Case Study 2

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

10/28/2021

Explication of abbreviations: https://s3.amazonaws.com/dl.ncsbe.gov/data/layout_ncvoter.txt

```
# contains info about the aggregate counts of voters who actually voted by demographic variables

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

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

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

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

# remove above three columns

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

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

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

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

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

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

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

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

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

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

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

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

EDA

```r
table(data2$vote_or_not)
```

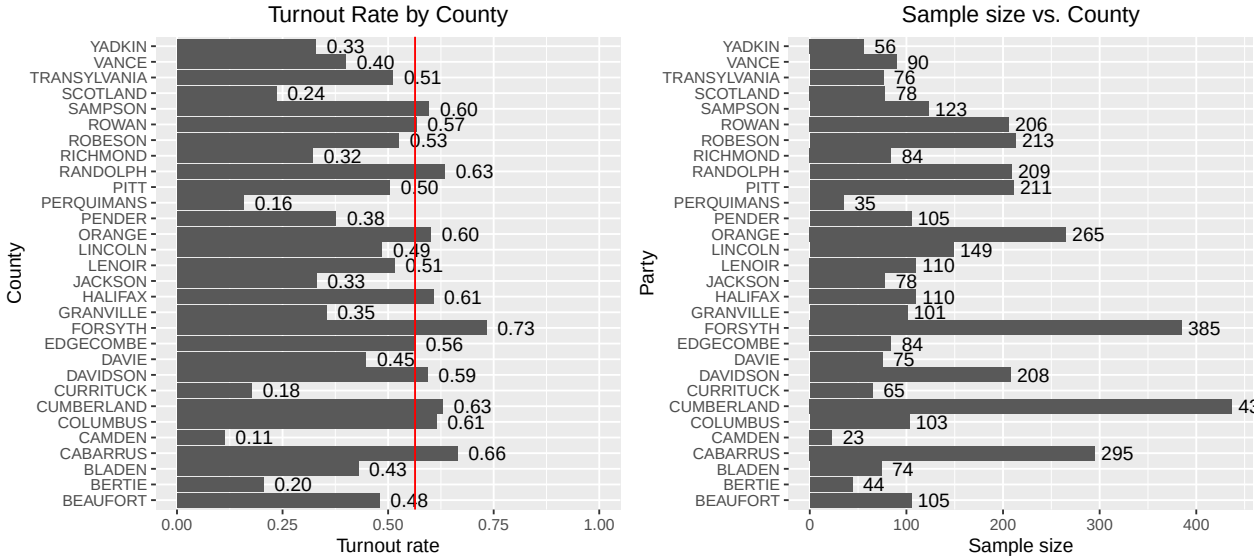
```
##
##     0     1
## 31888 41212
```

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

```
##   group turnout_rate
## 1 Total    0.5637756
```

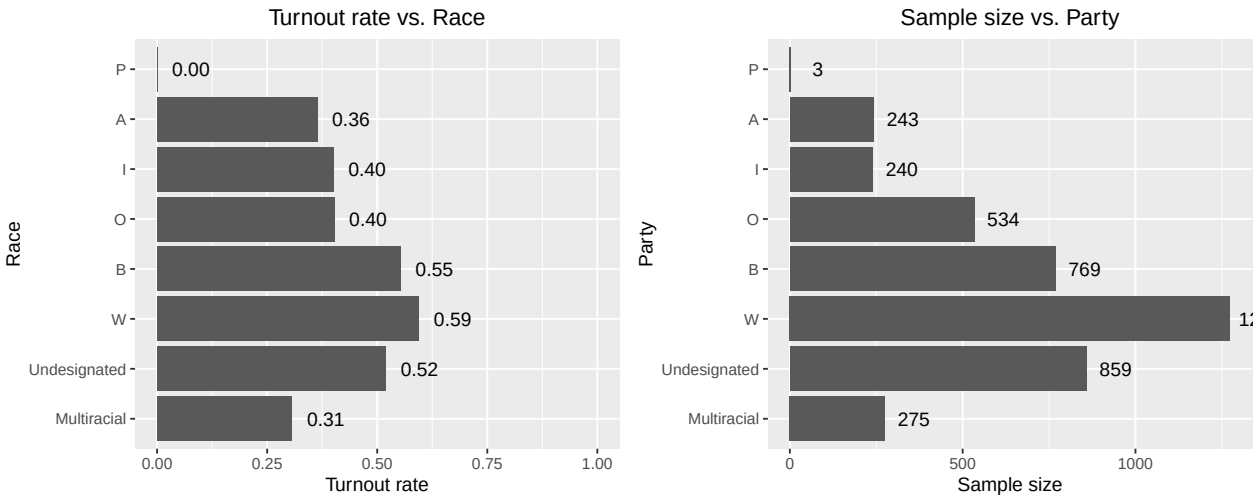
turnout rate

```
p1 <- data %>%
  group_by(county_desc) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
  ggplot( aes(x = county_desc, y = turnout_rate)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("County") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.3) +
  ggtitle("Turnout Rate by County") +
    theme(plot.title = element_text(hjust = 0.5)) +
  geom_hline(yintercept = turnout_rate, color = 'red') +
  coord_flip()

p2 <- data %>%
  group_by(county_desc) %>%
  summarize(n = n(), .groups="drop") %>%
  ggplot( aes(x = county_desc, y = n)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Sample size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.2) +
  ggtitle("Sample size vs. County") +
    theme(plot.title = element_text(hjust = 0.5)) +
  coord_flip()

