and campaigns have figured out how to use interest self-sorting to their advantage when targeting (Stier, Bleier, Lietz & Strohmaier 2018). These processes encourage more self-grouping of voters into segments.

D. Targeting

Another step to sorting and targeting is to appeal to voters who have not opted to self-select into groups. In this process, voter files and vote history data are extracted and used to fill in gaps of information where there is no self-selection. This data can come in the form of past vote history data, voter files that include information about the voter, and calculated voter scores. All of this data can be combined further with inexpensively purchased consumer data files (Nickerson & Rogers 2014).

Data on an individual voter is compiled from publicly available sources such as state voter files, precinct election day data, and additional consumer files (Rusch, Lee, Hornik, Jank & Zeileis 2013). This data is transformed into a usable format which consists primarily of voter scores. These scores can reflect overall likeliness to vote, and behavioral, support, or responsiveness scores. However, creating these “predictive scores” was, and still is, a niche business (Nickerson & Rogers 2014).

Typically, a campaign will look at a voter's likeliness to support the candidate in conjunction with their likeliness to turn out at the polls. A quadrant system can then be used to break up outreach into targeted messages and general mobilization. For example, if a voter is likely to both vote and support a candidate, campaigns will generally look to mobilize. However, if a voter is predicted to be less likely in one of those two criteria, targeted messages based on their personal data as discussed above will be used to send a message with the intent to mobilize or persuade. Finally, those that are both unlikely to support and turnout to vote are typically ignored by campaign efforts (Rusch, Lee, Hornik, Jank & Zeileis 2013).

Analysts are also able to use attributes of an aggregated geographical group's voting patterns to make inferences on how others within the group would likely turn out to vote as well. Grouping ecologically has shown "to increase predictive score accuracy" (Nickerson & Rogers 2014).

E. Measuring two parts of targeting

Success of targeting efforts is notably difficult to measure because electoral results are not actually able to be measured until after election day, unless proxy data like nightly polls are used as an estimate (Rusch, Lee, Hornik, Jank & Zeileis 2013). It has been broadly concluded that experienced data analysts are exponentially more valuable to a campaign than less-capable data analysts. Less experienced data analysts might not take care to control for additional constructed variables based on current issues or candidate specifics. Similarly, campaign message developers' roles have proven to be of increasing importance given that targeting has had the ability to become more pinpointed with access to more data (Nickerson & Rogers 2014).

Further, successful pinpointing of populations to persuade is not the only factor that dictates turnout. The second part of targeting is finding the optimal method to convince and

encourage a voter. Priming and persuading have been labeled as two types of message targeting. Priming is “directing attention of the public to evaluate the candidate on one specific issue over others,” and persuading is “defined as directly changing voters’ evaluation of the candidate” (Holman, Schneider & Pondel 2015).

The only way to send the correct messages to certain targeted groups is to keep traditional roles of campaigns, including phone conversations and grassroots successes, central to campaign tactics. The campaigns now develop messages while they collect voter data through polls and phone interviews (Nickerson & Rogers 2014).

F. Variable: Gender

Gender targeting is a form of identity-based targeting. This is important because women have traditionally supported the Democratic party more frequently than men in elections. Women are not dispersed evenly across the country and have higher voter turnout numbers than men. The power of descriptive representation provides female candidates with a separate appealing tactic to female voters. However, male candidates have used identity appeals and female “surrogates” to send messages as well.

Overall, persuading through identity-based political appeal proved more effective than issue-based messaging in the studies of female voters. Also, a female candidate’s effectiveness of appeal did not prove statistically significantly different from a male candidate or a female surrogate on behalf of a male candidate’s effectiveness of messaging to the female electorate. Finally, a female candidate’s messaging will “prime group-based social identity,” creating a significant vote increase as a product (Holman, Schneider & Pondel 2015).

However, it is important for candidates to ensure that they are not only appealing to women through women's issues. Usage of data needs to be used in the context of appealing to free thinking individuals and not strictly automating every message and interaction (Brooks 2014).

G. Variable: Issue-based targeting

Despite the agenda setting power of campaigns, their success is also dependent on how they match up with the existing policy agenda created by the public (Sides & Karch 2008). When targeting by issue, candidates are "priming" audiences through influencing voters to evaluate a candidate based on one issue (Holman, Schneider & Pondel 2015).

However, there is little evidence that shows that issue targeting based off of deemed interests is successful in mobilizing voters to the polls. For example, when looking at whether veterans showed increased mobilization when targeted messages pertaining to veterans' issues were sent to them, there was no statistically significant increase in veteran turnout as compared to before.

But, when a less "participatory" group, like parents of children under the age of 18, is mobilized on their respective issue, like education, through high "volume" targeting, there is a slightly more significant result in mobilization as a result. Substantively, the difference is small, but nonetheless, certain issue-based targeting might be able to act as an equalizer for turnout (Sides & Karch 2008). But all targeting does not necessarily focus on single issues that are expected to be associated with a group. The lack of correlation demonstrates the importance of layering identity, issues, and other factors to reach populations effectively through targeting.

H. Social media

Social media targeting has also been studied, but its effects are especially difficult to pinpoint because many social media accounts are not directly associated with an identifiable voter.

However, new technology, like social media, has created a forum that is available for all to use for free. This has evened the playing field for challengers who do not have access to travel allowances, mail, and other incumbent advantages. Technology usage is not contingent on holding office. Resources such as skills and technology are available to opponents and incumbents equally (Jacobson 1987).

As mentioned before, social media is helpful to campaigns because users often have the ability to self-select into certain groups. More broadly, the fact that a voter is on certain social media allows for campaigns to make certain assumptions and group them. Politicians use this to “discuss different topics on social media than those salient among a mass audience” (Stier, Bleier, Lietz & Strohmaier 2018). Campaign events are more frequently discussed on social media than election issues and policy plans, which is consistent with the notion that the internet is used as a more impersonal front.

Further, campaigns use different social media platforms to achieve different campaign goals. It is difficult to group all social media usage for campaign purposes into one umbrella. For example, it has been observed that Facebook is used for general mobilization efforts, whereas Twitter is used for more direct communication. Additionally, topics on Facebook are more disconnected from the interests labeled by survey respondents than issues on Twitter. Social media content still intersects with the interests of the public based on survey data, but also acts as an agenda setting platform like traditional media (Stier, Bleier, Lietz & Strohmaier 2018).

Different from using social media as a discussion platform, social media has also started to demonstrate itself as a platform that can house online political ad buys (“Of mud and money” 2012). These online political ad buys have “the greatest potential for microtargeting” simply

because you can direct ads to those searching for campaign information. These people are already indicating that they are interested in the candidate or the topic at hand (“Of mud and money” 2012).

However, the difficulty of actually reaching the intended viewer is becoming increasingly difficult. Viewers are now able to skip through or choose the ads they view online and on television with the rise of digital recorders (Keller 2009). As a result, YouTube is now offering the option for campaigns to opt in and only pay for advertisements that are actually viewed (“Of mud and money” 2012).

I. Television

Similar to social media, campaigns can also set agendas through television ads. However, television used for campaign purposes has had content restrictions placed on it.

Television traditionally has reached broader audiences given that most ad buys are through broadcast television that is based on region (Holbrook & McClurg 2005). Broadcast television is cheaper than cable and can be used as a “nuclear weapon” due to its traditionally lower price point than cable television (“Of mud and money” 2012). But, increased demand during election years and new promises of greater ability to reach intended viewers have increased prices for these ad buys (“Of mud and money” 2012). Additionally, traditional regional broadcast television can cause tensions between interests, given that regions are complex within themselves (Nickerson & Rogers 2014).

But television is frequently passed off as being strictly broadcast based. There are also cable networks that can be used to direct interest-specific targeting. They market themselves as being a precisely aimed option. This can mitigate potential adverse effects by weaving around local issue complexity (Nickerson & Rogers 2014). Microtargeting can take place through

television by placing ads on shows that would only appeal to certain subgroups of the population (“Of mud and money” 2012).

J. Comparison with consumer targeting

Voter targeting has been compared to studies done on consumer targeting. Traditionally, brands through mail vendors were on the “cutting edge” of using data as a targeting tool and were “at least a decade ahead of the political campaign learning curve” (Nickerson & Rogers 2014). Given that consumer targeting preceded voter targeting through data, consumer marketers have clear and concise tactics that have had longer to be tested and accepted. Brands understand that consumer perception and knowledge dictate successes. Knowledge and perception are based on brand loyalty, the quality/functionality of a product, and social perceptions within their social circle (Keller 2009).

As campaigns work to increase precision in the types of voter they attract, they must “become increasingly reliant on consumer-marketing techniques” (Holman, Schneider & Pondel 2015). Messages sent by brands also should have versatility, defined as those messages that will reach and affect customers in the intended way regardless of their group attachments. Using differentiated marketing techniques helps establish the perception of the brand that is desired. Further, interactive techniques of communication allow for versatility of messages (Keller 2009).

Candidates have implemented some of these techniques to cultivate perceptions of themselves. They do so by creating districts based on persona and campaign by emphasizing their personal qualities over party values. Given decreased partisan connections of the electorate, these “personal political franchises” are a good way to ensure safety for an incumbent (Jacobson 1987).

In this way, candidate-centered districts can be analogous to cultivating brand loyalty. In order to cultivate and target the most effectively, campaigns must also allocate their resources

optimally in the same way that brands do. Campaigns have limited financial resources, so they use targeted contact lists to pinpoint who they should direct outreach towards. Campaigns then also use results from these targeted outreaches to make broader decisions down the line (Nickerson & Rogers 2014).

If candidates are to use brand cultivation tactics, they must also direct focus to understand what knowledge and perception cultivation needs to happen at the most effective stages. This also would help “allocate communications dollars more efficiently” (Keller 2009). The optimal allocation is measured in the context of brand cultivation through effectiveness and an achieved increase in sales (Keller 2009). Or, put in terms of campaigns, voter knowledge of the candidate through campaign messaging is an electoral success.

K. Campaign money and resource acquisition

Individuals have the ability to donate to campaigns to increase their reserve of usable resources. As explained through the quadrant system, the voters that are determined likely to vote and support the candidate are solicited for volunteer time and donations (Rusch, Lee, Hornik, Jank & Zeileis 2013). Past donors will also be contacted again for additional donations. Reconnecting with past donors and volunteers has been a tactic that campaigns have used since well before the introduction of personal data. But with data availability, the process of reconnection is now streamlined because of easily accessible contact information and probabilistic donor scores based on past donation information (Nickerson & Rogers 2014).

