Part 2: Voting in 2020 Elections in North Carolina

Summary

The North Carolina State Board of Elections (NCSBE) has provided user voter registration and turnout data online. This project aims to estimate how different demographic groups voted in the 2020 national elections. Additionally, this study seeks to study the odds of voting across counties.A substantial effort went into data pre-processing since the original data set provided by NCSBE comprised of two separate documents which had county level data of aggregated counts of registered voters and aggregated counts of registered voters who actually voted in North Carolina by demographic groups. Before carrying out Exploratory Data Analysis (EDA), we randomly selected 25 counties out of 100 after we cleaned and merged the data sets.

We decided on fitting a random intercept and random slope hierarchical model that returned an AIC value of 49740.6, which was substantially less than all other models. Also, age had a different effect on the odds of voting by county and the baseline turnout is different for each county. The summary of the model can be found in the model section. The baseline voter for the final model was a White, Non-Hispanic/Latino, male voter who was affiliated with the Republican party and was between the ages of 18 and 25. The odds of turning out to vote for a White, Non-hispanic/Latino, Republican male voter between the ages of 18 and 25, in the 2020 elections was $2.01(e^{0.69})$.

The final model generated the following answers to the questions of interest. Keeping age, ethnicity, race, political party affiliation and all interactions at baseline, a female voter had 1.037 times ($e^{0.037}$) the odds of turning out to vote compared to a male voter. Voters belonging to the undesignated subgroup had 1.61 times (61% increase) the odds of turnout for male voters. For race, ethnicity, sex, political party affiliation and interactions at baseline levels, a voter over the age of 66 had 2.94 times (194% increase) the odds of voting compared to a voter in the 18 to 25 age bracket, whereas a voter in the 41 to 65 age bracket had 2.96 (196% increase) times the odds of voting compared to voters in the baseline age. Voters in the 25 to 40 age bracket had 1.29 (29% increase) times odds of turning out to voters in the baseline age. Keeping race, sex, ethnicity and age and all interactions at baseline, a voter affiliated with the Democratic party had 0.76 (24% decrease) times the odds of voting compared to a voter affiliated with the Republican party. Additional answers to the inferential questions and limitations of this study have been outlined in conclusions of this report.

Introduction

As the agency mandated to administer elections process, campaign finance disclosure and compliance, the North Carolina State Board of Elections (NCSBE) has provided user voter registration and turnout data online. This projects seeks to estimate how different demographic groups voted in the 2020 elections, using the above-mentioned data. Specifically, the project seeks to answer the following questions using a hierarchical model:

- How did different demographic subgroups vote in the 2020 general elections? That is, how did the turnout for one demographic subgroup compare with other demographic subgroups, controlling for other potential predictors?
- Did the overall probability or odds of voting differ by county in 2020? Which counties differ the most from other counties?

- How did the turnout rates differ between female and males for the different party affiliations?
- How did the turnout rates differ between age groups for the different party affiliations?