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



race

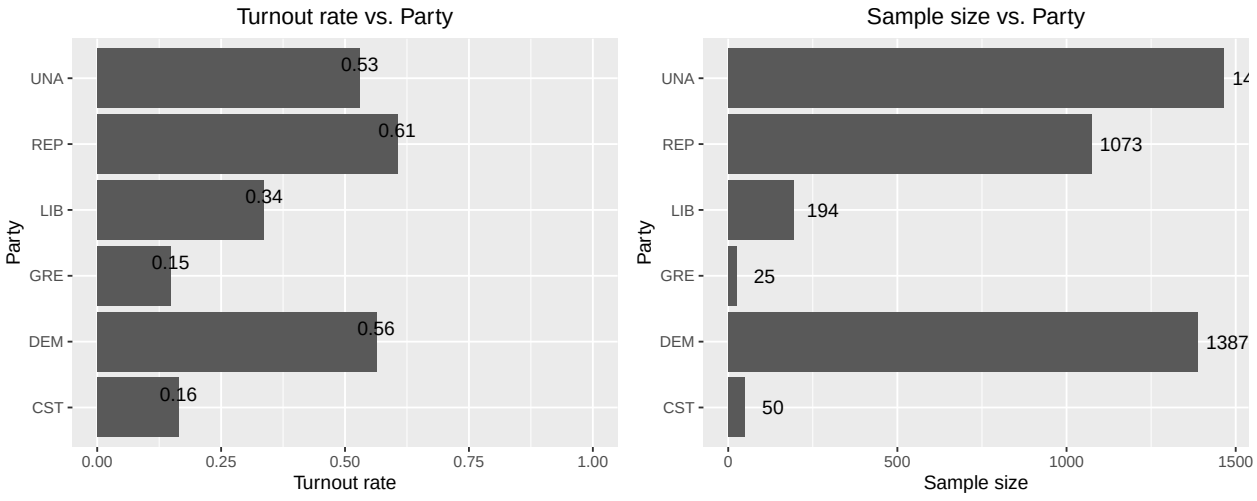
```
p1 <- data %>%
  group_by(race_code) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
  select(group = race_code, turnout_rate) %>%
  ggplot( aes(x = group, y = turnout_rate)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Race") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.4) +
  ggtitle("Turnout rate vs. Race") +
    theme(plot.title = element_text(hjust = 0.5)) +
    coord_flip()



p2 <- data %>%
  group_by(race_code) %>%
  summarize(n = n(), .groups="drop") %>%
  ggplot( aes(x = race_code, y = n)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Sample size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.2) +
  ggtitle("Sample size vs. Party") +
    theme(plot.title = element_text(hjust = 0.5)) +
  coord_flip()

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




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

party

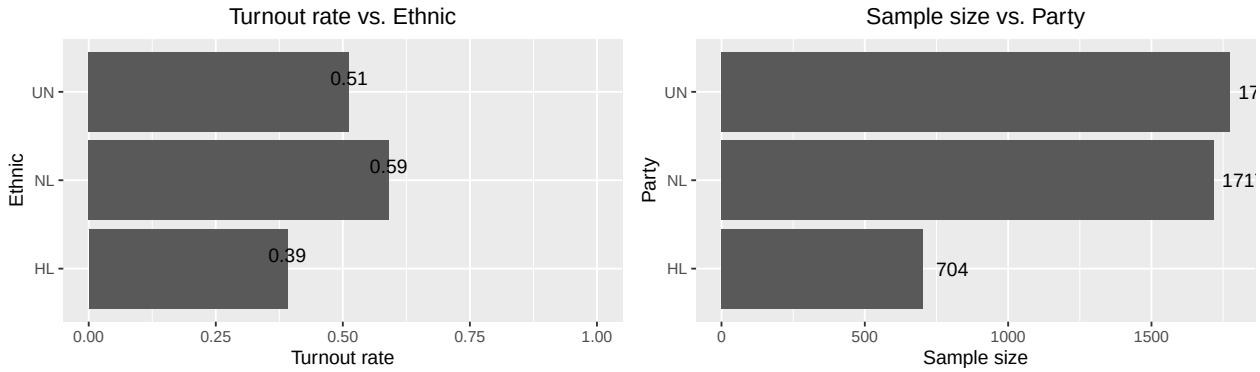
```
p1 <- data %>%
  group_by(party_cd) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
  select(group = party_cd, turnout_rate) %>%
  ggplot( aes(x = group, y = turnout_rate)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Party") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), vjust = -0.5) +
  ggtitle("Turnout rate vs. Party") +
    theme(plot.title = element_text(hjust = 0.5))  +
    coord_flip()

p2 <- data %>%
  group_by(party_cd) %>%
  summarize(n = n(), .groups="drop") %>%
  ggplot( aes(x = party_cd, y = n)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Sample size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.2) +
  ggtitle("Sample size vs. Party") +
    theme(plot.title = element_text(hjust = 0.5)) +
  coord_flip()

