Interim Report: Who Votes in NC?

Alexander Bendeck, Lynn Fan, Cathy Lee, Alice Liao, Justina Zou

24 October 2020

Introduction

The United States is getting closer to the 2020 Congressional Elections on November 3, 2020. With the current polarizing political landscape, the congressional election outcomes are significant to determine the next stage of this country. As it becomes critical for statisticians to help build forecasting models to predict the election outcomes, we need to first understand the patterns of voter turnout. Voting has special importance in democratic systems, but only about half of the eligible U.S. citizenry votes, and there are real political consequences when voters differ systematically from nonvoters (Uhlaner et al.). There has been abundant literature proving that variation in voter turnout will have electoral consequences (Hansford, et al.), in several ways. First of all, the media conventional wisdom is that "higher turnout benefits Democrats," although there has been mixed evidence about this theory (Weinschenk, 2019). Second, literature has proven certain demographic factors to statistically significantly benefit one party over the other, in both congressional elections and presidential races. For example, gender, race and party registration could help profile voting patterns for congressional elections (Uhlaner et al.). Election prediction models need to understand these patterns to predict the potential election outcomes.

Among all states, North Carolina has been a swing state in presidential and congressional elections for decades. In 2008, Obama won the state narrowly, but lost it narrowly after 4 years in 2012. Since 1996, the Republican statewide vote share in congressional elections has varied "from a low of 45% in 2008 to a high of 55% in 2014" (Perrin et al.). It makes North Carolina an interesting battleground in which voter demographic changes could potentially lead to significant impact on election outcomes and "an excellent site for those interested in partisan voting trends" (Perrin et al.). This report seeks to understand the voter turnout of North Carolina for 2020 NC Congressional Elections in order to predict who will vote in 2020.

Data Description

We are using public data provided by the NC State Board of Elections, which can all be accessed directly at the link https://dl.ncsbe.gov/list.html. The database contains voter history information for elections within the past 10 years in the NC voter history files, and all legally available voter specific information in the NC voter registration files. The NC voter registration files contain point-in-time snapshot voter registration data. For privacy concerns, names, birth dates and drivers license are not included, but the two types of files could be matched by North Carolina identification (NCID) number. The NC voter registration files have 8,007,192 data points available. The database was last updated on September 9, 2020. While we understand that voters might register later than that as the voter registration deadline for North Carolina is October 9, 2020, we believe it is sufficient to represent the majority of NC potential voters.

From the NC voter history files, we only kept the voters that voted for the 2016 general election for our analysis. Among 8,522,310 data points showing voters that voted either for the 2016 general election or the 2018 midterm election, we kept the 4,768,188 voter records for 2016. Studies have shown that presidential elections help mobilize voters, so voter turnout in presidential election years are significantly higher. In recent elections, voter turnout during presidential election years is around 60%, and only about 40% during midterm elections (FairVote.org). For North Carolina, voter turnout data in 2018 is also inappropriate to use because neither of

North Carolina's U.S. senators nor the governor was up for reelection, further demotivating voters (Perrin et al.). From the NC voter registration files, we filtered demographic factors that are supported by existing literature to be significant in understanding voting patterns, including gender, race, party registration, and age (Kim et al.). We also have their county and congressional district information available.

Additionally, we found relevant literature proving the relationship between voter turnout and wage (Charles et al.), so we added county-level median household income data from Economic Research Service under United States Department of Agriculture to our data set for analysis (https://www.ers.usda.gov/data-products/county-level-data-sets).

Data Munging

After binding NC voter history files and NC voter registration files by NCID and binding NC median household income by county, we started to process data for analysis. First of all, we identified those 4,389 data points older than 116 years old and removed them as the oldest person in NC is 116 years old and anyone older should be wrong data points. Many data points are also missing congressional district information. We imputed the missing districts by matching the voter's registered county with congressional district. We removed the 4% of voters who reside in counties that span across more than one congressional district. Because NC voter registration files show the voter registration data in 2020 while we are looking into voter turnout in 2016, it is important to treat those voters who were not eligible to vote in 2016. We subtract all voters' age by 4 years old to get the age information back in 2016, and then filtered out those 460,847 voters younger than 18 as they were unable to vote. We understand that doing so might introduce potential bias to the data set as the population 18-22 years old in 2020 might have different voting preferences than those 18-22 years old in 2016, but we believe the similarity of elections in 2016 and 2020 makes them comparable. Our **numerator** in voter turnout analysis is those who voted for the 2016 general election stored in NC voter history files, and our **denominator** is those registered voters in the NC voter registration files that were also eligible to vote back in 2016. The resulting binded data set has 7,229,595 rows.

After reading relevant literature in voter turnout analysis, we discovered that many papers used grouping to improve the estimation of the effects of individual and geographic predictors (Park, et al.). By aggregating demographics, researchers are also able to study collective political participation across groups, especially those who have been historically underrepresented (Uhlaner, et al.). Because we have eight million data points available, running models in a one-line-per-voter data set will also be very computationally expensive, and grouping can help us improve modeling efficiency. We instead decided to group data points by gender, race, party registration, county median income, and age, so that we can run models for the data set in a collapsed format. In the combined data set, there are party registrations for all parties, including the Libertarian Party and the Green Party. Because we are interested primarily only in the Republican Party and the Democratic Party and there are concerning class imbalance issues as the two parties take up the majority of registered voter population, we binded other parties as the third category `Other` for party. Similarly, because of class imbalance, we binded the races other than White and African Americans as `Other` for race as well. For those missing gender information, we binded them with `Unspecified`.

We divided (1) median county household income into four levels by the 25th, 50th, and 75th quantiles; (2) age into four levels for 18-29, 30-44, 45-59, and older than 60 years old, as it is a common way to analyze voter ages (McDonald, 2020); (3) gender into three categories, Female, Male and Other, and (4) race into three categories, Black, White, and Other. The final grouped data set has 432 subgroups (and therefore 432 rows).
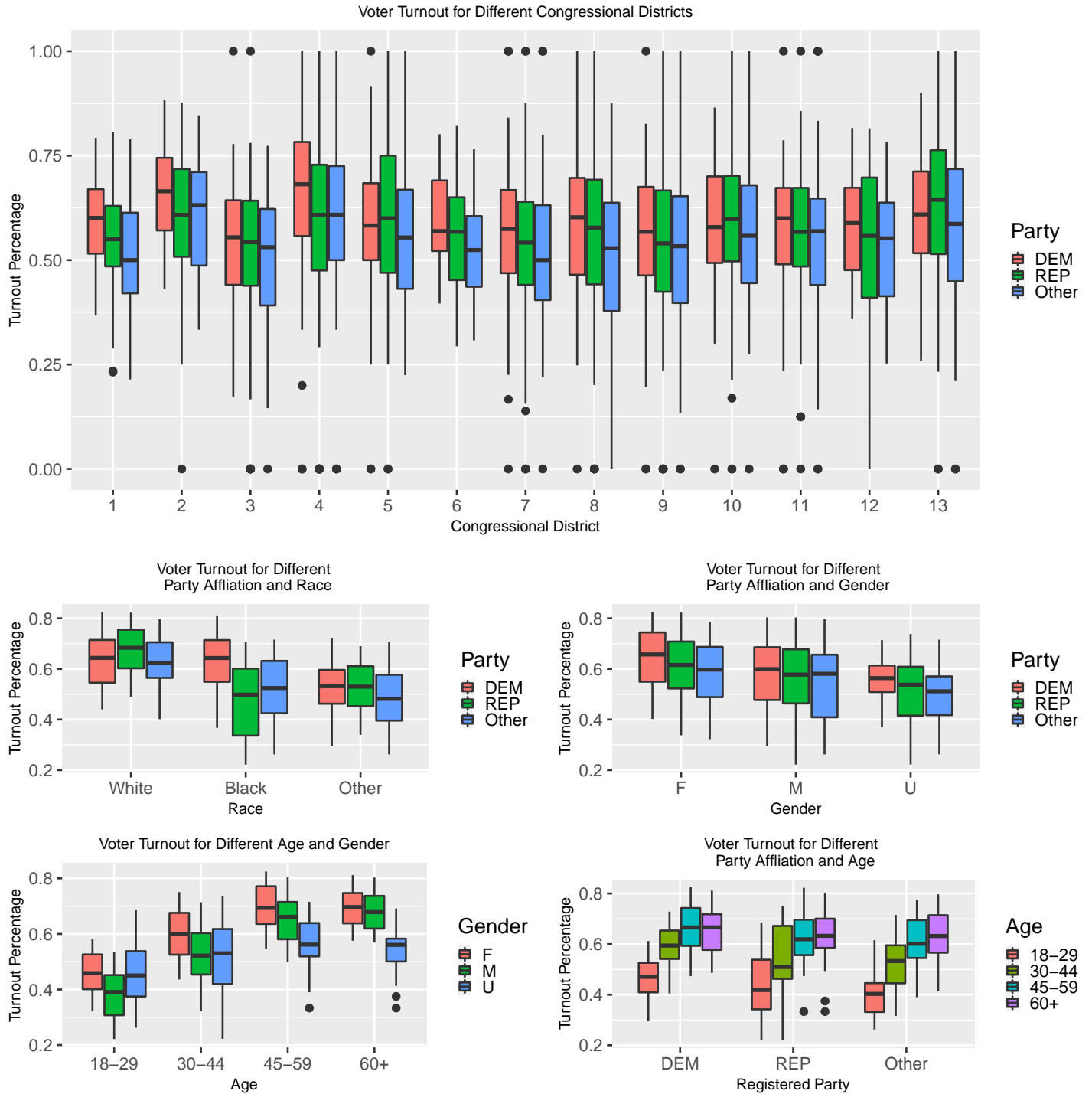
EDA

Figure 1: Grouped Feature Variable Exploration

After cleaning up data set and grouping demographic factors for analysis, Figure 1 aims to explore the the differences in voter turnout across different levels of party affiliation, gender, and age range. The upper left plot shows that white registered voters, on average, are more likely to vote than black registered voters, but black voters registered as Democrats have a comparable voter turnout as white voters. In the upper right plot, female registered voters are more likely to vote than male voters. The lower left plot shows that registered voters in the age range of 18-29 are the least likely to vote, which is consistent with existing literature (Kim et al.). Similar to the upper right plot, female voters have a higher voter turnout than male voters in all age groups displayed. The lower right plot shows a higher voter turnout among voters registered to be Democrats than those

Republicans. In the Appendix, exploratory data analysis for county median household income and ungrouped demographic variables are shown in the additional EDA section. After comparing the plots, we conclude that grouping accurately retains the differences of voter turnout across different levels of these predictors, and decide to continue using grouped data for the modeling process.

Method

We will take a Bayesian approach to not only predict if a voter with a certain profile would vote, but also understand quantitatively how the geographic and demographic information of a registered voter is associated with his or her likelihood of actually casting a ballot. Motivated by Y. Ghitza and A. Gelman's idea of grouping (2013), we divide the population into mutually exclusive categories according to their demographic and geographic characteristics and fit a Bayesian model with group-level predictors as well as their interactions. We will then be able to get average estimates for each of the subgroups from the posterior distributions. Bayesian viewpoint is an intuitive way of incorporating data to update the estimates and this method allows us to quantify the uncertainty about the model.

From Exploratory Data Analysis and literature, we identify voters' income (binned), gender, race, age (binned) and party affiliation as potentially useful predictors in predicting voter turnout. Furthermore, literature suggests possible interactions between some variables. Pew Research Center highlighted that historically party identification differs widely by gender (2020) and age (2016). J. Misra (2019) found while younger women voted at higher rates than younger men, women aged 65 and older voted at lower rates than men in the same age range. Walker et al (2019) found there are different effects across race and party subgroups on turnout especially in North Carolina due to its status as a swing state.

