Case Study 2

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

10/28/2021

Tools -> Global Options -> Code -> Display -> Show Margin

```
# 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 %in% c("", " ")] <- 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(-c(election_date, stats_type, update_date)) %>%
  mutate(total_voters2 = as.numeric(total_voters))

votes <- votes %>%
  select(-c(election_date, stats_type, update_date)) %>%
  mutate(total_voters = as.numeric(total_voters))
```

```
# aggregated_data <- aggregate(voter$total_voters,
#                              list(county_desc = voter$county_desc,
#                                   age=voter$age,
#                                   party_cd=voter$party_cd,
#                                   race_code = voter$race_code,
#                                   sex_code = voter$sex_code,
#                                   precinct_abbrv = voter$precinct_abbrv,
#                                   ethnic_code = voter$ethnic_code,
#                                   voting_method_desc = voter$voting_method_desc),
#                              sum)
#
# aggregated_data <- aggregated_data %>%
#   rename(total_voters = x)
```

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

```
## `summarise()` has grouped output by 'county_desc', 'age', 'party_cd', 'race_code', 'ethnic_code'. Yo
```

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

```
## `summarise()` has grouped output by 'county_desc', '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
# 13066 rows
data <- data %>%
  filter(!is.na(total_reg)) %>%
  mutate(
    total_vot = case_when(
      is.na(total_vot) ~ 0,
      total_vot > total_reg ~ total_reg,
      TRUE ~ total_vot
    )
  ) %>%
  drop_na()
```

```r
set.seed(1031)

counties <- sample(unique(data$county_desc), 30)

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

EDA

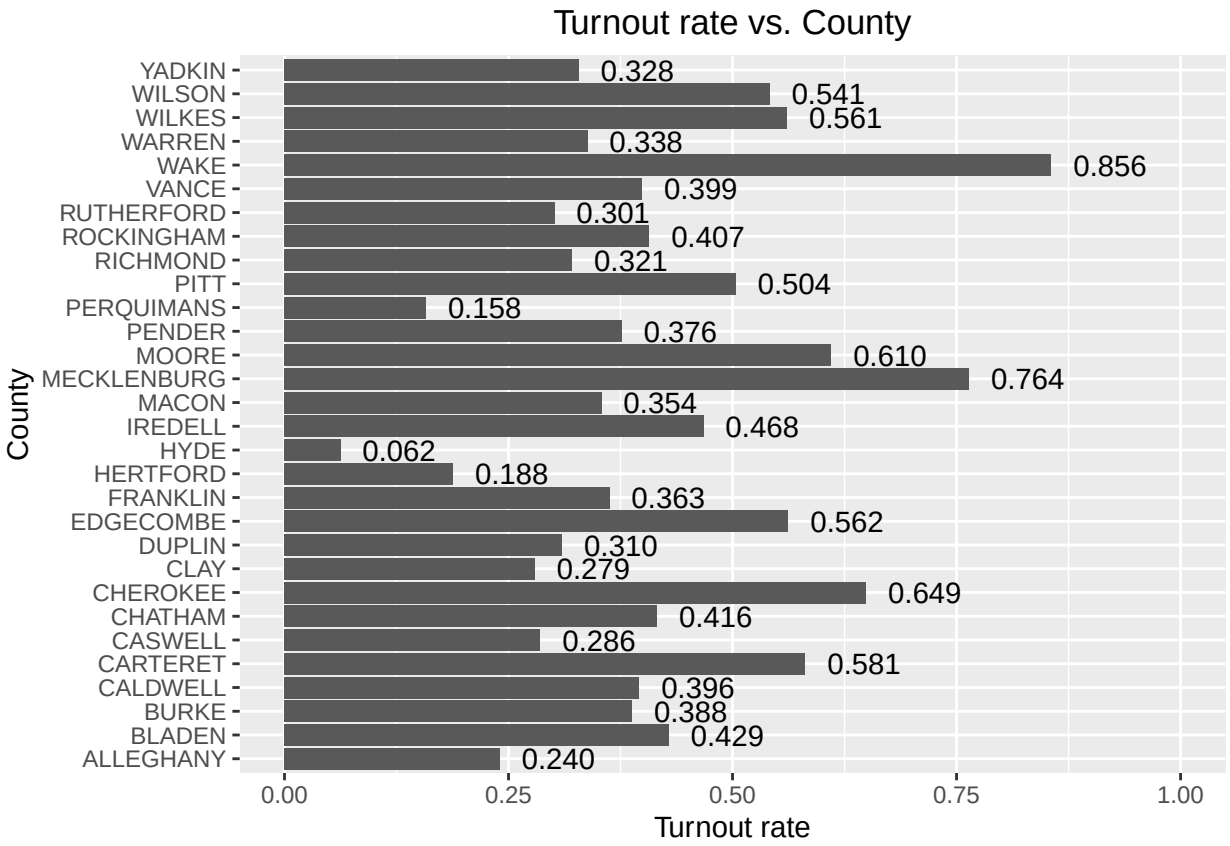
```r
turnout_rate <- sum(data$total_vot) / sum(data$total_reg)

