North Carolina Voter Turnout Analysis

Derek Tao

Introduction

Voter turnout data is one of the most valuable tools for measuring the health of a democratic society. By examining disparities in turnout among different demographic groups, we can obtain valuable insights into social equality, civic engagement, and representation within a society. This information is important for both citizens to assess the quality of their political system as well as politicians to assess the effectiveness of their campaigns. In this case study, I will examine data from the North Carolina State Board of Elections, which has provided voter registration and turnout data for North Carolina voters in the 2020 general elections. I will use a hierarchical model to examine differences in the voter turnouts of various demographic groups and party affiliations. I will also investigate if voter turnout differed depending on the county which a voter resided in.

Data Cleaning and EDA

There are two data files available. One contains the total number of registered voters, aggregated by county, precinct, voter tabulation district, party affiliation, race, ethnicity, sex, and age group. The other file contains the total number of people who actually voted, aggregated by the same features in addition to voting method. Prior to modeling, a few data cleaning steps are needed. The two data files must be merged in order to observe aggregated counts of registered voters and actual voters in the same data set. To do this, I first aggregated both data files by county, age, party affiliation, race, ethnicity, and sex. I then merged the data set containing total registered voters with the data set containing total actual voters using a left join. Finally, I took a random sample of 30 counties and subset the merged data to observations from only these 30 counties. These 30 counties include Pender, Henderson, Haywood, Onslow, Yancey, Polk, Brunswick, Union, Greene, Pasquotank, Ashe, Nash, Hertford, Sampson, Randolph, Robeson, Mcdowell, Davidson, Yadkin, Currituck, Bladen, Avery, Franklin, Wayne, Buncombe, Mecklenburg, Duplin, New Hanover, Camden, and Johnston.

When performing some basic EDA, I examined the ratio of actual voters to registered voters across different counties, demographic groups, and party affiliations. In short, it appears that there are some disparities in voter turnout for all of our features of interest (county, age, party, race, ethnicity, sex). We can also see that there are some potential interactions present between some of the features, particularly for sex/party and age/party. As a result of this, and in order to address all of our questions of interest, we should include these covariates and interactions in any final model we decide on. See figures 1-8 in the appendix for EDA plots.

Model Building

We can frame our election data as binomial count data, where we have a collection of binary trials (N), the number of successes from these trials (Y), and the vector of covariates associated with these trials (X). In this case, the trials are the total number of registered voters and the successes are how many of these registered voters actually voted in the election. Each observation in our data is a different collection of trials corresponding to a different grouping of county, demographic, and party affiliation covariates. Furthermore, we can model the probability that a registered voter within one of these groupings will actually vote using a logit() link function. More details on this will be displayed below in the final model. Finally, in order to establish a hierarchical structure in our model, we will implement a random intercept for each of the 30 counties in our data. This will allow us to address the question of whether the odds of voting differ across counties. All other covariates will remain as fixed effects.

We begin by testing a frequentist approach to our model, using a GLM of the binomial family. However, this model does not converge unless I remove the interaction terms, and doing so would prevent us from answering some of our questions of interest. Adding additional interaction terms to the model also does not result in convergence. Thus, we can move on to the Bayesian logistic model approach, which I have outlined below:

$$Y_{ij} \mid X_{ij} \sim \text{Binomial}(N_{ij}, \pi_{ij})$$

i: grouping of demographic/party affiliation

j: county

$$\begin{aligned}\text{logit}(\pi_{ij}) = \beta_0 &+ \sum_a \beta_{1a} I(\text{Age}_{ij} = a) + \sum_b \beta_{2b} I(\text{Party}_{ij} = b) + \sum_c \beta_{3c} I(\text{Race}_{ij} = c \\ &+ \sum_d \beta_{4d} I(\text{Ethnicity}_{ij} = d) + \sum_e \beta_{5e} I(\text{Sex}_{ij} = e) + \\ &\sum_{a,b} \beta_{6a,b} I(\text{Age}_{ij} = a, \text{Party}_{ij} = b) + \sum_{b,e} \beta_{7b,e} I(\text{Party}_{ij} = b, \text{Sex}_{ij} = e) + b_{0j}\end{aligned}$$

$b_{0j} \sim N(0, \sigma_0^2)$, random intercept for county j

Prior specification:

$$(\beta_0, \beta_{1a}, \beta_{2b}, \beta_{3c}, \beta_{4d}, \beta_{5e}, \beta_{a,b}, \beta_{7b,e}) \sim MVN(0, \sigma_1^2 I) \qquad \sigma_1^2 = 7$$

$$\sigma_0^2 \sim \text{Half-}t(df = 1, \mu = 0, \sigma = 10)$$

We can assume that individual voter registrations are independent, and that there is a common probability of voting for each individual grouping i of age, party, race, ethnicity, and sex in county j, which we denote by pi. We also include for each county a random intercept, which are independently normally distributed with mean 0 and constant variance. Additionally, we need to specify uninformative prior distributions for our fixed effects and between-county variance since we are not given any prior information on their behaviors in this study. For the fixed effects, we want the prior to center around 0 and have large variance to give significant coverage to a wide range of positive and negative values, so we choose multivariate normal. For the between-county variance, we reference lecture notes 2.6 and choose a half t prior with a large scale parameter and 1 degree of freedom.

The above model converges successfully. For the sake of reducing an already large computational load (several hours of runtime), we will not add additional interaction terms to the model. This model will answer all questions of interest and will be used as our final model.

Model Assessment and Validation

There are several methods that we can use to evaluate the quality of our model. We can start by performing convergence diagnostics. The traceplots for our model show that there is good mixing and stationarity in both of the MCMC chains for all of our parameters. Furthermore, our model summary gives Rhat values of 1.00 or 1.01 for all our covariates, indicating strong evidence for convergence. We can also use AUC-ROC to get an idea of how well our model does at correctly predicting whether an individual will vote. Our model yields an AUC of 0.6655, which is solid and far better than random guessing. Finally, we can also examine the behavior of our model residuals. Figure 9 shows a residuals vs. fitted plot, and we can see that there is some evidence of heteroscedascity in the residuals (ie. variance of residuals increases with increasing fitted values). However, this behavior is not too concerning given that there is a relatively small amount of points causing this pattern. For the vast majority of our observations, the residuals pattern is quite random.

Model Results and Interpretation

Figure 10 in the appendix shows the posterior parameter estimates of our model, as well as the corresponding standard errors and 95% confidence intervals. With this, we can answer our questions of interest:

1. How did different demographic subgroups vote in the 2020 general elections?

We can examine the predicted voter turnouts of different demographic subgroups while controlling for the other demographic predictors. Starting with age group, holding all other covariates constant, we see that an individual over 66 or between 41 and 65 will have the highest odds of voting, followed by 26-40 and 18-25. Specifically, individuals in these age groups will have a exp(0.87)=2.3869 factor increase in odds of voting compared to the baseline level age group of 18-25. For party affiliation, holding all other covariates constant, we see that Democrats have the highest odds of voting, followed by Republicans, Green Party (GRE), Unaffiliated, Libertarians, and finally Constitution Party (CST). Democrats will have a exp(0.81)=2.2479 factor increase in odds of voting compared to Constitution Party voters, and a exp(0.81)/exp(0.76)=1.0513 factor increase in odds of voting compared to Republican voters. For race, holding all other covariates constant, we see that Pacific Islanders have by far the highest odds of voting, followed by White, Undesignated, Asian, Multiracial, Black, Indian, and "Other" voters. Pacific Islanders will have a exp(1.05)=2.8577 factor increase in odds of voting compared to "Other" voters, and a exp(1.05)/exp(0.13)=2.5093 factor increase in odds of voting compared to White voters! For ethnicity, holding all other covariates constant, we see that Non-Hispanics have the highest odds of voting, followed by Undesignated and Hispanic voters. Non-Hispanics will have a exp(0.37)=1.4477 factor increase in odds of voting compared to Hispanic voters. Finally, for sex, holding all other covariates constant, we see that the Undesignated sex has the highest odds of voting, followed by males and then females. Undesignated sex will have a exp(0.24)=1.2712 factor increase in odds of voting compared to females, while males will have a exp(0.04)=1.0408 factor increase in odds of voting compared to females.

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

We can answer this question by examining the posterior treatment effects of each individual county on odds of voting (see Figure 11). From our 30 sampled counties, we can definitely conclude that there are noticeable differences in predicted odds of voting across different counties. The two counties that stand out the most are Onslow and Robeson; individuals from these two counties have by far the lowest odds of voting, holding all demographic and party affiliation characteristics constant. Pasquotank and Polk are two other counties with relatively low predicted voter turnout by our model. As for counties with high odds of voting, we see that Franklin, Yadkin, and Greene are the top three.

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

To answer this question, we want to look at the interaction term between sex and party. Interestingly, holding all other covariates constant, the only parties where the odds of voting are marginally higher for males than females are the Libertarian Party and Constitution Party, despite the overall odds of voting being higher for males than females, as seen before. The party with the highest odds of voting for females relative to males is the Democratic Party, followed by Unaffiliated, Green Party, and Republicans.

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

To answer this question, we want to look at the interaction term between age group and party. We can see that every party follows the general trend of increasing odds of voting with increasing age except for the Green Party, which has a higher odds of voting for the 26-40 age group than the 41-65 age group. We can also see that there are larger disparities in odds of voting between older age groups and younger age groups for the Democratic Party, Republican Party, and Unaffiliated, and smaller disparities in odds of voting between older age groups and younger age groups for the Green Party, Libertarians, and Constitution Party.

Potential Limitations

The main limitation in our analysis is the computational load during model runs. Because the Bayesian model takes many hours to run, under the current time constraints it is unfeasible to do model comparisons. Thus, we were unable to determine the significance of additional interaction terms outside of those pertinent to our questions of interest. Another limitation of our model is that we can only judge differences in odds of voting for a subset of all counties in North Carolina. Again, if we were to implement random intercepts for all of the counties in the data, the computational load would likely be too much. Another limitation with the data is that the age demographic is not continuous, but rather separated into four age groups. Continuous age data would likely give more insightful model results because we would be able to establish a more precise trend between age and odds of voting. With our current setup, we merely have arbitrary cutoffs for the age groups that may be hiding some underlying trends in the data.

