

Case Study 2

Emily Gentles

Weiyi Liu

Jack McCarthy

Qinzhe Wang

10/28/2021

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

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

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

```
# contains info about the aggregate counts of registered voters by demographic variables
```

```
voter <- read.table("voter_stats_20201103.txt", header = TRUE, fill = TRUE, sep = '\t')
```

```
# set "" or " " to NA
```

```
voter[voter == ""] <- NA
```

```
voter[voter == " "] <- NA
```

```
history[history == " "] <- NA
```

```
unique(voter$election_date) # "11/03/2020" NA
```

```
## [1] "11/03/2020" NA
```

```
unique(voter$stats_type) # "history" NA
```

```
## [1] "voter" NA
```

```
unique(voter$update_date) # "01/13/2021" NA
```

```
## [1] NA
```

```
# remove above three columns
```

```
voter <- voter %>%  
  select(-election_date, -stats_type, -update_date)
```

```
history <- history %>%  
  select(-election_date, -stats_type, -update_date)
```

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

```
voter <- voter %>%  
  mutate(total_voters2 = 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)
```

```
history <- history %>%
  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'. You

```
data <- voter %>%
  group_by(county_desc, party_cd, race_code, ethnic_code, sex_code, age) %>%
  summarize(total_reg = sum(total_voters2, na.rm = T)) %>%
  left_join(history, 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'

```
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))
```

```
data <- data %>%
  filter(!is.na(total_reg))

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

```
set.seed(1031)
counties <- sample(unique(data$county_desc), 30)

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

```
sum(is.na(data))
```

```
## [1] 5
```

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

EDA

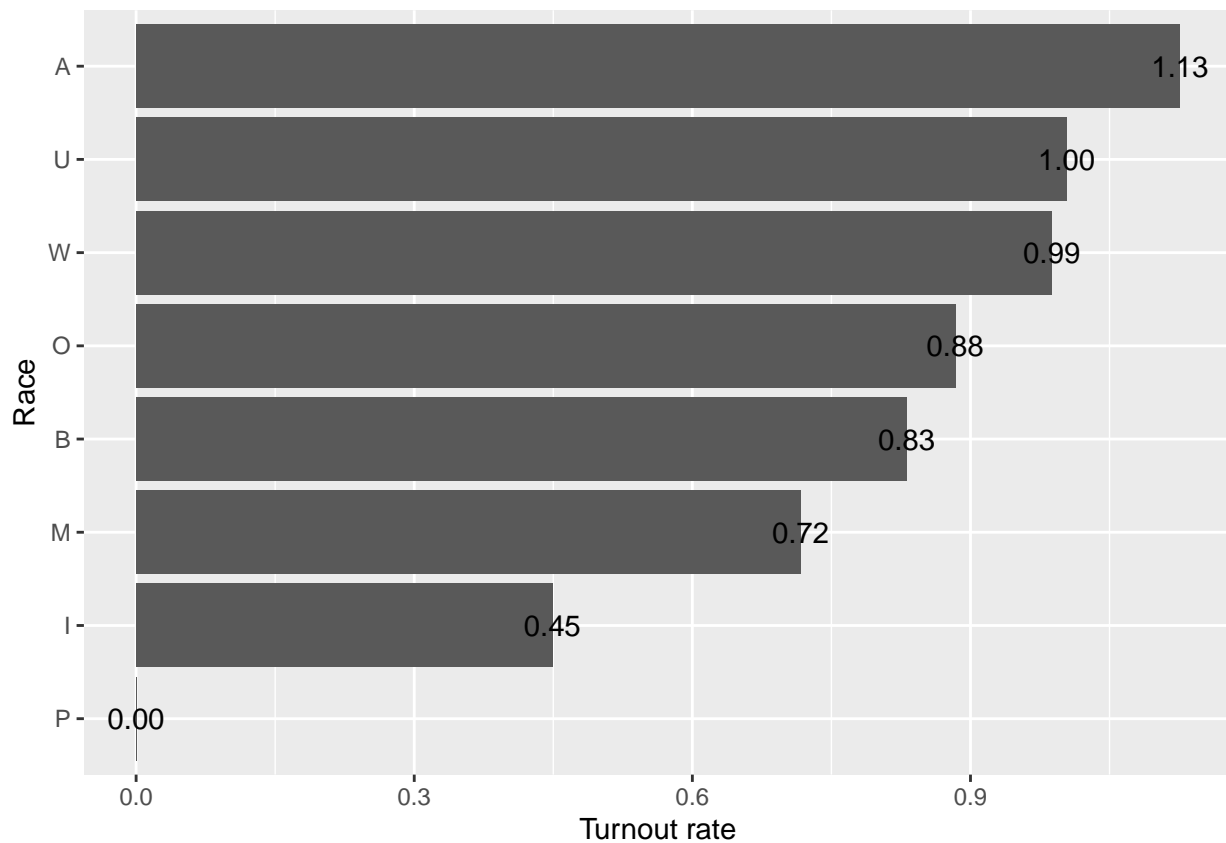
```
turnout_rate <- sum(data$total_vot) / sum(data$total_reg)
```

