January 2024

DEMOCRACY & POWER
INNOVATION FUND

Surfacing Missing Voters:

Addressing Data Systems, Tools and Engagement Models that Invisibilize Black and Brown Communities

Written by: Miriam McKinney Gray

Data visualization support by Tianyi Hu, editing support by Joy Cushman and Elizabeth McKenna, and invaluable contributions from the Democracy and Power Innovation Fund Organizing Data Lab's directors and managers

Contents:

Meet the Author

Miriam McKinney Gray, Founder and CEO of McKinney Gray Analytics, is an inquisitive data and research professional who studied psychology for her Bachelor of Arts degree at Loyola University Maryland, and completed the Quantitative Methods for Social Sciences Master of Arts degree at Columbia University. Miriam worked as a research data analyst at Johns Hopkins University, including collecting data for the now archived world-renowned Johns Hopkins University Coronavirus Resource Center. Miriam has assisted researchers at Columbia University, Howard University and Morgan State University, and created a new research and evaluation framework for The Kennedy Center. She has also authored articles for major publications on the topics of data science, algorithmic bias and politics.

McKinney Gray Analytics LLC is a research and data analytics firm that aligns itself with socially impactful data and research projects. For the last four years Miriam has worked with state-based power-building groups as Senior Data Analyst with the Democracy and Power Innovation Fund, working with innovators from all over the country. In that role she has spent thousands of hours collecting data, co-creating analyses, and reflecting on results with state-based data directors, analysts, and researchers. It is through interfacing with these practitioners in the field that Miriam has come to deeply understand both the urgent challenges and the emerging solutions outlined in this memo.

Executive Summary

According to my estimates, based on US Census data and on a recent Stanford study, 24.76 million Black and Latino eligible voters are currently missing or listed with incorrect information in voter databases sold by vendors, making them effectively unreachable. While 40 percent of Black and Latino people are invisible to voter outreach efforts, only 18 percent of white people are missing or mislisted (Jackman & Spahn, 2021).[1]

Finding and engaging these millions of missing Black and Latino voters will almost certainly determine the outcome of consequential elections in the coming decade. For example, in 2020 the Presidential margin of victory in Arizona was 10,457, in Georgia it was 11,779, and it was 20,682 in Wisconsin. Many US Senate and Congressional races, gubernatorial, as well as local and state races have been decided by similarly small margins.

So why is this happening and what can we do about it?

Black and brown individuals are being systematically sidelined by the seemingly-inclusive, data-driven, business-backed digital systems of voter engagement that currently exist in the United States — a phenomenon that, while existing within an abstract digital reality, is an artifact of and has the ability to deeply affect our physical reality and our future. State-provided voter files omit Black and brown voters at disproportionate rates because of aggressive voter purges. These issues are exacerbated by biases in vendor-generated models commonly appended to the voter file for targeting our movement's voter contact efforts.

The mere *ability to access* full, highly representative data–meaning data that reflect a community successfully enough that one can glean useful insights about or reach those within said community–is completely necessary for our shared democracy. We cannot systematically and perpetually ignore some eligible voters and engage with others indefinitely: doing so creates a reinforcing cycle that is nearly impossible to break. For example: even if you have voted consistently in the past, if you have recently moved or missed an election and are missing from a key voter file database, you may not be contacted to vote. You may miss another election, which then leads you to be assigned a low vote propensity score if you register to vote. When you have a low propensity score, it is less likely that you will be contacted by voter outreach campaigns. How, then, might you learn

[1] A summary of this current memo was published in The Hill in June 2023.

about an election, go vote, or be added to a voter file in the future? With targeted list-based voter contact methods alone, it is highly unlikely that we would ever reach a person who has been invisible to our data systems.

> *Incredibly, given the scale of the problem, there seems to be little awareness or alarm yet among political operatives and vendors. My intent with this memo is to raise the alarm, to identify some of the underlying problems, as well as solutions being generated through state-based programs in recent cycles.*

The concepts that I introduce and dissect at length in this report are: voter files, voter file-based models (such as vote propensity scoring, predictive algorithms for race and voter file "matching"), and related, widely-used data tools. I conclude with several propositions for future civic engagement work and investment strategy.

Philanthropic investments directed at surfacing missing voters will be necessary to increase community-based data collection, support organizational engagement in antiracist modeling, and bolster efforts to build better community-based strategies, including community-led ownership of data. Simultaneously, organizers and voter outreach organizations must critically examine the models that are intentionally marketed towards ourselves and our teams, understand caveats present within certain datasets, and lean into broad relational organizing rather than targeted voter contact alone.

> *As our lives continue to skew toward virtual and data-driven reality (featuring remote canvassing, virtual organizing meetings, algorithm-based decision-making, and continued, intentional data analysis for better program outcomes), it becomes more pressing to address data systems that have proven to be biased against Black and brown people.*

We may want algorithms and data systems to simply tell us who is most likely to vote, with whom organizations should be engaging, and what races people identify as, but we first need better data and fairer methods. Currently, voter outreach organizations are relying on data that are inaccurate and incomplete, voter file matching systems that fail or are unable to draw accurate conclusions about large swaths of people, and data tools that are exclusive and error-prone.

If not resolved, the sustained, systematic, effective erasure of Black and brown individuals from virtualized voter engagement systems will continue. Please join me in deeply understanding and working to ameliorate these issues.

The voter file:
A foundation

The voter file: A foundation

What is a *voter file*?

A voter file is a purchasable record of currently registered or previous voters in a given state.

A voter file database is a national data set that convenes the records across states for the purpose of understanding United States voters (DeSilver, 2018). Contact information including address and phone number given at the point of registration are included in a voter file. So too are pertinent demographics information, including but not limited to race, gender, and age. Other potentially useful information for determining overall voter status, such as vote history and voter registration status are commonly included. States differ as to which data are required when an individual registers to vote. Non-governmental variables will sometimes be appended to a voter file by vendors to fill in those gaps, such as data from credit bureaus and political organizations. Voter files can be purchased directly from Secretaries of State Offices, or through vendors, by people or organizations whose efforts center around voter registration or voter targeting.

A voter file is an official government record, whereas marketed voter file-based technology is not.

It is important to note the distinction between official voter files provided at the state level and supplements provided by vendors, including models and consumer data. Data and analytics firms profit from the algorithms and tools they provide to augment and manage voter file data.

Voter file technology is marketed by several firms as a large dataset that, in theory, gives users a chance to better understand voters and potential voters. And yet, try as we might, due to many factors, there just is no perfect government-sanctioned artifact that allows us to fully understand voters or potential voters in the United States.

There are several other forms of government or government-adjacent information collected that are conceptually similar to a voter file. For example, the US Census also applies the concept of counting or collecting information about people for the purpose of research or political engagement. Yet, Census counts frequently miss people with lower incomes, renters and those who may experience housing instability, those who speak languages other than English, those lacking internet, and those who distrust or fear the government and therefore choose not to actively participate in government counting techniques due to concerns about identification (Elliott et al., 2019; Kramer, 2019; Stanford, 2020; Vines, 2018).

In a recent report released by the US Census Bureau related to the 2020 Census, an overcount of white and Asian American people was present while a staggering undercount of other minority groups, Black, Hispanic, and Native American residents specifically, was also present. “The Black population in the 2020 census had an estimated net undercount of 3.3%, and the gap was almost 5% for Hispanics and 5.6% for Alaska Natives and Native Americans living on reservations. The non-Hispanic white population’s net overcount was 1.6%, and Asians had a net overcount of 2.6%.” By comparison, in the earlier 2010 census, “the Black population had a net undercount of more than 2%, and it was 1.5% for the Hispanic population. There was a nearly 4.9% undercount for Alaskan Natives and Native Americans living on reservations, and an undercount of 0.08% for Asians. The non-Hispanic white population had a net overcount of 0.8%.” (Schneider, 2022). A very real consequence of this missingness in the Census, as further detailed by researchers and professors G. Cristina Mora and Julie A. Dowling, is that this 5% underrepresentation could translate into at least $3 billion in lost funding for towns and cities with larger populations of Black and brown communities.

Similarly, a voter file, while widely-used by political strategists and organizations nationwide, is built on incomplete and sometimes biased data for targeting voters and their communities. Crucially: **state voter files, much like the Census, have blind spots and inaccuracies, and these have heavy consequences**. While we have documentation about Census undercounts–the Census Bureau systematically releases reports estimating their undercounts–voter file missingness is a topic much less referenced in existing research. Furthermore, Secretaries of State are not currently accountable for reporting missing or inaccurate data in their files.

What is *voter file missingness*?

Voter file missingness is the complete lack of representation of a living voting-age person or people within a voter file.

We have seen, through working with the data and through conversations reflecting organizations’ lived experiences, that competitive, commonly-recommended voter files actually lack representation – especially for many people of color. As Drew DeSilver and his colleagues at Pew explain,

> *“When voter files first came to prominence for political practitioners and researchers, many were just what the name suggests – lists of registered voters. But as use of voter files for research and targeting has become more widespread, most vendors have tried to cover all US adults, including those who aren’t registered to vote. Because the core component of the files is a combination of official state lists of registered voters,*

> *vendors have sought out commercial databases – available from sources such as credit rating agencies – to locate Americans missing from state voter rolls" (DeSilver, 2018).*

These commercially appended datasets can be useful for studying community level differences, but are problematic when we depend on them to reach the millions of eligible voters who are not accurately represented in the voter file.

> **"People of color are considerably less likely to be correctly registered than whites. Among Blacks and Hispanics, about 20% were unlisted and another 20% were mislisted." The result: "Two in five Blacks and Hispanics are unreachable" with voter file data alone (Jackman & Spahn, 2021).**

Missingness within the voter file is analogous to the Census' well-documented undercounts of people living in the United States. Some people are just missing from the count, regardless of the coordinated people-location tactics implemented by vendors.

According to Simon Jackman and Bradley Spahn in their recent study of voter file missingness: "At least 11% of the adult citizenry is unlisted. An additional 12% is mislisted. These groups are invisible to list-based campaigns and research, making them difficult or impossible to contact" (Jackman & Spahn, 2021). Furthermore, "people of color are considerably less likely to be correctly registered than whites. Among Blacks and Hispanics, about 20% were unlisted and another 20% were mislisted." The result: "Two in five Blacks and Hispanics are unreachable" with voter file data alone (Jackman & Spahn, 2021).

When we match this reality to Census data about voting age people, nearly 25 million Black and brown eligible voters are missing from commonly used voter file databases. Data from the 2021 ACS 1-Year Estimates Subject Tables provided by the Census shows 29.16 million Black citizens over the age of 18, and 32.74 million Latino citizens over the age of 18, for a total of 61.9 million Black and Latino voting age citizen voters. If 40% of them are missing or mislisted, **that is 24,761,762 missing Black and Latino voters.** Due to the very public fact that the US Census Bureau routinely undercounts Black and brown communities, even this number is likely a low portion of the people who are effectively invisible to the multi-billion dollar political industry.

Nearly half of all Black and brown eligible voters are not accurately or proportionally included in voter files. Of course, this reality snowballs as it informs whom organizations contact in voter engagement efforts. In 2020, as she began to consider the impact of this on her own work, PICO California Data Organizing Manager Lisa Thornton remarked, “For those who aren’t matched to the voter file, we do not contact them. We’re not contacting those folks for our Get Out The Vote work. And those people are not being reached by other campaigns either.”

Pew Research confirms this. In 2020 half of non-voters said they did not receive any or only one contact from a campaign in the weeks leading up to the November election. By comparison, those with the highest level of political activity received the highest volume of campaign outreach: 80% of highly active voters reported three or more contacts in the same time period (Daniller & Gilberstadt, 2020). Similarly, an Equis survey with DPI in 2022 found that “67% of Latino non voters said they had no contact about the election in 2020” (Equis Research, 2022).

