Case Study 2

Marc Brooks (Presenter), Bo Liu (Checker), Shirley Mathur (Programmer), Aasha Reddy (Writer)

11/2/2021

Introduction

The North Carolina State Board of Elections (NCSBE) is the agency charged with the administration of the elections process and campaign finance disclosure and compliance. Among other things, they provide voter registration and turnout data online. Our goal in this case study is to use the NC voter files for the general elections in November 2020 to identify/estimate how different groups voted in the 2020 elections (out of those registered).

Research Questions

1) How did different demographic subgroups vote in the 2020 general elections? For example, how did the turnout for males compare to the turnout for females after controlling for other potential predictors?
2) Did the overall probability or odds of voting differ by county in 2020? Which counties differ the most from other counties?
3) How did the turnout rates differ between females and males for the different party affiliations?
4) How did the turnout rates differ between age groups for the different party affiliations?

Data and cleaning

We have two datasets that we will merge: voter_stats_20201103.txt contains information about the aggregate counts of registered voters by the demographic variables, and history_stats_20201103.txt contains information about the aggregate counts of voters who actually voted by the demographic variables. We first clean the data.

Our outcome variable of interest needs to be created. We are examining the turnout of voters, and we have counts of total registered voters in one dataset (we will call this dataset "registered"), and counts of people who actually voted in another dataset (we will call this dataset "voted"). We will end up wanting one column for successes (number of people who actually voted), and one column for total number of people who were elligible to vote (total registered).

First, we note the variables we definitely would like to keep in our final dataset to answer our questions of interest. From the registered dataset, we would like to keep total registered voters. From the voted dataset, we would like to keep `total voters`. We would also like to keep `sex_code`, `age`, and `county_desc` from both datasets. Additional variables of interest include `voted_party_cd`, `race_code`, `ethnic_code` which appear in both datasets, so we can merge on these. From the voted dataset, we will not keep `voting_method`, and `voting_method_desc` since they do not appear in the registered dataset. From the prompt, we also note that we use the `voted_party_cd` variable in the voted dataset, and not the `party_cd` variable.

We note that in the registered dataset, `precinct_abbrv` and `vtd_abbrv` will not be useful as they contain over 1000 factors each. `update_date` is also only NAs, so we will exclude that variable as well. We also will exclude `election_date` and `stats_type` as indicated in the prompt.

Overall, in the voted dataset, we would like to keep the following variables: `county_desc`, `age`, `voted_party_cd`, `race_code`, `ethnic_code`, `sex_code`, `total_voters`, so we will aggregate to this level using a dplyr::group_by(). Overall, in the registered dataset, we would like to keep the following variables: `county_desc`, `voted_party_cd`, `race_code`, `ethnic_code`, `sex_code`, `age`, `total_registered`, so we will again aggregate to this level using a dplyr::group_by(). After this, we can merge the datasets.

We note that after merging, we have 6081 missing observations for `total_voters`. This is because there were demographic/geographic groups in registered dataset that did not exist in the voted dataset. This likely means that these demographic groups who registered to vote, did not actually vote, since there are no values of `total_voters` = 0 in the voted dataset. Thus, we impute 0 for the missing values of `total_voters` to indicate that these demographic groups did not vote.

Finally, we notice that some counties have instances where the total number of voters is larger than the total number of registered voters (12 observations in total), which is impossible and will not work in our model. For these counties, we reduce the total number of voters to be equal to the total number of registered voters.

Final dataset: 51,906 observations (after imputing 0 for NA values in `total_voters`)

Finally, we select a random sample of 30 counties to continue our modeling with. The 30 counties we choose are: Hyde, Edgecombe, Haywood, Clay, Montgomery, Durham, Cabarrus, Cherokee, Duplin, Orange, Bladen, Sampson, Transylvania, Davie, Surry, Stanly, Watauga, Caldwell, Anson, Robeson, Beaufort, Pender, Graham, Stokes, Martin, Lenoir, Hertford, Wilson, Randolph, and New Hanover.

The final dataset after sampling 30 counties has 14,935 observations after deleting missing values.

EDA

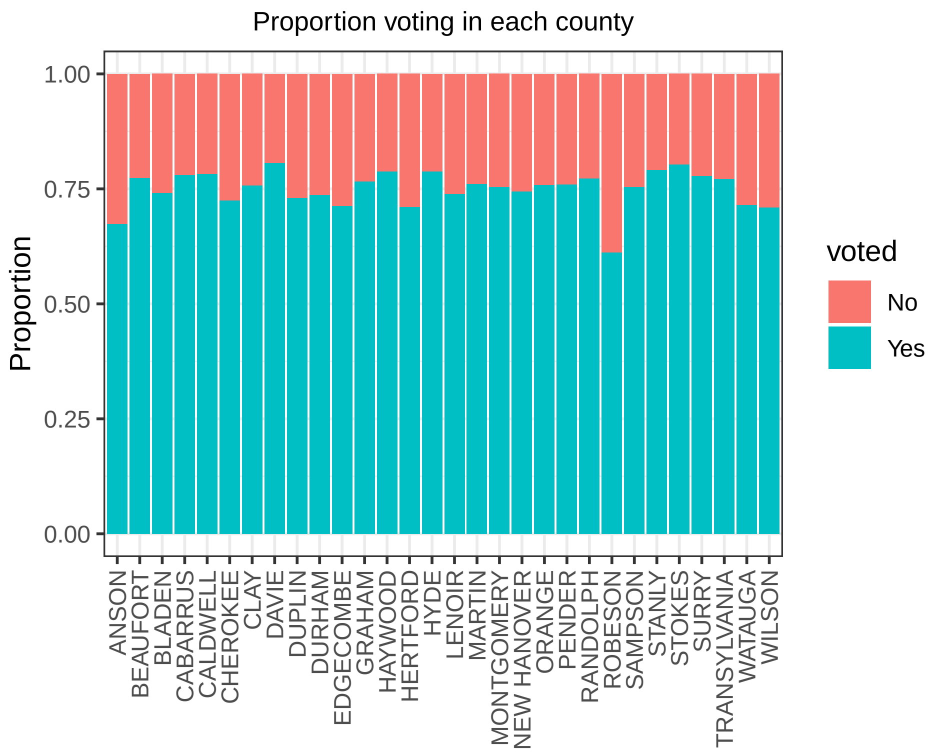
First we de-aggregate the dataset so that we can do some EDA on it, meaning we expand the dataset to have our outcome variable, whether someone voted or not (`voted`) be 0 or 1, rather than having an aggregated count. This increases our dataset to 1,590,202 observations. Note that we will use the aggregated dataset for modeling since having over 1 million observations is computationally infeasible.

Examine outcome variable: `voted`

Our outcome variable, `voted`, is binary, and equals "Yes" if the person in the dataset voted, and "No" if they did not. We see that there are 1,190,593 people who voted and 402,165 people who did not vote.

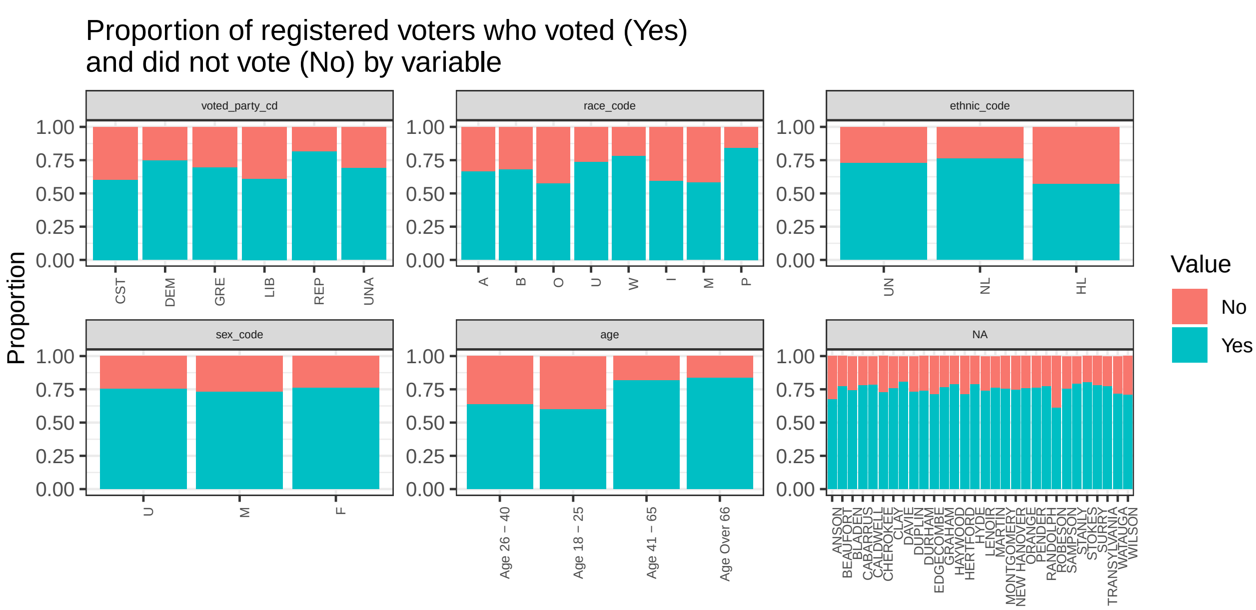
Research question 2: random intercept for county

Our 2nd research question asks us to examine whether the overall probability or odds of voting differed by county in 2020. This points to inclusion of a random intercept by county. We can see that there is a small amount variation in the proportion of registered voters who voted by county. We will include this variable as our grouping variable in our logistic hierarchical model based on our research questions.



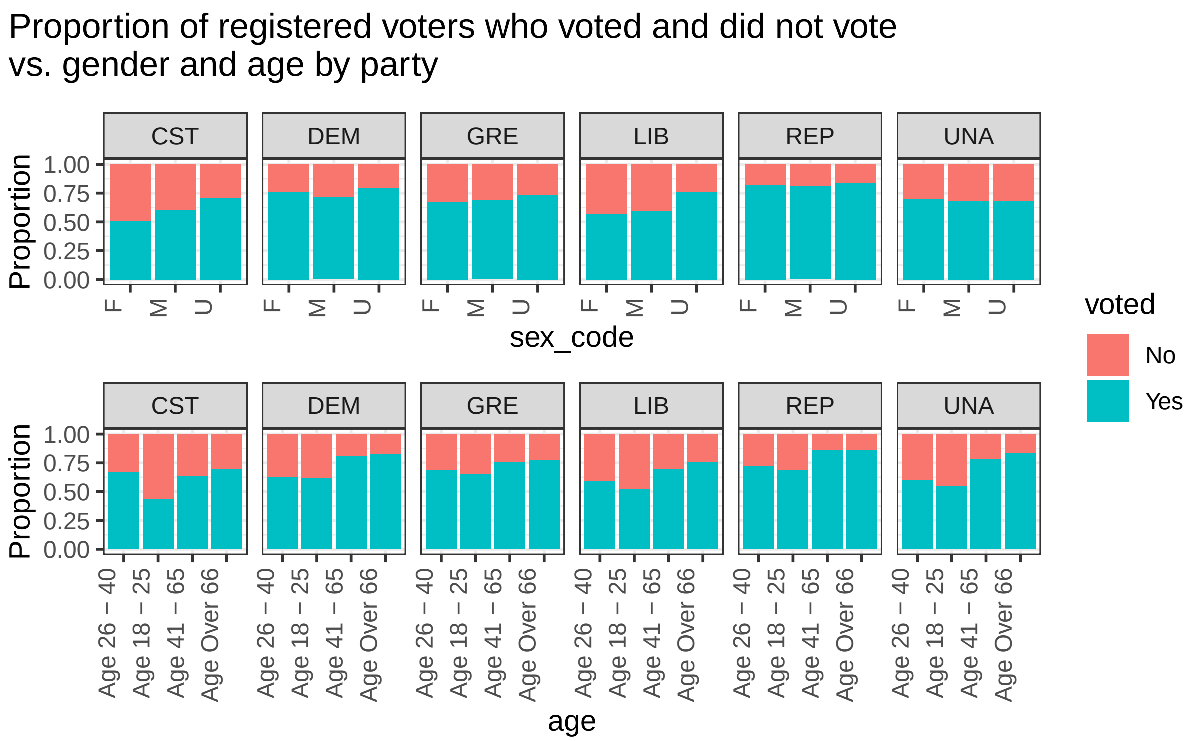
Research question 1: Examine fixed effects

Our first research questions asks to examine how different demographic subgroups voted, and how turnout for males compared to turnout for females after controlling for other predictors. This points to examination of different fixed effects, with a emphasis on examining gender. In our model selection process, we will test all fixed effects, but we use EDA to examine which fixed effects may be most related to our outcome. `age`, `voted_party_cd`, `race_code`, and `ethnic_code` seem to vary the most among their levels with the proportion of those who voted. Note that we will also definitely include `sex_code` as this is explicitly in our research question.