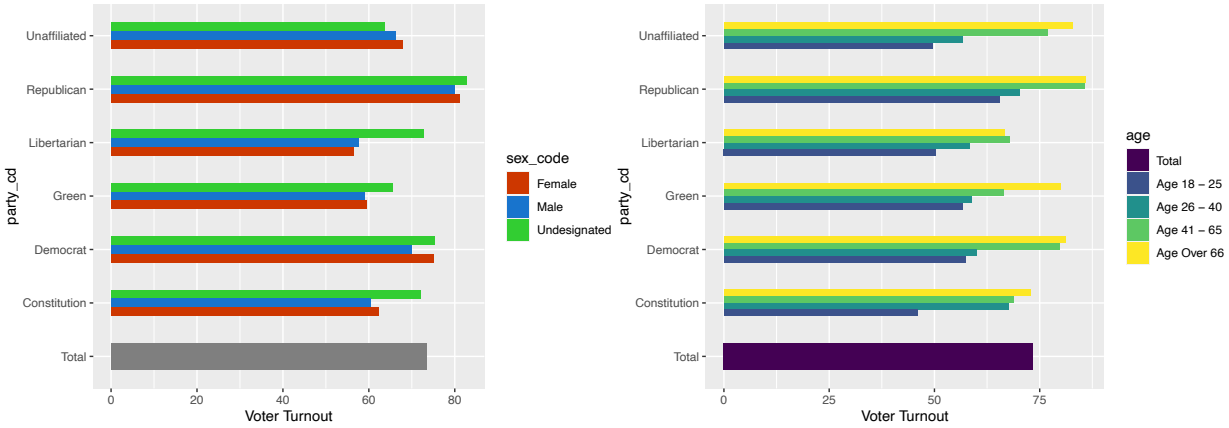
Data

We observed 11,164 observations in the *voters_stats_20201103.txt* data set and we identified the possibility of 43,200 possible unique combinations of demographic variables. It is only natural to assume that the demographic combinations missing from our data did not have any voters in them and thus the number of registered voters is zero for these observations. This is further evidenced by the fact that the minimum number of registered voters within demographic combinations is at least one. The same argument can be extended to *history_stats_20201103.txt* data set and it is natural to assume that the demographic combinations provided in *voters_stats_20201103.txt* data set but not in *history_stats_20201103.txt* data set had number of votes equal to zero. This enables us to include in our analysis, people who registered in one county or for a party and voted in another county or for another party, of which we have already seen evidence of after merging the data sets. Left join generated a more balanced data set that reflects the peculiar nuances of working with elections data. Before carrying left join of the two data sets, we had 11,164 observations in *voters_stats_20201103.txt*, which is the "left" data set, and 9,760 observations in the *history_stats_20201103.txt*. After the left join we had 11,164 observations in our merged data set.

To ensure clarity and interpretability of our analysis, we selected a random sample of 25 counties out of the 100 counties that existed in the data set. We also dropped categorical variables at the precinct voting district levels because we had scant data across all precincts because of their number, and they were also not necessary in answering the questions of interest. These twenty (25) randomly selected counties are Caldwell, Yancey, Stokes, Pender, Perquimans, Guilford, Greene, Granville, Martin, Robeson, Clay, Alamance, Jackson, Polk, Lincoln, Washington, Cherokee, Northampton, Davie, Jones, Hyde, Halifax, Warren, Micthell and Onslow.

Overall, voter turnout for the 2020 elections in North Carolina was 73%. This is above the previous record of 69.6% set in 2008, when Barack Obama carried the state. At the state-level, the highest turnout of 78% was recorded by Republicans, while Libertarians recorded the least voter turnout of 60%. Republicans have carried North Carolina in 11 of the last 13 presidential elections. The only two presidents to have won the state are Jimmy Carter in 1976 and Barack Obama in 2008. The Democratic party had a turnout of approximately 74%. Across races, pacific islanders had the strongest showing, with a turnout rate of 79%.

Interactions
We explored the effect that interaction between demographic subgroups had on voter turnout. Four out of five 5 (81%) female Republicans voted in the 2020 elections, a record for the demographic subgroup. The previous record for the subgroup was in the 2016 elections and it stood at 76%. Similarly, the highest turnout of males across all political parties, 80%, was from the Republican party.



Across all political parties, there was a trend of voters in the oldest age group turning out to vote in higher proportions than voters in the immediate younger age groups. Since the relationship between turnout and political party varied by age groups we considered this interaction during modeling.

Model

Initially we fitted a model with a hierarchical model with varying intercept by county and all predictor variables in the data set including interactions between age and political party affiliation and sex and political party affiliation. This model had a Akaike Information Criterion (AIC) value of 51110.9. We explored other models to see if we could achieve a lower AIC value. We decided on fitting a random intercept and random slope hierarchical model that returned an AIC value of 49740.6, which was substantially less than all other models. The random intercept by county and random slope by age were chosen because age has a different effect on the odds of voting by county and the baseline turnout is different for each county. To answer questions of interest, and model findings made during EDA, we included the interaction between sex and party affiliation along with the interaction between age and party affiliation in our final model. The summary of the model can be found on page 4 for easy referencing.

Mathematically, the final model can be expressed as:

$y_{ij}|x_i \ Bernoulli(\pi_i); \ \ i = 1, ..., n; j = 1, ..., 25(selcted \ counties); k = number \ of \ levels \ categorical \ variable$

$log(\frac{\pi_i}{1-\pi_i}) = (\beta_0+\gamma_{0j})+(\beta_{1k}+\gamma_{1jk})x_{i1j}(age)+\beta_{2k}x_{i2j}(politicalparty)+\beta_{3k}x_{i3j}(race)+\beta_{4k}x_{i4j}(ethnicity)+\beta_{5k}x_{i5j}(sex) + \beta_{6k}x_{i1j}x_{i2j}(age : politicalparty) + \beta_{7k}x_{i2j}x_{i5j}(politicalparty : sex)$

$(\gamma_{0j} + \gamma_{1j,k=1} + \gamma_{1j,k=2} + \gamma_{1j,k=3}) \sim N_4(\mathbf{0}, \Sigma)$

The baseline voter for the final model was a White, Republican, Non-Hispanic/Latino, male voter who was in the 18 to 25 age bracket.The odds of turning out to vote for a White, Non-hispanic/Latino, Republican male voter between the ages of 18 and 25 was $2.01(e^{0.69})$. The county-level standard deviation at the random intercept of the final model is estimated at 0.26, so voter turnout in counties do vary some, but not so much. A total of 0.56 describes the within-county standard deviation of turnout attributable to the random slope of age.

