

Projekt z Metody Reprezentacyjnej

Klaudia Wolniewicz, Alicja Bicka, Michel Voss, Mateusz Stanisławski

13 02 2020

Metoda reprezentacyjna: jest częściowym badaniem statystycznym opartym na próbie pobranej ze zbiorowości generalnej w sposób losowy. Z teoretycznego i praktycznego punktu widzenia metoda ta jest najbardziej prawidłową formą badania częściowego. W metodzie reprezentacyjnej dokonuje się wyboru próby na dwa sposoby. Może to być wybór przez losowanie, albo przez celową selekcję.

Na początku na potrzeby dalszej analizy instalujemy potrzebne biblioteki:

```
install.packages("survey")
install.packages("tidyverse")
install.packages("DataExplorer")
install.packages("labelled")
install.packages("devtools")
```

```
library(tidyverse)
library(survey)
library(DataExplorer)
library(labelled)
```

Kolejnym krokiem jest załadowanie pliku z danymi do pamięci:

```
load("ess.RData")
```

Obliczanie i dodanie do zbioru danych kolumny z wagami uzyskanych na podstawie wag analitycznych. Ludność została pobrana z Banku Danych lokalnych

```
ess_weight<-ess %>%
  mutate(weight = (dweight/nrow(.))*38411148)
```

Na potrzeby przeprowadzanie badania, ze zbioru danych wybraliśmy następujące zmienne:

cndry - państwo

gndr - płeć

polintr - zainteresowanie polityką

trstprl - Trust in country's parliament

trstep - Trust in the European Parliament

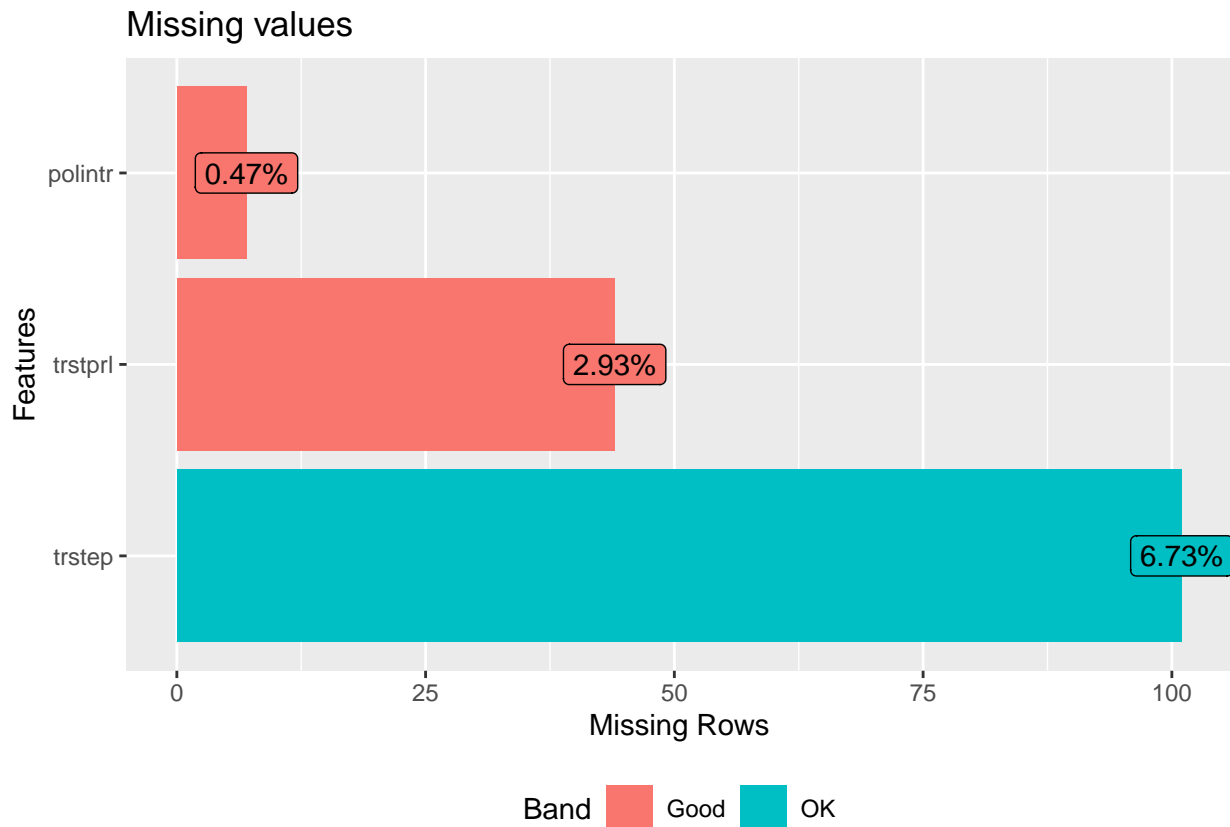
yrbrn - data urodzenia

```
variable_list<-c('cndry','gndr','polintr','trstprl','trstep','weight','yrbrn')
```