Research questions 3 and 4: interactions

Research question 3 asks how the turnout rates different between males and females for different parties, and question 4 asks how turnout rates differ between age groups for the different parties. These point to `sex_code:voted_party_cd` and `age:voted_party_cd` interactions, respectively. We will thus include these interactions in our final model. According to the below plots, we can see there is some initial evidence of different turnout rates of genders and ages by parties, and we will explore this further in our model.



We also assess other potential interactions, and find that `sex_code:race_code` and `voted_party_cd:race_code` could be helpful. With that being said, we decide to exclude all two-way interactions that do not address the state research questions. Additional two-way interactions increased the difficulty of fitting a model and greatly increase the difficulty of interpreting our results.

Assessing random slopes

We also assess potential random slopes of our variables by county. We do see some heterogeneity of voter turnout by `ethnic_code` among the counties (See appendix for plots). Ultimately we decided to not test for random slopes. When only considering random intercepts, we encountered convergence issues with frequentest model and excessive runtimes with the Bayesian implementation. Therefore, the addition of random slopes would only increase the difficulty of fitting a model and did not specifically address our research questions.

Model selection and specification

We first tried to do exhaustive search selection using BIC, starting with the base model of the random intercept for county, and a fixed effects for `sex_code`, `age`, `voted_party_cd`, `sex_code:voted_party_cd`, and `age:voted_party_cd` based on our research questions. We aimed to use BIC to select other potential fixed effect predictors and test all possible 2-way interactions, however, none of our models converged. BIC is thus an unreliable measure to compare models, and we choose to use our model that incorporates elements from the research question as our final model. As mentioned we do not fit any random slopes due to convergence issues. Due to these convergence issues we ultimately use a Bayesian implementation.

Our final model is thus:

$$y_{ij} \mid N_{ij}, \pi_{ij} \overset{\text{iid}}{\sim} \text{Bin}(N_{ij}, \pi_{ij}), \quad \text{logit}(\pi_{ij}) = \boldsymbol{x}_{ij}^{\text{T}}\boldsymbol{\beta} + b_i, \quad b_i \sim N(0, \sigma^2)$$
$$\beta \sim N(\boldsymbol{\mu}, \Sigma), \quad \sigma^2 \sim IG(\nu_0/2, \nu_0\sigma_0^2/2)$$

where y_{ij} is the number of people voted, and N_{ij} is the number of total voters for group j in county i (random intercept indexed by i). Each $\boldsymbol{x}_{ij}$ consists of 86 elements - the first one being 1 for intercept, and the next 85 elements being the value of the covariates (in one hot form) for group j in county i, namely each possible value of `race_code`, `ethnic_code`, `sex_code`, `age`, `voted_party_cd`, `sex_code:age`, `voted_party_cd:race_code`, `voted_party_cd:sex_code`, `voted_party_cd:age` except baseline. It should be noted that Σ is a diagonal matrix as `brms` only allows indpendent priors on the fixed effect coefficients.

Results and Interpretation

Interpretations

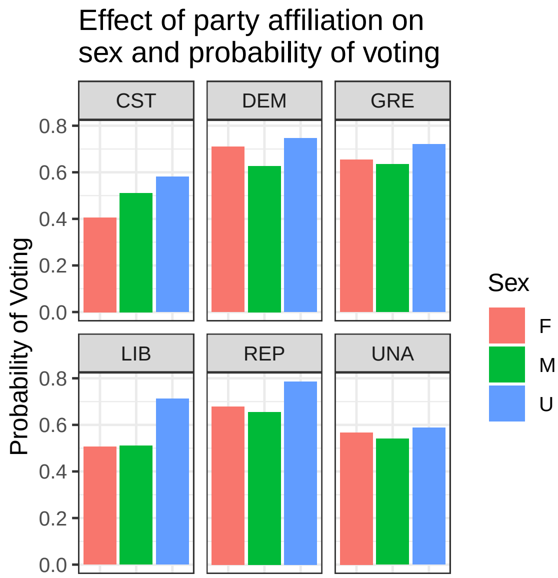
Main effects

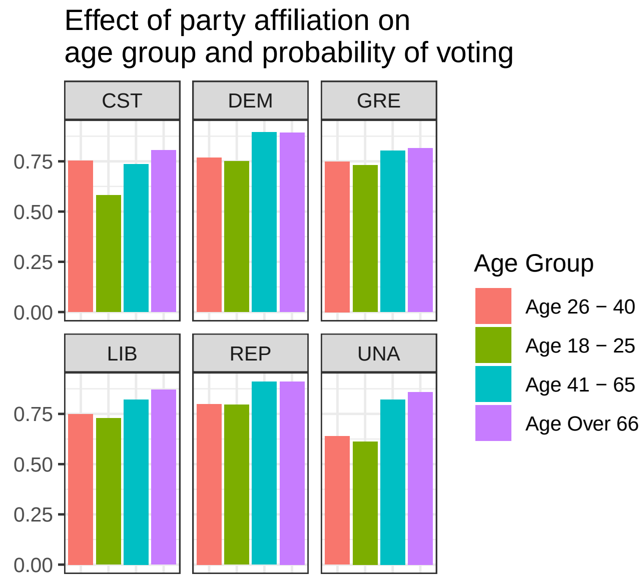
Our estimated intercept is -0.04. This is the expected log odds of someone voting, across all counties in our sample, who is Asian, ages 26-40, female, and did not specify their ethnicity. That means the odds of voting is approximately 1. The estimate, clearly is not significant and contains 1 in the posterior credible interval (on odds scale).

For the purposes of interpretation, statements of significance imply that the 95% posterior credible interval did not include 0. The following highlight significant results concerning the relationship between voting odds and various demographics.

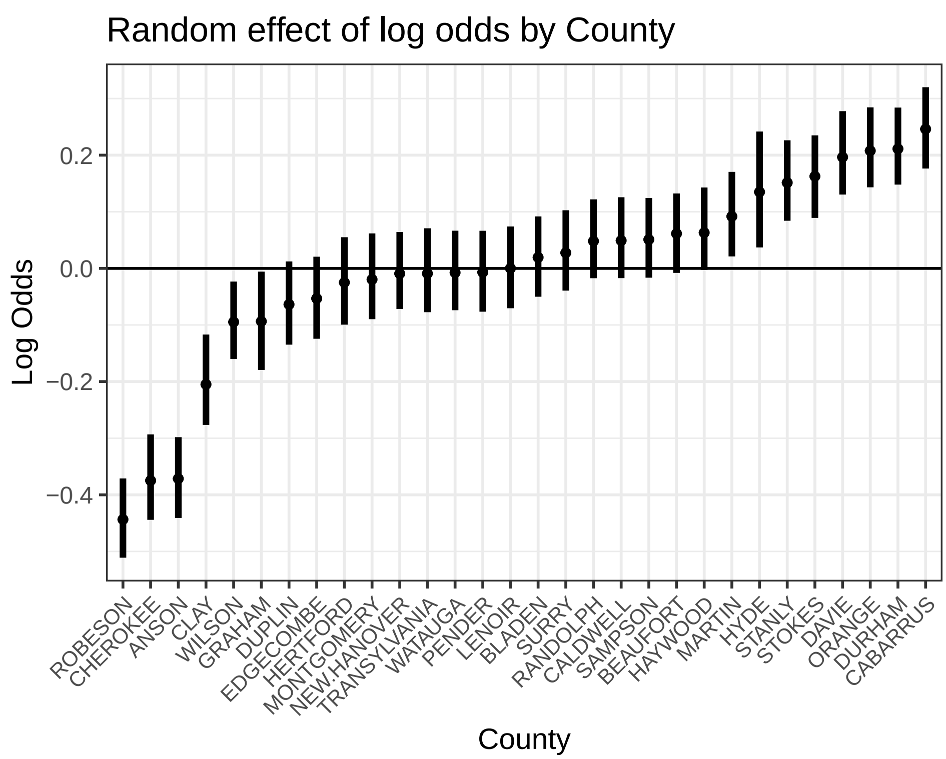
- On average, we expect the odds of voting for someone between the ages 18-26 to decrease by 56% when compared to some that is Ages 26-40.
- On average, we expect the odds of voting for someone who is White to increase by 46% when compared to some who is Asian. We identified a similar relationship for someone registered with an undesignated race.
- On average, we expect the odds of voting for males to increase by 42% when compared to females.
- On the other hand, we expect the odds of voting for someone who is registered under two or more races to decrease by 8% when compared to some who is Asian. We identified a similar relationship for someone registered as black.
- On average, we expect the odds of voting for someone who is not Hispanic or Latino to increase by 14% when compared to some who has an undesignated ethnicity. Someone who is Latino or Hispanic sees an expected decrease in odds of voting of 25% when compared to someone with an undesignated ethnicity.

Interaction Effects





From the above plots we see that older age groups consistently have a higher probability of voting across all parties. There is some variation across parties in probability of voting between age groups 26-40 and 18-25. Overall, there doesn't appear to be a strong interaction effect between age group and voter party. We see a similar pattern for gender and voter party. Undesignated gender consistently had the highest probability of voting, while male and females often had a similar probability of voting across parties with the exception of CST and DEM.