data.frame(group = "Total", turnout_rate = turnout_rate)
```

```
##   group turnout_rate
## 1 Total    0.6533629
```

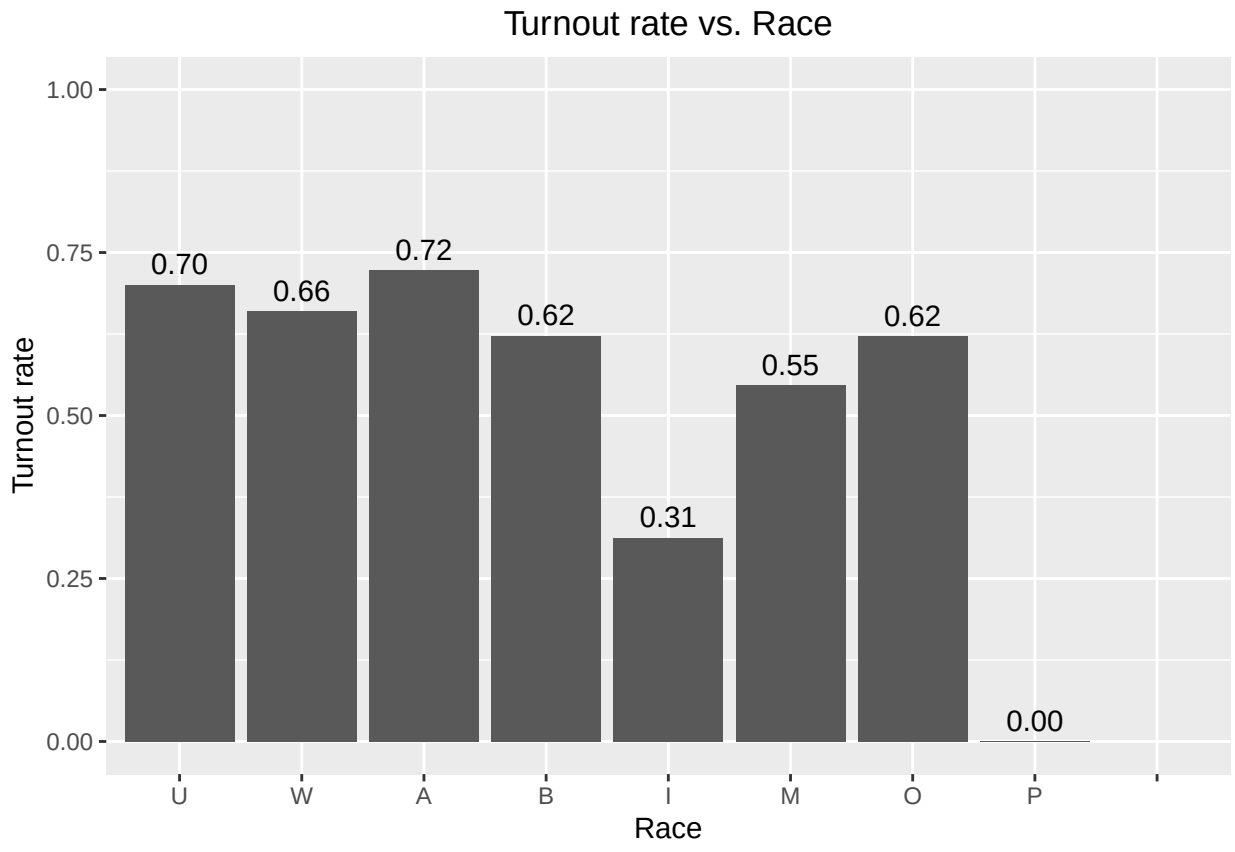
turnout rate

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



race

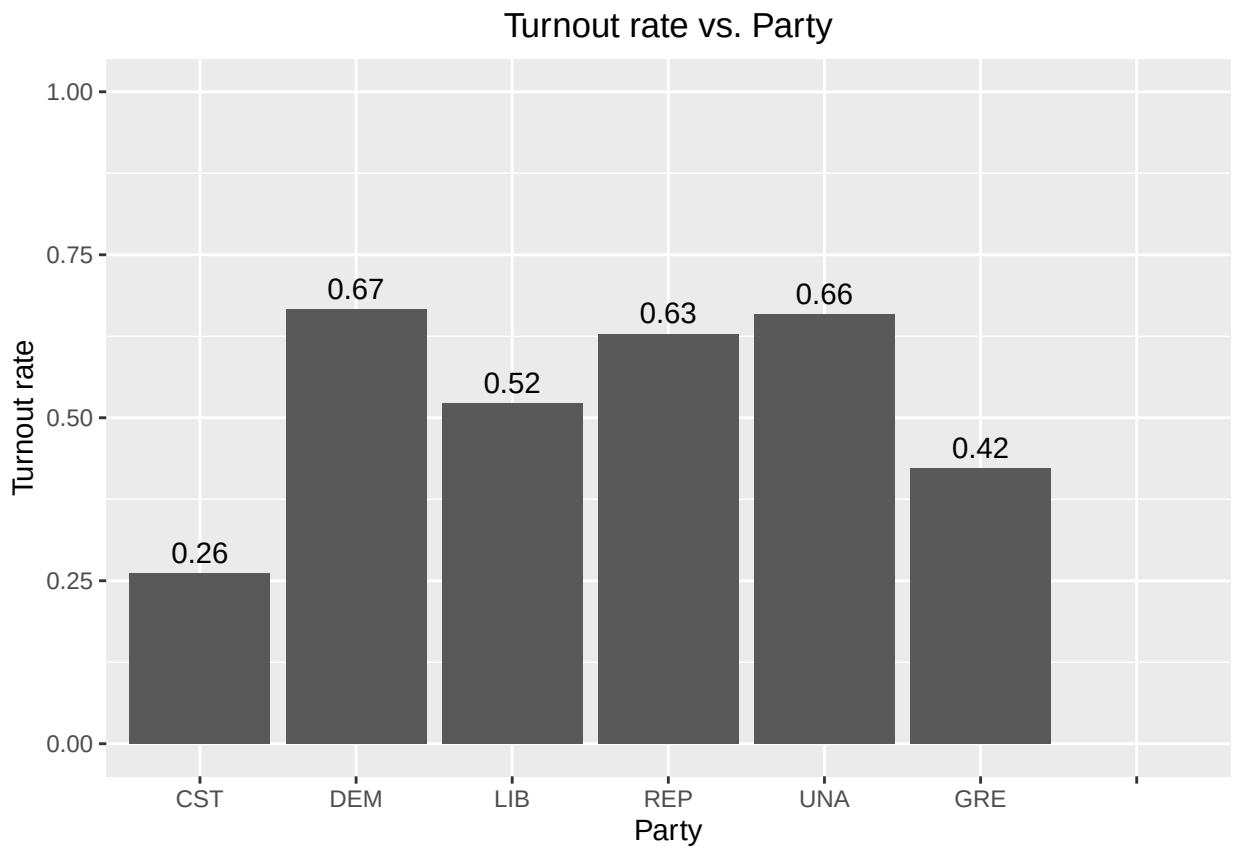
```
data %>%
  group_by(race_code) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
  select(group = race_code, turnout_rate) %>%
  ggplot( aes(x = group, y = turnout_rate)) +
  geom_bar(stat =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