The following are inline answers to the questions of interest, explained by our final model.

How did different demographic subgroups vote in the 2020 general elections? For example, how did the turnout for males compare to the turnout for females after controlling for other potential predictors?

Sex: Keeping age, ethnicity, race, political party affiliation and all interactions at baseline, a female voter had 1.037 ($e^{0.037}$) (4% increase) times the odds of turning out to vote compared to a male voter. Voters belonging to the Undesignated subgroup had 1.61 (61% increase) times the odds of turnout for male voters.
Age: For race, ethnicity, sex, political party affiliation and interactions at baseline levels, a voter over the age of 65 had 2.94 times the odds of voting compared to a voter in the 18 to 25 age bracket, whereas a voter in the 41 to 65 age bracket had 2.96 times the odds of voting compared to voters in the baseline age. Voters in the 25 to 40 age bracket had 1.29 times odds of turning out to voters in the baseline age.
Political Party Affiliation: Keeping race, sex, ethnicity and age and all interactions at baseline, a voter affiliated with the Democratic party had 0.76 times the odds of voting compared to a voter affiliated with the Republican party. Compared to the Republican party, the Constitution, Green, Libertarian and Unaffiliated parties had 0.54, 0.73, 0.55, 0.56 times the odds of turning out to vote compared to the Republican party, respectively.
Race: For age, sex, ethnicity, political party affiliation and interactions at baseline levels, a Black voter had 0.69 times the odds of voting compared to a white voter. Asian, Indian American, Pacific Islanders and other races had 0.83, 0.68, 3.06, 0.63 times the odds of turning out to vote compared to White voters, respectively. For voters who belonged to two or more races, the odds of turning out to vote was 0.68 times the odds of voting for White voters.

Model Summary

Fixed Effects	Estimate	z-value	p-value	Fixed Effects	Estimate	z-value	p-value
(Intercept)	0.699	12.943	2.00E-16	ageAge 26 - 40:party_cdDem	-0.107	-5.636	0.000
ageAge 26 - 40	0.254	8.373	2.00E-16	ageAge 41 - 65:party_cdDem	-0.017	-0.600	0.336
ageAge 41 - 65	1.089	28.384	2.00E-16	ageAge Over 66:party_cdDem	-0.024	-1.226	0.220
ageAge Over 66	1.080	19.739	2.00E-16	ageAge 26 - 40:party_cdGreen	-0.142	-0.661	0.509
party_cdConstitution	-0.610	-3.663	2.45E-04	ageAge 41 - 65:party_cdGreen	-0.675	-2.718	0.007
party_cdDemocrat	-0.276	-16.639	2.00E-16	ageAge Over 66:party_cdGreen	0.070	0.103	0.918
party_cdGreen	-0.311	-1.548	1.21E-01	ageAge 26 - 40:party_cdLibert	0.088	1.411	0.158
party_cdLibertarian	-0.589	-10.373	2.00E-16	ageAge 41 - 65:party_cdLibert	-0.408	-5.715	0.000
party_cdUnaffiliated	-0.583	-38.650	2.00E-16	ageAge Over 66:party_cdLibert	-0.441	-3.384	0.001
race_codeAsian	-0.190	-9.180	2.00E-16	ageAge 26 - 40:party_cdUnaffil	0.063	3.616	0.000
race_codeBlack	-0.380	-58.520	2.00E-16	ageAge 41 - 65:party_cdUnaffil	0.066	3.902	0.000
race_codeIndian American	-0.392	-26.853	2.00E-16	ageAge Over 66:party_cdUnaffil	0.372	18.449	0.000
race_codeOther	-0.474	-32.165	2.00E-16	party_cdConst:sex_codeFemale	0.042	0.222	0.824
race_codePacific Islander	1.122	3.012	3.00E-03	party_cdDem:sex_codeFemale	0.230	19.958	0.000
race_codeTwo or more	-0.386	-15.384	2.00E-16	party_cdGreen:sex_codeFemale	0.001	0.007	0.995
race_codeUndesignated	-0.068	-5.554	2.78E-08	party_cdLib:sex_codeFemale	-0.019	-0.362	0.718
ethnic_codeHispanic/Latino	-0.342	-24.236	2.00E-16	party_cdUnaffil:sex_codeFemale	0.053	4.555	0.000
ethnic_codeUndesignated	-0.062	-10.067	2.00E-16	party_cdConst:sex_codeUndesig	-0.086	-0.435	0.664
sex_codeFemale	0.037	4.220	2.44E-05	party_cdDem:sex_codeUndesig	-0.044	-1.613	0.107
sex_codeUndesignated	0.478	21.229	2.00E-16	party_cdGreen:sex_codeUndesig	-0.200	-0.790	0.430
ageAge 26 - 40:party_cdConst	0.548	2.761	6.00E-03	party_cdLib:sex_codeUndesig	0.315	3.724	0.000
ageAge 41 - 65:party_cdConst	-0.266	-1.307	1.91E-01	party_cdUnaffil:sex_codeUndesig	-0.355	-15.062	0.000
ageAge Over 66:party_cdConst	-0.209	-0.492	6.22E-01				

