Case Study 2

Emily Gentles (Writer)   Weiyi Liu (Coordinator and Checker)   Jack McCarthy (Presenter)
Qinzhe Wang (Programer)

06 December, 2021

Introduction

Voter turnout is a huge topic of interest for both pollsters and candidates, especially around presidential elections and with a particular focus on swing states, such as North Carolina. In this case study we will use North Carolina registration and turnout data, by county, for the 2020 election year. We wish to investigate what demographic factors such as age, race, and party affect turnout and whether turnout differs by county.

Data Cleaning

Note that we renamed the `history_stats_20201103` data as votes and the `voter_stats_20201103` data as registers. Within the votes and registers data we aggregated the total voters variable by taking the sum, ignoring NA values, and then left joined votes to registers on county and all the demographic features. We next replaced any NA values in total votes with 0. While these values are unlikely to be exactly 0, we chose to do this in order to preserve the information that was present as the alternative would be to throw the entire row away. After this we dropped all of the rows containing NA values in any of the other columns. Since total voters is one of the main variables we care about it made sense to preserve as many rows as possible but if we replaced every NA value in the data with 0 that may potentially introduce a lot of error into the data, as the values are unlikely to be exactly 0, so we instead dropped all of the rows containing NA values, leaving us with 51,883 rows. For some subgroups that have more actual voters than registered voters, we replaced the number of actual voters by the number of registers to ensure the turn out rate to be valid. We next randomly sampled 30 counties to utilize in our analysis, leaving our final data with 16,439 rows and, after de-aggregation, 1,908,907 individuals.

EDA

Below we see that most counties have a turnout rate over 70% with the average of all counties being 73%. The lowest turnout rate of 61% is from Robeson county. We also see that two counties, Forsyth and Cabarrus, have many more registered voters than the other sampled counties.

When looking at race, age, gender, ethnicity, and party individually (plots in appendix), we found that white people had the highest turnout and number of voters registered while multiracial people had the lowest turnout rate, along with people of race 'other', and the smallest number of voters registered. We also found the race 'P' has an extremely small sample size (115) compared with other race groups. Therefore, we decided to drop this group for computational efficiency.

The Republican party had the highest turnout while the Libertarian party had the lowest turnout. Both the Green party and the Constitution party had the lowest number of voters registered, less than 1500 each, so we combined these categories to provide our model with computational stability.

When looking at ethnicities, people of Hispanic/Latinx ethnicity had both the lowest number of voters registered and turnout, a 20 point difference compared to people who weren't of Hispanic/Latinx ethnicity.

Turnout rate varied very little by gender although people of unknown gender had the smallest number of voters registered.

Looking at age we see that the age group of people 18-25 had both the lowest number of voters registered and turnout rate, a difference of 27 points when compared to those 66 and older. However, the age group of people 41-65 had the largest number of voters registered, nearly triple that of the 18-25 age group.

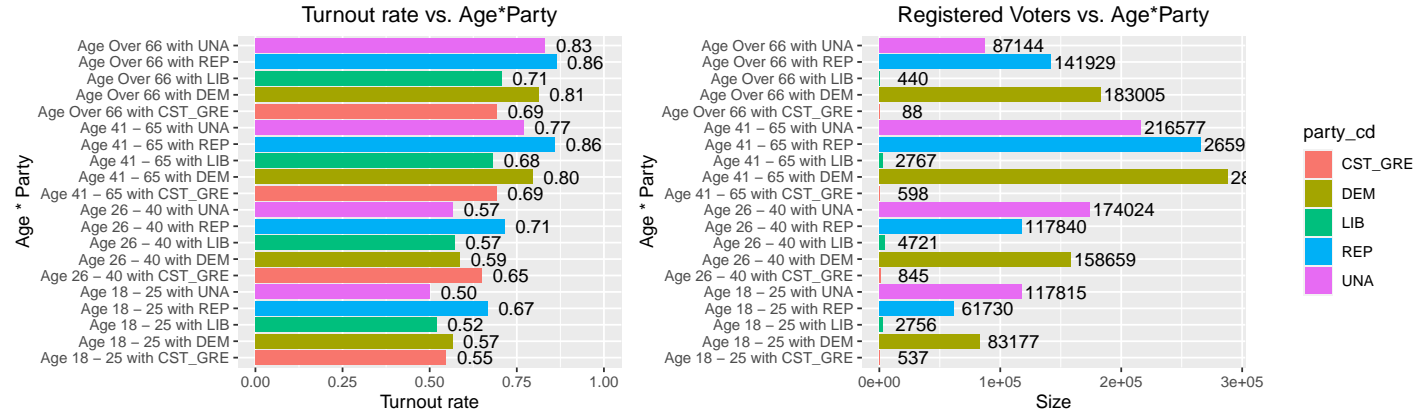## Turnout Rate by County
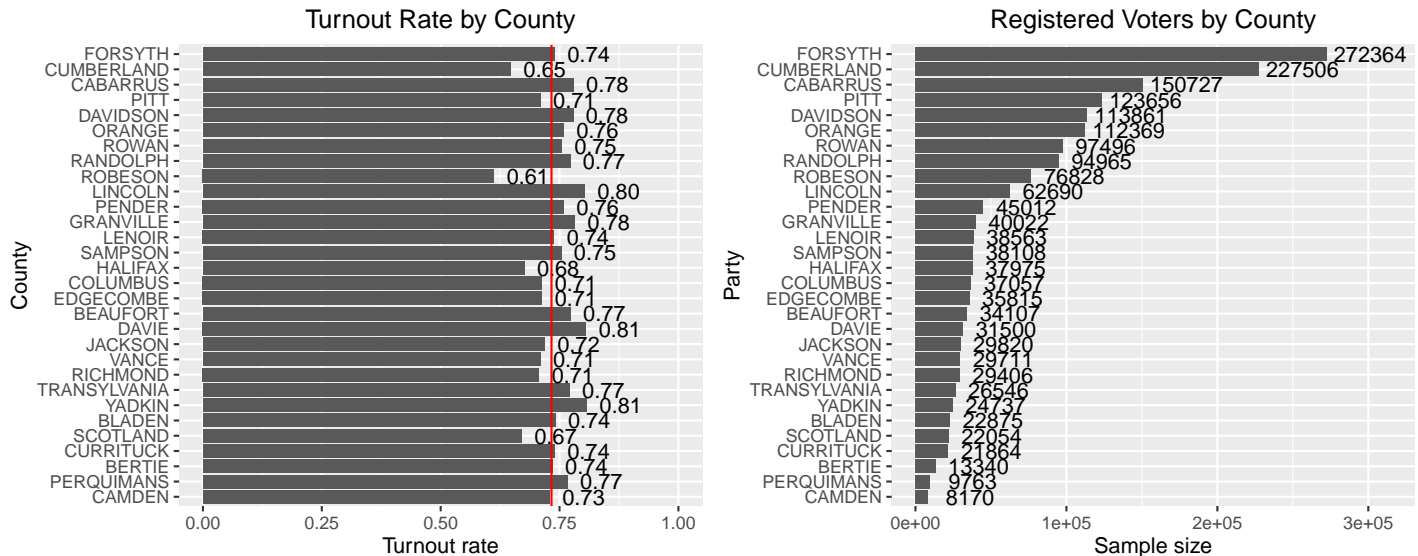
Gender & Party