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

```
##   group turnout_rate  
## 1 Total    0.9482582
```

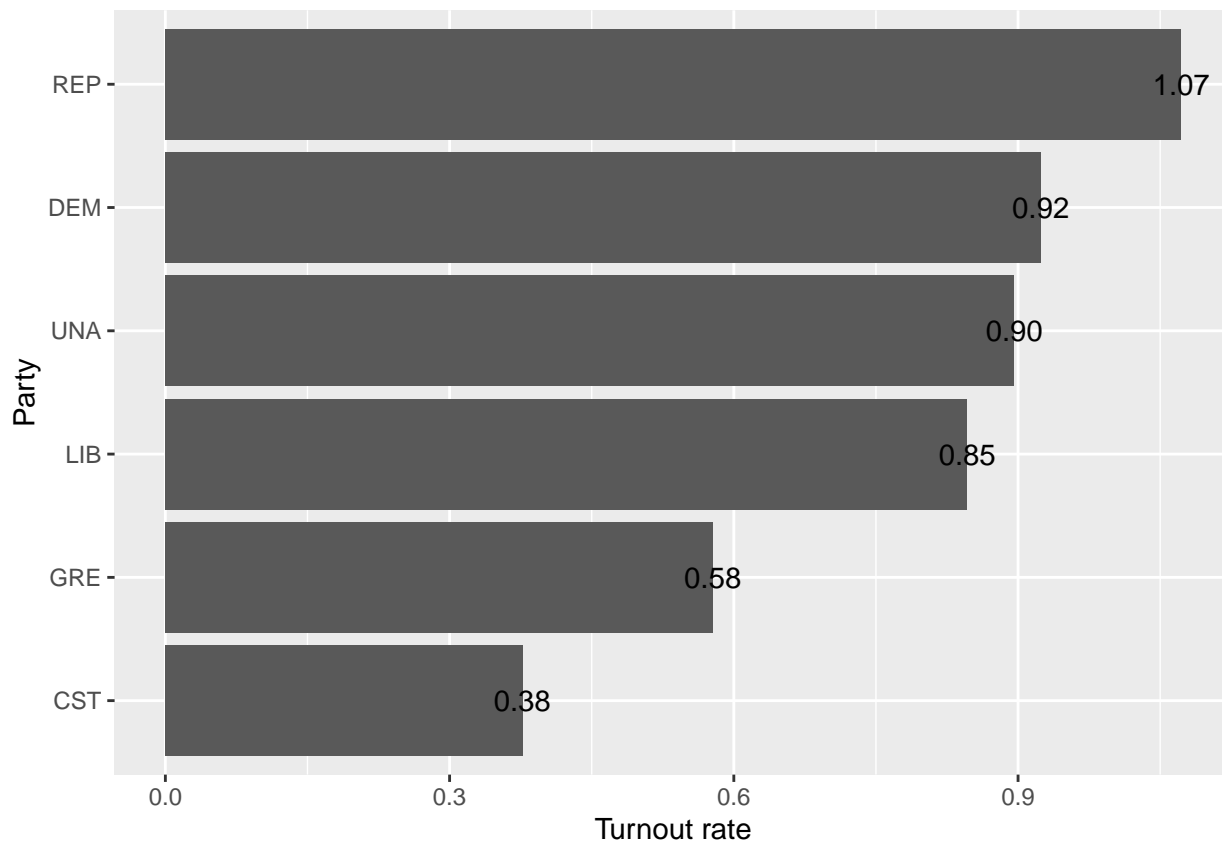
race

```
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 = reorder(group, turnout_rate), y = turnout_rate)) +  
  geom_bar(stat = "identity") +  
  xlab("Race") + ylab("Turnout rate") +  
  geom_text(aes(label=format(turnout_rate, digits = 2))) +  
  coord_flip()
```