As it stands, **we cannot encourage organizations to rely solely on the voter file to engage their members or communities in civic engagement because it is not adequate alone.** We need to architect more inclusive data that actually reflect the constituencies we want and need to organize. Many of these issues are related: by taking purposeful steps towards inclusion, we may also be supporting the creation of voter file-based models that are more accurate for all.

Who is responsible for issues related to the voter file?

Campaigns, philanthropists, companies, and those paying for voter files or voter file-based supplements have a responsibility to demand the amelioration of voter file issues by their vendors.

As mentioned above, state-level voter files are distinct from corporate-level, marketed tools that utilize voter file data. Responsibility for voter file-related issues falls on both state governments, who create these files and make them available to the public, and on political and data vendors, who build upon these files and sell their products to a wide variety of audiences.

At the state government level, voter missingness is exacerbated by overzealous and often politically motivated voter purges in states. In many states, even if a person has registered to vote, they are subject to removal if they remain inactive for too long. The definition of “too long” differs from state to state. When a resident of Ohio has not voted in the past six years, that person is removed from state voter rolls. When a resident of Oklahoma has not voted in the past two elections, that person is removed from voter rolls. When a resident of California is inactive, that person’s voter registration is removed (Valisolgambros, 2019; Smith, 2020). This and other, similar, practices result “in the deletion of hundreds of thousands of registrants each year. Very often, those people get energized to vote in a given election but find when they show up at the polls that they are no longer registered and cannot cast a ballot" (Smith, 2020). And although researchers have suggested that "there is no justification for using voter inactivity as an independent basis for eliminating registrations" (Smith, 2020), it still happens each year.

From November 2020 through July 2021 alone, over 8.6 million voters were purged by counties and states across the country. Data vendor TargetSmart found that 63 of the 73 counties with highest purge rates disproportionately purged voters of color (Garland, 2021). Addressing these problems will require changes at the state and federal level, as well as significant oversight. Civil and voting rights lawyers and advocates are clearly aware of and working on these challenges.

However, just as problematic, but much less widely understood, is the role that virtual redlining plays in further putting Black and brown eligible voters out of reach of the robust network of voter contact programs in our country. In order to keep people on voter lists, especially as voting laws continue to become more restrictive for many, independent organizations and campaigns must remind them to vote (Timm, 2021). We must inform them when to vote, how the process of voting works, and where they can go. Realistically, unless a person has the intense personal motivation and spare time to remain up-to-date on all state and local elections, traditional or community-based methods of voter contact will help them remain aware of upcoming civic duties. However, these reminders cannot happen efficiently or effectively when community organizations are working with biased models and error-prone tools. In this memo I focus on what can be done in the near term in the private and nonprofit sectors to ameliorate voter file missingness.

What are the most pressing voter file-related issues?

What are the most pressing voter file-related issues?

Issue 1: Relying primarily on algorithmic modeling as a method for targeting voter engagement is problematic

First, what is an *algorithm*?

An algorithm is a curated function, heavily based in mathematical concept, that uses data points to create interpretable results.

An *algorithm,* in its simplest form, is a process or a set of rules to be followed within a mathematical calculation. Inputs are data points that are fed into the algorithm and outputs are the results of said algorithm. Specifically, *machine-learning algorithms* are often implemented by computer or data scientists to analyze massive amounts of data in a short span of time. More and more often, algorithms and their outputs are used to make choices for us – with or without our knowledge of the process. Today, computer-based algorithms are quietly applied for various decisions in almost every facet of mainstream American life: for example, what we are shown on social media on any given day, which advertisement is suggested for your child, how your partner's credit score is computed.

Some machine-learning algorithms can be used to predict future behavior based on past data. As a society, we sometimes use predictive algorithms to make what we assume to be accurate guesses about people and their probable behaviors. We apply these generated predictions to make other, sometimes far-reaching decisions.

A predictive algorithm can be as harmless as guessing what show you might enjoy next on your streaming service. However, it can be as potentially harmful as forecasting whether or not a certain incarcerated person will commit another crime in the future and thus should or should not receive the option of bail from a judge. Since the inception of this type of application, countless researchers have publicly critiqued the problematic application of algorithms due to well-documented inequities (Angwin et al., 2016). **Of greatest concern are algorithms that are imperfectly designed and simultaneously have significant real life consequences**.

What happens when algorithms are infused with racial bias?

Predictive policing, facial recognition, and racial identity algorithms are all products of our imperfect scientific methods and racist realities.

In recent years academics have begun to study the insidious ways racial bias impacts algorithmic design and outcomes. Two examples of many are the *COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)* predictive policing algorithm, and present-day facial recognition technology.

COMPAS was developed to predict the likelihood that a defendant will re-offend in the future. As far as underlying assumptions, defendant responses to an 137-item questionnaire are weighted to produce a score. Results from this algorithm have been used to determine whether or not a person should remain incarcerated or receive bail, and model designers claim that these predictions will help to curb violent crime. In 2016, Julia Angwin and her ProPublica colleagues analyzed COMPAS. They found that the algorithm was biased against Black people in that, "Blacks are almost twice as likely as whites to be labeled higher risk but not actually re-offend… And COMPAS makes the opposite mistake among whites: They are much more likely than Blacks to be labeled lower-risk but go on to commit other crimes" (Yong, 2018). We should take issue with the underlying data points and mathematical assumptions made in this model, both of which lead to profound biases towards Black people.

Researcher Ben Green asserts that we must study issues of algorithmic unfairness more explicitly. To fully understand why an algorithm is much more biased than we might think, we should always consider the full political context informing a predictive algorithm. All algorithms are developed in contextualized, political societies whose data points are registered, collected and weighed with biases common to that culture. For example, "Predictive policing algorithms offer a particularly pointed example of how striving to remain neutral entrenches and legitimizes existing political conditions. The issue is not simply that the training data behind predictive policing algorithms are biased due to a history of overenforcement in minority neighborhoods. In addition, our very definitions of crime and how to address it are the product of racist and classist historical processes… Predictive policing represents a notably salient example of how data science cannot be neutral" (Green, 2021).

Present-day facial recognition technology is another machine-learning-based application (probably) developed with good intentions. However, because it is employed for uses as serious as "law enforcement surveillance, airport passenger screening, and employment and housing decisions" (Najibi, 2020), we have a duty to critically dissect its underlying data points. Inequality in face recognition algorithms has been documented by many researchers over recent years. "A growing body of research exposes divergent error rates across demographic groups, with the poorest accuracy consistently found in subjects who are female, Black, and 18-30 years old" (Najibi, 2020). Essentially: Research suggests that current facial recognition is biased, particularly against darker skinned young women. One suggested solution: "algorithms can train on diverse and representative datasets, as standard training datasets are predominantly white and male" (Najibi, 2020). Still, even if the underlying mathematical calculations of the algorithm are altered and training involves new, more inclusive data, the problematic uses of facial recognition remain.

A decision as generationally impactful as whether or not someone will remain imprisoned, or how police resources should be allocated should never be so profoundly influenced by an inanimate, scientifically-generated prediction, and yet it is. Below I argue that predicting how likely a person is to vote or assigning a person's race are similarly consequential, and, like COMPAS, policing algorithms and facial recognition, the consequences of using biased voter targeting models will potentially affect generations to come.

Now, what is *vote propensity*?

Vote propensity is an algorithm-based prediction of how likely a person is to vote. *Vote propensity* scoring is the output of an algorithm that produces a numeric score for an individual. The inputs include various data points surrounding voting history, demographic makeup, neighborhood and other factors intended to predict how likely this individual is to vote in an upcoming election. The output is one final numeric score between 0 and 100. In theory, assuming that a vote propensity scoring algorithm considers "vote history" to be a crucial input factor: a person who has voted consistently in all previous elections might receive a vote propensity score of 95, whereas a person who has never voted in the past might receive a vote propensity score of 10 or less.

A vote propensity score is often used as a "threshold" or arbitrary cut-off point by campaigns to determine whom to contact and whom to skip over in voter engagement programs. For example: Maria, a strapped-for-time field director, might choose to contact only individuals with scores of 60-100 because these folks have been algorithmically determined "most likely to vote" and thus assumed to be a "good investment of her time." Maria might simultaneously decide to ignore individuals with scores of 0-59 because (to her understanding, based on the algorithm) those contacts would not result in action. Maria might not know that, due to all the issues related to this scoring technique, younger people often receive lower vote propensity scores than older people because there is less historical voting data for them. Or that the same is true of renters and others who move frequently. If

Maria chooses to ignore individuals in the 0-59 vote propensity score range, she effectively and unknowingly ignores many of her potential voters who are simply mis-categorized by the algorithm as a result of inadequate input data.

Why might vote propensity scoring invisibilize voters?

Due to proprietary software restrictions, we cannot fully know the answer to this question; however, post-program analyses show that underlying data points and assumptions of popular vote propensity scoring algorithms are flawed.

"Data science exists on a political landscape. Whether articulated by their developers or not, machine-learning systems already embed political stances. Overlooking this reality merely allows these political judgments to pass without scrutiny, in turn granting data science systems with more credence and legitimacy than they deserve" (Green, 2021).

Like any other algorithm created by humans with data derived *from* humans, vote propensity scoring has bias. Firstly, an algorithm simply cannot be fair if the data used in its training are not representative. Statistically speaking, the algorithm will always hold some sort of bias because the data used to train it are steeped in context. Even if, in a perfect world, the data were fully representative of all people, some bias would remain because all data emerges from complex sociopolitical, historically contextualized moments. For example: It is now well established that medical clinical trials continue to consistently under-include non-white participants and women as subjects. As a result, our current level of medical knowledge is biased toward the experiences of white men. (Kelsey et al., 2022; Fultinaviciute, 2022).

Similarly in politics the context must be known and understood to treat data points with the thoughtfulness necessary for a fair algorithm today. As Ben Green argues, "Data science exists on a political landscape. Whether articulated by their developers or not, machine-learning systems already embed political stances. Overlooking this reality merely allows these political judgments to pass without scrutiny, in turn granting data science systems with more credence and legitimacy than they deserve" (Green, 2021).

Secondly, we do not know enough about the creation of any proprietary algorithm, according to currently implemented trade laws within the United States, as they are regularly considered assets of intellectual property. Although "We must ask ourselves whether the underlying math of the algorithm is appropriate and if the type of model has been chosen correctly" (Rentsch et al., 2019), if we cannot see the math underlying an algorithm it is difficult to judge its accuracy or bias. Given the widespread use of vote propensity models, it would be useful to know exactly how creators of this type of scoring weigh certain data points as opposed to others, what techniques were used, and so on. However, we simply do not currently have access to that detailed information. Instead, we have our rich and detailed lived experiences over multiple election cycles. The simple fact is: this method of scoring has proved time and again to be highly accurate in its predictions for some, but highly inaccurate for others.

In 2019, Faith in Action (a national community-organizing network) produced civic engagement research highlighting this very issue. They argue that "vote propensity scores redline neighborhoods out of political outreach based on bad data" and suggest that "we and other groups like us should target all the way to 0, and knock every door in some communities, early in our voter outreach efforts" (Faith in Action, 2019). They also argue that scores have proven to be "highly accurate for high propensity voters, but very inaccurate for mid and low propensity voters" (Faith in Action, 2019).

