

R-seminar 2 Oppgaver

STV1020 Vår 2021

Uke 11

1. Bruk koden `read.dta("ESS9NO.dta")` for å få ESS-datasettet med alle variablene igjen.

- Variabelen `"trstplt"` måler tillit til politikere og variabelen `"gndr"` er en dikotom variabel for respondentens kjønn. Gi disse to variablene, samt `"yrbrn"` navn som du synes er mer intuitive. Du skal bruke dine egne navn på variablene i de neste oppgavene der det står `"trstplt"`, `"gndr"`, og `"yrbrn"`.
- Gjør `"yrbrn"` (året man er født) om til alderen respondenten hadde i 2018.

```
library(tidyverse)
library(foreign)

df <- read.dta("ESS9NO.dta")

# Variabelen "trstplt" måler tillit til politikere og variabelen "gndr" er en
# dikotom variabel for respondentens kjønn. Gi disse to variablene, samt
# "yrbrn" (år født) navn som du synes er mer intuitive. Du skal bruke disse
# navnene i de neste oppgavene der det står "trstplt", "gndr", og "yrbrn".

df <- df %>%
  rename(trust = trstplt,
         gender = gndr,
         age = yrbrn)

# Gjør "yrbrn" (året man er født) om til alderen respondenten hadde i 2018.

df$age <- 2018-df$age

# Sjekker om det ble riktig
summary(df$age)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	15.00	32.00	47.00	46.54	61.00	90.00	32

2. Lag et datasett som bare inneholder `"trstplt"`, `"gndr"`, `"yrbrn"`, og `"vote"`.

```
df <- df %>%
  select(trust, gender, age, vote)

View(df)
```

3. Hvilken klasse har "trstplt" og "gndr"?

```
class(df$trust) # factor

## [1] "factor"

class(df$gender) # factor

## [1] "factor"
```

4. Hvor mange nivåer har "trstplt"? Hvor mange har "complete trust" til politikere? Er noen av nivåene mindre relevante?

```
levels(df$trust)

## [1] "No trust at all" "1" "2" "3"
## [5] "4" "5" "6" "7"
## [9] "8" "9" "Complete trust" "Refusal"
## [13] "Don't know" "No answer"

# 10, når man ser bort i fra "Refusal", "Don't Know", og "No answer".

summary(df$trust)

## No trust at all      1      2      3      4
##           32      28      75     120    152
##           5       6       7       8       9
##          308     290     240     109     36
## Complete trust      Refusal    Don't know    No answer    NA's
##           8         0         0         0         8

# 8 stk har "complete trust" til politikere. Ingen har verdiene "Refusal",
# "Don't Know" eller "No answer", og disse måler ikke politisk tillit.
```

5. Hvor mange nivåer har "gndr"? Hvor mange kvinner er det i datasettet?

```
levels(df$gender) # 2, når man ser bort i fra NA.

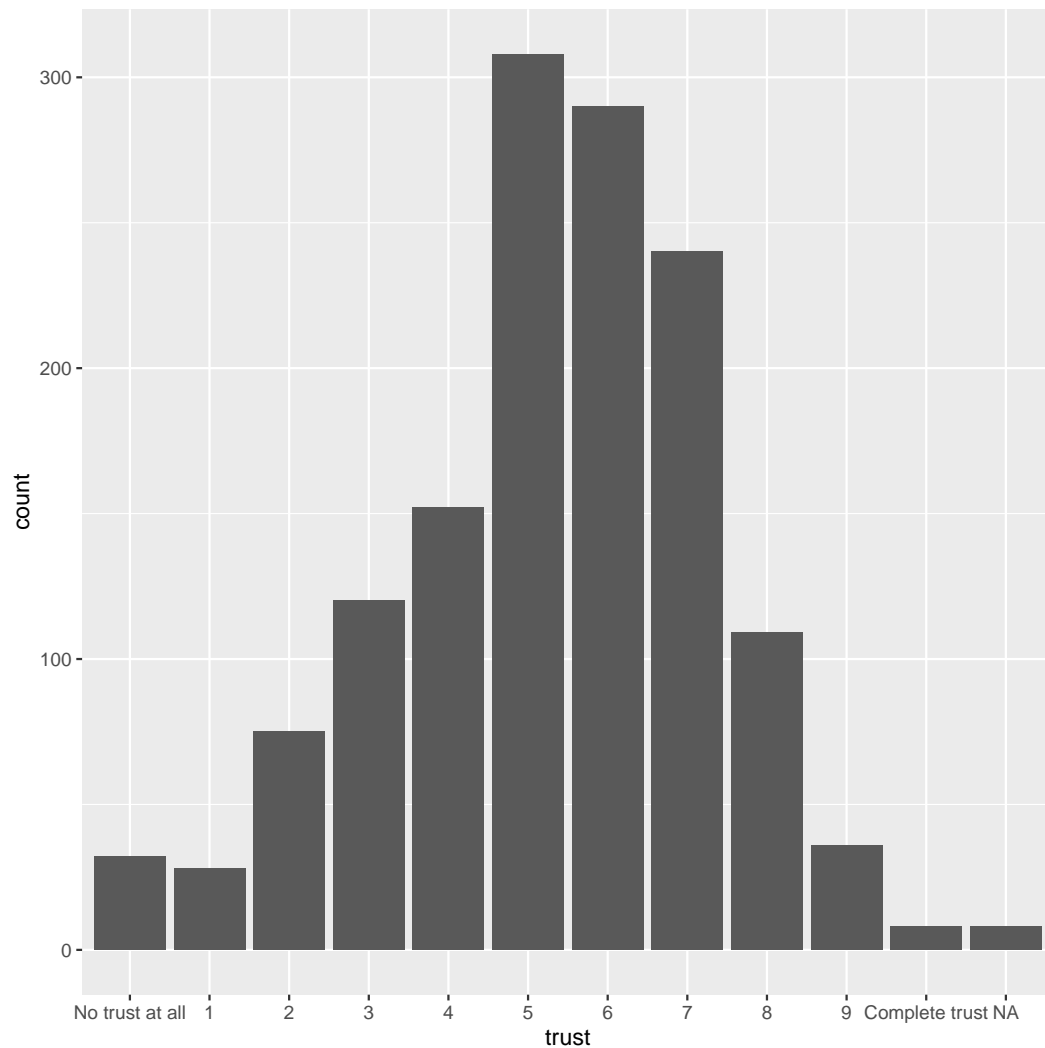
## [1] "Male" "Female" "No answer"