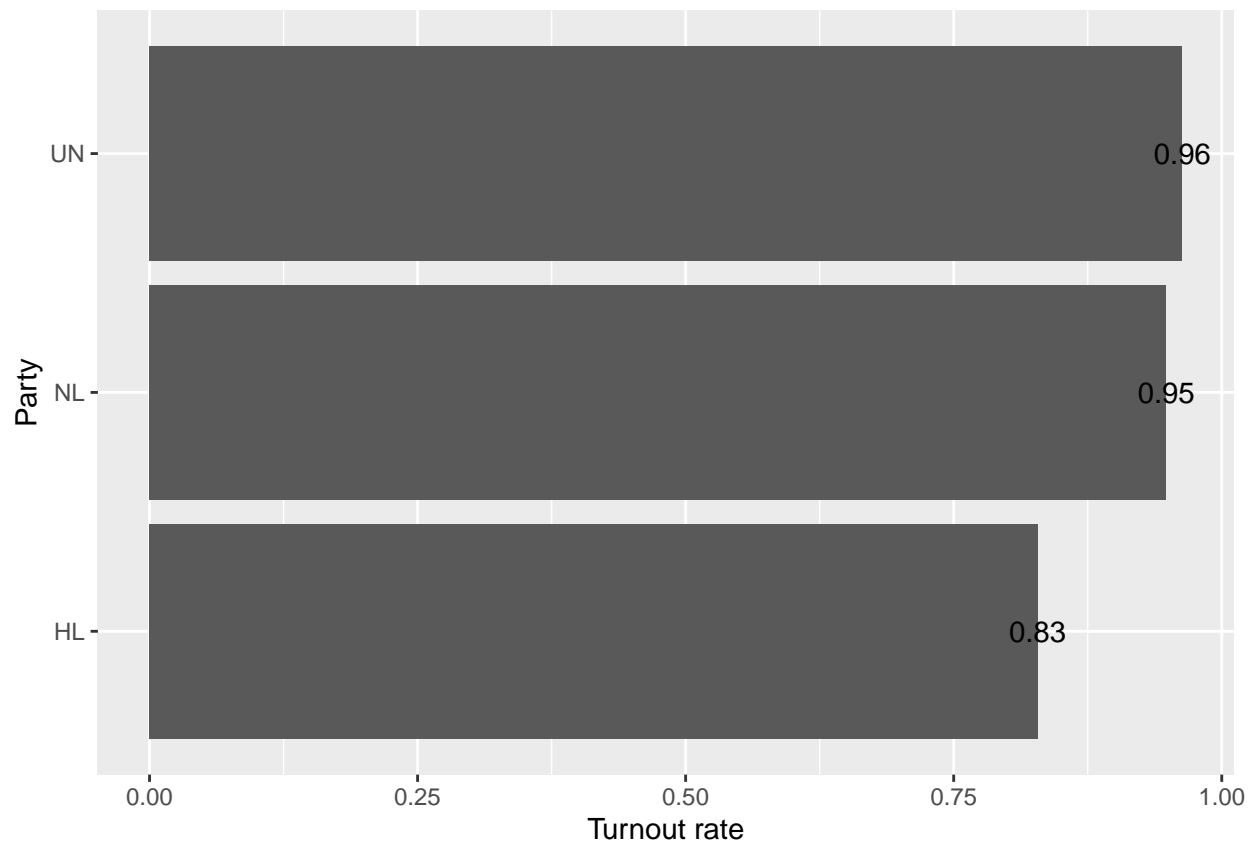
party

```
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 = reorder(group, turnout_rate), y = turnout_rate)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2))) +
  coord_flip()
```