Table 1: The Difference between Predicted Votes and Actual Votes (Faith in Action, 2019)

Nevada General Election: Contacted Voters and Net Votes Over Propensity Score Expectations				
Vote Propensity (Catalist score)	Number of People Contacted	Should Have Turned Out According to Score	Actual Turnout	Net Votes Beyond Score Expectations
0 to 9.99	2,141	107	748	641
10 to 19.99	3,258	489	1,204	715
20 to 29.99	1,159	290	591	301
30 to 39.99	2,507	877	1,738	861
40 to 49.99	1,747	786	1,351	565
50 to 59.99	1,085	597	891	294
60 to 69.99	1,171	761	1,036	275
70 to 79.99	650	488	570	82
80 to 89.99	650	553	607	54
90+	847	805	828	23
Unknown	68		10	10
Total	15,283	5,753	9,574	3,821

Table 1, showing figures from Faith in Action's 2018 Nevada program data, illustrates the very issue I invite you to critically assess. Even though national data vendors recommended that Faith in Action organizers reach out mainly to individuals with high propensity scores, Faith in Action expanded their targeting to include individuals with lower vote propensity scores. Due to this choice, they saw that individuals with 0 to 69.99 vote propensity scores gave the highest return on their efforts, as far as voter contact and turnout. In fact, 89% of the net votes beyond what the propensity model would have predicted came from those with scores under 60 - precisely those voters most campaigns would ignore.

It is also the case that the added value of an organizing or field program may be in the lower half of the voter propensity scale with less frequent voters, and not the upper half, where many campaigns are advised to target.

From this and other similar cases shared within the DPI data community, it is clear that a vote propensity score is not a fully accurate prediction of future voting behavior. It is also the case that the added value of an organizing or field program may be in the lower half of the voter propensity scale with less frequent voters, and not the upper half, where many campaigns are advised to target.

As mentioned earlier, there are two main issues that likely feed into the bias of vote propensity scoring: lack of representation in data and inadequate training of the model. One of the major reasons for inaccurate scoring on vote propensity is likely to be that the model excludes (whether intentionally, unknowingly, or both) certain segments of the population. If we return to the fact that 40% of Black and brown voters do not show up at all or show up inaccurately on the voter file, that is a significantly biased foundation from which to build vote propensity models. We suspect that, if this type of model were able to include all data points related to those excluded, it would more accurately score these people, since **accurate data representation is a necessary component for algorithm reliability**.

In addition to the representativeness of input data sources, the methodology behind this kind of model is also a challenge. For example: If the creators of a specific brand of vote propensity score decide to *impute* missing data — that is, assign a value based on inference or prediction as opposed to actual observation — they introduce additional bias to the model, because data imputation in itself is statistically biased. Similarly, the weighting of variables in the model introduces human bias. So too, will other assumptions a creator of a vote propensity model might consider, for example:

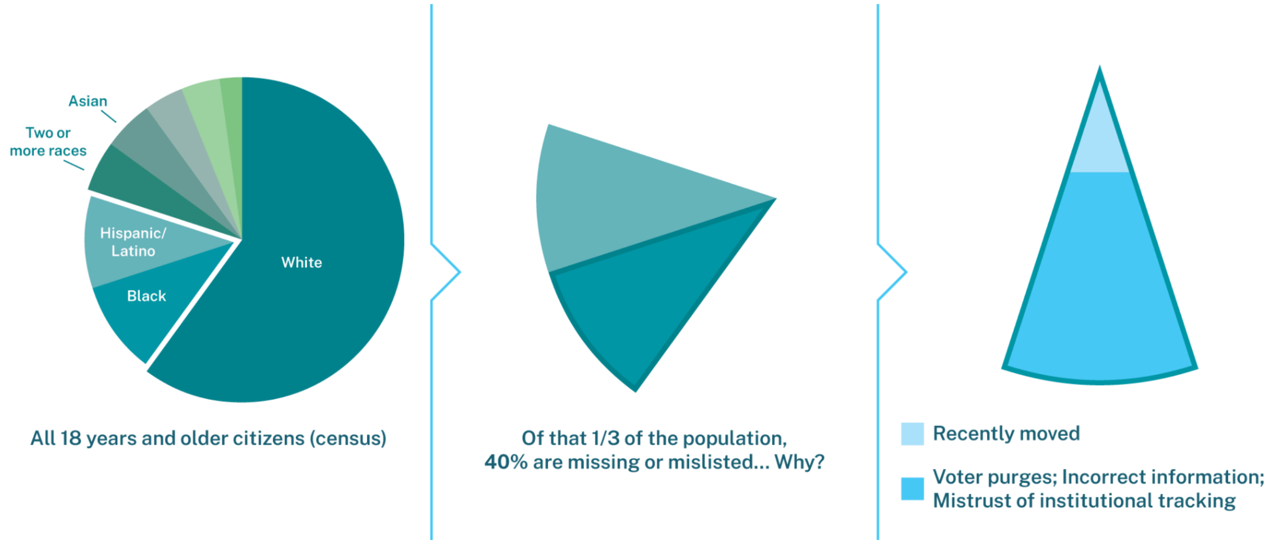
- How is frequent moving (and removal from the voter file) tracked and weighted?
- Are people who live in apartment complexes and who may move more often weighted differently than those who live in single-family homes?
- If a person lives in an area where the majority of residents have not voted in recent elections, how is that weighted?
- Is “last name” a proxy for race, such that ethnic-sounding last names are assumed Black and brown while those that sound white are not? How does that inform these models?

- How can we trust a race variable if it is at all based on large assumptions and may be subject to change over time (for example, Person A is deemed "Black" at Time 1 but then deemed "Caucasian" at Time 2)?
- In the end, how might all of these circumstances affect one's vote propensity score?

In thinking about fixing problems related to vote propensity score, it would be useful to require model vendors to answer transparently questions such as these.

Why is it so crucial to understand vote propensity?

The potential negative ramifications of vote propensity scoring are wide-reaching.



As mentioned earlier, issues with voter models are magnified when people or organizations actually act upon the model outputs. A recently-publicized DNC voter file model demonstrates just how critical it is to understand methodologies because their related models determine which people volunteers will contact (Turman, 2020). If an organization decides to only reach out to those with a propensity score of 70+, and volunteers agree to *only* contact those high score individuals, then those with numeric scores of 0-69 will be ignored. That logic is already unfair and self-perpetuating. If campaigns fail to contact low propensity voters, those voters will likely remain in that "low propensity range" as they are systemically not contacted. And what if the model's projection was just plain *wrong*? What if a person who receives a propensity score of 20 is actually highly likely to vote and, in reality, should have received a score of 90 by the model?

After the general election in 2020, Florida Rising created several visualizations based on outreach data to help them better understand better how their voter contact attempts were actually targeted. The time series graph shown below is based on data about the organization's total number of voter outreach attempts. Figure 1 shows that the organization's contact attempts were most heavily skewed towards the high vote propensity range.

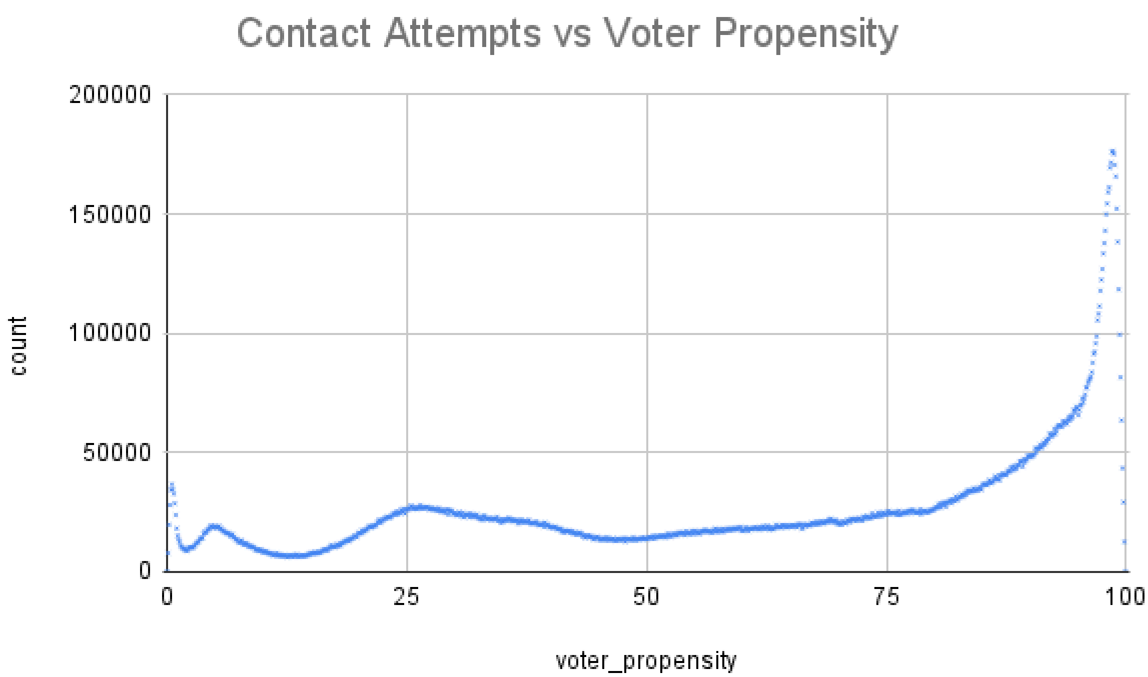


Figure 1: Florida Rising voting contact attempts by vote propensity (2020).

It is also incumbent on state data tables to take a critical look at the impact of their targeting decisions and how they may inadvertently perpetuate racial, economic or age exclusion with overuse of voter file appended models.

The data underlying the visualization above represents all call and text attempts. Each point on the graph represents a vote propensity score and how many times the person with that score was contacted by phone or text. It is skewed to the left and shows a long tail — only 13% of all attempts were made to contact voters with a vote propensity score below 25, whereas 50% of all attempts were directed at voters with a score of 75 or above. The implications are that the people assigned lower vote propensity scores were less likely to be encouraged to vote, at least by this particular program, which may lead them to be in fact less likely to vote and less likely to be contacted for future elections.

DEMOCRACY & POWER INNOVATION FUND

This analysis is not to say it was Florida Rising's *intent* to reach out to mostly high propensity voters. In fact, their intent was to reach primarily Black and brown voters who would not otherwise be contacted. However, in a movement data culture where models and over-targeting are the norm, it requires vigilance and funded data capacity on the part of organizing and voter outreach groups to regularly monitor their contacts and ensure they are reaching the people they most intend to reach. It is also incumbent on state data tables to take a critical look at the impact of their targeting decisions and how they may inadvertently perpetuate racial, economic or age exclusion with overuse of voter file appended models.

In the last four years, I have seen that this issue is pervasive across the country. In a November 2020 Organizing Data Lab call with 16 national power-building organization data managers and leaders, the topic of vote propensity scoring surfaced. Christina Marikos of LUCHA, a large Arizona-based organization with a predominantly young Latino base, pointed out how "good yet frustrating… it is to hear that so many other states are having the same issues." Another DPI data leader then said, "every time we dig into the vote propensity score logic, I have more to consider about who we are actually filtering out of our pools."

It is critical to inform organizing and voter outreach groups about data bias issues. We encourage groups to build or request universes of voters that have a diverse range of propensity scores or even lower vote propensity scores to focus where other party and candidate campaigns may not.

Missing voters, or those lacking representation within our systems of voter engagement, will reap long-term consequences from potentially inaccurate or harmful vote propensity scoring. These potential voters will continue to be overlooked and unseen by (perhaps unaware) organizers, canvassers and campaigns who lean on algorithm-based scoring to decide which people to contact about elections.

Compounded over time, the effects of these algorithm-informed decisions may mean that the benefits of an ever-growing political industry will accrue more and more to frequent (more often white and upper class) voters, creating a permanent underclass of infrequent voters.

What is the evidence of bias in other voter file models?

Another type of civic engagement model that surfaces similar problems is *race modeling*: algorithms that project a person's race when self-reported data are not available. Very few states require that eligible voters report their race at point of registration, and the majority of those are in the South. As a result, data vendors have built race models that can be appended to the voter file, and sell this information to campaigns and organizations to assist with their voter contact targeting.