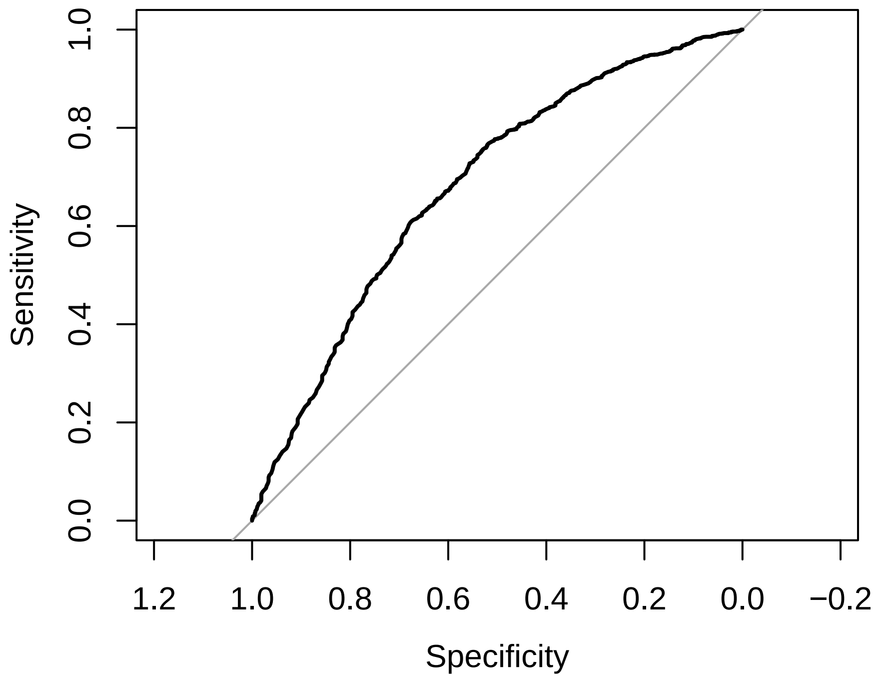


The interval plot demonstrates a large amount of heterogeneity in the log odds of voting across our sampled counties. For a registered voter who is Asian, between ages 26-40, female, and did not specify their ethnicity, the odds of voting decreases by more than 32.9% if they are in Robeson county and increases by more than

22% if they are in Cabarrus county.

ROC Curve for Model Evaluation

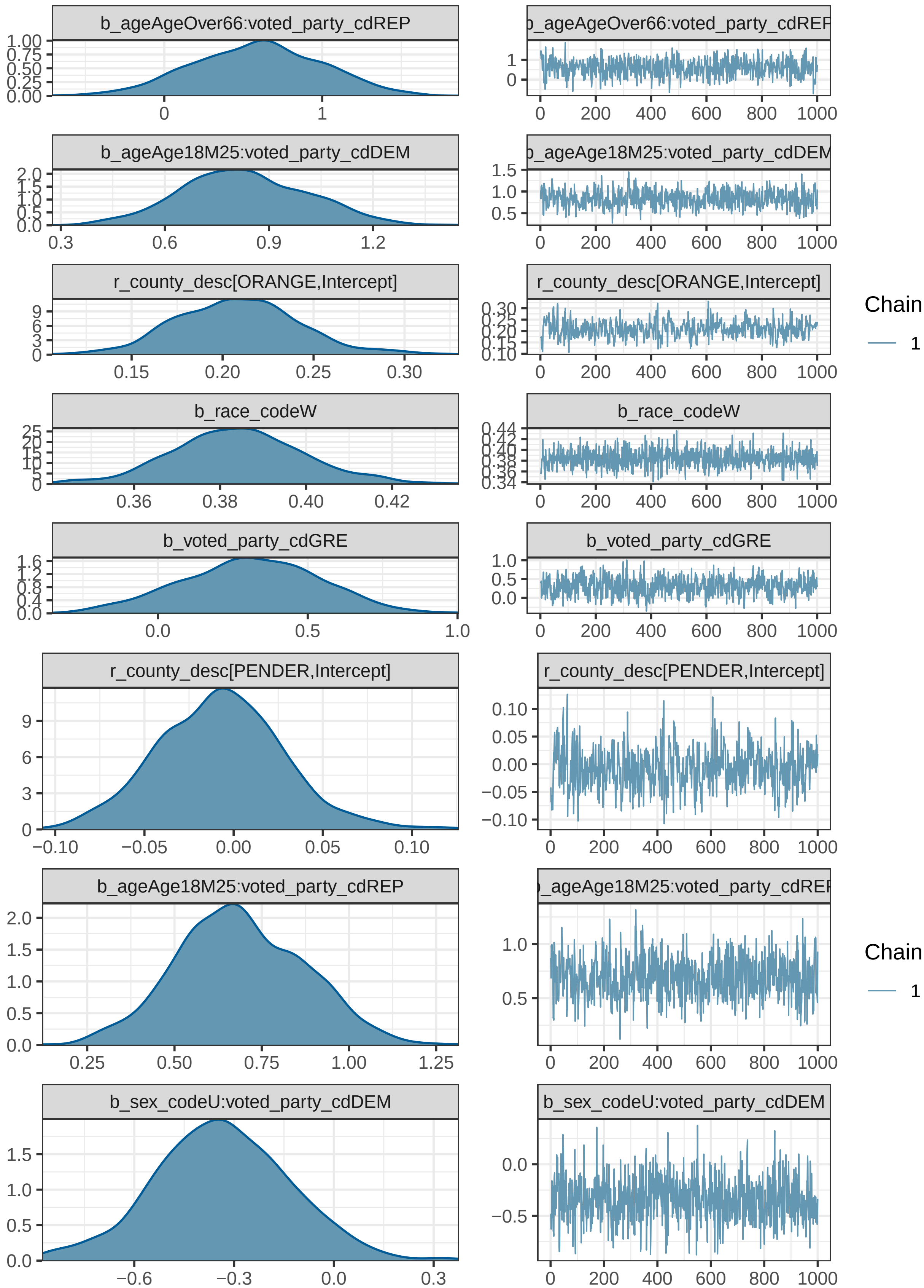
We plotted an ROC curve to evaluate our model and see how well it does at predicting the whether or not a particular voter voted. From the curve below, you can see that our model does better than simply taking random guesses for whether or not a voter belonging to a particular demographic within a particularly voted or not. Furthermore, the AUC from our ROC curve is 0.681, which indicates that our model does a decent job of prediction since the value is greater than 0.5, which is what the AUC would be had we simply taken random guesses as to whether an individual voted or not. Thus, our model seems to be useful in predicting whether or not individuals of certain demographics in various counties in NC voted or not.




```
## Area under the curve: 0.681
```

Posterior Checks for Model

We also examined the trace plots and posterior distributions of the parameters in our model to check that our model had indeed converged. Since we have several parameters, the posterior distributions and trace plots for only a few parameters are included below, but in general from our posterior checks, the model indeed converged.


b_ageAgeOver66:voted_party_cdREP
b_ageAge18M25:voted_party_cdDEM
r_county_desc[ORANGE,Intercept]
b_race_codeW
b_voted_party_cdGRE
r_county_desc[PENDER,Intercept]
b_ageAge18M25:voted_party_cdREP
b_sex_codeU:voted_party_cdDEM
Chain
1
Chain
1