party

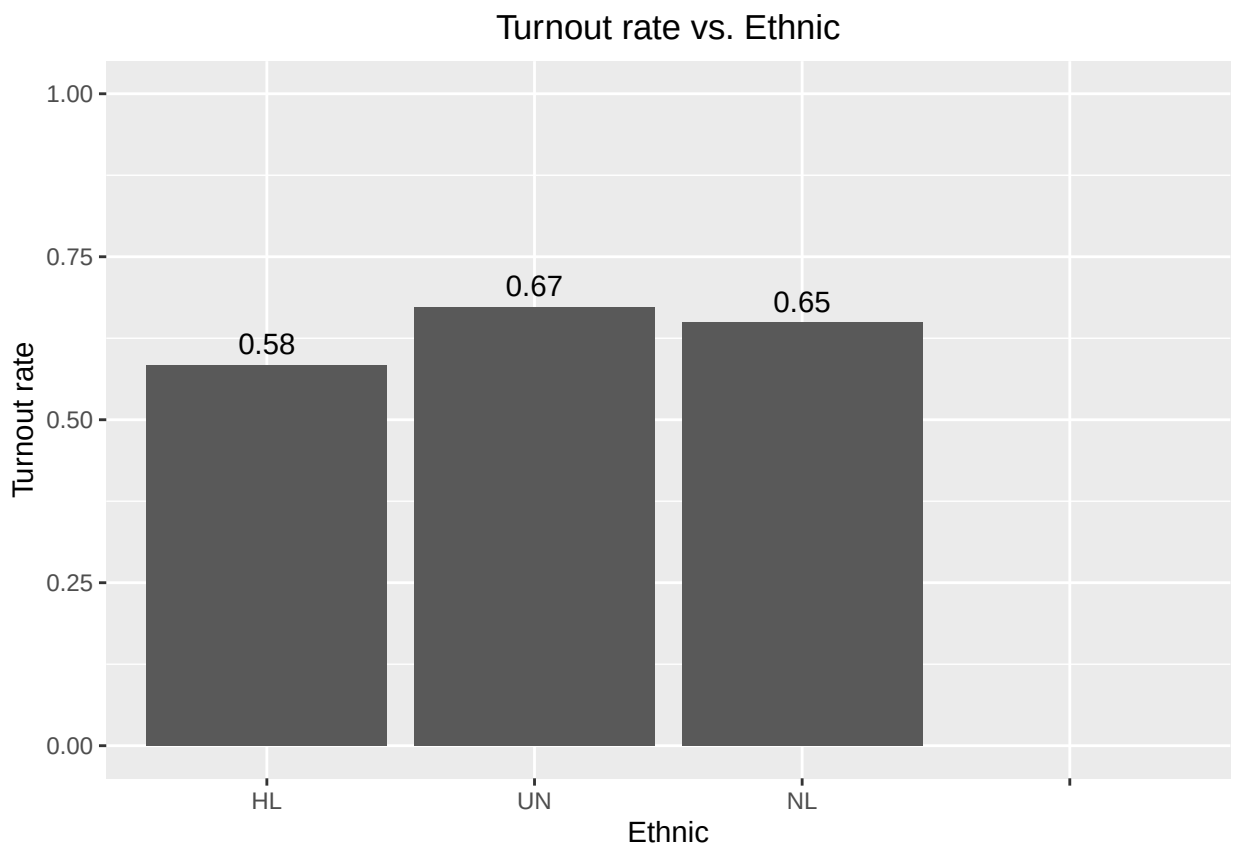
```
data %>%
  group_by(party_cd) %>%
  summarise(total_reg = sum(total_reg),
            total_vot = sum(total_vot), .groups = "drop") %>%
  mutate(turnout_rate = total_vot / total_reg) %>%
```

```
  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))
```