But imagine: In 21st century America, how might you train a model to assume a person is Black versus white, or Latino versus Native? What tips the scale of Blackness toward whiteness? How might the factors you are considering be problematic?

The problem is: race modeling is often inaccurate. Still, many organizations whose missions revolve around reaching otherwise excluded voters rely on these models to project voters' racial identities. But imagine: In 21st century America, how might you train a model to assume a person is Black versus white, or Latino versus Native? What tips the scale of Blackness toward whiteness? How might the factors you are considering be problematic?

In 2022, a company called Zest AI released its nascent "Zest Race Predictor." This machine-learning algorithm "estimates the race/ethnicity of an individual using only their full name and home address as inputs" (Zest AI, 2022). Although this company's honesty is refreshing given the wider landscape of "we cannot share this information" within the political technology space, Zest AI's assumption that a societal construct as complex and deeply personal as race can be predicted by only two factors is clearly biased. The underlying stereotypes informing this assumption are, in fact, the problem I am addressing here.

These types of business decisions, commonly-shared stereotypes, data biases, and remnants of covertly racist state decisions all combine to establish a set of widely used models that result in further exclusion for none-the-wiser organizations and the Black and brown people and potential voters they seek to engage.

In 2022 a large-scale community voter outreach program at the Ohio Organizing Collaborative provided a rare look into the accuracy of race models. In this program the OOC worked with 2,503 residents in Cincinnati, Columbus and Cleveland to build lists of friends and family they could reach out to in order to get them registered and turned out to vote. The lists ultimately included 66,076 people. The form they filled out asked them to write in the race of each person they were adding to their list. 81% of the people building lists identified themselves as Black. They also identified 70% of the names on their friends and family lists as Black as well. However, for the Black friends and family who matched to the voter file where the race model is applied, one third of them were wrongly predicted to be white (Gutierrez & McGarity, 2023).

The consequence of inaccurate race models is that voter outreach programs large and small who are working to find Black and brown voters are over-targeting their programs and missing outreach to millions of potential voters who are miscategorized in the voter file as the default race: white.

The result of bias in vote propensity models, race models, and other models dependent on these two is the further exclusion and invisibilization of Black and brown voters and potential voters. There are many reasons why racial bias exists, each reason likely compounded on the last. However, the systemic biases documented here stem from racially biased data systems. Because these models are proprietary and their underlying math assumptions hidden from the public, we cannot fully know all of the sources of the bias, however the documented result here is further racial exclusion.

I implore you to remember the common axiom of data managers across the country: "Garbage in, garbage out." A model will always reflect — and only be as good as — the underlying data used to train it. If there are underlying issues or assumptions made within the data or by the data system, the algorithm will be biased in the exact same way. In the case of vote propensity scoring and predictive race models, if systemic bias or racism are well-documented factors of the data and data system, we can assume the model based on these data points is similarly biased.

So what can we do about it?

> **We need to ask: "Who will be affected and how (i.e. at what scale) when this algorithm is deployed?"**

The first step to correct for algorithmic bias and the ways it exacerbates the exclusion of missing voters is to require model vendors to provide more transparency about their underlying data and the calculations and assumptions used to train their models. Model vendors could be required by donors and those with whom they hold major contracts to provide a caution to model users about the potential harmful ramifications of overdependence on their model. This might include reports on tested accuracy levels in Black and brown communities, as well as contexts for best use.

To be clear: researchers are not saying that predictive algorithms should never be used in practical applications. We are suggesting that a potential requirement of use is for model creators to outline possible unintended consequences. "You shouldn't need people like us to say: This doesn't work. You should have to prove that something works before hinging people's lives on it" (Yong, 2018).

In an algorithm-centered research project I collaborated on while working at Johns Hopkins University, one criteria we encouraged creators and users of algorithms to consider is: "who will be affected and how (i.e. at what scale) when this algorithm is deployed?" The answers to the questions found in this project, [The Ethics and Algorithms Toolkit](), help assign a certain level of response or urgency to said application.

A second critical step in our field is to train end users of models like these to think critically about their use. Our newfound digital realm of reality is a slippery slope. There are so many ways algorithms can improve lives and streamline our work flows online and in-person, and these options are often enticing. Yet we must encourage a spirit of skepticism in this field. To assume that a model's projection is inherently correct is to do a disservice to one's organization and program.

Issue 2: Voter file missingness and the futility of voter file matching disserve Black and brown power-building organizations

What is voter file matching?

Voter file matching is how an organization merges voter file information with current membership information according to shared data points across the two datasets.

Voter file matching is the process of matching a person in a membership base to an existing voter file, typically by name and address. Since voter files have the least accurate data for Black and brown people, the organizations attempting to broaden their membership in these communities, or to engage their communities in voter contact are often stymied by poor voter file match rates.

Many organizations in the DPI Organizing Lab have attempted matching their membership lists acquired through organizing and other forms of outreach to the voter file. They have consistently found that voter file data are simply incorrect or that any information on the people they need to reach is missing. These organizations have surfaced, time and time again, the futility of voter file matching. They report that their ability to match members to voter files has typically a 30% success rate or lower. In my work with DPI organizations, we have noted a match rate as low as 12% for majority Black and brown groups and 70%+ match rate for majority white groups.

The OOC's relational voter experience in 2020 confirmed this. "White voters, who are only 18%-20% of our relational program, but have an average match rate of 80%-90%," Prentiss J. Haney, co-executive director at Ohio Organizing Collaborative, once shared candidly. He continued, "Compared to Black voters, currently 76%-80% of our program, who only match at a 19%-20% rate with the voter file... You cannot increase voter turnout for voters you cannot find."

In early 2020, LUCHA could only match 14% of their membership base to the Arizona voter file. LUCHA had to create various workarounds for matching or reaching out to their missing members. If LUCHA had relied solely on platforms like NGP VAN to direct their 2020 general election outreach, they would have talked to a much smaller segment of their membership than they actually did, and those initially missing from the voter file or showing up with inaccurate contact information would have been ignored.

What are the consequences of poor voter file matching?

When people are missing from or do not match to the voter file, they become more difficult to contact.

When individuals are missing from voter file data, they become unseen by civic engagement institutions in multiple ways. It is difficult to get a clear picture of the community an organization wishes to organize without reliable data. Organizations have to work twice as hard to (only possibly) reach those invisible people. They do manual organizing and data entry work to find those not listed in voter files. Perhaps most frustrating of all, groups have to design and run separate programs to reach their members who do not match to the file, because they cannot easily integrate these people into phone or door to door canvass programs for voter contact.

These same organizations, whose capacities are often already stretched thin, need to then pivot and rally around more time- and energy-consuming voter registration and relational organizing tactics to connect with the invisibilized (yet *powerful*) voters within and adjacent to their bases. And they do all this in a dominant political culture that privileges cost efficient contacts to higher propensity voters rather than the laborious, but highly impactful work of finding and engaging missing voters.

Issue 3: Current tools for voter file management and analysis are exclusive and often unreliable

Compounding issues with model accuracy and voter file match rates are the challenges built into the current plethora of tools available to groups focused on engaging Black and brown voters. Newly developed online contact tools such as ThruTalk, ThruText, Empower, Hustle, Mobilize, and more are popular, commonly used tools in our movement because they appear to streamline workflows and automate previously arduous manual tasks. However, after using them and interfacing with others who used them as well, it is clear that each tool would benefit from many improvements.

It is the nature of technology that it is constantly adapting and requiring updates or patches, which means these tools, like so many other types of digital support, are iterative. However, when these changes happen frequently mid-program, we find that organizations working with the hardest to reach voters bear the brunt of the impact. As one former executive-level leader in a large-scale voter contact organization, who chooses to remain anonymous, remarked in frustration: “We’re not even beta testing these tools, we’re alpha testing them. Then we’re paying [these vendors] through all that.”

In a cohort-wide conversation on data tools, one data director remarked: “ThruTalk has no process integrity for mapping data into VAN[2]. Our data was missing, incorrect, or incomplete. They completely purged their archive, so it’s hard to fix later.” During that same conversation, another data manager shared, “ThruTalk’s customer service has been incredibly unreliable. They intermittently shut down their chat function entirely, even during peak calling time.” The conversation and the list went on . . . .

[2] NGP VAN is a voter file interface tool that allows organizing groups to purchase access to and use their version of states’ voter file information. NGP VAN is currently owned by Every Action, which was acquired in 2021 by private equity firm Apax Partners.

Some voter contact-related data tools currently used by organizations in the DPI Organizing Lab:

- **Every Action: the dominant tool for acquiring voter file access, and for engaging voters and members.**
- **ThruTalk and ThruText: phone banking and texting tools (respectively) that can legally call cell phones and landlines at scale and send thousands of text messages simultaneously.**
- **Empower: a relational organizing app that allows users to organize people within their personal networks to register and vote.**
- **Hustle: a text messaging distribution tool that allows users to target specific contacts with personalized messages.**
- **Mobilize: a tool for member and volunteer recruitment and management.**
- **Action Network: a tool for email outreach to and event management with members**

Not only are there issues with usability, some of these tools have proven to be more exclusive than expected for organizations working in a multiracial, cross-class movement. "These tools are built for white, college-educated men," said Lisa Thornton of PICO California. "Many are not conducive if you speak another language, are on a less expensive phone, utilize non-western character languages for communication, and so on." Several organizations have found that for many reasons their white members have been significantly more likely to adopt relational outreach tools than their Black and brown members. This is what led the Ohio Organizing Collaborative to develop their own paper list plus QR code methods for their robust, scaled relational program in 2022.

In addition to the challenges with each tool's accessibility and functionality, it is very difficult for organizations to integrate their voter contact data across tools for voters on the voter file, let alone for those invisible or missing. Carter Kalchik documented these programs in detail in his DPI Fund report last year: *Surveying the Landscape of Data Integration for State and Local Organizing Groups*.

With so much investment in so many tools marketed to power-building organizations (see Figure 2), it is reasonable to expect higher quality outputs. If the tools are not working properly, or if their data do not integrate across platforms, they serve as yet another barrier to effective voter engagement work — catalyzing yet another workaround that organizations are forced to develop just so they might find and reach their own people.

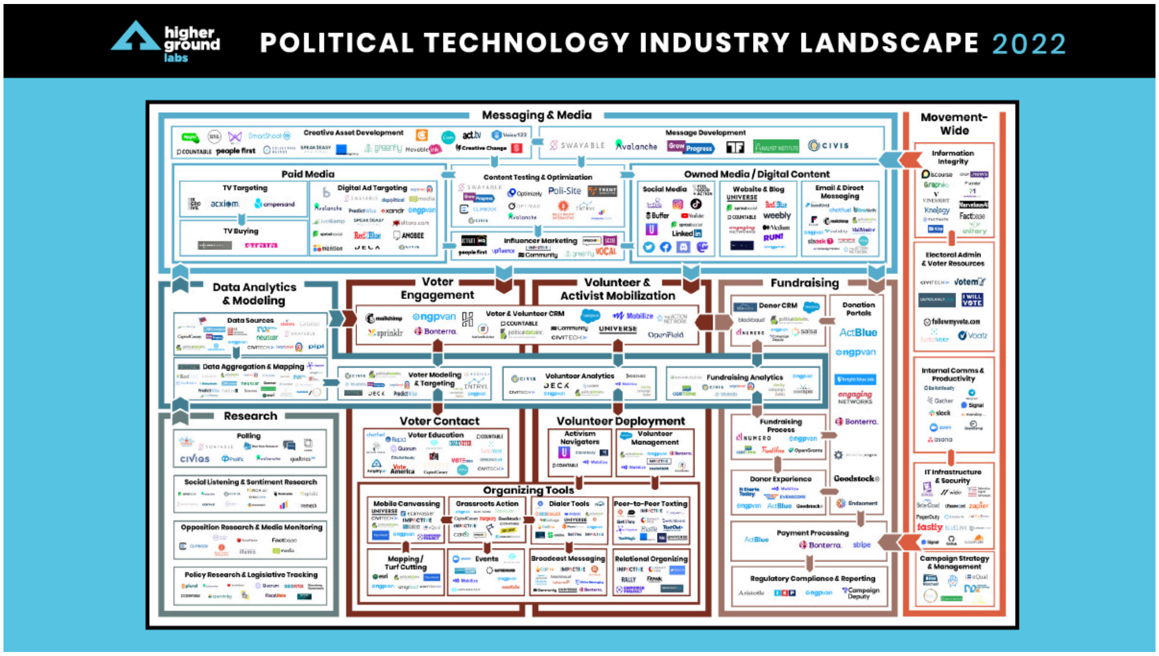


Figure 2: The huge landscape of data tools currently available to power-building organizations. (Higher Ground Labs, 2023.)