Like campaigns, national political parties can also spend money through transferring money down to state and local levels. Analysis done on the effects of party transfers indicates that spending party money increases voter behavior and electoral outcomes but does not necessarily change the composition of the electorate (Holbrook & McClurg 2005). In the 1950s, incumbents

and challengers alike relied on party messaging and partisanship of the electorate to help them maintain or reach their position in office. But since the 1970s, there has been little evidence of this kind of “partisan tide” aiding those reaching or maintaining office (Jacobson 1987).

L. Negative implications of targeting

The implications of using voter information to target are dependent on whether campaign actions actually have a systemic effect on voters. Studies confirm that there is correlation between what campaigns message and how voters act. Campaigns do matter, but the context under which they matter is still being debated (Sides & Karch 2008).

However, to the extent that citizens are being affected, there is also the possibility of systematic disenfranchisement that is dependent on the groups of the populations that are repeatedly targeted or left out of the targeting. Some subsection of the population is all together unreachable because they are not on registration lists that are widely used by campaigns (Jackman & Spahn 2018). Moreover, the lists are diverse and not always consistent (McDonald 2007).

Trends within the lists indicate that minorities, the poor, and those with liberal ideology are underrepresented. Additionally, those on the list, either registered or unregistered, report more contact than those that are unlisted. Those that are young, low income, and with a residence that is not stable are notably difficult to reach because they are frequently unlisted. Voters that are “mislisted” and do not have accurate contact information on file at compiling firms are difficult to contact as well (Jackman & Spahn 2018).

Visibility in the American political system is dependent on perceived or actual difficulty of registration, which results in underrepresentation of the poor. Further, knowledge of a voter’s consumer preferences increases visibility in the system past registration through luxury items like

credit cards, stable housing, and usage of bank accounts. Overall, being socially and economically disadvantaged correlates with political invisibility (Jackman & Spahn 2018).

Further, it is much easier for campaigns to collect information on those who are likely to support the campaign than those who are more likely to support the opposition. Voter files are publicly available for those that have already voted, but not for those that have not registered or voted. This disparity of data information has the potential to emphasize inequalities between those that are already politically inclined and those that are not (Nickerson & Rogers 2014).

Identity-based targeting like gender targeting can also have implications on descriptive representation. For example, male candidates are able to “claim solidarity” with female groups through liaisons and have relatively similar effects on voter mindset as female candidates do with identity-based targeting (Holman, Schneider & Pondel 2015). This creates an imbalance in descriptive representation preferences for those that might not be as issue aware as they need to be to look past these appeals. Also, it has been shown that there is no disincentive for male candidates to use these kinds of appeals of solidarity to women because there is no backlash from male voters (Holman, Schneider & Pondel 2015).

M. Positive implications of targeting

These normative implications can also work in the opposite way as well for an uninformed female population. Claiming solidarity with the female population as a female candidate may incentivize voters to overlook other issues that the candidate has taken a stance on. This kind of issue targeting has also been discussed as having an equalizing effect for populations that do not express their voice enough. Traditionally high turnout voters like veterans and seniors will continue to turn out to vote regardless of campaign messaging, whereas lower turnout voters will be encouraged to turn out when they would not have otherwise (Sides & Karch 2008).

Discussion about turnout being negatively affected when there are negative ads run has been met with research that suggests that any campaign outreach promotes turnout. Turnout is increased by campaign advertisements because they create interest and make information more easily accessible. Partisans are the most activated as a result of campaign messaging. Importantly, campaign advertisements also have the ability to change the composition of the electorate from presidential election to election (Holbrook & McClurg 2005).

Also, party money transfers have significant positive effects on election outcomes through changing the composition of the electorate as well as possibly mobilizing certain independents. The findings suggest that party money should not be simply restricted; instead, attention should be turned to understanding how these campaigns actually allocate their funds (Holbrook & McClurg 2005). About four in ten Americans classify themselves as independents, but only about seven percent do not claim to lean one way or the other (Pew Research Center 2019). Given that so many Americans are leaning one way or the other, “latent” party and election interests of independents might be activated through campaign messaging (Holbrook & McClurg 2005).

In the United States, this kind of voter data is all publicly available in aggregated forms from states. The U.S. is on the front lines of transparency in this sense. In many other countries, this type of data is only available from market research companies. American campaigns have access to similar information and mediums to express messages created by different analysis tactics (Rusch, Lee, Hornik, Jank & Zeileis 2013).

However, this kind of power through data only affects those campaigns that were already effective in their outreach and grassroots capabilities (Nickerson & Rogers 2014). The campaigns that do not have the creative skillset needed to balance messaging tactics with policy and demographic blocks with individual thinking will not be successful despite technological

superiority. Overuse of data disregards concepts like free-thinking that are central to elections and the American political system (Brooks 2014).

Nonetheless, using predictive scores derived from voter data has shown the highest net benefit for more resource constrained campaigns. Having access to better information about audiences can encourage more pleasant campaign experiences, leading to a potential rise in desire to become politically involved (Nickerson & Rogers 2014).

N. Conclusions

Given the findings and implications regarding campaign data use, the future of the voter data processing industry itself should be an area of interest. The government has begun to express an interest in regulating voter data through the Voter Privacy Act of 2019, and it is uncertain as to where these regulations will end. Campaigns use data, and the usage and potential effects of usage of data to shape messaging have been and are being studied.

However, there remains a lack of information on the firms involved in the data collection. The political data industry is responsible for collecting data on voters, compiling into a usable format, and suggesting usable targeting setups for campaigns (Jackman & Spahn 2018). Similar collection and compilation firms have a longstanding place in consumer targeting (Keller 2009). The line between personal informational privacy of voter data and ensuring the security of elections has yet to be clearly defined by the government, and thus political data firms' practices should be studied to help create a clearer picture of the industry.

Chapter III. Methods

The data industry's processes are important to understand in relation to the lifecycle of data and its usage by campaigns and message developers. In particular, microtargeting is a specialized tool used by campaigns among other services to improve messaging tactics that may have potential implications on democratic ideals. This research examines the industry through both interviews and case analysis.

The first portion of the research consists of original interviews with professionals that operate within the political data industry. Interviewees were chosen based on voluntary participation in response to an email blast that was distributed throughout a firm. These professionals' backgrounds range in seniority level and responsibilities. All quotes and ideas in the data section below (A) were stated by an interviewee in the process.

Interviewing was done over the phone without recording. All interviewees were asked a list of predetermined questions and notes on their answers were recorded. The predetermined questions can be found in appendix section (I). After this, the notes were transcribed into digital format minus all identifying information.

The analysis was carried out inductively, primarily due to the lack of research on this specific industry that is in existence at this point. The research was guided by the findings and questions raised in prior research. However, the data and findings were approached without any set goals or assumptions, allowing the interviewees themselves to identify key insights. The data was sorted based on interviewees and topics that were prevalent. The sorted responses can be found in appendix section (II). From this, patterns were recognized and used to make conclusions about commonality of certain ideals and responses. If there were clear disparities, those were also indicated.

The second aspect of my research is a case study of the behavior of Ohio voters. Ohio is a sought-after state by candidates running for presidential office due to the state's large number of electoral college votes. The state's electorate is also known to be a group that largely consists of "swing voters," or voters whose behavior is not certain from election year to election year ("Ohio Presidential Election Voting History"). The methods used for my case study are used to give an example of voter isolation using data.

The goal of this research is to fill current gaps in knowledge about the political data industry. The two methods of inquiry used as the basis of this analysis—interviews with industry experts and case analysis demonstrating how data is used to target voters—are designed to provide a clear and specific picture of what the political data industry actually does and to consider the implications of these practices for democracy.

Chapter IV. Original Research Interviews

Ten professional data analysts were interviewed to gain insight on the nature of the industry and their personal perspectives on how this work affects campaigns and elections. Interviewees were asked about the details of their actual work (what they actually do); exclusivity of entry into this specialized data industry; whether or not they saw their work as a form of advanced campaign work; and their views on the safety of the data, including the need for or likelihood of future government regulation.

A. Details of job

The interviews provide insight on the technicality of the work on which political data industry professionals compete. Compiled voter files are the primary product for which the political data industry is responsible. Professionals in technical roles within the industry make requests to Secretaries of State or county clerks for public information, depending on the group of voters on which they are seeking information.

The specificity of data collected can range from state level to precinct level, with each state and level releasing different public information on voters. For example, the firm will build a state's voter file using historical election results by precinct, but in the form of vote totals for president. They will collect vote totals typically from the federal level down to the state assembly level. These state builds are updated whenever state information is updated. There are also roles that are delegated responsibilities to send out firm updates when pertinent updates go on in the voter file.

After the data is acquired in its respective format, depending on the state it is being received from, it is then ingested into the database. Most of the data is in different formats and their job is to refine it into a usable format. Attention to detail is of importance in this step, due to the different formats and flaws each state's data might have. One professional used the example of a state

leaving in voters who have cancelled their registration or moved. The value added by the industry directly refers to their ability to recognize these types of mistakes and make the data outwardly usable for clients. Ingesting the data into a uniform company format allows for ease of product usage and retrieval of specific data on voters for clients.

Once the information is uploaded, professionals are then able to fill client requests. During this process the data gets pushed from the internal side of the company to the external facing side that handles client requests. Client requests can be visualizations in the form of maps or graphics to help explain composition of the electorate. They also might be in the form of lists for polling, lists of raw voter file data, or lists of emails of specific voters. Occasionally, the firm will also need to put in additional specific requests to legally collect other data for a client.

There is a subsection of professionals in the industry that specialize in outward facing client relationships. The purpose of those roles is to typically predict what the current and future client's biggest needs are regarding data usage. Client requests can be for raw voter data, access to consumer data, and data visualizations, among other products. Though, they are typically specific to certain regions and characteristics of voters. These requests are then transmitted to the client in the format of a one-time email or online access link, depending on whether the request is ongoingly updated information.

In order to properly choose and manage clients, there are also professionals that complete research on potential clients. Political clients could be data analytics groups, advocacy groups, digital agencies attached to campaigns, trade associations, a national party, and polling companies. These are all organizations that are attached to campaigns and work to provide specific advertisement buying and campaign messaging. All clients are not necessarily campaigns, but they are "engaged in the political process." Many clients are "politically skewed," otherwise they would

not have a significant interest in the product in the first place. Clients also range in budgets, and the interviewees noted that their ability to cater to clients of differing budgets through lower cost research gives lower-budget campaigns a chance in the political sphere.