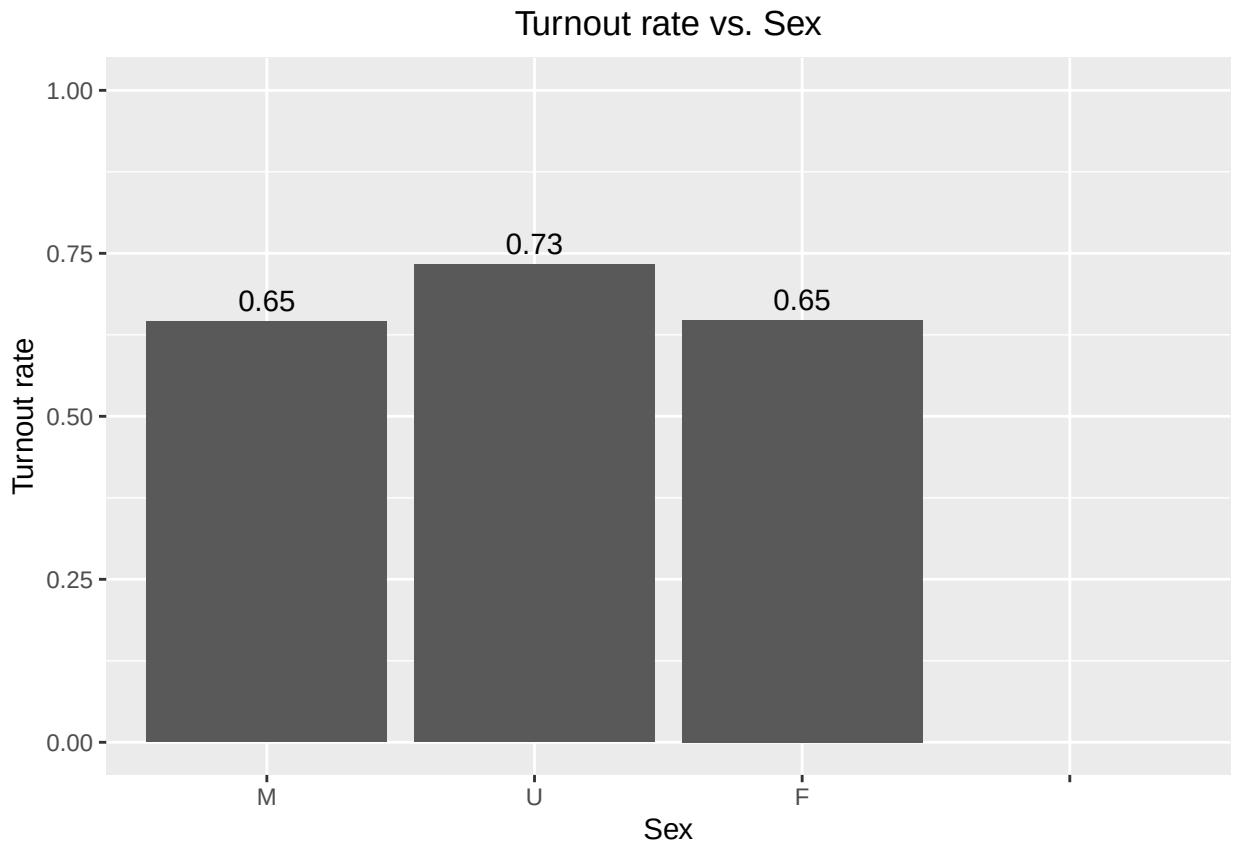
ethnic groups

```
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))
```