We decided to use a fixed-effect logistic regression model. Since we are analyzing whether or not an individual votes, the range that the response variable can take is between 0 and 1; therefore, logistic regression using grouped summary data is the most reasonable model. Coefficients and confidence intervals will help us answer the research question on potential associations of predictor variables. We propose a model with the following equation:

$$Vote_i \sim Binomial(n_i, \pi_i)$$

$$logit(\pi_i) = \alpha_0 + \Sigma_{s \in S}\beta_j * I(Income_i = Income\ Category_s) + \Sigma_{g \in G}\beta_g * I(Gender_i = Gender\ Category_g) +$$
$$\Sigma_{r \in R}\beta_r * I(Race_i = Race\ Category_r) + \Sigma_{a \in A}\beta_a * I(Age_i = Age\ Category_a) +$$
$$\Sigma_{p \in P}\beta_p * I(Party_i = Party\ Category_p) +$$
$$\Sigma_{g \in G, p \in P}\beta_{gp} * I(Gender_i = Gender\ Category_g)I(Party_i = Party\ Category_p) +$$
$$\Sigma_{r \in R, p \in P}\beta_{rp} * I(Race_i = Race\ Category_r)I(Party_i = Party\ Category_p) +$$
$$\Sigma_{g \in G, a \in A}\beta_{ga} * I(Gender_i = Gender\ Category_g)I(Age_i = Age\ Category_a) +$$
$$\Sigma_{p \in P, a \in A}\beta_{pa} * I(Party_i = Party\ Category_p)I(Age_i = Age\ Category_a)$$

where:

$$S = \{Median\ Income\ \$46864 - \$52798,\ Median\ Income\ \$52798 - \$64509,\ Median\ Income > \$64509\}$$
$$G = \{Male, Unspecified\}$$
$$R = \{Black, Other\}$$
$$A = \{Age\ 30 - 44,\ Age\ 45 - 59,\ Age\ 60+\}$$
$$P = \{Republican,\ Other\}$$

Where π_i is the probability that subgroup i votes. $Vote_i$ is the number of people in the ith subgroup that voted in the 2016 election. The baseline is `Median Income < $46864` for income, `Female` for gender, `White` for race, `18-29` for age, and `Democrat` for party affiliation.

Due to the lack of strong prior knowledge, we will use `brm` function's default priors: flat prior for all the β coefficients and a Student t distribution for α (i.e. $\alpha \sim Student_3(0, 2.5)$).

In a later section, we will compare this Bayesian model with two additional models: one is a frequentist logistic regression model with the same predictors and interactions and the other is a similar Bayesian model with an additional random effect at the congressional district level. In this way, we hope to assess if the Bayesian framework is superior to a frequentist approach when predicting voter turnout and if there is any salient unexplained variation within each congressional district.

Results and Interpretations

Table 1: Coefficient Table Part 1

	Est	SE	2.5% Q	97.5% Q
Intercept	0.00	0.00	-0.01	0.00
46.9k-52.8k	0.10	0.00	0.09	0.10
52.8k-64.5k	0.09	0.00	0.08	0.09
>64.6k	0.36	0.00	0.35	0.36
Male	-0.43	0.00	-0.44	-0.42
Unspecified	-0.08	0.01	-0.09	-0.06
Black	0.03	0.00	0.03	0.04
Other Race	-0.41	0.00	-0.42	-0.40
30-44	0.62	0.00	0.61	0.63
45-59	1.08	0.00	1.07	1.09
60+	0.94	0.00	0.93	0.95
Republican	0.19	0.01	0.18	0.20
Other Party	-0.24	0.00	-0.25	-0.23
Male:Rep	0.18	0.00	0.17	0.19
Unspec:Rep	0.10	0.01	0.08	0.13
Male:Other Pty	0.21	0.00	0.20	0.22
Unspec:Other Pty	0.01	0.01	-0.01	0.02

Table 2: Coefficient Table Part 2

	Est	SE	2.5% Q	97.5% Q
Black:Rep	-0.98	0.01	-1.00	-0.96
Other Race:Rep	-0.28	0.01	-0.30	-0.27
Black:Other Pty	-0.42	0.01	-0.43	-0.41
Other Race:Other Pty	-0.08	0.01	-0.10	-0.07
Male:30-44	0.02	0.00	0.01	0.03
Unspec:30-44	-0.41	0.01	-0.43	-0.40
Male:45-59	0.11	0.00	0.10	0.12
Unspec:45-59	-0.60	0.01	-0.62	-0.58
Male:60+	0.25	0.00	0.24	0.26
Unspec:60+	-0.47	0.01	-0.50	-0.45
30-44:Rep	0.04	0.01	0.03	0.05
45-59:Rep	-0.05	0.01	-0.06	-0.04
60+:Rep	-0.06	0.01	-0.07	-0.05
30-44:Other Pty	0.05	0.01	0.04	0.06
45-59:Other Pty	0.01	0.01	0.00	0.02
60+:Other Pty	0.27	0.01	0.25	0.28

Table 1 and Table 2 show that each coefficient has small standard errors (no indication of identifiability issues). In addition, we see that all variables except for the interaction between unspecified gender and third/unregistered party have credible intervals that do not contain 0, indicating significance in determining who votes. We are not particularly concerned about the this interaction not being significant since the interaction between gender and party for the other levels are significant, so we still include this interaction in the model. We can interpret the intercept and coefficients as follows:

- For a white, female Democrat aged 18-29 whose county median household income is in the lowest category ($< \$46.9k$), the odds of voting is 1 with a 95% CI of (0.9, 1).
- For a white, female Democrat aged 18-29 whose county median household income is in the second quartile ($\$46.9k - \$52.8k$), the odds of voting increases by a multiplicative factor 1.11 with a 95% CI of (1.09, 1.11) compared to the first quartile ($< \$46.9k$).
- Holding county median household income ($< \$46.9k$), race (White), party (Democrat), and age (18-29) at baseline, the odds of voting among men is 0.65 times that among women with a 95% CI of (0.64, 0.66). We can also note that, holding county median household income and age at baseline, the odds of voting among black Republican men is 0.29 times that among black Republican women.

- For white, female Democrat voters with county median household income less than $\$46.9k$, the odds of voting among those aged 30-44 would be 1.86 times higher than among those aged 18-29 with 95% CI of (1.84, 1.88). In fact, if we hold all other variables at baseline, the odds of voting is the lowest in the

lowest age group. The odds of voting among older (aged 60+) Republican white male voters with median income less than $46.9k$ is 2.92 times more likely than the odds of voting among voters with all baseline demographic characteristics.
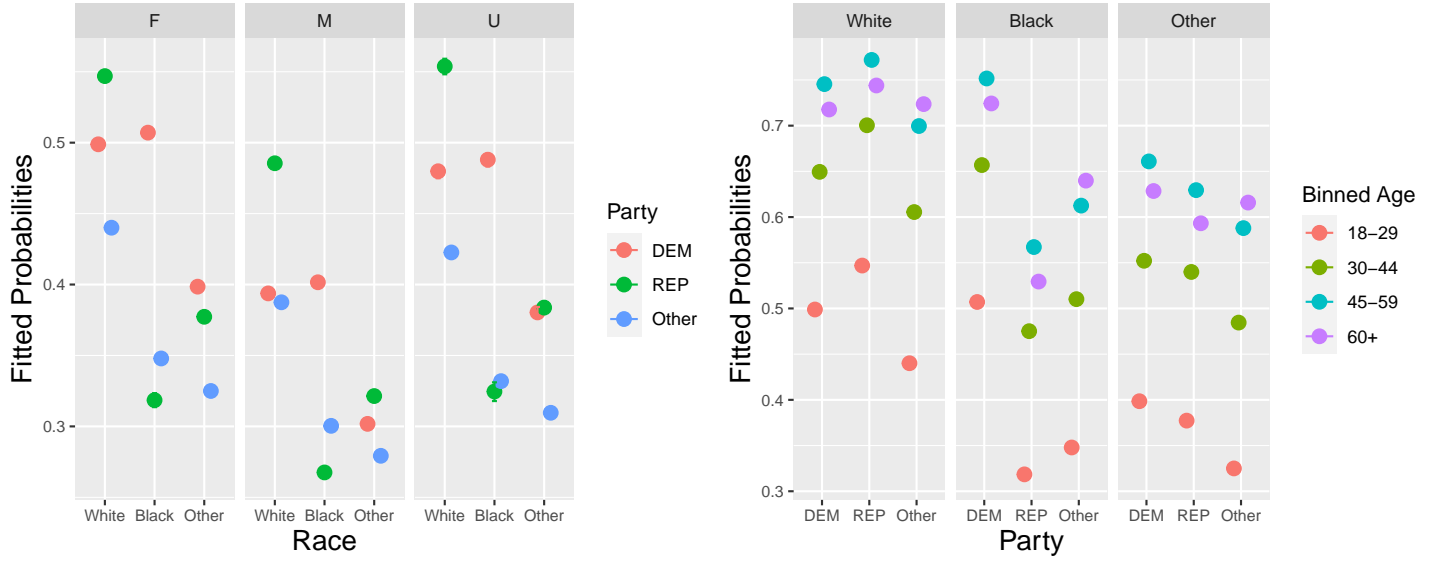

Figure 2: Marginal Effect of Party by Age and Race

We can interpret each point in the left plot of Figure 2 as follows: holding county median household income at baseline (less than $46,864$) and age at baseline (ages 18-29), the y-axis value is the expected probability that a person of a particular race (x-axis), party (color), and gender (facet) votes. For example, the expected probability that a black, male, Democrat votes is 0.4, whereas the expected probability that a black, male, Republican votes is approximately 0.28. We can also see that women, regardless of race and party, are expected to be more likely to vote than men. The plot on the right in Figure 2 shows that, holding county median household income at baseline (less than $46,864$) and gender at baseline (female), the expected probability of voting for white Democrats within each age group is less than that for white Republicans. However, the expected probability of voting among black Democrats within each age group is higher than that among black Republicans.

Model Validation

Figure 3 displays the posterior predictive checks for our model. On the left, the dark black line is the observed distribution and the light blue line is the simulated distribution. On the right, the black dots are the actual voter turnout for each subgroup, and the grey points are the predicted turnout with 95% credible intervals. The plots show that our simulated distribution is similar to the observed distribution. Our simulations also produce few outliers. It appears that most of the true values are covered by the 95% confidence intervals of the predicted values, indicating model fit.

Most of the standardized residuals (shown in Figure 4) are within $\pm$ 2, but some points have somewhat larger values. This means that for the majority of the groups, the model predicts fairly well.