Analiza braków danych w wybranych zmiennych

```
ess_weight %>%
  select(variable_list) %>%
  plot_missing(
    missing_only = TRUE,
    title = "Missing values",
    ggtheme = theme_gray(),
```

```
theme_config = list(legend.position = c("bottom"))
)
```



Sprawdzenie informacji o danych, w których występują braki

```
print(ess_weight$polintr)
print(ess_weight$trstprl)
print(ess_weight$trstep)
```

```
typeof(ess_weight$polintr)
typeof(ess_weight$trstprl)
typeof(ess_weight$trstep)
```

Analiza wybranych zmiennych pod względem braków danych pokazała, że w trzech wybranych zmiennych pojawiły się braki danych. Jednakże są one dopuszczalne, by je uzupełnić przeprowadziliśmy w kolejnym etapie inputację danych, do imputacji posłużyliśmy się medianą.

#imputacja

```
imp_polintr<- median(ess_weight$polintr, na.rm = TRUE)
```

```
ess_weight <- ess_weight %>%
  mutate(polintr=if_else(is.na(polintr), imp_polintr, polintr))
```

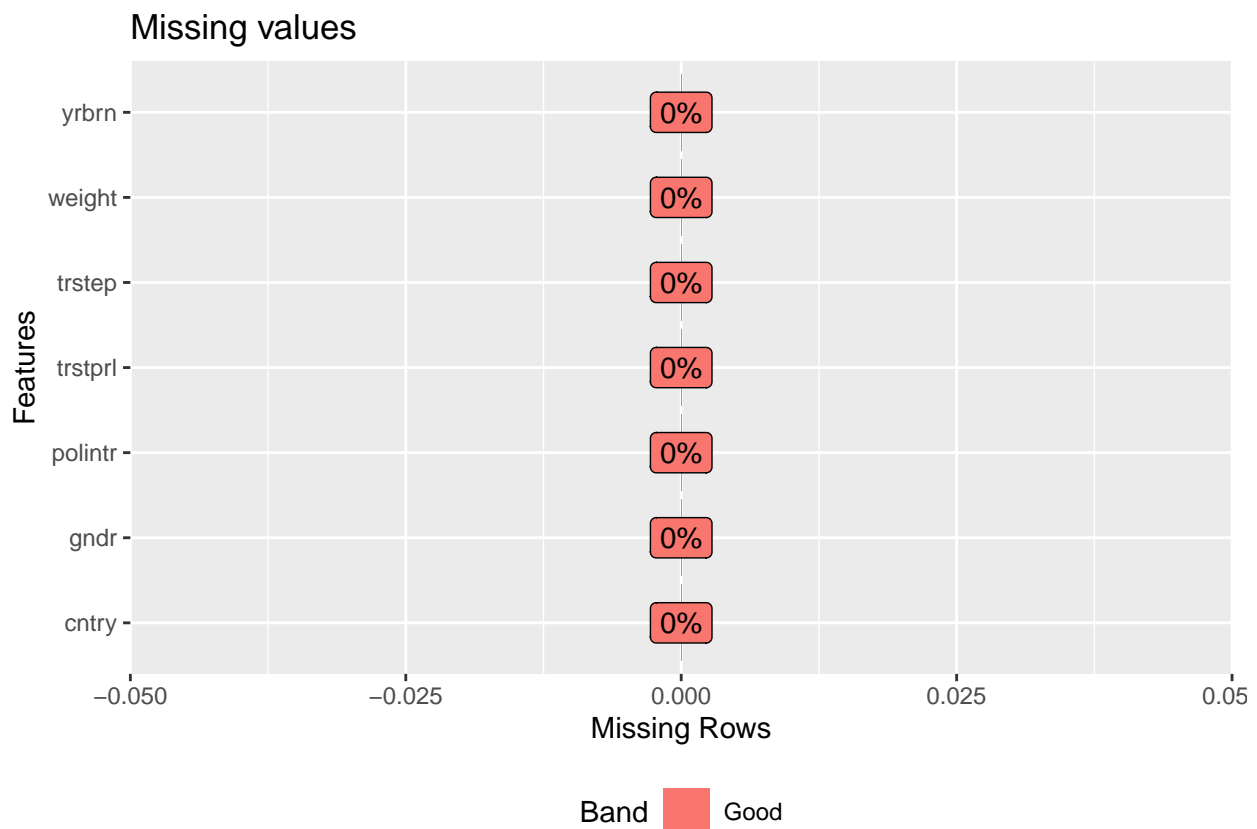
```
imp_trstprl<- median(ess_weight$trstprl, na.rm = TRUE)
ess_weight <- ess_weight %>%
  mutate(trstprl=if_else(is.na(trstprl), as.numeric(imp_trstprl), as.numeric(trstprl)))
```

```
imp_trstep<- median(ess_weight$trstep, na.rm = TRUE)

ess_weight <- ess_weight %>%
  mutate(trstep=if_else(is.na(trstep), imp_trstep, trstep))
```