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




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

ethnic groups

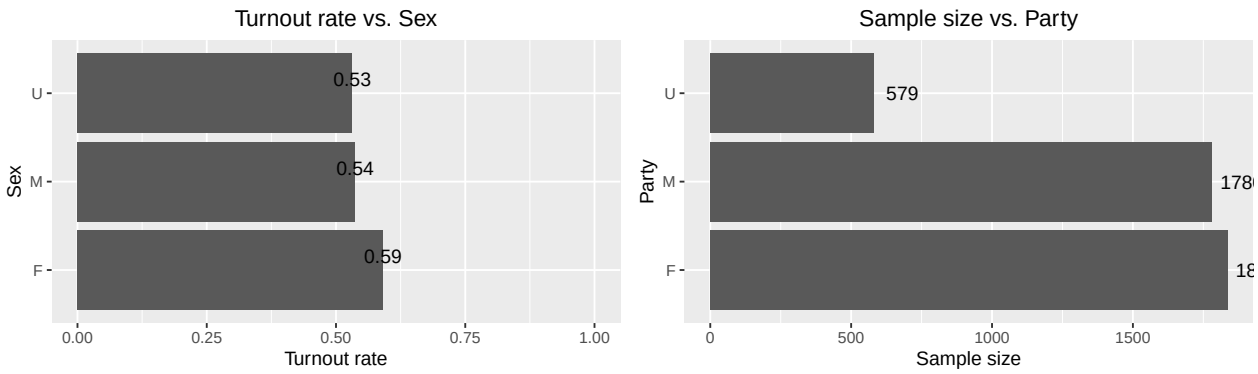
```
p1 <- data %>%
  group_by(ethnic_code) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
  select(group = ethnic_code, turnout_rate) %>%
  ggplot( aes(x = group, y = turnout_rate)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Ethnic") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), vjust = -0.5) +
  ggtitle("Turnout rate vs. Ethnic") +
    theme(plot.title = element_text(hjust = 0.5))+
    coord_flip()

p2 <- data %>%
  group_by(ethnic_code) %>%
  summarize(n = n(), .groups="drop") %>%
  ggplot( aes(x = ethnic_code, y = n)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Sample size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.2) +
  ggtitle("Sample size vs. Party") +
    theme(plot.title = element_text(hjust = 0.5)) +
  coord_flip()

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



Sex

```
p1 <- data %>%
  group_by(sex_code) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
  select(group = sex_code, turnout_rate) %>%
```

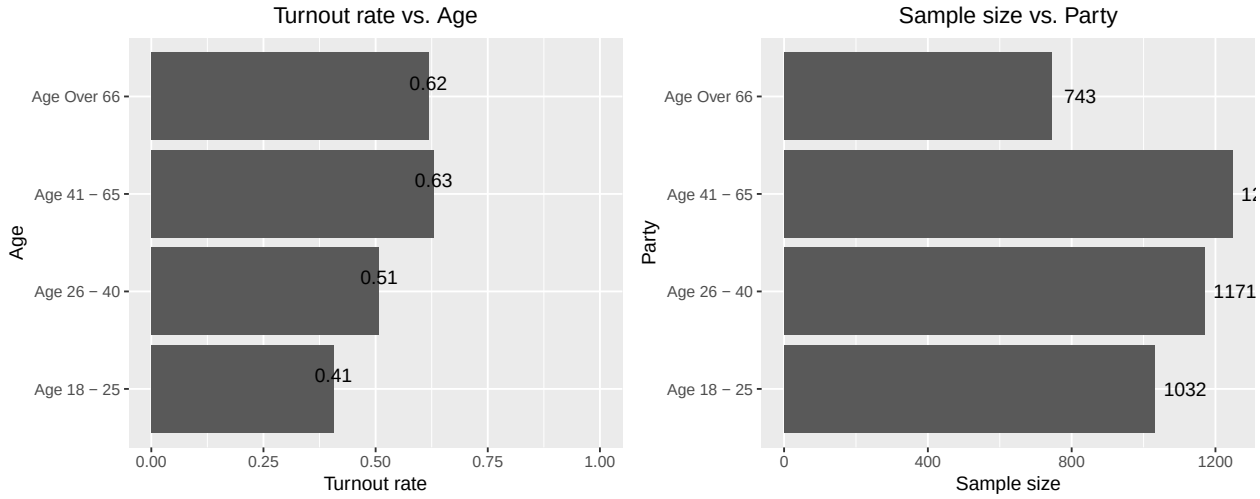
```
  ggplot( aes(x = group, y = turnout_rate)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Sex") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), vjust = -0.5) +
  ggtitle("Turnout rate vs. Sex") +
    theme(plot.title = element_text(hjust = 0.5)) +
    coord_flip()

p2 <- data %>%
  group_by(sex_code) %>%
  summarize(n = n(), .groups="drop") %>%
  ggplot( aes(x = sex_code, y = n)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Sample size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.2) +
  ggtitle("Sample size vs. Party") +
    theme(plot.title = element_text(hjust = 0.5)) +
  coord_flip()

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



age

```
p1 <- data %>%
  group_by(age) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
  select(group = age, turnout_rate) %>%
  ggplot( aes(x = group, y = turnout_rate)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Age") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), vjust = -0.5) +
  ggtitle("Turnout rate vs. Age") +
    theme(plot.title = element_text(hjust = 0.5))+
    coord_flip()
```

