Audience Analysis for Vote-By-Mail Text Messages

Team 1
Steve Bao
Ryan Brockway
Keith Tyser
Nicholas Zotalis

Abstract

VoteAmerica is a non-partisan organization with the mission of increasing voter turnout, particularly among voters of color. Our team investigated the results of a campaign where vote-reminder texts were sent to voters who had received absentee ballots. We studied the demographic makeup of the groups that were contacted, both at a national level and on a state-by-state basis. We also investigated the effect that these texts had on voting rates. We checked the predictive power of vote propensity and contactability on voting rates, as well as what voting methods corresponded to these factors.

Data Processing

The majority of data was provided to us by VoteAmerica. We were given csv files that contained the messages that VoteAmerica sent to the voters, the demographics of these voters, as well as information on which voters opted out from receiving these text messages. To process this data, we combined these different tables into one, and then from there, began to analyze the data. We had approximately 8.8 million text message records. We also obtained census data in order to examine how different states were targeted.

Project Description

VoteAmerica sent over 100 million text messages to voters across the country during the 2020 general election cycle. Over 8 million messages were reminders to voters to submit their mail or absentee ballots. This project is focused on analyzing the demographics of who was targeted, who was reached, and voting rates and presenting that information in an appropriate, informative manner.

Table of Contents

National-level Voter Demographics:

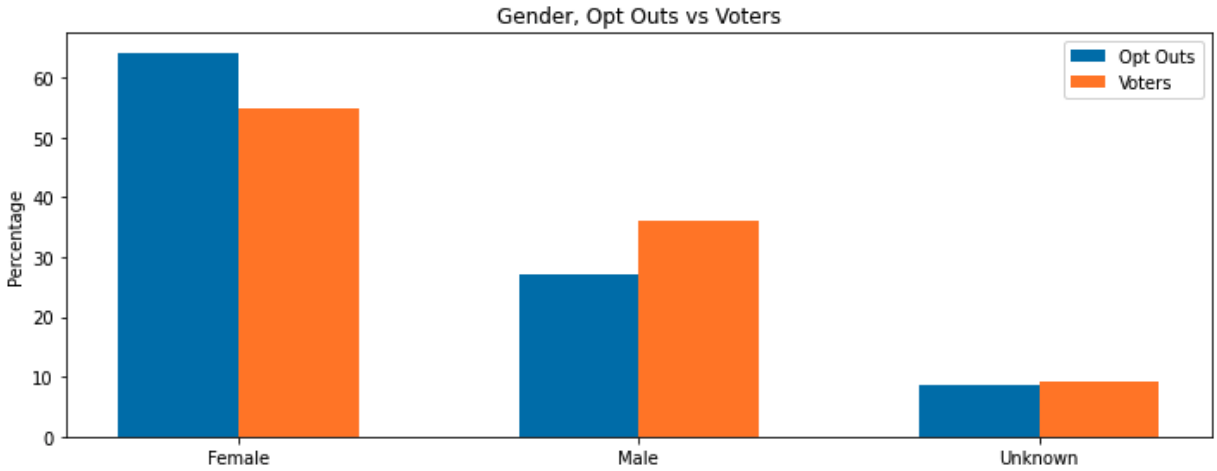
This section contains information on targeted-voter demographics and opt-outs by group, at the national level.

New American Majority Group		Number of People Texted	Percentage of Total Texted
African-American		425791	14.9%
Asian		103452	3.6%
Hispanic		608517	21.3%
Native American		15708	0.6%
Other		36759	1.3%
White or Race Unknown	Unmarried Women 30-49	308245	10.8%
	Young Men (18-29)	352280	12.4%
	Young Women (18-29)	431216	15.1%
	Young, Gender Unknown (18-29)	102744	3.6%
Not Part of New American Majority		466139	16.4%
Total		2850851	100.0%

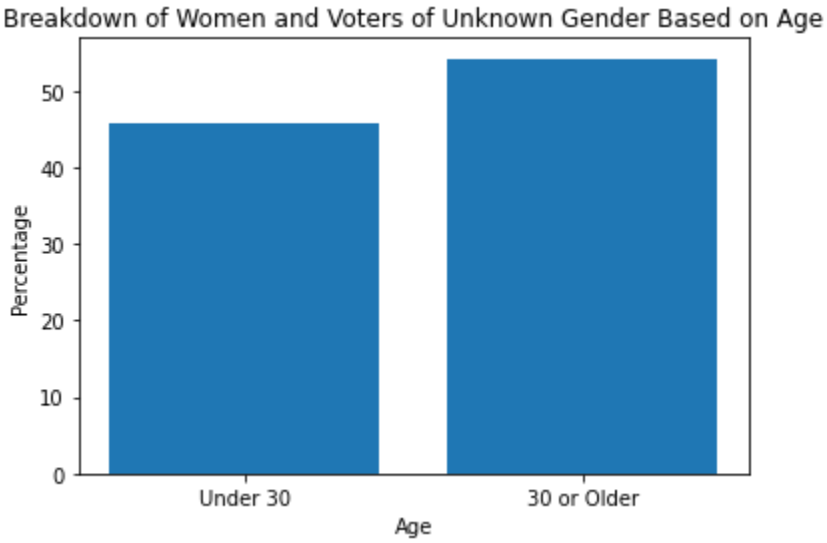
First, we examined the voters targeted based on if they were part of the New American Majority. Looking at all of the voters texted, 84% of all voters texted were part of the New American Majority and about 42% were part of the POC group (everyone but Caucasian).

Of all groups, Hispanic voters were the New American Majority Group with the most voters targeted. VoteAmerica is reaching New American Majority Groups and are not sending many text messages to voters who are not a part of the New American Majority.

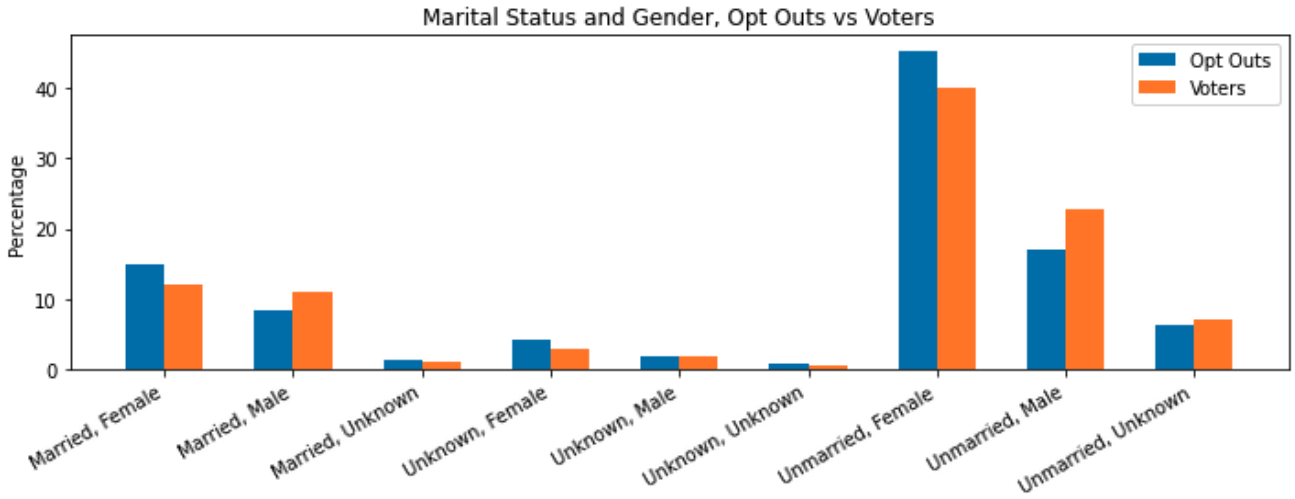
Next, we examined the general demographic breakdown of voters targeted, as well as how often each group opted out of receiving the text messages.