Appendix

Code Appendix

```
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE,
                      fig.align = 'center')
library(tidyverse)
library(lme4)
library(rstan)
library(brms)
library(knitr)
library(kableExtra)
library(patchwork)
library(lubridate)
library(gridExtra)
library(influence.ME)
library(pROC)
library(tidybayes)


options(scipen = 0, digits = 4)
ggplot2::theme_set(ggplot2::theme_bw())
#load data
registered <- read.delim("../data/voter_stats_20201103.txt")
voted <- read.delim("../data/history_stats_20201103.txt")
registered <- registered %>%
  dplyr::rename(total_registered = total_voters) %>%
  dplyr::rename(voted_party_cd = party_cd) %>%
  mutate(total_registered = as.numeric(total_registered))

voted <- voted %>%
  select(-voting_method_desc) %>%
  mutate(total_voters = as.numeric(total_voters))
v <- voted %>%
  group_by(county_desc, age, voted_party_cd, race_code, ethnic_code, sex_code) %>%
  summarize(total_voters = sum(total_voters))

r <- registered %>%
  group_by(county_desc, voted_party_cd, race_code, ethnic_code, sex_code, age) %>%
  summarize(total_registered = sum(total_registered))

vote <- r %>%
  left_join(v, by = c("county_desc", "age", "voted_party_cd", "race_code",
                      "ethnic_code", "sex_code")) %>%
  mutate(total_registered = as.numeric(total_registered),
         total_voters = as.numeric(total_voters))

# apply(vote, 2, function(x) sum(is.na(x))) %>%
#   kbl()

vote <- vote %>%
  mutate(total_voters = as.numeric(total_voters),
         total_registered = as.numeric(total_registered),
```

```
         county_desc = as.factor(county_desc),
         voted_party_cd = as.factor(voted_party_cd),
         race_code = as.factor(race_code),
         sex_code = as.factor(sex_code),
         ethnic_code = as.factor(ethnic_code),
         age = as.factor(age)
         ) %>%
  mutate(total_voters = case_when(
    total_voters > total_registered ~ total_registered,
    TRUE ~ total_voters
  )) %>%
  mutate(total_voters = ifelse(is.na(total_voters), 0, total_voters))
# test that all the NAs in the final dataset in voted_party_cd is correct - it is
registered %>%
  filter(county_desc == "ALAMANCE", voted_party_cd == "CST", race_code == "B",
         ethnic_code == "NL", sex_code == "F", age == "Age 18 - 25")

voted %>%
  filter(county_desc == "ALAMANCE", voted_party_cd == "CST", race_code == "B",
         ethnic_code == "NL", sex_code == "F", age == "Age 18 - 25")
set.seed(99)
county_samp <- as.character(sample(unique(vote$county_desc), size = 30))

# this is a random sample, save it so report is reproducible
county_samp <- c("HYDE", "EDGECOMBE", "HAYWOOD", "CLAY", "MONTGOMERY", "DURHAM", "CABARRUS", "CHEROKEE"

vote <- vote %>%
  filter(county_desc %in% c(county_samp))

# create a 0 variable for people who didnt vote
vote <- vote %>%
  mutate(total_nonvoters = total_registered - total_voters)

# elongate vote dataset to make it not aggregate
vote_long <- vote %>%
  pivot_longer(cols = c("total_voters", "total_nonvoters"), names_to = "voted", values_to = "freq") %>%
  mutate(voted = ifelse(voted == "total_voters", "Yes", "No")) %>%
  mutate(obs = map(freq, ~rep_len(1, .x))) %>%
  unnest() %>%
  select(-freq, -obs)
table(vote_long$voted)
ggplot(vote_long) +
  geom_bar(aes(x = county_desc, fill = voted),
           position = "fill") +
  labs(y = "Proportion",
       title = "Proportion voting in each county") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5),
        axis.title.x = element_blank())

names <- vote_long %>%
  select(-county_desc, -total_registered) %>%
  colnames
```

```r
names <- names[-1]

vote_long %>% select(-county_desc, -total_registered) %>%
  pivot_longer(-voted) %>%
  mutate(name = factor(name, levels = names)) %>%
  mutate(voted = as.factor(voted)) %>%
  mutate(type = "barplot") %>%
  ggplot() +
  geom_bar(
    aes(x = value, fill = voted),
    position = "fill",
    data = ~subset(.x, type == "barplot")
  ) +
  facet_wrap(~name, scale = "free", ncol = 3) +
  xlab("") +
  ylab("Proportion") +
  labs(title = "Proportion of registered voters who voted (Yes) \nand did not vote (No) by variable") +
  guides(fill=guide_legend(title="Value")) +
  theme(strip.text.x = element_text(size = 5),
        legend.position = "right",
        axis.text.x = element_text(angle = 90, size = 6, hjust=1))
sex_code_int <- ggplot(vote_long) +
  geom_bar(aes(x = sex_code, fill = voted),
           position = "fill") +
  labs(y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, vjust = -0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5),
        legend.position = "none") +
  facet_wrap(~voted_party_cd, ncol= 6)

age_int <- ggplot(vote_long) +
  geom_bar(aes(x = age, fill = voted),
           position = "fill") +
  labs(y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, vjust = -0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5)) +
  facet_wrap(~voted_party_cd, ncol = 6)

patchwork <- sex_code_int / age_int +
  plot_layout(guides = "collect", widths = c(1, 1))

patchwork + plot_annotation(
  title = "Proportion of registered voters who voted and did not vote \nvs. gender and age by party")
exhaustive_search <- function(raw_model, vars, data) {
  y_name <- deparse(raw_model[[2]], width.cutoff = 500L)
  group_name <- deparse(raw_model[[3]], width.cutoff = 500L)
  id <- 0 : (2^length(vars) - 1)
  construct_model <- function(.id){
    subset <- (.id %/% 2^(0:(length(vars) - 1))) %% 2 == 1
    if (all(subset == F)){
      RHS <- paste0(c("1", group_name), collapse = ' + ')
    }
    else {
```

```
      RHS <- paste0(c(vars[subset], group_name), collapse = ' + ')
    }
    paste(y_name, RHS, sep = ' ~ ')
  }
  run_model <- function(.id){
    model_str <- construct_model(.id)
    model_formula <- as.formula(model_str)
    res <- glmer(model_formula, data = data, family = binomial("logit"))
    return (summary(res)$AICtab)
  }
  models <- sapply(id, construct_model)
  result <- pbapply::pbsapply(id, run_model)
  bind_cols(
    model = models,
    as.data.frame(t(result))
  )
}
# may take 2-3 hrs. result saved to 'ex_result_int.RDS'
ex_result_int <- exhaustive_search(
  raw_model <- cbind(total_voters, total_nonvoters) ~ sex_code + age + voted_party_cd +
    sex_code:voted_party_cd + age:voted_party_cd + (1 | county_desc),
  vars = c("race_code", "ethnic_code", "sex_code:age", "voted_party_cd:race_code"),
  data = vote
)
saveRDS(ex_result_int, "ex_result_int.RDS")
ex_result_int <- readRDS("case_study_02/ex_result_int.RDS")
ex_result_int %>% arrange(BIC) %>% head(10) %>%
  select(1, 3) %>%
  kbl(caption = "Exhaustive search of fixed effects using BIC") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
                latex_options = "HOLD_position")
bayesian_base_model <- readRDS("bayesian_result.RDS")
set.seed(5562)
new_data_sex <- expand.grid(unique(vote$sex_code),
                            unique(vote$voted_party_cd)) %>%
  rename(sex_code = Var1, voted_party_cd = Var2) %>%
  mutate(race_code = as.factor(sample(unique(vote$race_code),1)),
         ethnic_code = as.factor(sample(unique(vote$ethnic_code),1)),
         age = as.factor(sample(unique(vote$age),1)),
         county_desc = as.factor(sample(unique(vote$county_desc),1)),
         sex_code = as.factor(sex_code),
         total_registered = 1
         )
new_data_sex$race_code <- droplevels(new_data_sex$race_code)

new_data_sex$probs <- predict(bayesian_base_model, new_data_sex, type="response")[,1]

p1 <- ggplot(new_data_sex, aes(x = sex_code, y = probs, fill = sex_code)) +
  geom_bar(stat = "Identity") + facet_wrap(~voted_party_cd) +
  labs(title = "Effect of party affiliation on\nsex and probability of voting",
       y = "Probability of Voting", x = "") +
  theme(axis.ticks.x = element_blank(),
        axis.text.x = element_blank()) +
```

```r
  guides(fill=guide_legend(title="Sex"))
set.seed(5562)
new_data_age <- expand.grid(unique(vote$age),
                            unique(vote$voted_party_cd)) %>%
  rename(age = Var1, voted_party_cd = Var2) %>%
  mutate(race_code = sample(unique(vote$race_code),1),
         ethnic_code = sample(unique(vote$ethnic_code),1),
         sex_code = sample(unique(vote$sex_code),1),
         county_desc = sample(unique(vote$county_desc),1),
         total_registered = 1
         )
new_data_age$race_code <- droplevels(new_data_age$race_code)

new_data_age$probs <- predict(bayesian_base_model, new_data_age, type="response")[,1]

p2 <- ggplot(new_data_age, aes(x = age, y = probs, fill = age)) +
  geom_bar(stat = "Identity") + facet_wrap(~voted_party_cd) +
  labs(title = "Effect of party affiliation on\nage group and probability of voting",
       y = "", x = "") +
  theme(axis.ticks.x = element_blank(),
        axis.text.x = element_blank()) +
  guides(fill=guide_legend(title="Age Group"))
p1 + p2
# dotplot
library(lattice)
tmp <- bayesian_base_model %>%
  spread_draws(r_county_desc[county_desc,]) %>%
  median_qi(`Group Means` = r_county_desc)
sorted_county_desc <- tmp %>%
  arrange(`Group Means`, desc = T) %>% pull(county_desc)
p1 <- tmp %>%
  ggplot(aes(y = factor(county_desc, levels = sorted_county_desc),
             x = `Group Means`, xmin = .lower, xmax = .upper)) +
  geom_pointinterval(orientation = "horizontal", fatten_point = .8) +
  labs(title = "Random effect of log odds by County",
       y = "County",
       x = "Log Odds") +
    geom_vline(aes(xintercept = 0)) +
  theme(axis.text.x = element_text(angle = 45, hjust=1, size = 8)) +
  coord_flip()
p1
set.seed(4)
test_subset <- filter(vote_long, county_desc %in% county_samp)
test_subset <- test_subset[sample(c(1:nrow(test_subset)), 3000), ]
test_subset <- test_subset %>%
  select(sex_code, age, voted_party_cd, race_code, ethnic_code, county_desc, voted) %>%
  mutate(race_code = as.factor(race_code),
         age = as.factor(age),
         sex_code = as.factor(sex_code),
         ethnic_code = as.factor(ethnic_code),
         voted_party_cd = as.factor(voted_party_cd),
         total_registered = 1,
         voted = if_else(voted == "Yes", 1, 0),
```

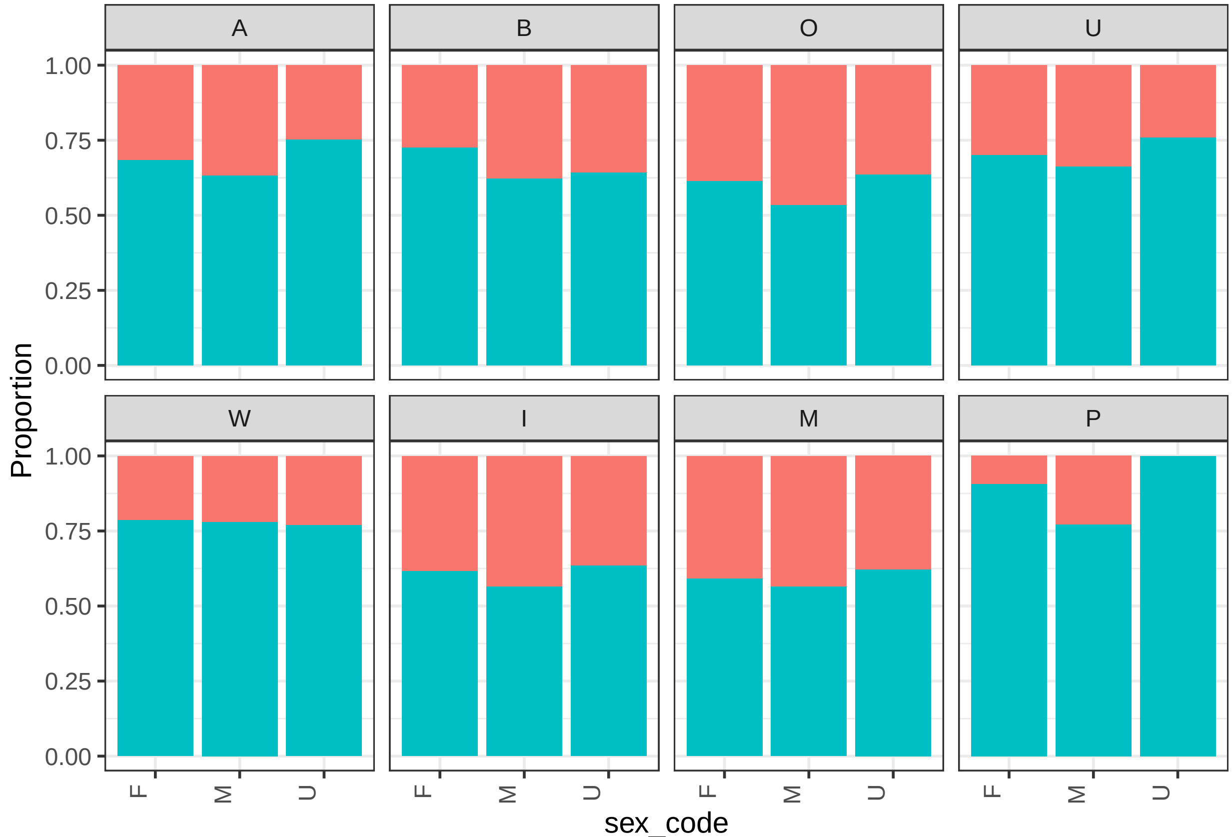
```
         predicted = NA)
test_subset$race_code <- droplevels(test_subset$race_code)
set.seed(4)
test_subset[["predicted"]] <- predict(bayesian_base_model, newdata = test_subset, type="response")[,1]
roc_data <- roc(data = test_subset, response = "voted", predictor = "predicted")
plot.roc(roc_data)
auc(roc_data)
set.seed(7)
variables <- sample(parnames(bayesian_base_model), 8)
plot(bayesian_base_model, variable = variables)

# Interactions
####### sex:code ################################################################
# sex_code:race_code
ggplot(vote_long) +
  geom_bar(aes(x = sex_code, fill = voted),
           position = "fill") +
  labs(y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, vjust = -0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5),
        legend.position = "none") +
  facet_wrap(~race_code, ncol= 4)


# sex_code:ethnic_code
ggplot(vote_long) +
  geom_bar(aes(x = sex_code, fill = voted),
           position = "fill") +
  labs(y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, vjust = -0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5),
        legend.position = "none") +
  facet_wrap(~ethnic_code, ncol= 4)

# sex_code:age
ggplot(vote_long) +
  geom_bar(aes(x = sex_code, fill = voted),
           position = "fill") +
  labs(y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, vjust = -0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5),
        legend.position = "none") +
  facet_wrap(~age, ncol= 4)

######## party_cd ##############################################################
# party_cd:race_code
ggplot(vote_long) +
  geom_bar(aes(x = voted_party_cd, fill = voted),
           position = "fill") +
  labs(y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, vjust = -0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5),
        legend.position = "none") +
```