Kontrolne sprawdzenie, czy pozbyliśmy się braków danych

```
ess_weight %>%
  select(variable_list) %>%
  plot_missing(
    missing_only = FALSE,
    title = "Missing values",
    ggtheme = theme_gray(),
    theme_config = list(legend.position = c("bottom"))
  )
```



Na podstawie powyższego wykresu możemy stwierdzić, że wszystkie braki danych zostały uzupełnione. Następnym krokiem jest zadeklarowanie **schematu losowania**.

```
ess_selected<-ess_weight %>%
  select(variable_list)

ess_selected<-ess_selected %>%
  mutate(age=2018-yrbrn)

schemat <-
  svydesign(
    id = ~1 ,
    data = ess_selected ,
```

```
weight = ~weight ,
strata = NULL
)
```

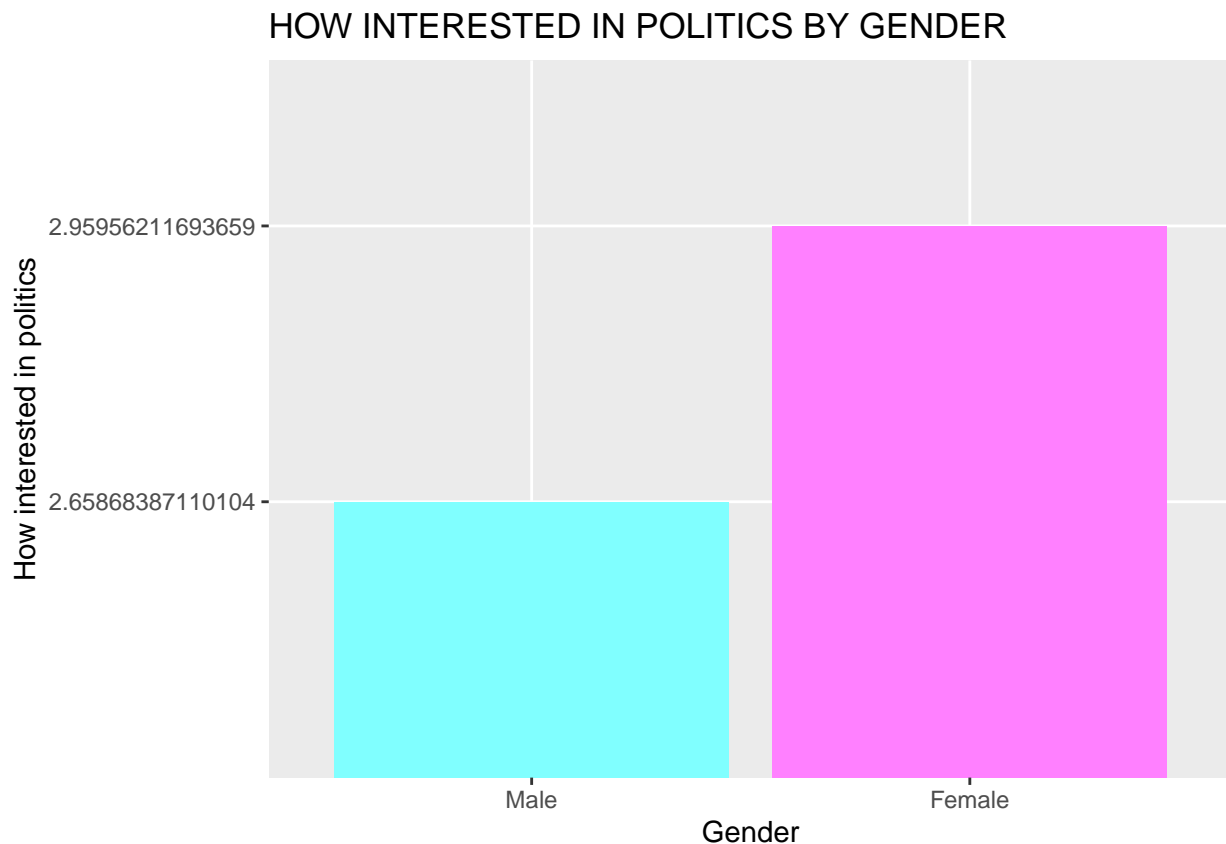
Sprawdzamy w podziale na płeć jak bardzo respondenci są zainteresowani polityką