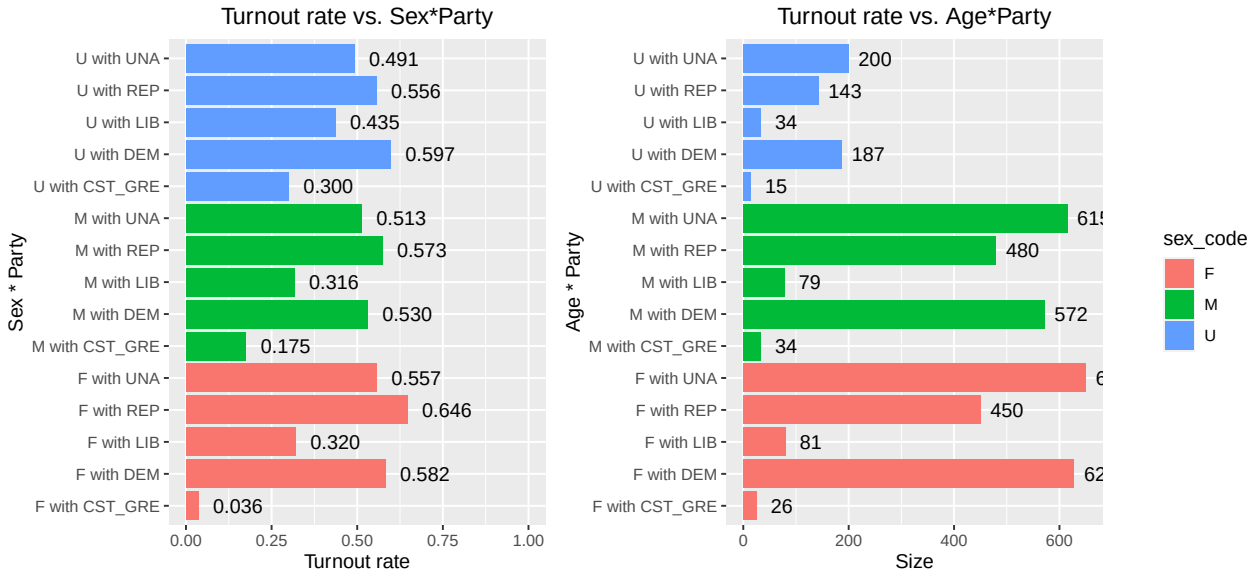
```
p2 <- data %>%
  group_by(age) %>%
  summarize(n = n(), .groups="drop") %>%
  ggplot( aes(x = age, y = n)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Sample size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.2) +
  ggtitle("Sample size vs. Party") +
    theme(plot.title = element_text(hjust = 0.5)) +
  coord_flip()

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



sex & party

```
p1 <- data %>%
  group_by(sex_code, party_cd) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg,
         group = paste0(sex_code, " with ", party_cd)) %>%
  select(group = group, turnout_rate, sex_code, party_cd) %>%
  ggplot( aes(x = group, y = turnout_rate, fill = sex_code)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Sex * Party") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.3) +
  coord_flip() +
  ggtitle("Turnout rate vs. Sex*Party") +
    theme(plot.title = element_text(hjust = 0.5))

p2 <- data %>%
  group_by(sex_code, party_cd) %>%
  summarize(n = n(), .groups = "drop") %>%
  mutate(group = paste0(sex_code, " with ", party_cd)) %>%
```

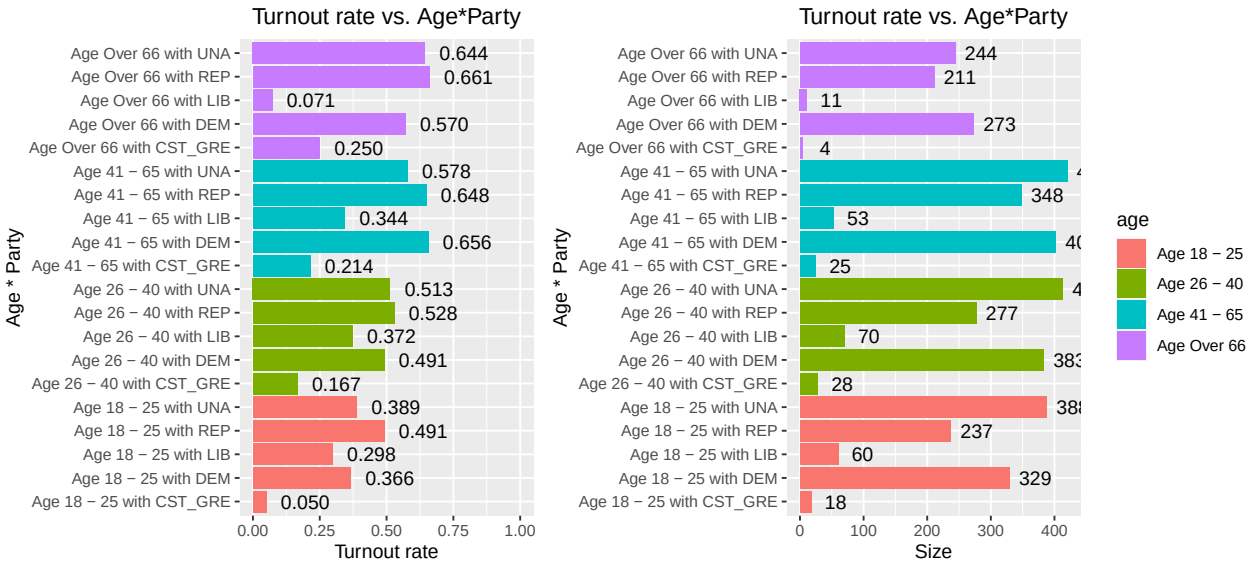
```
  ggplot( aes(x = group, y = n, fill = sex_code)) +
  geom_bar(stat = "identity") +
  xlab("Age * Party") + ylab("Size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.3) +
  coord_flip() +
  ggtitle("Turnout rate vs. Age*Party") +
    theme(plot.title = element_text(hjust = 0.5))

legend <- get_legend(p1)

grid.arrange(widths = c(3, 3, 1),p1 + theme(legend.position="none"),
             p2 + theme(legend.position="none"), nrow= 1,legend)
```



age & party