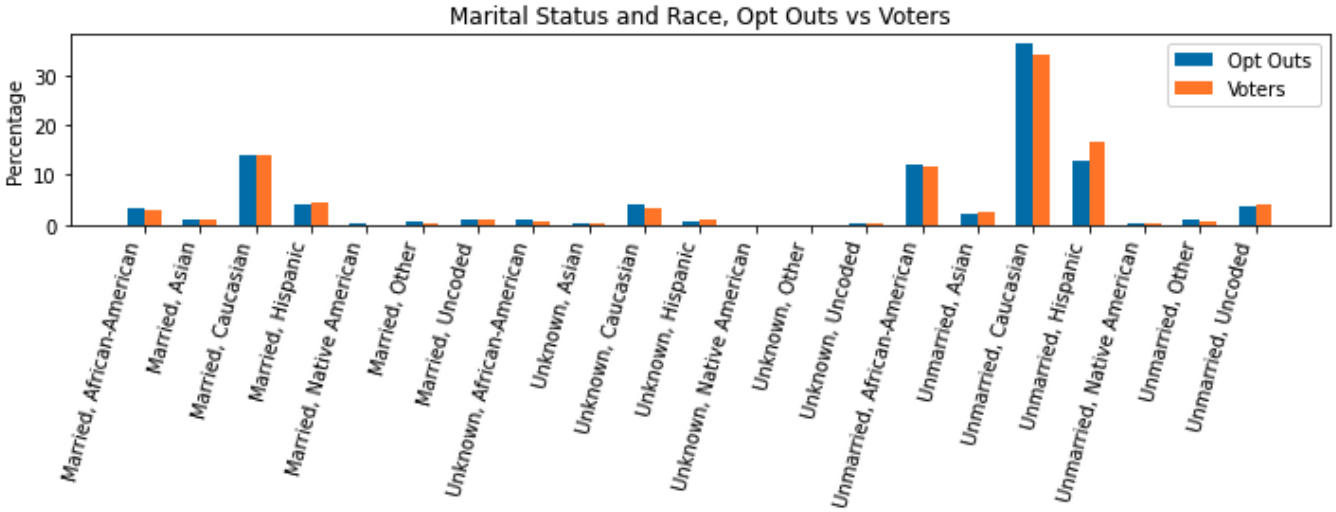
First, we analyzed the gender of people VoteAmerica reached out to. Women were contacted the most, followed by men. Based on this graph, women tended to opt out more, while men tended to vote more.



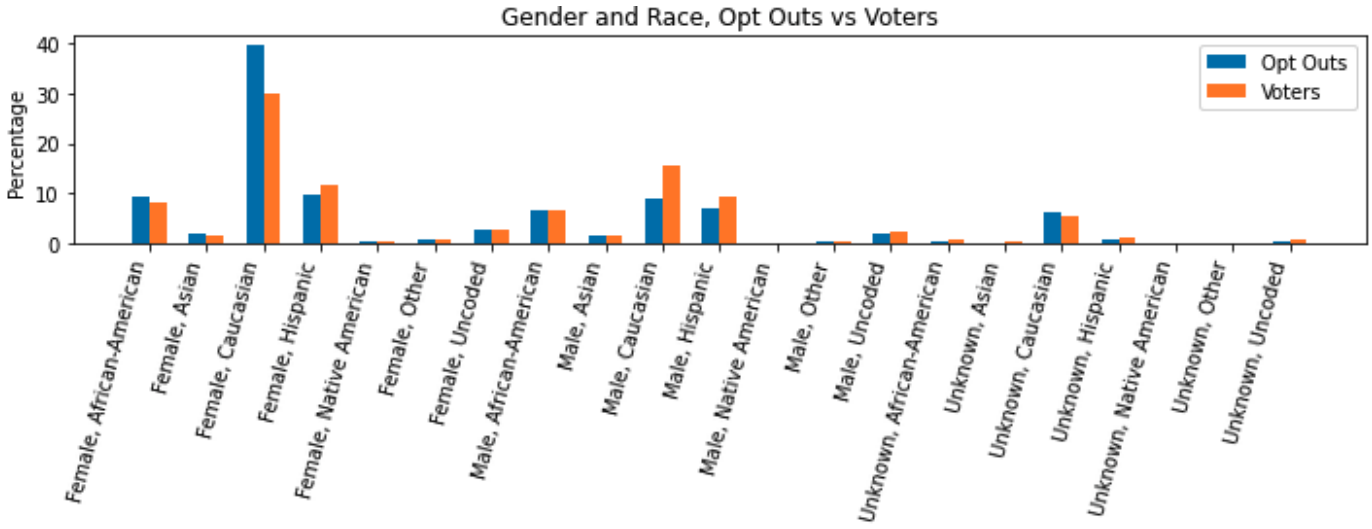
Then, looking at just women and voters of unknown gender, voters under the age of 30 only made up 45.8% of the voters contacted. If we look at the entire dataset, women and voters of unknown gender under the age of 30 only made up 29.3% of voters.



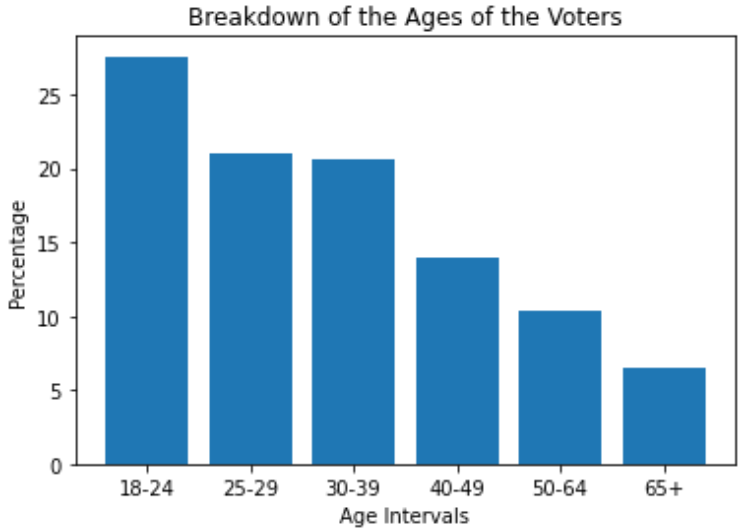
We then began to analyze combined demographics of people VoteAmerica reached out to. We first analyzed the marital status and gender of people in the campaign. Unmarried women were contacted the most, followed by unmarried men.