We conducted external validation on the 2012 dataset, cleaned the same way the 2016 dataset was cleaned. The RMSE for the 2012 dataset is 1660.38, which is larger than the RMSE from internal validation for the 2016 dataset, 613.52. This is to be expected because the candidates on the ballot between the two elections may mobilize different types of voters. For example, Hillary Clinton being on the ballot prompted many women to come out to vote, whereas Barack Obama in the 2012 election prompted many African Americans to vote (Weinschenk, et al).
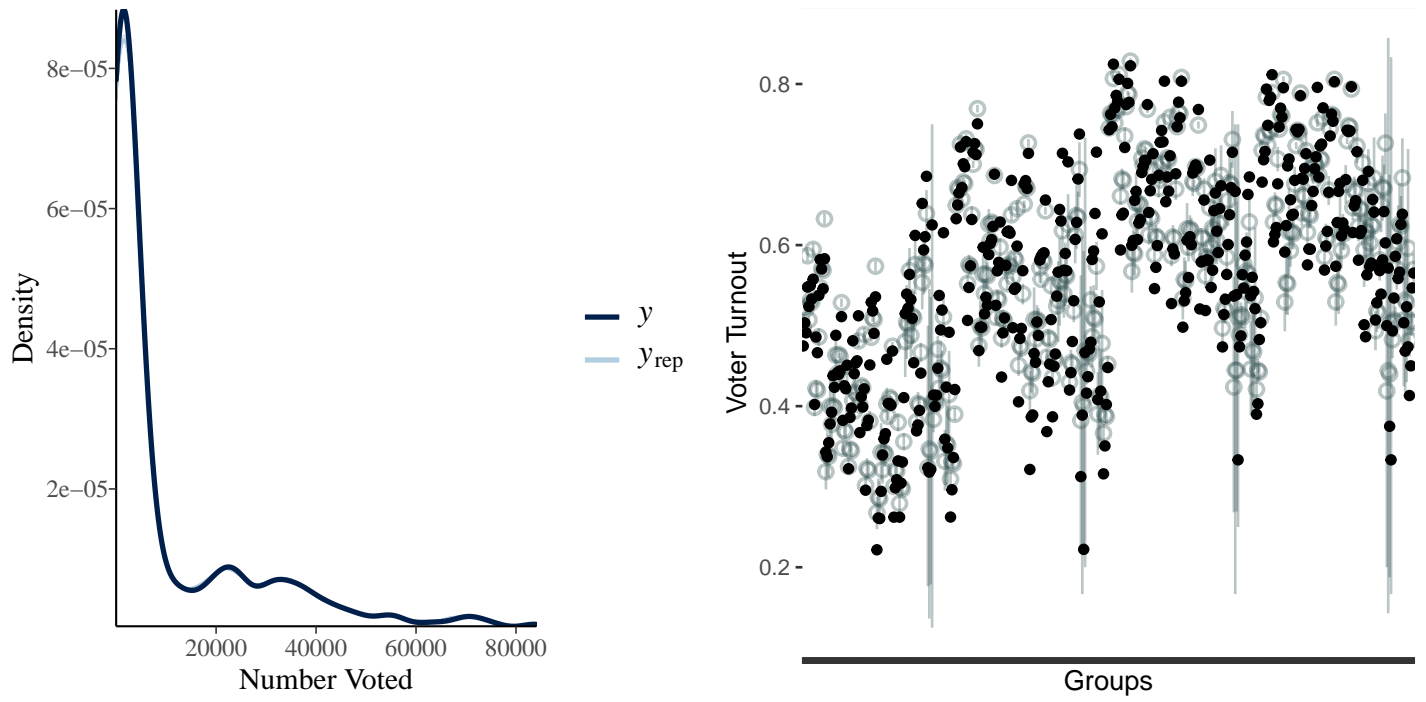
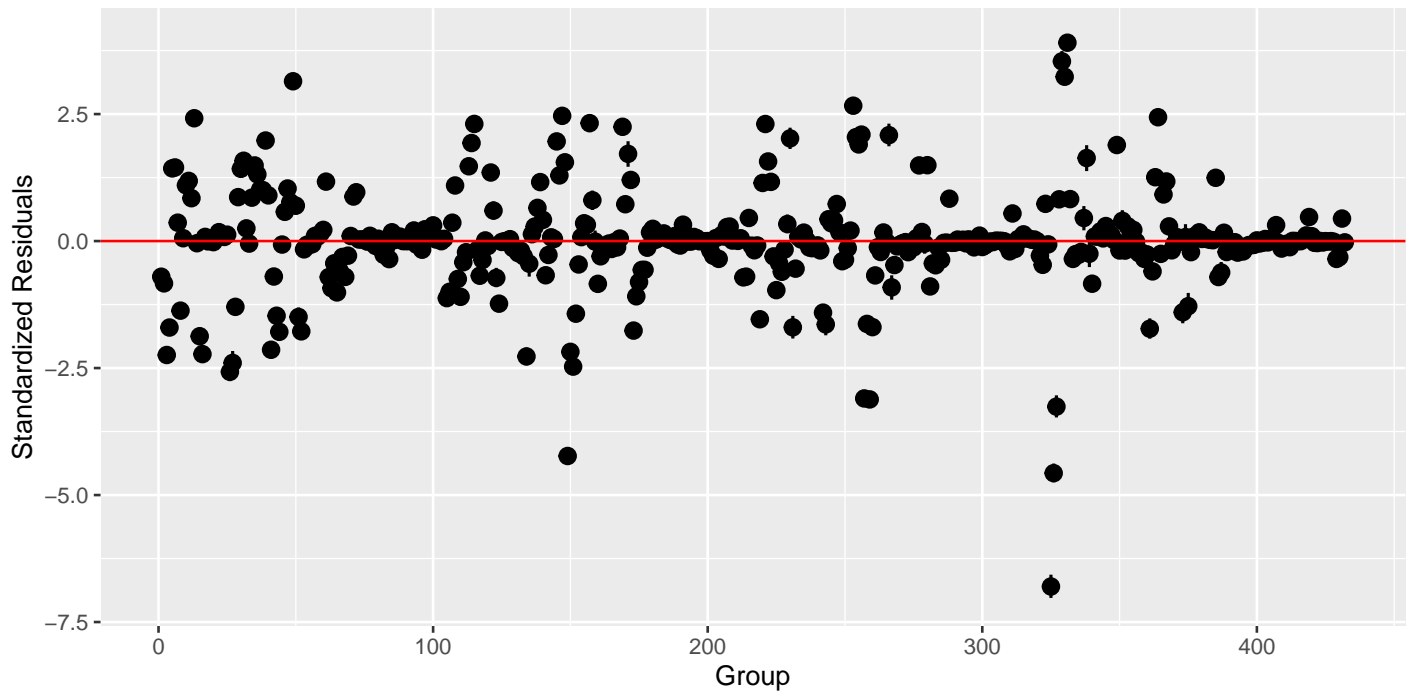Figure 3: Posterior Predictive Checks Plots

Figure 4: Residual Plot

Sensitivity Analysis

As part of our sensitivity analysis, we fit a model with identical predictor variables to our main model but with different priors. Based on literature review, we chose normal priors for the main effects that aligned with expectations that older voters tend to vote more often (McDonald, 2020), men tend to vote less often than women (Igielnik, 2020), party-unaffiliated voters tend to vote less often (Killian, 2012), and wealth tends to correlate positively with voter turnout (Akee, 2019). For these main effects, we chose the normal priors to have means somewhere between 0.5 and 1, and we chose a variance of 3. For all other terms (including interactions and intercept), we chose normal priors with a mean of 0 and a variance of 3. Changing these priors had minimal effect on the resulting coefficients and confidence intervals (see model output in Appendix).

We also fit another model with random intercept and slopes with all of the predictors for congressional districts to capture possible variation between voters in different geographic parts of North Carolina. The coefficients for the main effects (included in the Appendix, along with group-level standard deviation) look largely the same compared to the main model, with the exception that the coefficients for income are slightly different in the two models. This is perhaps because some of the variations between different income groups can be attributed to differences between counties. We did not choose this model as our main model because it has a higher WAIC (246407584) compared to the main model (25039.1). Although the RMSE when predicting on 2012 data for the random effects model (285.9272189) is better, we ultimately decided to keep the main effects model because of the higher level of stability in its coefficients. The main effects model also had smaller residuals and better coverage (see Appendix for residual plot and coverage plot for the random effects model).

With the addition of random slopes, we see that for black race, congressional districts 1, 2, 8, 10, and 11 have random slopes that are significant (determined by the fact that their 95% CI's do not contain zero). Looking at both main and random effects, holding all other variables at baseline, the expected odds of voting increases by a multiplicative factor for black people from districts 8 and 10, while those odds decreases by a multiplicative factor for black people from districts 1, 2, and 11. For the 45-59 age group, the random slopes for districts 2, 3, 4, 7, 8, 9, and 11 are significant, and the expected odds of voting increases by a multiplicative factor for 45-59 year-olds from those districts, holding all else at baseline. For the Republican party, random slopes for districts 2, 4, 5, 8, 10, and 13 are significant, and the expected odds of voting increases by a multiplicative factor for Republicans in those districts, holding all else at baseline.

Finally, since all of the models discussed above were fit using a Bayesian framework via the `brms` package, we decided to fit a frequentist model for comparison. The model output (included in the Appendix) shows little difference from the results of our main Bayesian model.

Conclusion and Limitations

We fit a Bayesian logistic regression model to model the binary outcome of whether or not a demographic subgroup, defined by their race, gender, party affiliation, age, and median county household income (such as white female Democrats between 30-44 years of age whose county median household income is greater than $64,509) votes. Based on results from literature review, we added interactions between gender and party, race and party, age and gender, and age and party. After model comparison, we did not include random effects in our model. Our MCMC diagnostics and model diagnostics do not indicate a lack of fit and show that our model coefficients converge well.

All of the demographic variables, as well as the interactions, were significant except for the interaction between unspecified gender and the third/unspecified party category. From the plots of the marginal effects of the covariates, holding county median household income at the $< \$46,864$ category and holding age at the 18-29 level, we found that females have a higher expected probability of voting than male registered voters. For those with the same county median household income as above and are female, we note that among white voters, Republicans are more likely to vote than Democrats, while the opposite is true among black voters. We also found that for those whose county median household income falls in the highest category ($> \$64,509$), the expected probability of voting is larger relative to any other income category.

The biggest strength of this study is its focus on grouping data. Grouping data to analyze voting patterns makes results helpful for future analysis of the same demographic groups, and candidates can use this result to adjust their resource allocation during campaigns. Another strength is the substantial amount of diagnostics, validation, and sensitivity analysis done to evaluate model fit and demonstrate the robustness of our final model.

Many limitations were already addressed previously. Others include the weaknesses of the data set, namely that the voting patterns in 2016 will not completely match that of 2020 even though Trump has been the Republican candidate for both races. This is because the Democratic candidate may mobilize voters differently and the political landscape has changed between 2016 and 2020 (Weinschenk et al.). Also, grouping prevents us from analyzing individual-level demographic variations and predictors not yet included in our analysis. As voter turnout is a very complicated subject and a lot of predictors might be relevant in shaping voting patterns, we might be oversimplifying the study with the number of predictors we have on the group level.

Finally, while demographic variables are key to explain election voter turnout, elections are complicated events and more facets of elections and diverse data sources, analyzed by different analyses, are necessary to understand voter turnout and election outcomes (Weinschenk et al.).
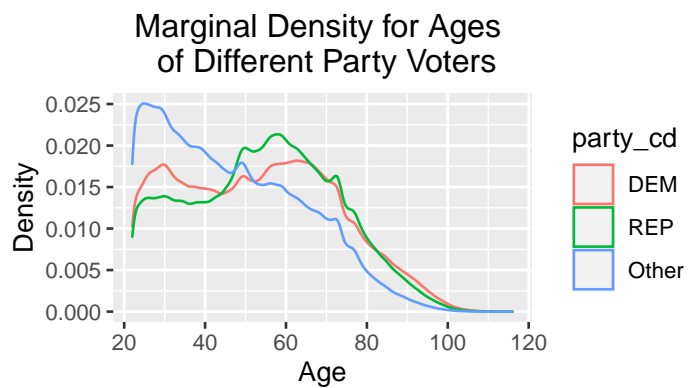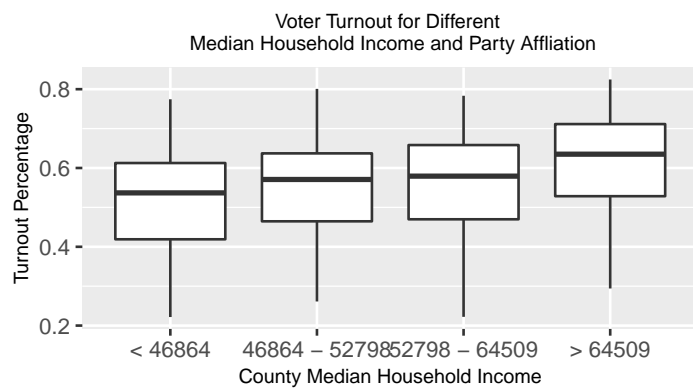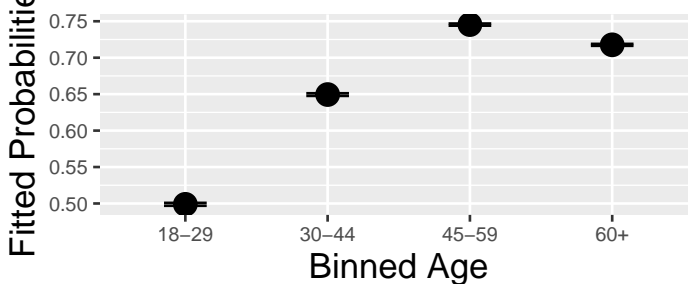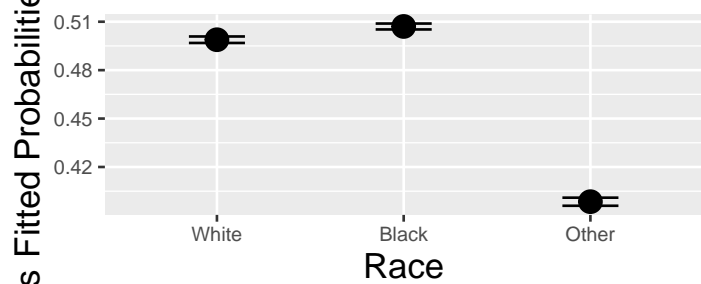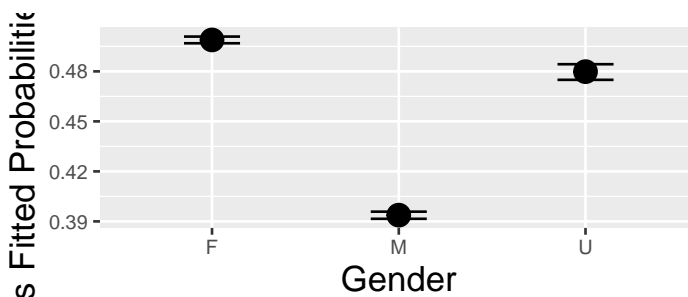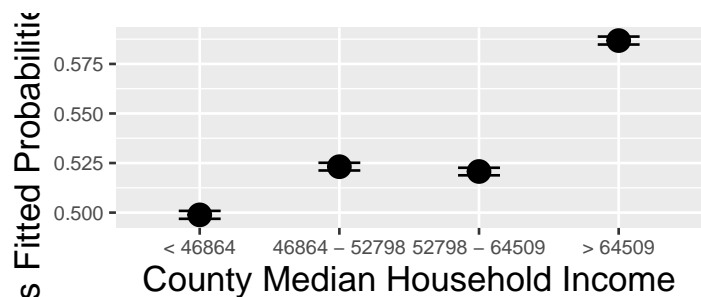
Appendix