summary(df$gender) # Det er 629 kvinner i datasettet.

##      Male      Female No answer
##      777      629      0
```

6. Lag et histogram som viser fordelingen av "trstplt". Si noe om hvordan verdiene fordeler seg.

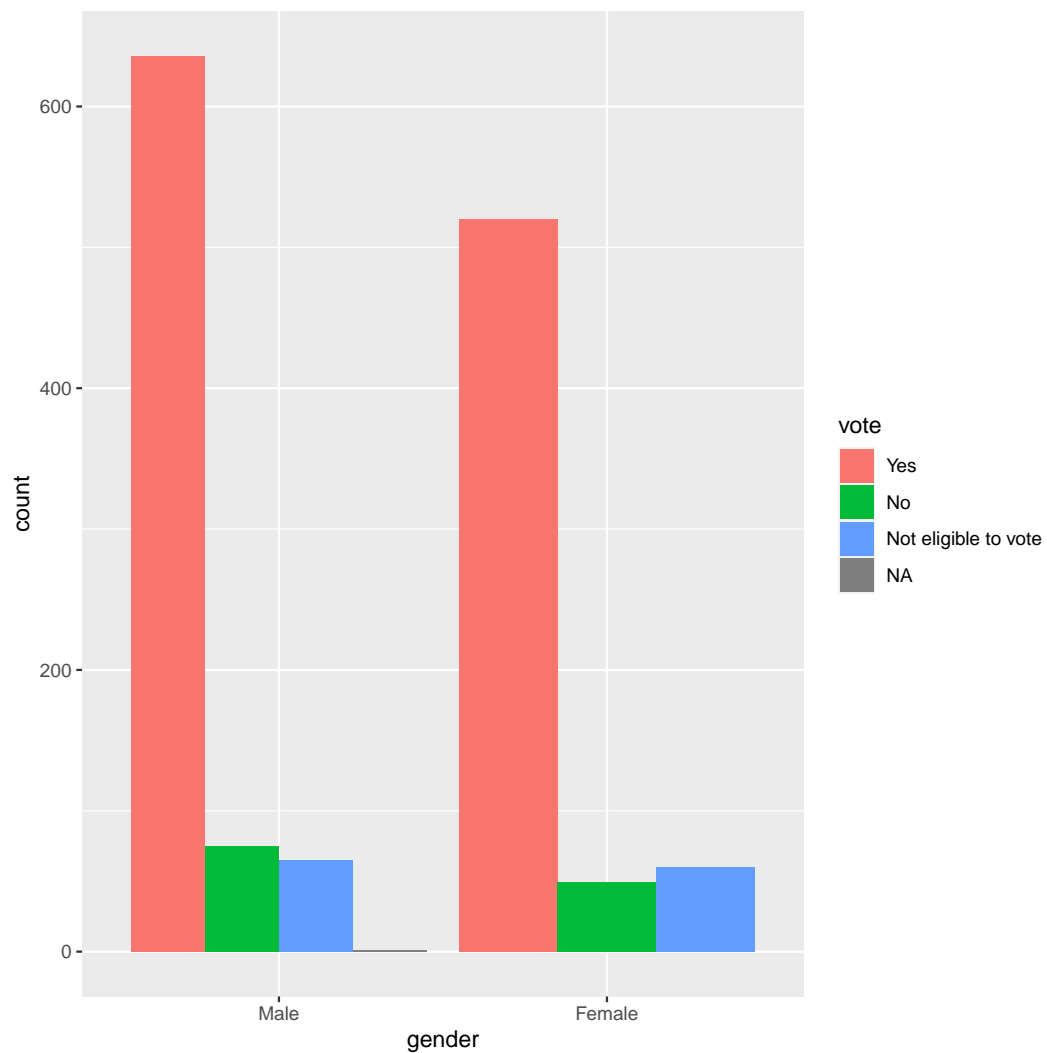
```
# _bar  
ggplot(data = df, aes(x = trust)) +  
  geom_bar()
```



```
# Normalfordeling eller ikke? Median rundt 5-6
```

7. Er det noe forskjell mellom hvor mange menn og kvinner som stemte? Vis dette ved hjelp av et diagram.