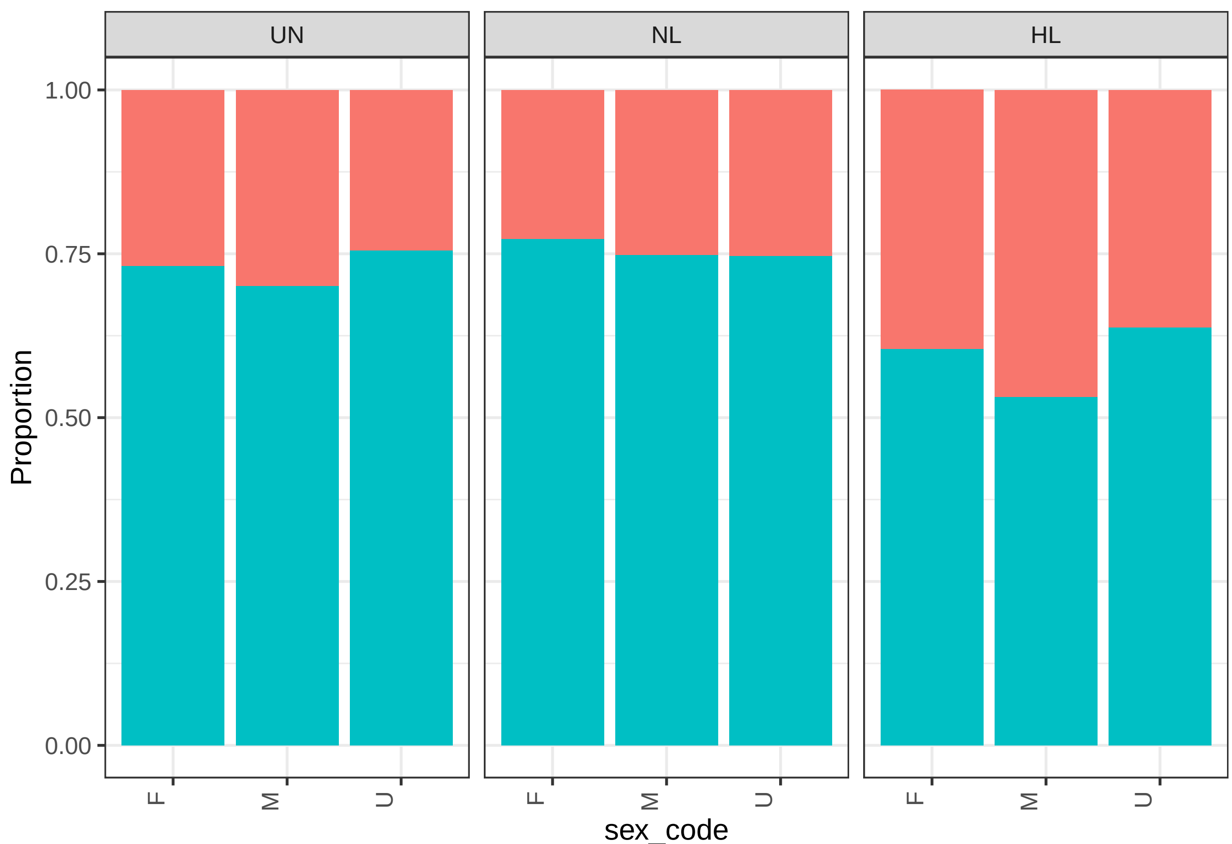
```r
  facet_wrap(~race_code, ncol= 4)

# party_cd:ethnic_code
ggplot(vote_long) +
  geom_bar(aes(x = voted_party_cd, fill = voted),
           position = "fill") +
  labs(y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, vjust = -0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5),
        legend.position = "none") +
  facet_wrap(~ethnic_code, ncol= 4)

###### ethnic_code ###############################################################
# ethnic_code:race_code
ggplot(vote_long) +
  geom_bar(aes(x = ethnic_code, fill = voted),
           position = "fill") +
  labs(y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, vjust = -0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5),
        legend.position = "none") +
  facet_wrap(~race_code, ncol= 4)

# ethnic_code:age
ggplot(vote_long) +
  geom_bar(aes(x = ethnic_code, fill = voted),
           position = "fill") +
  labs(y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, vjust = -0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5),
        legend.position = "none") +
  facet_wrap(~age, ncol= 4)
# MOVE TO APPENDIX
# Random slopes

# voted_party_cd
ggplot(vote_long) +
  geom_bar(aes(x = county_desc, fill = voted),
           position = "fill") +
  labs(y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, vjust = -0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5)) +
  facet_wrap(~voted_party_cd, ncol = 4)

# race_code
ggplot(vote_long) +
  geom_bar(aes(x = county_desc, fill = voted),
           position = "fill") +
  labs(y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, vjust = -0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5)) +
  facet_wrap(~race_code, ncol = 4)
```

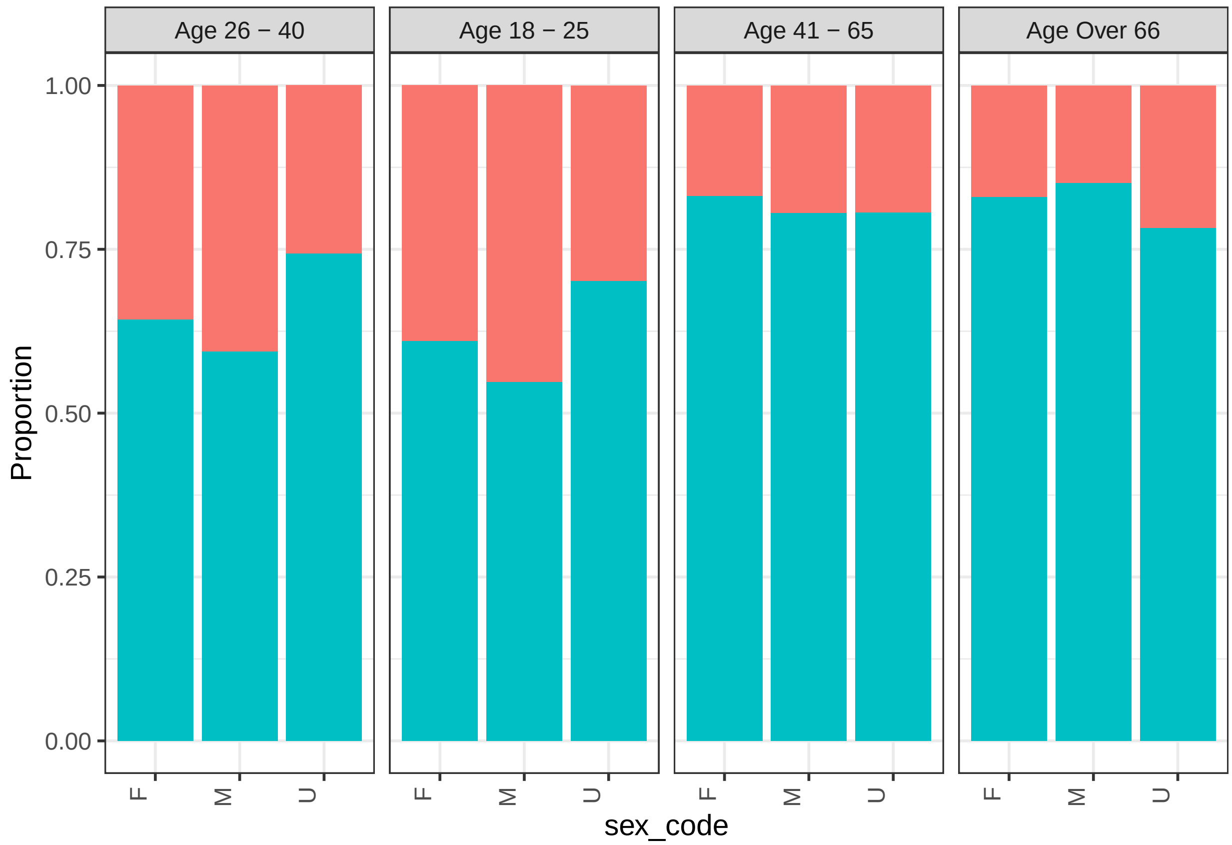
```r
# ethnic_code
ggplot(vote_long) +
  geom_bar(aes(x = county_desc, fill = voted),
           position = "fill") +
  labs(y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, vjust = -0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5)) +
  facet_wrap(~ethnic_code, ncol = 4)

# sex_code
ggplot(vote_long) +
  geom_bar(aes(x = county_desc, fill = voted),
           position = "fill") +
  labs(y = "Proportion") +
  theme(axis.text.x = element_text(angle = 90, vjust = -0.5, hjust=1),
        plot.title = element_text(size = 10, hjust = 0.5)) +
  facet_wrap(~sex_code, ncol = 4)
base_model <-glmer(
              cbind(total_voters, total_nonvoters) ~ sex_code + age + voted_party_cd +
              sex_code:voted_party_cd + race_code + ethnic_code +
              age:voted_party_cd + (1 |county_desc), data = vote,
              family = binomial("logit"),
              control=glmerControl(optimizer="bobyqa",optCtrl=list(maxfun=2e5))
              )

saveRDS(base_model, "base_model.RDS")
base_model <- readRDS('base_model.RDS')
coef(summary(base_model)) %>%
  knitr::kable(caption = "Fixed effect estimates on log odds scale") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"),
               latex_options = "HOLD_position")
bayesian_base_model <-brm(
              total_voters | trials(total_registered) ~ sex_code + age + voted_party_cd +
              sex_code:voted_party_cd + race_code + ethnic_code +
              age:voted_party_cd + (1 |county_desc), data = vote,
              family = binomial("logit"),
              chains = 1)

summary_pars <- summary(bayesian_base_model)$fixed[, c(1,3,4)]
# rownames(summary_pars) <- c("Intercept (grand mean)", "numbeds", "sigma", "sd_control", "sd_state")
colnames(summary_pars) <- c("Est", "Lwr", "Upr")
summary_pars %>%
  knitr::kable(caption = "Estimated posterior parameters",
               digits = 3)
bayesian_<- readRDS("bayesian_result.RDS")
```