The relationship that data compilation firms have with these clients sets them apart from other analysis firms due to their versatility. “Other firms tend to specialize a little more,” but are still just “value adds” to the data. One professional stated that a compilation firm is a “central hub” for comprehensive data. Another professional used the analogy of a library in describing their role in the process. They have the ability to distribute large amounts of thorough voter file information to others for more specialized usage.

There are also troubleshooting roles that some employees are required to fill due to the technical nature of both the compilation and the product that is being dispensed. Some roles handle assigning credentials to give proper access based on the specifics of the agreement with the client. Additionally, many parts of the dispensable data are offline, and some firms are looking to take it online. This would allow clients to access it in different formats with credentials that are personalized to their appropriate needs. Client requests sometimes require more visualizations that make the data more readable and understandable. Overall, the initial data compilation is the “product that is then dispensable” to other parts of the company and outwardly to clients in different formats.

B. Exclusivity of Entry into the Industry

Professionals in the political data industry were clear that the state voter data collected is legally obtainable public information. Although it is public information, one professional said, “it depends on what portion of the public you are in” given that “some of the data that we process is

only available for request by a registered voter in that state.” Nonetheless, in all fifty states a private person can request a voter file.

This discrepancy in the definition of “public” created an interesting discussion of restrictions. One professional pointed out that taking the definition of public as “anyone can go online and download it” means that only two or three states make voter information publicly available. For example, Michigan requires a Freedom of Information Act (FOIA) request to be filed, and in Pennsylvania, you have to be registered to vote in that state. Further, if public means that the data is also free, then data is public in only about fifty percent of states. Industry firms can receive data made available to other organizations under licensing agreements.

Additionally, there is a concept of proprietary data and “restrictions around usage” that must be adhered to. These restrictions are all indicated and enforced by the state of origin of the data. It was stressed by one professional that the industry is “not operating outside of any Secretary of State mandate.” Typically, these mandates outline usage to be restricted to “all political reasons.” For example, Verizon cannot use this voter data for any business development.

Professionals also felt it was important to confirm that the data never includes information on minors or other information that is secure, such as social security numbers. Further, the data is only inclusive of certain political identifiers like party affiliation if the originating state mandates that that information be released.

However, there is also another source of data they use in their work and are able to provide to others. They purchase some consumer data from large consumer data firms like InfoGroup and Axiom. This allows some gaps in information given by state voter data to be filled in. But, the consumer data itself only gives probability ranges for all categories. For example, it may give a percentage on the likelihood for a certain purchasing habit to occur in a specific individual.

This data is relatively expensive to purchase, and one professional mentioned that they think InfoGroup imposes fairly strict vetting procedures and security measures on its licensees. But another professional made the point that lower-budget campaigns can utilize portions of the consumer data files as well. If the campaign contracts with an industry firm to use only the portions of the consumer data files that are applicable, it is nonetheless less expensive than the price of accessing the entire file.

Another way that the industry is able to supplement state voter file data is through campaign collected data. For example, this data can come from information that event attendees have provided to a campaign while attending an event. The industry firms are then responsible for the privacy of the data collected under these terms. This is not always public information.

C. Is this a form of advanced campaigning?

One professional in this industry views this work as “not a whole new thing in itself.” Data analysis and usage of past results to determine future plans translates to efficiency. One professional stated, “adding a component of data analysis to [campaigning] will make campaigning of increasing efficiency and effectiveness.” Data usage increases the ability for campaigns to direct their communication efforts to the most receptive people. Campaigns have always maintained lists of optimal people they should be interacting with, and this analysis confirms their assumptions or redirects them. Another professional made a comment about the level of campaigns’ tenacity to find the voters that will show up for them.

The opposite scenario of blanket marketing a candidate was brought up by many professionals. One asked, “am I expected to call every voter regardless?” Another professional brought up the inefficiency associated with sending messages to people who have been already predetermined that they just will never bite. For example, there is a reason why Clinton did not

make significant efforts in Alabama. A republican candidate being forced to talk to all registered democrats too is inefficient. Data usage is “just taking this distinction to an advanced level.”

Further, if this kind of campaigning were to be removed, it would be “shutting out lower-dollar campaigns and fundraisers.” If targeted advertising were to be removed it would cost more for all candidates to participate and thus create a higher threshold for someone to have a chance at winning an office. A professional also stated that it would make more political noise and potentially make the public more averse to political participation.

Another interviewee stated that money is still the center of election results. Further, a different professional stated that nothing has changed due to data collection and “the candidate with more money can allocate more resources to make the campaign more effective.” This resource strain on democratic ideals was not a “spigot that was turned on because of data usage.”

When asked if data usage had the ability to “skew” election results, most professionals came back with the argument that this is the point of any sort of campaign effort. Anything that relates to campaigns and political organizations’ execution of political actions will skew electoral results. One professional stated that for the clients that use political data, reaching desired outcomes is the goal.

Data usage by campaigns is obviously political in nature. One question to interviewees was about informational advantages and whether disproportionately collecting and spreading information was significant. The professionals highlighted the fact that campaigns will often work to contact people regardless of likelihood of voting primarily because campaigning is becoming extremely competitive. Additionally, one pointed out that because of competitiveness of elections, everyone is going to hear from both sides of the spectrum. It is more likely for the electorate to hear from both sides because of the speed of distribution of information over media. Further,

another professional stated, “I don’t think people are learning a lot of new things from targeted advertising.” But the data does provide new insights for researchers to conduct further survey work.

In response to follow up questions about those voters that do not show up in many data systems, one professional responded that campaigns will always find people who are likely to support them, regardless of data. When questioned about the worry of political invisibility and a perceived advantage to those voters who are visible in data systems, other professionals responded with a “free will argument.” They noted that it is insulting to someone’s free will to make this argument, which starts with speculating about why someone had or had not voted. They pointed out that voting is ultimately the voters’ decision and that inducing turnout of any population is part of democratic ideals.

Professionals also connected having increased information on the electorate to increased effectiveness on “get out the vote” efforts. This created a basis for discussion about the democratic ideals that they perceive this type of targeting is furthering. One interviewee commented that when you push someone to vote that usually doesn’t, you are “helping get people to the polls.” The ability to deliver effective political messaging improves turnout and democratic ideals. This type of targeting and advertising “helps connect people to democracy and to the people they support.” Another professional made a connection between very high turnout in the 2018 midterm elections and the increase in data usage. Data is not dissuading people from turning out.

Many of the beliefs about freedom to campaign and use data to induce a certain outcome mimics goals in using consumer data. One professional pointed to the fact that marketing departments at big corporations are doing the exact same thing political data firms are doing when they “leverage statistics about individuals” and “deliver [messages] in the most tailored

advertisements they can.” Another professional stated, “In both cases you’re trying to find a niche of people that have certain qualities and traits.” The end goal is to speak with a specific group of people about a specific thing, and using historical information is just a way to create an audience. A different interviewee made the connection that if campaigns are not getting the results they want, they should step up their data efforts. It is fair game if one campaign knows how to use the data better. This is analogous to Apple doing a better job than Samsung selling a phone to a consumer.

Political data firms have a goal of merging the two processes of targeting and are working towards it by integrating the voter data lists with consumer files for easier usage. The consumer data is not political in any way. Additionally, the final product that is being “pushed” to the end audience is different. One interviewee said that political data clients are trying to “convey a message” while product marketers are trying to push “a repurchase of granola bars.” Another interviewee made the statement that the political data they use is not anything worse than what consumer data companies collect. Having the two data sources integrated as one is useful to understanding populations.

D. Safety of the data and the future of government regulation

Interviewees stressed that it was a common misunderstanding that the data even needs security for privacy reasons. Rather, security is in place to ensure the protection of the work of the company. Additionally, security prevents unlawful misuse of information outside of Secretary of State mandates.

Security is something that is taken seriously. Protections that were commonly mentioned were two-factor authentication for all users and virtual private networks needed for access. The interviewees referenced contracted technology companies that help make sure these and other efforts are being properly implemented.

To protect data after it has been transmitted to clients, firms implement questionnaires regarding client hardware security and send exploding links that expire after a certain time period. They have also started adding client codes into the data, so they are able to tell who leaked it if there were to be a breach. This allows them to maintain “power and say over what our data is doing” and track the “chain of custody.” These are matters that protect a firm’s reputation for security. After the hacking in the 2016 election, professionals said they “definitely need to keep it as secure as possible to prevent that from happening in the future.”

The need for this data to be kept secure has led to the question of whether the government has an interest in protecting it by regulating it. Many professionals stated that they cannot foresee the industry ever being regulated to the point of disappearance. One professional stated that “I don’t think there is a lot of government regulation that could impact us, that could get rid of data information as a whole.” A few interviewees welcome regulation because with regulation there are often more efficient adaptations and innovations that come from it. Despite potential regulation, there will always be a demand for constant information. As one professional observed, the “desire to know things will always exist,” especially who to talk to and how to talk to them.

There is also skepticism about the likelihood of meaningful regulation actually being successful in limiting data usage. An interviewee pointed out one of the reasons by referring to an April 2018 Facebook congressional hearing. When Facebook founder Mark Zuckerberg appeared in a joint hearing before the U.S. Senate Judiciary & Commerce and Science & Transportation Committees on Facebook’s privacy policy and use of data, it became apparent that lawmakers were confused about what Facebook does and what problem needs to be solved. It was noted that the discussion in that hearing showed how little members of Congress appear to understand basic technology. Further, it is unlikely that anyone in Congress will propose significant changes to the

structure of data usage because the people in office are the primary beneficiaries from such technology. To one of the interviewees, the biggest indicator of lack of future government regulation is the lack of current government regulation in the consumer data sphere.

However, another professional did make the point that many consumers are unaware that they are being tracked, specifically in the context of Facebook. Further, a different interviewee indicated that government regulation might take the form of either opting in or out of being listed. Nonetheless, interviewees stressed that there actually is nothing to be afraid of regarding voter data collection. An industry interviewee recognized that there is a “lack of understanding on what they know about you.” They only have information about you that you have willingly given to them. But it also was noted that many people are not aware of voter data collection generally, so “voter report cards that are sent to your house can be kind of startling if you aren’t expecting it.”

Chapter V. Case study of Ohio

Examination of the State of Ohio's voter data system is instructive on how the political data industry can access and use campaign-relevant information. Ohio's Secretary of State publishes up-to-date voter files on a State of Ohio website in downloadable format, making it accessible to all with no payment. The ease of accessing information about Ohio voters is high relative to many other states, which may require a request form, subscription, or payment.