Next we move on to EDA related to interactions. We're interested in any potential differences in turnout rates for females and males with various party affiliations. In the graph below we see that females have higher turnout rates than males for each party. People with unknown genders, however, often have the highest turnout rate. For parties with substantial turnout rates, the Republican and Democrat parties, as well as those unaffiliated with a party, females had a larger number of voters registered than males and people with unknown genders had a much smaller number of voters registered. Note that, even after combining the Green and Constitutional parties, the number of registered voters for these parties is still quite small, relative to the other parties.
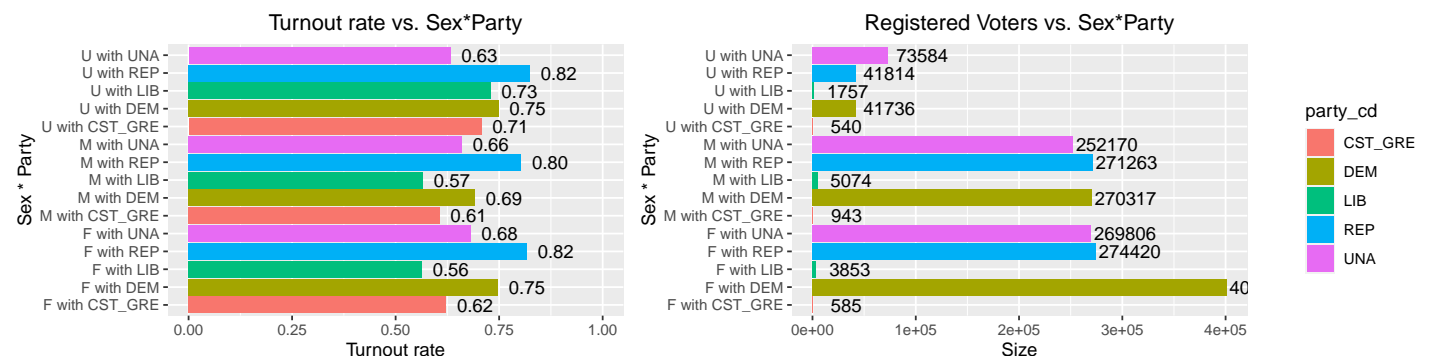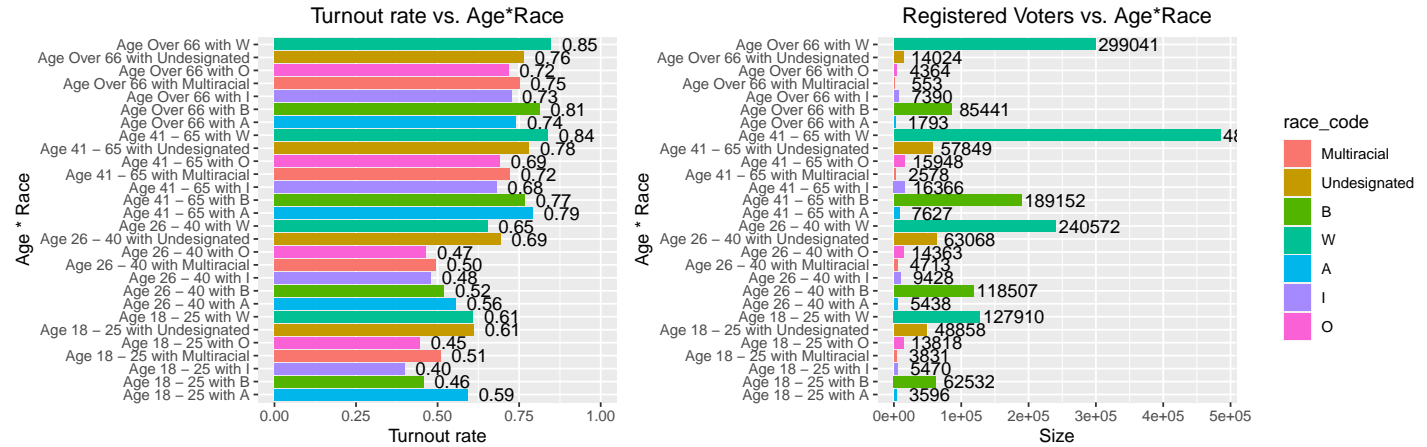

Age & Party

We are also interested in the potential difference between age groups in various political parties. From the graphs below we see that the age group of 66 and older have the highest turnout rate for every political party and the age group 18-25 has the lowest. In general, it seems that, across parties, as age increases, the turnout

rate also increases. As noted previously, the age group 41-65 has the largest number of registered voters for both of the main parties, Republican and Democrat, and for unaffiliated voters.

Turnout rate vs. Age*Party

Age Over 66 with UNA — 0.83
Age Over 66 with REP — 0.86
Age Over 66 with LIB — 0.71
Age Over 66 with DEM — 0.81
Age Over 66 with CST_GRE — 0.69
Age 41 − 65 with UNA — 0.77
Age 41 − 65 with REP — 0.86
Age 41 − 65 with LIB — 0.68
Age 41 − 65 with DEM — 0.80
Age 41 − 65 with CST_GRE — 0.69
Age 26 − 40 with UNA — 0.57
Age 26 − 40 with REP — 0.71
Age 26 − 40 with LIB — 0.57
Age 26 − 40 with DEM — 0.59
Age 26 − 40 with CST_GRE — 0.65
Age 18 − 25 with UNA — 0.50
Age 18 − 25 with REP — 0.67
Age 18 − 25 with LIB — 0.52
Age 18 − 25 with DEM — 0.57
Age 18 − 25 with CST_GRE — 0.55

Registered Voters vs. Age*Party

Age Over 66 with UNA — 87144
Age Over 66 with REP — 141929
Age Over 66 with LIB — 440
Age Over 66 with DEM — 183005
Age Over 66 with CST_GRE — 88
Age 41 − 65 with UNA — 216577
Age 41 − 65 with REP — 2659...
Age 41 − 65 with LIB — 2767
Age 41 − 65 with DEM — 28...
Age 41 − 65 with CST_GRE — 598
Age 26 − 40 with UNA — 174024
Age 26 − 40 with REP — 117840
Age 26 − 40 with LIB — 4721
Age 26 − 40 with DEM — 158659
Age 26 − 40 with CST_GRE — 845
Age 18 − 25 with UNA — 117815
Age 18 − 25 with REP — 61730
Age 18 − 25 with LIB — 2756
Age 18 − 25 with DEM — 83177
Age 18 − 25 with CST_GRE — 537

party_cd
CST_GRE
DEM
LIB
REP
UNA

Age & Race

We are also interested in a potential relationship between age and race. In the graphs below we see a general trend that, across races, as the age increase so too does the turnout rate. We also see that, within each age group, white people make up the majority of registered voters.

Turnout rate vs. Age*Race

Age Over 66 with W — 0.85
Age Over 66 with Undesignated — 0.76
Age Over 66 with O — 0.72
Age Over 66 with Multiracial — 0.75
Age Over 66 with I — 0.73
Age Over 66 with B — 0.81
Age Over 66 with A — 0.74
Age 41 − 65 with W — 0.84
Age 41 − 65 with Undesignated — 0.78
Age 41 − 65 with O — 0.69
Age 41 − 65 with Multiracial — 0.72
Age 41 − 65 with I — 0.68
Age 41 − 65 with B — 0.77
Age 41 − 65 with A — 0.79
Age 26 − 40 with W — 0.65
Age 26 − 40 with Undesignated — 0.69
Age 26 − 40 with O — 0.47
Age 26 − 40 with Multiracial — 0.50
Age 26 − 40 with I — 0.48
Age 26 − 40 with B — 0.52
Age 26 − 40 with A — 0.56
Age 18 − 25 with W — 0.61
Age 18 − 25 with Undesignated — 0.61
Age 18 − 25 with O — 0.45
Age 18 − 25 with Multiracial — 0.51
Age 18 − 25 with I — 0.40
Age 18 − 25 with B — 0.46
Age 18 − 25 with A — 0.59

Registered Voters vs. Age*Race

Age Over 66 with W — 299041
Age Over 66 with Undesignated — 14024
Age Over 66 with O — 4364
Age Over 66 with Multiracial — 553
Age Over 66 with I — 7390
Age Over 66 with B — 85441
Age Over 66 with A — 1793
Age 41 − 65 with W — 48...
Age 41 − 65 with Undesignated — 57849
Age 41 − 65 with O — 15948
Age 41 − 65 with Multiracial — 2578
Age 41 − 65 with I — 16366
Age 41 − 65 with B — 189152
Age 41 − 65 with A — 7627
Age 26 − 40 with W — 240572
Age 26 − 40 with Undesignated — 63068
Age 26 − 40 with O — 14363
Age 26 − 40 with Multiracial — 4713
Age 26 − 40 with I — 9428
Age 26 − 40 with B — 118507
Age 26 − 40 with A — 5438
Age 18 − 25 with W — 127910
Age 18 − 25 with Undesignated — 48858
Age 18 − 25 with O — 13818
Age 18 − 25 with Multiracial — 5831
Age 18 − 25 with I — 5470
Age 18 − 25 with B — 62532
Age 18 − 25 with A — 3596

race_code
Multiracial
Undesignated
B
W
A
I
O

Model

With our exploration of the data complete, we now begin our model building. From the research questions we have an idea of what must be included in the model; this includes the main effect for party, sex, and age, as well as random intercepts by county. This will be our base model. From the research questions we also know that we likely need an interaction term between age and party as well as sex and party. Additionally, since our dependent variable, turnout rate, is between 0 and 1, we know we should use a logistic mixed effects model. Below we see the output from the forward selection. We see that the last model, with interactions between sex and party, age and party, as well as age and race, has the lowest BIC and AIC so this will be our final model.