Code/Figure Appendix

```
registered<-read.delim(file='voter_stats_20201103.txt')
```

```
## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec = dec, :
## embedded nul(s) found in input
```

```
voted<-read.delim(file='history_stats_20201103.txt')
```

```
#aggregation and merging
aggregated_voted <- aggregate(voted$total_voters,list(county_desc=voted$county_desc,
                                                       age=voted$age,
                                                       party_cd=voted$voted_party_cd,
                                                       race_code=voted$race_code,
                                                       ethnic_code=voted$ethnic_code,
                                                       sex_code=voted$sex_code),sum)
names(aggregated_voted)[names(aggregated_voted) == 'x'] <- 'total_voted'
aggregated_registered<-aggregate(registered$total_voters,list(county_desc=registered$county_desc,
                                                       age=registered$age,
                                                       party_cd=registered$party_cd,
                                                       race_code=registered$race_code,
                                                       ethnic_code=registered$ethnic_code,
                                                       sex_code=registered$sex_code),sum)
names(aggregated_registered)[names(aggregated_registered) == 'x'] <- 'total_registered'
merged<-aggregated_registered%>%left_join(aggregated_voted,by=c("county_desc",
                                          "party_cd","race_code","ethnic_code",
                                          "sex_code","age"))
```

```
#county sampling
set.seed(422)
all_counties<-unique(c(unique(registered$county_desc),unique(voted$county_desc)))
subset<-sample(all_counties,size=30)
sub_df<-merged[merged$county_desc %in% subset,]
```

```
#cleaning
sub_df[is.na(sub_df)] <- 0
sub_df$total_voted[which(sub_df$total_registered-sub_df$total_voted<0)]<-sub_df$total_registered[which(sub
```

```
#EDA
sub_df$turnout<-sub_df$total_voted/sub_df$total_registered
sub_df%>%ggplot(aes(x=county_desc,y=turnout))+geom_boxplot()+
  theme(axis.text.x=element_text(angle = 90, vjust = 0.5, hjust = 1))+
  labs(title="Voter turnouts by county",x='County',y="Turnout",caption="Figure 1")
```

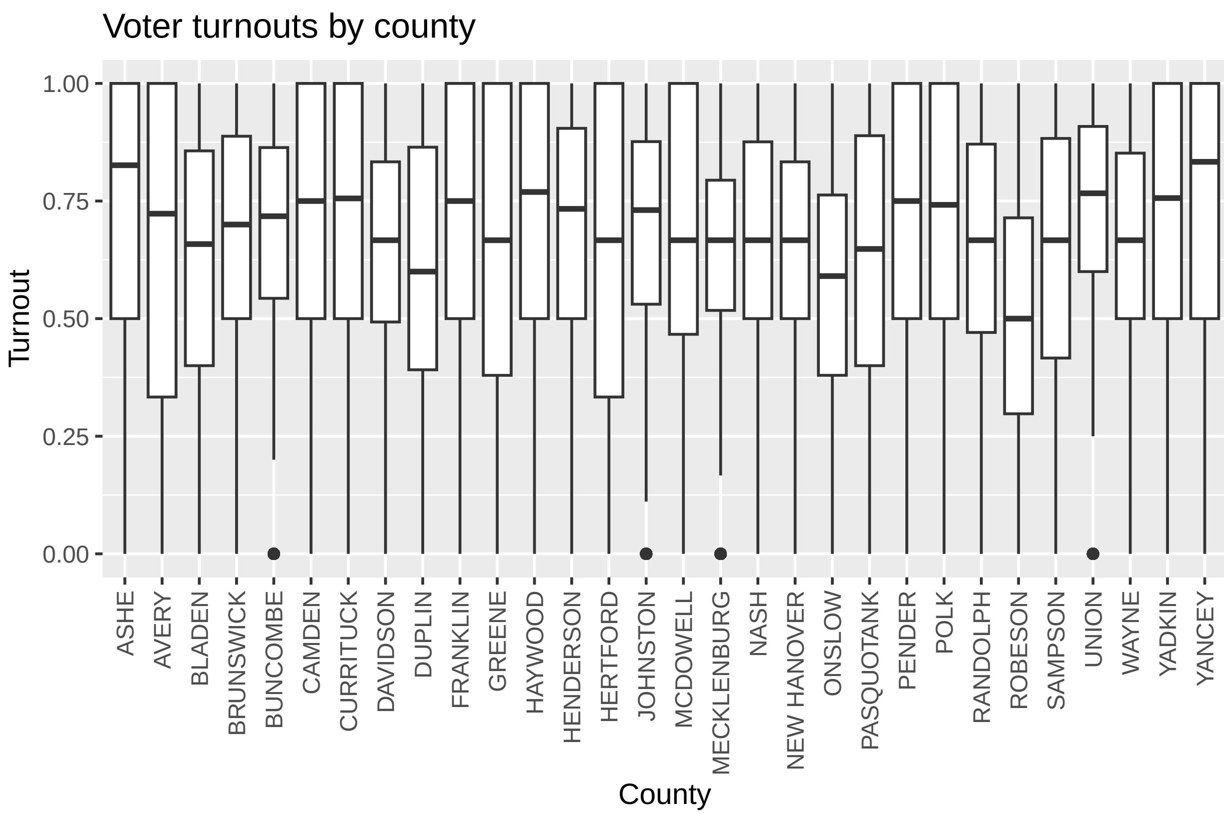


Figure 1

```
sub_df%>%ggplot(aes(x=age,y=turnout))+geom_boxplot()+
  labs(title="Voter turnouts by age group",x='Age group',y="Turnout",caption="Figure 2")
```

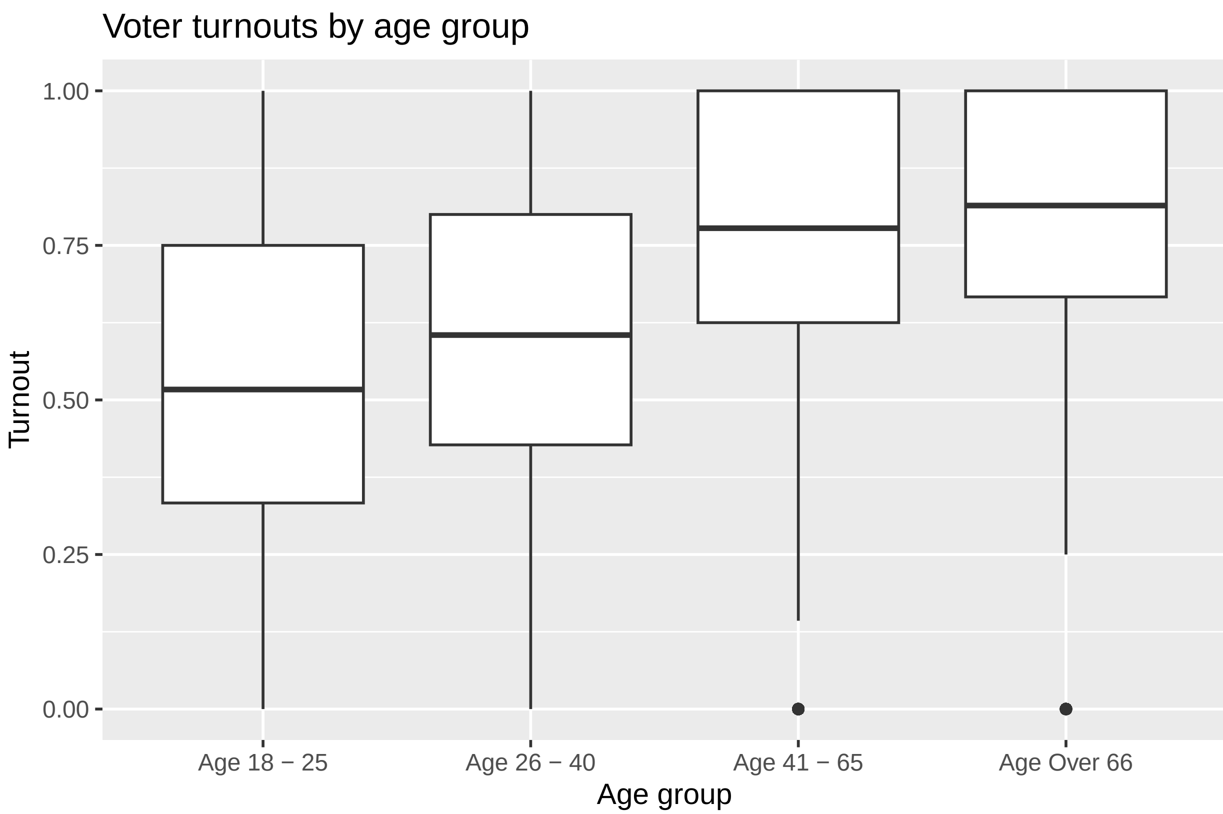
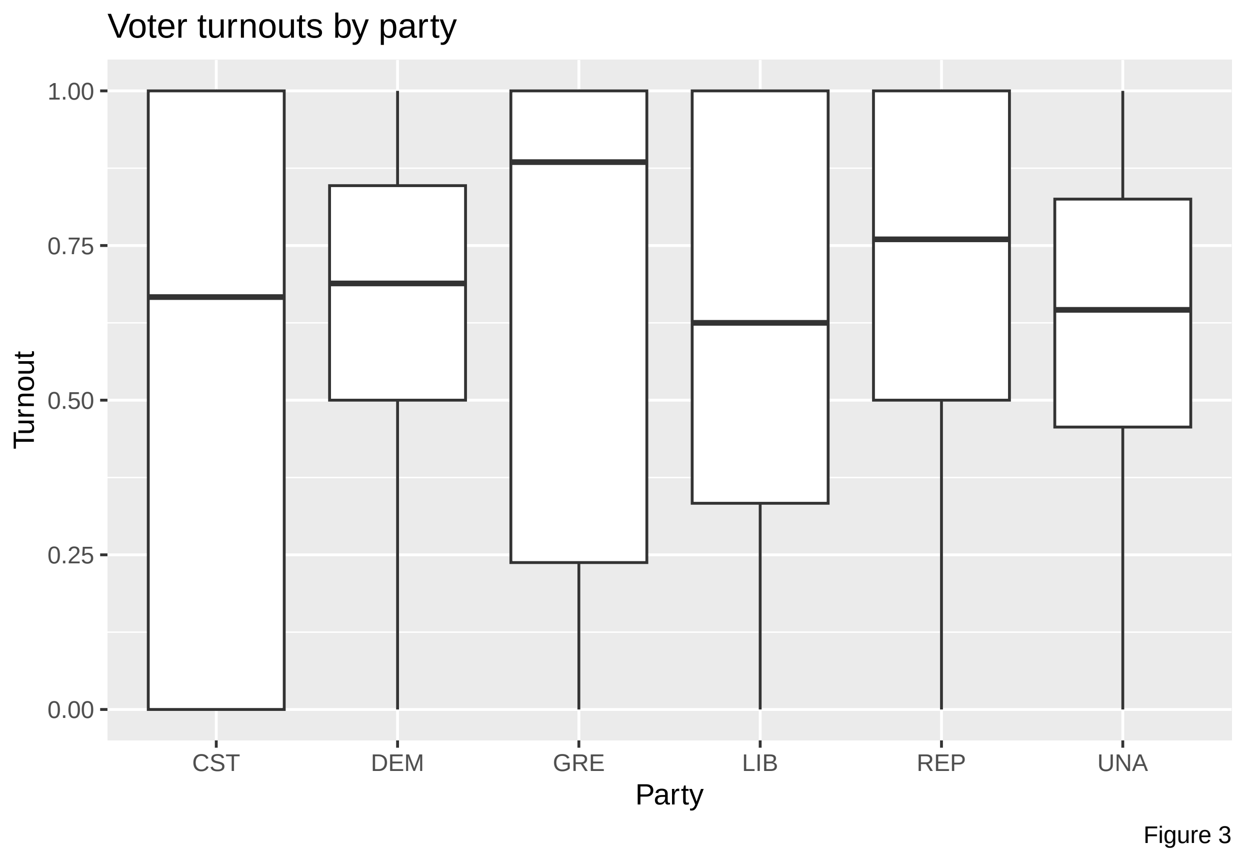