Ohio's primaries are semi-closed, where a registered voter declares their party affiliation on primary election day by choosing the ballot they would like to use to vote. There also is an option for unaffiliated voting, but these ballots do not include national party candidates ("Register to Vote and Update Your Registration").

By analyzing some of Ohio state voters' behavior based on the state's voter file, the usage of these voter files by analytical firms will be demonstrated. Two commonly used variables of geographical location and party affiliation will be isolated and used to group voters. The ability to isolate voters will indicate effectiveness of segmenting the population when interacting on a campaign level. The STATA do-file images clarify cleaning steps and can be found in appendix section (III).

A. Model 1: Primary voters' turnout in the respective general election

Introduction: John Kasich was the former Governor of Ohio, before he began to run for the office of President of the United States. He was the presidential candidate who won the Ohio Republican primary in the election on March 15, 2016. However, he was not the final national Republican candidate in the presidential election held on November 8, 2016. Donald Trump was the final Republican candidate on the ballot in that election and was listed next to the Democratic

Party's candidate, Hillary Clinton. In this first model, the goal is to isolate the population of voters that participated in the primary election but did not turnout in the general election.

The first step in cleaning the data was to drop all voters whose voter status was not confirmed. Then, the data collected about each voter's participation in the primary election was coded to nominal variables. The Ohio voter file provides information that someone voted in the primary and which ballot they took when they voted. Each party's ballot in the case study data is a different value depending on which ballot the voter took. Next, the general election dummy variable was coded to hold a value of 1 if the voter voted in the general election and 0 if they did not.

To isolate the population of interest, a new dummy variable called *nogenvote2016* was generated that identifies whether a person voted in the primary election but not in the corresponding presidential general election in 2016 by labeling them with a 1. Everyone else was labeled with a 0. To indicate political interest in the 2016 presidential election, any voter who did not vote in the 2016 general election or the corresponding primary was dropped.

Data: After cleaning the data, there were a total of 3,093,371 voter observations left who showed interest in the 2016 presidential election by casting a vote. Out of those that showed interest in the 2016 presidential election, 3.37% voted in the primary and not the general election. This statistic may indicate that 3.37% of voters were unhappy with the choices of candidates offered to them on the final general election ballot.

In order to geographically isolate the population that did not vote in the general election, but did vote in the primary election in 2016, statistics by county regarding the *nogenvote2016* variable were generated. These statistics were the number of people who hold a 1 in the *nogenvote2016* variable and the percentage of people that voted in the primary election but not the

general election out of the total number of people that showed interest in the 2016 presidential election.

There were four counties out of a total of 88 counties that had greater than four percent of those that showed interest in the 2016 presidential election who did not vote in the general but did vote in the primary. The county out of those four counties that had the highest number and highest percentage of voters not voting in the general but voting in the primary is the southernmost county of the state, Lawrence County. This county had 5.8% of voters interested in the 2016 presidential election turn out on primary election day, but not general election day.

The other three counties had above four percent of the voters with no general election vote despite showing interest in the election in the primary. They are Mieses, Vinton, and Athens counties. These three counties are clustered in southeast Ohio and separated from Lawrence county by just one county.

Analysis: This area of Ohio is contained in the Appalachian region and is characterized by its traditional reliance on coal. As the former governor of Ohio, John Kasich could have been perceived to have understood the region's reliance on coal better than other candidates. More voters in this region might have been likely to lose interest in the election after his candidacy ended after the primaries. If another candidate had been able to appeal to the interests of these voters who chose not to show on election day, they might have had success in instigating turnout of those voters.

B. Model 2: "Vote switchers"

Introduction: During the 2016 election in Ohio, there was coverage of voters who traditionally had voted for Democratic candidates, but then switched to vote for Donald Trump

(Jacobs 2016). The second model examines the importance of recognizing party affiliation as a variable that can affect a voter's experience in an election cycle.

Data: Ohio's provided voter file indicates in which primary a voter voted, but not how they voted in the general election. The indication of the primary in which the voter voted was used to help generate a new variable of *voteswitch*. "Vote switchers" are those voters that had voted in the Democratic primary in either the 2008 or 2012 presidential primaries but in the Republican primary in the 2016 presidential primary. The variable was labeled with a 1 if the voter is a "vote switcher" and they also turned out in the general election of 2016. All others that voted in the 2016 general election were labeled with a 0.

Analysis: Out of those that voted in the general election of 2016, 7.93% were "vote switchers." These are voters who had voted in one or both Democratic primaries in 2008 and 2012, the Republican primary in 2016, and then turned out to vote in the 2016 general election. There were additional "vote switchers" that did not turn out in the general election.

Ohio's semi-closed primary system has a voter's registration updated on primary day, depending on which ballot a voter chooses to take. So those "vote switchers" that had voted in the 2008 and 2012 Democratic primary elections had been receiving Democratic National Committee ("DNC") messaging until they chose to vote in the Republican primary of 2016. Since the switching of their party affiliation, these voters have been receiving Republican National Committee ("RNC") messaging. This creates implications about information they have received for the past four years as opposed to the preceding years. Information about party achievements and candidates traditionally get distributed to voters that are party affiliated. "Vote switchers" have been registered as a part of the Republican party for the past four years. As of the data download

in January 2020, 84.64% of “vote switchers” are still registered Republican and have been receiving messaging from the opposite party that they were in years prior to 2016.

As a first step, the DNC might want to contact these voters again to try to counteract the recent RNC messaging. Out of all registered republicans, these voters were most outwardly supportive of democrats in relatively recent elections. Ohio has a system through which their party affiliation indicates past votes and not necessarily their personal identification with a party. In Ohio, a voter’s affiliation with a party is connected to their choice in their primary vote. This system is different than some other states, where party affiliation structures provide more of an identity than Ohio.

Chapter VI. Discussion

A. Overview of the Process

Whether their goal is to isolate certain party voters, geographical areas, or other relevant subsets of potential voters, the political data industry's participation within the chain is to provide requested information to campaigns. Holding and delivering the information allows for others to dive into the meaning of that information and then determine the best way to message. The interviewed firm, as well as other firms in the political data industry, complete requests very similar but frequently more in depth to the two Ohio case studies above. A request would be filed by a political entity for the isolation of certain voters, and a list would be produced with contact information, historical voting information, and any other requested details. The political data industry then sends the list back to the requesting organization, where they are able to formulate the most efficient messaging to try and push these voters to provide the desired turnout.

Political data firms have access to information provided by the voter file as well as other variables that might help clients reach out to populations of interest. As discussed in the interviews, the professionals in the political data industry compile the voter information provided by the State and merge it with other proxy data. The proxy data might include more specifics on the voter than the two case studies suggest. Both of these cases could have been isolated further by other indicators, such as race, age, party switches at the national party level, and recent movers, for example. The process is intended to answer the questions of where and why candidates and political parties should concentrate their efforts. Answering these questions provides the proper framework for campaigns to answer how communication should be approached.

B. Barriers to entry

The discussion with professionals concerning the definition of what constitutes public voter data suggests a need for further study on the operationalization of “public.” Definitionally, “public” does not explicitly have anything to do with monetary value. The word is defined as the opposite of “private,” and reflects the general view and relates to a whole group (“Public”). Professionals noted that some states do have limitations on the state citizenship of the requesting party, and some do require payment. Further, it was mentioned that consumer data groups have the ability to refuse leasing to certain groups because of safety measures, confidentiality, and other concerns. Some entry into this specialized data analysis role requires access to proprietary information. However, certain data is readily available and usable in the way demonstrated through the Ohio case studies. If barriers to entry into the political data industry due to restrictions on data are deemed a reason for competitive market concern, appropriate legislation to even the playing field for all industry entrants should proportionately be applied.

C. Costs

Notably, the industry’s current role in the campaign process is to compile and hold large amounts of information and disperse it to clients accordingly. This is a specialization role in the market, making it more efficient for all campaigns and organizations to participate in the usage of data. Requesting organizations can be any kind of political organization, notably with varying budgets. This type of specialization of a technical service that supports campaigning creates lower cost for clients of all budgets to access data services. Once the data is produced by one firm, the marginal cost of providing it to many clients is relatively low. Not having this type of specialized industry creates an environment where each individual campaign must undertake their own data operations, thus limiting the usage of data to high budget campaigns.

D. Political Aspects of the Industry

The political data industry is a client-focused service industry similar to that of marketing agencies, print shops, and polling. These firms are currently operating under client mandates and are thus serving as a resource for those engaged in the political process. Some firms demonstrate political affiliation based on the types of clients with whom they choose to work, and some do not. The political leaning of a specific firm is determined by the types of clients with whom they choose to interact. Some firms have specific binding contracts with political entities, and thus take on political goals of the respective clients, but do not require a company-wide viewpoint of the political process at the employee level. As indicated in the interviews, operationally, industry employees are tasked with technical jobs that require understanding of the political process but are not required to take a uniform stance.

On the individual level, the interviews demonstrated that these professional industry jobs are not political. However, political beliefs of those working in the industry do play a role in the types of clients for whom certain individuals will choose to work. Certain firms do attract specific sides of the aisle, depending on their past work or their determination to only be of aide to those groups of a certain political party.

Making the distinction between choosing clients and working outside of client relationships is also important to understand the political effects of the industry. These industry firms, as demonstrated by interviewee discussions about leaking and security, have an incentive to keep confidentiality and work inside of contracts. Aiding anyone outside of contracted client relationships would damage the reputation of the firm in the specialized service market. Therefore, they do not have a monetary interest in acting on their own political beliefs within the company.

The only way there would be an incentive to do so is if an outside political group contracted with them for the same interest.

Nonetheless, these professionals do have access to high-end information that could be of use for their personal political goals. If they or the firm as a whole were to operate outside of client requests, either would be operating to serve their own interest to affect the political process. This would no longer be a specialization role in the chain of campaigning. Instead, it would be another political organization seeking to affect the process and should be regulated as such.

E. Democracy

If the political data industry were directly acting to achieve political goals, democratic ideals that strive to create equality of vote would be tested. Data usage as a singular form of recognizing people to communicate with could exacerbate differences in political engagement choices. Operating solely on a large dataset that presumably favors those that are registered and locatable creates implications for those voters who are not listed. Among those who are not listed are people who are frequently residentially mobile, people of color, and people with low median incomes. This creates implications for the policy considerations that are important to them. If they disproportionately receive less contact, their interests will not be communicated as effectively to those in office or running for office (Jackman & Spahn 2018). However, as one interviewee noted, he doesn't feel that much, if any, new information about political matters is being communicated due to data distribution. Further, the specialization of the industry deals more with the type of messaging that is used to get communications across.