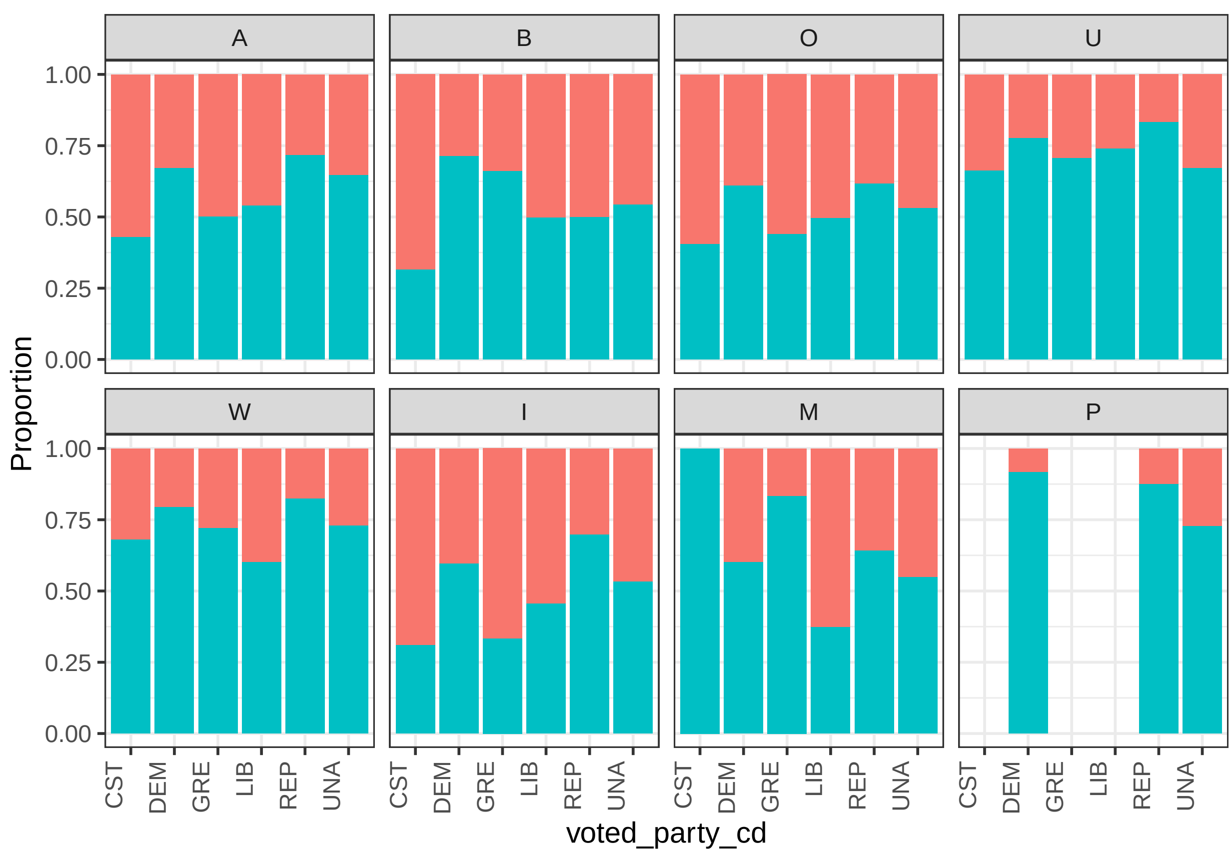
Plot Appendix

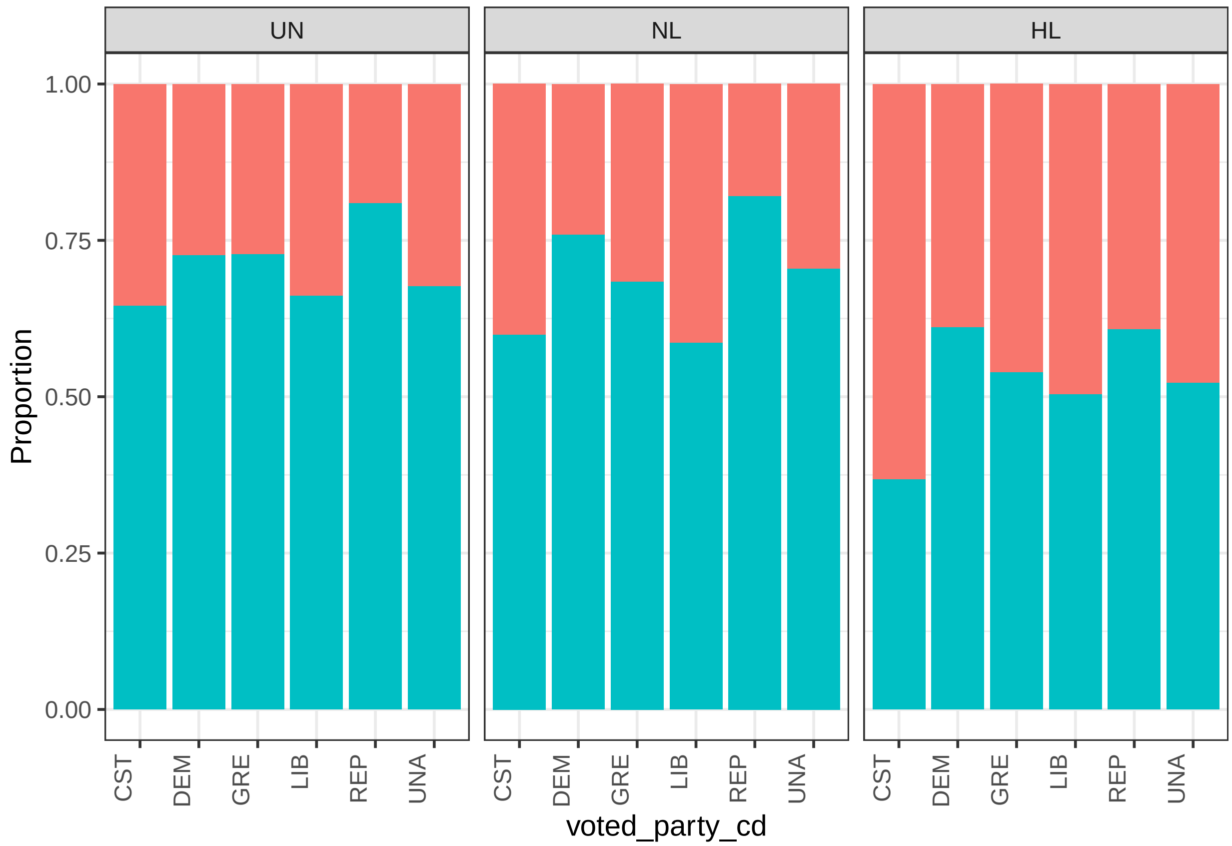
Testing Interactions

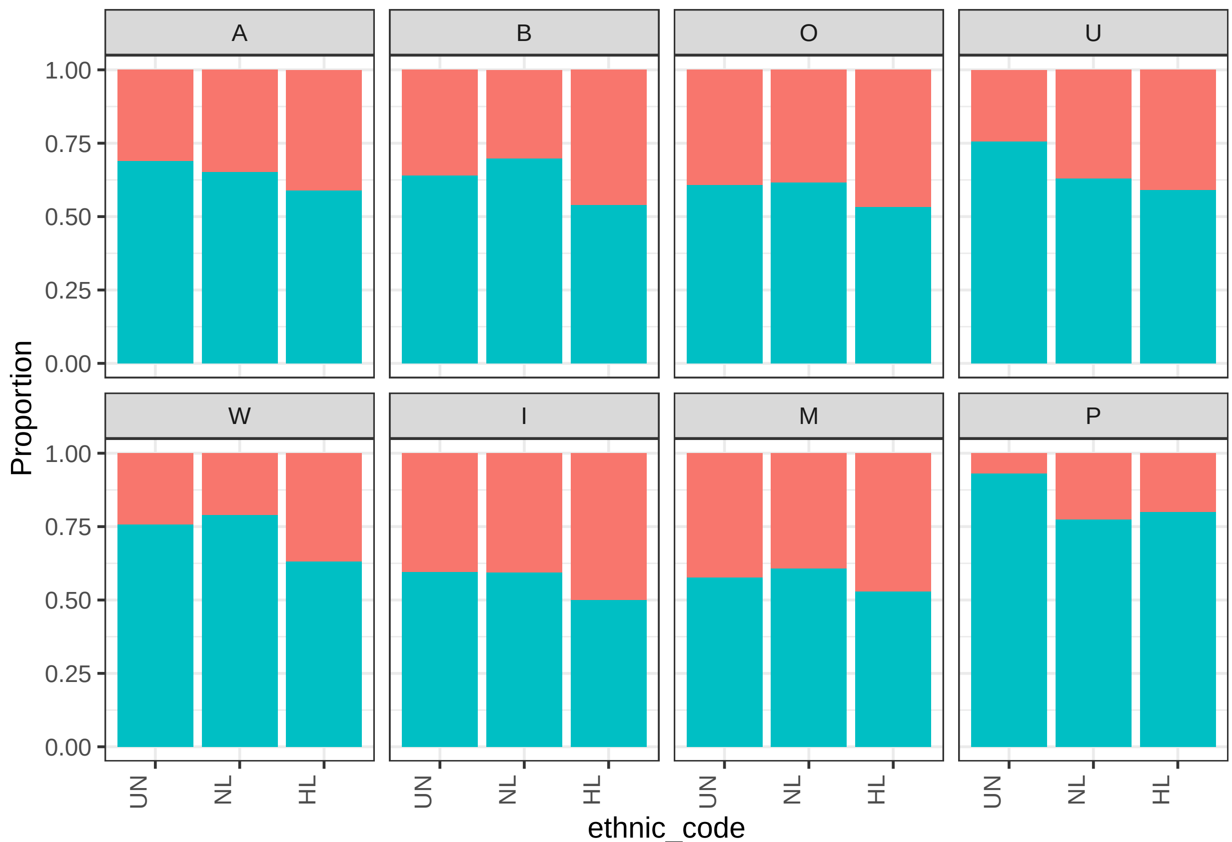

A
B
O
U
W
I
M
P
Proportion
1.00
0.75
0.50
0.25
0.00
F
M
U
sex_code

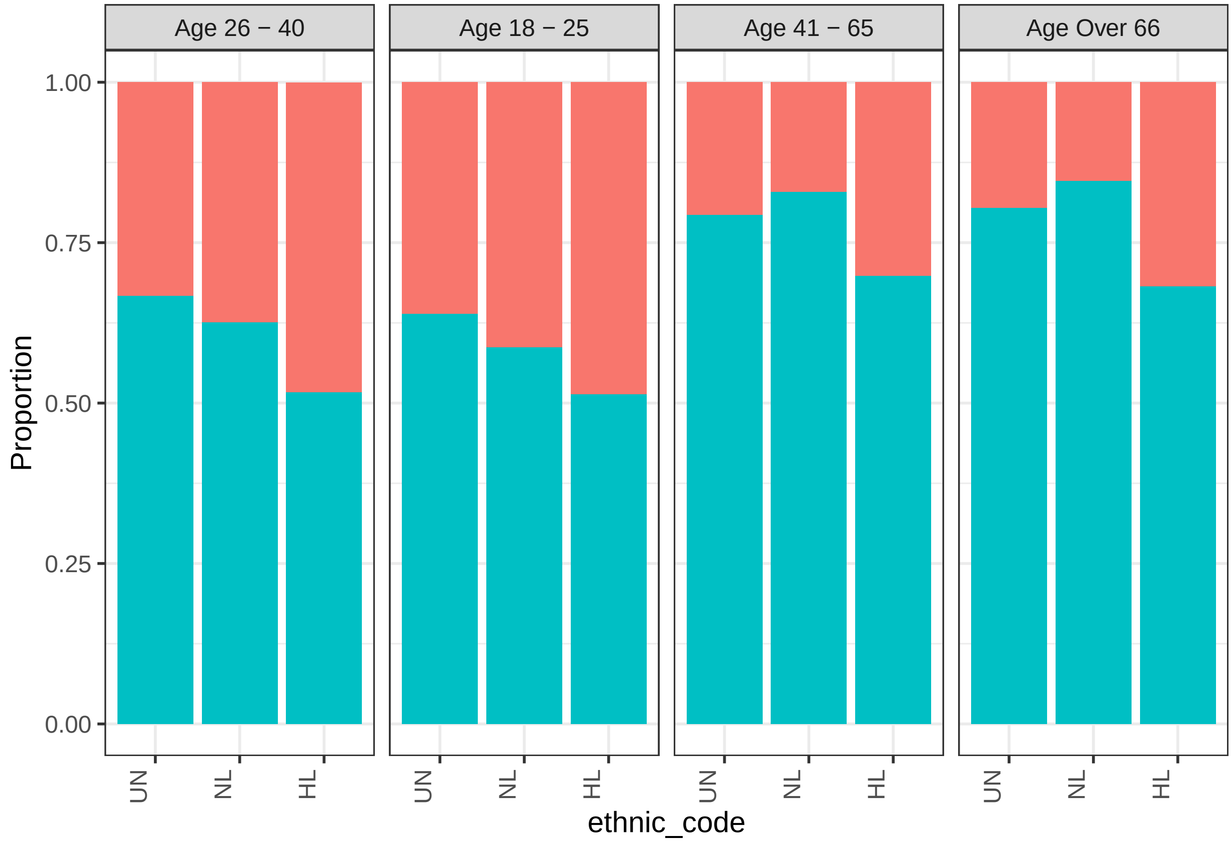

UN
NL
HL
Proportion
1.00
0.75
0.50
0.25
0.00
F
M
U
sex_code


Age 26 − 40
Age 18 − 25
Age 41 − 65
Age Over 66
Proportion
1.00
0.75
0.50
0.25
0.00
F
M
U
sex_code