$$log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \sum_{k_p}\beta_{1k_p}\mathbb{I}[P_{ij}=k_p] + \sum_{k_r}\beta_{2k_r}\mathbb{I}[R_{ij}=k_r] + \sum_{k_e}\beta_{3k_e}\mathbb{I}[E_{ij}=k_e] + \sum_{k_s}\beta_{4k_s}\mathbb{I}[S_{ij}=k_s]$$

$$+ \sum_{k_a}\beta_{5k_a}\mathbb{I}[A_{ij}=k_a] + \sum_{k_s,k_p}\beta_{6k_sk_p}\mathbb{I}[S_{ij}=k_s]\mathbb{I}[P_{ij}=k_p] + \sum_{k_a,k_p}\beta_{7k_ak_p}\mathbb{I}[A_{ij}=k_a]\mathbb{I}[P_{ij}=k_p]$$

$$+ \sum_{k_a,k_r}\beta_{8k_ak_r}\mathbb{I}[A_{ij}=k_a]\mathbb{I}[R_{ij}=k_r] + b_{0j}$$

3

where $\pi_{ij} = Pr(y_{ij} = 1)$ and $b_{0j} \overset{\text{iid}}{\sim} N(0, \sigma^2)$ and

- j indexes county

- i indexes observation

- P_{ij} is the Party code ($k_p \in \{DEM,\ LIB,\ REP,\ UNA\}$)

- R_{ij} is the Race code ($k_r \in \{Undesignated,\ B,\ W,\ A,\ I,\ O\}$)

- E_{ij} is the Ethnic code ($k_e \in \{NL,\ UN\}$)

- S_{ij} is the sex code ($k_s \in \{M,\ U\}$)

- A_{ij} is the Age group ($k_a \in \{Age\ 20 - 40,\ Age\ 41 - 65,\ Age\ Over\ 66\}$)

Table 1: Forward model selection

Model	LRT.p.value	AIC	BIC
Base model		96904.01	96988.73
Add race	0	88191.94	88322.86
Add race and ethnic	0	86526.87	86673.20
Without intercept	0	99839.91	99955.43
Add the interaction of sex:party, and age:party	0	83825.28	84125.65
Add the interaction of sex:party, age:party, and age:race	0	79688.45	80127.45

Model Assessment

From our EDA it can reasonably be assumed that the linearity condition is met for our model and that multicollinearity is not an issue. As we see in the binned residual plot below, there may cause for some concern as some points with expected values between 0.8 and 0.9 fall out of the range of the confidence bands. However, relative to the amount of data we have, these 8 points, in addition to 7 other scattered points outside the confidence bands, don't point to our model being incorrect. We see that the model has an AUC around 0.7 which is satisfactory and much better than a coin flip. Our model seems to more accurately identify positives than negatives as well.
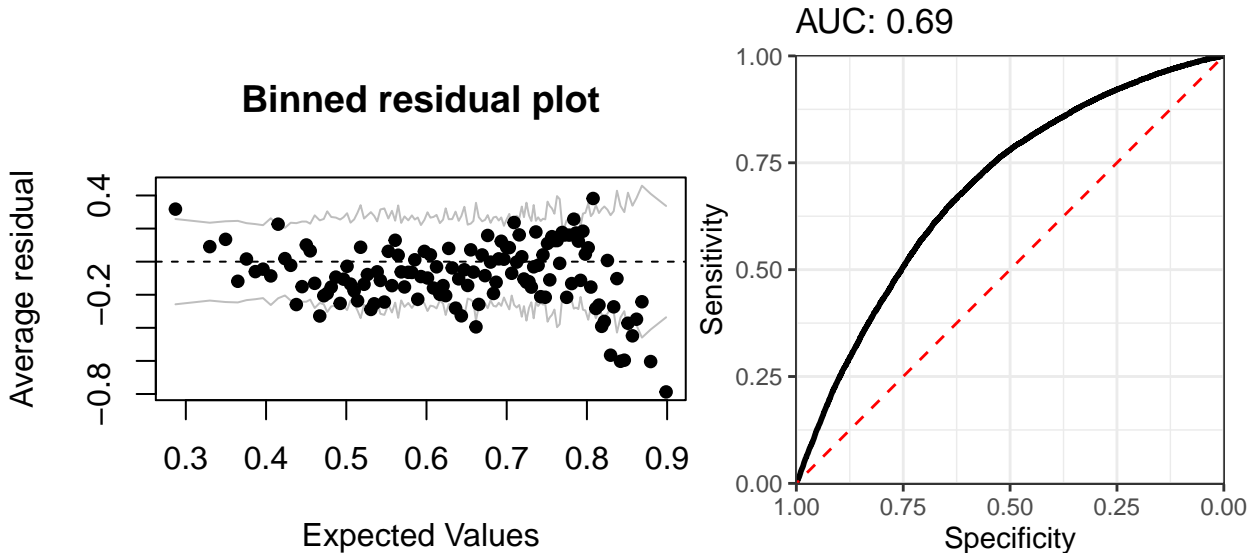
Conclusion

Now, we can answer the research questions based on the model results.

Q1: How did different demographic subgroups vote in the 2020 general elections?

Party

Notice that our final model incorporated the interaction terms `sex:party` and `age:party`, we should interpret the effects of party at the baseline levels of the other two terms (female and age 18-25).

Table 2: Estiamtes and 95% CI of Fixed Effects for Party

	Estimate	Std. Error	95% lower	95% upper
party_cdDEM	0.3477	0.1128	0.1267	0.5688
party_cdLIB	-0.2886	0.1217	-0.5271	-0.0501
party_cdREP	0.3248	0.1129	0.1034	0.5461
party_cdUNA	-0.1791	0.1127	-0.4000	0.0417

Holding all other predictors unchanged:

an 18-25 year old female DEM voter has $e^{0.3477} = 1.4158$ times (a 41% increase) the odds of voting

an 18-25 year old female LIB voter has $e^{-0.2886} = 0.7493$ times (a 25% decrease) the odds of voting

an 18-25 year old female REP voter has $e^{0.3248} = 1.3838$ times (a 38% increase) the odds of voting

an 18-25 year old female UNA voter has $e^{-0.1791} = 0.8360$ times (a 17% decrease) the odds of voting (this effect is not significant since the 95% CI covers 0)

compared with an 18-25 year old female GRE/CST voter.

Race

Table 3: Estiamtes and 95% CI of Fixed Effects for Race

	Estimate	Std. Error	95% lower	95% upper
race_codeUndesignated	0.3584	0.0350	0.2899	0.4270
race_codeB	-0.3441	0.0341	-0.4110	-0.2772
race_codeW	0.2318	0.0337	0.1658	0.2979
race_codeA	0.1517	0.0477	0.0581	0.2453
race_codeI	-0.2409	0.0437	-0.3264	-0.1553
race_codeO	-0.1098	0.0374	-0.1831	-0.0364

Holding all other predictors unchanged:

an 18-25 year old voter with undesignated race has $e^{0.3584} = 1.4310$ times (a 43% increase) the odds of voting

an 18-25 year old Black voter has $e^{-0.3441} = 0.7089$ times (a 29% decrease) the odds of voting

an 18-25 year old White voter has $e^{0.2318} = 1.2609$ times (a 26% increase) the odds of voting

an 18-25 year old Asian voter has $e^{0.1517} = 1.1638$ times (a 16% increase) the odds of voting

an 18-25 year old Indian American voter has $e^{-0.2409} = 0.7859$ times (a 21% decrease) the odds of voting

an 18-25 year old voter in other race has $e^{-0.1098} = 0.8960$ times (a 10% decrease) the odds of voting

compared with an 18-25 year-old multiracial voter.

Ethnic

Table 4: Estiamtes and 95% CI of Fixed Effects for Ethnic

	Estimate	Std. Error	95% lower	95% upper
ethnic_codeNL	0.4438	0.0107	0.4228	0.4647
ethnic_codeUN	0.3618	0.0109	0.3404	0.3833

Holding all other predictors unchanged:

a non Hispanic/Latino voter has $e^{0.4438} = 1.5586$ times (a 56% increase) the odds of voting

a voter with unkown ethnic has $e^{0.3618} = 1.4359$ times (a 44% increase) the odds of voting

compared with a Hispanic/Latino voter.

Sex

Table 5: Estiamtes and 95% CI of Fixed Effects for Sex

	Estimate	Std. Error	95% lower	95% upper
sex_codeM	-0.0728	0.1112	-0.2908	0.1451
sex_codeU	0.1790	0.1302	-0.0762	0.4342

Holding all other predictors unchanged:

a male CST/GRE voter has $e^{-0.0728} = 0.9298$ times (a 7% decrease) the odds of voting

a CST/GRE voter with undesignated gender has $e^{0.1790} = 1.1960$ times (a 20% increase) the odds of voting

compared with a female CST/GRE voter.

However, neither of these coefficients are significant since none of the 95% CIs cover 0.