Random Effects

Groups	Name	Variance	Std. Dev
county_desc	(Intercept)	6.70E-02	0.260
	ageAge 26 - 40	1.60E-02	0.127
	ageAge 41 - 65	3.00E-02	0.174
	ageAge Over 66	6.70E-02	0.259

Akaike Information Criterion (AIC)

49740.6

Ethnicity: Keeping race, sex, age, political affiliation and interactions at baseline levels, a Hispanic/Latino voter had 0.71 times the odds of voting than a Non Hispanic/Latino voter whereas a voter belonging to an Undesignated ethnicity had 0.94 times the odds of voting compared to the baseline ethnicity.

Did the overall probability or odds of voting differ by county in 2020? Which counties differ the most from other counties?

A dot plot of our final model, which can be found in the appendix, visualizes the county by county differences in the odds of voting. Yes, the overall odds of voting did differ by county in 2020 and it differed significantly. From the dot plot, we observed that the probability for voting was significantly lower in the following seven (7) counties: Onslow, Robeson, Washington, Cherokee, Halifax, Clay and Northamption. The odds of voting was not significantly different in the following seven (7) counties: Perquimans, Warren, Jones, Pender, Martin, Polik and Greene. In the following remaining eleven (11) counties, voter turnout was significantly higher than other counties: Caldwell, Guilford, Jackson, Lincoln, Stokes, Davie, Alamance, Granville, Hyde, Mitchel and Yancy.

How did the turnout rates differ between females and males for the different party affiliations?

The baseline for the interaction between sex and party affiliation was a male Republican. Keeping race,

ethnicity, age and other interactions at baseline, the odds of a female affiliated with the Constitution party voting was $0.59(e^{(0.037+(-0.61)+0.0424))})$ times the odds of voting of a male affiliated with the Republican party. Compared to the baseline, the odds of a female affiliated with the Democratic party voting was $0.99(e^{(0.037+(-0.28)+0.23))})$ times or 1% less. The chances of a female Green party member voting was 0.76 times the odds of voting of a male Republican. The odds of female affiliated with the Libertarian party voting was 0.56 times the odds of voting of a male Republican. Compared to the baseline, the chances of a voter, of undesignated sex affiliated with the Constitution party, turning out to vote was 0.80 times or 20% less. The odds of a voter of undesignated sex affiliated with the Democratic party voting was 1.17 the odds of a male Republican voting. For a person of undesignated sex who is affiliated with the Green party, the odds of turning out to vote was 0.98 times the odds of voting of a male Republican. Compared to the baseline, the odds of voting for a person of undesignated sex affiliated with the Libertarian party is 1.23 times or 23% more. Lastly, the odds of voting for a person of undesignated sex not affiliated with any political party is 0.64 times or 36% less than the odds of voting for a male Republican.

How did the turnout rates differ between age groups for the different party affiliations?