The power of comprehensive data on the American electorate and the power of skills possessed by the professionals that understand its complexities are not necessarily measurable through the type of informative research undertaken in this project. But data usage is alarming if

massive amounts of information are directed towards one group's interests and certain populations disproportionately receive no contact.

One professional pointed out that this should not be the center of an argument for restriction on communication. Arguments that are contingent on questioning the free will of a citizen's choice to vote are based on undemocratic principles. Who is one to speculate on the choice making ability of any individual who chooses to exercise his or her constitutional right? Or, if they had not turned out, what would the result have looked like? The democratically optimal situation is to have increased turnout in any format.

One point made by the professionals should ease the worry about disproportionate contact. Many noted that although it is not in their purview, campaigns still take it upon themselves to contact those that are not listed. It is efficient for campaigns to do this because they have not had to work as hard to reach the first portion of voters that are listed because of widespread data usage. This gives them the ability to take on the humanizing aspect of campaign work and connect with voters as they try to find others who could help them with electoral success, despite their absence on voter lists. Further, those that are not listed are nonetheless citizens with agency to become known to a political process by showing up to vote or to remain invisible by choice. Overall, campaigns have every incentive to find them and have them show up to the polls.

In connection with discussions about disproportionately contacting the electorate to vote, many professionals emphasized that data analysis has drastically increased "Get Out The Vote" efficacy. People are being properly identified and contacted with uniform, up to date information, thereby creating the environment for election participation encouragement to flourish. An interviewee referenced the 2018 midterm election's high turnout rate as being correlated to ever-increasing data usage.

Enticing more people to the polls to cast a ballot in their best interest is consistent with furthering democratic ideals on which the United States was founded. However, further research should be completed to assess the true impact that voter data lists have on disproportionately contacting the electorate.

F. Future Regulation

Professionals familiarized the data usage process by relating it to older campaign methods that also recognized specific voter behavior. Recognizing certain voter behavior was then, and is now, crucial to winning an election. Using lists of people to contact is not a new idea. It has been around since the concentric circles model of voter behavior and possibly before (Sigelman & Jewell 1986).

In response to questions about consumer data, interviewees also familiarized the process to the way corporations try to push messages out to consumers. All forms of targeting, regardless of data usage or not, are intended to isolate certain populations with certain interests and communicate with them effectively to promote a certain type of behavior. In the end, data usage is not creating a new complication of campaigns' messaging to the electorate promoting certain behavior.

Communication between doorknockers, neighbors, and the highest forms of campaign messaging and corporate actions will all have an effect on the way the electorate will turn out. If regulation were to be determined to be in the best interest of the electorate, it would need to be proportionately applied to all forms of messaging. However, this type of regulation might have further implications on free speech principles and their relationship to elections and should be addressed in this context. Overall, an arbitrary line drawn between data usage and other forms of messaging without justification would be discriminatory.

In 2019, WinRed was created as a Republican donor platform to match the Democrat ActBlue platform that had been in operation since 2004 (Hakim & Thrush 2020; Willis 2014). Both of these platforms are singular entities that are used to store donor data and information while also being a platform for campaign donations. These two functions—one regulated and one unregulated—being synthesized into one platform and working in conjunction with each other is novel. Campaign financing is regulated currently, and the close relationships between data collection and picking out donors represented by these donor platforms should be carefully assessed and regulated accordingly.

Chapter VII. Conclusion

This paper has discussed the power that comes from holding massive amounts of data on the electorate. However, the current monetary interest of these firms is to operate strictly inside of client contracts, which are presumably political. At their core, political data industry firms do not represent personal political interests. Instead, they demonstrate political interests through serving the interests of their clients.

Within campaign processes, there are services such as those that the political data industry offer that are in operation to serve the needs of campaigns in an efficient and specialized manner. The power held by some of these servicing firms has increased with the rise in technological specialization, such as the political data industry. There is no barrier to entry to access most of the data, but the skills and technology infrastructure is much different than that of a bumper sticker shop's printer.

Campaigning has become too large for the comfort of some, and its effects on expression of political interests at the polls should be addressed as a whole. To the extent that there is a discomfort about increased power maintained by some campaign services, the discomfort should be quantified and studied further. However future studies should be specific about their quantification of data usage and campaign messaging in order to properly represent any impositions on the electorate.

These services are contracted to achieve a certain outcome and have the technological power to often instigate that outcome if the proper steps are taken. If the public is under the perception that current regulations are not able to adequately address their concerns, concerns should be quantified and studied. Reconstruction of rules as a whole for campaign services would be of benefit to those with the belief that this type of service is adversely affecting political

representation on election day. Those populations could take steps to elect lawmakers that would implement such regulations. These regulations could include transparency measures similar to regulations of interest groups' finances.

There is a new reality in politics that includes specialized outsourcing of services. The country is no longer operating in the world that existed when campaigns were still being undertaken out of politicians' homes. The overall nature of campaigning has changed, and it is possible that new appropriate regulations should be implemented if the public feels it is in their best interest and the country's best interest to do so. Potentially, a new definition of the political service industry and attached regulations may be needed.

But, once again, this is not just an issue of data. This is an issue of service industry advancement. An arbitrary line that is drawn to regulate data usage for campaigns will not be effective in treating the industrial nature of current campaigns. These regulations would have to apply to all outsourced services to achieve the effect of redistributing equality to all voters. If only one specialized outsourced service were to be restricted, others would fill the legally mandated void with an alternative. This would not achieve the goal of protecting the electorate, if such a need were to be determined.

Appropriate follow up analysis on the effectiveness of regulation on the subcategory of political services as a whole would be essential. Further research and effects on the entire lifespan of data usage from compilation to final messaging should be completed. The effects of messages on the electorate should be understood in the context of the environment in which they were formulated.

Bibliography

Brooks, David. 2014. "Death by Data." *The New York Times*.
<https://www.nytimes.com/2014/11/04/opinion/david-brooks-death-by-data.html> (November 6, 2019).

This opinion piece argues that campaigns that have overly relied on data have experienced significant downfalls in outcome. Strict data tactics rely on Impersonalism which disregards character and creativity of individual voters and sees them as demographic slices instead. It assumes that voters are not free thinking and that they value messaging tactics over policy substance. It also references the downfall of a race in Colorado where a candidate spent too much time appealing to women's issues and too little time on other policy. Overall, this article argues that it is important to keep creativity of campaigns evolving alongside technology innovation.

Hakim, Danny, and Glenn Thrush. 2020. "How the Trump Campaign Took Over the G.O.P." *The New York Times*. <https://www.nytimes.com/2020/03/09/us/trump-campaign-brad-parscale.html?referringSource=articleShare> (March 13, 2020).

This article discusses the irony between Donald Trump's involvement in the launch of a strictly G.O.P. donation platform called WinRed. Involvement of data industry firms are referenced as players in making the launch a success. There is also discussion about the data that is used to pinpoint potential donors.

Holbrook, T. M., & McClurg, S. D. (2005). The Mobilization of Core Supporters: Campaigns, Turnout, and Electoral Composition in United States Presidential Elections. *American Journal of Political Science*, 49(4), 689–703. doi: 10.2307/3647691

This study examines impacts of presidential campaigns on voter turnout and electorate composition. Persuading is needed before mobilization of independents and they comment that is the reason that campaigns typically look to partisans to direct campaign expenditure. Also similar to the "Election campaigning on social media" article, they mention self-selection of the population into subgroups based on rallies and candidate appearances. The study finds that there is change within the number of Democrats in each state election to election, with little change in Republicans. Additionally, national party money being transferred down to state and local levels has the largest effect on outcome and is notably effecting the behavior of independents.

Holman, M., Schneider, M., & Pondel, K. (2015). Gender Targeting in Political Advertisements. *Political Research Quarterly*, 68(4), 816-829.

This paper defined targeting as sending a tailored message with the intent to have “the group evaluate favorably” and identity-based targeting as targeting but through their attachments to “politicized social groups.” There is also higher turnout for females and females more frequently use their gender as a factor in their vote decision. The study finds that identity targeted messaging towards females increases the likelihood for voting for a candidate, that there is no difference between effectiveness of female versus male candidates’ targeting in likeliness to change voter turnout. However, females who viewed the identity-targeted video from a female candidate were more likely to reference their closeness to females as a group as an aspect of decision making. The paper mentions normative implications these findings have on future descriptive and substantive representation.

Jackman, Simon, and Bradley Spahn. 2018. “Politically Invisible in America (working paper).” <https://www.usc.edu.au/analysis/politically-invisible-in-america>.

In this working paper, the authors investigate trends that are associated with being politically unreachable. Consumer files are frequently paired with state voter files to create campaign lists that notably have imperfections. Campaigns are unable to contact those not on these “lists,” and thus registered and unregistered people on the list report much larger levels of contact than those that are mislisted. These lists also systematically exclude minorities, the poor, and liberal ideology. The paper covers normative policy concerns that arise through underrepresentation and emphasizes the constitutional right for those to be politically represented despite financial circumstance.

Jacobs, Ben. 2016. “Donald Trump Taps into Manufacturing Regions to Extract Democratic Voters.” *The Guardian*. <http://www.theguardian.com/us-news/2016/jul/03/donald-trump-campaigns-ohio-pennsylvania-democratic-voters> (March 15, 2020).

This is an example of coverage for “vote switchers.” There is a proportion of the population that had previously voted Democrat for years and then decided to vote for Donald Trump in the 2016 election.

Jacobson, G. C. (1987). The Marginals Never Vanished: Incumbency and Competition in Elections to the U.S. House of Representatives, 1952-82. *American Journal of Political Science*, 31(1), 126–141. doi: 10.2307/2111327

Suggested “safety” of incumbents in Congress was challenged in this article, citing that probability of incumbent loss has not actually changed since the 1960s-1970s. Marginality was arbitrarily defined in these studies. In turn, incumbents are increasingly using all campaign methods available, including social media, larger campaign funds, and draining travel allowances. The article mentions that lack of safety of incumbents is also heavily related to the fact that social media as a resource is universally available, making it an even playing ground for challengers. This has forced many Congresspersons to turn to creating districts off of their specific persona rather than relying on partisanship. These districts are important to the change in interaction because now there is less reliance on party messaging.