```
p1 <- data %>%
  group_by(age, party_cd) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg,
         group = paste0(age, " with ", party_cd)) %>%
  select(group = group, turnout_rate, age, party_cd) %>%
  ggplot( aes(x = group, y = turnout_rate, fill = age)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Age * Party") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.3) +
  coord_flip() +
  ggtitle("Turnout rate vs. Age*Party") +
    theme(plot.title = element_text(hjust = 0.5))

p2 <- data %>%
```

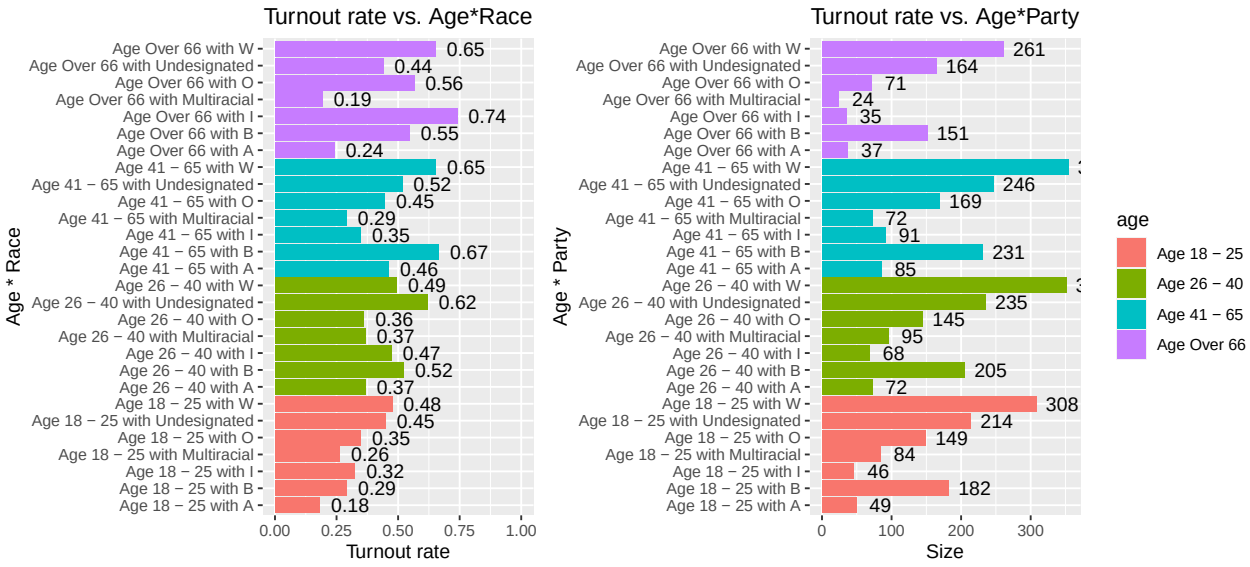
```
  group_by(age, party_cd) %>%
  summarize(n = n(), .groups = "drop") %>%
  mutate(group = paste0(age, " with ", party_cd)) %>%
  ggplot( aes(x = group, y = n, fill = age)) +
  geom_bar(stat = "identity") +
  xlab("Age * Party") + ylab("Size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.3) +
  coord_flip() +
  ggtitle("Turnout rate vs. Age*Party") +
    theme(plot.title = element_text(hjust = 0.5))

legend <- get_legend(p1)

grid.arrange(widths = c(3, 3, 1),p1 + theme(legend.position="none"),
             p2 + theme(legend.position="none"), nrow= 1,legend)
```



age & race

```
p1 <- data %>%
  group_by(age, race_code) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg,
         group = paste0(age, " with ", race_code)) %>%
  select(group = group, turnout_rate, age, race_code) %>%
  ggplot( aes(x = group, y = turnout_rate, fill = age)) +
  geom_bar(stat = "identity") +
  ylim(0,1) +
  xlab("Age * Race") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2)), hjust = -0.3) +
  coord_flip() +
  ggtitle("Turnout rate vs. Age*Race") +
    theme(plot.title = element_text(hjust = 0.5))
```

```
p2 <- data %>%
  group_by(age, race_code) %>%
  summarize(n = n(), .groups = "drop") %>%
  mutate(group = paste0(age, " with ", race_code)) %>%
  ggplot( aes(x = group, y = n, fill = age)) +
  geom_bar(stat = "identity") +
  xlab("Age * Party") + ylab("Size") +
  geom_text(aes(label=format(n, digits = 2)), hjust = -0.3) +
  coord_flip() +
  ggtitle("Turnout rate vs. Age*Party") +
    theme(plot.title = element_text(hjust = 0.5))

legend <- get_legend(p1)

grid.arrange(widths = c(3, 3, 1),p1 + theme(legend.position="none"),
             p2 + theme(legend.position="none"), nrow= 1,legend)
```




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

Model

```
# initial
mod1 <- glmer(cbind(total_vot, total_reg - total_vot) ~
```

```
                      party_cd  + sex_code + age + (1|county_desc),
                    data = data, family = binomial,
              control=glmerControl(optimizer = "bobyqa"))


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


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


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


mod5 <- glmer(cbind(total_vot, total_reg - total_vot) ~
                  party_cd + race_code + ethnic_code + sex_code + age +
                  sex_code:party_cd + age:party_cd + (1|county_desc),
                data = data, family = binomial,
              control=glmerControl(optimizer = "bobyqa", optCtrl=list(maxfun=2e5)))