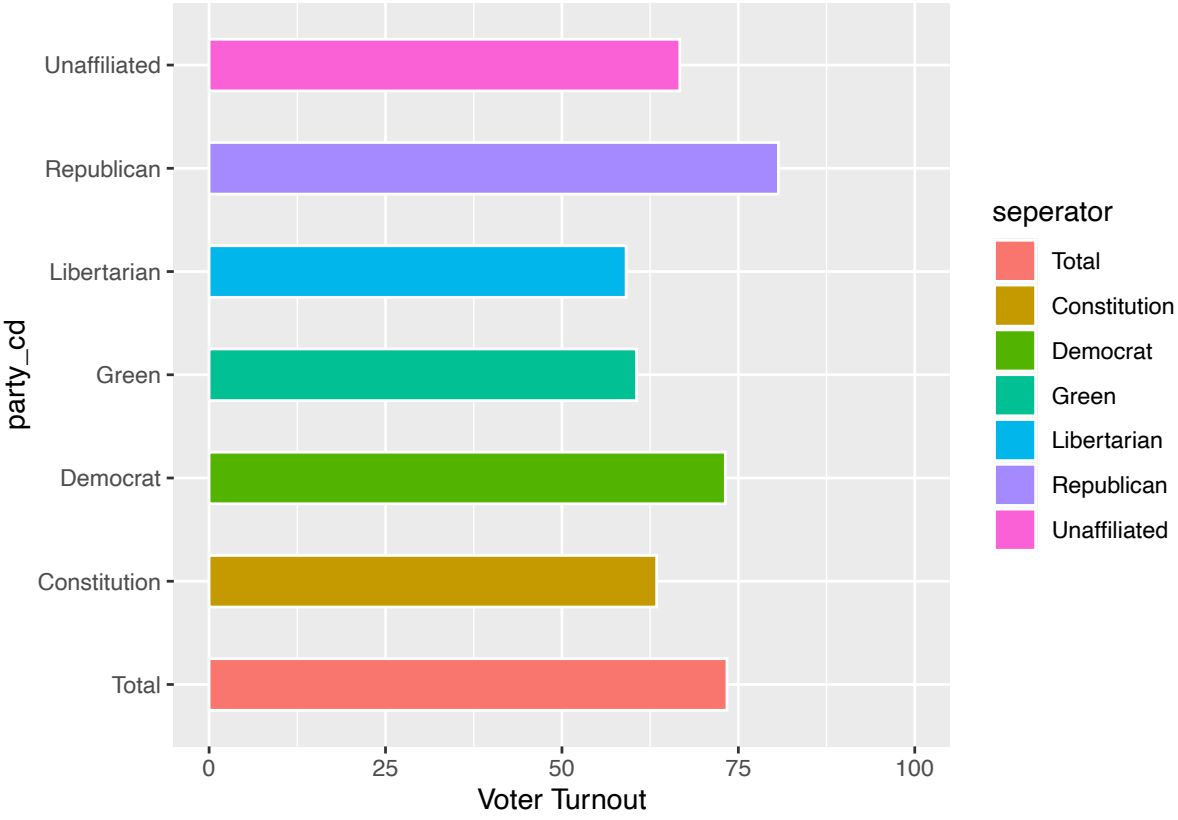
The baseline for the interaction between age groups and different party affiliations was a person affiliated with the Republican party in the 18 to 25 age bracket. Keeping race, ethnicity, sex and other interactions at baseline, the odds of a Democratic voter between the ages of 26 and 40 turning out to vote was $0.87(e^{(0.25+(-0.28)+(-0.11))})$ times, or 13% less than, the odds of voting of a Republican voter in the 18 to 25 age group. Compared to the baseline, the odds of voting of a Democratic voter in the 41 to 65 age bracket is 2.21 times or 121% more. For a Democratic voter over the age of 66, the odds of turning out to vote was 2.17 times the odds of voting for a Republican between the ages of 18 and 25. For a voter affiliated with the Green party between the ages of 26 and 40, the odds of voting was 0.82 times the odds of a Republican in the 18 to 25 age group. On the other hand, the odds of voting of a Green party member in the 41 to 65 age bracket was 1.11 times the odds of voting of a Republican party member in the 15 to 25 age bracket. Compared to the baseline, the chances of a Green party member over 66 years voting was 2.31 times or 131% more. The odds of a Libertarian in the 26 and 40 age bracket voting was 0.78 times the odds of voting of a Republican in the 18 and 25 age bracket. For a Libertarian in the 41 and 65 age bracket, the odds of voting was 1.09 times the odds of voting of the baseline subgroup. For Libertarians over the age of 66, the chances of voting was 1.05 times the chances of a Republican in the 18 to 25 age bracket. For a voter affiliated with the Constitution party between the ages of 26 and 40, the odds of voting was 1.21 times the odds of a Republican in the 18 to 25 age group. Compared to the baseline, the chances of a Constitution party member in the 41 to 65 age bracket voting was 1.23 times or 23% more. For a voter affiliated with the Constitution party between over the age of 66, the odds of voting was 1.30 times the odds of a Republican in the 18 to 25 age bracket. For a person in the 26 to 40 age bracket who was politically unaffiliated, the odds of turning out to vote was 0.77 times the odds of voting of a male Republican. Compared to the baseline, the odds of voting for a person who is politically unaffiliated in the 41 to 65 age bracket is 1.78 times or 78% more. Lastly, the odds of voting for a person not affiliated with any political party over the age of 66 was 2.39 times or 139% more than the odds of voting for a male Republican.