Im mniej tym bardziej zainteresowany(a), gndr = 1 to mężczyzna

```
polintr_gndr <- svyby(formula = ~polintr, by = ~gndr, design = schemat,
FUN = svymean, na.rm = TRUE) %>%
mutate(cv=se/polintr)
```

Przedstawienie wykłów za pomocą wykresu kolumnowego:

```
ggplot(data = polintr_gndr, aes(x = haven::as_factor(gndr),
y = haven::as_factor(polintr))) +
geom_col(fill = cm.colors(2))+
ylab("How interested in politics") +
xlab("Gender") +
ggtitle("HOW INTERESTED IN POLITICS BY GENDER") +
ggsave('HOW INTERESTED IN POLITICS BY GENDER.png')
```

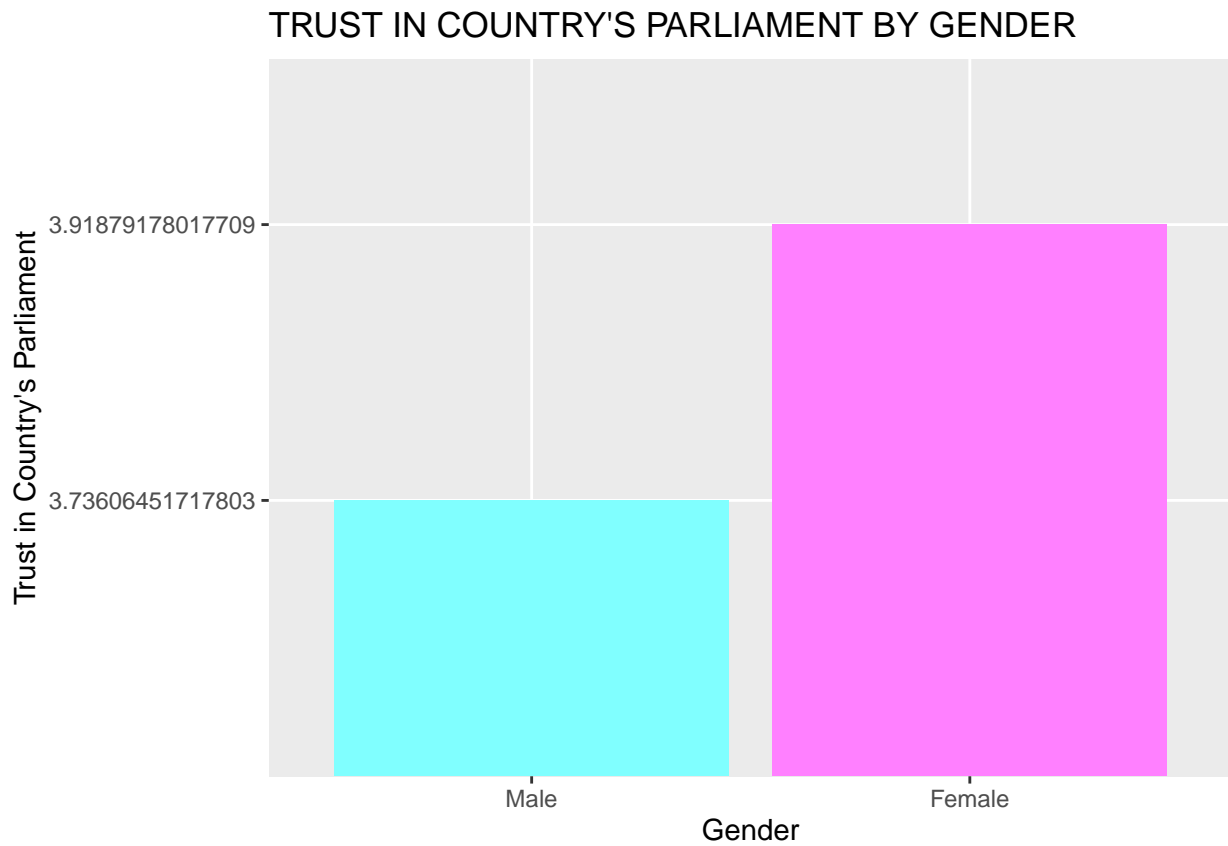


Mając na uwadze fakt, że mniejsza im mniejsza wartość tym większe zainteresowanie polityką, możemy stwierdzić, że zdecydowanie więcej mężczyzn jest zainteresowanych polityką.