Figure 2

```
sub_df%>%ggplot(aes(x=party_cd,y=turnout))+geom_boxplot()+
  labs(title="Voter turnouts by party",x='Party',y="Turnout",caption="Figure 3")
```




```
sub_df%>%ggplot(aes(x=race_code,y=turnout))+geom_boxplot()+
  labs(title="Voter turnouts by race",x='Race',y="Turnout",caption="Figure 4")
```

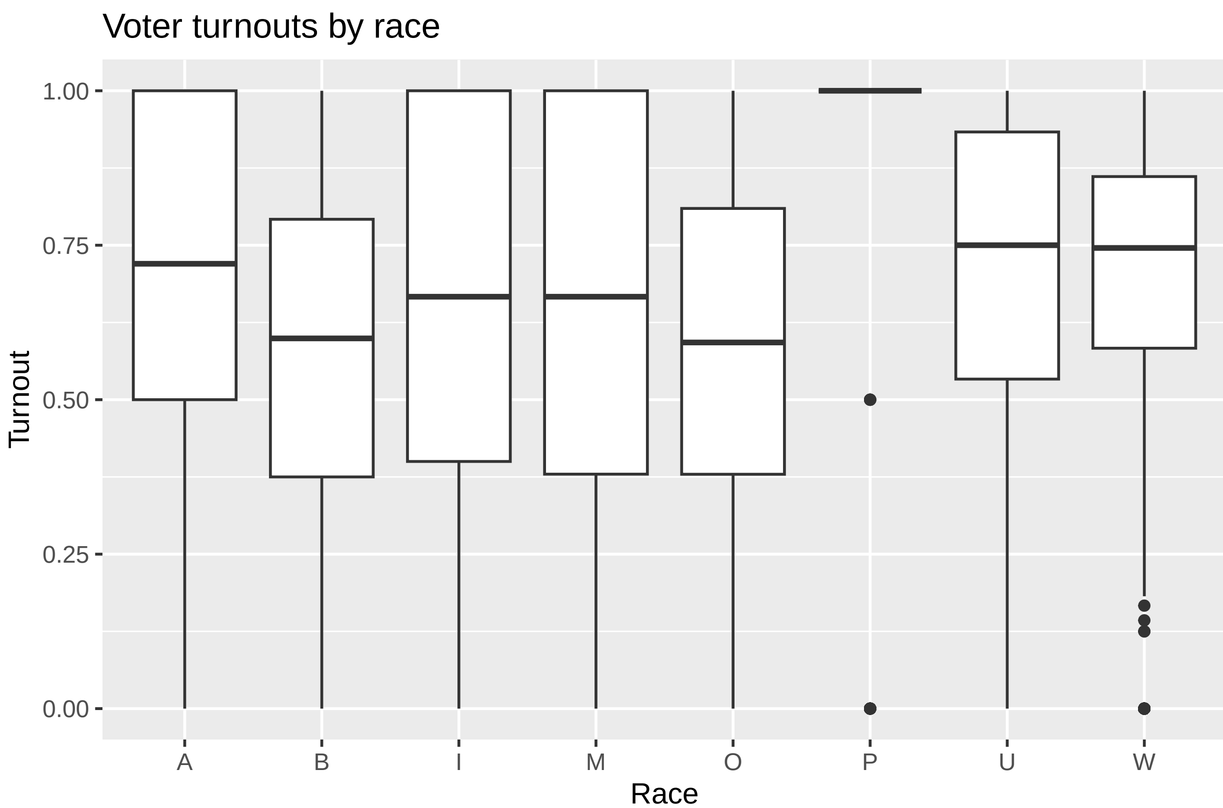
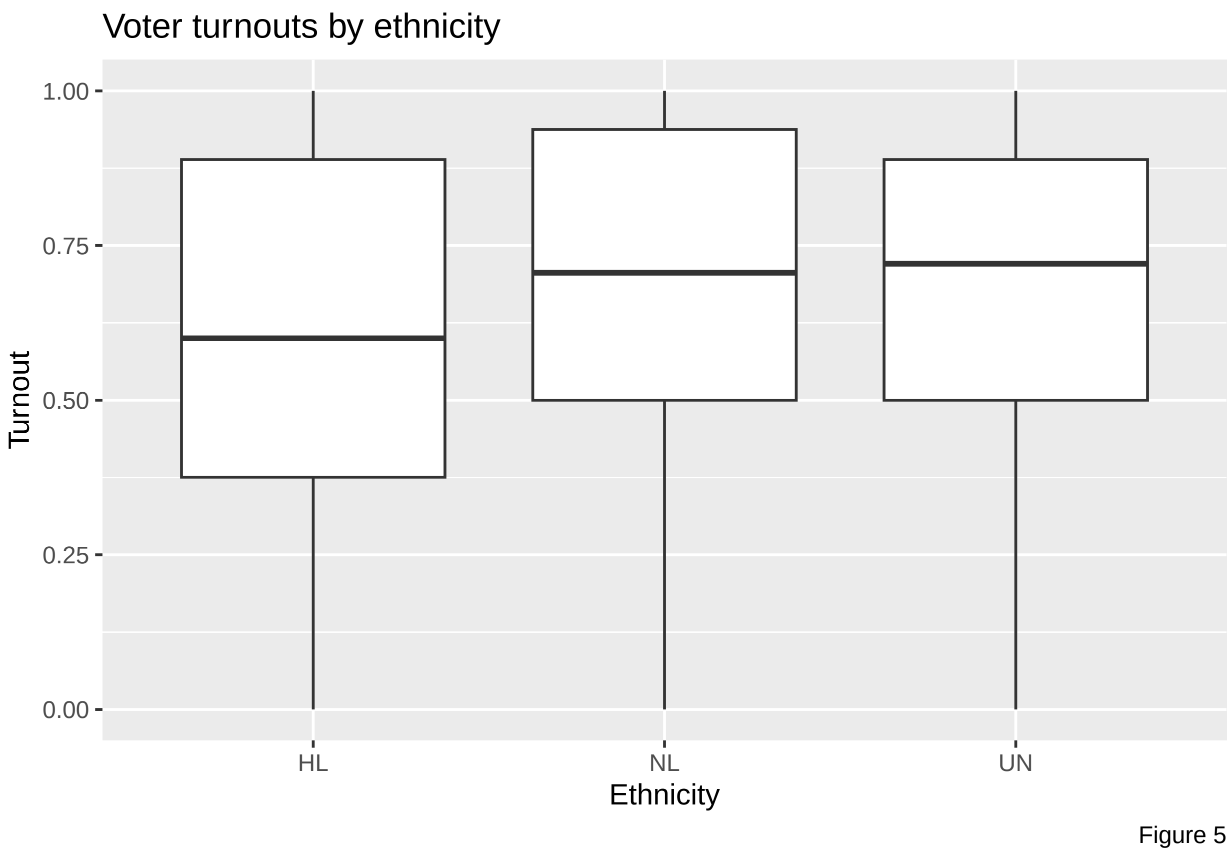


Figure 4

```
sub_df%>%ggplot(aes(x=ethnic_code,y=turnout))+geom_boxplot()+
  labs(title="Voter turnouts by ethnicity",x='Ethnicity',y="Turnout",caption="Figure 5")
```




```
sub_df%>%ggplot(aes(x=sex_code,y=turnout))+geom_boxplot()+
  labs(title="Voter turnouts by sex",x='Sex',y="Turnout",caption="Figure 6")
```

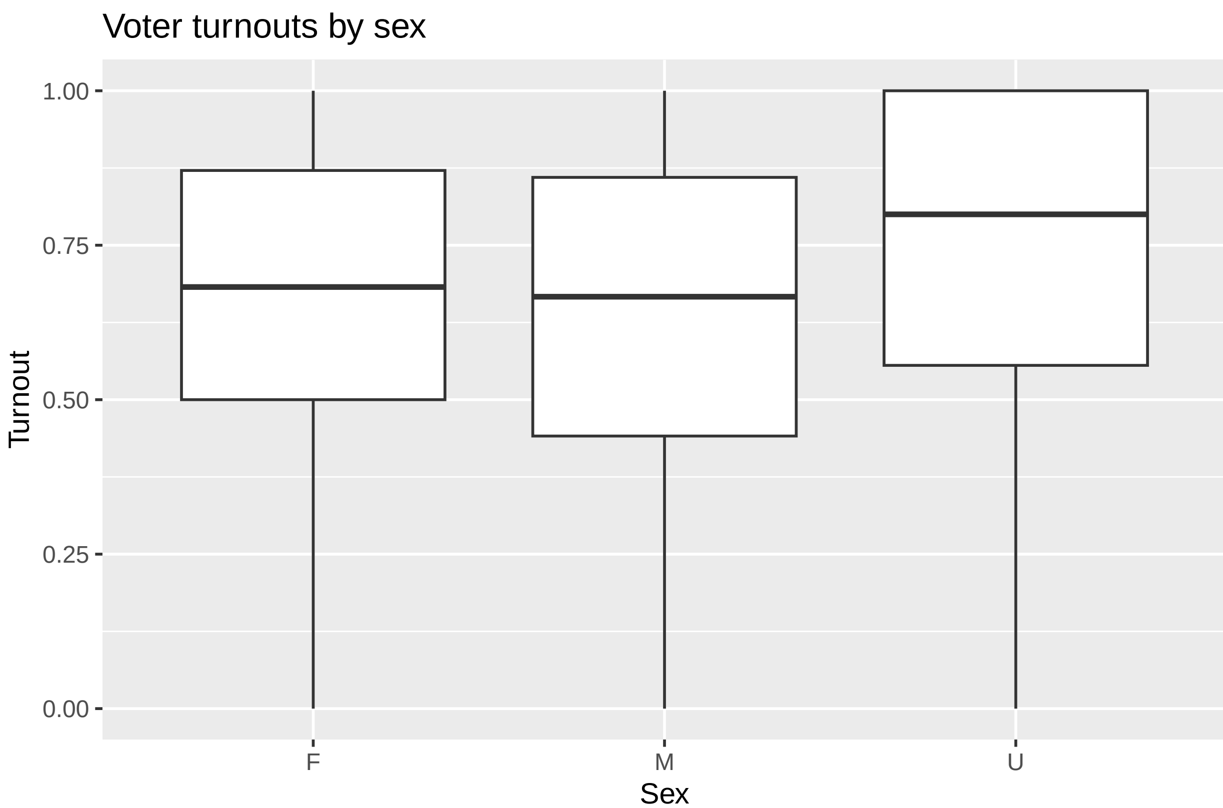


Figure 6

```
sub_df%>%ggplot(aes(x=sex_code,y=turnout))+geom_boxplot()+
  labs(title="Voter turnouts by sex for each party affiliation",
       x='Sex',y="Turnout",caption="Figure 7")+facet_grid(~party_cd)
```