References

1. Linzer, D. A. (2013). Dynamic Bayesian Forecasting of Presidential Elections in the States. Journal of the American Statistical Association, 108(501), 124-134. doi:10.1080/01621459.2012.737735

2. Hansford, T. G., & Gomez, B. T. (2010). Estimating the Electoral Effects of Voter Turnout. American Political Science Review, 104(2), 268-288. doi:10.1017/s0003055410000109

3. 2020 Election. (2020, October 20). Retrieved October 20, 2020, from https://fivethirtyeight.com/politics/elections/

4. Park, D. K., Gelman, A., & Bafumi, J. (2006). State-Level Opinions from National Surveys:. Public Opinion in State Politics, 209-228. doi:10.2307/j.ctvr33bdg.17

5. Mahler, V. A., Jesuit, D. K., & Paradowski, P. R. (2013). Electoral Turnout and State Redistribution. Political Research Quarterly, 67(2), 361-373. doi:10.1177/1065912913509306

6. Uhlaner, C. J., & Scola, B. (2015). Collective Representation as a Mobilizer. State Politics & Policy Quarterly, 16(2), 227-263. doi:10.1177/1532440015603576

7. Godbout, J. (2012). Turnout and presidential coattails in congressional elections. Public Choice, 157(1-2), 333-356. doi:10.1007/s11127-012-9947-7

8. Kim, S. S., Alvarez, R. M., & Ramirez, C. M. (2020). Who Voted in 2016? Using Fuzzy Forests to Understand Voter Turnout. doi:10.33774/apsa-2020-xzx29

9. Weinschenk, A. C. (2019) That's Why the Lady Lost to the Trump: Demographics and the 2016 Presidential Election, Journal of Political Marketing, 18:1-2, 69-91, DOI: 10.1080/15377857.2018.1478657

10. Charles, K. K., & Stephens, M. (2011). Employment, Wages and Voter Turnout. doi:10.3386/w17270

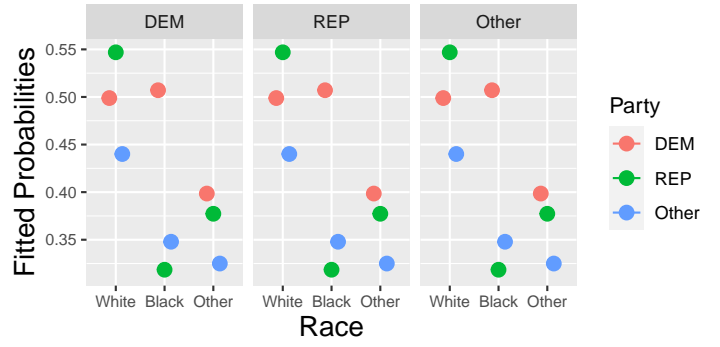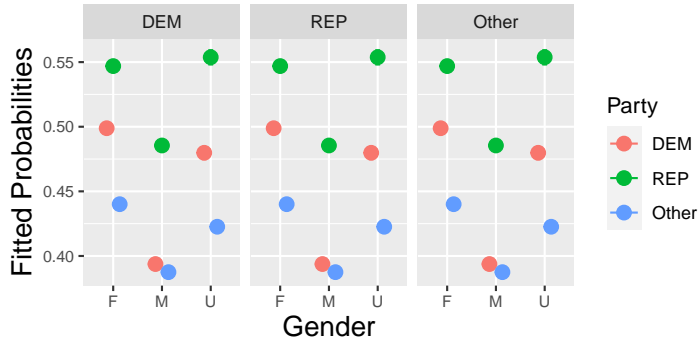11. Hills, M. (2020, September 25). US election 2020: A really simple guide. Retrieved October 20, 2020, from https://www.bbc.com/news/election-us-2020-53785985
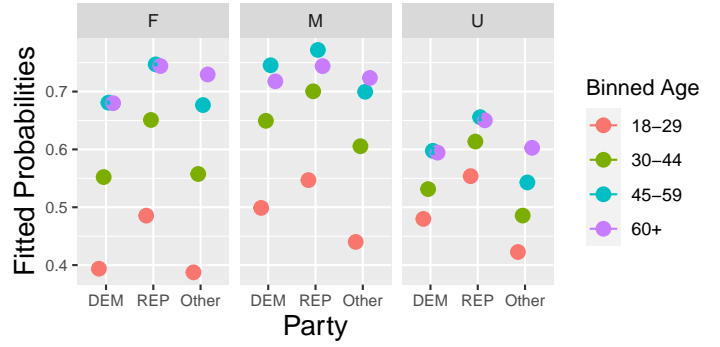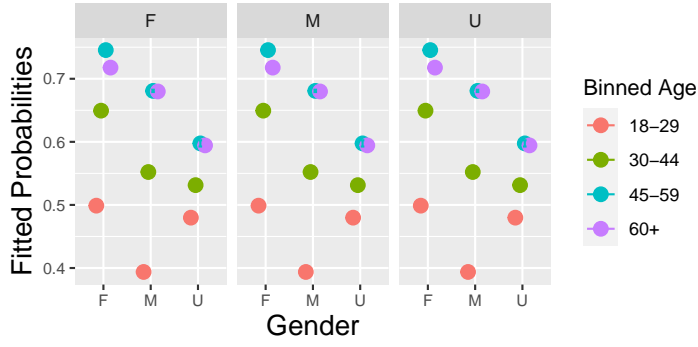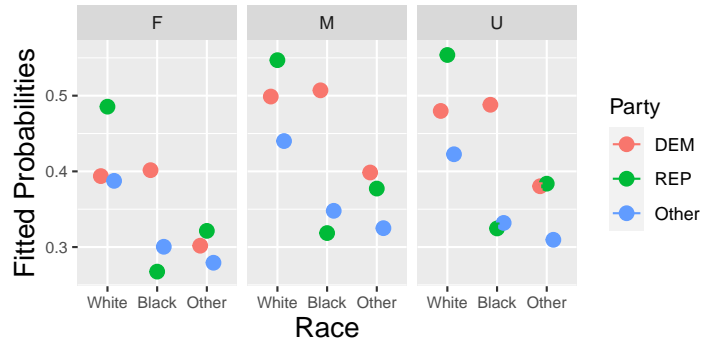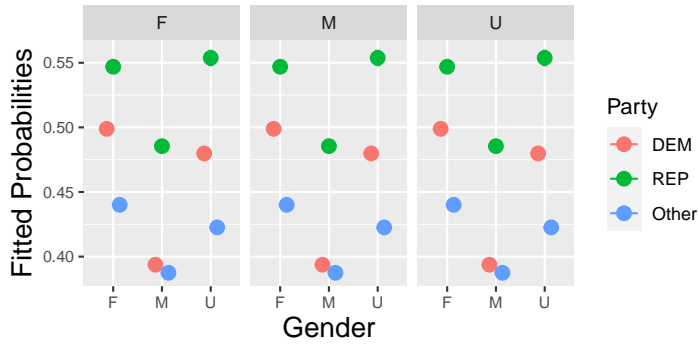
12. Railey, K. (2016). Federal Judges Let Stand North Carolina,'s New Congressional Map. The Hotline. https://link.gale.com/apps/doc/A498010836/ITOF?u=duke_perkins&sid=ITOF&xid=119d6ad9

13. Perrin, A. J., & Ifatunji, M. A. (2020). Race, Immigration, and Support for Donald Trump: Evidence From the 2018 North Carolina Election. Sociological Forum, 35(S1), 941-953. doi:10.1111/socf.12600

14. Redistricting in North Carolina. (2020). Retrieved October 21, 2020, from https://ballotpedia.org/Redistricting_in_North_Carolina

15. § 132-1. Public Records. https://www.ncleg.gov/EnactedLegislation/Statutes/PDF/BySection/Chapter_132/GS_132-1.pdf

16. § 163-82.10. Official record of voter registration. https://www.ncleg.gov/EnactedLegislation/Statutes/PDF/BySection/Chapter_163/GS_163-82.10.pdf

17. FairVote.org. (n.d.). Voter Turnout. Retrieved October 22, 2020, from https://www.fairvote.org/voter_turnout

18. McDonald, M. P. (2020). Voter Turnout Demographics. Retrieved October 23, 2020, from http://www.electproject.org/home/voter-turnout/demographics

19. Y. Ghitza, A. Gelman's (2013). Deep interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups. American Journal of Political Science 57 762– 776.

20. Walker, H.L., Herron, M.C. & Smith, D.A. Early Voting Changes and Voter Turnout: North Carolina in the 2016 General Election. Polit Behav 41, 841–869 (2019). https://doi-org.proxy.lib.duke.edu/10.1007/s11109-018-9473-5

21. Pew Research Center (2020). Men and women in the U.S. continue to differ in voter turnout rate, party identification. https://www.pewresearch.org/fact-tank/2020/08/18/men-and-women-in-the-u-s-continue-to-differ-in-voter-turnout-rate-party-identification/

22. Pew Research Center (2016). Party affiliation among voters: 1992-2016. https://www.pewresearch.org/politics/2016/09/13/2-party-affiliation-among-voters-1992-2016/

23. J. Misra (2019). Voter Turnout Rates Among All Voting Age and Major Racial and Ethnic Groups Were Higher Than in 2014. https://www.pewresearch.org/fact-tank/2020/08/18/men-and-women-in-the-u-s-continue-to-differ-in-voter-turnout-rate-party-identification/

24. Killian, L. (2012). Five Myths About Independent Voters. Retrieved October 22, 2020, from https://www.washingtonpost.com/opinions/five-myths-about-independent-voters/2012/05/17/gIQAZmGyWU_story.html

25. Igielnik, R. (2020). Men and Women in the U.S. Continue to Differ in Voter Turnout Rate. Retrieved October 22, 2020, from https://www.pewresearch.org/fact-tank/2020/08/18/men-and-women-in-the-u-s-continue-to-differ-in-voter-turnout-rate-party-identification/