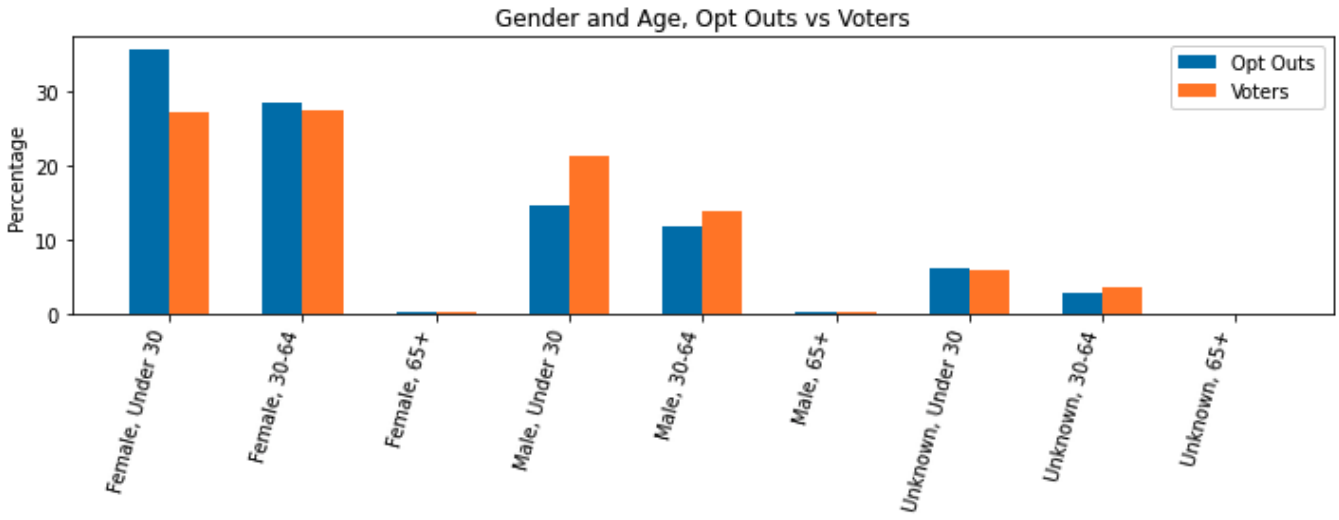
Next, we analyzed the marital status and race of people in the campaign. Unmarried Caucasians were contacted the most, followed by unmarried Hispanics, unmarried African-Americans, and married Caucasians.



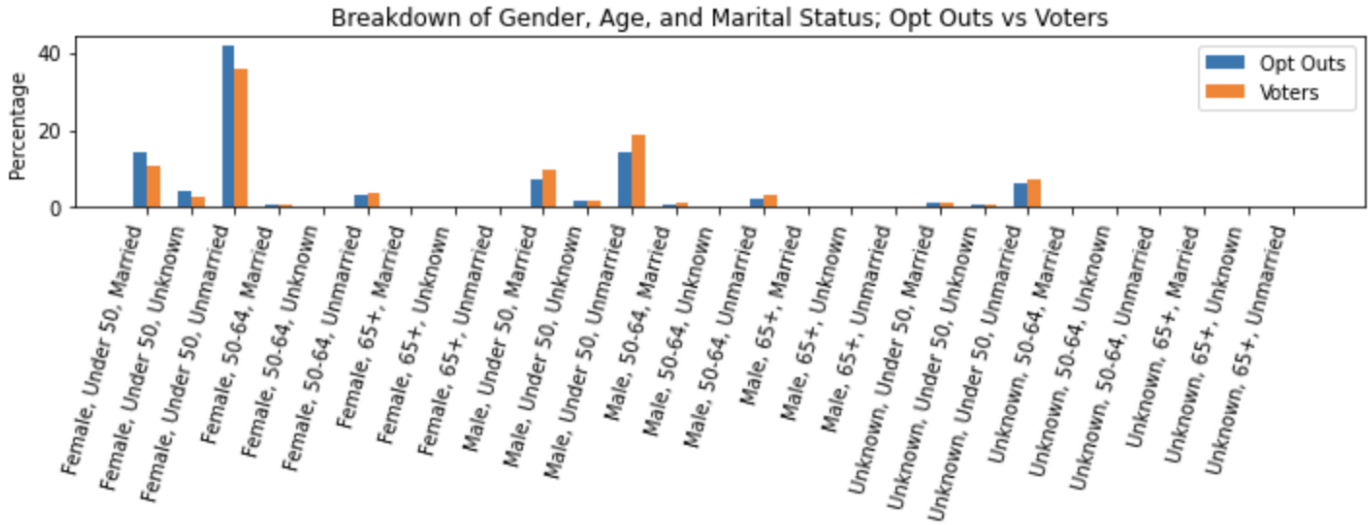
Here we can see the gender and race of people in the campaign. Caucasian women were contacted the most by VoteAmerica.



The most voters contacted were between the ages of 18-24 (27.58%), followed by ages 25-29 (21.05%), and ages 30-39 (20.67%).



One of VoteAmerica's target demographics is women under 30, so we looked at gender and age of people in the campaign. This graph shows that this demographic is the one they texted the most. Therefore, they are reaching one of their targeted audiences.



Another one of VoteAmerica's target demographics is unmarried women under 50 so we looked at the gender, age, and marital status of people in the campaign. This graph shows that unmarried women under 50 are the demographic that they texted the most, so they are reaching one of their desired demographics.