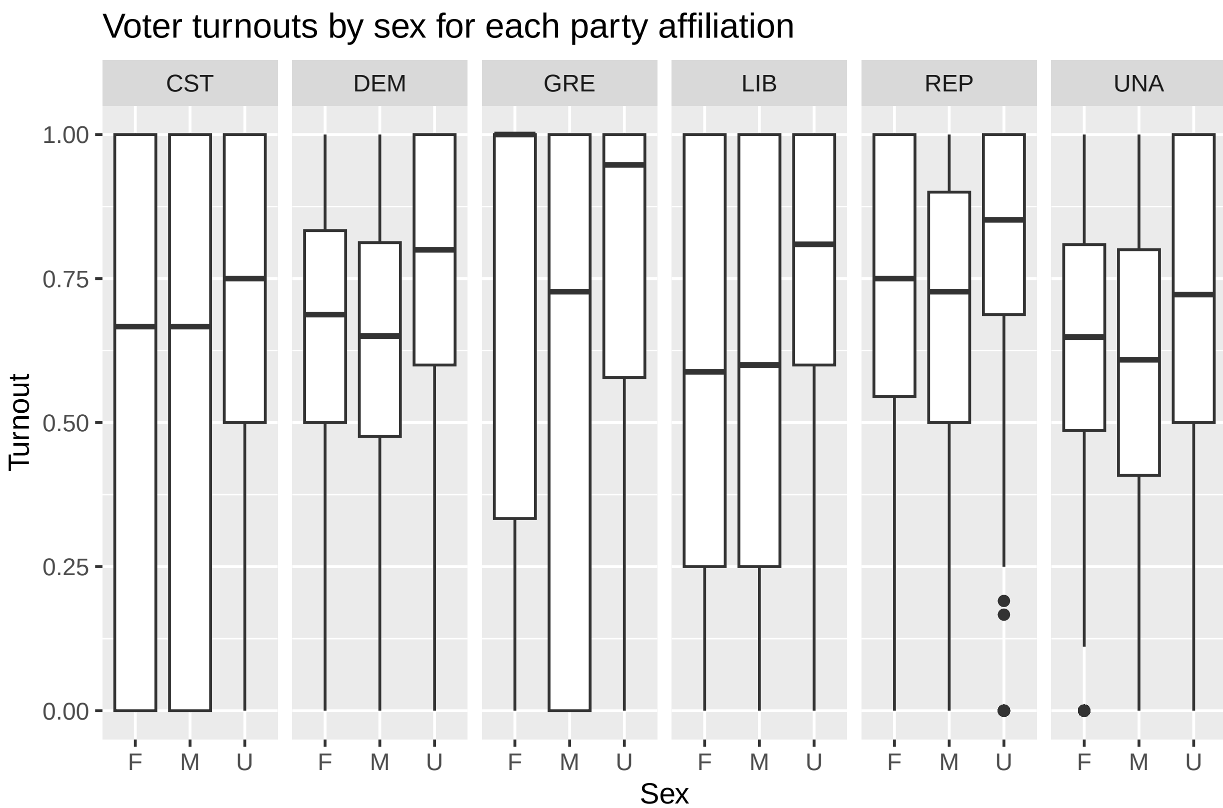


Figure 7

```
sub_df%>%ggplot(aes(x=age,y=turnout))+geom_boxplot()+
  theme(axis.text.x=element_text(angle = 90, vjust = 0.5, hjust = 1))+
  labs(title="Voter turnouts by age group for each party affiliation",
       x='Age Group',y="Turnout",caption="Figure 8")+facet_grid(~party_cd)
```

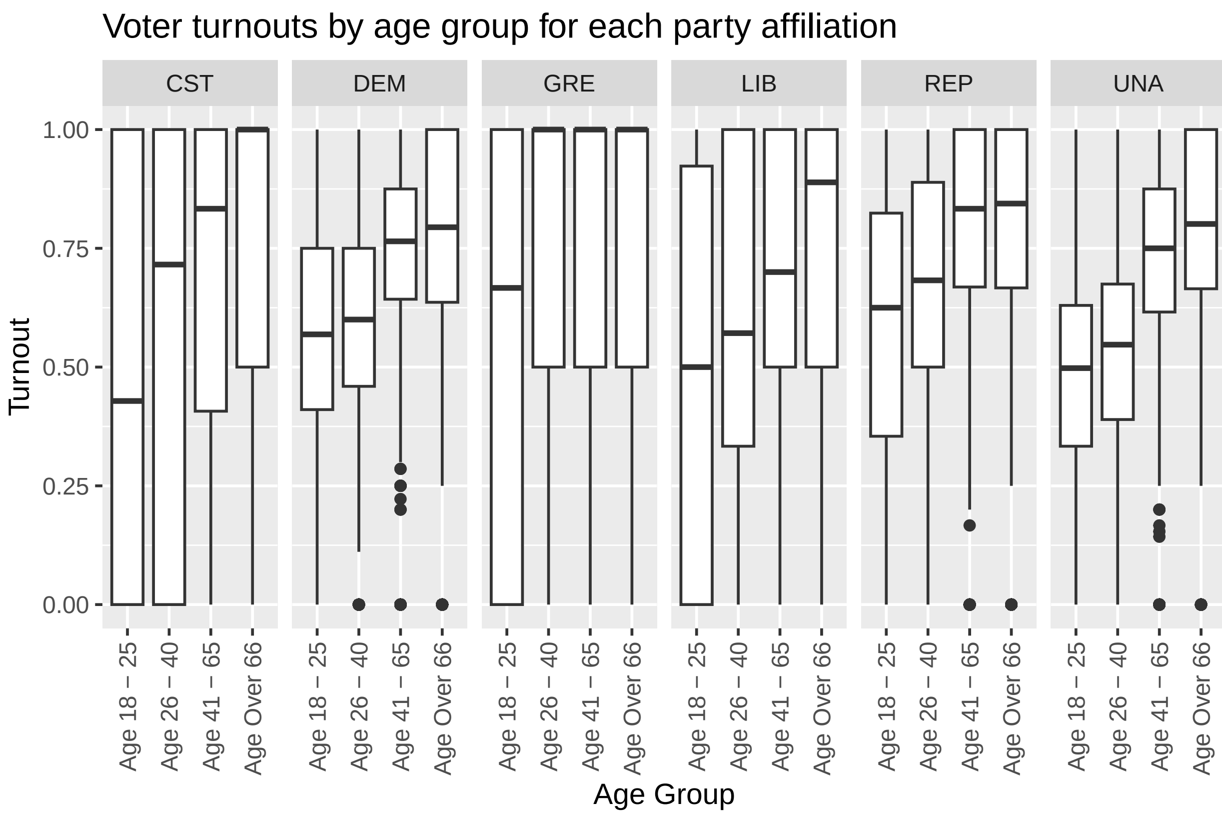


Figure 8

```
fit0 <- glmer(cbind(total_voted, total_registered-total_voted) ~age +party_cd +
              race_code+ethnic_code+sex_code+age*party_cd+sex_code*party_cd+(1|county_desc),
            data=sub_df, family=binomial())
summary(fit0)
```

```
fit1 <- glmer(cbind(total_voted, total_registered-total_voted) ~age +party_cd +
              race_code+ethnic_code+sex_code+(1|county_desc),
            data=sub_df, family=binomial())
summary(fit1)
```

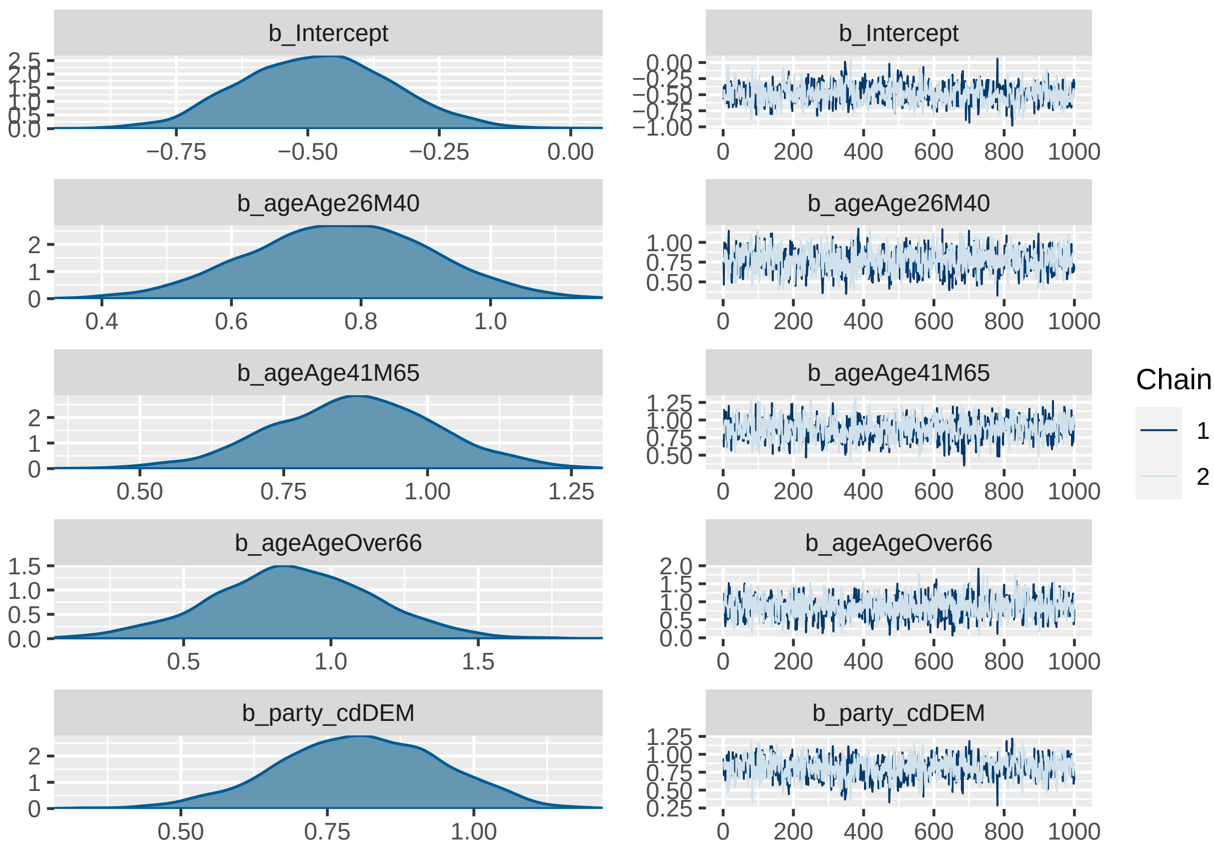
```
fit2 <- glmer(cbind(total_voted, total_registered-total_voted) ~age +party_cd +
              race_code+ethnic_code+sex_code+age*party_cd+sex_code*party_cd+
              race_code*sex_code+age_code*sex_code+(1|county_desc),
            data=sub_df, family=binomial())
summary(fit2)
```