All of these issues around voter file databases, models and tools inadequacies can interact, compound, and exacerbate the inequities many communities of color already face in the political arena. **With traditional voter-file based contact methods alone, we will never be able to reach a person who was invisible to begin with. This is the paradox of the invisibilized.**

State-Based Solutions for Surfacing Missing Voters

State-Based Solutions for Surfacing Missing Voters

Many of the possible solutions for finding and engaging missing voters are being generated in real time by organizing groups in states who are committed to full representation of their communities in the democratic process.

Evidence from Arizona: LUCHA

A case from LUCHA in Arizona shows the compounding effects of the issues raised above during the 2020 election. During the summer of 2020, LUCHA canvassers reported having on their voter outreach lists significant numbers of white and often very conservative voters. When they ran an analysis of their America Votes voter contact universes in early fall, LUCHA was able to see that high propensity voters (people with scores of 70 or higher) of all races made up the majority of their universe. They also found that nearly half of their universe were white voters, even though their mission was to engage Latino, Black, Native and AAPI voters. In fact, in a state with over 300,000 Native Americans, almost none appeared in LUCHA's voter contact list. (See figure 3).

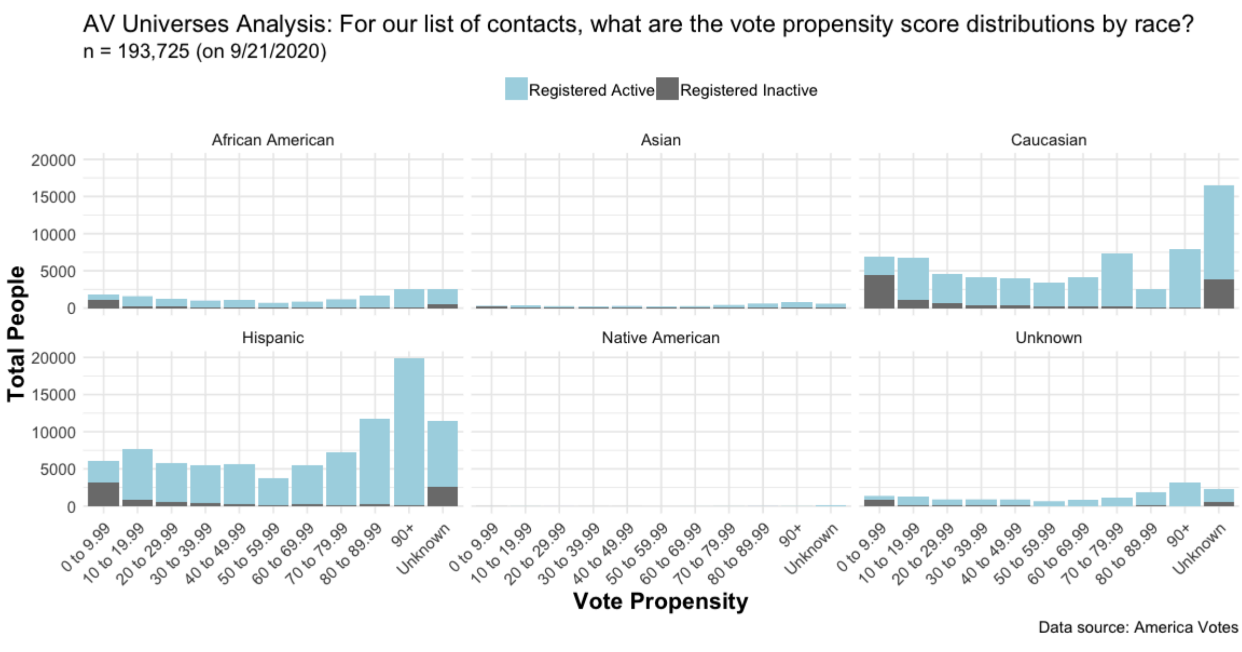


Figure 3: Analysis of LUCHA's 2020 targeted voter universe by race and vote propensity.

One way of finding some of the voters missing from this targeted list was for LUCHA to engage their own large membership base in voter contact to friends and family. That became difficult, though, when only 14% of their member lists matched to the voter file. Of those who did match, the majority were Hispanic, which makes sense given LUCHA's focused organizing program. However Caucasian was the second largest category, even though Adolfo Solorio, Data Director at LUCHA, knew from experience that there were more Black and Native members in LUCHA than white. It is possible that white members matched the file at a higher rate because they were older, because there are fewer errors in name spelling on the voter file, or for other reasons. The impact, though, was that matching their member list to the voter file did little to help LUCHA connect their most powerful spokespeople–their members–with the right lists of voters to talk with in their communities.

Realizing that they should also be reaching younger voters with lower propensity scores in a targeted effort, LUCHA leadership returned to their America Votes table managers to request additional contacts. With supplemented lists in Maricopa and Pima counties, LUCHA was able to correct for the initial targeting bias and reach thousands of new Black and Latino voters, but only in the final weeks before the election.

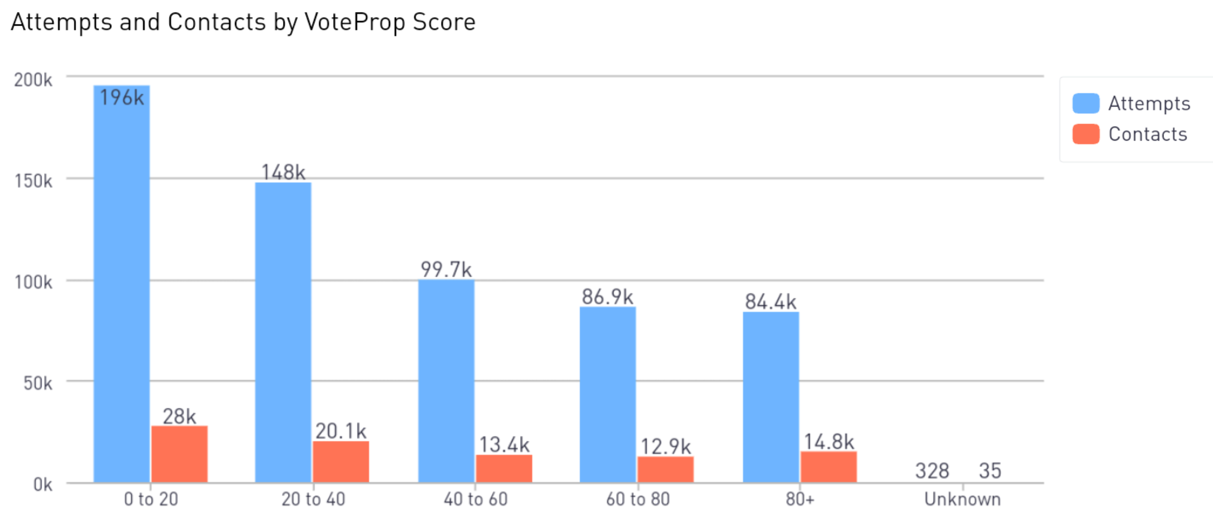
After the 2020 election, in an effort to increase their membership match rate in the future, LUCHA focused on gathering more contact and identity information from their members online and in person. By July 2022, LUCHA's Adolfo Solorio was able to successfully match 47% of LUCHA's membership base to their voter file database, a significant improvement over 2020.[3] When he and organizers looked at the lists of members still not matching, they found that although their base is quite large, many profiles were missing address data that would improve record matching. They are continuing to work on collecting address data from members. They also noted that people who live on tribal lands may have addresses not formatted in a way that facilitates voter file matching. In addition, there are some members who will not match because they are undocumented, or because they are under age 18 and unregistered. However, those members are still critical to LUCHA's efforts to register and turnout their eligible friends and family to vote, so their tools for 2024 will need to better facilitate that level of contact.

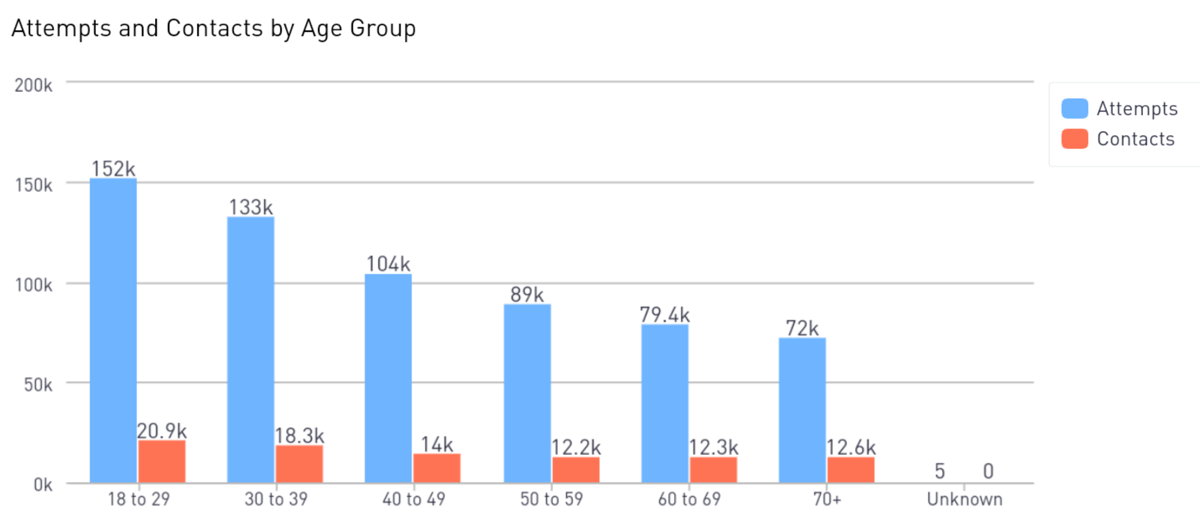
Adolfo also led the effort to help LUCHA reach the voters they wanted to reach in the next major election. On November 8, 2022, just ahead of national midterm elections, the

[3] All voter file data came from TargetSmart. The match was based on first and last names, phone numbers, emails and addresses. The LUCHA base was matched with a .25 match threshold. The match threshold could have been dropped to further increase matching rates, but at the expense of having bad matches.

visualizations of voter contact on LUCHA's post-election dashboard told a very different story than those of 2020: The majority of attempts and successful contacts in 2022 were with young and low propensity voters, in an almost completely inverse pattern to their initial 2020 progarm. (See Figure 4.)

By zeroing in on finding and engaging missing voters, LUCHA was better equipped to visualize and monitor outreach data in 2022. They also made a more concerted effort to reach new voters those with lower assigned propensity scores. LUCHA's internal analysis showed that "In 2022, about 19.5% of voters canvassed had no previous voting history. Of these, over 75% were people of color" (LUCHA, 2022). Of the 12,253 first-time voters they contacted, the average propensity score was 8.4, and 65% of them were Hispanic.