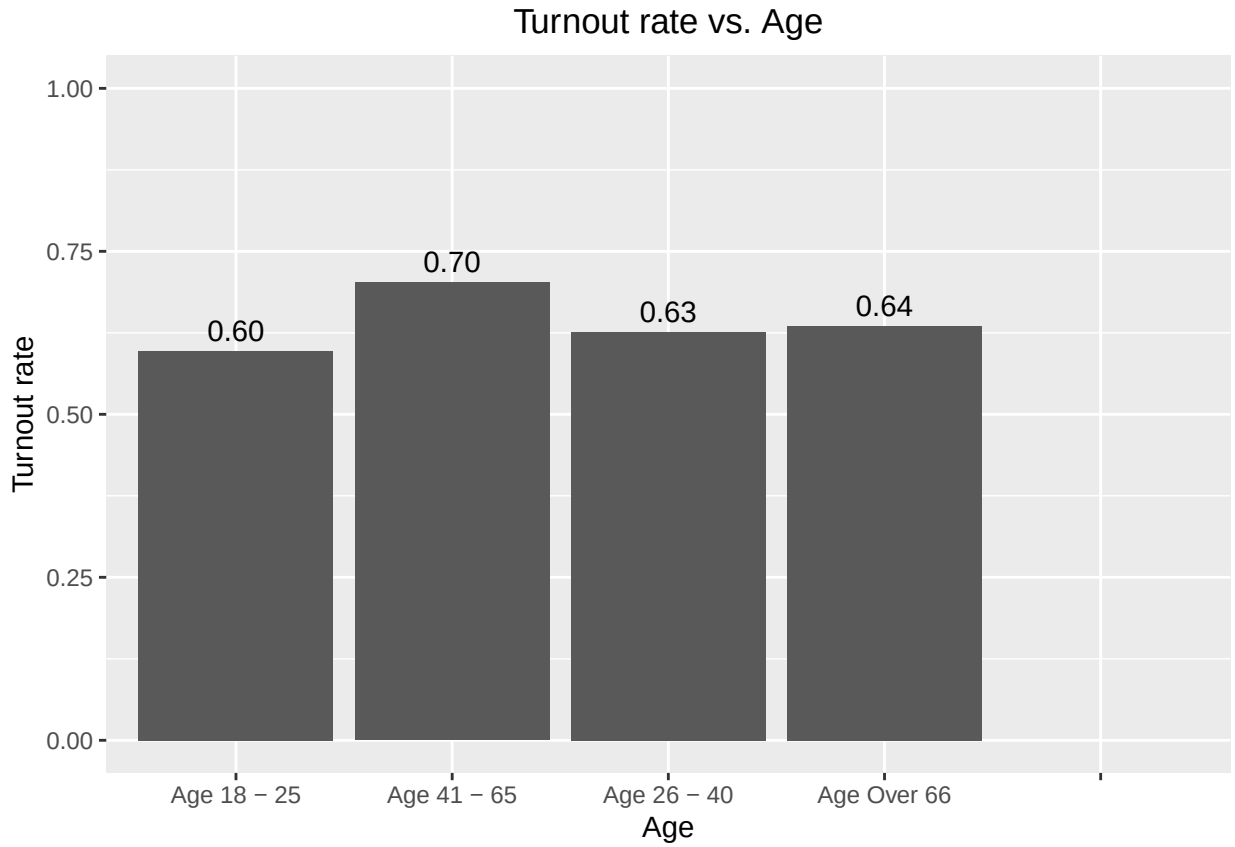
Sex

```
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))
```



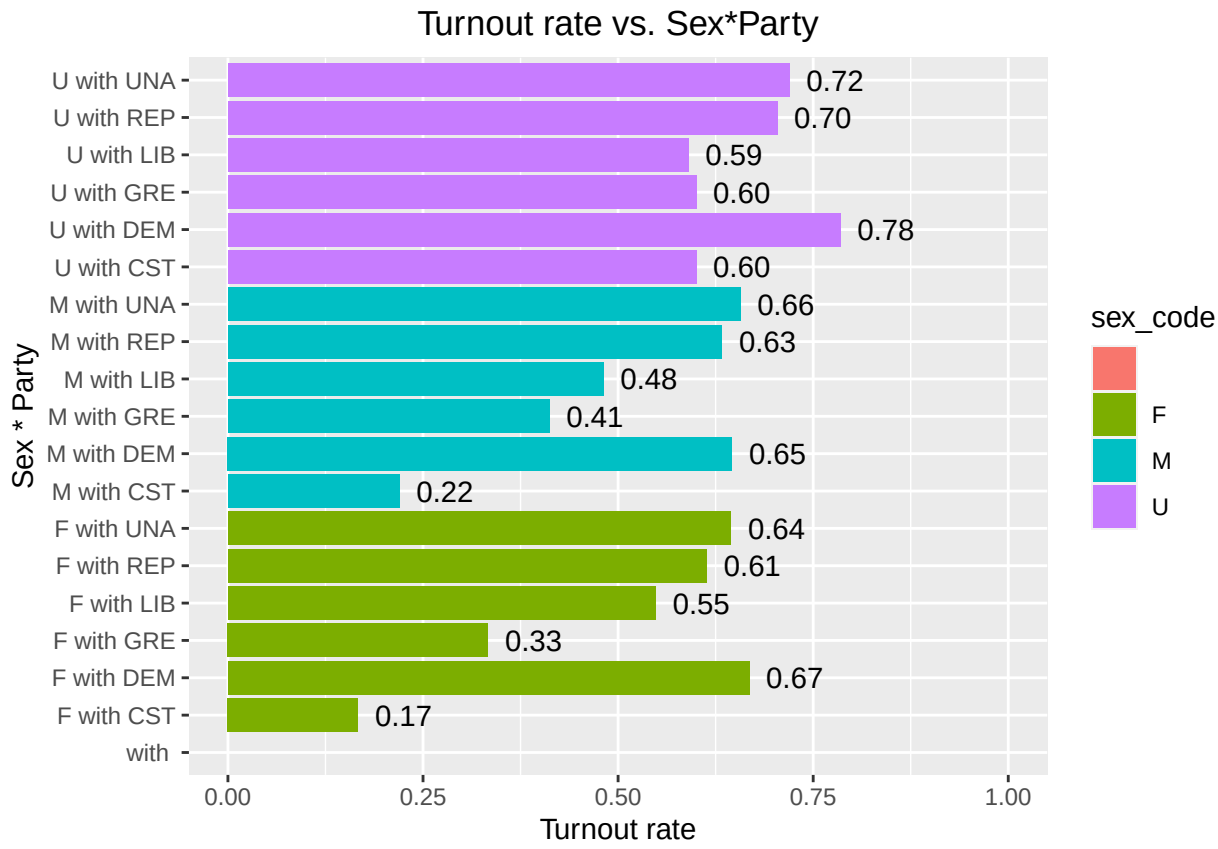
age

```
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))
```