```
# Forslag til diagram man kan bruke:  
ggplot(data = df, aes(x = gender)) +  
  geom_bar(aes(fill = vote),  
    position = "dodge")
```

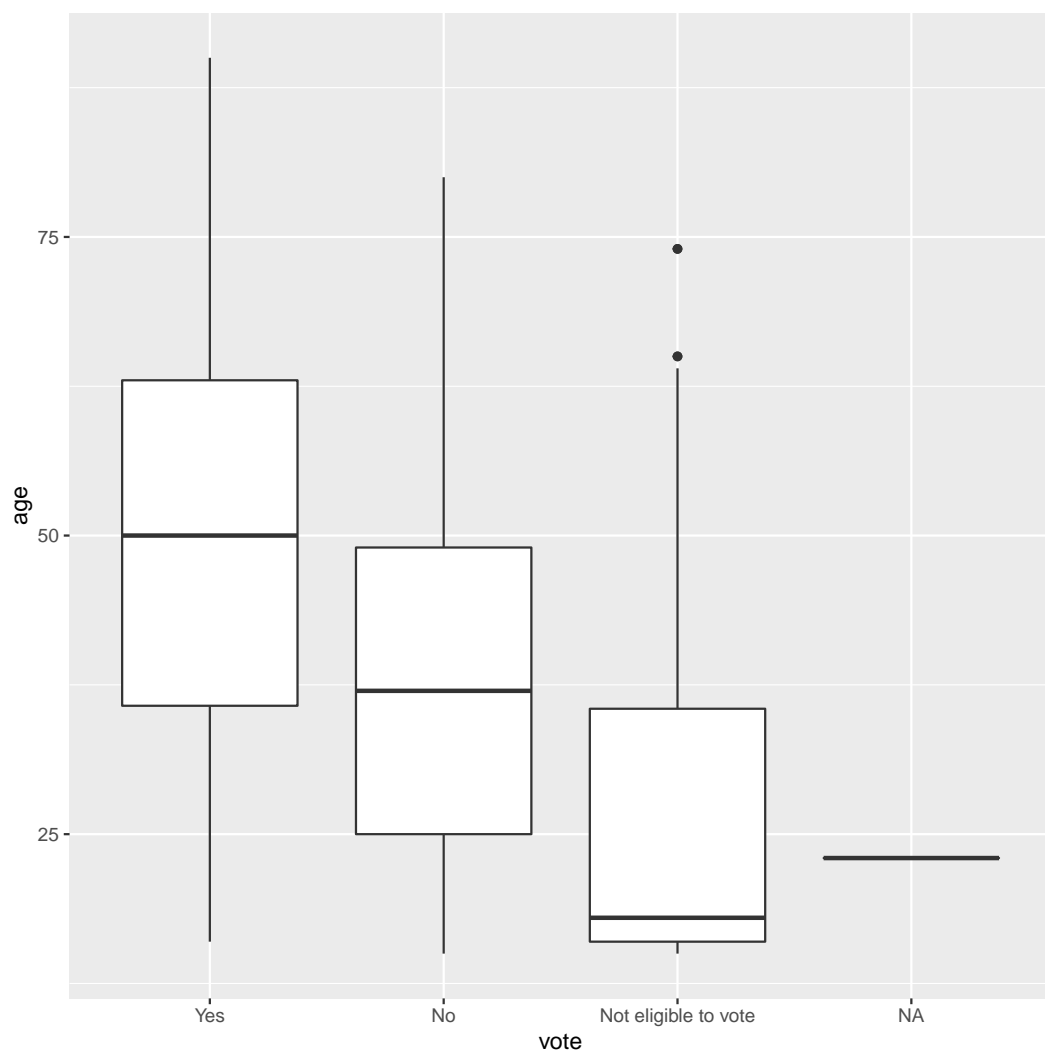


```
# Tolk
```

8. Er det en forskjell i spredningen av alder for de som stemte og de som ikke stemte? Vis dette ved hjelp av et diagram.

```
# Forslag til diagram man kan bruke:
ggplot(data = df, aes(x = vote, y = age)) +
  geom_boxplot()

## Warning: Removed 32 rows containing non-finite values (stat_boxplot).
```



De som ikke stemte er yngre enn de som stemte.

Noen "uteliggere" på not eliglbe to vote alternativet, men majoriteten senterer seg rundt 1

av den tykke streken i boksen (medianen).