Age

Table 6: Estiamtes and 95% CI of Fixed Effects for Age

|	Estimate	Std. Error	95% lower	95% upper
ageAge 26 - 40	0.0492	0.1239	-0.1936	0.2920
ageAge 41 - 65	0.3871	0.1379	0.1168	0.6575
ageAge Over 66	0.4609	0.2720	-0.0723	0.9940

Holding all other predictors unchanged:

a 26-40 year old multiracial CST/GRE voter has $e^{0.0492} = 1.0504$ times (a 5% increase) the odds of voting (this effect is not significant since the 95% CI covers 0)

a 41-65 year old multiracial CST/GRE voter has $e^{0.3871} = 1.4727$ times (a 47% increase) the odds of voting

an over 66 year old multiracial CST/GRE voter has $e^{0.4609} = 1.5855$ times (a 59% increase) the odds of voting (this effect is not significant since the 95% CI covers 0)

compared with an 18-25 year old multiracial CST/GRE voter.

Q2: Did the overall probability or odds of voting differ by county in 2020? Which counties differ the most from other counties?

county_desc
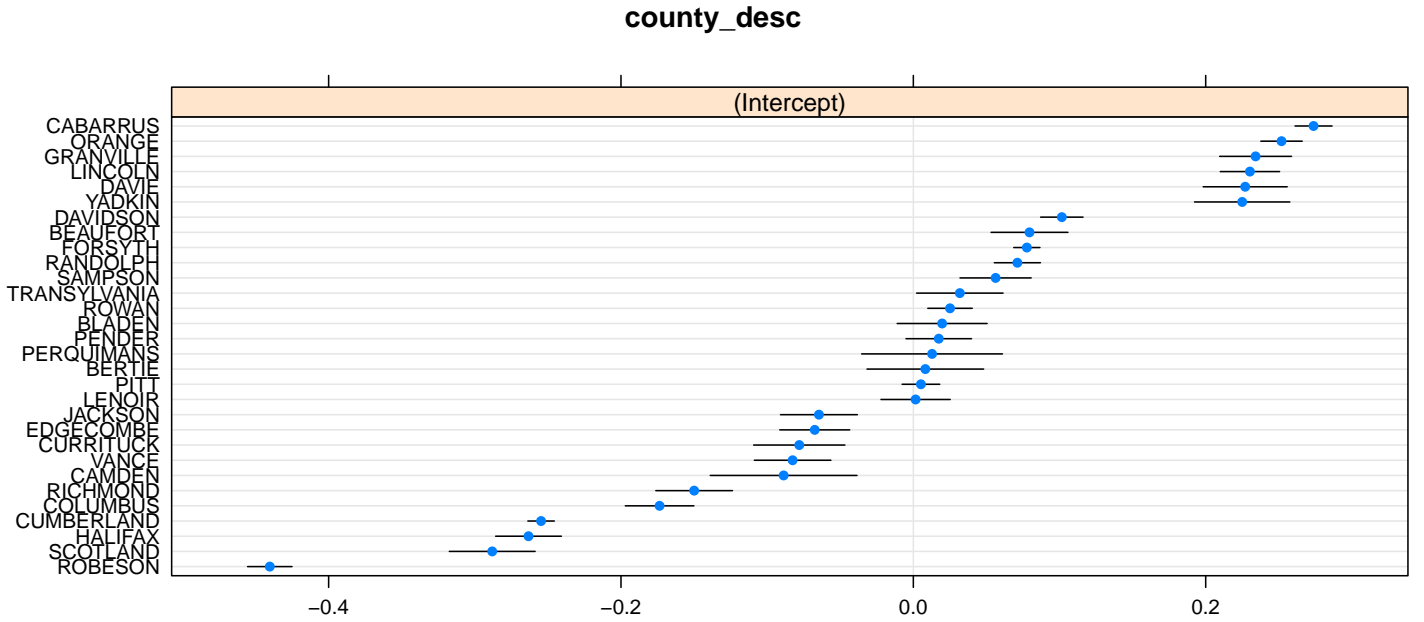
The above dotplot displays the random effects for each county. Since we observe clear heterogeneity across counties, we can conclude the overall probability of voting differed by county in 2020. Robeson County has the smallest estimate, while Cabarrus has the largest estimate. Furthermore, Robeson differs the most from all other counties. A table with the estimates and corresponding confidence intervals is included in the appendix.

Q3: How did the turnout rates differ between females and males for the different party affiliations?

Table 7: Estiamtes and 95% CI of Fixed Effects for Party:Sex

	Estimate	Std. Error	95% lower	95% upper
party_cdDEM:sex_codeM	-0.2085	0.1113	-0.4268	0.0097
party_cdLIB:sex_codeM	0.0302	0.1197	-0.2044	0.2648
party_cdREP:sex_codeM	0.0227	0.1114	-0.1956	0.2411
party_cdUNA:sex_codeM	-0.0401	0.1114	-0.2584	0.1781
party_cdDEM:sex_codeU	-0.0766	0.1305	-0.3324	0.1792
party_cdLIB:sex_codeU	0.4754	0.1447	0.1917	0.7591
party_cdREP:sex_codeU	0.1535	0.1307	-0.1027	0.4096
party_cdUNA:sex_codeU	-0.2499	0.1302	-0.5051	0.0054

From the table above, only the fixed effects for `party_cdLIB:sex_codeU` is significant (the 95% CI does not cover 0). For simplicity, we only interpret this fixed effect: A Libertarian voter with undesignated gender has $e^{0.4754} = 1.6087$ times the odds of voting.

Q4: How did the turnout rates differ between age groups for the different party affiliations?

Table 8: Estiamtes and 95% CI of Fixed Effects for Party:Age

	Estimate	Std. Error	95% lower	95% upper
party_cdDEM:ageAge 26 - 40	-0.2391	0.1161	-0.4666	-0.0115
party_cdLIB:ageAge 26 - 40	-0.1267	0.1258	-0.3733	0.1199
party_cdREP:ageAge 26 - 40	-0.0635	0.1163	-0.2913	0.1644
party_cdUNA:ageAge 26 - 40	-0.0556	0.1159	-0.2828	0.1717
party_cdDEM:ageAge 41 - 65	0.4162	0.1267	0.1679	0.6646
party_cdLIB:ageAge 41 - 65	0.1036	0.1386	-0.1682	0.3753
party_cdREP:ageAge 41 - 65	0.5018	0.1268	0.2533	0.7504

7

	Estimate	Std. Error	95% lower	95% upper
party_cdUNA:ageAge 41 - 65	0.5563	0.1266	0.3082	0.8044
party_cdDEM:ageAge Over 66	0.2917	0.2516	-0.2015	0.7848
party_cdLIB:ageAge Over 66	0.2045	0.2754	-0.3353	0.7443
party_cdREP:ageAge Over 66	0.5102	0.2517	0.0169	1.0035
party_cdUNA:ageAge Over 66	0.8719	0.2516	0.3788	1.3651

From the table above, we see some of the fixed effects are significant, we only interpret them:

A 26-40 year old Democratic voter has $e^{-0.2391} = 0.7873$ times the odds of voting.

A 41-65 year old Democratic voter has $e^{0.4162} = 1.5162$ times the odds of voting.

A 41-65 year old Republican voter has $e^{0.5018} = 1.6517$ times the odds of voting.

A 41-65 year old voter with undesignated party has $e^{0.5563} = 1.7442$ times the odds of voting.

An over 66 year old Republican voter has $e^{0.5102} = 1.6656$ times the odds of voting.

An over 66 year old voter with undesignated party has $e^{0.8719} = 2.3915$ times the odds of voting.