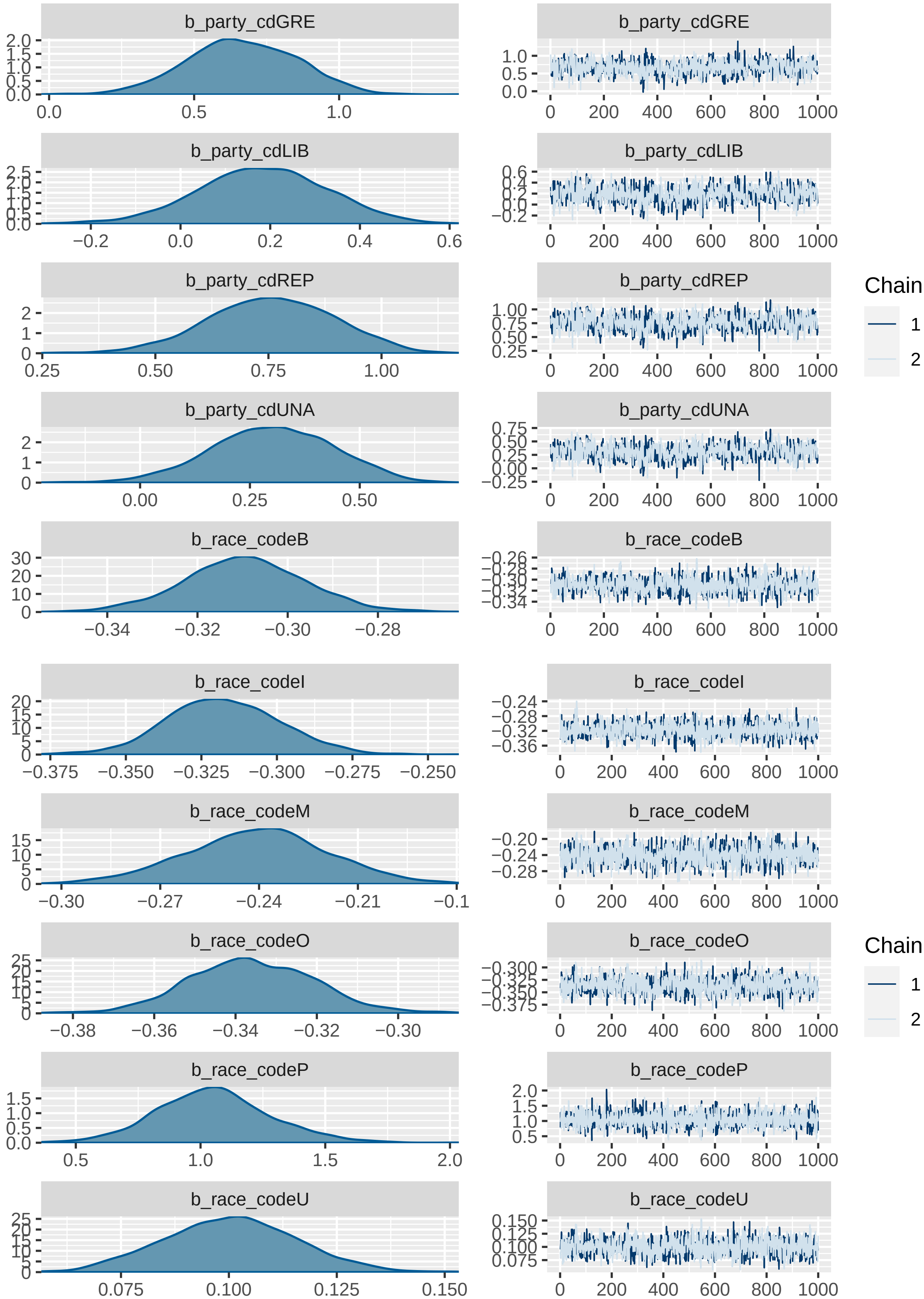
```
fit3<-brm(formula='total_voted|trials(total_registered)~age+party_cd+race_code+
ethnic_code+sex_code+age*party_cd+sex_code*party_cd+(1|county_desc)',
data=sub_df,family=binomial(link='logit'), prior=c(
  set_prior('normal(0,7)',class='Intercept'),
  set_prior('normal(0,7)',class='b'),
  set_prior('student_t(1,0,10)',class='sd')
),
iter=2000,seed=422,file='final_model',cores=getOption("mc.cores", 1), chains=2,
silent=0
)
```

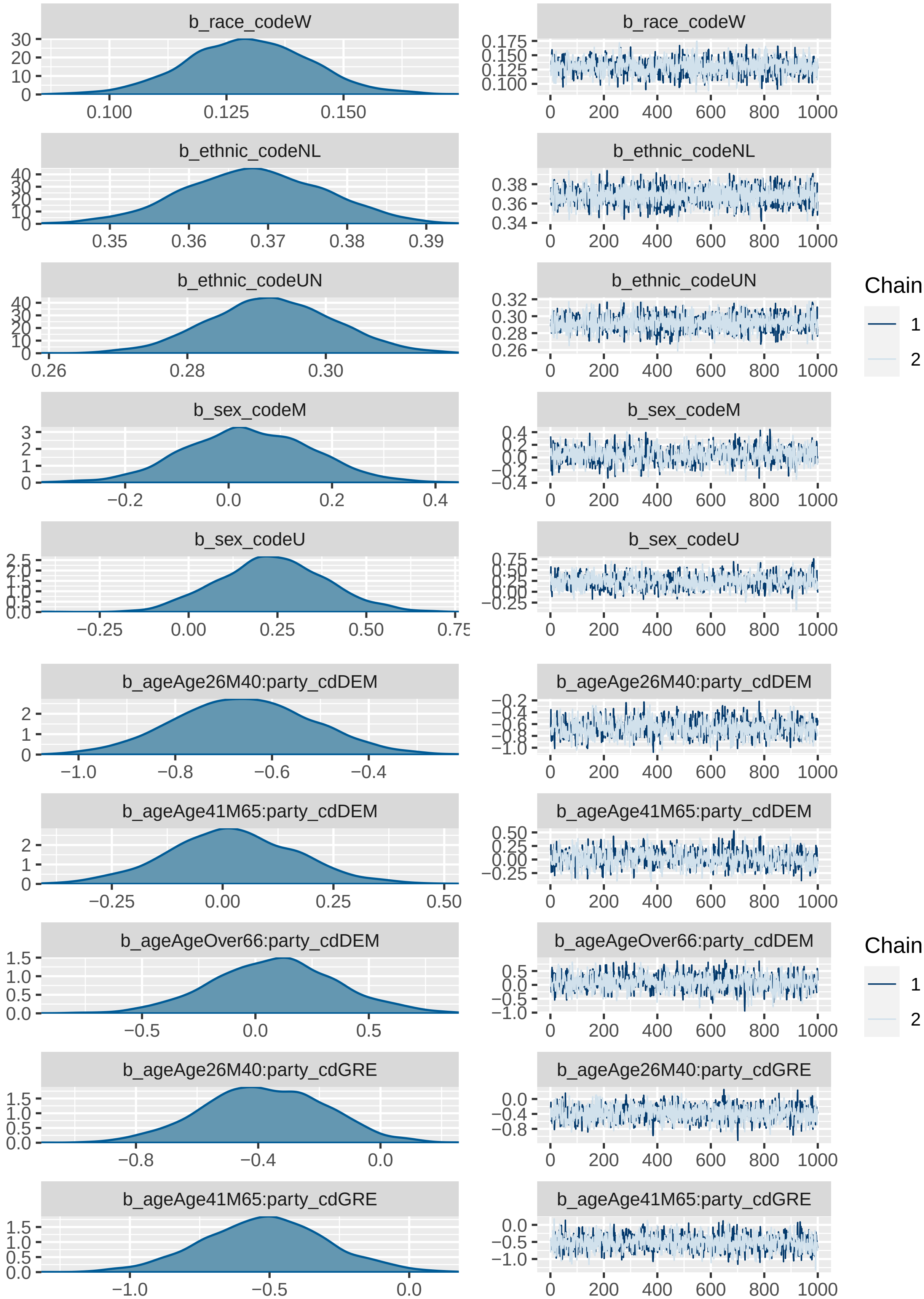
```r
final_model<-readRDS('final_model.rds')
```