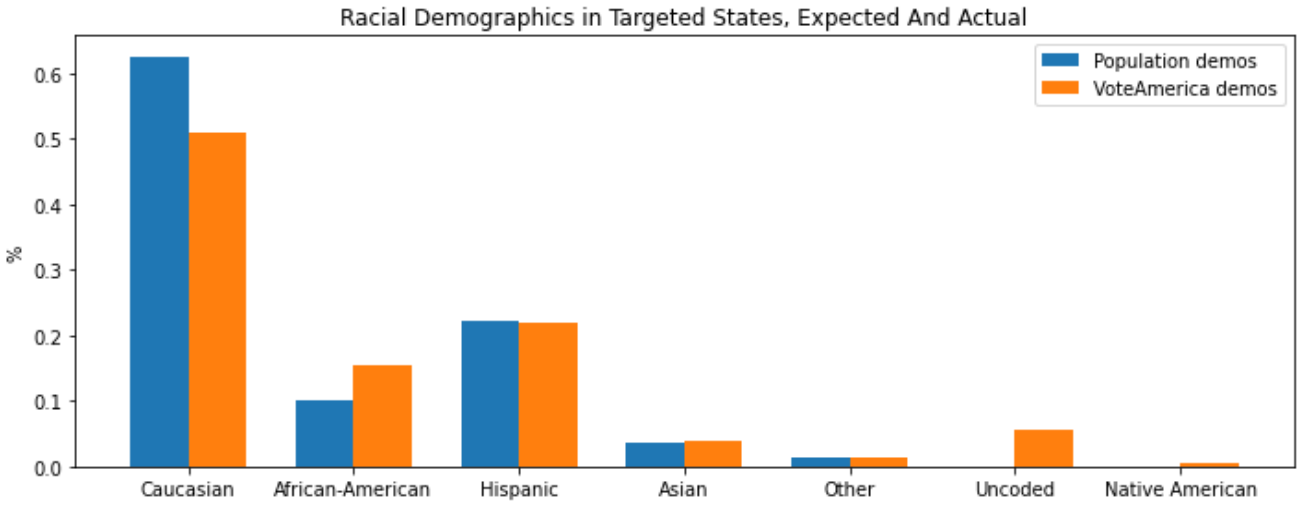
State-level Voter Demographics:

In this section we look at the demographics of the states that were targeted most frequently. We obtained the racial demographics on a state by state basis from this dataset: https://worldpopulationreview.com/states/states-by-race.

The top 25 states contacted by VoteAmerica account for 99% of all the texts sent, so we will use their values to approximate the entire demographic set.

Taking a weighted average we obtain the following racial demographics:
(White: 62.68%, Black: 10.2%, Hispanic: 22.2%, Asian: 3.5%, Other: 1.5%)
These are the true demographics of the states targeted by VoteAmerica, based on the proportion of texts sent to each state.

The VoteAmerica messages were actually sent to demographic groups in the following proportions:
(White: 51.1%, Black: 15.6%, Hispanic: 22.1%, Uncoded: 5.6%, Asian: 3.8%, Other: 1.4%, Native American: 0.56%)



What we can see is that at the national level, the VoteAmerica campaign is under-representing white voters, by about 12%. It is over-representing African American voters by about 5%, and tracking at the baseline for Hispanic, Asian, and Other voters.

Then, we look into the racial breakdown of each of the targeted states

Table: VoteAmerica demographic targets by state

State	Caucasian	African-American	Hispanic	Asian	Other	Native American	Uncoded
FL	31.00%	22.00%	36.00%	5.00%	5.00%	0.40%	0.50%
CO	59.00%	4.00%	27.00%	1.00%	0.00%	0.30%	8.00%
AZ	50.00%	2.00%	37.00%	1.00%	0.00%	2.00%	7.00%
NV	50.00%	8.00%	29.00%	5.00%	0.00%	0.10%	7.00%
PA	60.00%	18.00%	7.00%	5.00%	0.00%	0.02%	10.00%
MI	60.00%	28.00%	2.00%	4.00%	0.00%	0.10%	7.00%
UT	81.00%	0.30%	10.00%	1.00%	0.00%	0.40%	7.00%
GA	24.00%	54.00%	7.00%	10.00%	4.00%	0.30%	1.00%
OH	66.00%	24.00%	2.00%	2.00%	0.00%	0.20%	7.00%
NC	27.00%	47.00%	9.00%	7.00%	6.00%	1.00%	2.00%
MT	87.00%	0.20%	1.00%	0.10%	0.00%	5.00%	7.00%
KS	70.00%	7.00%	9.00%	3.00%	0.20%	0.20%	10.00%
IA	84.00%	3.00%	4.00%	2.00%	0.01%	0.06%	7.00%
MN	63.00%	12.00%	3.00%	9.00%	0.02%	0.30%	12.30%
TX	21.00%	17.00%	44.00%	10.00%	0.30%	0.10%	8.00%
NE	78.00%	7.00%	6.00%	1.00%	0.03%	0.20%	8.00%
WI	60.00%	18.00%	8.00%	4.00%	0.04%	0.40%	1.00%
ME	93.00%	1.00%	0.20%	1.00%	0.10%	0.10%	0.40%
CA	58.00%	9.00%	19.00%	7.00%	1.00%	0.10%	5.00%
NY	49.00%	20.00%	18.00%	6.00%	3.40%	0.30%	3.70%

Table: True demographic values of each of these states:

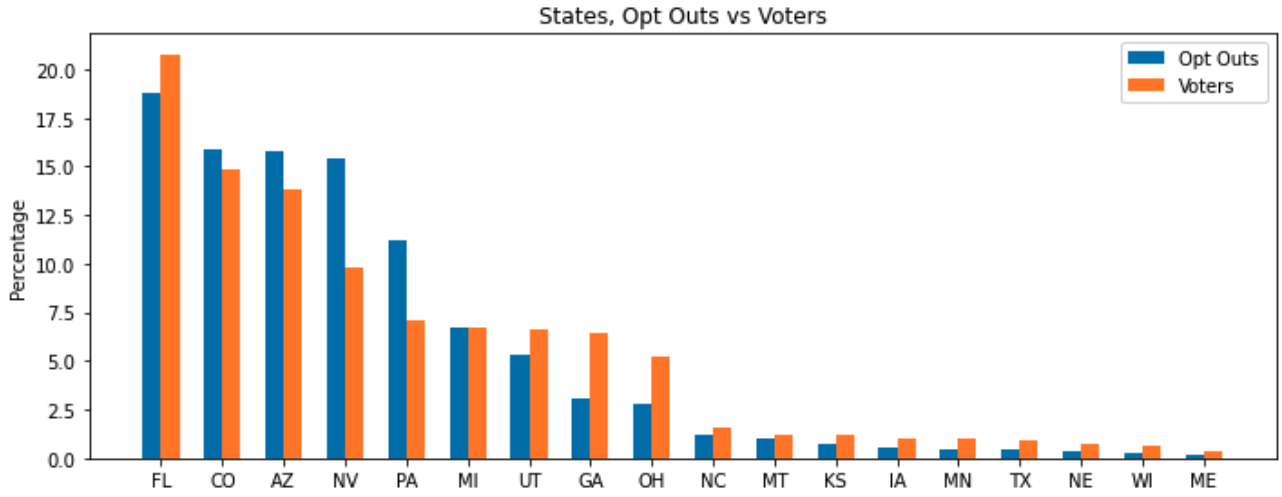
State	Caucasian	African-American	Hispanic	Asian	Other
AZ	54.71%	4.21%	31.34%	3.21%	6.53%
CA	37.18%	5.52%	39.02%	14.28%	4.00%
CO	68.06%	3.95%	21.53%	3.11%	3.35%
FL	53.90%	15.32%	25.58%	2.68%	2.52%
GA	52.73%	31.18%	9.54%	3.95%	2.60%
IA	85.68%	3.62%	6.00%	2.40%	2.30%
KS	75.88%	5.65%	11.87%	2.92%	3.68%
ME	93.19%	1.30%	1.65%	1.12%	2.74%
MI	75.03%	13.63%	5.08%	3.12%	3.14%
MN	79.85%	6.29%	5.37%	4.79%	3.70%
MT	86.09%	0.44%	3.85%	0.77%	8.85%
NE	78.98%	4.69%	10.88%	2.37%	3.08%
NV	49.23%	8.74%	28.70%	7.99%	5.34%
NY	55.61%	14.26%	19.00%	8.35%	2.78%
NC	63.08%	21.09%	9.38%	2.83%	3.62%
OH	78.91%	12.23%	3.80%	2.20%	2.86%
PA	76.43%	10.66%	7.31%	3.39%	2.21%
TX	41.95%	11.78%	39.34%	4.74%	2.19%
UT	78.33%	1.12%	14.04%	2.29%	4.22%
WI	81.26%	6.29%	6.80%	2.78%	2.87%