Strengths and Limitations

Strengths: A major strength of our model is the inclusion of random intercepts by county which allows us to understand the difference in turnout rate between each county.
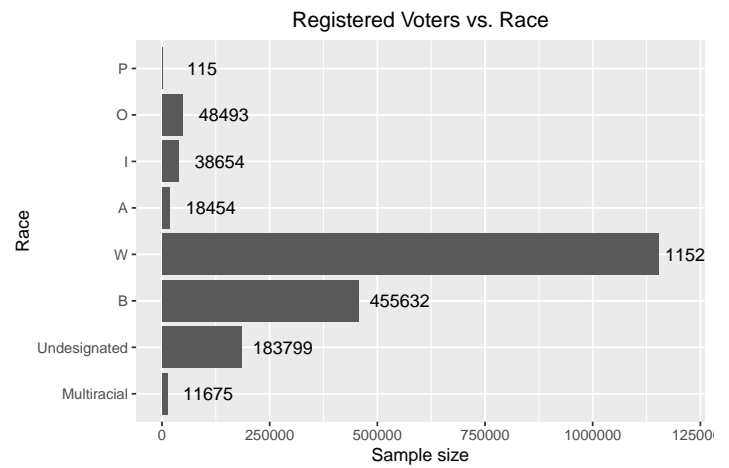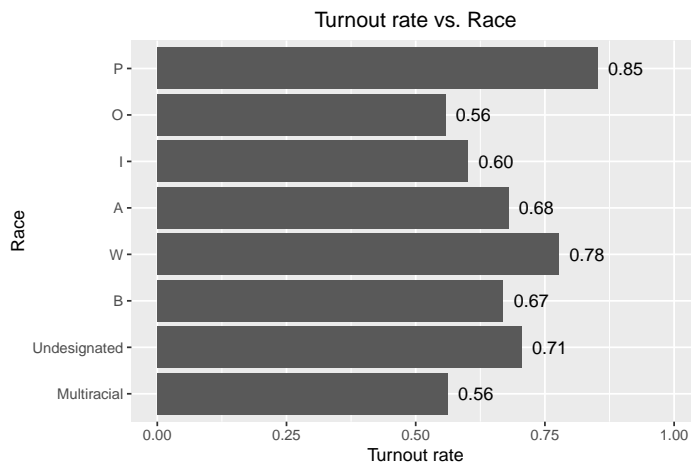
Limitations: Firstly, we dropped rows that contained missing values during data cleaning . While we are working with almost 2 million observations, we do have quite a few small demographic groups, especially for our interaction terms. A possible alternative would have been to impute values for the missing data so that we could keep as many rows as possible. Obviously imputation comes with it's own set of assumptions and limitations. In addition, we did not try poststratification to re-adjust the proportions among demographic groups. This could lead to slightly biased result.
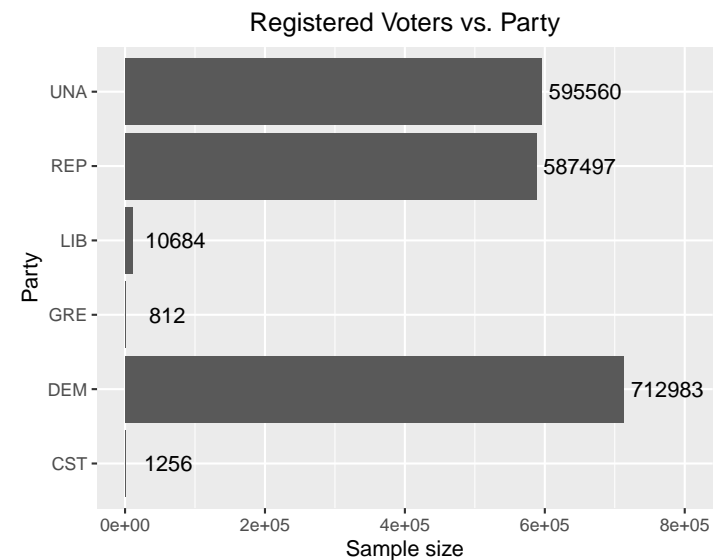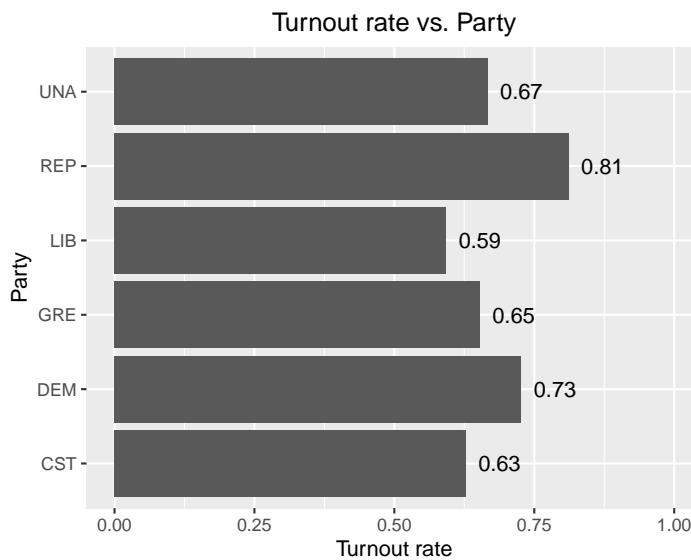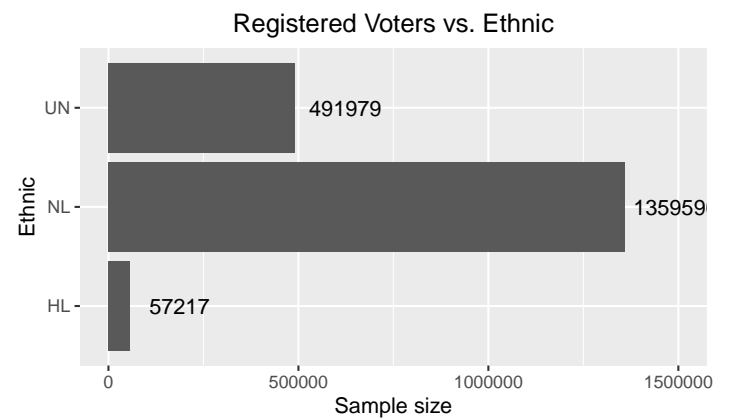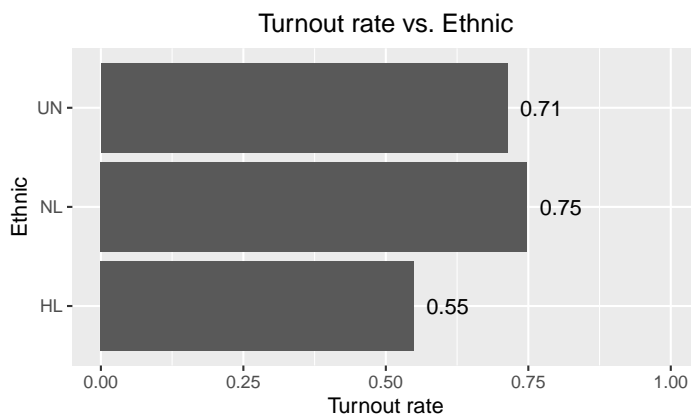
Appendix

EDA

Race

Turnout rate vs. Race

Race	Turnout rate
P	0.85
O	0.56
I	0.60
A	0.68
W	0.78
B	0.67
Undesignated	0.71
Multiracial	0.56

Registered Voters vs. Race

Race	Sample size
P	115
O	48493
I	38654
A	18454
W	1152
B	455632
Undesignated	183799
Multiracial	11675

Party

Turnout rate vs. Party

Party	Turnout rate
UNA	0.67
REP	0.81
LIB	0.59
GRE	0.65
DEM	0.73
CST	0.63

Registered Voters vs. Party

Party	Sample size
UNA	595560
REP	587497
LIB	10684
GRE	812
DEM	712983
CST	1256

Ethnic

Turnout rate vs. Ethnic

Ethnic	Turnout rate
UN	0.71
NL	0.75
HL	0.55

Registered Voters vs. Ethnic

Ethnic	Sample size
UN	491979
NL	135959
HL	57217

Sex

Turnout rate vs. Sex

Sex	Turnout rate
U	0.71
M	0.72
F	0.75

Registered Voters vs. Sex

Sex	Sample size
U	159431
M	799767
F	94959

Age

Turnout rate vs. Age

Age	Turnout rate
Age Over 66	0.83
Age 41 – 65	0.81
Age 26 – 40	0.61
Age 18 – 25	0.56

Registered Voters vs. Age

Age	Sample size
Age Over 66	412606
Age 41 – 65	774082
Age 26 – 40	456089
Age 18 – 25	266015

Model Estimates

Table 9: Estiamtes and 95% CI of Fixed Effects

	Estimate	Std. Error	95% lower	95% upper
(Intercept)	-0.2854	0.1216	-0.5237	-0.0472
party_cdDEM	0.3477	0.1128	0.1267	0.5688
party_cdLIB	-0.2886	0.1217	-0.5271	-0.0501
party_cdREP	0.3248	0.1129	0.1034	0.5461
party_cdUNA	-0.1791	0.1127	-0.4000	0.0417
race_codeUndesignated	0.3584	0.0350	0.2899	0.4270
race_codeB	-0.3441	0.0341	-0.4110	-0.2772
race_codeW	0.2318	0.0337	0.1658	0.2979
race_codeA	0.1517	0.0477	0.0581	0.2453
race_codeI	-0.2409	0.0437	-0.3264	-0.1553
race_codeO	-0.1098	0.0374	-0.1831	-0.0364
ethnic_codeNL	0.4438	0.0107	0.4228	0.4647
ethnic_codeUN	0.3618	0.0109	0.3404	0.3833