26. Akee, R. (2019). Voting and Income. Retrieved October 22, 2020, from https://econofact.org/voting-and-income
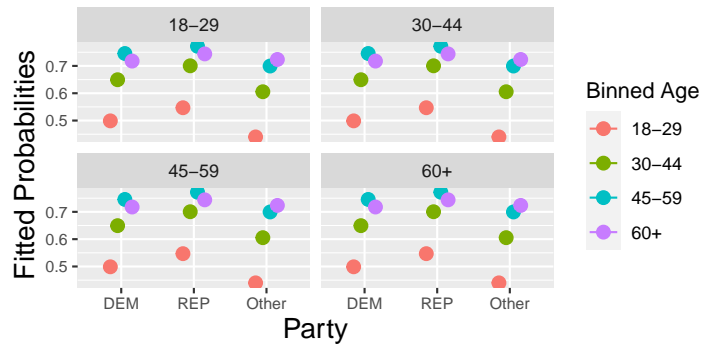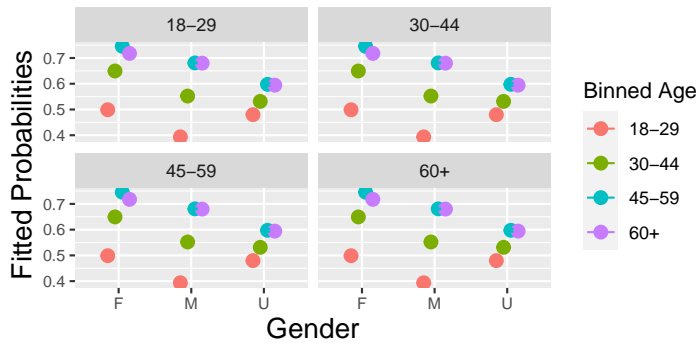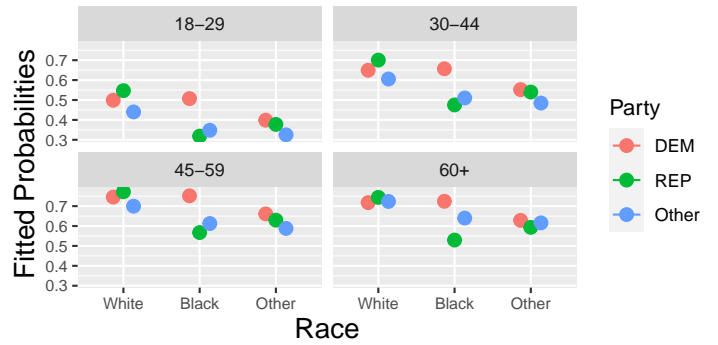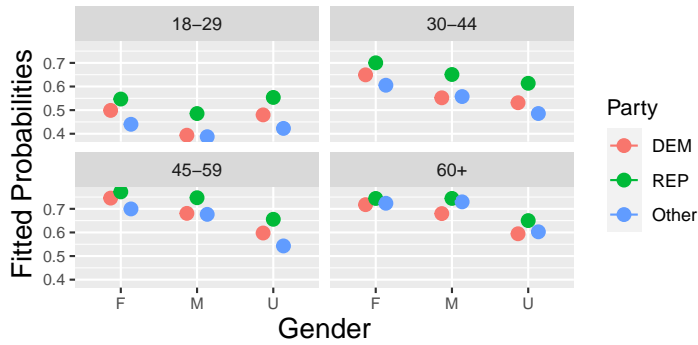
Additional EDA

Voter Turnout for Different Median Household Income and Party Affliation

Marginal Density for Ages of Different Party Voters

Race v. Party

Age v. Races
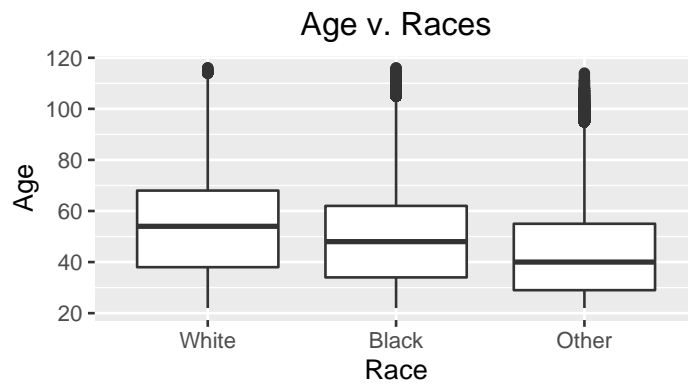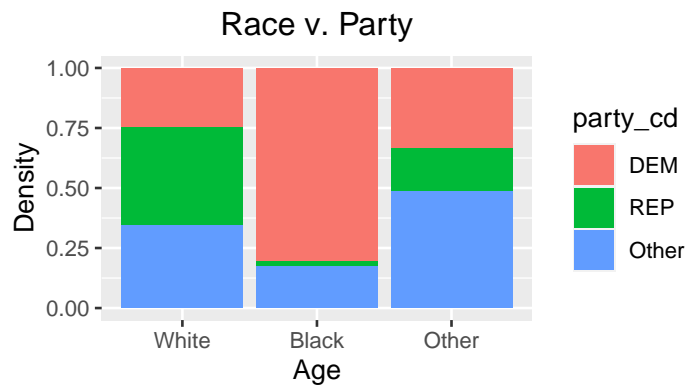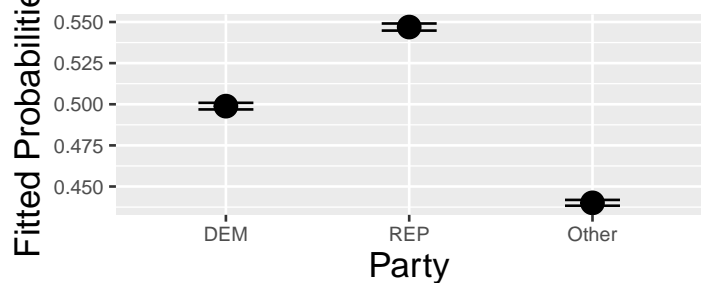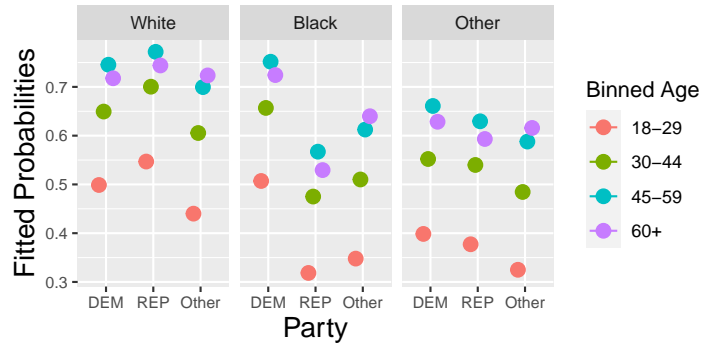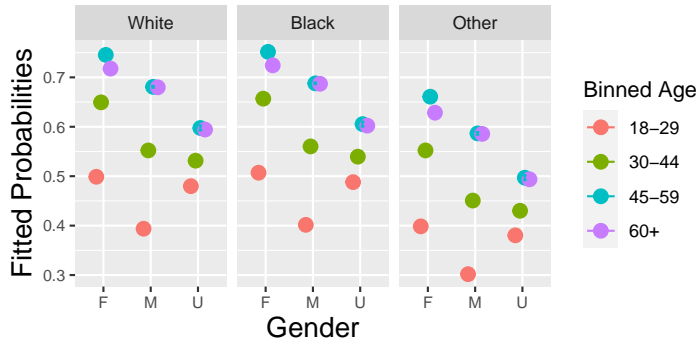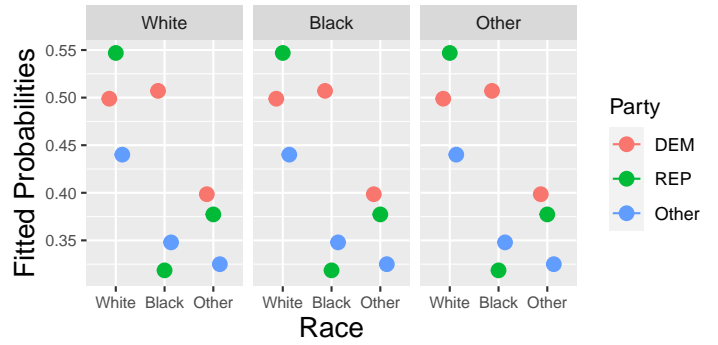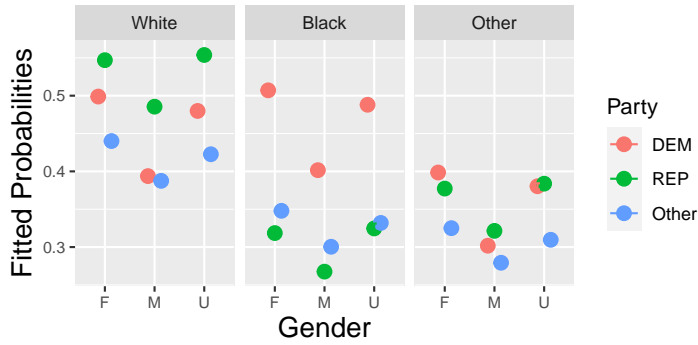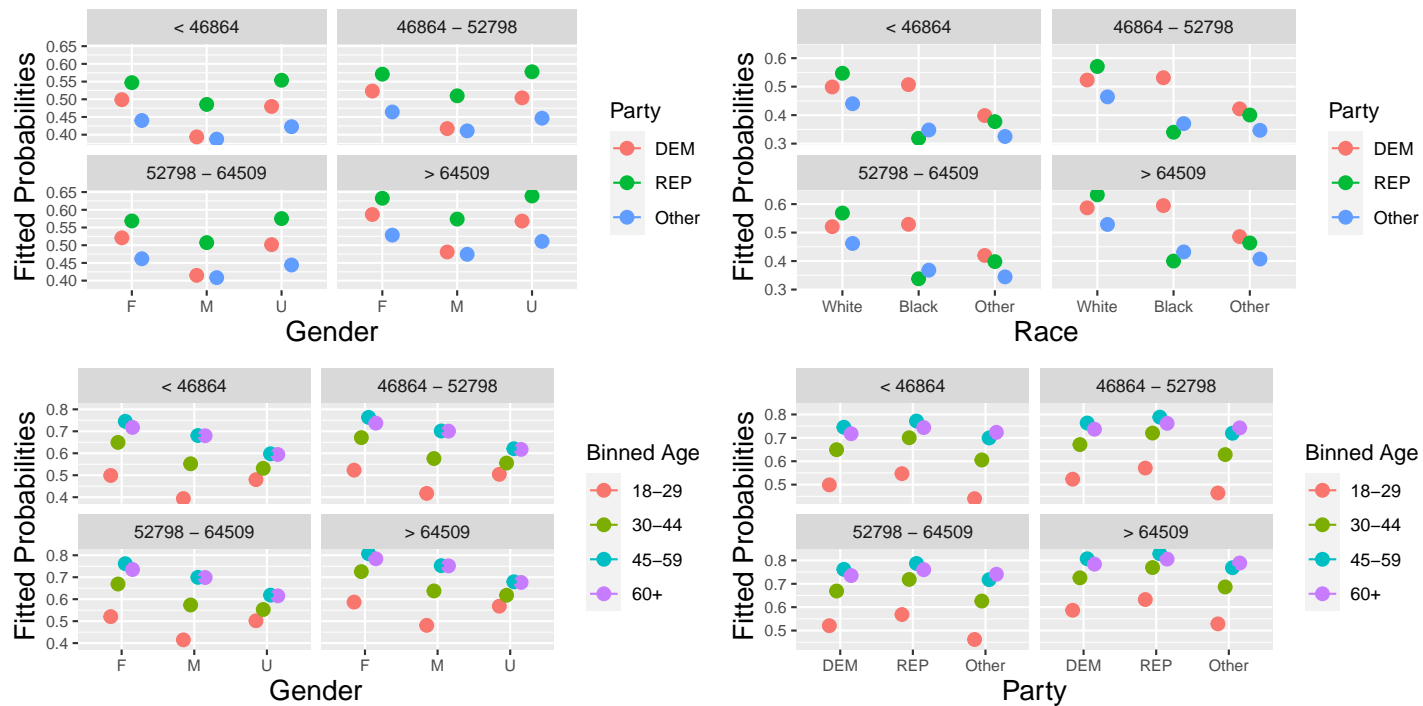
Plots