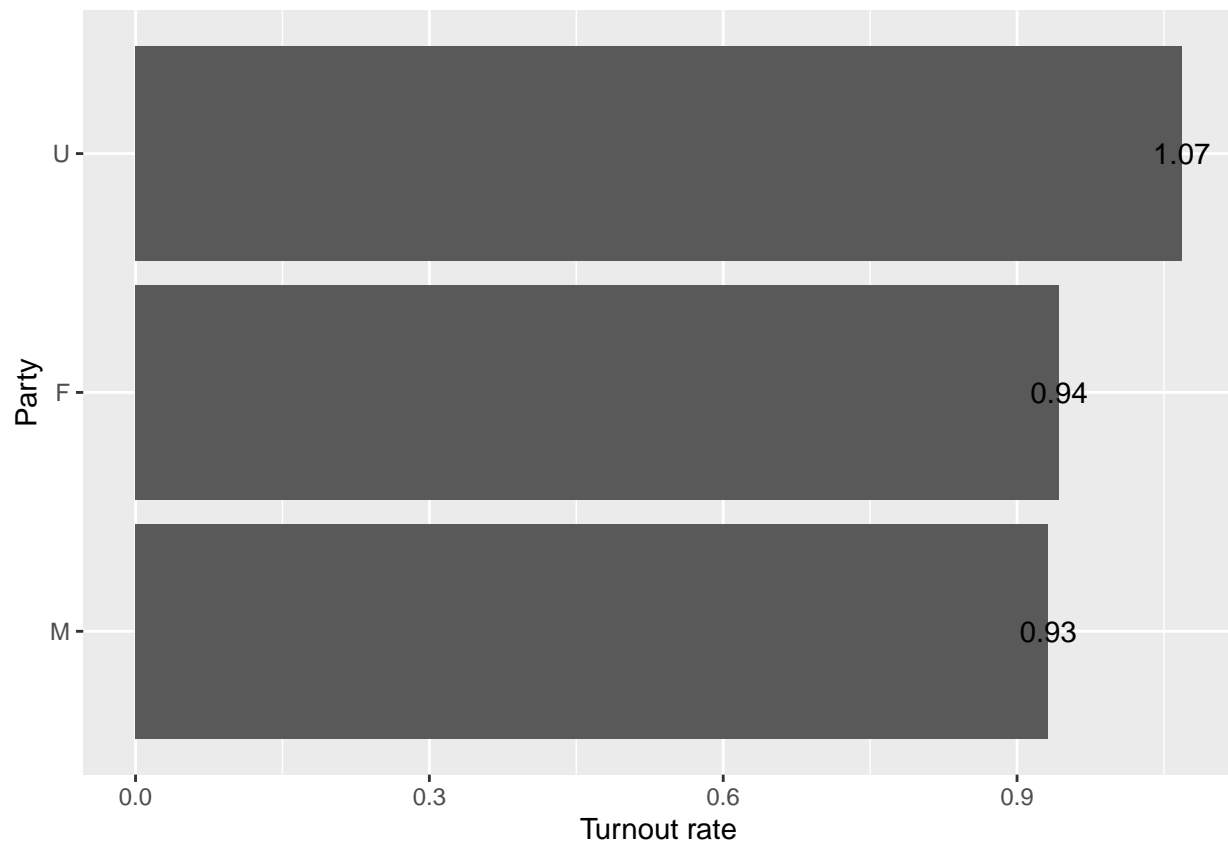
ethnic groups

```
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 = reorder(group, turnout_rate), y = turnout_rate)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2))) +
  coord_flip()
```