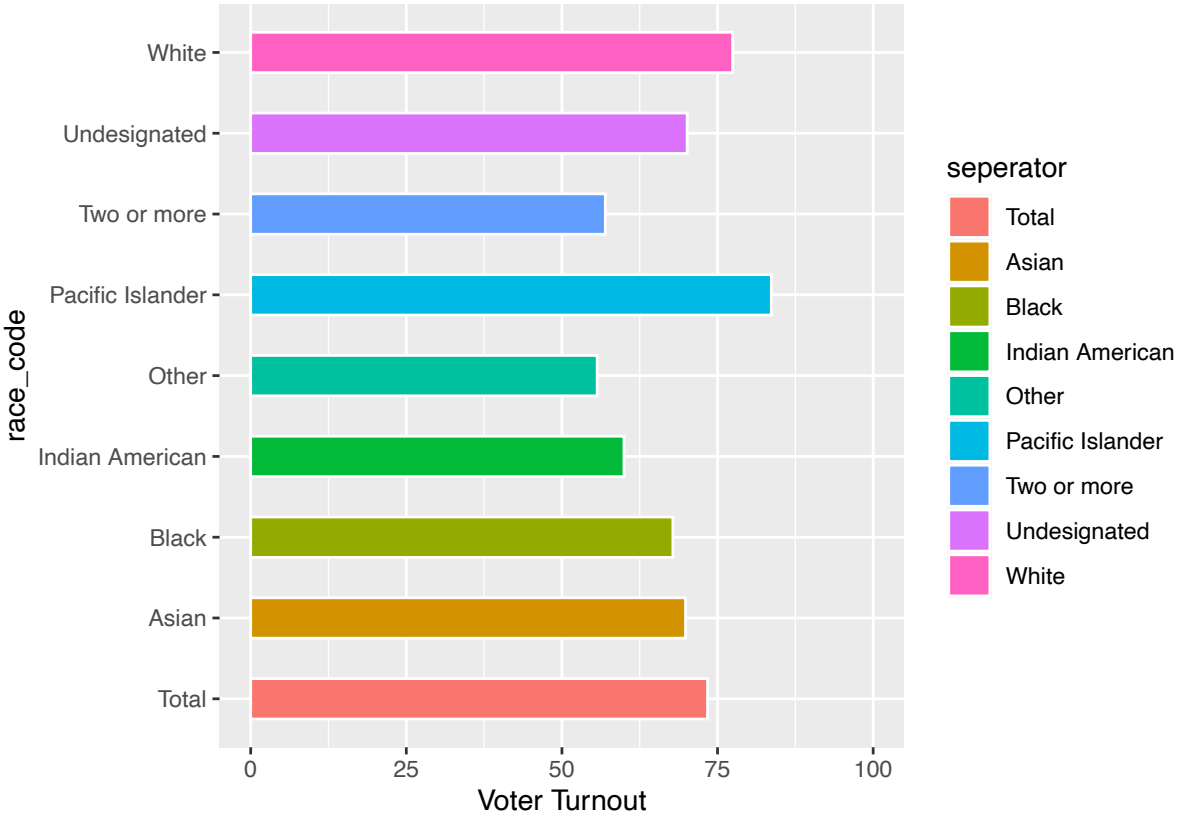
Conclusion & Limitations

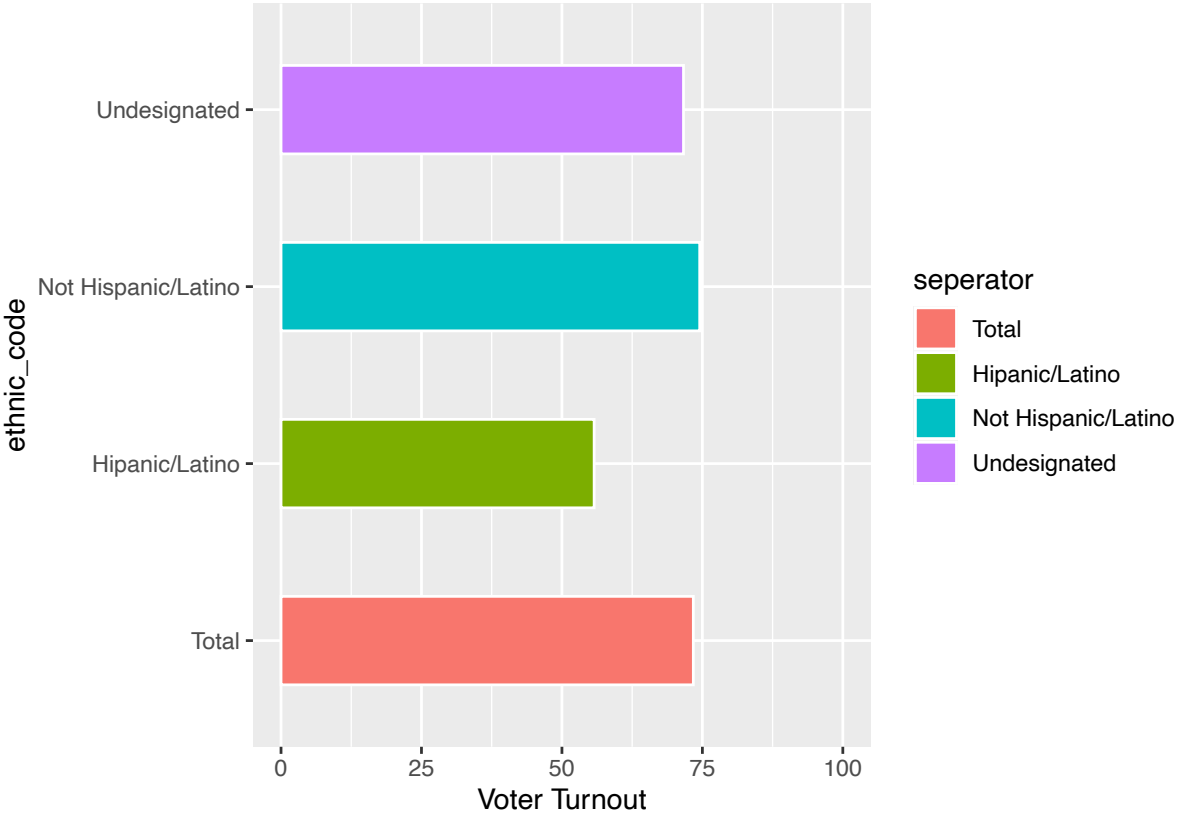
As mentioned earlier, the nature of the data set presented some noteworthy challenges. There were observations which did not have aggregated entries for the number of registered voters who actually voted. Even though we had evidence that this was partly attributable to the fact that some individuals voted in counties in which they did not register, there are other unexplained reasons for this occurrence. We did not have enough information to make corrections to the data set. To make the data set ready for analysis, we were forced to set a cut off voter turnout limit of 100% for the counties which recorded a turnout greater than 100%. Also, the fact that same day registration of voters further limited the accuracy of this analysis.Finally, our analysis is based on 25% of the 100 counties in North Carolina so there are limitations on the inferences that can be drawn from this study.