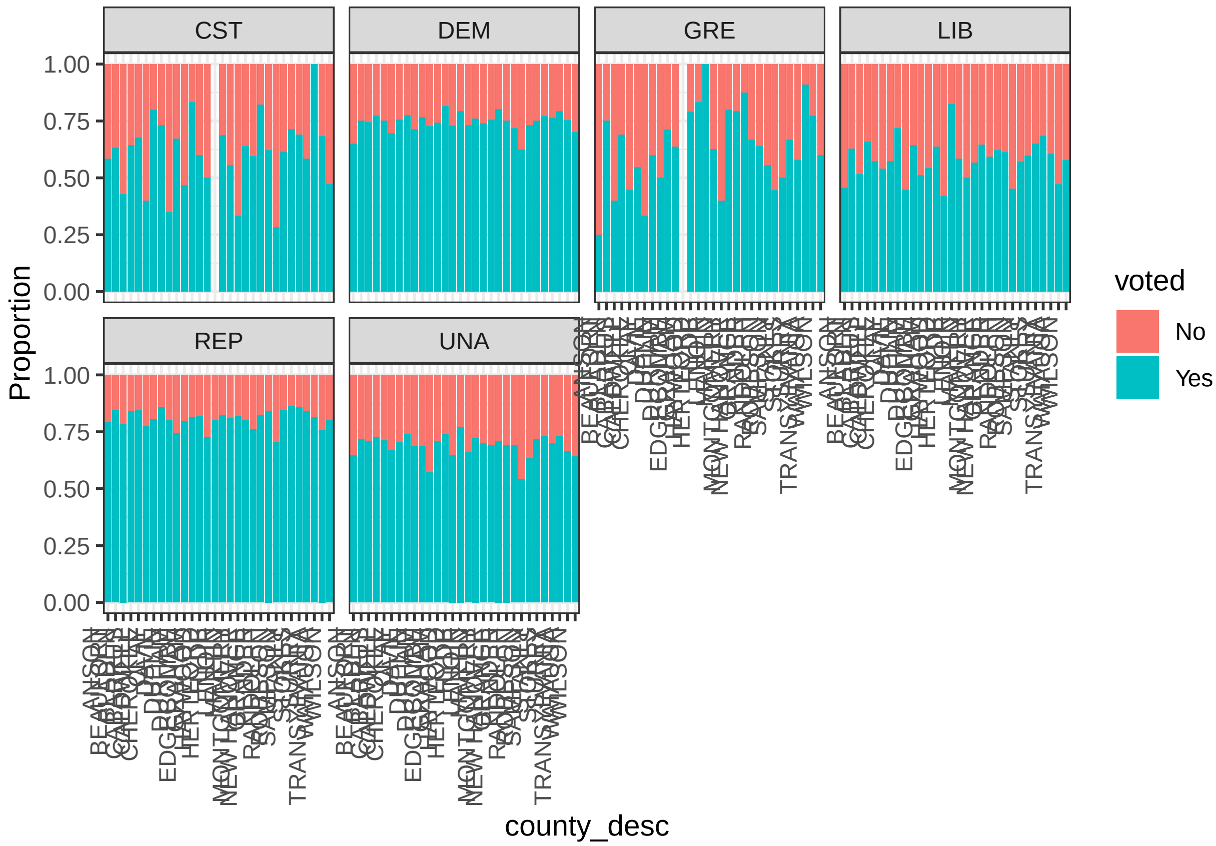

A
B
O
U
W
I
M
P
Proportion
1.00
0.75
0.50
0.25
0.00
CST
DEM
GRE
LIB
REP
UNA
voted_party_cd

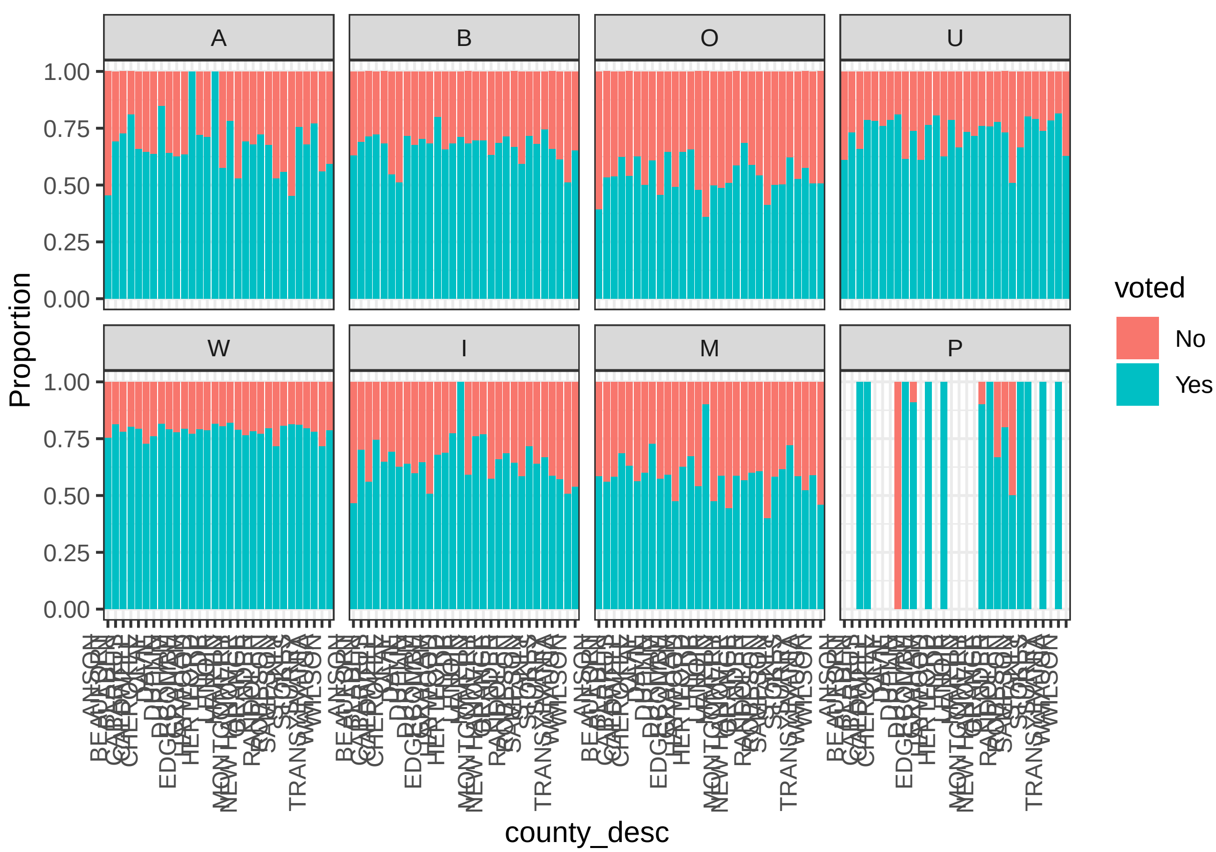

UN
NL
HL
Proportion
1.00
0.75
0.50
0.25
0.00
CST
DEM
GRE
LIB
REP
UNA
voted_party_cd

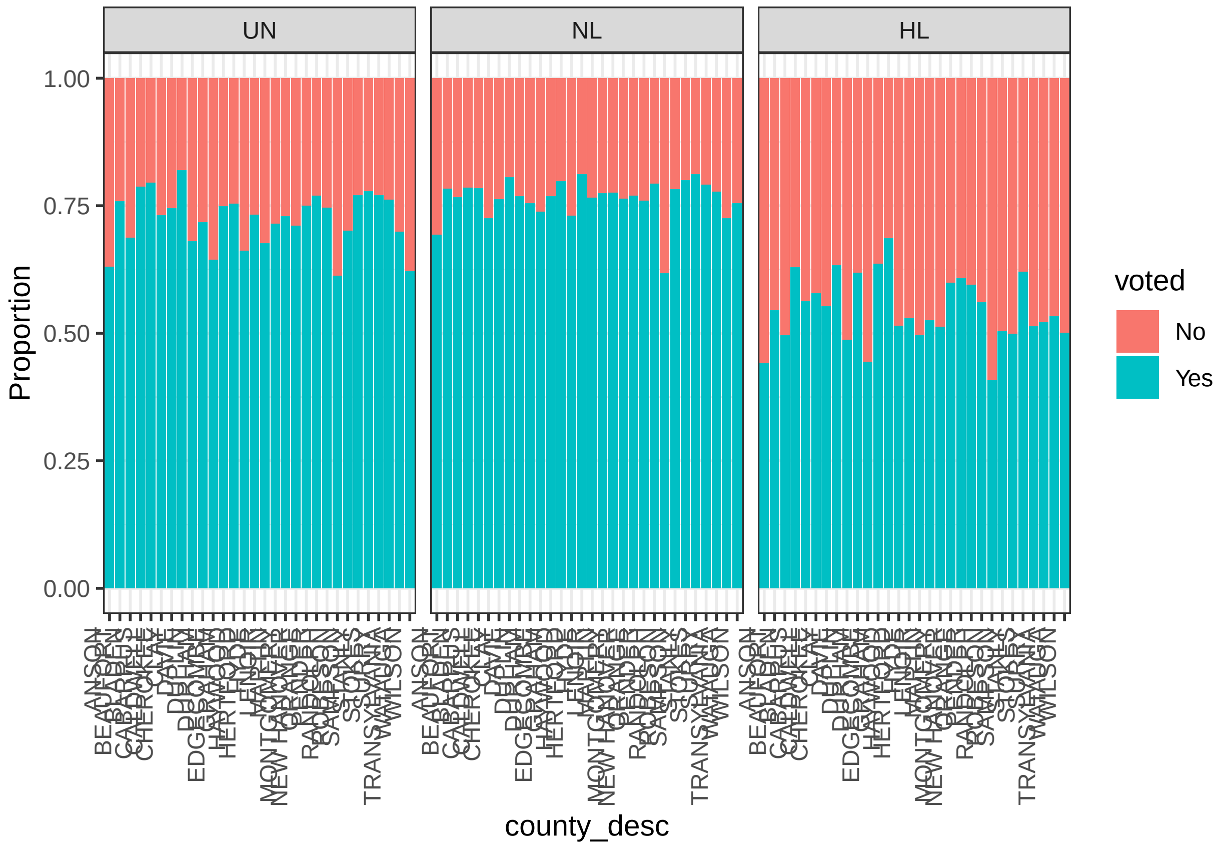

A
B
O
U
W
I
M
P
Proportion
1.00
0.75
0.50
0.25
0.00
UN
NL
HL
ethnic_code



Random slopes


CST
DEM
GRE
LIB
REP
UNA
1.00
0.75
0.50
0.25
0.00
Proportion
county_desc
voted
No
Yes