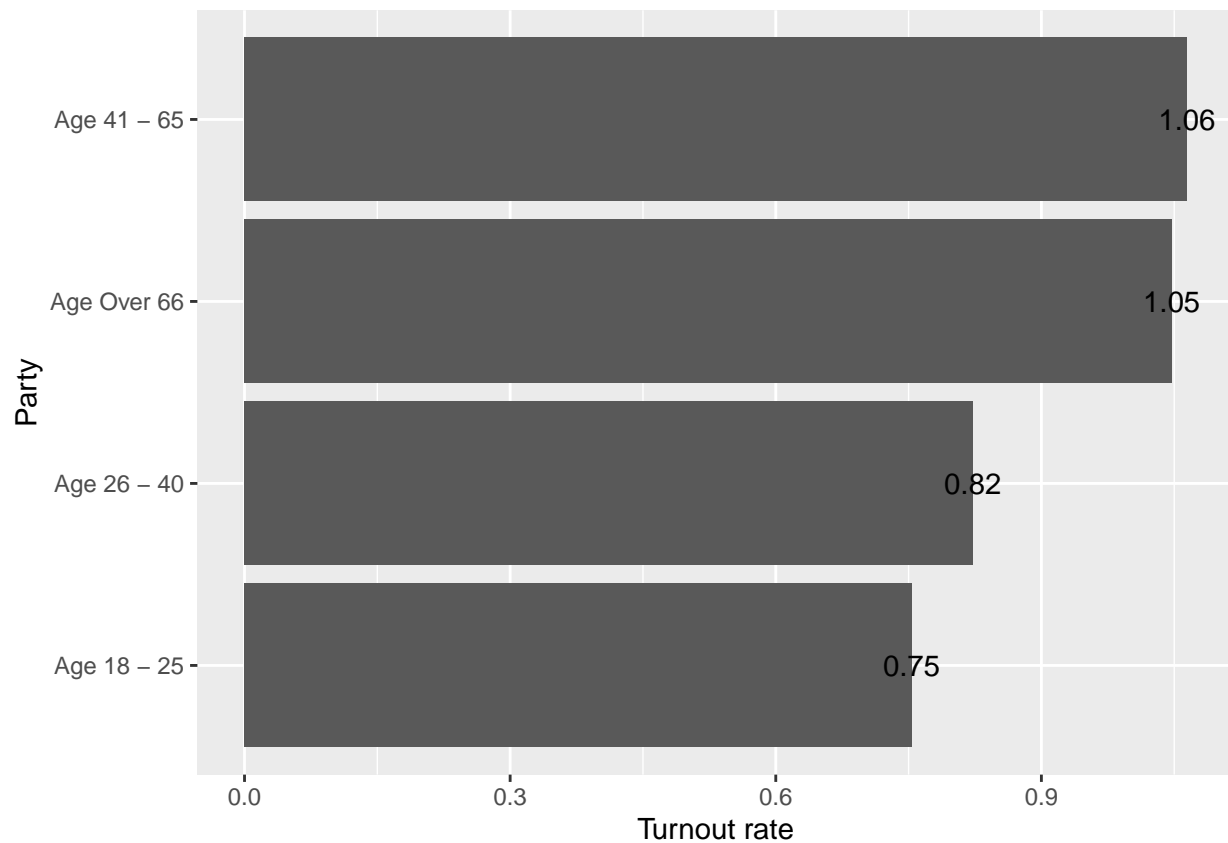
sex

```
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 = reorder(group, turnout_rate), y = turnout_rate)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2))) +
  coord_flip()
```



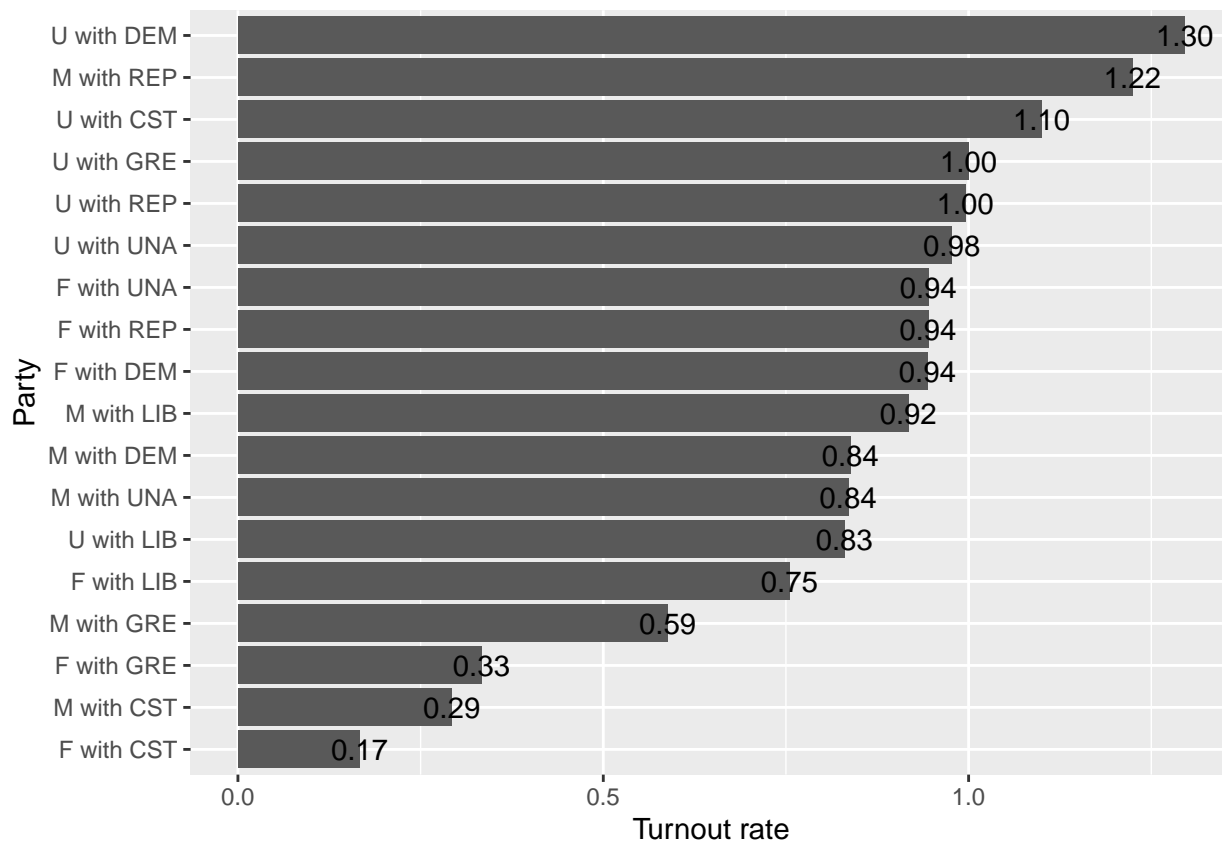
age

```
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 = reorder(group, turnout_rate), y = turnout_rate)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2))) +
  coord_flip()
```