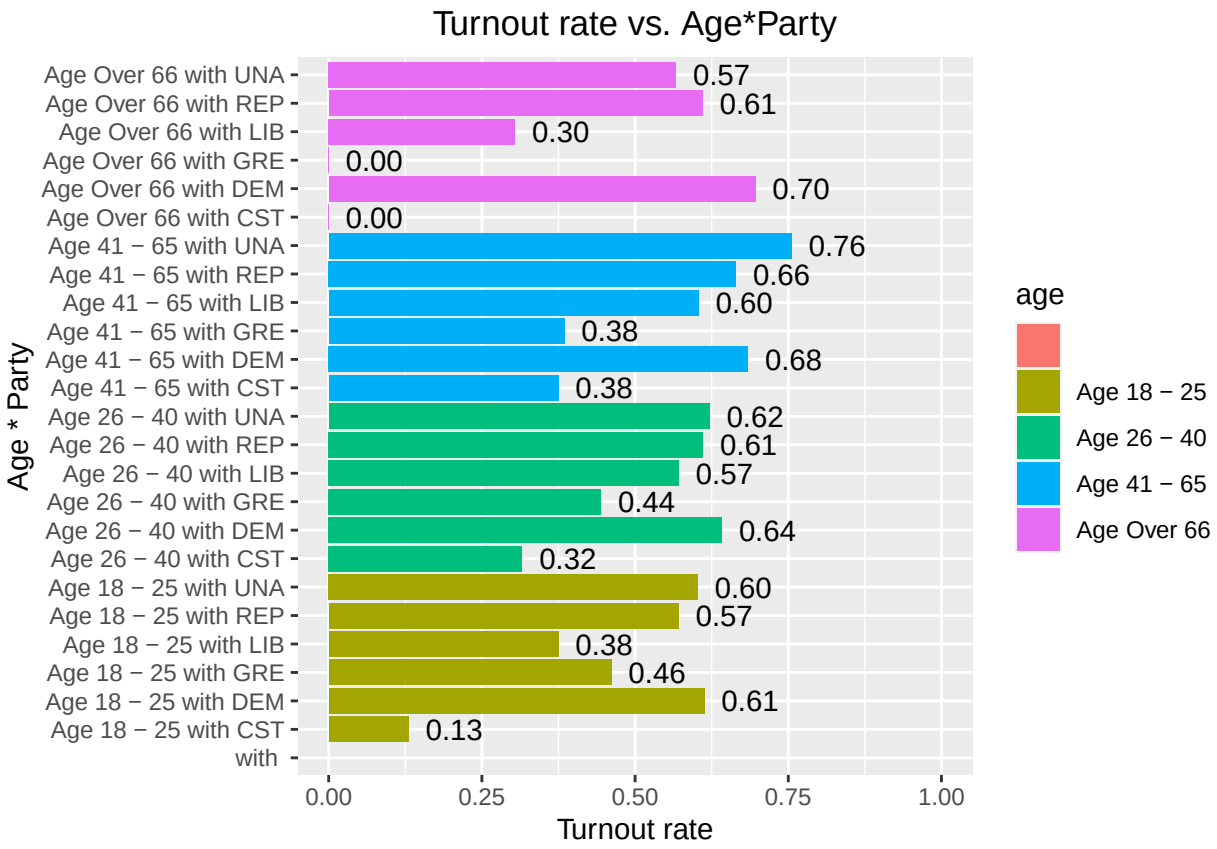
sex & party

```
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))
```