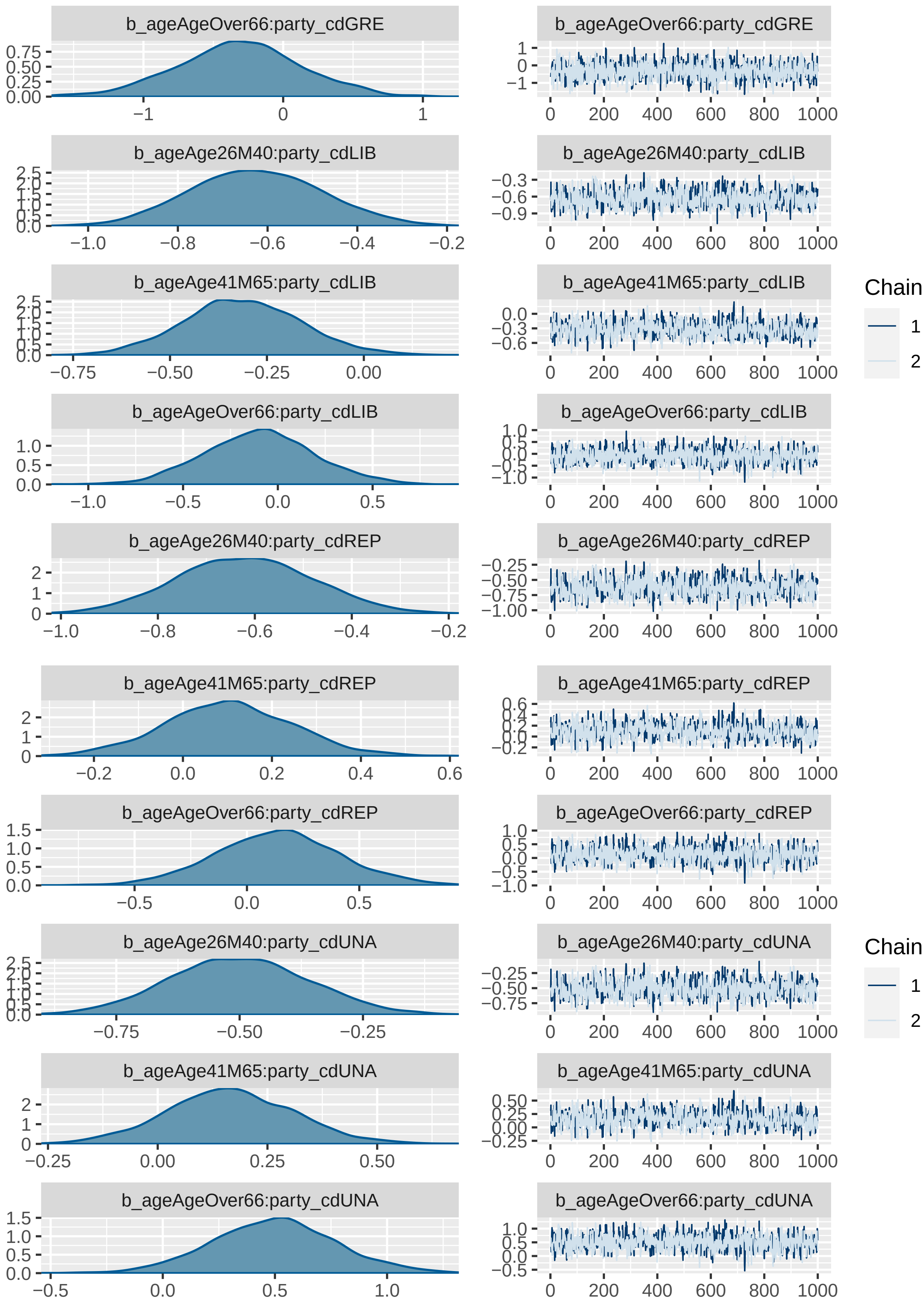
‘

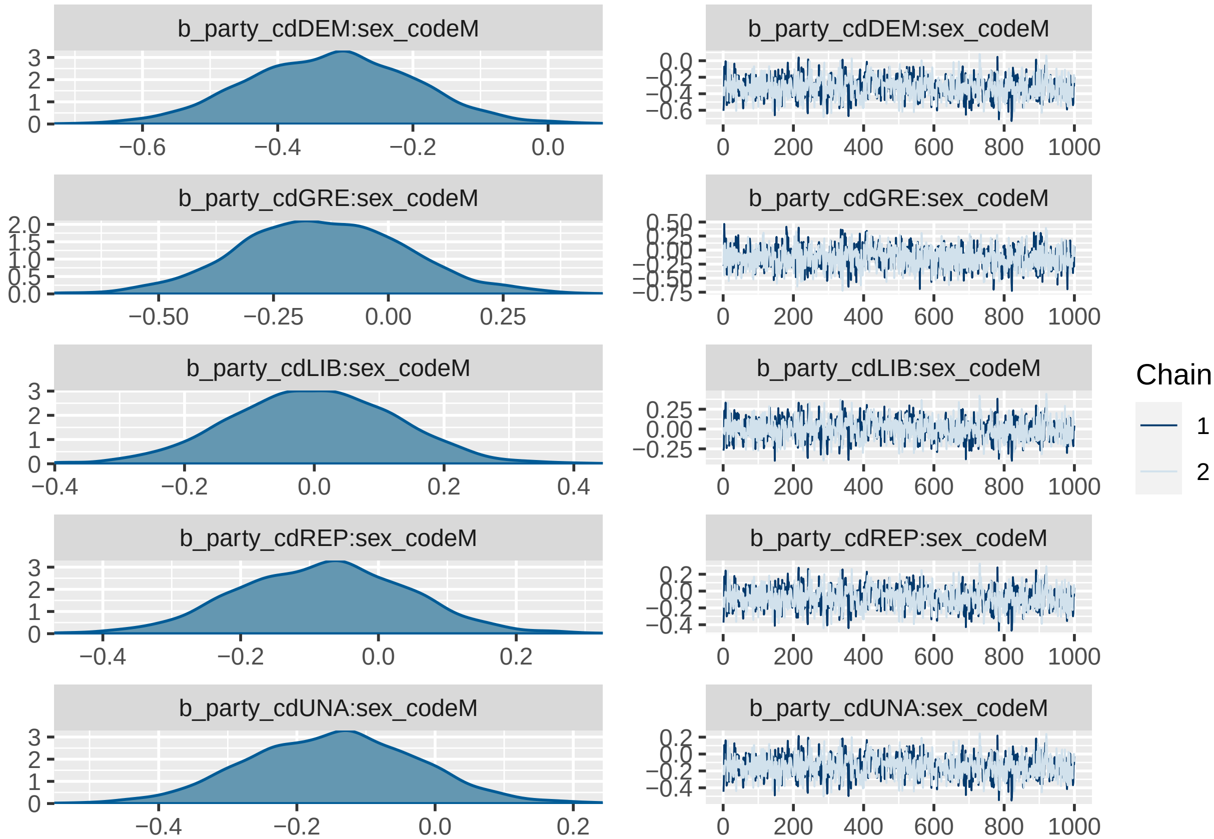
```r
#traceplots
plot(final_model,ask=FALSE)
```

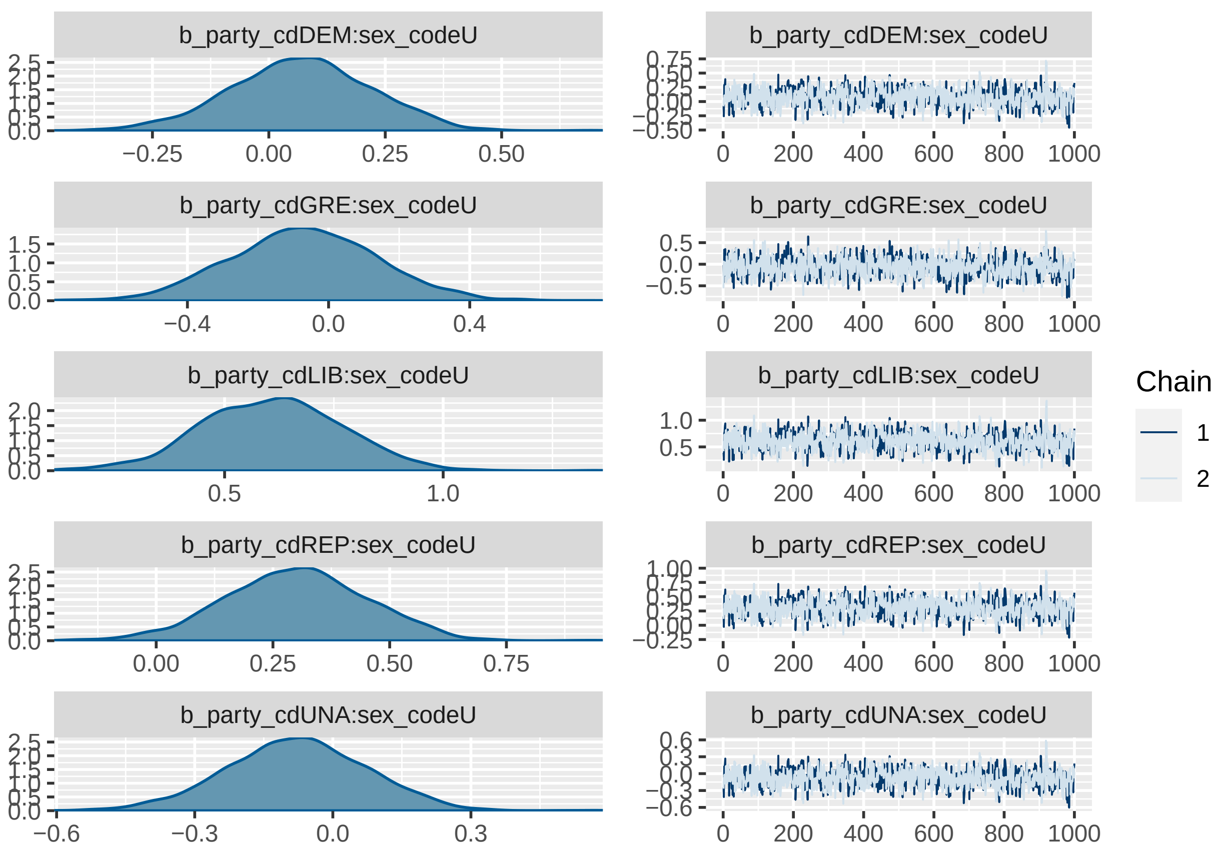


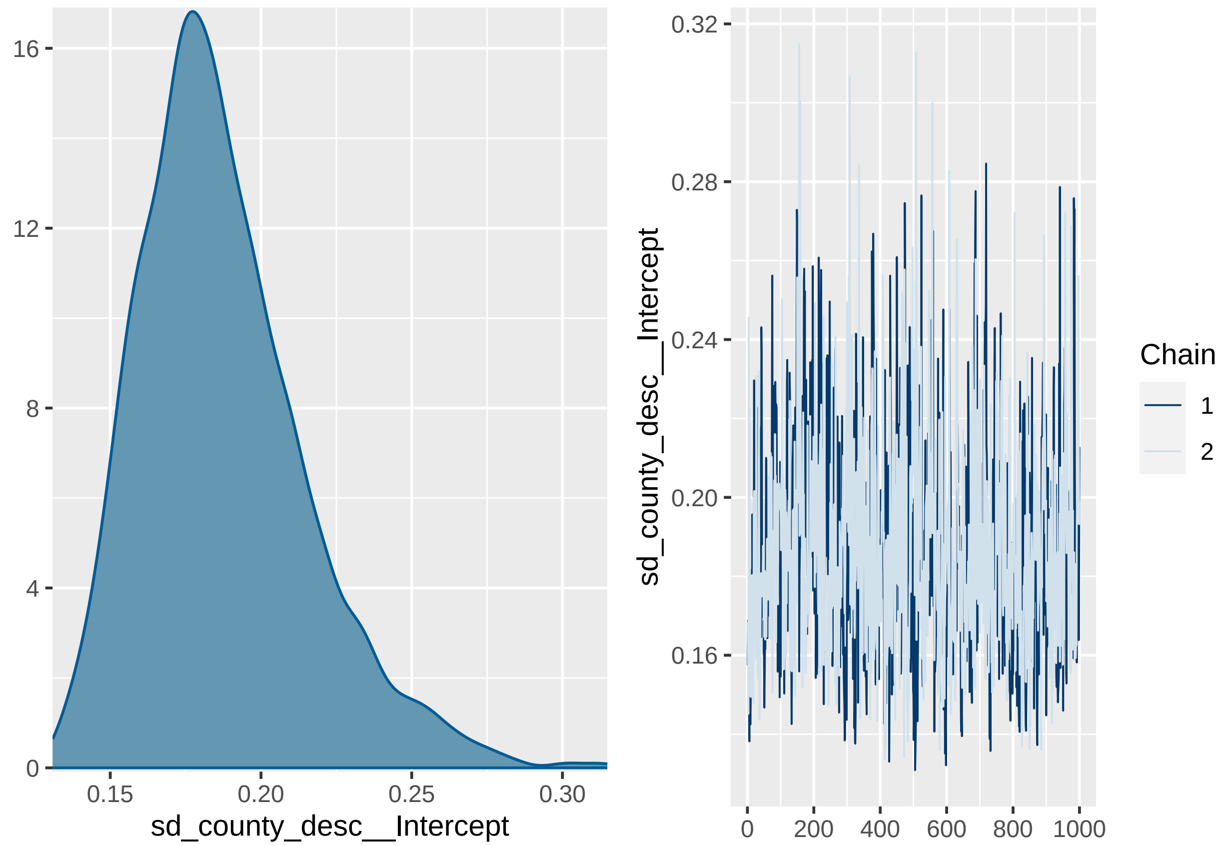

b_party_cdGRE
b_party_cdLIB
b_party_cdREP
b_party_cdUNA
b_race_codeB
b_race_codeI
b_race_codeM
b_race_codeO
b_race_codeP
b_race_codeU
Chain
1
2