The next figure presents the same information in the form of differences between VoteAmerica demographics values and expected values. Negative values correspond to underrepresented groups, and positive values correspond to overrepresented groups. In this case the colors represent the value of each cell relative to its column (A green value in the African-American column means a high value relative to other African-American values.)

State	Caucasian Diff	African-American	Hispanic Diff	Asian Diff	Other Diff
FL	-22.90%	6.68%	10.42%	2.32%	2.48%
CO	-9.06%	0.05%	5.47%	-2.11%	-3.35%
AZ	-4.71%	-2.21%	5.66%	-2.21%	-6.53%
NV	0.77%	-0.74%	0.30%	-2.99%	-5.34%
PA	-16.43%	7.34%	-0.31%	1.61%	-2.21%
MI	-15.03%	14.37%	-3.08%	0.88%	-3.14%
UT	2.67%	-0.82%	-4.04%	-1.29%	-4.22%
GA	-28.73%	22.82%	-2.54%	6.05%	1.40%
OH	-12.91%	11.77%	-1.80%	-0.20%	-2.86%
NC	-36.08%	25.91%	-0.38%	4.17%	2.38%
MT	0.91%	-0.24%	-2.85%	-0.67%	-8.85%
KS	-5.88%	1.35%	-2.87%	0.08%	-3.48%
IA	-1.68%	-0.62%	-2.00%	-0.40%	-2.29%
MN	-16.85%	5.71%	-2.37%	4.21%	-3.68%
TX	-20.95%	5.22%	4.66%	5.26%	-1.89%
NE	-0.98%	2.31%	-4.88%	-1.37%	-3.05%
WI	-21.26%	11.71%	1.20%	1.22%	-2.83%
ME	-0.19%	-0.30%	-1.45%	-0.12%	-2.64%
CA	20.82%	3.48%	-20.02%	-7.28%	-3.00%
NY	-6.61%	5.74%	-1.00%	-2.35%	0.62%

From this table, we can see that VoteAmerica tends to overrepresent African Americans, underrepresent Hispanic, while the representation of Asian is a mix of over and under representations.

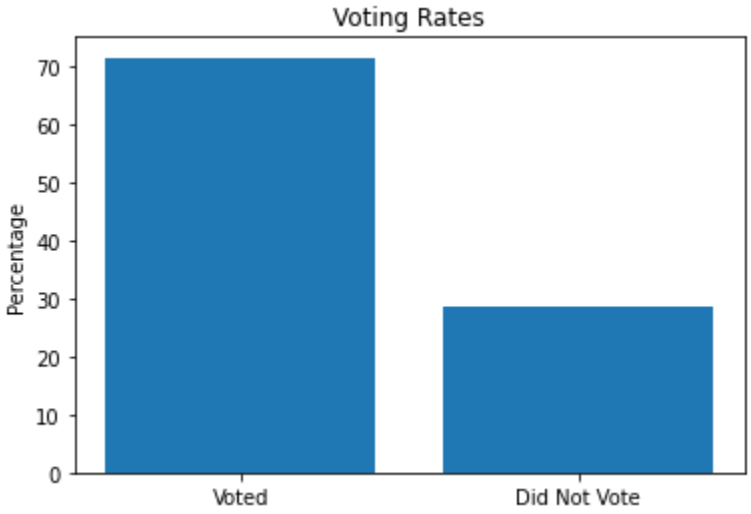
We are informed that VoteAmerica's goal is to overrepresent the minority races, which would include African-American, Hispanic, Asian, and possibly other minority racial groups.We can see that they have been successful in their representation of African-Americans.



People who lived in CO, AZ, NV, and PA voted the most in this campaign, while people from FL, UT, GA, and OH tended to opt out more.

Voting Rates Across Demographics:

1.General Breakdown



Looking at the voting rates of all the voters reached out to, about 71% of them voted.

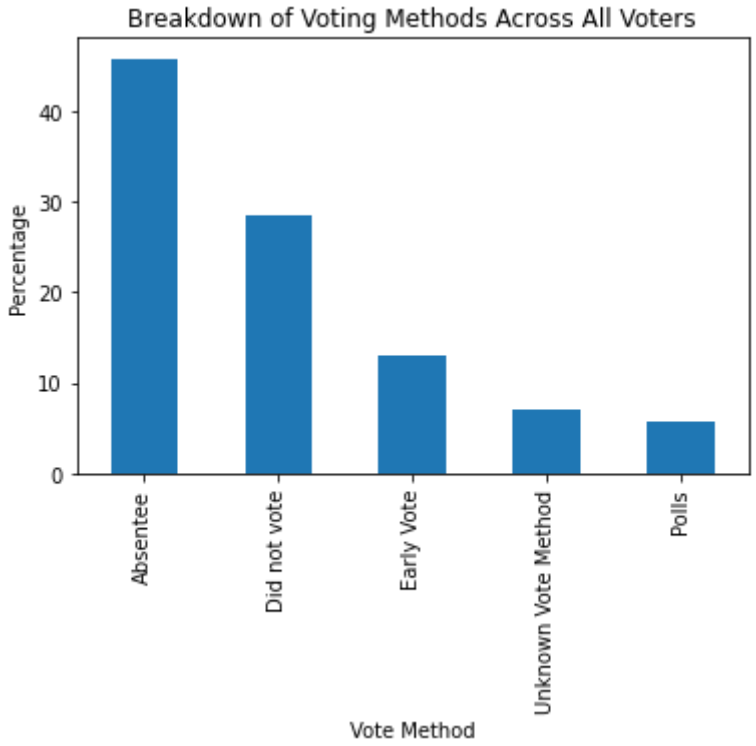
2.Voting Rates of New American Majority Groups

New American Majority Group		Voted	Did Not Vote
African-American		76.3%	23.7%
Asian		81.1%	18.9%
Hispanic		68.2%	31.8%
Native American		62.6%	37.4%
Other		84.3%	15.7%
White or Race Unknown	Unmarried Women 30-49	69.5%	30.5%
	Young Men (18-29)	71.7%	28.3%
	Young Women (18-29)	77.0%	23.0%
	Young, Gender Unknown (18-29)	65.1%	34.9%

In the New American Majority Groups, voters with a race classified as Other had the highest percentage of people who voted, at 84.3%, while Native Americans were the least likely to vote, with only 62.6% of Native Americans voting.