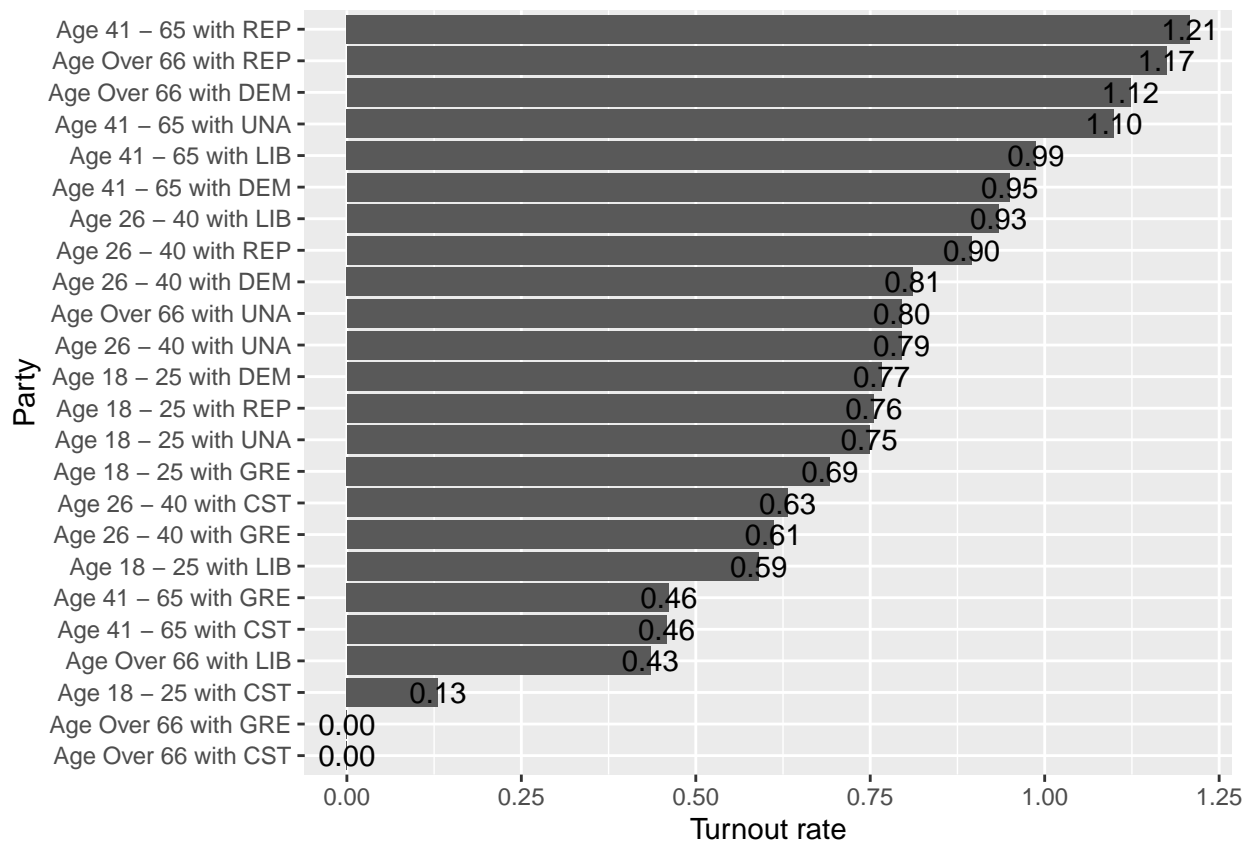
sex & party

```
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) %>%
  ggplot( aes(x = reorder(group, turnout_rate), y = turnout_rate)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2))) +
  coord_flip()
```