10

	Estimate	Std. Error	95% lower	95% upper
sex_codeM	-0.0728	0.1112	-0.2908	0.1451
sex_codeU	0.1790	0.1302	-0.0762	0.4342
ageAge 26 - 40	0.0492	0.1239	-0.1936	0.2920
ageAge 41 - 65	0.3871	0.1379	0.1168	0.6575
ageAge Over 66	0.4609	0.2720	-0.0723	0.9940
party_cdDEM:sex_codeM	-0.2085	0.1113	-0.4268	0.0097
party_cdLIB:sex_codeM	0.0302	0.1197	-0.2044	0.2648
party_cdREP:sex_codeM	0.0227	0.1114	-0.1956	0.2411
party_cdUNA:sex_codeM	-0.0401	0.1114	-0.2584	0.1781
party_cdDEM:sex_codeU	-0.0766	0.1305	-0.3324	0.1792
party_cdLIB:sex_codeU	0.4754	0.1447	0.1917	0.7591
party_cdREP:sex_codeU	0.1535	0.1307	-0.1027	0.4096
party_cdUNA:sex_codeU	-0.2499	0.1302	-0.5051	0.0054
party_cdDEM:ageAge 26 - 40	-0.2391	0.1161	-0.4666	-0.0115
party_cdLIB:ageAge 26 - 40	-0.1267	0.1258	-0.3733	0.1199
party_cdREP:ageAge 26 - 40	-0.0635	0.1163	-0.2913	0.1644
party_cdUNA:ageAge 26 - 40	-0.0556	0.1159	-0.2828	0.1717
party_cdDEM:ageAge 41 - 65	0.4162	0.1267	0.1679	0.6646
party_cdLIB:ageAge 41 - 65	0.1036	0.1386	-0.1682	0.3753
party_cdREP:ageAge 41 - 65	0.5018	0.1268	0.2533	0.7504
party_cdUNA:ageAge 41 - 65	0.5563	0.1266	0.3082	0.8044
party_cdDEM:ageAge Over 66	0.2917	0.2516	-0.2015	0.7848
party_cdLIB:ageAge Over 66	0.2045	0.2754	-0.3353	0.7443
party_cdREP:ageAge Over 66	0.5102	0.2517	0.0169	1.0035
party_cdUNA:ageAge Over 66	0.8719	0.2516	0.3788	1.3651
race_codeUndesignated:ageAge 26 - 40	0.3327	0.0462	0.2422	0.4232
race_codeB:ageAge 26 - 40	0.3603	0.0455	0.2712	0.4494
race_codeW:ageAge 26 - 40	0.2176	0.0450	0.1293	0.3059
race_codeA:ageAge 26 - 40	-0.0583	0.0625	-0.1808	0.0642
race_codeI:ageAge 26 - 40	0.3734	0.0563	0.2630	0.4838
race_codeO:ageAge 26 - 40	0.1393	0.0505	0.0403	0.2383
race_codeUndesignated:ageAge 41 - 65	-0.1918	0.0572	-0.3039	-0.0798
race_codeB:ageAge 41 - 65	0.4707	0.0564	0.3603	0.5812
race_codeW:ageAge 41 - 65	0.2480	0.0560	0.1383	0.3578
race_codeA:ageAge 41 - 65	0.1140	0.0712	-0.0255	0.2535
race_codeI:ageAge 41 - 65	0.2272	0.0643	0.1012	0.3533
race_codeO:ageAge 41 - 65	0.0672	0.0606	-0.0516	0.1860
race_codeUndesignated:ageAge Over 66	-0.4665	0.1071	-0.6764	-0.2566
race_codeB:ageAge Over 66	0.7329	0.1055	0.5261	0.9397
race_codeW:ageAge Over 66	0.1776	0.1051	-0.0283	0.3836
race_codeA:ageAge Over 66	-0.3342	0.1229	-0.5751	-0.0933
race_codeI:ageAge Over 66	0.3896	0.1115	0.1710	0.6082
race_codeO:ageAge Over 66	-0.0054	0.1115	-0.2240	0.2131

Table 10: Estimated random intercepts

grpvar	term	grp	condval	condsd	exp(condval)
county_desc	(Intercept)	CABARRUS	0.2738	0.0065	1.3149
county_desc	(Intercept)	ORANGE	0.2519	0.0072	1.2864

grpvar	term	grp	condval	condsd	exp(condval)
county_desc	(Intercept)	GRANVILLE	0.2341	0.0125	1.2638
county_desc	(Intercept)	LINCOLN	0.2303	0.0104	1.2589
county_desc	(Intercept)	DAVIE	0.2270	0.0147	1.2549
county_desc	(Intercept)	YADKIN	0.2250	0.0166	1.2523
county_desc	(Intercept)	DAVIDSON	0.1016	0.0074	1.1069
county_desc	(Intercept)	BEAUFORT	0.0795	0.0134	1.0827
county_desc	(Intercept)	FORSYTH	0.0777	0.0046	1.0808
county_desc	(Intercept)	RANDOLPH	0.0712	0.0080	1.0738
county_desc	(Intercept)	SAMPSON	0.0563	0.0124	1.0579
county_desc	(Intercept)	TRANSYLVANIA	0.0318	0.0151	1.0323
county_desc	(Intercept)	ROWAN	0.0251	0.0078	1.0254
county_desc	(Intercept)	BLADEN	0.0197	0.0157	1.0199
county_desc	(Intercept)	PENDER	0.0174	0.0114	1.0175
county_desc	(Intercept)	PERQUIMANS	0.0128	0.0246	1.0129
county_desc	(Intercept)	BERTIE	0.0082	0.0204	1.0082
county_desc	(Intercept)	PITT	0.0052	0.0066	1.0052
county_desc	(Intercept)	LENOIR	0.0015	0.0121	1.0015
county_desc	(Intercept)	JACKSON	-0.0646	0.0134	0.9375
county_desc	(Intercept)	EDGECOMBE	-0.0674	0.0122	0.9348
county_desc	(Intercept)	CURRITUCK	-0.0780	0.0159	0.9249
county_desc	(Intercept)	VANCE	-0.0826	0.0134	0.9207
county_desc	(Intercept)	CAMDEN	-0.0887	0.0256	0.9151
county_desc	(Intercept)	RICHMOND	-0.1499	0.0134	0.8608
county_desc	(Intercept)	COLUMBUS	-0.1736	0.0120	0.8406
county_desc	(Intercept)	CUMBERLAND	-0.2547	0.0046	0.7752
county_desc	(Intercept)	HALIFAX	-0.2633	0.0115	0.7685
county_desc	(Intercept)	SCOTLAND	-0.2881	0.0150	0.7497
county_desc	(Intercept)	ROBESON	-0.4402	0.0078	0.6439