Keller, Kevin Lane. 2009. "Building Strong Brands In A Modern Marketing Communications Environment." *Journal of Marketing Communications* 15(2-3): 139-155.

This piece focuses on consumer responses to branding and marketing in the marketplace and defined many issues that have to be covered by brand marketers. Brands have to cultivate consumer knowledge and perception, and consumers do this frequently off of loyalty, the quality or functional use, and outside social perceptions within their social circle. In order to best do this, they have to allocate their resources efficiently at the steps in the process where communications interference matters most. The messages also need to have versatility to have an effect on different groups of consumers. Additionally, the internet has posed both some complication and efficiency now for brand marketers, because there is the introduced ability to skip and tune out ads but also opportunity for more “interaction and individualization.” I chose to include this piece because the techniques and difficulties prove to be very similar to early-stage voter targeting.

McDonald, Michael P. 2007. “The True Electorate: A Cross-Validation of Voter Registration Files and Election Survey Demographics.” *Public Opinion Quarterly* 71(4): 588–602. doi: 10.1093/poq/nfm046.

This article assesses the validity of certain voter information databases. It compares state voter files, the Current Population Survey (CPS), and the National Election Polls (NEP) on the basis of their representation of age, gender, and race. Overall, the CPS and state voter files represent demographics fairly similarly, while the NEP represents the voting population as younger and less white. However, all of these databases have the potential to be biased through collection means. However, it was noted that state voter files prove to be of greatest value given that they provide the most contact information about a voter.

Nickerson, D. W., & Rogers, T. (2014). Political Campaigns and Big Data. *Journal of Economic Perspectives*, 28(2), 51–73. doi: 10.1257/jep.28.2.51

Campaigns have the end goal of victory. The paper describes data as having improved efficiency by letting campaigns allocate resources efficiently. Behavior, support, and responsiveness scores are typically what are used, and these are created from an array of publicly available data. The paper makes a point to mention that these are not causal but only correlational discoveries because these are not controlled experiments. It also discusses a couple conclusions regarding the high effectiveness of smaller campaigns using data and inherent tensions of geographical mobilization and policy outcomes. At the root of the paper, it is understood that data is only useful if there are politically smart campaign employees that understand how to most effectively conduct outreach to populations on the ground given certain analyst conclusions.

Of mud and money. (2012, September 8). *The Economist*. Retrieved from <https://www.economist.com/business/2012/09/08/of-mud-and-money>

“Of mud and money” touches on the progression of campaign usage of digital advertising. During the time the article was written (September 2012), broadcast TV and website banner ads were heavily used, but the article does mention that there is an increase in cable TV and video ads on the internet. The article also defines micro targeting as “tailoring messages to specific groups of voters” and notes that this type of advertising was initially used to fundraise but is now primarily used to mobilize voters to the polls.

“Ohio Presidential Election Voting History.” *270toWin.com*.
<http://www.270towin.com/states/Ohio> (February 18, 2020).

This page shows Ohio’s voting history specifically in the context of the state being a sought-after battleground. Voters are fairly split, and the state has a relatively high number of electoral college votes.

Pew Research Center, March, 2019, “Political Independents: Who They Are, What They Think”

Frequently, Americans that describe themselves as politically independent lean to either one party or the other. As of 2018, 38% of Americans described themselves as independent, but just seven percent did not have a lean one way or the other. This number of “true” independents that are frequently not politically inclined has remained less than or equal to ten percent since 1994. The majority of Republican and Democrat leaners have a favorable opinion of their own party and are almost as likely as claimed partisans to have an unfavorable opinion of the opposite party. These leaners on both sides have increasingly described their conservative and liberal ideologies to match their parties, respectively. Men overall are more likely to be leaners of both parties and are more likely to be true independents. Both leaners and independents are less likely to be politically engaged than partisans.

“Public.” *Merriam-Webster*. <https://www.merriam-webster.com/dictionary/public> (March 1, 2020).

This page gives a definition for “public.”

“Register to Vote and Update Your Registration.” *Ohio Secretary of State*. <http://www.ohiosos.gov/elections/voters/register/#gref> (February 18, 2020).

This page indicates how registered party information is updated in Ohio. It is updated on the day of an election, based on which primary ballot a voter chooses.

Rusch, T., Lee, I., Hornik, K., Jank, W., & Zeileis, A. (2013). Influencing elections with statistics: Targeting voters with logistic regression trees. *The Annals of Applied Statistics*, 7(3), 1612–1639. doi: 10.1214/13-aos648

This paper creates a framework to prove the effectiveness of using voter information to target during elections. It defines two steps to microtargeting: the first is identifying people to mobilize, and the second is identifying the most effective mobilization tactics for those subsections of people. They conduct a case study in Ohio and conclude that in order to use the quadrant setup where likeliness to turnout and support issues intersect, you must use both vote history data (past turnout numbers), and voter file profile or “proxy” data. They also determine that logistic regression tree models do better in reality because they allow for different regression models to fit different subsections of the population based on certain set criteria. Through the trees you can make predictions at the individual level.

Sides, J., & Karch, A. (2008). Messages that Mobilize? Issue Publics and the Content of Campaign Advertising. *The Journal of Politics*, 70(2), 466-476. doi:10.1017/s0022381608080432

This study seeks to determine the breadth of effect that campaigns and mobilizing have on population turnout. Mobilization has been proven more significant for turnout above education and other factors. The article cites that campaigns use “sophisticated databases” to help pinpoint markets they hope to be successful in mobilizing. Specifically, when campaigns target their ads towards the populations that the specific policy is deemed to have the most effect on, it was determined that there was very little or not any correlation to success in mobilizing those populations to turnout. The scope of the article is also important given that targeting has been suggested to “exacerbate” systemic inequalities among voters. Despite mostly null results, the article helps define what campaigns can and cannot accomplish through using data and targeting.

Sigelman, Lee, and Malcolm E. Jewell. 1986. "From Core to Periphery: A Note on the Imagery of Concentric Electorates." *The Journal of Politics* 48(2): 440–49. doi: 10.2307/2131103

This paper tests the theory of modelling the electorate in concentric circles that encompass each other. Modelling voters based on core to periphery was proven to be unsuccessful in this paper. Though, using this model of segmenting voters shows the consistent goals of campaigns to create different messages based on certain traits of the electorate.

Stier, S., Bleier, A., Lietz, H., & Strohmaier, M. (2018). Election campaigning on social media: Politicians, audiences and the mediation of political communication on Facebook and Twitter. *Political Communication*, 35(1), 50–74. doi: 10.31235/osf.io/58u4c

This article uses a study on a German federal election to emphasize variation between campaign messaging from social media platform to platform. It is useful to understand that social media is comprised of self-chosen networks. Politicians have the ability to use this self-selection aspect to send certain messages to subsections of the masses. The study found through examining keywords that social media messaging is notably different than mass messaging, and interests are distinct from the masses. Additionally, campaign events are discussed more than policy on social media. Topics on Facebook are often more disconnected from the audience interest than Twitter.

U.S. Department of Commerce. Bureau of Census. 2018. *Quick Facts*. Washington, D.C.: Department of Commerce.

This page clarifies the current population of the United States in 2018.

"Voter Files Download Page." *County Voter Files Download Page*.

<https://www6.ohiosos.gov/ords/f?p=VOTERFTP:HOME::::::> (January 4, 2020).

This is Ohio Secretary of State's page where the Ohio voter data was downloaded from.

Voter Privacy Act of 2019, S.2398, 116th Congress. (2019).

Dianne Feinstein introduced this bill on the Senate floor on July 31, 2019 and it has since been referred to the Senate Rules and Administration Committee. The findings list that targeting services are not "changing minds through persuasion but 'informational dominance'" and the bill works to give voters access to their personal data, the ability to erase and prohibit transfer, have notice of receipt, and prohibit targeting of themselves based on their personal data. The definitions subsection is also useful in breaking down the process from a "covered entity" (i.e. a candidate) to a targeting service, to the individual. It also lays out the government's broad definition of "personal data."

Willis, Derek. 2014. "How ActBlue Became a Powerful Force in Fund-Raising." *The New York Times*. <https://www.nytimes.com/2014/10/09/upshot/how-actblue-became-a-powerful-force-in-fund-raising.html> (February 18, 2020).

This article is cited with the year ActBlue was launched as a donation application. It also discusses some of the data usage the application implements to pinpoint donors.

Appendix

(I) Interview Questions

1. What is your position within the company?
2. What tasks do you perform day to day?
3. What types of clients do you typically provide services to?
4. Describe your part in the chain of input to client.
5. What differences, if any, do you see between this company and other political data firms
in the ecosystem?
6. Please elaborate on any differences with regard to focus and tasks performed.
7. Can you describe efforts put in place to prevent a breach of privacy?
8. Is all data used publicly available?
9. Are there similarities between consumer targeting and voter targeting?
10. Some scholars have noted that this kind of analysis may skew electoral results. Do you have
a response to the literature claiming this?
11. Do you have an opinion on whether the use of political data has substantive representation
implications?
12. Do you have concerns for the future of this industry? What about security of the data?
Government regulation?

Follow up questions:

Does it involve ____ or ____?

Please clarify your answers. Please expand on that answer.

Can you expand on that? Can you provide a hypothetical example?