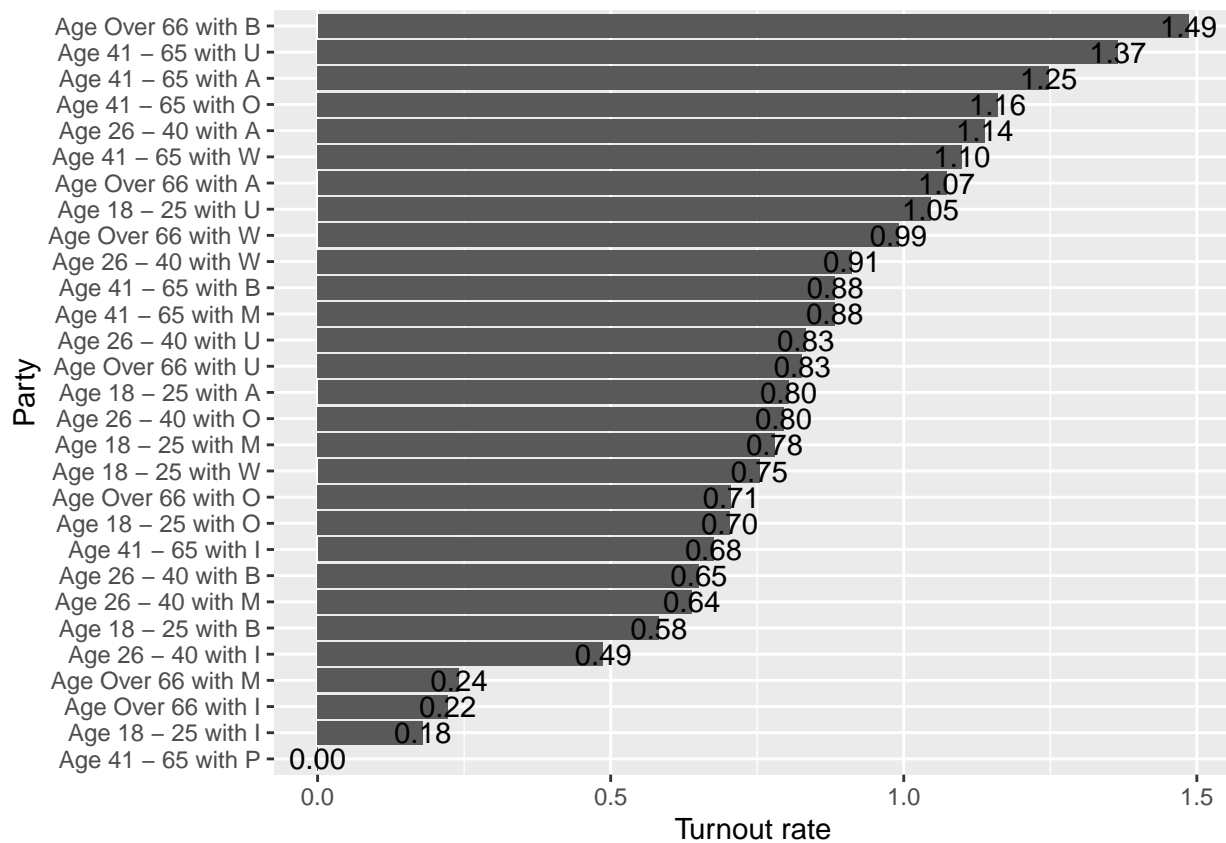
age & party

```
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) %>%
  ggplot( aes(x = reorder(group, turnout_rate), y = turnout_rate)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2))) +
  coord_flip()
```



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) %>%
  ggplot(aes(x = reorder(group, turnout_rate), y = turnout_rate)) +
  geom_bar(stat = "identity") +
  xlab("Party") + ylab("Turnout rate") +
  geom_text(aes(label=format(turnout_rate, digits = 2))) +
  coord_flip()
```



Model

```
# attach(data)
# mod1 <- 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"))
```

```
# 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)
```