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

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

Table 1: Forward model selection

Model	LRT.p.value	AIC	BIC
Base model		41563.94	41633.70
Add race	0	40816.85	40924.66
Add race and ethnic	0	40495.44	40615.93
Without intercept	0	40617.17	40712.29
Add the interaction of sex and party_cd, and age and party	0	40051.42	40298.74
Add the interaction of sex and party_cd, age and party, and age and race	0	39115.44	39476.90

```
summary(mod6)$coefficients
```

```
##                                       Estimate Std. Error     z value
## (Intercept)                        -4.10770560 1.03678123  -3.9619791
## party_cdDEM                         2.53213509 1.01359983   2.4981605
## party_cdLIB                         1.84815777 1.05096566   1.7585330
## party_cdREP                         2.62985205 1.01357975   2.5946178
## party_cdUNA                         2.42948949 1.01313454   2.3979929
## race_codeUndesignated               0.75693724 0.18950720   3.9942401
## race_codeW                          0.92827433 0.18420543   5.0393428
## race_codeB                         -0.06776191 0.18586623  -0.3645735
## race_codeO                          0.55663523 0.20512495   2.7136398
## race_codeI                          0.19964916 0.22622135   0.8825390
## race_codeA                         -0.77203297 0.28168096  -2.7408064
## ethnic_codeHL                      -0.71634343 0.05637892 -12.7058743
## ethnic_codeUN                      -0.34210175 0.02179287 -15.6978778
## sex_codeF                          -1.57591039 1.09262512  -1.4423157
## sex_codeU                           0.89680258 0.73414766   1.2215561
## ageAge 41 - 65                      1.01422728 1.10640276   0.9166890
## ageAge Over 66                      1.23935762 1.60972360   0.7699195
## ageAge 26 - 40                      0.88207559 1.06869135   0.8253792
## party_cdDEM:sex_codeF               1.80057334 1.09299097   1.6473817
## party_cdLIB:sex_codeF               1.65678848 1.12124291   1.4776356
## party_cdREP:sex_codeF               1.76578356 1.09302989   1.6154943
## party_cdUNA:sex_codeF               1.70028935 1.09304551   1.5555522
## party_cdDEM:sex_codeU              -0.36357836 0.73556084  -0.4942873
## party_cdLIB:sex_codeU              -0.22506626 0.80290481  -0.2803150
## party_cdREP:sex_codeU              -0.65764205 0.73487784  -0.8948998
## party_cdUNA:sex_codeU              -0.88628936 0.73420140  -1.2071475
## party_cdDEM:ageAge 41 - 65         -0.98003377 1.07548590  -0.9112474
## party_cdLIB:ageAge 41 - 65         -1.75962534 1.12154685  -1.5689272
## party_cdREP:ageAge 41 - 65         -1.27722107 1.07548598  -1.1875758
## party_cdUNA:ageAge 41 - 65         -1.15928291 1.07494611  -1.0784568
## party_cdDEM:ageAge Over 66         -2.02686216 1.52277614  -1.3310309
## party_cdLIB:ageAge Over 66         -4.22538693 1.88115399  -2.2461675
## party_cdREP:ageAge Over 66         -1.90878692 1.52304574  -1.2532696
## party_cdUNA:ageAge Over 66         -1.31377714 1.52279866  -0.8627386
## party_cdDEM:ageAge 26 - 40         -0.76259857 1.04724154  -0.7281974
## party_cdLIB:ageAge 26 - 40         -0.51396081 1.08313799  -0.4745109
## party_cdREP:ageAge 26 - 40         -0.44149382 1.04735012  -0.4215341
## party_cdUNA:ageAge 26 - 40         -0.35144977 1.04660048  -0.3358013
## race_codeUndesignated:ageAge 41 - 65  0.50539581 0.28385065   1.7804990
```