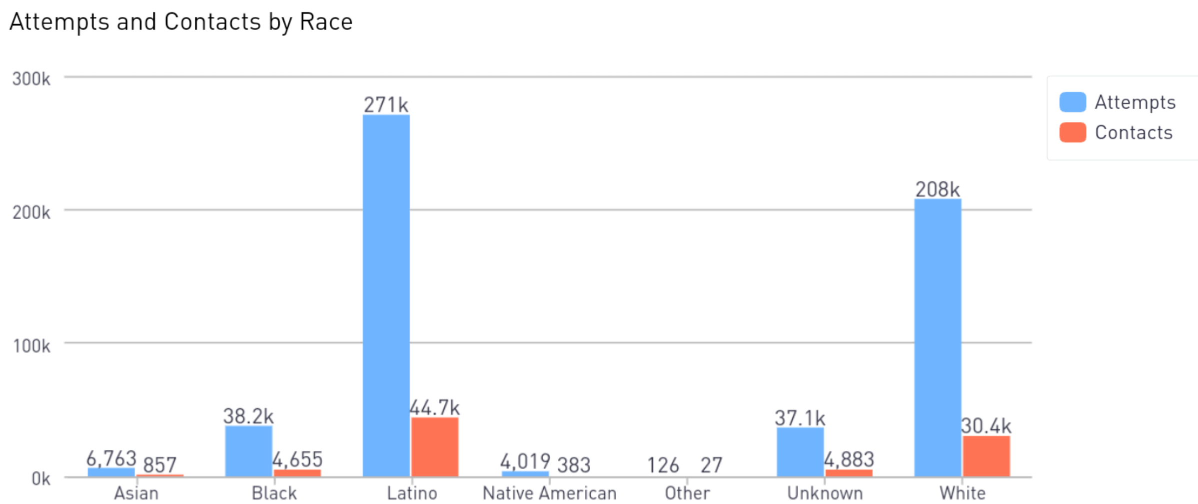


Figure 4: LUCHA Attempts and Contacts with Voters by Vote Propensity, Age & Race, 2022

Evidence from Wisconsin: Voces de la Frontera

Voces de la Frontera is a base building organizing group in Wisconsin seeking to improve the lives of Latino families, workers, immigrants and their allies through mutual aid programs as well as legislative campaigns. When they joined the DPI Organizing Lab in the spring of 2020, the Voces team were seeking to build their organization's power by bringing eligible voters into the electorate and growing the Latino share of the electoral pie in a state with razor thin margins for local, state and national elections. They were doing this through one of the largest relational outreach programs in the country, targeting their friends, family, and community members. They also did voter registration focused in the counties and wards in the state with the highest concentration of eligible Latino voters. By late summer, however, they wondered if any of that work was paying off. Why, when they compared current registration rates with those from a similar period in 2016, did they see little improvement or even decline in overall registration rates when they had done so much work in the interim?

Tianyi Hu, Senior Analyst with the State Power Fund, worked with Alan Nichols from Voces to find monthly registration rates by county and ward from 2016 through 2020 and to chart those for the places where Voces was working. What they found was alarming. In nearly every ward they studied they found a voter registration cliff between January 2016 and January 2018 (the first two bars for each ward in the chart below). With further research they found that immediately after the 2016 Presidential election, the Wisconsin Election

Commission, which had been initiated by Republican Governor Scott Walker, had purged almost 700,000 voters, or 19% of the electorate, from the rolls (PRWatch Editors, 2018). While this seems to have been evident to political elite in Wisconsin, it was news to Voces and many of their allies to learn that their voter registration efforts were simply filling the void that the purge had created. (See Figure 5.) This was part of a multi-state voter purge strategy between the 2016 and 2020 elections. Around the same time, in 2017, over 500,000 people were eliminated from Georgia's voter rolls as well (Niesse, 2021).

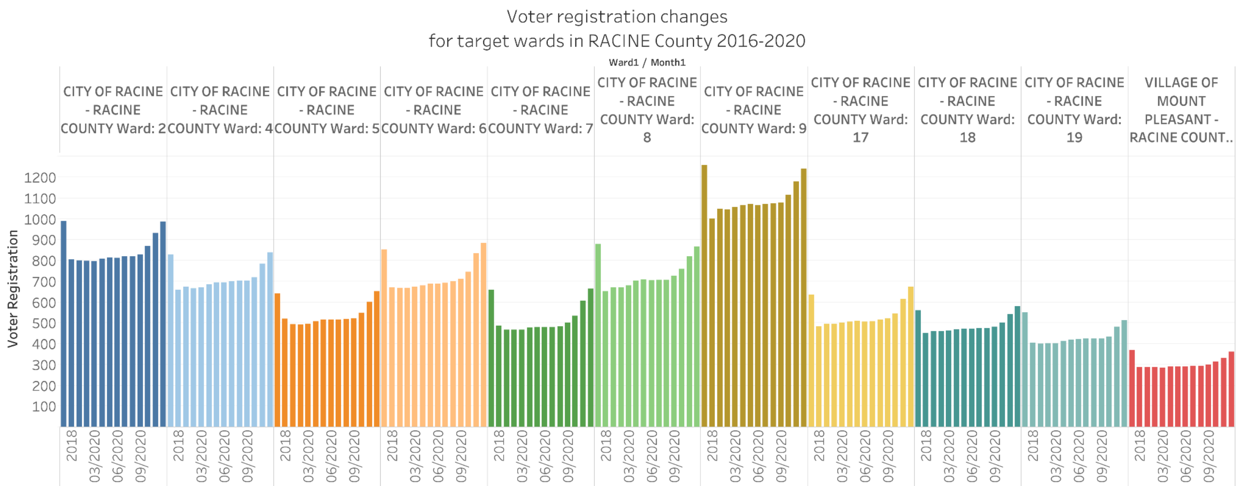
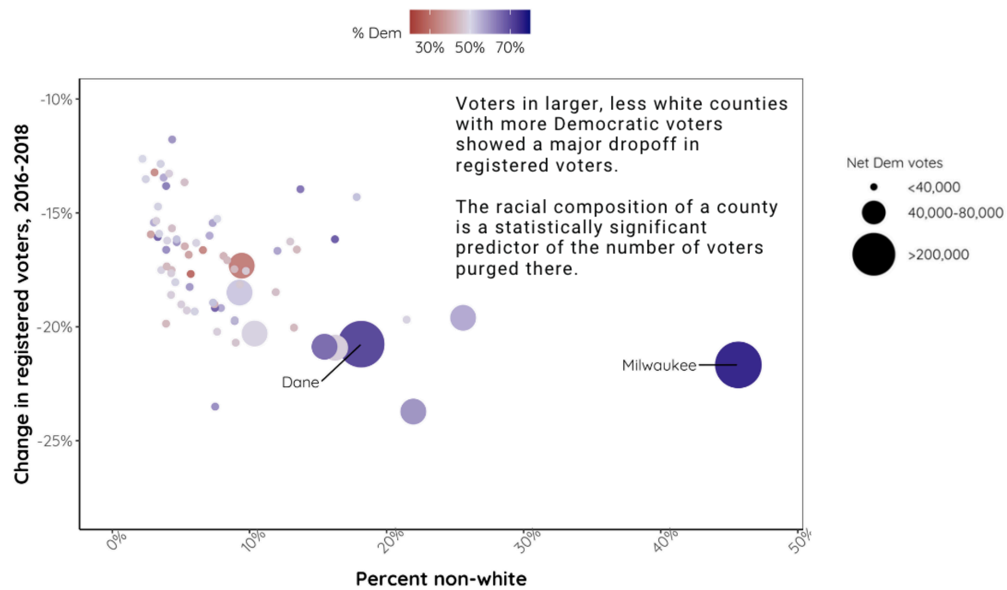


Figure 5: Voter registration changes for target wards in Racine County, Wisconsin, 2016-2020

Deeper analysis by Elizabeth McKenna, now at Harvard University, showed the Wisconsin purges were racially biased at a statistically significant level.

WISCONSIN VOTER PURGES BY COUNTY AND PERCENT NON-WHITE POST-2016 ELECTION



Source: Wisconsin Secretary of State and LTSB Open Data. Menominee, a Native American reservation that is 89% non-white and voted 91% for Clinton in 2016, is not shown for scale. 21% of its voters were purged from the rolls in this period. Dem votes are measured as presidential totals (Clinton v. Trump).

Figure 6: Wisconsin Voter Purges by county and percent non-white population post 2016 election

By looking critically at their state's registration data and asking "is our voter registration program having the intended impact" during the cycle instead of after, Voces became clearer that they were operating inside a complex, behind the scenes, mostly invisible power arena where the voices of hundreds of thousands of Wisconsinites could be silenced without a peep. They had to reorient, not just to registering voters, but to engaging in these power relationships directly with the Wisconsin Election Commission and local election officials. They also had the evidence to craft a motivational narrative for their constituents that was not just about the generic good of registering to vote, but about joining an active campaign against powerful opponents trying to silence voters in Wisconsin. **Through organizing, the proactive choice to register friends and family to vote became a deeply agentic and heroic one to build the power of their organization vis-a-vis opponents who had purged so many from the rolls.**

Over the next several months and years, Voces has doubled down on a set of precincts across the state that have large Latino populations. They have added thousands of voters

DEMOCRACY & POWER INNOVATION FUND

back onto the rolls through the hard work of voter registration and member outreach to friends and family.

> **Those removed from voter lists cease to exist to the political industry, in a way, as they become unseen by standard voter contact — that is, until organizations like Voces find innovative ways to find and engage the invisibilized.**

When purges happen and millions of people are actively invisibilized according to trivial criteria (including not voting in the past two election cycles, changing address, or having election mail returned to sender as "undeliverable"), the systems that many rely on for their voter contact work become less complete, trustworthy, and fair. Those removed from voter lists cease to exist to the political industry, in a way, as they become unseen by standard voter contact — that is, until organizations like Voces find innovative ways to find and engage the invisibilized.

Additional State Program-Generated Solutions

Like LUCHA and Voces de la Frontera, there are dozens of local and state-based organizing programs across the country working with each other and national allies and donors to generate grassroots solutions to voter file missingness. Here are some of the innovations groups are using to address the issues discussed in this paper.

1. Curate in-house voter files through intimate community engagement and relational strategies.

> **Lead Organizer Abdulahi Farah shared that they "were able to get access to community WhatsApp groups, partner with childcare and Islamic centers, and small businesses. We asked community leaders to download their phone contacts."**

Rather than rely solely on voter file data, ISAIAH, a broad base-building organization in Minnesota, focuses on reaching members in precincts with large Somali populations through relational and institution-based organizing via childcare centers, mosques, community centers, and small businesses. In 2020, the goal was to focus on expanding the electorate in Black, Muslim, and Latino communities statewide. Data Manager Amity Foster found that one key solution is to "rely on who you know to be your people." By enlisting their members to build new, internal lists of voters, ISAIAH increased their eligible contacts by 90,000 people, engaged 269,000 voters of color, and increased voter turnout in Somali precincts by 20%. Their focus was on conversations with "irregular voters," less than half of whom were listed in the state voter file, but who were eligible to vote.

How were they able to do this? ISAIAH team members were very flexible in how they gathered personal data because they knew that all communities are not the same and that "traditional" voter contact methods are only marginally effective at increasing voter turnout. Their Muslim canvassers used WhatsApp, pledge cards, and other community-based strategies to gather up-to-date personal information, then updated their in-house voter file database either manually or using an automated script. Lead Organizer Abdulahi Farah shared that they "were able to get access to community WhatsApp groups, partner with childcare and Islamic centers, and small businesses. We asked community leaders to download their phone contacts."