(II) Interview Data Table

Subject	Barriers to entry	Public Data	Data Restrictions
Subject 1	There is a common misunderstanding with voter data, this is actually public information, anyone can request it from the state.	Performs daily tasks of making requests of public data typically from Secretaries of State or counties.	"Yes public, but depends on what portion of the public you are in." This is regarding that "some of the data that we process is only available for request by a registered voter in that state." Once you have it, it can be used for any purpose outlined in state statutes, typically "all political reasons"
Subject 2		They are "in the middle" regarding their role in using additional information that campaigns might have collected from events etc. They are responsible for that information and that isn't always public	
Subject 3	Consumer data is expensive. BUT this also brings the ability for lower budget campaigns to succeed. If all campaigning was blanket messaging, low budgets couldn't afford it and people would be upset.	Voter file data is all publicly available, in some states its free, others its not. They coordinate agreements with state parties to license data on their behalf. Campaigns also collect data from traffic on their website.	There is no way to actually know who you voted for. Your party affiliation is only given out if you live in a state that does that.
Subject 4	A significant amount of it you do have to pay for	Drew a distinction about public information "Can anyone go online and download it? No." Ex: In Pennsylvania in order to get the voter file you have to be a registered voter in that state. In Michigan you have to fill out a FOIA request. If you take publicly available as above definition, only two or three states have it publicly available as above	But mentioned that they aren't storing social security numbers or other secret information. However, brought up that voter model scores are just a "snapshot in time" and that they are constantly changing and are therefore only good until the next build of the voter file.
Subject 5		"Believes that" consumer data is not publicly available. Brought up the distinction between publicly available and paid for again. Historical results are the "backbone" of labeling a piece of geography like this (referring to walk-app).	
Subject 6		All 50 states a private person can request a voter file if they want it.	
Subject 7		There is a concept of proprietary data "restrictions around usage" Ex: Verizon cannot use this voter data for business development. "We are not operating outside of any SoS mandate." What information they provide state to state varies... hinted at that being a state protection deal.	
Subject 8	InfoGroup vets vendors pretty hard, not sure if anyone can necessarily buy it.	They only get data from public places. Also "helpful that it is all public data" and not mined. Anyone can go and get this political data.	Very helpful legally that there is no information on minors.
Subject 9	Mentioned that if others are not getting results, they should step up the data efforts.	Doesn't know much, 50/50 he thinks. If public=free then 50/50.	
Subject 10		No, not all, but most. Most is from the SoS but they do buy some from consumer data companies to fill in the gaps on demographic issues. Also, can make relevant consumer data available to clients. Most of the consumer data is modelled and gives ranges	

Subject	Understanding of client relations	Understanding of technical jobs	Clients
Subject 1	Does not provide direct client services.	"Once the data is received I cleanse the data so it can be ingested into the database."	Political clients, pollsters, data analytics groups, advocacy groups
Subject 2	Has a main roll in billing and contract terms with clients.	Deals with onboarding personal information and budgeting.	Political campaigns and organizations. From a cost perspective, they are "low-cost which reduces barriers to entry" for small political campaigns to use their services.
Subject 3	Works with digital agencies.		Works directly with the digital agencies attached to campaigns. The digital agencies are the ones actually placing ad buys on Google or Facebook for campaigns.
Subject 4	The internal team doesn't interface with clients frequently. They do get requests from clients to build a specific file and then they have to put in the requests to the state or state party to legally collect it	Transform, cleanse, and compile precinct data... collects raw files and builds the voter file. Also troubleshoots problems with queries and running tables. Audits voter files to make sure they aren't odd: i.e. states will sometimes leave voters that have cancelled and he will check that.	
Subject 5	They also will fulfill requests from people they have data sharing agreements with. External team will ask for something specific that will then go on to the client. Direct asks from partners by their regional data directors, they will ask him directly for results.	Building states using historical election results. Ex: New York 2018 by precinct but get vote totals for president. They start at federal level and collect vote totals up until the state house level typically	
Subject 6			They don't really do much of the client stuff, but from what interviewee knows, the [national party], campaigns, advocacy groups.
Subject 7	He typically tries to figure out what the current/new client's biggest needs are to decipher how data plays into their goals or problems that need solving. Do they care about working with new data ongoingly or getting a view/answer to the question day to day? Then gives access to appropriate technology or assigns one person from staff to be in house, etc. Just looking at the "best way"		A mix of political consultants that handle the modeling, polling, and digital advertisements. Also trade associations that are passionate about different issues. SuperPACs are handled as "large outside groups." They "are not campaigns but engaged in the political process."
Subject 8	Getting everyone all set up, making sure that the states and the counts get updated. Deals with special requests, ex: adding emails to the lists. Or ex: a client is requesting a different form of delivery, how can they best integrate	Re: clients makes sure people have the right credentials, etc. re: geographies, parts of consumer**access focused, whereas external is focused on helping clients use the data	Helps with technical difficulties. Front facing advocacy groups or campaigns. The clients are "politically skewed."
Subject 9	Others clean and put into the tables, external gets the requests. Then he does the SQL query that gets it to match to other tables. Uploads to Tableau and builds dashboards, etc. and passes along to client in form of PDF or online access		Political organizations, campaigns, getting more into membership groups.
Subject 10	He receives an email from a client. Ex: We want this many people from Alabama that have voted in 1 out of the last 4 elections, then he will create that table.		Polling companies are the number one client. Also, campaign consultants that also do the polls, he gets the feeling that they also do more than that.

Subject	Comparison with other data firms	Safety	Consumer Data comparison
Subject 1	Personally no experience with other data firms.	Common misunderstanding that this even needs security... primarily for protection of the work of the company.	"One of the goals is to merge those two processes so there's not just a political or consumer list but one that includes both." Indicates that consumer style data is useful
Subject 2	\	"Security is something we take very seriously." There will always be a new way to hack though.	"In the sense that we are using historical information to create an audience."
Subject 3		They vet the people that the data travels to through heavy questionnaires. People that have access to the data need to have proper security and not have bad interests .	Big consumer data groups don't actually track what you buy, its "not exactly what you bought, but it buckets it into a category" that's all based off of likelihood to buy something or shop somewhere.
Subject 4	Most firms in the industry lean one way or the other	"We patch all of our Azure machines twice a month and require a VPN to access the Azure environment." Multifactor authentication They also have the ability to do endpoint data discovery through which they can monitor the access of the Azure AND the hardware. Reputation would be devastated if there was a breach. End product is proprietary and it would be "devastating"	When you use consumer data to target in the voter sphere, you are not pushing products, you are pushing messages. You are not asking for someone to repurchase the granola bars they got last week "The clients are trying to convey a message" "The end goals are very different"
Subject 5	"seems like other firms tend to specialize a little more" regarding communications, polling, etc. They are a "central hub." They spin off information to other people that need it. They can do a lot of things and are not necessarily into one category	"We've definitely stepped up." After 2016 there was an obvious need to preemptively step up. They frequently test for the usual sources like phishing, has personally seen phishing tests and trainings, although cannot speak on the infrastructural changes higher up. Mentioned up tightened up usage of the VPN and secured ways to retrieve shared information. "Kept on our toes." Laptops can be remotely disabled now if they are lost or stolen.	"Believes that" consumer data is not publicly available. What he deals with is all public, but it might cost money to get it in the format that you want it in. Drew a comparison between the end results of both-not super equipped to be commenting on consumer data usage.
Subject 6	We update the voter file more frequently than other voter file providers, and don't provide as much analysis. Tend to "stay out of that" and let the experts do it.		
Subject 7	They are the only ones that build up as complete, comprehensive, and up-to-date databases. Others are "not as accurate or enhanced as us." Others are "value adds" to our data. They fill a dual role of being client facing but also working product relations. No one else creates and delivers product of this size and scope.	He doesn't specifically deal with the implementation. Security is important because while a lot of information is a public record there are a lot of restrictions around how that data can be utilized. Safeguarding on the individual to the corporate level from being passed into the wrong hands.	Marketing departments at big corporations are doing the exact same thing political data firms are. "leverage statistics about individuals" "deliver them in the most tailored advertisements they can" It does differ though on what data they connect and how they use it.
Subject 8	"Wish I knew more." Used analogy of a library, firms can come use information for analytics and targeting.	They have now started adding client codes into the data, so they are able to tell who leaked it. Need to keep the "power and say over what our data is doing"	Consumer data is difficult in that there are a lot of different endpoints that they have to match people back to. Notes though that the consumer data is not political in any way. "In both cases you're trying to find a niche of people that have certain qualities and traits." The goal is to speak with a specific group about a specific thing, and this is just another way to know who you are looking at.
Subject 9	Not much other experience so can't really say. But thinks that they have a lot more data than other data companies.	Change passwords every 30-60 days. Have a tech company that they use for antivirus and antihacker practices. Must use secure networks at all times. There will always be a concern for security—they don't want people getting into the data that shouldn't have their hands on it. Mentioned that the 2016 meddling was an issue and proper measures should be taken to ensure that doesn't happen again. "definitely need to keep it as secure as possible to prevent that from happening in the future"	
Subject 10	This is his first data job so unsure.	IT security on actual devices. Must use a VPN. On his part, every file sent to a client is sent with an "exploding link" that expires after 7-14 days. New thing that they have started is that all files have an identifying column to track the "chain of custody"	Yes, basically the same thing. Product-candidate. Targeting likely voters is the same as targeting those that are more likely to act on buying our product. Doesn't think that the political data they have is anything worse than consumer data companies

Subject	Another form of campaigning (efficiency)	Campaign money	Not skewing
Subject 1	"Adding a component of data analysis to it will make campaigning of increasing efficiency and effectiveness." This is "not a whole new thing in itself."	Money is still the center of election results, nothing has changed because of data collection. "The candidate with more money can allocate more resources to make the campaign more effective."	"Any sort of campaigning skews election results."
Subject 2			"disagree"
Subject 3	Every campaign already has a list of who they should be talking to, this just creates more "transparency." Campaigns can be directed to talk to the right people.	"Shutting out lower-dollar campaigns and fundraisers when you remove campaign targeted advertising."	Its "refining" not skewing.
Subject 4	"Am I expected to call every voter regardless?" Ex: In charge of campaigning for a Republican but I have to talk to all registered Democrats too... this is inefficient. "Just taking it to an advanced level"		When it comes to democracy, "where do you draw the line?"
Subject 5	"If there are voters who will show up for someone, campaigns will find them" "working to contact the people regardless".	Ex: McCain-Feingold regulating money --thinks that these "concerns are misguided" though. Billionaires on the Democratic side are not even able to "move the needle." If money meant a win that would not be the case.	The point of campaigning is to influence the election in your favor. If campaigning is skewing... running an ad is also targeting. Ex: Without campaigning we would only be able to have the two candidates have a neutral TV debate, even then how do we make it even, they are still getting their names out there, in this world that doesn't exist.
Subject 6	For the people that use our data that's kind of the point. They use the data to reach their desired outcomes. Mentioned the opposite side of the coin of skewing being it provides information for those to conduct further outreach and reach the campaign outcome that they would like.		
Subject 7	Its fair game if one campaign knows just how to do it best. Ex: If Apple sells me a phone better than Samsung that's on them.		Anything that relates to campaigns and political organizations execution of political actions will skew electoral results. Went into talking about the polarization and difficulty the two sides are having talking with each other. The "use of data does not exacerbate the problem there."
Subject 8	There is no reason to waste time talking to people who will just never bite. It's the same thing where they show truck commercials on ESPN and not on E! The other team at this point wasn't trying if they can't be heard.	The one person one vote issue was not a spicket that was turned on because of data usage—alluded this is more of a campaign finance issue.	That's the point, advertising may skew, not a bad thing necessarily. You can see the success in results. Everyone now is media targeted, you are going to hear some of both sides no matter what.
Subject 9			
Subject 10	Ex: there's a reason why Clinton didn't try for Alabama...No unfair advantage		He doesn't think data adds any sort of new wrinkle to campaigns skewing. No, as much as they can say, targeting in itself does not really actually put much new information out there. "I don't think people are learning a lot of new things from targeted advertising."