b_race_codeW
b_ethnic_codeNL
b_ethnic_codeUN
b_sex_codeM
b_sex_codeU
b_ageAge26M40:party_cdDEM
b_ageAge41M65:party_cdDEM
b_ageAgeOver66:party_cdDEM
b_ageAge26M40:party_cdGRE
b_ageAge41M65:party_cdGRE
Chain
1
2


b_ageAgeOver66:party_cdGRE
b_ageAgeOver66:party_cdGRE
b_ageAge26M40:party_cdLIB
b_ageAge26M40:party_cdLIB
b_ageAge41M65:party_cdLIB
b_ageAge41M65:party_cdLIB
Chain
1
2
b_ageAgeOver66:party_cdLIB
b_ageAgeOver66:party_cdLIB
b_ageAge26M40:party_cdREP
b_ageAge26M40:party_cdREP
b_ageAge41M65:party_cdREP
b_ageAge41M65:party_cdREP
b_ageAgeOver66:party_cdREP
b_ageAgeOver66:party_cdREP
b_ageAge26M40:party_cdUNA
b_ageAge26M40:party_cdUNA
Chain
1
2
b_ageAge41M65:party_cdUNA
b_ageAge41M65:party_cdUNA
b_ageAgeOver66:party_cdUNA
b_ageAgeOver66:party_cdUNA


b_party_cdDEM:sex_codeM
b_party_cdGRE:sex_codeM
b_party_cdLIB:sex_codeM
b_party_cdREP:sex_codeM
b_party_cdUNA:sex_codeM
Chain
1
2


b_party_cdDEM:sex_codeU
b_party_cdGRE:sex_codeU
b_party_cdLIB:sex_codeU
b_party_cdREP:sex_codeU
b_party_cdUNA:sex_codeU
Chain
1
2




```r
#model summary
summary(final_model)
```

```
##  Family: binomial
##   Links: mu = logit
## Formula: total_voted | trials(total_registered) ~ age + party_cd + race_code + ethnic_code + sex_code +
##    Data: sub_df (Number of observations: 16636)
##   Draws: 2 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup draws = 2000
##
## Group-Level Effects:
## ~county_desc (Number of levels: 30)
##               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sd(Intercept)     0.19      0.03     0.14     0.25 1.00      478      756
##
## Population-Level Effects:
##                         Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
## Intercept                  -0.48      0.14    -0.75    -0.20 1.01      644
## ageAge26M40                 0.77      0.14     0.50     1.03 1.00      762
## ageAge41M65                 0.87      0.14     0.58     1.15 1.00      953
## ageAgeOver66                0.87      0.27     0.32     1.41 1.00     1103
## party_cdDEM                 0.81      0.14     0.52     1.06 1.01      643
## party_cdGRE                 0.67      0.19     0.30     1.02 1.00      804
## party_cdLIB                 0.18      0.14    -0.11     0.46 1.01      718
## party_cdREP                 0.76      0.14     0.48     1.02 1.01      638
## party_cdUNA                 0.30      0.14     0.02     0.56 1.01      639
## race_codeB                 -0.31      0.01    -0.34    -0.28 1.00     1437
## race_codeI                 -0.32      0.02    -0.35    -0.28 1.00     1681
```

```
## race_codeM                 -0.24      0.02    -0.28    -0.20 1.00     1796
## race_codeO                 -0.34      0.02    -0.37    -0.30 1.00     1530
## race_codeP                  1.05      0.23     0.61     1.53 1.00     2705
## race_codeU                  0.10      0.02     0.07     0.13 1.00     1565
## race_codeW                  0.13      0.01     0.10     0.16 1.00     1372
## ethnic_codeNL               0.37      0.01     0.35     0.39 1.00     2336
## ethnic_codeUN               0.29      0.01     0.27     0.31 1.00     2270
## sex_codeM                   0.04      0.12    -0.20     0.28 1.00      818
## sex_codeU                   0.24      0.15    -0.05     0.54 1.00      881
## ageAge26M40:party_cdDEM    -0.66      0.14    -0.93    -0.39 1.00      770
## ageAge41M65:party_cdDEM     0.02      0.14    -0.26     0.32 1.00      955
## ageAgeOver66:party_cdDEM    0.09      0.27    -0.45     0.64 1.00     1107
## ageAge26M40:party_cdGRE    -0.38      0.20    -0.77    -0.00 1.00     1048
## ageAge41M65:party_cdGRE    -0.52      0.22    -0.95    -0.07 1.00     1284
## ageAgeOver66:party_cdGRE   -0.31      0.45    -1.20     0.55 1.00     1378
## ageAge26M40:party_cdLIB    -0.62      0.15    -0.90    -0.34 1.00      794
## ageAge41M65:party_cdLIB    -0.31      0.15    -0.61    -0.01 1.00      984
## ageAgeOver66:party_cdLIB   -0.10      0.29    -0.65     0.50 1.00     1132
## ageAge26M40:party_cdREP    -0.62      0.14    -0.89    -0.34 1.00      752
## ageAge41M65:party_cdREP     0.10      0.14    -0.18     0.39 1.00      953
## ageAgeOver66:party_cdREP    0.14      0.28    -0.40     0.69 1.00     1107
## ageAge26M40:party_cdUNA    -0.50      0.14    -0.77    -0.23 1.00      759
## ageAge41M65:party_cdUNA     0.17      0.14    -0.11     0.46 1.00      949
## ageAgeOver66:party_cdUNA    0.50      0.27    -0.04     1.05 1.00     1112
## party_cdDEM:sex_codeM      -0.32      0.12    -0.56    -0.08 1.00      818
## party_cdGRE:sex_codeM      -0.14      0.18    -0.50     0.23 1.00     1178
## party_cdLIB:sex_codeM      -0.00      0.13    -0.25     0.24 1.00      833
## party_cdREP:sex_codeM      -0.08      0.12    -0.32     0.16 1.00      808
## party_cdUNA:sex_codeM      -0.15      0.12    -0.38     0.09 1.00      814
## party_cdDEM:sex_codeU       0.07      0.15    -0.24     0.35 1.00      883
## party_cdGRE:sex_codeU      -0.07      0.21    -0.47     0.35 1.00     1087
## party_cdLIB:sex_codeU       0.61      0.16     0.28     0.92 1.00      892
## party_cdREP:sex_codeU       0.30      0.15    -0.01     0.59 1.00      866
## party_cdUNA:sex_codeU      -0.08      0.15    -0.38     0.21 1.00      878
##                          Tail_ESS
## Intercept                     976
## ageAge26M40                  1068
## ageAge41M65                  1061
## ageAgeOver66                 1246
## party_cdDEM                   973
## party_cdGRE                  1236
## party_cdLIB                   924
## party_cdREP                   940
## party_cdUNA                   964
## race_codeB                   1264
## race_codeI                   1472
## race_codeM                   1365
## race_codeO                   1333
## race_codeP                   1351
## race_codeU                   1147
## race_codeW                   1122
## ethnic_codeNL                1295
## ethnic_codeUN                1504
## sex_codeM                    1111
## sex_codeU                    1255
## ageAge26M40:party_cdDEM      1132
```

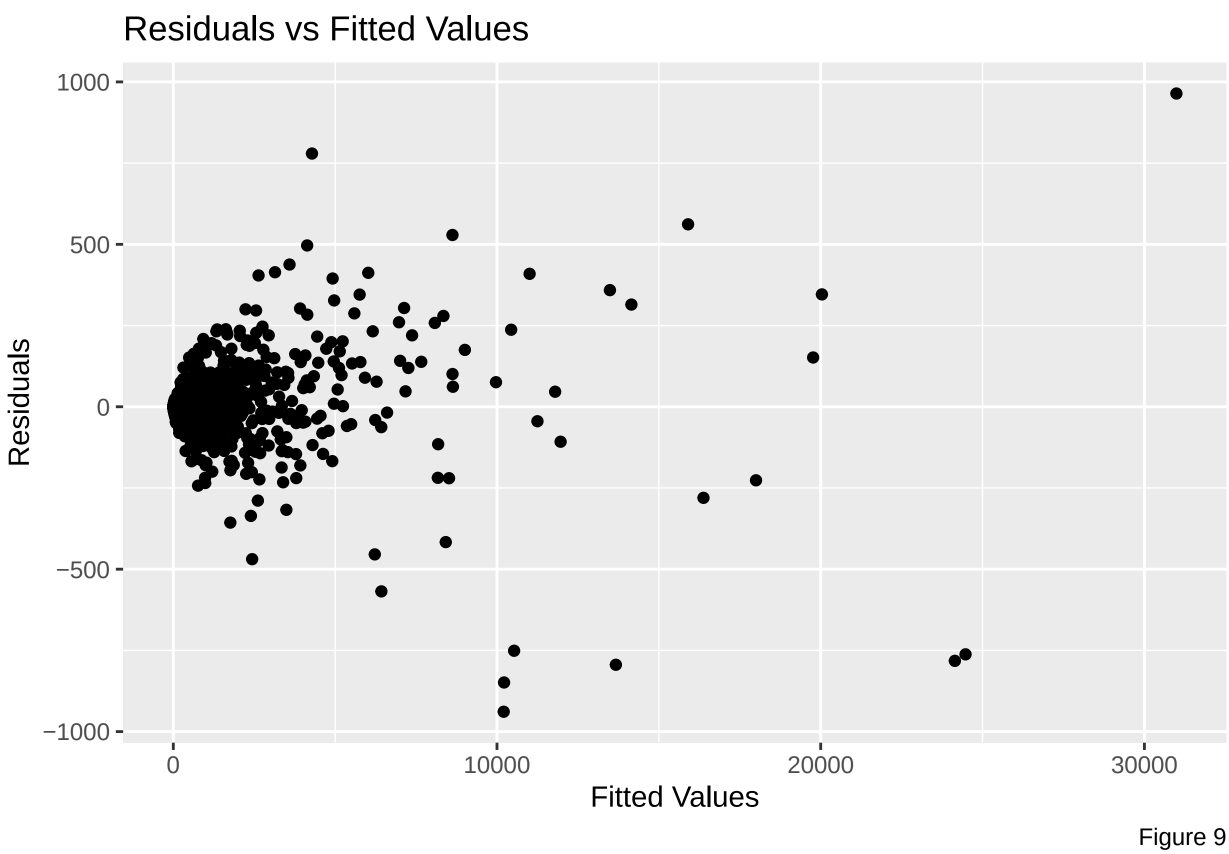
```
## ageAge41M65:party_cdDEM      1039
## ageAgeOver66:party_cdDEM     1263
## ageAge26M40:party_cdGRE      1218
## ageAge41M65:party_cdGRE      1402
## ageAgeOver66:party_cdGRE     1249
## ageAge26M40:party_cdLIB      1094
## ageAge41M65:party_cdLIB       972
## ageAgeOver66:party_cdLIB     1258
## ageAge26M40:party_cdREP      1088
## ageAge41M65:party_cdREP      1132
## ageAgeOver66:party_cdREP     1266
## ageAge26M40:party_cdUNA      1056
## ageAge41M65:party_cdUNA      1081
## ageAgeOver66:party_cdUNA     1264
## party_cdDEM:sex_codeM        1142
## party_cdGRE:sex_codeM        1200
## party_cdLIB:sex_codeM        1324
## party_cdREP:sex_codeM        1171
## party_cdUNA:sex_codeM        1073
## party_cdDEM:sex_codeU        1230
## party_cdGRE:sex_codeU        1430
## party_cdLIB:sex_codeU        1297
## party_cdREP:sex_codeU        1199
## party_cdUNA:sex_codeU        1246
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