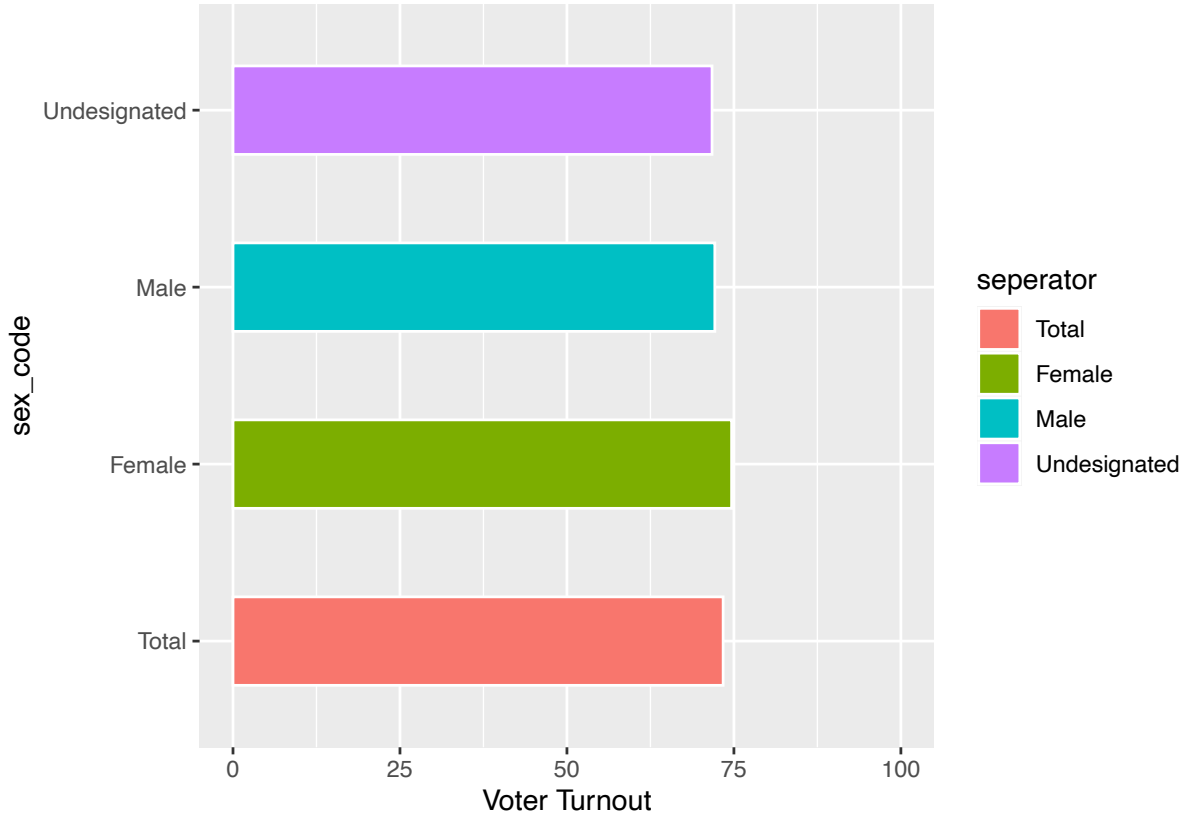
Appendix

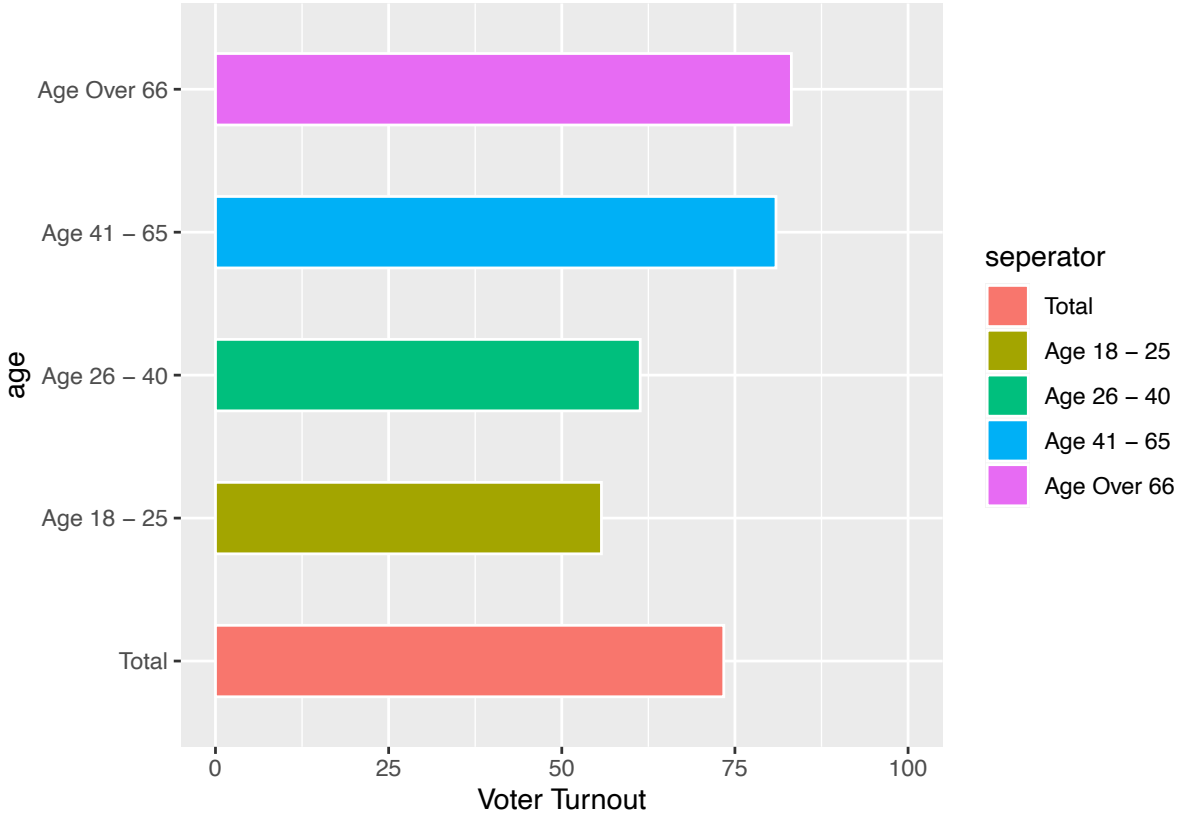
Visualization of Associative Effect of Demographic Groups on Voter Turnout




White
Undesignated
Two or more
Pacific Islander
Other
Indian American
Black
Asian
Total
race_code
0
25
50
75
100
Voter Turnout
seperator
Total
Asian
Black
Indian American
Other
Pacific Islander
Two or more
Undesignated
White


Undesignated
Not Hispanic/Latino
Hipanic/Latino
Total
ethnic_code
0
25
50
75
100
Voter Turnout
seperator
Total
Hipanic/Latino
Not Hispanic/Latino
Undesignated


Undesignated
Male
Female
Total
sex_code
0
25
50
75
100
Voter Turnout
seperator
Total
Female
Male
Undesignated


Age Over 66
Age 41 – 65
Age 26 – 40
Age 18 – 25
Total
age
0
25
50
75
100
Voter Turnout
seperator
Total
Age 18 – 25
Age 26 – 40
Age 41 – 65
Age Over 66

Dot Plot of Final Model

$county_desc