Subject	Government Regulation	People who are scared about data collection	Description of their industry's role (A-political?)
Subject 1	Government regulation can be the way they will let you opt in or out of collection and beign listed on a registration list.	A personal decision to be afraid. Subject has resigned to the fact that companies and political organizations will have increasing data accesability. Ex: Oregon allows you to opt out	
Subject 2		There are huge efficiencies to this type of data usage in life and in voting.	
Subject 3	When you have have hearings like the one about Facebook, it is clear that Congress isn't clear on what they are actually discussing.	"People get paranoid when they hear about data usage." There is a "lack of understanding on what they know about you." They only have information about you that you have willingly given them. There should be a clear way for government to address usage because its "not actually scary"	Most of the time that data is in different formats, their job is to refine it down and put into a usable format.
Subject 4	Did talk about restrictions on Google and Facebook that maybe are worrisome. But it is a potential positive given that "with regulation comes innovation." Potentially a new service would allow them to even better accomplish what the client wants... more regulation=more innovation.		Doesn't do much analysis. Overall a less political interview so couldn't get much on the government regulation aspect I can get information compiled and from one place to another but not very political of a job. "I'm not one to really follow the news"
Subject 5	Government regulation-welcomes it, the industry just has always adapted. Regarding regulation on privacy: the data they have isn't a surprise to anyone... "that's my name, address, etc. yep that's me"	Everyone is generally skeptical of whatever privacy is promised digitally. That's "kind of the world now." A lot of states have VH collection that does not include who you voted for, so "voter report cards that are sent to your house can be kind of startling if you aren't expecting it.	Chopping up places for immediate usage by a campaign. Deals with election results down to the precinct level.
Subject 6	Mentioned that there will be a constant demand for information, and that those in government are unlikely to do anything to limit the usage because they are the ones primarily using the benefit to hold office. Doesn't foresee pushback.		Regarding analysis for political ends: Tend to "stay out of that" and let the experts do it.
Subject 7	No, with the way that the electoral system currently works. Some people say that polling might be dead, but no there will always be a want to know more, it will just be in a different way. "Desire to know things will always exist." In "lockstep" with that, we will always want to know who to talk to and how to talk to them. Finds it hard to believe that in 20 year this will be gone.		They don't do polling or modeling like others do but they make sure they have a good jump off when they do. Well first, it's very difficult to introduce voting as a behavior. Mentioned their own focus on persuasion.
Subject 8	No. We have been trending in a way where this isn't going anywhere. Facebook is different, these people didn't know that this was being tracked. Could see consumer data possibly falling off of the map.	You can call all of these (consumer data) places that collect and opt out of collection, but they get their data from tons of different sources.	Clients are politically skewed, but this interviewee is just is technical. This firm doesn't really venture into actual targeting or analyzing. They have definitely carved out their own spot and are not "competing but working with other firms." ***This industry is the library, not creating the data itself like Facebook.
Subject 9	It will be interesting to see how regulated it gets.		
Subject 10	Compared his expectation of regulation to the fact that consumer data isn't regulated. He thinks the biggest risk posed is if more social media turns into Twitter and bans political advertisements. "I don't think there is a lot of regulation that could impact us, that could get rid of data information as a whole."		

Subject	Description of job process	Political ends	This is democracy
Subject 1	Make requests of public information, acquire lists of registered voters, typically from Secretaries of State or county clerks. Then cleanses.		
Subject 2		GOTV efforts to be more effective. Voting information from last elections directs survey work and GOTV efforts to be more effective.	"helping to get people to the polls"
Subject 3	Create and onboard digital client relationships. A lot of data is offline, taking it online and making it actionable on digital platforms for clients. Works directly with the digital agencies attached to campaigns, those are the people placing actual ad buys for campaigns on platforms like Facebook or Google.		If you remove the ability to target people and require all campaigning to be blanketing, this creates more noise and makes people upset.
Subject 4			Further example, if you can turn out a voter to vote that is 90% Republican but only 10% likely to vote, you are still significantly improving turnout and democratic ideals.
Subject 5	An election happens, and then the results are certified 3-4 weeks later, calls the SoS or the Board of Elections to see if the results are available from there, obtains a PDF or a CSV and puts it through the consistent company formatting. When there is a request for that state, they will run a SQL query that returns the information with the specific parameters given. They will pass what the request returns along to someone with the respective final skill (whether its making maps, media markets, etc.)	"I've seen the walk app's ability" to label certain houses and to label if this area is changing or which houses have regular voters, etc.	It "may seem a bit scary, but from an electoral standpoint it helps connect people to democracy and to the people they support." Usage of data is a "worthwhile approach to targeting" if you're getting people to show up and turnout in democracy. This is participation. *****Regarding the follow up explanation of "politically invisible": Even if they are not banked or haven't voted, etc., but they are still likely to support someone, campaigns WILL find them to make them turnout. The point of targeting is to drive people out to the polls.
Subject 6	Maintain the national voter file and update states as new information is given by the SoS. Develops the voter file that clients use to target, that others in external portions of the company use to query and look for people, what those in Business Intelligence use to create dashboards, what those working with GIS use to create maps. External department provides accounts and analysis for clients. "I create a product that is then dispensable" to other portions of the company or other companies.		
Subject 7	Mix of writing proposals and doing background market research on trade associations, campaigns, and superpacs. Manages 2 employees at client sites and makes sure they have what they need. Works a little bit on marketing materials used to preserve the branding, noted this used to not be an issue but now in the more recent space it is. Doesn't work much with the database specifically but will sometimes pull a few counts for specific clients or propositions.		There needs to be an ability to deliver effective political messaging, because that affects everybody. If anything, data usage allows people to not waste their time talking to someone that thinks that they are a heretic, speeds up and makes the process more efficient. Mentioned 2018 midterm as having some of the highest turnout, need to recognize the correlation. Can't say that usage is going to dissuade people from voting.
Subject 8	They say they have slow days when things are working properly. Interviewee pushes data from internal to the client side. Also acts as a catch all, manages the company API. Deals with everything consumer data. Completes some voter file cleansing with addresses		There is no reason to waste time talking to people who will just never bite. The flipside of this is the "well if they hadn't voted..."—well who are you to say that. It is ultimately in the voter's decision; people can hear something all day and still chose something different. It's a "free will argument"—insulting to someone's free will. Its the voter's decision.
Subject 9	Updates the national voter file. Sends out reports covering the changes in the voter file. Fills requests from external regarding visualizations. Makes data more readable and understandable—not everyone understands tables right off the bat.	"the goal is to win"	But the goal is to get as many people to vote as they can, with the data that they have. Yes skewing, but that's what happens when you push someone to vote that doesn't usually.
Subject 10	Creating polling samples for polling clients. Creating data counts for those clients. Filling email appends requests (ex: a list of emails that came to a campaign event).		

(III) Case study STATA do-file

Ohio case study Model 1 data cleansing and analysis:

```
1 /*2016:no general vote but a primary vote, out of all the people that voted in the primary and/or general*/
2
3 use "/Users/raganjones/Downloads/Ohio1and2and3and4.dta"
4 drop if voter_status=="CONFIRMATION"
5
6 encode general11082016, generate(general11082016num)
7 replace general11082016num=0 if general11082016num==.
8
9 encode primary03152016, generate(primary03152016num)
10 replace primary03152016num=0 if primary03152016num==.
11
12 gen nogenvote2016=1 if primary03152016num>=1 & general11082016num==0
13 replace nogenvote2016=0 if primary03152016num>=1 & general11082016num==1
14 drop if nogenvote2016==.
15
16 sum nogenvote2016
17 tabstat nogenvote2016, by(county_number) stat (n mean sd)
18
19
```

Ohio case study Model 2 data cleansing and analysis:

```
21 /*2008 and 2012 vs. 2016 */
22 clear
23 use "/Users/raganjones/Downloads/Ohio1and2and3and4.dta"
24 drop if voter_status=="CONFIRMATION"
25
26 encode general11042008, generate(general11042008num)
27 replace general11042008num=0 if general11042008num==.
28
29 encode general11062012, generate(general11062012num)
30 replace general11062012num=0 if general11062012num==.
31
32 encode general11082016, generate(general11082016num)
33 replace general11082016num=0 if general11082016num==.
34
35 encode primary03042008, generate(primary03042008num)
36 replace primary03042008num=0 if primary03042008num==.
37
38 encode primary03062012, generate(primary03062012num)
39 replace primary03062012num=0 if primary03062012num==.
40
41 encode primary03152016, generate(primary03152016num)
42 replace primary03152016num=0 if primary03152016num==.
43
44 /*generating primary dummy variables*/
45 generate rvote2016prim=1 if primary03152016num==4
46 replace rvote2016=0 if primary03152016num!=0 & rvote2016prim!=1
47
48 generate dvote2008prim=1 if primary03042008num==2
49 replace dvote2008prim=0 if primary03042008num!=0 & dvote2008prim!=1
50
51 generate dvote2012prim=1 if primary03062012num==2
52 replace dvote2012prim=0 if primary03062012num!=0 & dvote2012prim!=1
53
54 generate dvote20_prim=1 if dvote2008prim==1 | dvote2012prim==1
55
56 generate voteswitch=1 if rvote2016prim==1 & dvote20_prim==1
57 replace voteswitch=0 if general11082016num==1 & voteswitch==.
58 replace voteswitch=. if general11082016num==0 & voteswitch==1
59
60 tabulate voteswitch
61
62 encode party_affiliation, generate(party_affiliation_num)
63 generate dpartyaffil=1 if party_affiliation_num==1
64 replace dpartyaffil=0 if party_affiliation_num!=1
65 generate rpartyaffil=1 if party_affiliation_num==4
66 replace rpartyaffil=0 if party_affiliation_num!=4
67 generate nopartyaffil=1 if party_affiliation_num==.
68 replace nopartyaffil=0 if party_affiliation_num!=.
69
70 tab2 voteswitch nopartyaffil, row nofreq
71 tab2 voteswitch dpartyaffil, row nofreq
72 tab2 voteswitch rpartyaffil, row nofreq
73
74 tabulate rvote2016prim dvote20_prim
75
```