```
#compute AUC
fitted<-fitted(final_model)
est_p<-fitted[,'Estimate']/sub_df$total_registered
full_vote<-sapply(1:nrow(sub_df),function(i){
  voters<-rep(1,sub_df$total_voted[i])
  nonvoters<-rep(0,sub_df[i,'total_registered']-sub_df$total_voted[i])
  counts<-c(voters,nonvoters)
})
full_vote<-unlist(full_vote)
preds<-rep(est_p,sub_df$total_registered)
preds<-prediction(preds,full_vote)

auc<-performance(preds,measure='auc')
auc@y.values[[1]]
```

```
## [1] 0.6654713
```

```
#residual plot
residual_df <- data.frame(
  Fitted_Values = fitted(final_model),
  Residuals = residuals(final_model)
)

ggplot(residual_df, aes(x = Fitted_Values.Estimate, y = Residuals.Estimate)) +
  geom_point() +
  labs(
    title = "Residuals vs Fitted Values",
```

```
    x = "Fitted Values",
    y = "Residuals",
    caption = "Figure 9"
  )
```




```
#posterior parameter estimates
summary(final_model)$fixed[,c(1,2,3,4)]%>%
  knitr::kable(caption="Figure 10: Posterior Parameter Estimates and 95% Confidence Intervals", digits=4)
```

Table 1: Figure 10: Posterior Parameter Estimates and 95% Confidence Intervals

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	-0.4820	0.1432	-0.7485	-0.2019
ageAge26M40	0.7694	0.1396	0.4972	1.0334
ageAge41M65	0.8733	0.1437	0.5811	1.1514
ageAgeOver66	0.8721	0.2747	0.3157	1.4124
party__cdDEM	0.8064	0.1391	0.5229	1.0640
party__cdGRE	0.6660	0.1894	0.2987	1.0210
party__cdLIB	0.1833	0.1444	-0.1056	0.4600
party__cdREP	0.7564	0.1391	0.4781	1.0184
party__cdUNA	0.3017	0.1389	0.0243	0.5603
race__codeB	-0.3095	0.0135	-0.3365	-0.2828
race__codeI	-0.3176	0.0187	-0.3540	-0.2803
race__codeM	-0.2405	0.0214	-0.2842	-0.1990
race__codeO	-0.3359	0.0155	-0.3658	-0.3045

	Estimate	Est.Error	l-95% CI	u-95% CI
race_codeP	1.0523	0.2280	0.6147	1.5280
race_codeU	0.1005	0.0152	0.0717	0.1305
race_codeW	0.1298	0.0132	0.1037	0.1562
ethnic_codeNL	0.3682	0.0088	0.3509	0.3854
ethnic_codeUN	0.2922	0.0090	0.2742	0.3096
sex_codeM	0.0390	0.1228	-0.2006	0.2768
sex_codeU	0.2363	0.1515	-0.0516	0.5436
ageAge26M40:party_cdDEM	-0.6575	0.1400	-0.9273	-0.3867
ageAge41M65:party_cdDEM	0.0229	0.1442	-0.2623	0.3153
ageAgeOver66:party_cdDEM	0.0863	0.2747	-0.4521	0.6408
ageAge26M40:party_cdGRE	-0.3793	0.1993	-0.7740	-0.0031
ageAge41M65:party_cdGRE	-0.5162	0.2200	-0.9517	-0.0743
ageAgeOver66:party_cdGRE	-0.3135	0.4467	-1.2048	0.5522
ageAge26M40:party_cdLIB	-0.6248	0.1450	-0.8960	-0.3361
ageAge41M65:party_cdLIB	-0.3125	0.1514	-0.6078	-0.0078
ageAgeOver66:party_cdLIB	-0.0964	0.2911	-0.6499	0.5009
ageAge26M40:party_cdREP	-0.6199	0.1399	-0.8877	-0.3440
ageAge41M65:party_cdREP	0.1037	0.1437	-0.1776	0.3941
ageAgeOver66:party_cdREP	0.1364	0.2750	-0.4044	0.6884
ageAge26M40:party_cdUNA	-0.5008	0.1399	-0.7689	-0.2289
ageAge41M65:party_cdUNA	0.1653	0.1440	-0.1149	0.4607
ageAgeOver66:party_cdUNA	0.4967	0.2748	-0.0412	1.0497
party_cdDEM:sex_codeM	-0.3213	0.1231	-0.5602	-0.0805
party_cdGRE:sex_codeM	-0.1415	0.1821	-0.4992	0.2324
party_cdLIB:sex_codeM	-0.0012	0.1273	-0.2530	0.2402
party_cdREP:sex_codeM	-0.0798	0.1231	-0.3194	0.1606
party_cdUNA:sex_codeM	-0.1472	0.1229	-0.3831	0.0919
party_cdDEM:sex_codeU	0.0699	0.1513	-0.2352	0.3546
party_cdGRE:sex_codeU	-0.0692	0.2090	-0.4663	0.3464
party_cdLIB:sex_codeU	0.6087	0.1616	0.2846	0.9208
party_cdREP:sex_codeU	0.3008	0.1519	-0.0063	0.5887
party_cdUNA:sex_codeU	-0.0781	0.1518	-0.3828	0.2119

```r
#random effects
final_model %>%
  spread_draws(r_county_desc[county_desc,]) %>%
  median_qi(`Group Effects` = r_county_desc) %>%
  ggplot(aes(y = county_desc, x = `Group Effects`, xmin = .lower, xmax = .upper)) +
  geom_pointinterval(orientation = "horizontal")+
  labs(y='County',title='County treatment effects on odds of voting',
       caption='Figure 11')
```

```
## Warning: ‘gather_()‘ was deprecated in tidyr 1.2.0.
## i Please use ‘gather()‘ instead.
## i The deprecated feature was likely used in the tidybayes package.
##   Please report the issue at <https://github.com/mjskay/tidybayes/issues/new>.
```

```
## Warning: Using the ‘size‘ aesthietic with geom_segment was deprecated in ggplot2 3.4.0.
## i Please use the ‘linewidth‘ aesthetic instead.
```

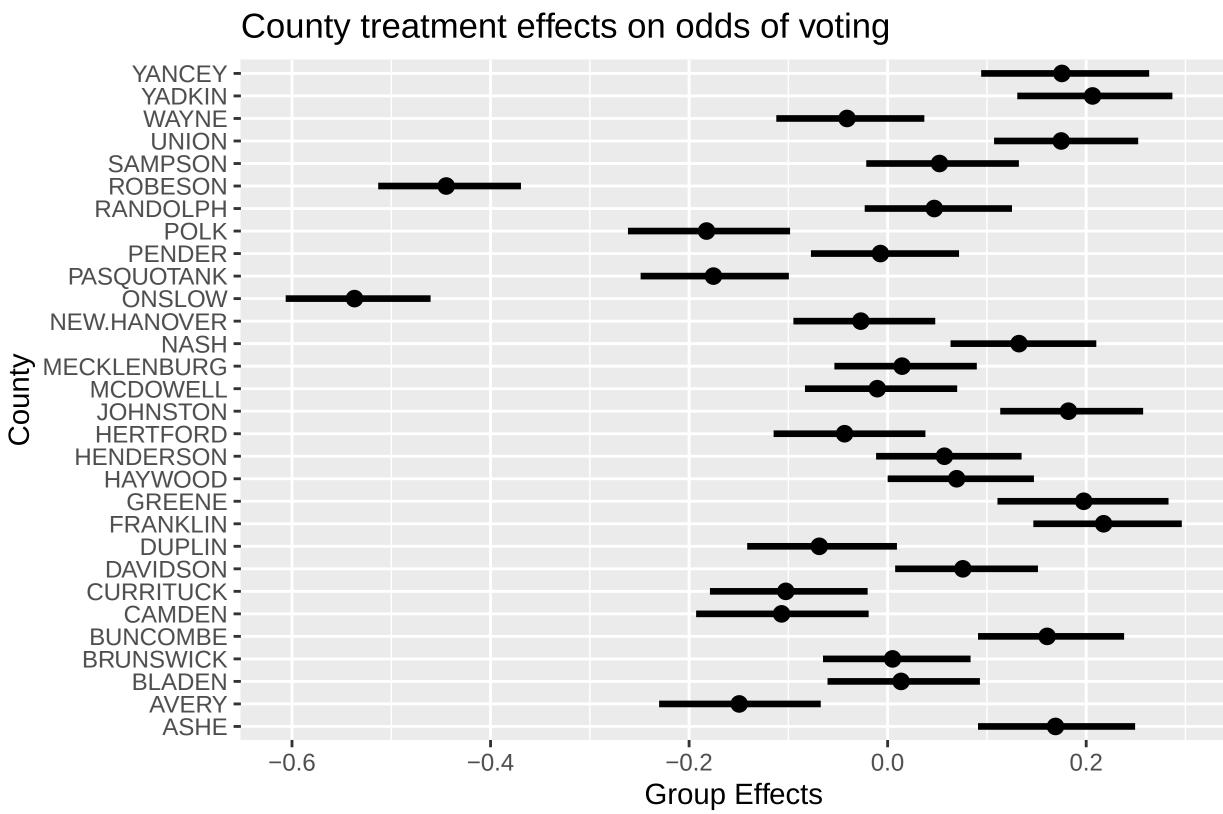

County treatment effects on odds of voting
YANCEY
YADKIN
WAYNE
UNION
SAMPSON
ROBESON
RANDOLPH
POLK
PENDER
PASQUOTANK
ONSLOW
NEW.HANOVER
NASH
MECKLENBURG
MCDOWELL
JOHNSTON
HERTFORD
HENDERSON
HAYWOOD
GREENE
FRANKLIN
DUPLIN
DAVIDSON
CURRITUCK
CAMDEN
BUNCOMBE
BRUNSWICK
BLADEN
AVERY
ASHE
County
−0.6
−0.4
−0.2
0.0
0.2
Group Effects

Figure 11