Codes

```
knitr::opts_chunk$set(cache = T, message = F, warning =F)
library(tidyverse)
library(lme4)
library(pROC)
library(kableExtra)
library(performance)
library(lattice)
library(brms)
library(reshape2)
library(dplyr)
library(gridExtra)
library(cowplot)
library(arm)
library(knitr)
```

```
# Explanation of abbreviations: https://s3.amazonaws.com/dl.ncsbe.gov/data/layout_ncvoter.txt
# contains info about the aggregate counts of voters who actually voted by demographic variables

votes <- read.table("history_stats_20201103.txt", header = TRUE, fill = TRUE, sep = '\t')

# contains info about the aggregate counts of registered voters by demographic variables
registers <- read.table("voter_stats_20201103.txt", header = TRUE, fill = TRUE, sep = '\t')
```

```
# set "" or " " to NA
registers[registers == ""] <- NA
registers[registers == " "] <- NA
votes[votes == " "] <- NA

# unique(registers$election_date) # "11/03/2020" NA
# unique(registers$stats_type) # "history" NA
# unique(registers$update_date) # "01/13/2021" NA

# remove above three columns

registers <- registers %>%
  dplyr::select(-election_date, -stats_type, -update_date)
votes <- votes %>%
  dplyr::select(-election_date, -stats_type, -update_date)

votes <- votes %>%
  mutate(total_voters = as.numeric(total_voters))

registers <- registers %>%
  mutate(total_voters2 = as.numeric(total_voters))


votes <- votes %>%
  group_by(county_desc, age, party_cd, race_code, ethnic_code, sex_code) %>%
  summarize(total_vot = sum(total_voters,na.rm =T))
```

```r
data <- registers %>%
  group_by(county_desc, party_cd, race_code, ethnic_code, sex_code, age) %>%
  summarize(total_reg = sum(total_voters2, na.rm= T)) %>%
  left_join(votes, by = c("county_desc", "age", "party_cd", "race_code",
                          "ethnic_code", "sex_code"))
```

```r
data$total_vot <- ifelse(is.na(data$total_vot), 0, data$total_vot)
```

```r
data$total_vot <- ifelse(data$total_vot <= data$total_reg, data$total_vot, data$total_reg)
```

```r
data <- data %>%
  drop_na()

# 51883 rows
```

```r
set.seed(10)
counties <- sample(unique(data$county_desc), 30)

data <- data %>%
  filter(county_desc %in% counties)

# 16439 rows
```

```r
data$race_code <- recode_factor(data$race_code, M = "Multiracial", U = "Undesignated")
data2 <- data %>%
  drop_na() %>%
  mutate(total_not_vote = total_reg - total_vot) %>%
  pivot_longer(cols = c("total_vot", "total_not_vote"), names_to = "vote_or_not", values_to = "prob"
  mutate(vote_or_not = ifelse(vote_or_not == "total_vot", 1, 0)) %>%
  mutate(row_expand = map(prob, ~rep_len(1, .x))) %>%
  unnest(cols = c(row_expand)) %>%
  dplyr::select(-prob, -row_expand)

# 1,908,907 individuals
```

```r
table(data2$vote_or_not)

turnout_rate <- sum(data$total_vot) / sum(data$total_reg)
data.frame(group = "Total", turnout_rate = turnout_rate)
```

```r
p1 <- data %>%
  group_by(county_desc) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop",
            n = sum(total_reg)) %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
  ggplot( aes(x = fct_reorder(county_desc,n), y = turnout_rate)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("County") + ylab("Turnout rate") +
```

```r
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.3) +
  ggtitle("Turnout Rate by County") +
    theme(plot.title = element_text(hjust = 0.5)) +
  geom_hline(yintercept = turnout_rate, color = 'red') +
  coord_flip()

p2 <- data %>%
  group_by(county_desc) %>%
  summarize(n = sum(total_reg), .groups="drop") %>%
  ggplot( aes(x = fct_reorder(county_desc,n), y = n)) +
  geom_bar(stat = "identity") +
  ylim(0, 300000) +
  xlab("Party") + ylab("Sample size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.1) +
  ggtitle("Registered Voters by County") +
    theme(plot.title = element_text(hjust = 0.5)) +
  ylim(0,3.15e5) +
  coord_flip()

 grid.arrange(p1,p2, nrow= 1)
```

```r
p3 <- data %>%
  group_by(race_code) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
  dplyr::select(group = race_code, turnout_rate) %>%
  ggplot( aes(x = group, y = turnout_rate)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Race") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.3) +
  ggtitle("Turnout rate vs. Race") +
    theme(plot.title = element_text(hjust = 0.5)) +
    coord_flip()


p4 <- data %>%
  group_by(race_code) %>%
  summarize(n = sum(total_reg), .groups="drop") %>%
  ggplot( aes(x = race_code, y = n)) +
  geom_bar(stat = "identity") +
  xlab("Race") + ylab("Sample size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.1) +
  ggtitle("Registered Voters vs. Race") +
    theme(plot.title = element_text(hjust = 0.5)) +
  coord_flip()

grid.arrange(p3,p4, nrow= 1)
```

```r
data$race_code <- droplevels(data$race_code, 'P')
data <- data %>%
  drop_na()


p5 <- data %>%
  group_by(party_cd) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
  dplyr::select(group = party_cd, turnout_rate) %>%
  ggplot( aes(x = group, y = turnout_rate)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Party") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.3) +
  ggtitle("Turnout rate vs. Party") +
    theme(plot.title = element_text(hjust = 0.5))  +
    coord_flip()

p6 <- data %>%
  group_by(party_cd) %>%
  summarize(n = sum(total_reg), .groups="drop") %>%
  ggplot( aes(x = party_cd, y = n)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Sample size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.1) +
  ggtitle("Registered Voters vs. Party") +
    theme(plot.title = element_text(hjust = 0.5)) +
  coord_flip()

grid.arrange(p5,p6, nrow= 1)



# "CST","GRE" have small smaple size (50,25) and similar turnout rate (0.16,0.15)
data$party_cd <- ifelse(data$party_cd %in% c("CST","GRE"),"CST_GRE",as.character(data$party_cd))

p7 <- data %>%
  group_by(ethnic_code) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
  dplyr::select(group = ethnic_code, turnout_rate) %>%
  ggplot( aes(x = group, y = turnout_rate)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Ethnic") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.3) +
  ggtitle("Turnout rate vs. Ethnic") +
    theme(plot.title = element_text(hjust = 0.5))+
    coord_flip()

p8 <- data %>%
```

```
  group_by(ethnic_code) %>%
  summarize(n = sum(total_reg), .groups="drop") %>%
  ggplot( aes(x = ethnic_code, y = n)) +
  geom_bar(stat = "identity") +
  xlab("Ethnic") + ylab("Sample size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.1) +
  ggtitle("Registered Voters vs. Ethnic") +
    theme(plot.title = element_text(hjust = 0.5)) +
  coord_flip()

grid.arrange(p7,p8, nrow= 1)
```

```
 p9 <- data %>%
  group_by(sex_code) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
  dplyr::select(group = sex_code, turnout_rate) %>%
  ggplot( aes(x = group, y = turnout_rate)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Sex") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.3) +
  ggtitle("Turnout rate vs. Sex") +
    theme(plot.title = element_text(hjust = 0.5)) +
    coord_flip()

p10 <- data %>%
  group_by(sex_code) %>%
  summarize(n = sum(total_reg), .groups="drop") %>%
  ggplot( aes(x = sex_code, y = n)) +
  geom_bar(stat = "identity") +
  xlab("Sex") + ylab("Sample size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.1) +
  ggtitle("Registered Voters vs. Sex") +
    theme(plot.title = element_text(hjust = 0.5)) +
  coord_flip()

grid.arrange(p9,p10, nrow= 1)
```

```
p11 <- data %>%
  group_by(age) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
  dplyr::select(group = age, turnout_rate) %>%
  ggplot( aes(x = group, y = turnout_rate)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Age") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.3) +
  ggtitle("Turnout rate vs. Age") +
```

```r
    theme(plot.title = element_text(hjust = 0.5))+
    coord_flip()

p12 <- data %>%
  group_by(age) %>%
  summarize(n = sum(total_reg), .groups="drop") %>%
  ggplot( aes(x = age, y = n)) +
  geom_bar(stat = "identity") +
  xlab("Age") + ylab("Sample size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.1) +
  ggtitle("Registered Voters vs. Age") +
    theme(plot.title = element_text(hjust = 0.5)) +
  ylim(0,8.8e05) +
  coord_flip()

grid.arrange(p11,p12, nrow= 1)
```

```r
p13 <- data %>%
  group_by(sex_code, party_cd) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg,
         group = paste0(sex_code, " with ", party_cd)) %>%
  dplyr::select(group = group, turnout_rate, sex_code, party_cd) %>%
  ggplot( aes(x = group, y = turnout_rate, fill = party_cd)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Sex * Party") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.3) +
  coord_flip() +
  ggtitle("Turnout rate vs. Sex*Party") +
    theme(plot.title = element_text(hjust = 0.5))

p14 <- data %>%
  group_by(sex_code, party_cd) %>%
  summarize(n = sum(total_reg), .groups = "drop") %>%
  mutate(group = paste0(sex_code, " with ", party_cd)) %>%
  ggplot( aes(x = group, y = n, fill = party_cd)) +
  geom_bar(stat = "identity") +
  xlab("Sex * Party") + ylab("Size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.05) +
  coord_flip() +