Voting Methods Across Demographics:

1.General Breakdown



Looking generally at the voting methods of all voters contacted, absentee was the most popular voting method (45.85% of voters), followed by not voting (28.47% of voters).

2.Voting Method Breakdown Based on Race:

Race	Absentee	Early Vote	Polls	Did not vote	Unknown Vote Method
African-American	55.23	6.81	5.12	23.95	8.90
Asian	62.40	9.69	4.36	19.27	4.28
Caucasian	44.37	11.95	5.71	28.96	9.03
Hispanic	40.27	19.83	6.60	32.04	1.25
Native American	28.34	24.70	8.10	37.77	1.10
Other	69.07	8.96	4.79	15.89	1.29
Uncoded	40.15	14.17	6.11	30.87	8.70

This table shows the breakdown of voting methods across races. In general, absentee voting was the most popular. However, this was not true for the Native American race. Did not vote was the most common voting method, and Early Vote and Absentee were very close to one another, which was not true for any other race.

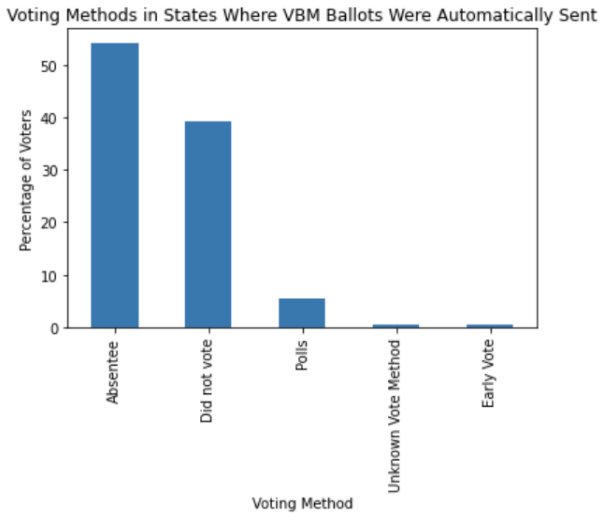
3.Voting Method Breakdown Based on Age:

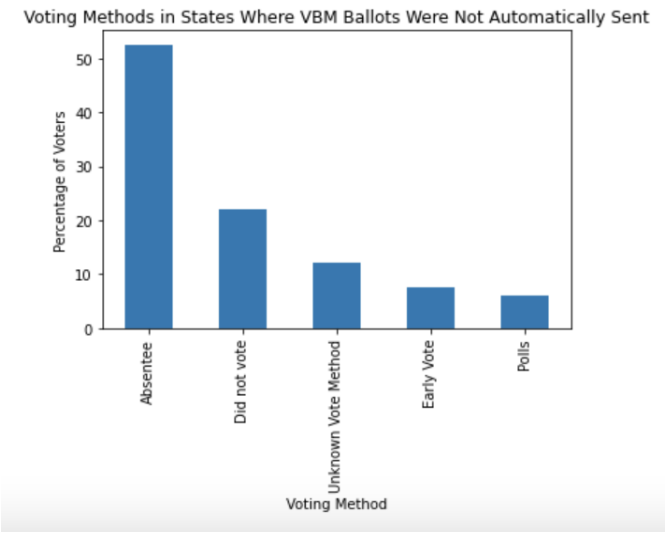
Age Interval	Absentee	Early Vote	Polls	Did not vote	Unknown Vote Method
18-24	46.38	13.40	5.79	25.56	8.87
25-29	43.07	11.84	6.09	32.11	6.90
30-39	42.09	12.94	6.69	32.26	6.02
40-49	45.28	14.12	6.08	28.16	6.37
50-64	51.95	13.67	4.87	23.74	5.76
65+	56.07	11.07	2.73	25.12	5.00

This table shows the breakdown of voting methods based on the voter's age. The results have the same general trend across all age intervals, in which absentee is the most popular voting method, and did not vote was the second most popular.

4.Voting Method Breakdown Based on State:

The following analysis looks at the voting methods in the states of priority. These were then broken into two groups, one where voters were automatically sent VBM ballots (AZ, MI, WI, CO, MT, UT), and the states in which this was not the case (FL, GA, IA, KS, ME, MN, NC, NE, NV, OH, PA, TX). However, the graph below for the states that were automatically sent VBM ballots does not include AZ, as it was an outlier with regards to absentee voting, as shown in the table following these graphs.





Absentee voting was very similar in states where VBM ballots were and were not automatically sent. Voters in states where VBM ballots were automatically sent were also less likely to vote. Absentee and not voting account for the vote methods of 94% of voters in states where VBM ballots were automatically sent, while only accounting for 75% of voters in states where VBM ballots were not automatically sent. Early voting was also very uncommon for voters who lived in states where VBM ballots were automatically sent, only accounting for 0.5% of all voters.

5.Voting Method Breakdown for States Where VBM Ballots Were Sent Automatically

State	Absentee	Early Vote	Polls	Did not vote	Unknown Vote Method
AZ	0.77	63.19	5.94	29.86	0.23
MI	74.46	0.09	9.26	16.07	0.12
WI	68.93	0.25	10.41	20.20	0.21
CO	47.95	0.74	6.24	44.31	0.76
MT	63.94	0.32	1.07	34.31	0.36
UT	45.05	0.43	0.22	53.89	0.43

Looking at the state by state breakdown, Arizona is an outlier with regards to their voting methods, as the voters who voted absentee were such a low percentage of the total voters at only 0.77% of all voters. In all of the other states, absentee was the most popular voting method and in all states not voting was the second most popular voting method.

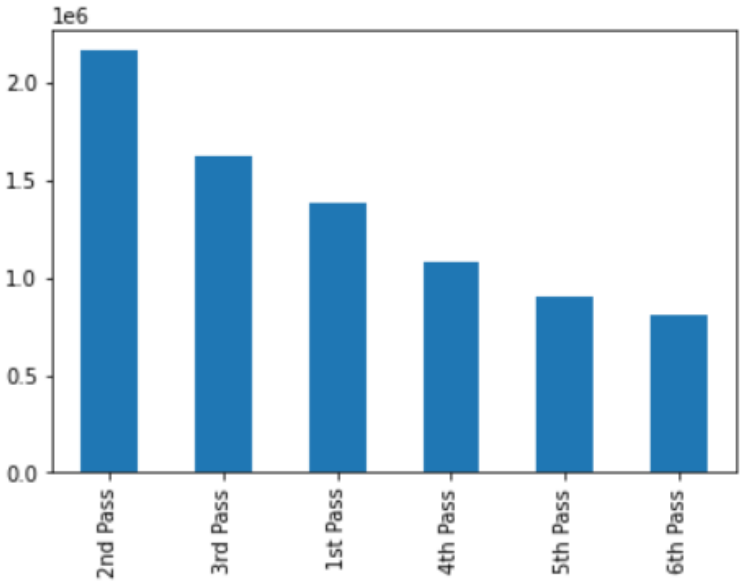
6.Voting Method Breakdown for States Where VBM Ballots Were Not Sent Automatically