Expanding the muti-racial majority also meant expanding ISAIAH's internal data team and bolstering skills within the organization. They created a Community Data Team made up of 5 data field organizers who worked directly with volunteers to build updated, clean, and accurate lists of high potential Muslim, Black, and Latino voters across Minnesota.

Reflecting on all of these efforts to increase Minnesota's multi-racial majority by not relying solely on voter file data, Amity noted "that curating our own lists and files meant long-term base-building work." ISAIAH has been investing deeply in the leadership of Somali communities in the state for almost ten years now, and this electoral community list building is just one part of that community power-building effort. She shares that this time investment may not be easy or make sense to everyone at first, but will pay dividends with time.

Amity also noted that data collection and analysis can begin with developing strong leadership practices: "It is very important to name the growth and development of our leaders. That is where our work begins and ends. It's how we are able to run these kinds of in-depth programs. Our power-building is not just reliant on the number of conversations had and with what type of voter. It is essentially that our leaders are developed into claiming their own agency in democracy and doing voter outreach – that is where the power is."

2. Lean deeply into relational organizing during election seasons.

The Ohio Organizing Collaborative found that the solution to having an estimated two million missing voters in the state was to lean into voter registration and relational organizing. With this mindset, organizers were able to register 142,722 voters in 2018 (50%

of whom were millennials and 64% of whom were people of color). They also piloted their first relational voter outreach program in 2018.

In 2020, as documented above, OOC volunteers made lists of potential voters from their friends and family network. The organization checked if these people matched to the voter file and discovered that the majority of potential voters in their base's social networks qualified as "missing" (e.g., missing from the voter file) — even if they had voted before.

Figure 5 visualizes their 2020 relational outreach data. Here, the x axis represents age, the y axis represents projected turnout score, blue points represent Caucasian voters, red points represent Black voters, and yellow points represent Native voters. The area at the top shows voters with high turnout scores who are likely to be contacted by conventional, national campaigns; this cluster of points is skewed older. The points within the gray area encompass voters who did not match to the voter file at all and therefore did not have a turnout score. This cluster skews younger, based on age estimates provided by volunteer list-builders.

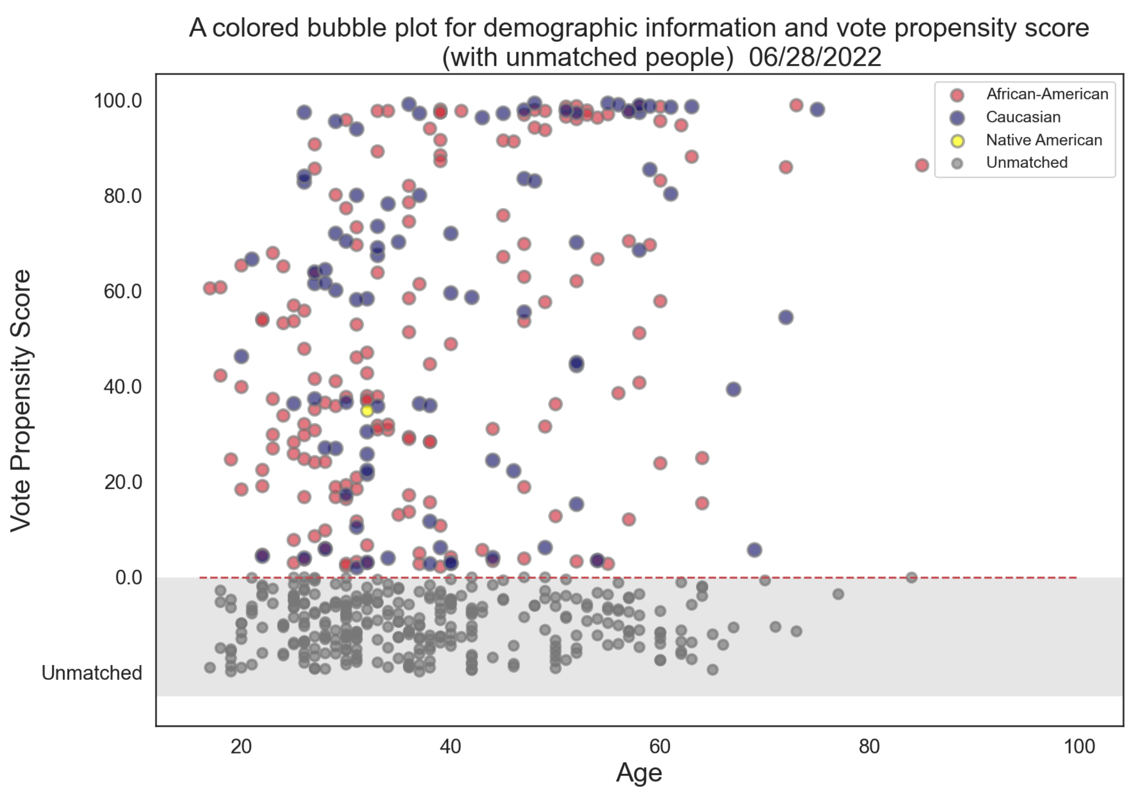


Figure 5: Demographic Information and Vote Propensity Scores, 2022

DEMOCRACY & POWER
INNOVATION FUND

The scatter plot visualization represents the relationship between low propensity score and missingness from the voter file. As can be extrapolated from the gray section towards the bottom: The majority of voters who are unmatched are young voters of color.

To combat this pressing issue, OOC led a relation program in 2020 in which Democracy Builders recruited volunteers in Cleveland, Cincinnati, and Columbus to text their friends and family to remind them of the upcoming election. After gathering data about these people, they found that 60% of all of these relational contacts were younger than 40 and 80% were “missing” voters. Additionally, of the voters who matched to the voter file, 70% had incorrect phone numbers on the voter file or no listed phone number at all.

“By the end of the relational program [in 2020], we made 28,000 relational contacts. Research suggests that this is equal in voter turnout impact to 23 million phone calls or 14 million cold texts,” Derrick shared. “This is the value of relational organizing,” Derrick continued. “People know how to reach their friends and family. They will have better phone contact information than what is available in the voter file.” Additionally, several studies have shown relational organizing increases turnout more than other voter contact tactics do.[4]

3. Actively remediate voter file errors at the organizational level.

The workaround deployed by AAPI Force, a coalition of AAPI organizing groups in California that focuses on engaging people of color, is founded in the acceptance of our current reality: We know that the voter file is wrong. And: We must fix it.

In their 2020 efforts to find AAPI voters, Executive Director Timmy Lu “knew that self-reported data was not enough.” Volunteers “found” thousands of AAPI voters who had been invisible by combing through outreach, surname, and genealogical data. They were able to grow AAPI voter lists from 180,000 to 330,000 voters, correct voter file errors (such as spelling errors and incorrect phone numbers), and, importantly, complete outreach in Asian languages. That personal touch was extremely effective and they had strong response rates. In the end, these efforts were supported by Timmy’s frank conclusion, that “the voter file is not built for us. We must understand this and work around it.”

Like AAPI Force, in 2020 OOC also found underrepresentation of their members and relational voter lists in state voter file data. An Analyst Institute study of their data post election found that even though the OOC had complete contact information for 87% of the people on their relational lists, only 47% of them matched to the voter file (Gutierrez &

[4] These studies have been led by the Analyst Institute and are available to their members.

McGarity, 2023). Similar to AAPI Force, OOC created a workaround for their relational organizing work by inputting the most up to date contact information manually.

4. Focus on registering new voters and maximizing same day registration.

Clearly when millions of voters are being purged from the voter rolls every year, organizations need the staff and resources to run large-scale effective voter registration programs that add new voters, and that re-enroll current voters with up to date contact information. Each cycle the Center for Information & Research on Civic Learning and Engagement at Tufts University does an in-depth analysis of voter registration across states. The most recent report showed that in 2020 field programs had the highest percentage of people of color registrations (68%) with online and mail lower at 41% and 48% respectively. In 2024 investment in these in-person field registration drives needs to grow significantly to continue to engage missing Black and brown potential voters.

Furthermore, 22 states and Washington, D.C. now have same day registration laws, allowing eligible voters to register and vote at the same time. This is a critical advancement in election law in the past decades. However, as a DPI report coming out later this spring shows, voters in states that have recently passed same day registration utilize the law at a very low rate. This may be in part because the people who would be eligible to use the law are left out of voter contact efforts because of the problems described in this memo. 2024 will be a critical opportunity for local and state community groups to find and engage the millions of missing Black and brown voters in their communities, to let them know about this new registration opportunity.

Recommendations and Conclusion

Recommendations

In the short term (within the next 6 months)...

ORGANIZING AND VOTER ENGAGEMENT GROUPS FOCUSED ON ENGAGING BLACK AND BROWN COMMUNITIES:

Clean up and fill in your membership database now. It is likely full of missing voters. You can run an assessment for accuracy and completeness of your member databases using analytics toolkits available for free from DPI. You can also use pooled community-driven databases to supplement your member data, such as the *TMC_Activist_Pool* schema produced by The Movement Cooperative. Several DPI partner groups use this data pool, such as New Georgia Project and Color of Change, to better understand and analyze membership at scale because: "you get a better and higher match rate using the activist pool rather than your own Civis match."[5] Do not rely only on voter file matching for contact data that can be incorrect or out of date for your members. Instead launch your own internal phone banks or canvasses or relational outreach to engage your members about your mission and to collect contact and demographic data from them over time. Develop a habit of using member sign-in sheets or sign-in apps at all events to keep your base's contact information up to date year round. And of course, provide the chance regularly for them to register to vote.

Negotiate to get targeted voter outreach lists that match your organization's power-building mission. Then analyze those lists closely when you get them. If your voter lists are being created by others, then you can run counts by demographic or geographic categories and by vote history to ensure your list reflects the type of people you want to reach. If you need a new list, then you can be more specific with the group providing your targeting services about which types of voters you want to contact. Remember, your highest return on program may be with the people other campaigns have neglected to reach, and your outreach could help keep them on the voter rolls.

[5] Civis is a useful online querying tool frequently used by many DPI Organizing Data Lab groups for voter file matching, saving data queries from certain projects, and referring back to past data exports. When an organization references "Civis matching," they typically mean that they have made an effort to join two datasets using Civis (frequently: membership data and voter file data).

Pivot to prioritize relational organizing work. We already know how crucial community organizations are in their ability to create real social change (Christens, 2010). As Paul Speer argues, "Community organizing, as a particular type of mediating institution, cultivates sociopolitical development by elevating psychological empowerment and civic engagement over time" (Speer et al., 2021). If we want to bring missing voters into community and long term civic engagement with others, it starts with base-building organizing groups. Furthermore, as referenced earlier, there is evidence that intimate, within-network, relational organizing work–as compared with large scale phone, canvas or text campaigns based on incomplete voter files–has the most value for organizations seeking to find missing voters, increase voter turnout, and to have deep, sustainable impact on voter engagement within Black and brown communities over time. As the OOC pilot showed, relational outreach also generates the most accurate list of missing voters, and can be done at significant scale if it is a resourced organizational priority during election cycles. If you aim to unearth deeper and more meaningful power for communities of color and other voters commonly ignored by consultants and political parties, it is crucial to implement and grow effective relational organizing programs that are rooted in permanent power organizations this year.

Design programs that take full advantage of same day registration and other accessible voting laws in your state. Do not rely solely on voter file lists for your voter contact program, particularly in the 22 states plus DC, where eligible voters can register and vote at the same time.

Remember, your highest return on program may be with the people other campaigns have neglected to reach, and your outreach could help keep them on the voter rolls.

DONORS:

Invest in programs that are identifying, engaging, registering and turning out missing voters. This year will provide significant opportunities to engage the millions of missing Black and brown voters in our country, *but only if we are intentionally oriented in that direction*.

Support the long-term base-building organizations that are filling in the holes created by targeted voter purges. Each state has an ecosystem of groups large and small who are engaging missing voters daily in their service and organizing work. They need year-round resources to continually register and engage a high scale of potential voters every day.