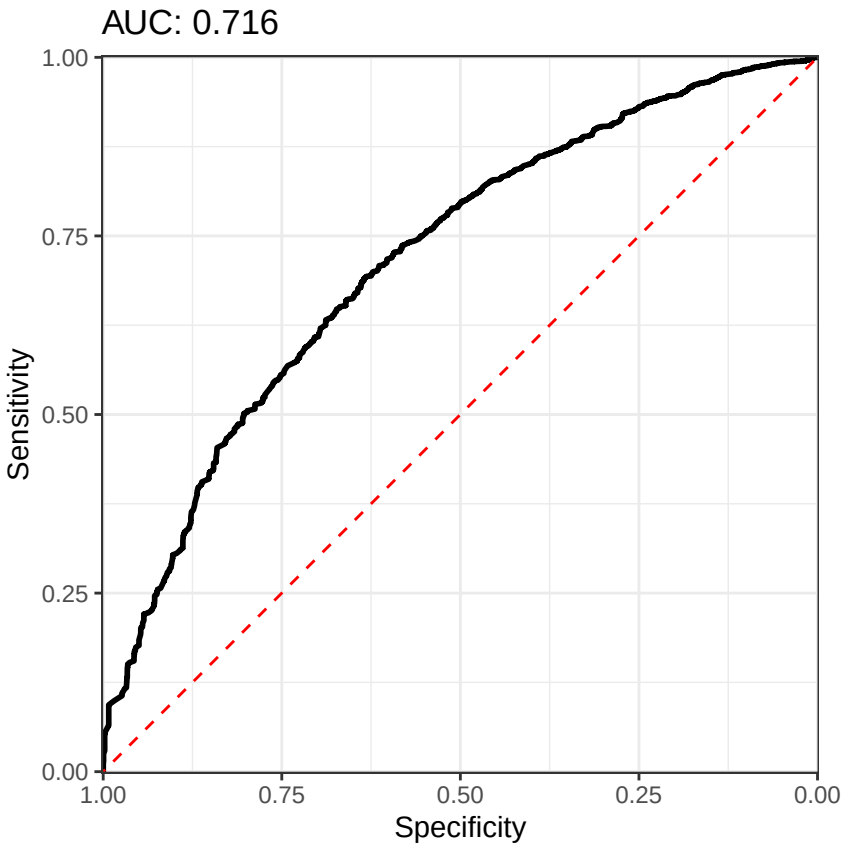
```
## race_codeW:ageAge 41 - 65             0.93851921 0.27986724   3.3534443
## race_codeB:ageAge 41 - 65             1.84728091 0.28131760   6.5665316
## race_codeO:ageAge 41 - 65             0.39488170 0.30551933   1.2924933
## race_codeI:ageAge 41 - 65             0.06010935 0.31629017   0.1900449
## race_codeA:ageAge 41 - 65             1.52298171 0.36870341   4.1306418
## race_codeUndesignated:ageAge Over 66  0.67357834 0.54479264   1.2363940
## race_codeW:ageAge Over 66             1.32805781 0.53707170   2.4727756
## race_codeB:ageAge Over 66             2.01324345 0.53893772   3.7355772
## race_codeO:ageAge Over 66             1.19008976 0.57100131   2.0842154
## race_codeI:ageAge Over 66             2.41249166 0.59131680   4.0798632
## race_codeA:ageAge Over 66             0.86591289 0.65674597   1.3184898
## race_codeUndesignated:ageAge 26 - 40  0.58623961 0.24575697   2.3854445
## race_codeW:ageAge 26 - 40            -0.39687347 0.24006271  -1.6532075
## race_codeB:ageAge 26 - 40             0.78774576 0.24270860   3.2456442
## race_codeO:ageAge 26 - 40            -0.33055305 0.27117590  -1.2189618
## race_codeI:ageAge 26 - 40             0.33277279 0.30019617   1.1085178
## race_codeA:ageAge 26 - 40             0.58063143 0.35115377   1.6534962
##                                          Pr(>|z|)
## (Intercept)                           7.433103e-05
## party_cdDEM                           1.248396e-02
## party_cdLIB                           7.865686e-02
## party_cdREP                           9.469616e-03
## party_cdUNA                           1.648518e-02
## race_codeUndesignated                 6.490207e-05
## race_codeW                            4.671331e-07
## race_codeB                            7.154298e-01
## race_codeO                            6.654848e-03
## race_codeI                            3.774854e-01
## race_codeA                            6.128861e-03
## ethnic_codeHL                         5.485295e-37
## ethnic_codeUN                         1.563867e-55
## sex_codeF                             1.492133e-01
## sex_codeU                             2.218755e-01
## ageAge 41 - 65                        3.593056e-01
## ageAge Over 66                        4.413476e-01
## ageAge 26 - 40                        4.091564e-01
## party_cdDEM:sex_codeF                 9.947961e-02
## party_cdLIB:sex_codeF                 1.395053e-01
## party_cdREP:sex_codeF                 1.062037e-01
## party_cdUNA:sex_codeF                 1.198146e-01
## party_cdDEM:sex_codeU                 6.211033e-01
## party_cdLIB:sex_codeU                 7.792358e-01
## party_cdREP:sex_codeU                 3.708406e-01
## party_cdUNA:sex_codeU                 2.273754e-01
## party_cdDEM:ageAge 41 - 65            3.621650e-01
## party_cdLIB:ageAge 41 - 65            1.166649e-01
## party_cdREP:ageAge 41 - 65            2.350006e-01
## party_cdUNA:ageAge 41 - 65            2.808300e-01
## party_cdDEM:ageAge Over 66            1.831788e-01
## party_cdLIB:ageAge Over 66            2.469328e-02
## party_cdREP:ageAge Over 66            2.101076e-01
## party_cdUNA:ageAge Over 66            3.882812e-01
## party_cdDEM:ageAge 26 - 40            4.664928e-01
## party_cdLIB:ageAge 26 - 40            6.351356e-01
```

```
## party_cdREP:ageAge 26 - 40           6.733651e-01
## party_cdUNA:ageAge 26 - 40           7.370207e-01
## race_codeUndesignated:ageAge 41 - 65 7.499433e-02
## race_codeW:ageAge 41 - 65            7.981247e-04
## race_codeB:ageAge 41 - 65            5.150073e-11
## race_codeO:ageAge 41 - 65            1.961863e-01
## race_codeI:ageAge 41 - 65            8.492739e-01
## race_codeA:ageAge 41 - 65            3.617519e-05
## race_codeUndesignated:ageAge Over 66 2.163122e-01
## race_codeW:ageAge Over 66            1.340683e-02
## race_codeB:ageAge Over 66            1.872851e-04
## race_codeO:ageAge Over 66            3.714057e-02
## race_codeI:ageAge Over 66            4.506221e-05
## race_codeA:ageAge Over 66            1.873397e-01
## race_codeUndesignated:ageAge 26 - 40 1.705850e-02
## race_codeW:ageAge 26 - 40            9.828864e-02
## race_codeB:ageAge 26 - 40            1.171852e-03
## race_codeO:ageAge 26 - 40            2.228587e-01
## race_codeI:ageAge 26 - 40            2.676383e-01
## race_codeA:ageAge 26 - 40            9.822991e-02
```

```r
chart_fixef <- function(model) {
  summary(model)$coefficients %>%
    as.data.frame() %>%
    mutate(`95% CI` = paste0(
      '[',
      round(Estimate - 1.96 * `Std. Error`, 3), ', ',
      round(Estimate + 1.96 * `Std. Error`, 3), ']' )
    ) %>%
    select(Estimate, `95% CI`) %>%
    kable(digits=3) %>%
    kable_styling(full_width=FALSE) %>%
    kable_classic()
}
```