State	Absentee	Early Vote	Polls	Did not vote	Unknown Vote Method
FL	67.51	11.58	6.04	14.69	0.17
GA	72.20	0.14	0.27	21.41	5.98
IA	83.52	0.07	5.89	10.40	0.13
KS	0.67	0.13	0.06	10.25	88.90
ME	87.17	0.07	3.38	9.23	0.14
MN	73.68	0.19	12.56	13.49	0.08
NC	52.30	23.42	6.50	17.46	0.32
NE	0.76	0.10	0.12	14.05	84.98
NV	25.72	13.22	9.07	51.35	0.64
OH	0.37	0.07	0.05	13.25	86.26
PA	71.26	0.08	12.60	15.92	0.13
TX	56.98	10.81	2.19	29.38	0.64

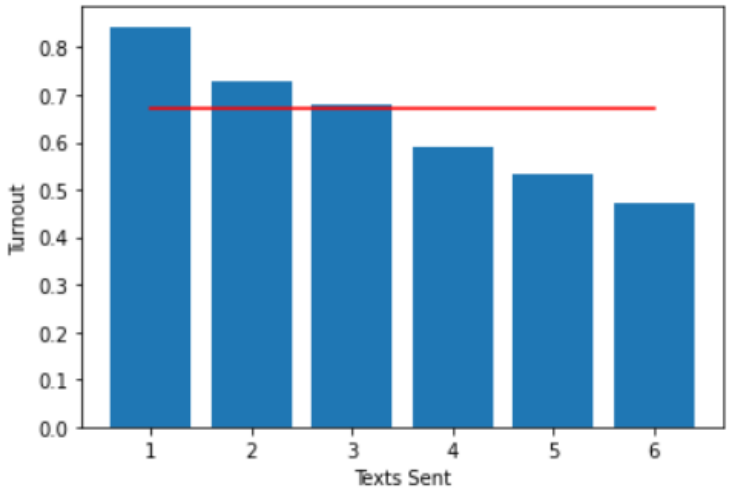
Absentee was generally the most popular vote method among states where VBM ballots were not sent automatically . Outliers were Ohio, Nebraska, and Kansas because the unknown vote method was the most common. Also, in Nevada 51% of voters did not vote, which was much higher than any of the other states.

Voting Trends Related to Texts Sent:

First, we analyzed how many users received one text, two texts, three, etc. The following graph shows raw counts of how many times each pass appeared. We get the somewhat surprising result that by far the most common outcome is a user receiving two texts and opting out. Next most common is three texts before opting out, and opting out after just one text is only the third most common outcome.

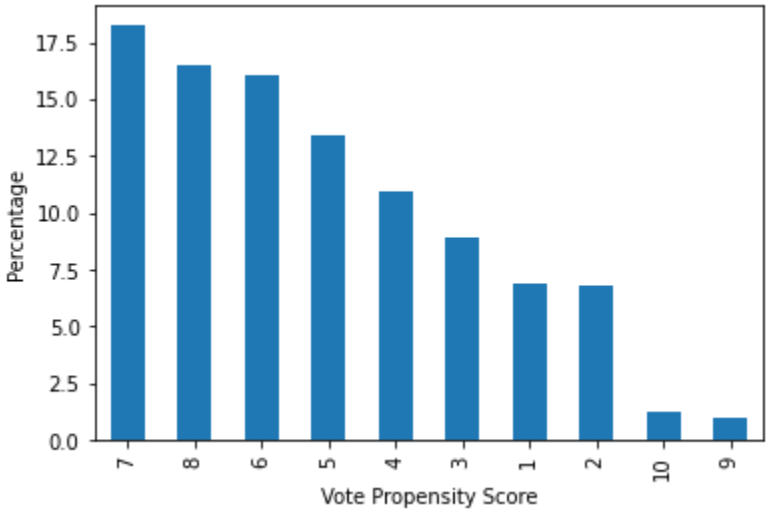


The following graph then shows the correlation between increased numbers of passes and voting rates. The red line is the national turnout for 2020, which was around 67% (https://www.brookings.edu/research/turnout-in-2020-spiked-among-both-democratic-and-republican-voting-groups-new-census-data-shows/). We can see that voters who received a first and second pass outperformed the baseline, while the third pass is very close to baseline, and passes beyond that actually underperform the baseline, suggesting that sending more than three passes might not have a large effect on turnout. This analysis groups all voters who received at least one pass together, including voters who also received further passes.



Vote Method vs. Vote Propensity:

This is a general breakdown of the proportion of voters rated for each vote propensity level.

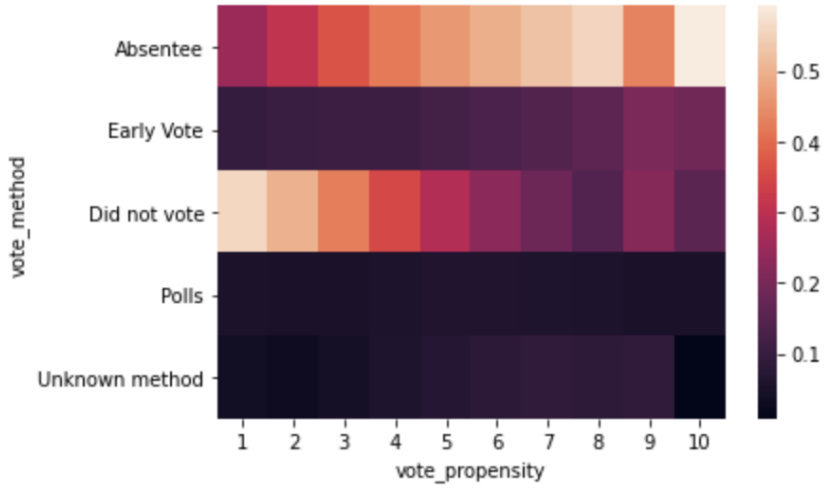


We can see that very few people rated for 10 and 9 in vote propensity (at about 5% combined)

To break down the voting method of voters of each propensity score, one thing to note is that the difference in voters rated with each propensity score is largely different. So, if we break down the vote methods by the counts, it can be largely biased. For example, the counts of any method in scores of 10 and 9 are certainly low, since the total count of voters with scores of 9 and 10 is low.

Thus, when we assessed the vote method of voters with each propensity score, we looked into the percentage that each vote method took, within each propensity level. This can fix the imbalance of total counts of each propensity level.

The following heatmap breaks down the counts for each propensity score. The heatmap should be read vertically, as each column represents the breakdown of the vote methods of the voters with a certain propensity level.