Następnie sprawdzimy zaufanie do parlamentu krajowego w podziale na płeć. Im większa wartość tym zaufanie jest większe.

```
trstprl_gndr <- svyby(formula = ~trstprl, by = ~gndr, design = schemat,
FUN = svymean, na.rm = TRUE) %>%
mutate(cv=se/trstprl)
```

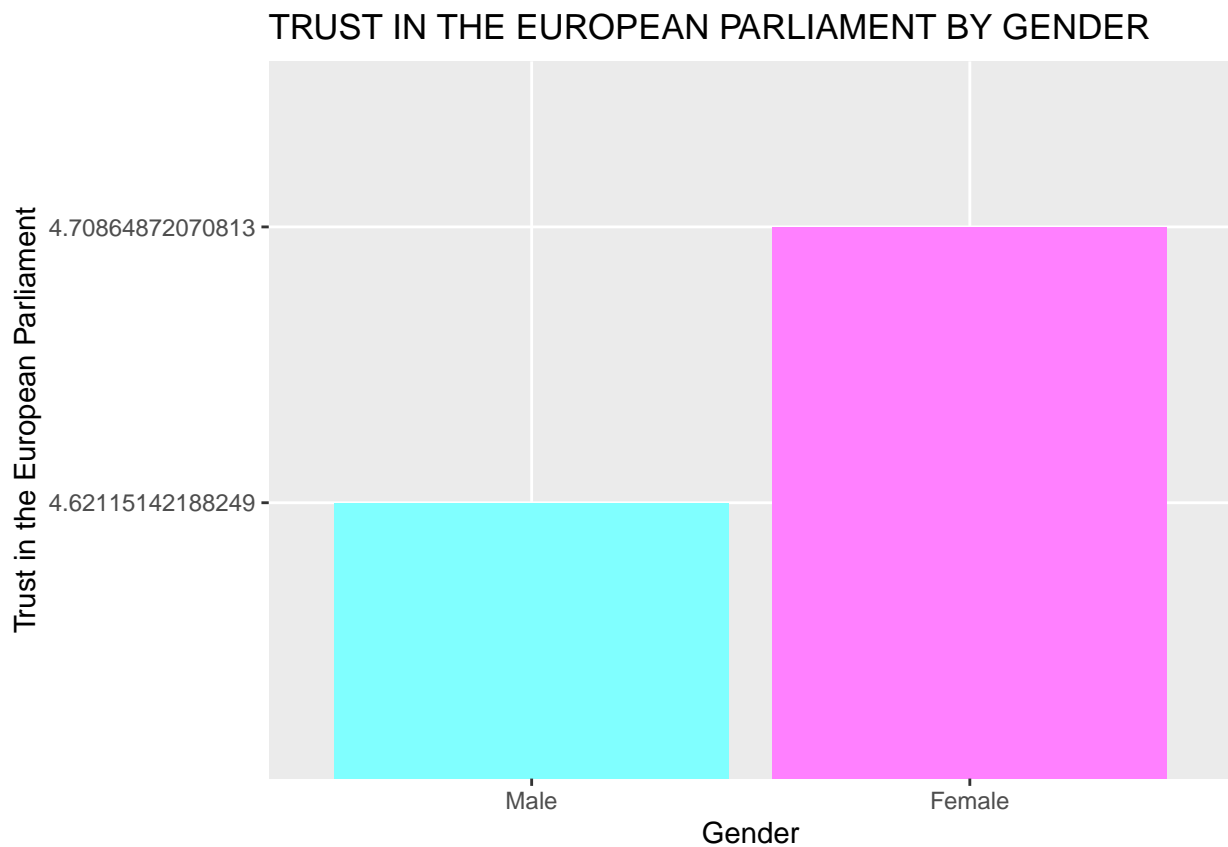
```
ggplot(data = trstprl_gndr, aes(x = haven::as_factor(gndr),
                                y = haven::as_factor(trstprl))) +
  geom_col(fill = cm.colors(2))+
  ylab("Trust in Country's Parliament") +
  xlab("Gender") +
  ggtitle("TRUST IN COUNTRY'S PARLIAMENT BY GENDER") +
  ggsave('TRUST IN COUNTRYS PARLIAMENT BY GENDER.png')
```