A
B
O
U
W
I
M
P
1.00
0.75
0.50
0.25
0.00
Proportion
county_desc
voted
No
Yes


UN
NL
HL
1.00
0.75
0.50
0.25
0.00
Proportion
county_desc
voted
No
Yes

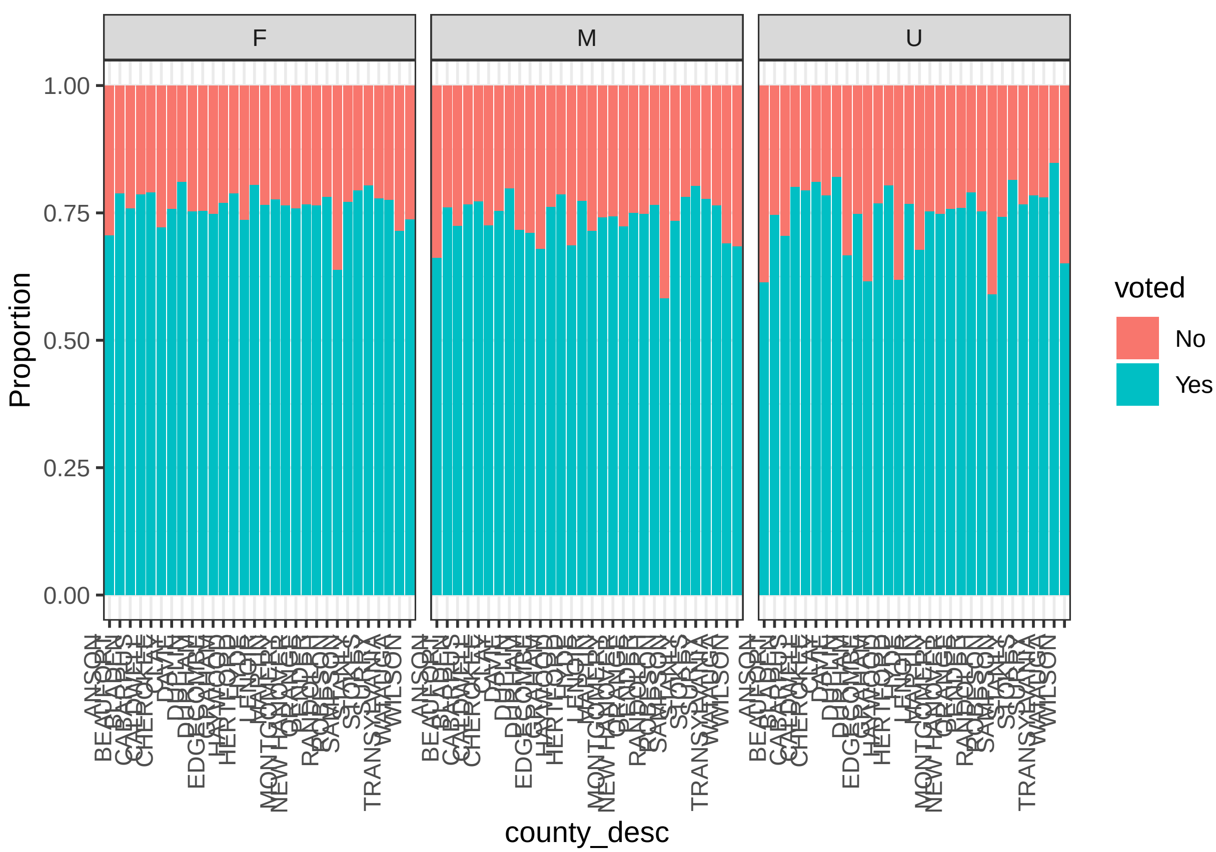

F
M
U
1.00
0.75
0.50
0.25
0.00
Proportion
county_desc
voted
No
Yes

Table of Main Effects

Table 1: Fixed effect estimates on log odds scale

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.0435	0.1743	-0.2495	0.8030
sex_codeM	0.3525	0.1654	2.1313	0.0331
sex_codeU	0.6595	0.2073	3.1816	0.0015
ageAge 18 - 25	-0.8270	0.1793	-4.6129	0.0000
ageAge 41 - 65	-0.0588	0.1650	-0.3566	0.7214
ageAge Over 66	0.2218	0.3844	0.5770	0.5639
voted_party_cdDEM	0.3396	0.1710	1.9858	0.0471
voted_party_cdGRE	0.3080	0.2341	1.3159	0.1882
voted_party_cdLIB	-0.2506	0.1766	-1.4188	0.1560
voted_party_cdREP	0.4798	0.1711	2.8035	0.0051
voted_party_cdUNA	0.0218	0.1710	0.1276	0.8984
race_codeB	-0.0810	0.0162	-4.9888	0.0000
race_codeO	-0.0869	0.0202	-4.3059	0.0000
race_codeU	0.3831	0.0186	20.6408	0.0000
race_codeW	0.3852	0.0158	24.3504	0.0000
race_codeI	-0.0855	0.0213	-4.0153	0.0001
race_codeM	-0.0594	0.0261	-2.2804	0.0226
race_codeP	1.2265	0.3377	3.6321	0.0003
ethnic_codeNL	0.1306	0.0050	26.0137	0.0000
ethnic_codeHL	-0.3007	0.0125	-24.0112	0.0000
sex_codeM:voted_party_cdDEM	-0.6178	0.1655	-3.7328	0.0002
sex_codeU:voted_party_cdDEM	-0.3260	0.2076	-1.5707	0.1163
sex_codeM:voted_party_cdGRE	-0.3006	0.2481	-1.2115	0.2257
sex_codeU:voted_party_cdGRE	-0.3596	0.2924	-1.2299	0.2188
sex_codeM:voted_party_cdLIB	-0.2985	0.1722	-1.7332	0.0831
sex_codeU:voted_party_cdLIB	0.2724	0.2200	1.2385	0.2155
sex_codeM:voted_party_cdREP	-0.3971	0.1656	-2.3983	0.0165
sex_codeU:voted_party_cdREP	-0.2037	0.2078	-0.9807	0.3267
sex_codeM:voted_party_cdUNA	-0.4604	0.1655	-2.7819	0.0054
sex_codeU:voted_party_cdUNA	-0.5352	0.2073	-2.5817	0.0098
ageAge 18 - 25:voted_party_cdDEM	0.8138	0.1795	4.5329	0.0000
ageAge 41 - 65:voted_party_cdDEM	1.0538	0.1652	6.3793	0.0000
ageAge Over 66:voted_party_cdDEM	0.8103	0.3845	2.1072	0.0351
ageAge 18 - 25:voted_party_cdGRE	0.6935	0.2559	2.7097	0.0067
ageAge 41 - 65:voted_party_cdGRE	0.4282	0.2736	1.5654	0.1175
ageAge Over 66:voted_party_cdGRE	0.2036	0.6224	0.3271	0.7436
ageAge 18 - 25:voted_party_cdLIB	0.6323	0.1875	3.3724	0.0007
ageAge 41 - 65:voted_party_cdLIB	0.6033	0.1742	3.4627	0.0005
ageAge Over 66:voted_party_cdLIB	0.6748	0.4021	1.6781	0.0933
ageAge 18 - 25:voted_party_cdREP	0.6733	0.1797	3.7466	0.0002
ageAge 41 - 65:voted_party_cdREP	0.9502	0.1653	5.7484	0.0000
ageAge Over 66:voted_party_cdREP	0.6438	0.3846	1.6739	0.0941
ageAge 18 - 25:voted_party_cdUNA	0.6588	0.1795	3.6707	0.0002
ageAge 41 - 65:voted_party_cdUNA	0.9620	0.1652	5.8238	0.0000
ageAge Over 66:voted_party_cdUNA	0.9987	0.3846	2.5969	0.0094

Bayesian Modeling

Table of Main Effects

Table 2: Estimated posterior parameters

	Est	Lwr	Upr
Intercept	-0.040	-0.367	0.282
sex_codeM	0.357	0.044	0.654
sex_codeU	0.669	0.262	1.082
ageAge18M25	-0.843	-1.206	-0.488
ageAge41M65	-0.056	-0.389	0.280
ageAgeOver66	0.277	-0.498	1.067
voted_party_cdDEM	0.338	0.020	0.657
voted_party_cdGRE	0.310	-0.148	0.738
voted_party_cdLIB	-0.251	-0.590	0.086
voted_party_cdREP	0.479	0.165	0.798
voted_party_cdUNA	0.021	-0.294	0.340
race_codeB	-0.082	-0.111	-0.049
race_codeO	-0.088	-0.127	-0.049
race_codeU	0.382	0.344	0.418
race_codeW	0.384	0.355	0.416
race_codeI	-0.087	-0.127	-0.045
race_codeM	-0.061	-0.112	-0.010
race_codeP	1.270	0.643	2.008
ethnic_codeNL	0.131	0.120	0.141
ethnic_codeHL	-0.301	-0.328	-0.274
sex_codeM:voted_party_cdDEM	-0.622	-0.924	-0.311
sex_codeU:voted_party_cdDEM	-0.336	-0.740	0.075
sex_codeM:voted_party_cdGRE	-0.312	-0.805	0.157
sex_codeU:voted_party_cdGRE	-0.366	-0.935	0.220
sex_codeM:voted_party_cdLIB	-0.302	-0.616	0.026
sex_codeU:voted_party_cdLIB	0.262	-0.175	0.685
sex_codeM:voted_party_cdREP	-0.402	-0.699	-0.091
sex_codeU:voted_party_cdREP	-0.213	-0.631	0.190
sex_codeM:voted_party_cdUNA	-0.465	-0.764	-0.150
sex_codeU:voted_party_cdUNA	-0.545	-0.944	-0.136
ageAge18M25:voted_party_cdDEM	0.830	0.476	1.196
ageAge41M65:voted_party_cdDEM	1.050	0.712	1.385
ageAgeOver66:voted_party_cdDEM	0.755	-0.044	1.527
ageAge18M25:voted_party_cdGRE	0.715	0.200	1.208
ageAge41M65:voted_party_cdGRE	0.436	-0.098	0.969
ageAgeOver66:voted_party_cdGRE	0.210	-1.013	1.494
ageAge18M25:voted_party_cdLIB	0.648	0.288	1.020
ageAge41M65:voted_party_cdLIB	0.599	0.242	0.954
ageAgeOver66:voted_party_cdLIB	0.625	-0.150	1.449
ageAge18M25:voted_party_cdREP	0.690	0.340	1.050
ageAge41M65:voted_party_cdREP	0.947	0.614	1.284
ageAgeOver66:voted_party_cdREP	0.589	-0.224	1.353
ageAge18M25:voted_party_cdUNA	0.675	0.323	1.042
ageAge41M65:voted_party_cdUNA	0.959	0.622	1.292
ageAgeOver66:voted_party_cdUNA	0.944	0.133	1.727