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

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

# fit model on train data

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

predicted <- predict(mod6, test, type="response")
```

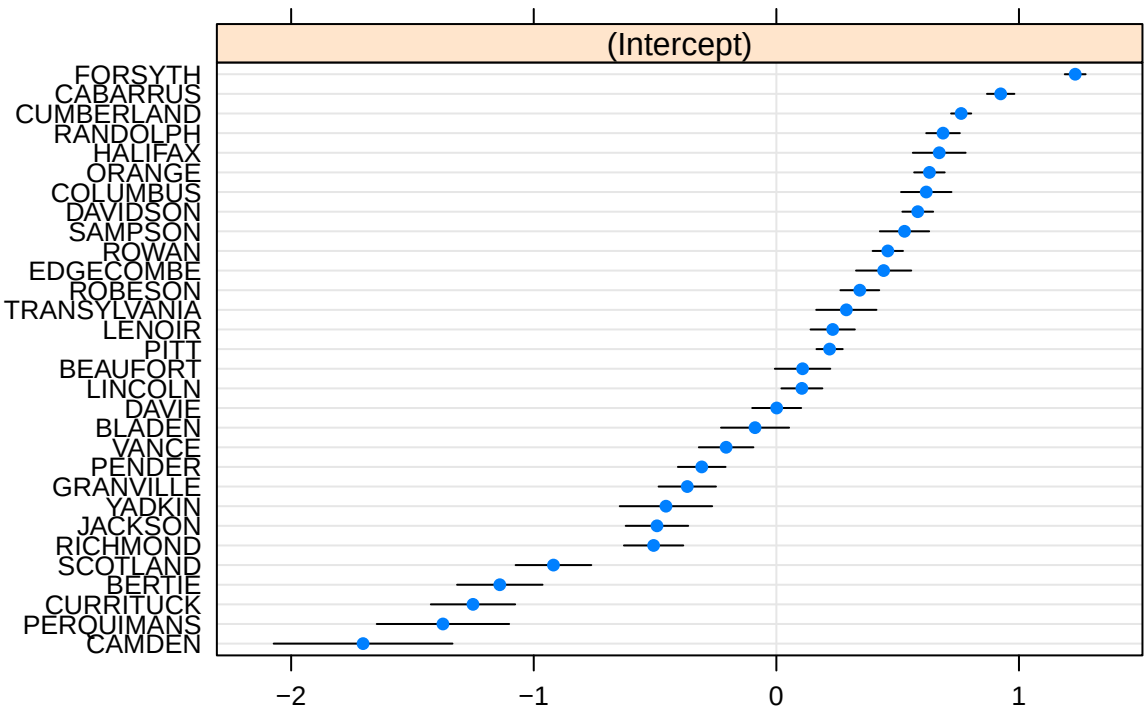
```
rocobj <- roc(test$resp, predicted)
ggroc(rocobj, size=1) +
  coord_fixed() +
  scale_x_reverse(
    name = "Specificity",
    limits = c(1,0),
    expand = c(0.001,0.001)
  ) +
  scale_y_continuous(
    name = "Sensitivity",
    limits = c(0,1),
    expand = c(0.001, 0.001)
  ) +
  geom_abline(intercept=1,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




```
# fit model on full data

mod6 <- glmer(cbind(total_vot, total_reg - total_vot) ~
                1 + party_cd + sex_code:party_cd + (1|county_desc),
              data = data, family = binomial,
              control=glmerControl(optimizer = "bobyqa"))

# Fixed Effects
chart_fixef(mod6)
```

	Estimate	95% CI
(Intercept)	-1.988	[-2.859, -1.117]
party_cdDEM	1.633	[0.801, 2.466]
party_cdLIB	0.667	[-0.224, 1.559]
party_cdREP	1.821	[0.988, 2.653]
party_cdUNA	1.609	[0.777, 2.442]
party_cdCST_GRE:sex_codeF	-1.812	[-3.96, 0.336]
party_cdDEM:sex_codeF	0.238	[0.186, 0.291]
party_cdLIB:sex_codeF	0.043	[-0.433, 0.52]
party_cdREP:sex_codeF	0.243	[0.184, 0.302]
party_cdUNA:sex_codeF	0.179	[0.12, 0.237]
party_cdCST_GRE:sex_codeU	0.507	[-0.776, 1.79]
party_cdDEM:sex_codeU	0.178	[0.06, 0.297]
party_cdLIB:sex_codeU	0.616	[-0.001, 1.233]
party_cdREP:sex_codeU	-0.119	[-0.225, -0.014]
party_cdUNA:sex_codeU	-0.292	[-0.378, -0.206]

```
# Random Effects
dotplot(ranef(mod6))$county_desc
```




```
# test <- binned_residuals(mod6)
```

```
# mod <- brm(
#   data = data, family = binomial,
```

```
#   total_vot | trials(total_reg) ~ 1 + party_cd + (1 | county_desc),
#   prior = c(prior(normal(0, 10), class = Intercept),
#             prior(normal(0, 1), class = b)),
#   iter = 2500, warmup = 500, cores = 4, chains = 2,
#   seed = 10
# )
```

```
# fixef(mod) %>%
#   as.data.frame() %>%
#   mutate(`95% CI` = paste0(
#     "[", round(Q2.5, 3), ",", round(Q97.5, 3), "]"
#   )) %>%
#   select(-c(Est.Error, Q2.5, Q97.5)) %>%
#   kable(digits=3) %>%
#   kable_classic() %>%
#   kable_styling(full_width=FALSE)
#
# ranef(mod)$county_desc %>%
#   as.data.frame() %>%
#
#   tibble::rownames_to_column(var="County") %>%
#   ggplot(aes(y=reorder(County, Estimate.Intercept), x=Estimate.Intercept)) +
#     geom_point(color="blue", size=2) +
#     geom_linerange(aes(
#       xmin=Q2.5.Intercept, xmax=Q97.5.Intercept
#     )) +
#     labs(
#       x="Intercept",
#       y="County"
#     ) +
#     theme_bw()
```