Na powyższym wykresie widać, że zdecydowanie wyższym zaufaniem do parlamentu krajowego mają kobiety. Następnie sprawdzimy zaufanie do Parlamentu Europejskiego w podziale na płeć. Im większa wartość tym zaufanie jest większe.

```
trstep_gndr <- svyby(formula = ~trstep, by = ~gndr, design = schemat,
                     FUN = svymean, na.rm = TRUE) %>%
  mutate(cv=se/trstep)
```

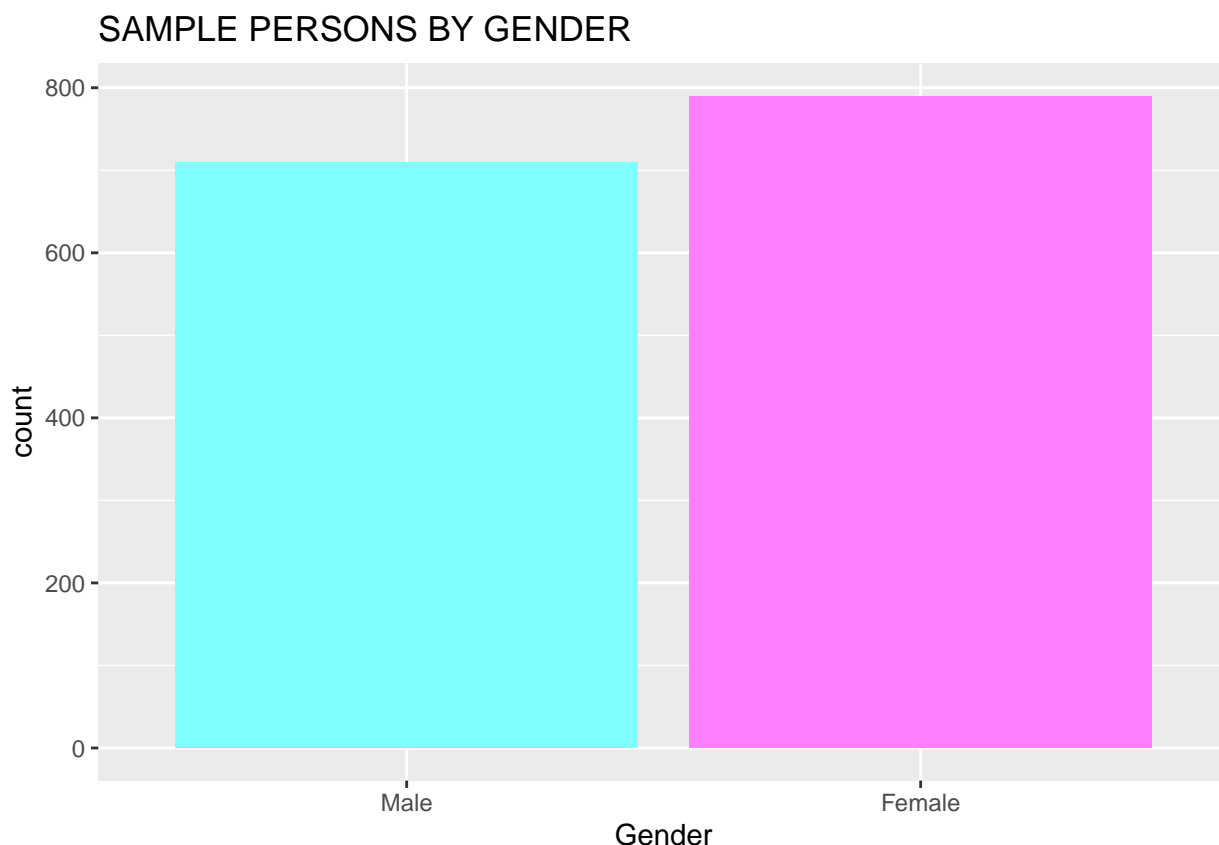
```
ggplot(data = trstep_gndr, aes(x = haven::as_factor(gndr),
                                y = haven::as_factor(trstep))) +
  geom_col(fill = cm.colors(2))+
  ylab("Trust in the European Parliament") +
  xlab("Gender") +
  ggtitle("TRUST IN THE EUROPEAN PARLIAMENT BY GENDER") +
  ggsave('TRUST IN THE EUROPEAN PARLIAMENT BY GENDER.png')
```