Analysis of the heatmap:

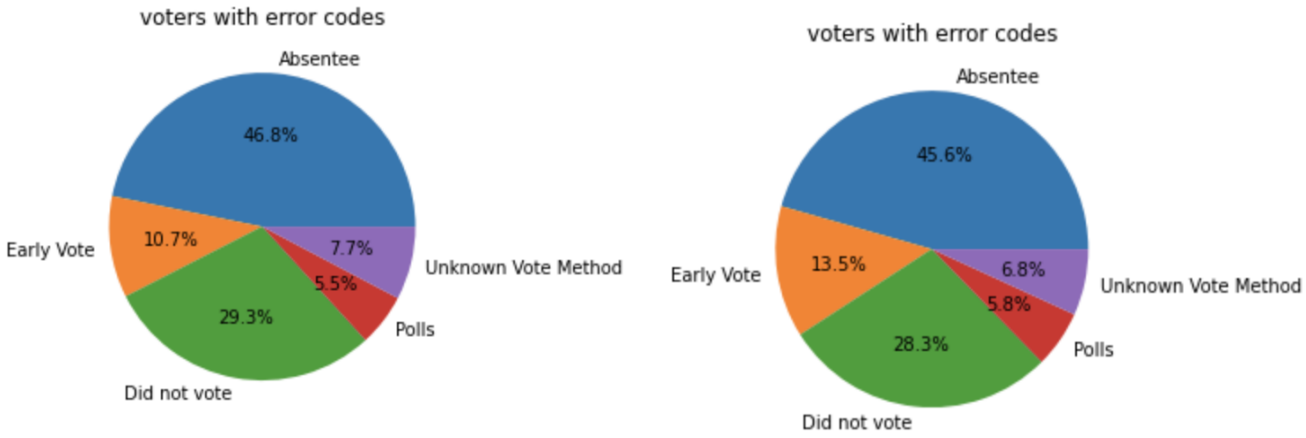
a) Proportion of Absentee, Early Vote and Unknown method goes higher in high propensity scores
   → people with higher propensity scores are more likely to vote by absentee ballot, early voting, and unknown method (slightly)
b) Proportion of Polls is pretty much consistent
   → whether people who prefers to vote at poll would vote doesn't matter with the vote propensity score
   → vote propensity is not an effective measure for Poll voters.
c) Proportion of No information goes down at higher propensity scores
   → people with higher propensity scores are less likely to have no information
   → this confirms that no information can be taken as did not vote.

Vote Methods vs. Contactability:

Note: error codes indicate that the message sent did not reach our targeted voters' phone for some reason.

1. General overview on effect of error codes

523803 out of the total 2,804,686 voters (about 18.68%) experienced at least an error code situation. This also means that the messages reached 81.32% of the targeted voters.



("No vote method information" is treated as "did not vote")

We can see that error codes only avoided 1% voters from voting, which may be a negligible amount even considered against the sample size. However, with previous experience in the racial analysis by states, we understand that the breakdown in each state can be vastly different from the overall figure.

2. Comparing percentage voted with and without error codes in each states

We focus on the states in which we targeted more than 20000 voters.

In the table below, a comparison between percentage voted targeted voters with and without the experience of error codes are shown. The “% extra vote” means how many more percent voted among voters free of error codes, compared to the percentage of voters with the experience of error codes.

	state	% extra vote	ttl				
0	FL	0.019454	573252	**8**	OH	0.011650	144566
1	CO	0.025679	409427	**9**	NC	0.035333	43152
2	AZ	0.052683	381066	**10**	MT	0.040974	33504
3	NV	0.038889	269434	**11**	KS	0.029162	33268
4	PA	0.007797	194373	**12**	IA	0.014166	28730
5	MI	0.002236	184981	**13**	MN	0.011379	27738
6	UT	0.036917	181958	**14**	TX	0.040316	25107
7	GA	0.006909	178208	**15**	NE	0.022423	21003

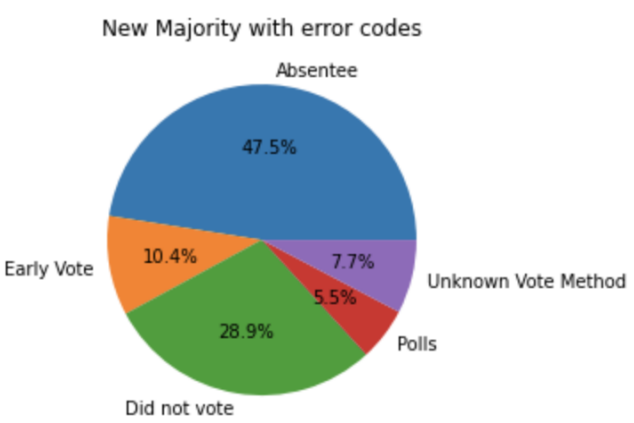
(ttl - total targeted voters in the respective state)
(% extra vote = % voted improvement in targeted voters by free of error code)

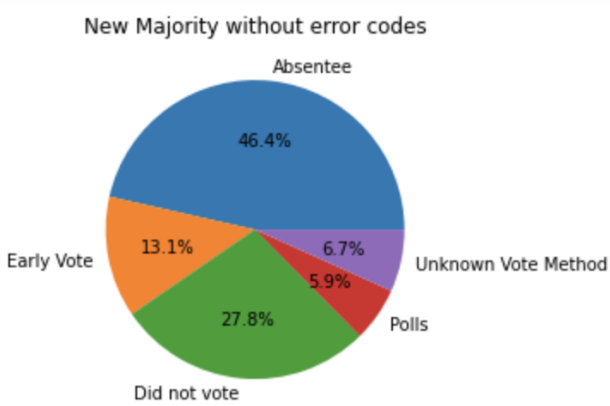
We can see that in targeted voters who receive the text messages, there are at least some percentage increases, from 0.7% to 5.3% increases in voting.

Error codes is an interesting topic to study since having an error code means that the targeted voters who experienced error codes did not receive the text messages. By comparing the breakdown in voters without error codes against the voters with error codes, it can be indicative in estimating the effectiveness in motivating voters. We can rethink the 1% general increase and a 5% maximum in the more targeted states in this way. Then, it means that generally, our messages motivated 1% voters to vote (3-5% in some states)

This can be huge if the scale of the targeted voters is largely increased, this extra percentage voters motivated would increase the vote turnout in some closer contested swing states. And in the above table, we do see some of the swing states. Arizona (5.3% increase), Texas (4%), Nevada (3.9%), Florida (1.9%).

3. Overview of error codes in New Majority Demographics





The size of target New Majority demographics is 2,454,108, as previously calculated, taking up about 84% of the total targeted voters

Among these about 2.5 million targeted New Majorities, 1,990,553 reached the text messages free of error codes, which indicates about 81% of targeted New Majorities being reached.

The breakdown of the vote method and whether they vote is the same as the overall population. Thus, we would not break this down to specific states.