Traceplots
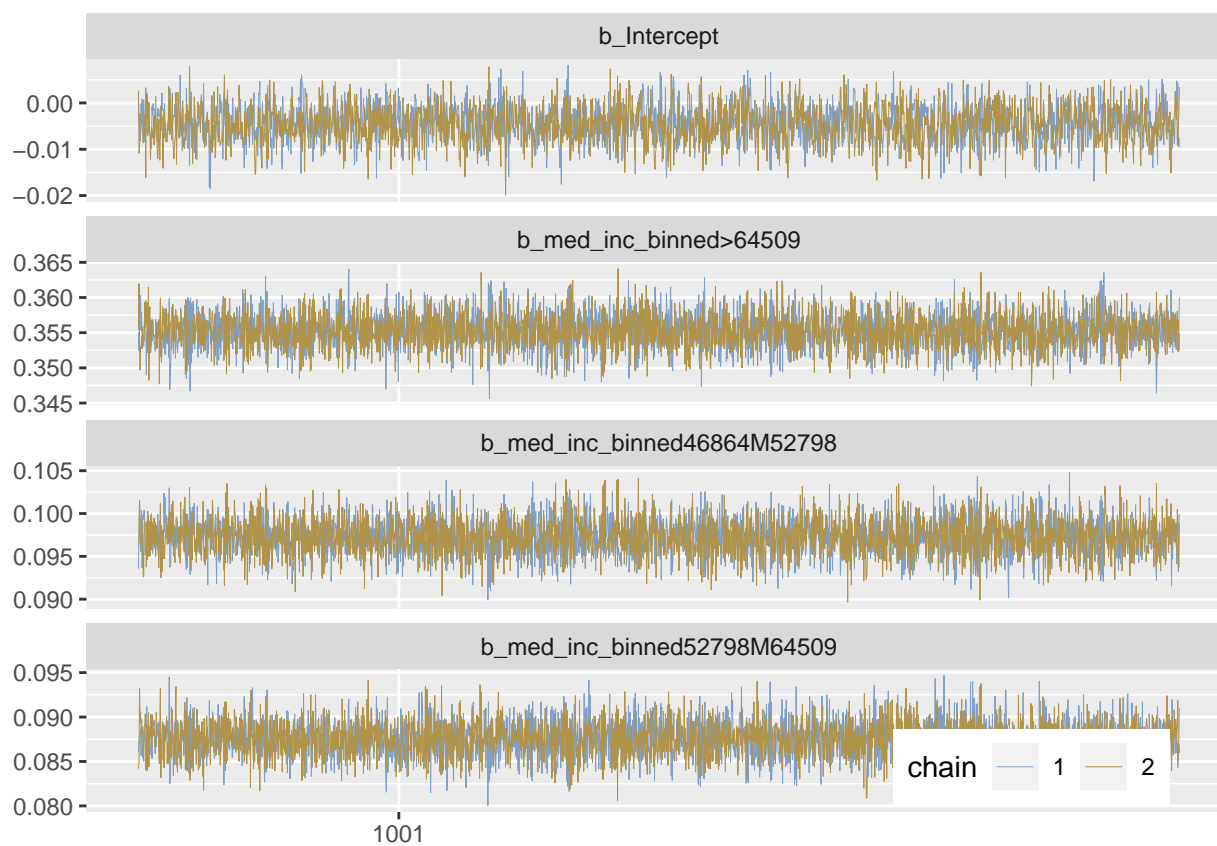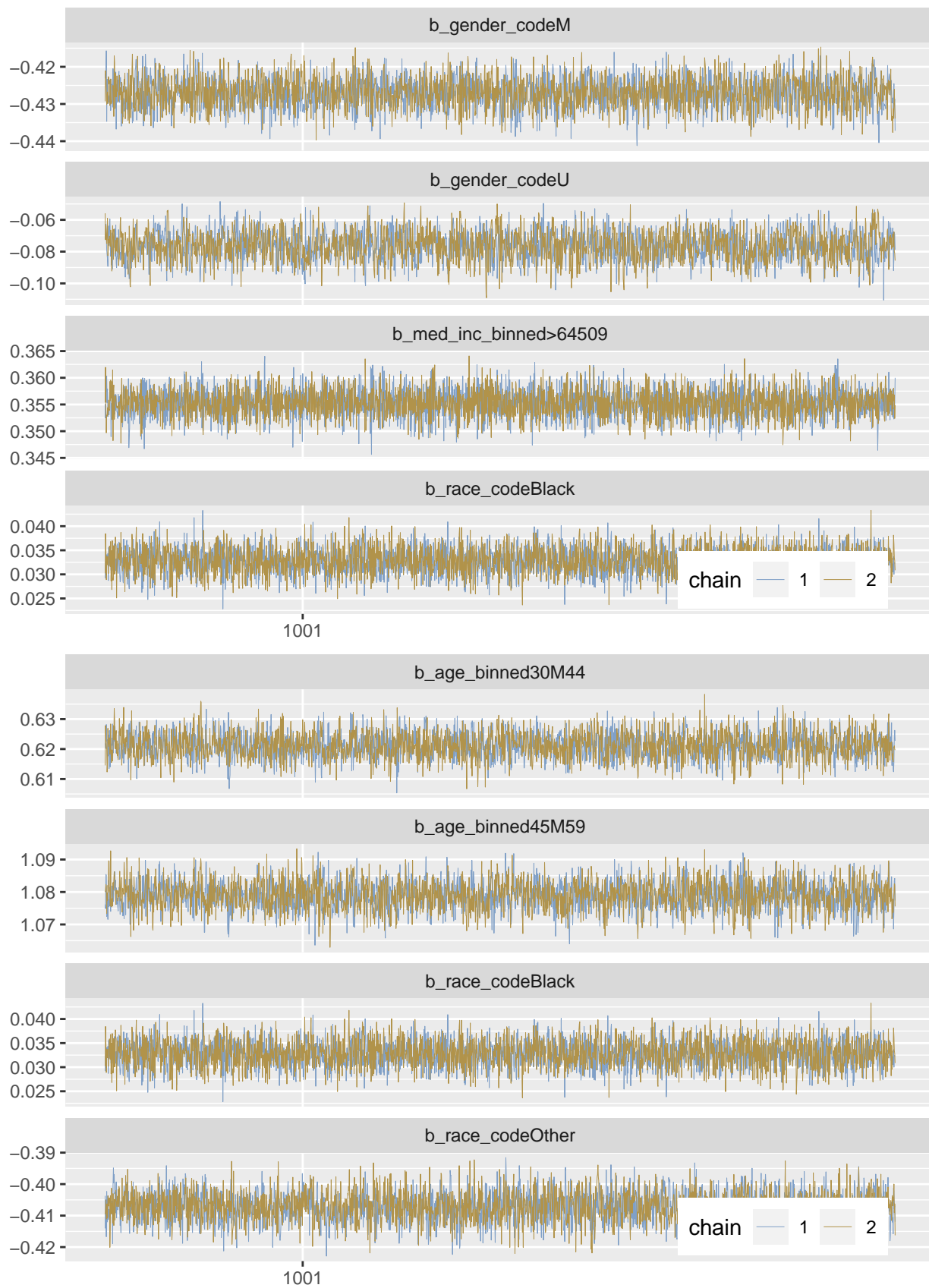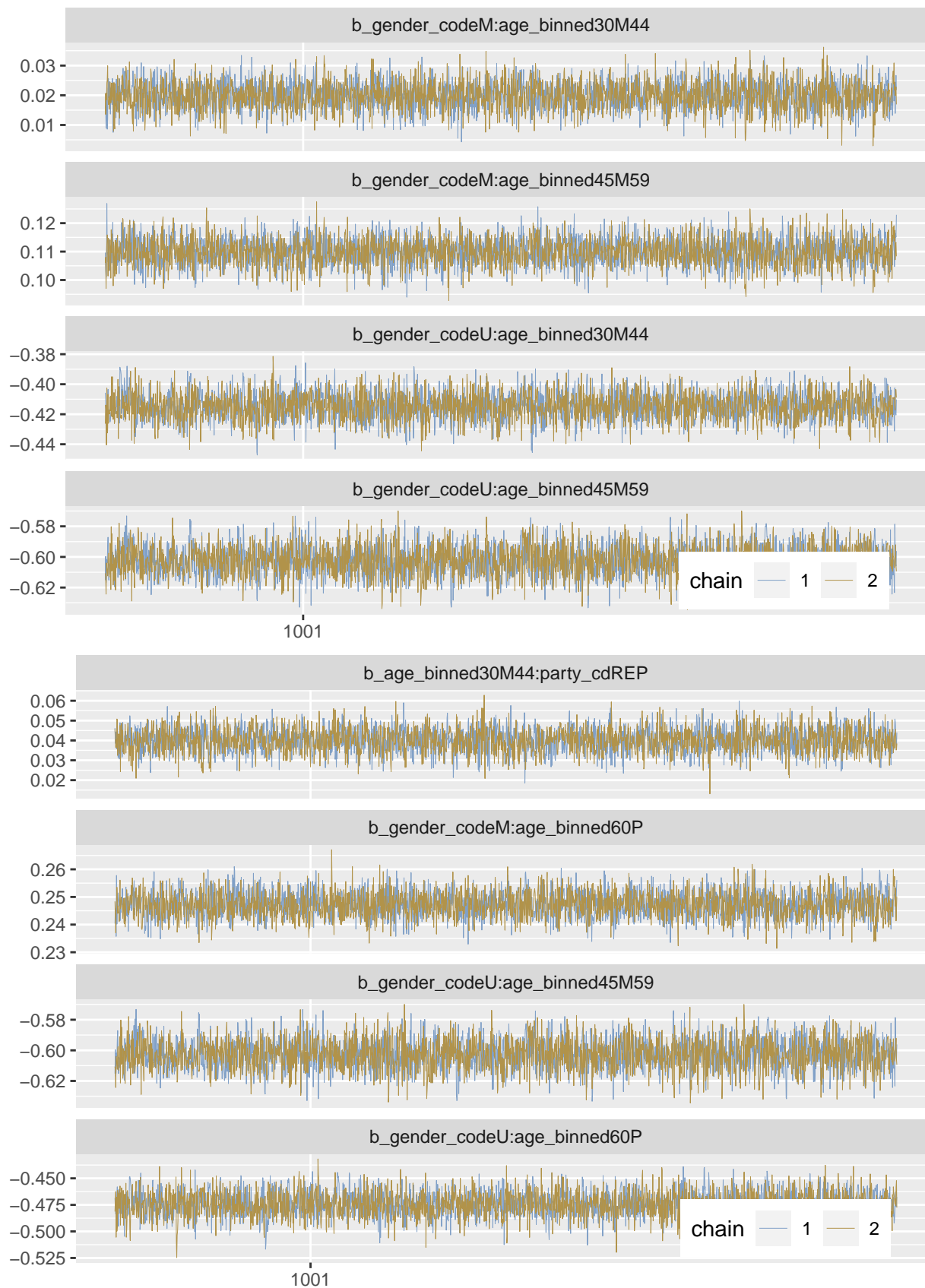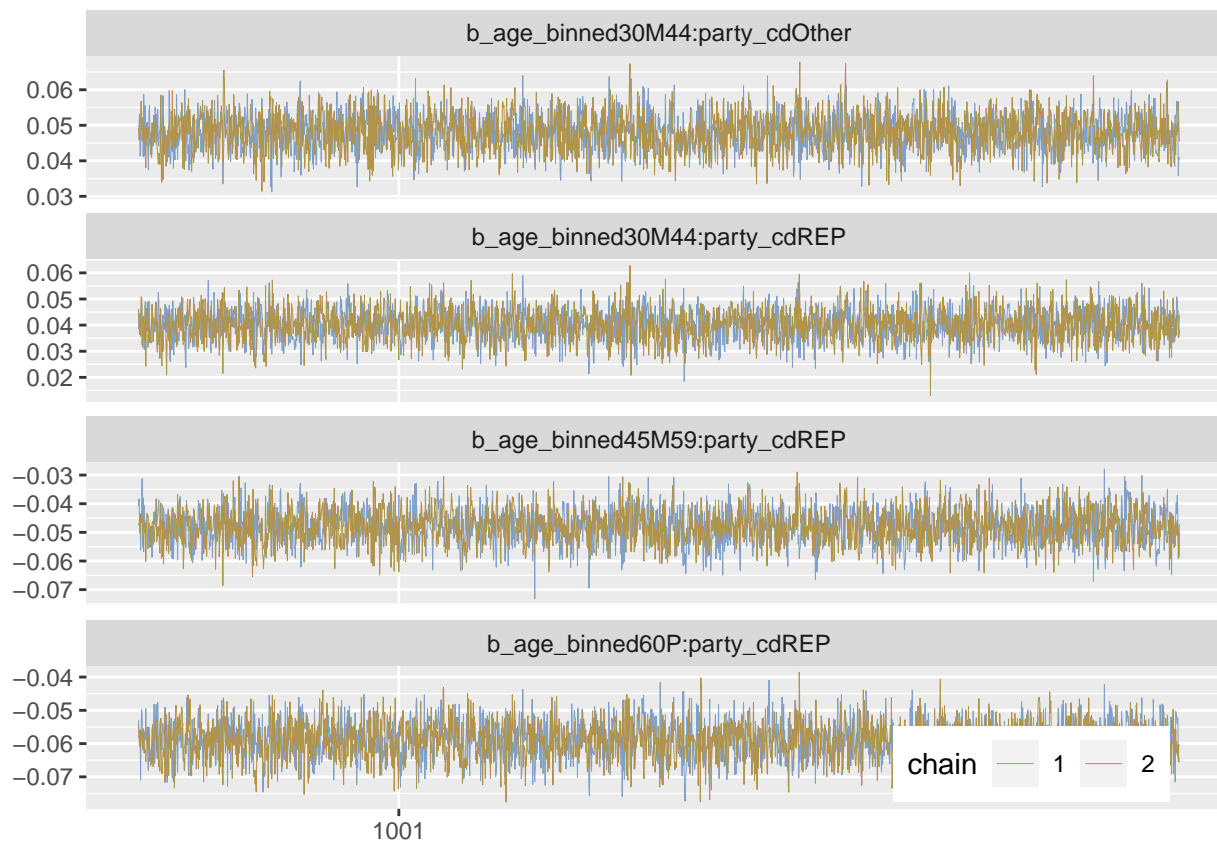
Sensitivity Analysis Model Output