  ggtitle("Registered Voters vs. Sex*Party") +
    theme(plot.title = element_text(hjust = 0.5))

legend <- get_legend(p13)

grid.arrange(widths = c(3, 3, 1),p13 + theme(legend.position="none"),
             p14 + theme(legend.position="none"), nrow= 1,legend)
```

```
p15 <- data %>%
  group_by(age, party_cd) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg,
         group = paste0(age, " with ", party_cd)) %>%
  dplyr::select(group = group, turnout_rate, age, party_cd) %>%
  ggplot( aes(x = group, y = turnout_rate, fill = party_cd)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Age * Party") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.3) +
  coord_flip() +
  ggtitle("Turnout rate vs. Age*Party") +
    theme(plot.title = element_text(hjust = 0.5))

p16 <- data %>%
  group_by(age, party_cd) %>%
  summarize(n = sum(total_reg), .groups = "drop") %>%
  mutate(group = paste0(age, " with ", party_cd)) %>%
  ggplot( aes(x = group, y = n, fill = party_cd)) +
  geom_bar(stat = "identity") +
  xlab("Age * Party") + ylab("Size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.05) +
  coord_flip() +
  ggtitle("Registered Voters vs. Age*Party") +
    theme(plot.title = element_text(hjust = 0.5))

legend <- get_legend(p15)

grid.arrange(widths = c(3, 3, 1),p15 + theme(legend.position="none"),
             p16 + theme(legend.position="none"), nrow= 1,legend)

p17 <- data %>%
  group_by(age, race_code) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg,
         group = paste0(age, " with ", race_code)) %>%
  dplyr::select(group = group, turnout_rate, age, race_code) %>%
  ggplot( aes(x = group, y = turnout_rate, fill = race_code)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Age * Race") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.3) +
  coord_flip() +
  ggtitle("Turnout rate vs. Age*Race") +
    theme(plot.title = element_text(hjust = 0.5))

p18 <- data %>%
  group_by(age, race_code) %>%
  summarize(n = sum(total_reg), .groups = "drop") %>%
```

```r
  mutate(group = paste0(age, " with ", race_code)) %>%
  ggplot( aes(x = group, y = n, fill = race_code)) +
  geom_bar(stat = "identity") +
  xlab("Age * Race") + ylab("Size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.05) +
  coord_flip() +
  ggtitle("Registered Voters vs. Age*Race") +
    theme(plot.title = element_text(hjust = 0.5))

legend <- get_legend(p17)

grid.arrange(widths = c(3, 3, 1),p17 + theme(legend.position="none"),
            p18 + theme(legend.position="none"), nrow= 1,legend)
```

```r
data <- data %>%
  mutate(total_vot = as.numeric(total_vot),
         total_reg = as.numeric(total_reg),
         county_desc = as.factor(county_desc),
         party_cd = as.factor(as.character(party_cd)),
         race_code = as.factor(race_code),
         sex_code = as.factor(sex_code),
         ethnic_code = as.factor(ethnic_code),
         age = as.factor(age))
```

```r
# initial
mod1 <- glmer(cbind(total_vot, total_reg - total_vot) ~
                   party_cd  + sex_code + age + (1|county_desc),
              data = data, family = binomial,
          control=glmerControl(optimizer = "bobyqa"))


mod2 <- glmer(cbind(total_vot, total_reg - total_vot) ~
                   party_cd + race_code  + sex_code + age + (1|county_desc),
              data = data, family = binomial,
          control=glmerControl(optimizer = "bobyqa"))


mod3 <- glmer(cbind(total_vot, total_reg - total_vot) ~
                   party_cd + race_code + ethnic_code + sex_code + age + (1|county_desc),
              data = data, family = binomial,
          control=glmerControl(optimizer = "bobyqa"))


mod4 <- glmer(cbind(total_vot, total_reg - total_vot) ~
                   -1 + race_code + ethnic_code + sex_code + age + (1 | county_desc),
              data = data, family = binomial,
          control=glmerControl(optimizer = "bobyqa", optCtrl=list(maxfun=2e5)))


mod5 <- glmer(cbind(total_vot, total_reg - total_vot) ~
                 party_cd + race_code + ethnic_code + sex_code + age +
                 sex_code:party_cd + age:party_cd + (1|county_desc),
```

```r
                  data = data, family = binomial,
               control=glmerControl(optimizer = "bobyqa", optCtrl=list(maxfun=2e5)))


mod6 <- glmer(cbind(cbind(total_vot, total_reg - total_vot)) ~
                  party_cd + race_code + ethnic_code + sex_code + age +
                  sex_code:party_cd + age:party_cd + age:race_code + (1|county_desc),
               data = data, family = binomial,
       control=glmerControl(optimizer = "bobyqa",optCtrl=list(maxfun=2e5)))


model <- c("Base model",
           "Add race",
           "Add race and ethnic",
           "Without intercept",
           "Add the interaction of sex:party, and age:party",
           "Add the interaction of sex:party, age:party, and age:race")
LRT <- c("",
         round(anova(mod1,mod2)$`Pr(>Chisq)`[2],4),
         round(anova(mod2,mod3)$`Pr(>Chisq)`[2],4),
         round(anova(mod3,mod4)$`Pr(>Chisq)`[2],4),
         round(anova(mod3,mod5)$`Pr(>Chisq)`[2],4),
         round(anova(mod3,mod6)$`Pr(>Chisq)`[2],4))
BIC_score <- sapply(c(mod1, mod2, mod3, mod4, mod5, mod6), BIC)
AIC_score <- sapply(c(mod1, mod2, mod3, mod4,mod5, mod6), AIC)
data.frame("Model" = model, 'LRT p-value' = LRT, 'AIC' = AIC_score, 'BIC' = BIC_score) %>%
  kable(caption = "Forward model selection") %>%
    kable_styling(latex_options = c("HOLD_position","striped"))


chart_fixef <- function(model) {
 summary(model)$coefficients %>%
    as.data.frame() %>%
    mutate(`95% lower` = round(Estimate - 1.96 * `Std. Error`, 4),
           `95% upper` = round(Estimate + 1.96 * `Std. Error`, 4)) %>%
    dplyr::select(Estimate, `Std. Error`, `95% lower`, `95% upper`)
}


# test <- binned_residuals(mod6)
binnedplot(fitted(mod6), y = resid(mod6))


# expand data from aggregated format
data_expand <- data %>%
  mutate(resp = map2(total_vot, total_reg, ~ c(
    rep(1, .x), rep(0, .y - .x)
  ))) %>%
  unnest(cols = c(resp)) %>%
  dplyr::select(-c(total_reg, total_vot))

# split into train and test
sample <- sample(c(TRUE, FALSE), nrow(data_expand), replace=TRUE, prob=c(0.7,0.3))
train <- data_expand[sample,]
```

```r
test <- data_expand[!sample,]

# fit model on train data

### fit any model you like on unaggregated train data
### will take a while...

predicted <- predict(mod6, test, type="response")
rocobj <- roc(test$resp, predicted)


ggroc(rocobj, size=1) +
  coord_fixed() +
  scale_x_reverse(
    name = "Specificity",
    limits = c(1,0),
    expand = c(0.001,0.001)
  ) +
  scale_y_continuous(
    name = "Sensitivity",
    limits = c(0,1),
    expand = c(0.001, 0.001)
  ) +
  geom_abline(intercept=1, slope=1, linetype="dashed", color="red") +
  labs(title=paste("AUC:", round(rocobj$auc, 3))) +
  theme_bw()
```

```r
chart_fixef(mod6)[2:5,] %>%
  kable(caption = "Estiamtes and 95% CI of Fixed Effects for Party",
        digits=4, format = "markdown") %>%
  kable_styling(latex_options = c("HOLD_position","striped"))
```

```r
chart_fixef(mod6)[6:11,] %>%
  kable(caption = "Estiamtes and 95% CI of Fixed Effects for Race",
        digits=4, format = "markdown") %>%
  kable_styling(full_width=FALSE,latex_options = c("HOLD_position", "striped")) %>%
    kable_classic()
```

```r
chart_fixef(mod6)[12:13,] %>%
  kable(caption = "Estiamtes and 95% CI of Fixed Effects for Ethnic",
        digits=4, format = "markdown") %>%
  kable_styling(full_width=FALSE,latex_options = c("HOLD_position", "striped")) %>%
    kable_classic()
```

```r
chart_fixef(mod6)[14:15,] %>%
  kable(caption = "Estiamtes and 95% CI of Fixed Effects for Sex",
        digits=4, format = "markdown") %>%
  kable_styling(full_width=FALSE,latex_options = c("HOLD_position", "striped")) %>%
    kable_classic()
```

```r
chart_fixef(mod6)[16:18,] %>%
  kable(caption = "Estiamtes and 95% CI of Fixed Effects for Age",
        digits=4, format = "markdown") %>%
  kable_styling(full_width=FALSE,latex_options = c("HOLD_position", "striped")) %>%
    kable_classic()
```

```r
dotplot(ranef(mod6))$county_desc
```

```r
chart_fixef(mod6)[19:26,] %>%
  kable(caption = "Estiamtes and 95% CI of Fixed Effects for Party:Sex",
        digits=4, format = "markdown") %>%
  kable_styling(full_width=FALSE,latex_options = c("HOLD_position", "striped")) %>%
    kable_classic()
```

```r
chart_fixef(mod6)[27:38,] %>%
  kable(caption = "Estiamtes and 95% CI of Fixed Effects for Party:Age",
        digits=4, format = "markdown") %>%
  kable_styling(full_width=FALSE,latex_options = c("HOLD_position", "striped")) %>%
    kable_classic()
```

```r
chart_fixef(mod6) %>%
  kable(caption = "Estiamtes and 95% CI of Fixed Effects", longtable = T,
        digits=4, format = "markdown") %>%
  kable_styling(full_width=FALSE,latex_options = c("HOLD_position", "repeat_header")) %>%
    kable_classic()
```

```r
ranef(mod6, condVar = TRUE) %>%
  as.data.frame() %>%
  arrange(desc(condval)) %>%
  mutate(`exp(condval)` = exp(condval)) %>%
  kable(caption = "Estimated random intercepts",longtable = T,
        digits =4, format = "markdown") %>%
  kable_styling(latex_options = c("HOLD_position", "repeat_header"))
```