Tu znów zdecydowanie większe zaufanie wykazują kobiety.

Udział w badaniu ze względu na płeć

```
ggplot(data=ess_selected, aes(x = haven::as_factor(gndr))) +
  geom_bar(fill = cm.colors(2)) +
  xlab("Gender") +
  ggtitle("SAMPLE PERSONS BY GENDER") +
  ggsave("SAMPLE PERSONS BY GENDER.png")
```



Ostatnim elementem jest sprawdzenie jak zmienia się zainteresowanie polityką ze względu na płeć.

#Policzenie średniego zainteresowania polityką ze względu na wiek i płeć

```
polint_gndr_age<-svyby(formula = ~polintr, ~interaction(gndr,age), design = schemat, data=ess_selected,
```

#Ustawienie danych jako data frame

```
polint_gndr_age<- as.data.frame(polint_gndr_age)
```

#Zmiana typu danych kolumny

```
polint_gndr_age$`interaction(gndr, age)` <- as.character(polint_gndr_age$`interaction(gndr, age)`)
```

#Zmiana nazwy kolumny

```
polint_gndr_age <- polint_gndr_age %>%
```

```
  rename(
```

```
    'interaction_gndr_age'=`interaction(gndr, age)`
```

```
)
```

#Rozdzielenie kolumny interaction_gndr_age na dwie kolumny

```
polint_gndr_age<-polint_gndr_age %>%
```

```
separate(interaction_gndr_age, c('Gender', 'Age'))
```

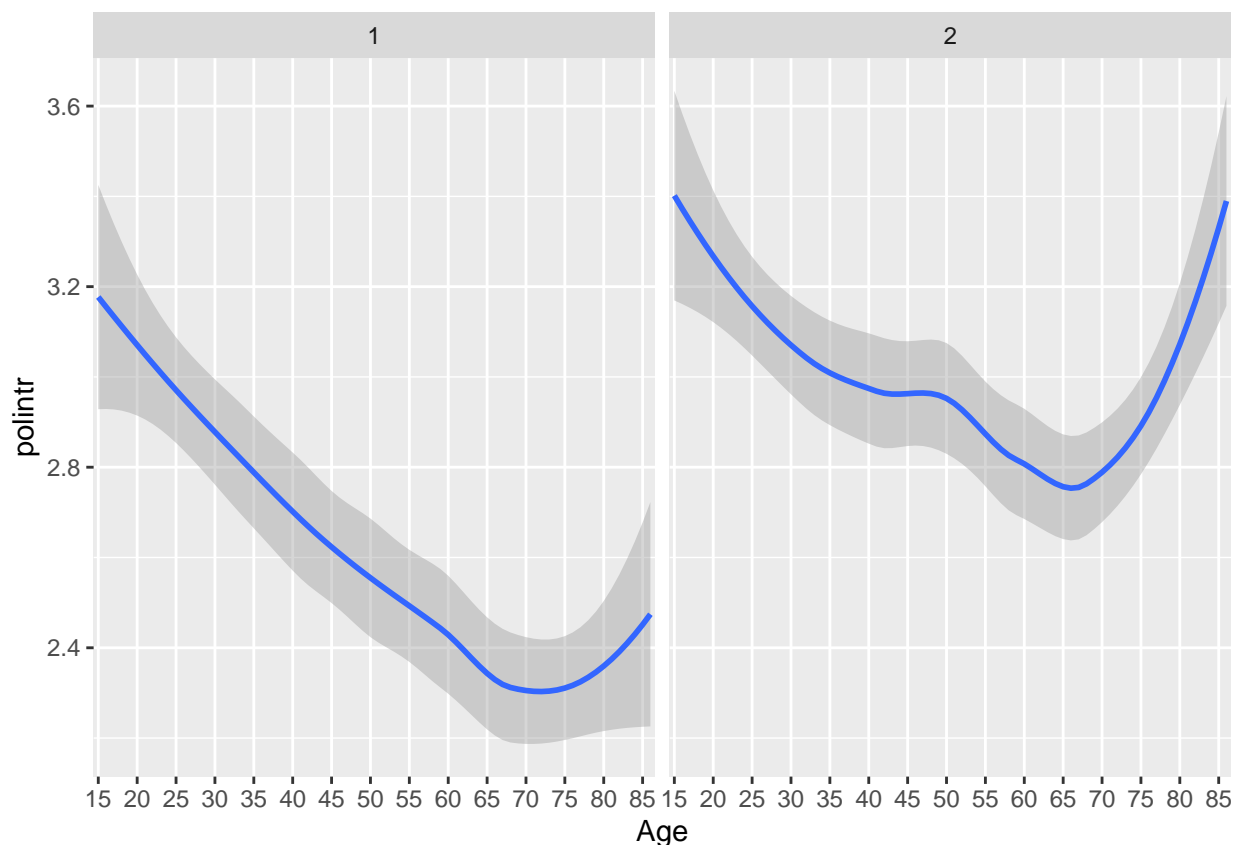
#Usunięcie nazw wierszy

```
rownames(polint_gndr_age) <- c()
```