Model with More Informative Priors

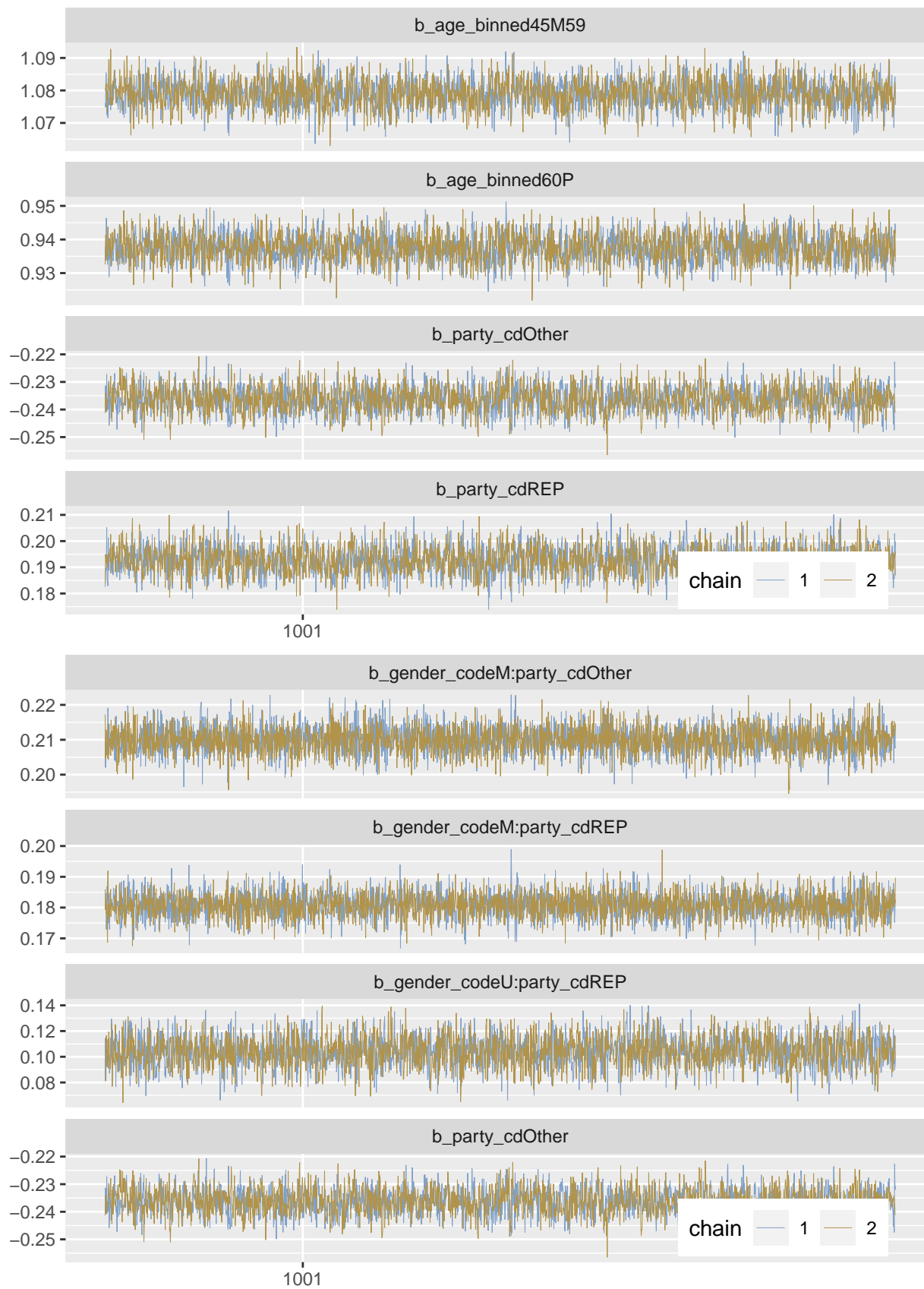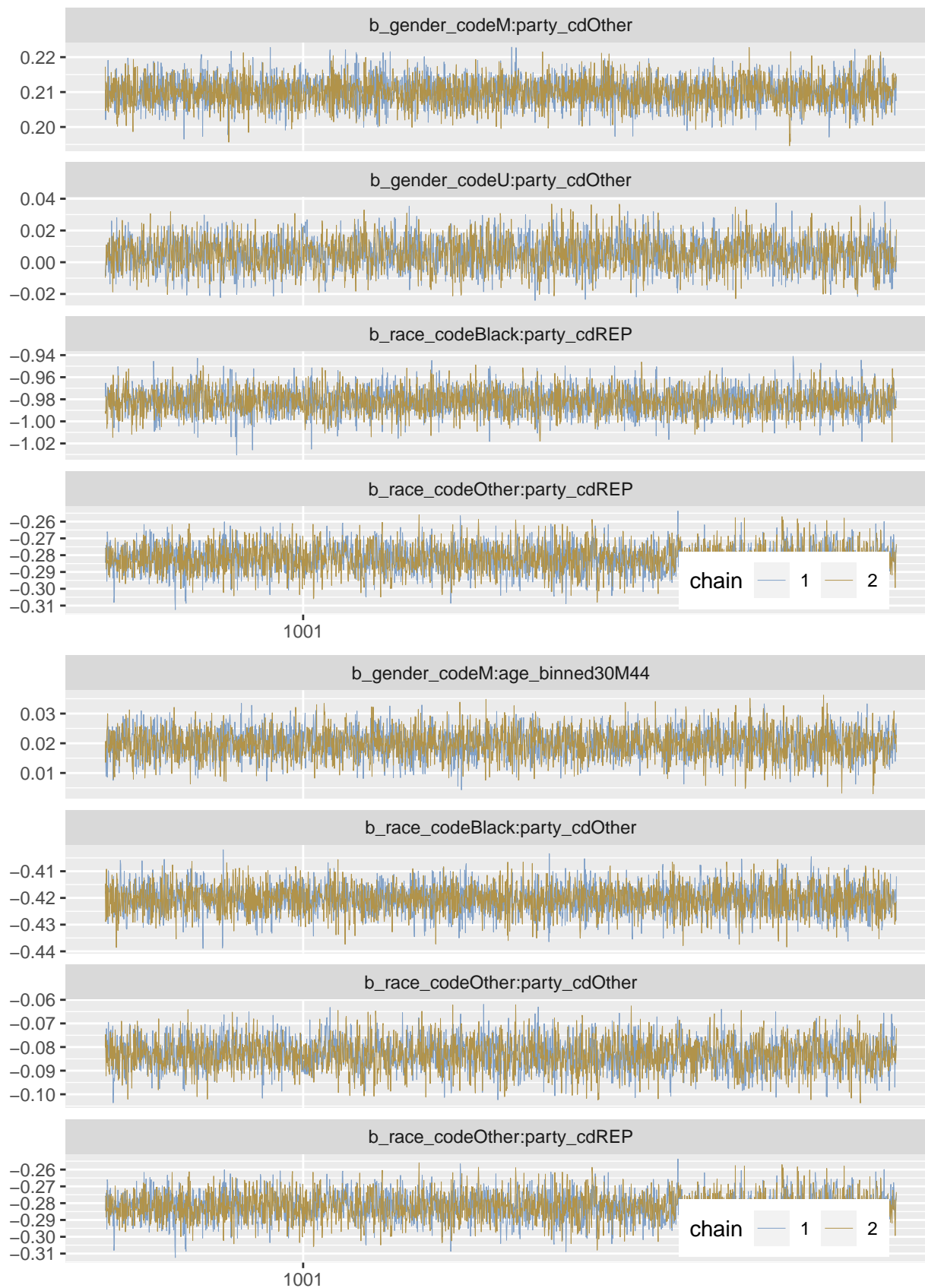|	Estimate	Std. Error	2.5% Quantile	97.5% Quantile
Intercept	0.00	0.00	-0.01	0.00
med_inc_binned46864M52798	0.10	0.00	0.09	0.10
med_inc_binned52798M64509	0.09	0.00	0.08	0.09
med_inc_binned>64509	0.36	0.00	0.35	0.36
gender_codeM	-0.43	0.00	-0.44	-0.42
gender_codeU	-0.08	0.01	-0.09	-0.06
race_codeBlack	0.03	0.00	0.03	0.04
race_codeOther	-0.41	0.00	-0.42	-0.40
age_binned30M44	0.62	0.00	0.61	0.63
age_binned45M59	1.08	0.00	1.07	1.09
age_binned60P	0.94	0.00	0.93	0.95
party_cdREP	0.19	0.01	0.18	0.20
party_cdOther	-0.24	0.00	-0.25	-0.23
gender_codeM:party_cdREP	0.18	0.00	0.17	0.19
gender_codeU:party_cdREP	0.10	0.01	0.08	0.13
gender_codeM:party_cdOther	0.21	0.00	0.20	0.22
gender_codeU:party_cdOther	0.01	0.01	-0.01	0.02
race_codeBlack:party_cdREP	-0.98	0.01	-1.00	-0.96
race_codeOther:party_cdREP	-0.28	0.01	-0.30	-0.27
race_codeBlack:party_cdOther	-0.42	0.01	-0.43	-0.41
race_codeOther:party_cdOther	-0.08	0.01	-0.10	-0.07
gender_codeM:age_binned30M44	0.02	0.00	0.01	0.03
gender_codeU:age_binned30M44	-0.41	0.01	-0.43	-0.40
gender_codeM:age_binned45M59	0.11	0.00	0.10	0.12
gender_codeU:age_binned45M59	-0.60	0.01	-0.62	-0.58
gender_codeM:age_binned60P	0.25	0.00	0.24	0.26
gender_codeU:age_binned60P	-0.47	0.01	-0.50	-0.45
age_binned30M44:party_cdREP	0.04	0.01	0.03	0.05
age_binned45M59:party_cdREP	-0.05	0.01	-0.06	-0.04
age_binned60P:party_cdREP	-0.06	0.01	-0.07	-0.05
age_binned30M44:party_cdOther	0.05	0.01	0.04	0.06
age_binned45M59:party_cdOther	0.01	0.01	0.00	0.02
age_binned60P:party_cdOther	0.27	0.01	0.25	0.28