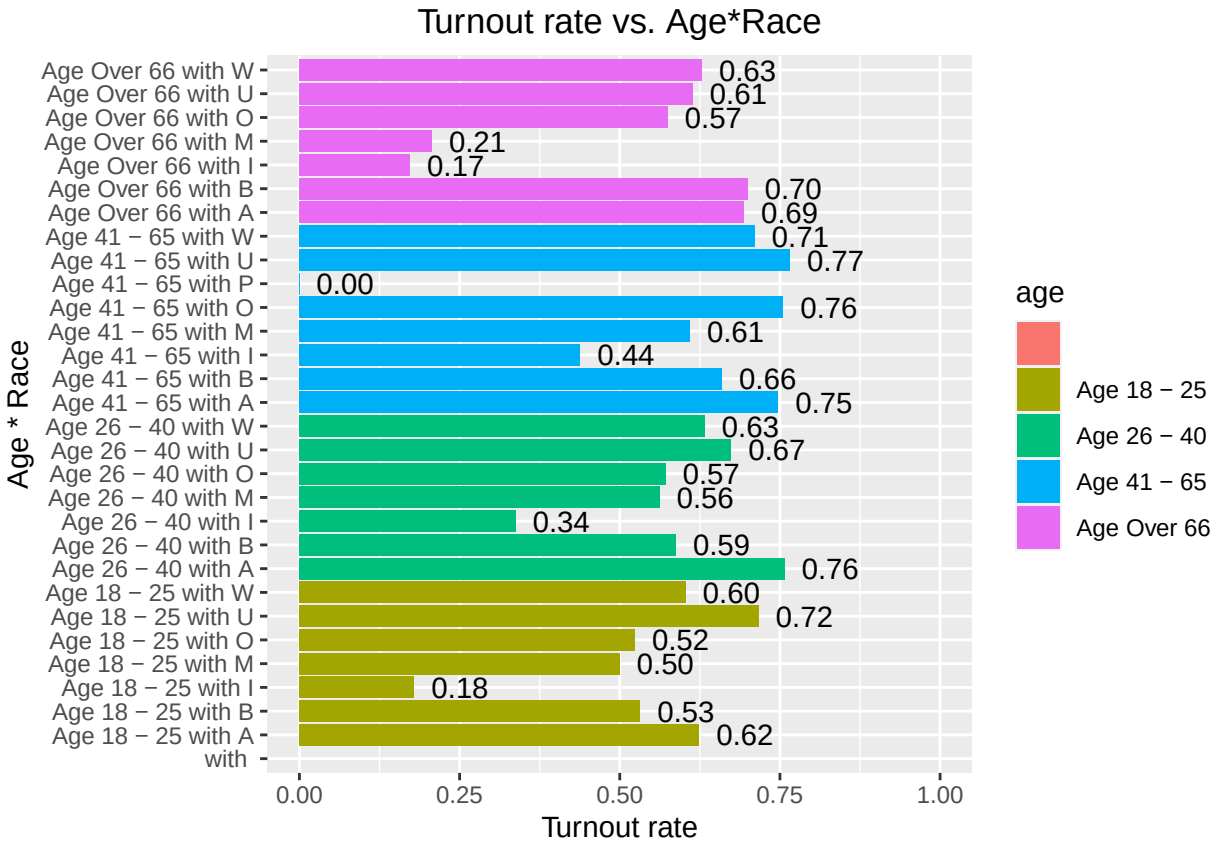
age & party

```
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))
```



age & race

```
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))
```



Model

In order to answer the questions of interest, we require the following features in our model:

1. Fixed effects for various demographic subgroups (race, ethnicity, age, sex)
2. Random effects for county
3. Sex-Party interaction
4. Age-Party interaction

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

```
as.data.frame(summary(mod1)$coefficients) %>%
  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()
```

	Estimate	95% CI
(Intercept)	-0.372	[-0.659, -0.086]

```
# 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)
```