#Funkcja, która przyjmuje wartość na wejściu. Co podana wartość będzie wyświetlana na wykresie

```
every_nth = function(n) {
  return(function(x) {x[c(TRUE, rep(FALSE, n - 1))]}})
}
```

```
ggplot(data=polint_gndr_age) +
  geom_smooth(mapping = aes(x = Age, y=polintr, group=1)) +
  facet_grid(~Gender) +
  scale_x_discrete(breaks = every_nth(n = 5))
```



Na poniższym wykresie możemy zauważyć, że wraz z wiekiem rośnie zainteresowanie polityką wśród respondentów (im mniejsza wartość tym większe zainteresowanie).

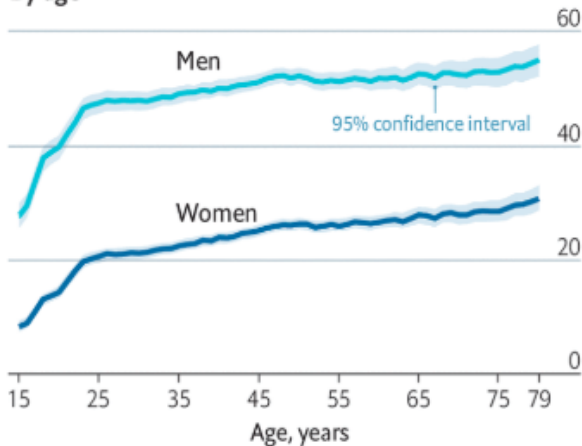
Rośnie do ok 70 roku życia wśród mężczyzn (1) jak i wśród kobiet (2).

Podobne wyniki zostały opublikowane przez The Economist W czerwcu 2019. W badaniu przeprowadzonym przez Marta Fraile okazało się, że mężczyźni bardziej interesują się polityką oraz, że zainteresowanie polityką rośnie wraz z wiekiem.

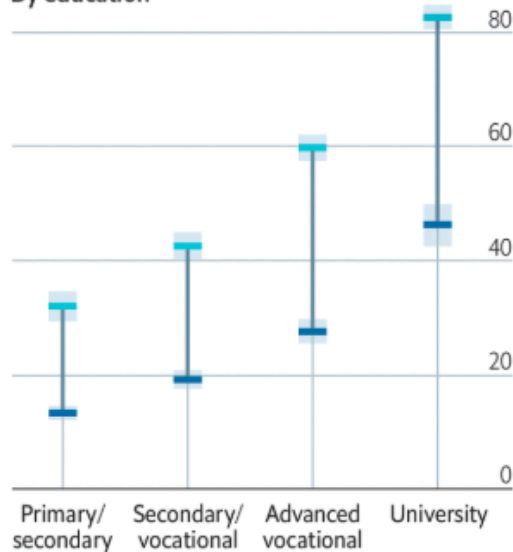
Polls apart

Britain, share of men and women declaring themselves interested in politics, 1991-2008, %

By age



By education



LINK DO BADANIA