Model with Random Effect for Congressional District

Table 4: Race Black Random Slope

	Estimate	Std. Error	2.5% Quantile	97.5% Quantile
Intercept	0.03	0.10	-0.15	0.23
med_inc_binned46864_52798	0.00	0.11	-0.24	0.22
med_inc_binned52798_64509	0.15	0.13	-0.11	0.42
med_inc_binned>64509	0.30	0.25	-0.20	0.82
gender_codeM	-0.42	0.02	-0.45	-0.39
gender_codeU	-0.09	0.05	-0.19	0.00
race_codeBlack	0.01	0.04	-0.06	0.08
race_codeOther	-0.40	0.03	-0.46	-0.34
age_binned30M44	0.64	0.02	0.60	0.67
age_binned45M59	1.10	0.03	1.03	1.16
age_binned60P	0.98	0.05	0.88	1.08
party_cdOther	-0.22	0.02	-0.25	-0.18
party_cdREP	0.21	0.04	0.13	0.29
gender_codeM:party_cdOther	0.20	0.00	0.19	0.21
gender_codeU:party_cdOther	0.00	0.01	-0.02	0.02
gender_codeM:party_cdREP	0.17	0.00	0.17	0.18
gender_codeU:party_cdREP	0.09	0.01	0.07	0.12
race_codeBlack:party_cdOther	-0.41	0.01	-0.42	-0.40
race_codeOther:party_cdOther	-0.10	0.01	-0.11	-0.09
race_codeBlack:party_cdREP	-0.96	0.01	-0.98	-0.94
race_codeOther:party_cdREP	-0.27	0.01	-0.28	-0.25
gender_codeM:age_binned30M44	0.02	0.00	0.01	0.03
gender_codeU:age_binned30M44	-0.41	0.01	-0.43	-0.40
gender_codeM:age_binned45M59	0.11	0.00	0.10	0.12
gender_codeU:age_binned45M59	-0.62	0.01	-0.64	-0.60
gender_codeM:age_binned60P	0.25	0.00	0.24	0.26
gender_codeU:age_binned60P	-0.52	0.01	-0.54	-0.49
age_binned30M44:party_cdOther	0.04	0.01	0.03	0.05
age_binned45M59:party_cdOther	0.00	0.01	-0.01	0.01
age_binned60P:party_cdOther	0.27	0.01	0.26	0.28
age_binned30M44:party_cdREP	0.02	0.01	0.01	0.04
age_binned45M59:party_cdREP	-0.07	0.01	-0.08	-0.05
age_binned60P:party_cdREP	-0.08	0.01	-0.09	-0.06

Table 5: Age 45-59 Random Slope

	Estimate	Est.Error	Q2.5	Q97.5
1	-0.09	0.04	-0.16	-0.01
2	-0.23	0.04	-0.30	-0.16
3	0.06	0.04	-0.01	0.14
4	0.02	0.04	-0.06	0.09
5	-0.02	0.04	-0.09	0.05
6	0.00	0.04	-0.07	0.07
7	0.03	0.04	-0.04	0.10
8	0.18	0.04	0.11	0.25
9	0.01	0.04	-0.06	0.09

	Estimate	Est.Error	Q2.5	Q97.5
10	0.13	0.04	0.05	0.20
11	-0.12	0.04	-0.19	-0.04
12	0.01	0.04	-0.06	0.08
13	0.01	0.04	-0.06	0.09

Table 6: Party Republican Random Slope

	Estimate	Est.Error	Q2.5	Q97.5
1	-0.07	0.03	-0.13	0.00
2	0.12	0.03	0.06	0.18
3	0.10	0.03	0.04	0.17
4	0.06	0.03	0.00	0.12
5	-0.07	0.03	-0.13	0.00
6	-0.06	0.03	-0.12	0.01
7	-0.12	0.03	-0.18	-0.05
8	0.10	0.03	0.04	0.16
9	0.08	0.03	0.02	0.15
10	0.03	0.03	-0.03	0.10
11	-0.21	0.03	-0.27	-0.14
12	0.03	0.03	-0.04	0.09
13	-0.02	0.03	-0.08	0.04

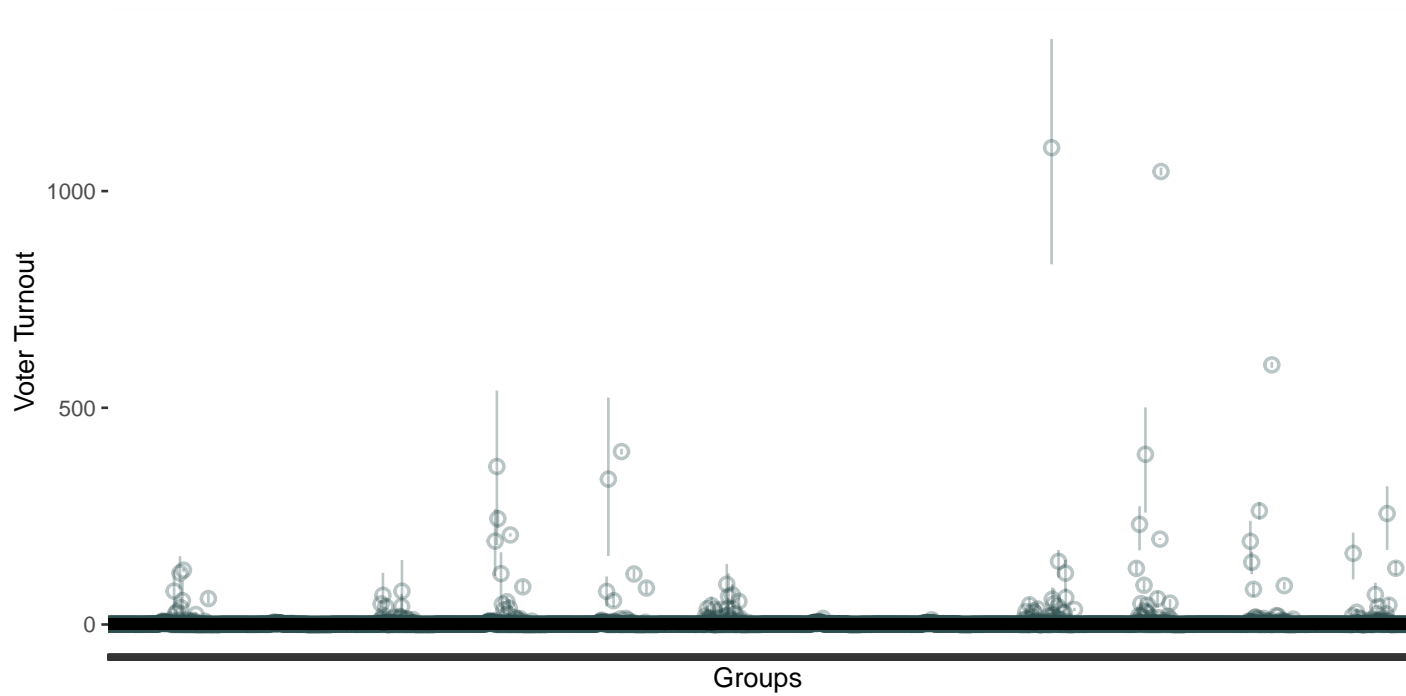|	Estimate	Est.Error	Q2.5	Q97.5
1	-0.04	0.04	-0.12	0.04
2	-0.18	0.04	-0.26	-0.10
3	0.03	0.04	-0.05	0.11
4	-0.17	0.04	-0.25	-0.09
5	0.09	0.04	0.01	0.17
6	-0.03	0.04	-0.11	0.05
7	0.05	0.04	-0.04	0.12
8	0.10	0.04	0.02	0.18
9	0.03	0.04	-0.05	0.11
10	0.17	0.04	0.09	0.25
11	-0.13	0.04	-0.21	-0.05
12	-0.12	0.04	-0.20	-0.04
13	0.19	0.04	0.11	0.27

Figure 5: Coverage Plot for Random Effects Model
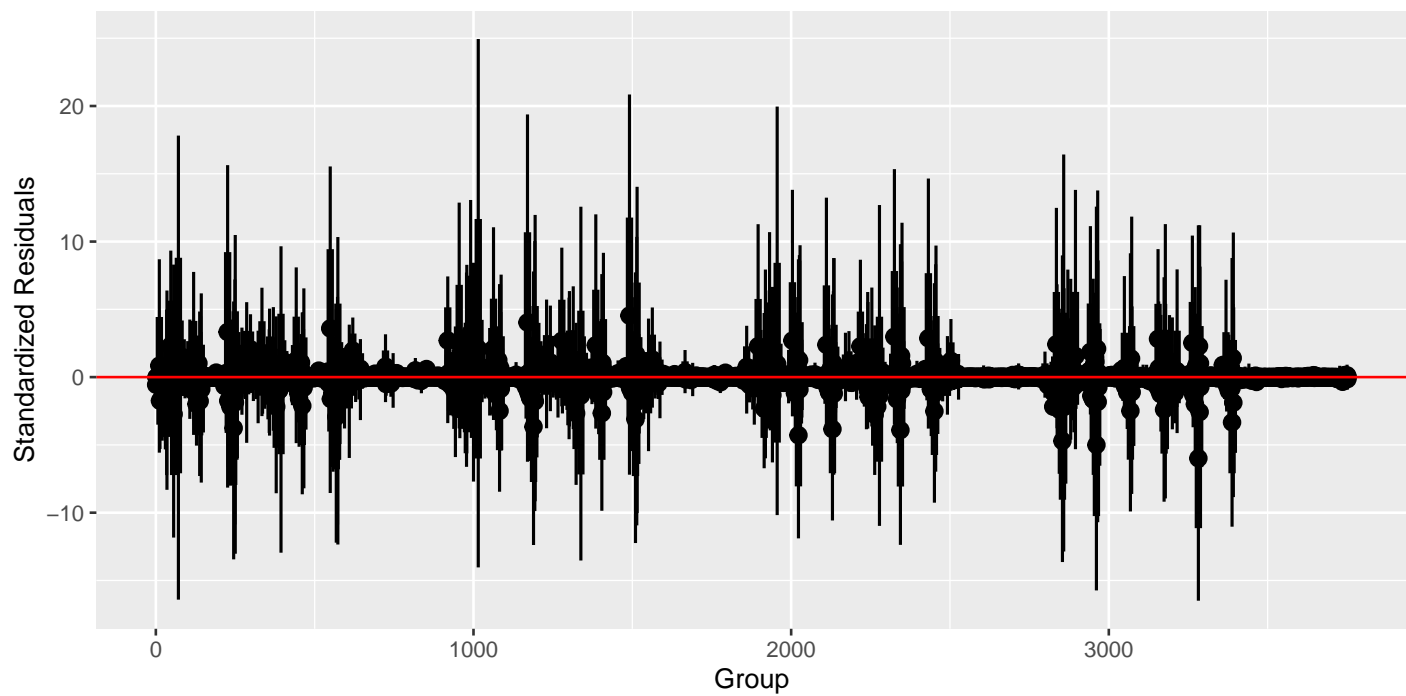
Figure 6: Random Effects Model Residual Plot

Frequentist Model

	Estimate	Std. Error	2.5 %	97.5 %
(Intercept)	0.00	0.00	-0.01	0.00
med_inc_binned> 64509	0.36	0.00	0.35	0.36
med_inc_binned46864 _ 52798	0.10	0.00	0.09	0.10
med_inc_binned52798 _ 64509	0.09	0.00	0.08	0.09
gender_codeM	-0.43	0.00	-0.44	-0.42
gender_codeU	-0.08	0.01	-0.09	-0.06
race_codeBlack	0.03	0.00	0.03	0.04
race_codeOther	-0.41	0.00	-0.42	-0.40
age_binned30-44	0.62	0.00	0.61	0.63
age_binned45-59	1.08	0.00	1.07	1.09
age_binned60+	0.94	0.00	0.93	0.95
party_cdREP	0.19	0.01	0.18	0.20
party_cdOther	-0.24	0.00	-0.25	-0.23
gender_codeM:party_cdREP	0.18	0.00	0.17	0.19
gender_codeU:party_cdREP	0.10	0.01	0.08	0.13
gender_codeM:party_cdOther	0.21	0.00	0.20	0.22
gender_codeU:party_cdOther	0.01	0.01	-0.01	0.03
race_codeBlack:party_cdREP	-0.98	0.01	-1.00	-0.96
race_codeOther:party_cdREP	-0.28	0.01	-0.30	-0.27
race_codeBlack:party_cdOther	-0.42	0.01	-0.43	-0.41
race_codeOther:party_cdOther	-0.08	0.01	-0.10	-0.07
gender_codeM:age_binned30-44	0.02	0.00	0.01	0.03
gender_codeU:age_binned30-44	-0.41	0.01	-0.43	-0.40
gender_codeM:age_binned45-59	0.11	0.00	0.10	0.12
gender_codeU:age_binned45-59	-0.60	0.01	-0.62	-0.58
gender_codeM:age_binned60+	0.25	0.00	0.24	0.26
gender_codeU:age_binned60+	-0.47	0.01	-0.50	-0.45
age_binned30-44:party_cdREP	0.04	0.01	0.03	0.05
age_binned45-59:party_cdREP	-0.05	0.01	-0.06	-0.04
age_binned60+:party_cdREP	-0.06	0.01	-0.07	-0.05
age_binned30-44:party_cdOther	0.05	0.01	0.04	0.06
age_binned45-59:party_cdOther	0.01	0.01	0.00	0.02
age_binned60+:party_cdOther	0.27	0.01	0.25	0.28