Ask the programs you invest in how they are finding and engaging missing voters. The dominant culture in the political industry is to focus on regular voters with higher vote propensity scores. Do not assume that the programs you engage with are focused on finding missing voters. In fact the opposite is likely true. You could play a pivotal role in bringing their attention to this critical problem, and giving them the programmatic resources to do the off-file work necessary to surface and connect with missing voters.

This year will provide significant opportunities to engage the millions of missing Black and brown voters in our country, *but only if we are intentionally oriented in that direction*.

In the long term (over the span of 1-3 years)...

Move away from using vote propensity scores for the purpose of including or excluding voters for contact. Use vote propensity scores only as a frame for reference, rather than a cut-off for engagement, understanding the bias in them that is documented above. Try to focus on people who are new voters or who have not voted regularly in the past, starting with the millions of members who belong to movement organizations.

Invest in community oversight of commonly used data models and implement anti-racist modeling practices. Data vendors should transparently answer a set of questions about the biases that could be implicit in their models and the data underlying them. They should test their models with Black and brown led organizations and communities and adjust accordingly. They should also provide end users of the models with intended use scenarios, including cautions about risks and the potential for compounded racial exclusion over time. Organizations using voter file databases and models should be able to understand clearly the purpose and limits of the models they employ.

Hold movement data and tech vendors accountable to end users. It should not be such an arduous task for organizations to download and analyze their own data in their efforts to generate full civic inclusion. Carter Calchik's 2023 memo details ways data and technology vendors can be held accountable to end users, investing more time in quality assurance, increased functionality and integration with other platforms to better support users' needs. This is critical, for example, for groups who register thousands of new voters captured in one database that does not sync with common voter file databases and interfaces for their get out the vote efforts.

Embrace new metrics for analyzing effectiveness, including power metrics and inclusion of missing voters. When time is limited, it is understandable that organizations and donors aim to make voter contacts as effective as possible. While the ease of using only scaled contact to people on the voter file creates the illusion of effectiveness, it is leaving millions of potential voters on the sidelines. DPI's 2023 report on Power Metrics, together with the recommendations in this memo, provide a new lens on metrics that groups and donors can use to measure their impact and not simply their programmatic inputs this year (Cushman & McKenna, 2023).

Invest in organizations' internal talent by focusing on analytics, critical thinking, and cultural competency skills – The organizations at the frontlines of voter engagement are ultimately best positioned to develop plans for finding and engaging missing voters. However, they need talented analysts in their own organizations or in state to do this who have familiarity with and a healthy skepticism about data tools and systems. They also need the opportunity to do open-ended pilot learning programs that would allow them to try new ways to reach missing voters. This is how the state-based solutions above came to fruition.

Conclusion

Our country continues to quickly shuttle towards becoming a highly virtual, data-driven, and technology-dependent democracy. Because of this reality, we cannot continue to ignore the issues outlined in this report. We cannot continue to leave significant segments of our communities behind over the long term in our efforts to achieve short-term efficiencies. **If we rely solely on currently problematic yet commonplace methods for engagement, we continue to erase the young, frequently mobile, potentially outcome-changing Black and brown electorate.** We must also remember that the demographic landscape of the United States continues to shift as it becomes larger, more connected, and more diverse. As people of color become not just the majority of the American population but the majority of the American electorate, our ability to build an inclusive, multiracial democracy depends on our ability to find and bring these communities fully into our reach for contact and engagement.

In a country with nearly 25 million missing Black and brown voters, the organizations that can find and engage these citizens consistently will unleash a new power in American politics. As Joy Cushman, Senior Advisor to DPI said, “The people unseen by voter files are still capable, if organized, to make moves and wield their latent power. People deemed ‘low propensity’ by models and the political industry are defying the odds and still turning out to vote. And many are doing even more than that: they are becoming active members and leaders in power-building organizations, mobilizing their friends and family to vote as well.” Imagine what is possible if they are embraced, encouraged and resourced to civically engage.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine Bias.* ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Anderson, D., Bonaguro, J., McKinney, M., Nicklin, A., & Wiseman, J. (2018). *Ethics and Algorithms Toolkit.* Center for Government Excellence. https://ethicstoolkit.ai/

Christens, B.D. (2010, August 3). Public relationship building in grassroots community organizing: relational intervention for individual and systems change. *Journal of Community Psychology*. https://onlinelibrary.wiley.com/doi/abs/10.1002/jcop.20403

Cushman, J. and McKenna, E. (2023). Power metrics: Measuring what matters to build a multiracial democracy. *Democracy and Power Innovation Fund*. https://community.dpifund.org/files/g51u5vtdkfy0wmp8nw83dv0dyl0c

Daniller, A. & Gilberstadt, H. (2020, December 14). *Key findings about voter engagement in the 2020 election*. Pew Research Center. https://www.pewresearch.org/short-reads/2020/12/14/key-findings-about-voter-engagement-in-the-2020-election/

DeSilver, D. (2018, February 23). *Q&A: The growing use of 'voter files' in studying the U.S. electorate.* Pew Research Center. https://www.pewresearch.org/fact-tank/2018/02/15/voter-files-study-qa/

Elliott, D., Santos, R., Martin, S., & Runes, C. (2019). *Assessing miscounts in the 2020 census.* Urban Institute. https://www.urban.org/sites/default/files/publication/100324/assessing_miscounts_in_the_2020_census_1.pdf

Equis Research. (2022). *DPI political participation poll.* https://community.dpifund.org/files/cel83s8u4enhl693aqxa5mmomowz

Faith in Action. (2019). *Civic engagement history and research.*

Frey, W. H. (2021, August 13). *New 2020 census results show increased diversity countering decade-long declines in America's white and youth populations.* Brookings Institution.

https://www.brookings.edu/research/new-2020-census-results-show-increased-diversity-countering-decade-long-declines-in-americas-white-and-youth-populations/

Fultinaviciute, U. (2022, October 17). Sex and science: Underrepresentation of women in early-stage clinical trials. *Clinical Trials Arena*. https://www.clinicaltrialsarena.com/features/underrepresentation-women-early-stage-clinical-trials/?cf-view

Garland, J. (2021, October 18). *Voter purges among communities of color*. TargetSmart. https://qa-insights.targetsmart.com/voter-purges-among-communities-of-color.html#_ftn1

Green, B. (2021). Data science as political action: Grounding data science in a politics of justice. *Journal of Social Computing, 2*(3), 249-265. https://doi.org/10.23919/JSC.2021.0029

Gutierrez, M., McGarity, B. (2023). Ohio Organizing Collaborative 2020 and 2022 relational voter turnout program descriptive analyses. Analyst Institute. https://community.dpifund.org/files/qcob33189b6nhw5ul914b285grft

Higher Ground Labs (2023, March 31). 2022 Political tech landscape report. https://highergroundlabs.com/political-tech-landscape-report-2022/

Jackman, S., & Spahn, B. (2021). Politically invisible in America. *PS: Political Science and Politics, 54*(4), 623-629. https://doi.org/10.1017/S1049096521000639

Kalchik, C. (2023). Surveying the landscape of data integration for state and local organizing groups. *Democracy and Power Innovation Fund*. https://community.dpifund.org/files/88orff6xtjrt0sadhp4fxm20kejz

Kelsey, M. et al. (2022, May). Inclusion and diversity in clinical trials: Actionable steps to drive lasting change. *Contemporary Clinical Trials*, vol. 116.

Kramer, M. (2019, April 1). *Census 2020: Everything you need to know about north carolina's hard-to-count communities.* https://carolinademography.cpc.unc.edu/2019/04/01/census-2020-everything-you-need-to-know-about-north-carolinas-hard-to-count-communities/

McKinney Gray, M. (2023, June 9). Black and Latino voters are being erased from databases: Here's how to fix it. *The Hill*. https://thehill.com/opinion/civil-rights/4040613-black-and-latino-voters-are-being-erased-from-databases-heres-how-to-fix-it/

Mora, G. C., & Dowling, J. A. (2022, March 11). What can we do about the latino undercount in the 2020 census? *The Los Angeles Times.* https://www.latimes.com/opinion/story/2022-03-11/census-undercount-latinos-race-ethnicity-donald-trump?_amp=true&fbclid=IwAR2os7XDE-RZR-mz-GlymOfdfntns5zamM6EwhVx2s6glIaZ9GqaRDjsHj0

Najibi, A. (2020, October 24). Racial discrimination in face recognition technology. *Science in The News, Harvard University.* https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/

Niesse, M. (2021, June 18). Georgia releases list of 102K voter registrations to be canceled. *The Atlanta Journal-Constitution.* https://www.ajc.com/politics/georgia-releases-list-of-102k-voter-registrations-to-be-canceled/FKYR6VYAQVENRHX5PC3F7ITL7Y/

PRWatch Editors (2018, October 25). Hundreds of thousands of Wisconsin voters purged from rolls since 2016 general election. *The Center for Media and Democracy's PRWatch*. https://www.prwatch.org/news/2018/10/13417/hundreds-thousands-wisconsin-voters-purged-rolls-2016-general-election

Rentsch, A., Schaffner, B. F., & Gross, J. H. (2019). The elusive likely voter: Improving electoral predictions with more informed vote-propensity models. *Public Opinion Quarterly, 83*(4), 782-804. https://doi.org/10.1093/poq/nfz052

Schneider, M. (2022, March 10). Latinos undercounted by 5% in latest census, more than three times the amount in 2010. *The Los Angeles Times.* https://www.latimes.com/world-nation/story/2022-03-10/2020-census-bureau-undercount-hispanic-black-native-american

Smith, P. M. (2020, February 9). "Use it or lost it": The problem of purges from the registration rolls of voters who don't vote regularly. *The American Bar Association.* https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/voting-rights/-use-it-or-lose-it---the-problem-of-purges-from-the-registration0/

Speer, P. W., Christens, B. D., & Peterson, N. A. (2021). Participation in community organizing: Cross-sectional and longitudinal analyses of impacts on sociopolitical development. *Journal of Community Psychology, 49*(8), 3194-3214. https://doi.org/10.1002/jcop.22578

Stanford, J. (2020, March 9). *Why some Americans don't trust the census.* Carolina Demography.

https://carolinademography.cpc.unc.edu/2020/03/09/why-some-americans-dont-trust-the-census/

Tamir, C (2021, March 25). *The Growing Diversity of Black America.* Pew Research Center. https://www.pewresearch.org/social-trends/2021/03/25/the-growing-diversity-of-black-america/

Timm, J. C. (2021, December 21). *19 states enacted voting restrictions in 2021. What's next?* NBC News. https://www.nbcnews.com/politics/elections/19-states-enacted-voting-restrictions-2021-rcna8342

Turman, J. (2020, July 10). *DNC hopes to reach more voters with new voter file model.* CBS News. https://www.cbsnews.com/news/dnc-hopes-to-reach-more-voters-with-new-voter-file-model/

Vasilogambros, M. (2019). *The messy politics of voter purges.* The Pew Charitable Trusts. https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2019/10/25/the-messy-politics-of-voter-purges

Vines, M. (2018, December 12). *2020 census barriers, attitudes, and motivators study (cbams) survey and focus groups: Key findings* [PowerPoint slides]. US Census Bureau. https://becountedmi2020.com/wp-content/uploads/CBAMS_Presentation_MNA_121218.pdf

Yong, E. (2018, January 17). A popular algorithm is no better at predicting crimes than random people. *The Atlantic.* https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/

Zest AI releases new race prediction model to reduce systemic bias in lending. (2022, March 1). Cision PR Newswire. https://www.prnewswire.com/news-releases/zest-ai-releases-new-race-prediction-model-to-reduce-systemic-bias-in-lending-301492688.html