Group Project

Evan McKee, Roman, Jordan

12/06/2022

Input data

```
## setting working directory
setwd("C:/git/statsGroupProject/R")
## input data
load("ces.RData")

library(ggplot2)
```

Variables to factor

```
data <- load("ces.RData")


ces$abortionRaw <- ces$abortion
ces$minwageRaw <- ces$minwage


ces$polengage <- ces$voted + ces$meeting + ces$sign +
    ces$campaign + ces$protest + ces$contact + ces$donate


## Region
ces$region <- factor(ces$region,
    levels = 1:4,
    labels = c("Northeast", "Midwest", "South", "West")
)

## Gender
ces$gender <- factor(ces$gender,
    levels = 1:2,
    labels = c("Male", "Female")
)

## Education
ces$educ <- factor(ces$educ,
    levels = 1:6,
    labels = c(
```

```r
        "No HS", "High School Graduate", "Some college",
        "2-year degree", "4-year degree", "Post-grad"
    )
)

## Race
ces$race <- factor(ces$race,
    levels = 1:8,
    labels = c(
        "White", "Black", "Hispanic", "Asian",
        "Native American", "Middle Eastern", "Two or more races", "Other"
    )
)

## Hispanic
ces$hispanic <- factor(ces$hispanic,
    levels = 1:2,
    labels = c("Yes", "No")
)

## Homeowner
ces$ownhome <- factor(ces$ownhome,
    levels = 1:3,
    labels = c("Own", "Rent", "Other")
)

## Neighborhood Type
ces$urbancity <- factor(ces$urbancity,
    levels = 1:5,
    labels = c("City", "Suburb", "Town", "Rural Area", "Other")
)

## Union
ces$unionhh <- factor(ces$unionhh,
    levels = 1:2,
    labels = c("Yes", "No")
)

## Religious
ces$religious <- factor(ces$religious,
    levels = 1:4,
    labels = c(
        "Very important", "Not too important",
        "Not too important", "Not at all important"
    )
)

## Family income
ces$faminc2 <- factor(ces$faminc,
    levels = 1:16,
    labels = c(
        "< 10,000", "10,000 - 19,999", "20,000 - 29,999",
        "30,000 - 39,999", "40,000 - 49,999",
```

```r
        "50,000 - 59,999", "60,000 - 69,999", "70,000 - 79,999",
        "80,000 - 99,999", "100,000-119,000", "120,000 - 149,999",
        "150,000 - 199,999", "200,000 - 249,999", "250,000 - 349,999",
        "350,000 - 499,999", "500,000 or more"
    )
)

## Voted
ces$voted <- factor(ces$voted,
    levels = 1:2,
    labels = c("Yes", "No")
)

## Meeting
ces$meeting <- factor(ces$meeting,
    levels = 1:2,
    labels = c("Yes", "No")
)

## Signed
ces$sign <- factor(ces$sign,
    levels = 1:2,
    labels = c("Yes", "No")
)

## Campaign
ces$campaign <- factor(ces$campaign,
    levels = 1:2,
    labels = c("Yes", "No")
)

## Protest
ces$protest <- factor(ces$protest,
    levels = 1:2,
    labels = c("Yes", "No")
)

## Contact
ces$contact <- factor(ces$contact,
    levels = 1:2,
    labels = c("Yes", "No")
)

## Donate
ces$donate <- factor(ces$donate,
    levels = 1:2,
    labels = c("Yes", "No")
)

## ideo5
ces$ideo5 <- factor(ces$ideo5,
    levels = 1:5,
    labels = c(
```

```r
        "Very liberal", "Liberal", "Moderate",
        "Conservative", "Very Conservative"
    )
)

## pid3
ces$pid3 <- factor(ces$pid3,
    levels = 1:5,
    labels = c(
        "Democrat", "Republican", "Independent",
        "Other", "Not Sure"
    )
)

## pid7
ces$pid7 <- factor(ces$pid7,
    levels = 1:7,
    labels = c(
        "Strong Democrat", "Not very strong Democrat",
        "Lean Democract", "Independent", "Lean Republican",
        "Not very strong Republican", "Strong Republican"
    )
)

## medicare
ces$medicare <- factor(ces$medicare,
    levels = 1:2,
    labels = c("Support", "Oppose")
)

## ACA
ces$ACA <- factor(ces$ACA,
    levels = 1:2,
    labels = c("Support", "Oppose")
)

## abortion
ces$abortion <- factor(ces$abortion,
    levels = 1:2,
    labels = c("Support", "Oppose")
)

## EPA
ces$EPA <- factor(ces$EPA,
    levels = 1:2,
    labels = c("Support", "Oppose")
)

## Increase the number of police
ces$police_incr <- factor(ces$police_incr,
    levels = 1:2,
    labels = c("Support", "Oppose")
)
```

```r
## Decrease the number of police
ces$police_decr <- factor(ces$police_decr,
    levels = 1:2,
    labels = c("Support", "Oppose")
)

## Trade Tariffs
ces$trade <- factor(ces$trade,
    levels = 1:2,
    labels = c("Support", "Oppose")
)

## Raise minimum wage
ces$minwage <- factor(ces$minwage,
    levels = 1:2,
    labels = c("Support", "Oppose")
)

## Work requirement for food stamps
ces$work_req <- factor(ces$work_req,
    levels = 1:2,
    labels = c("Support", "Oppose")
)

ces[1:2, ]
```

```
## # A tibble: 2 x 37
##    ...1 gender educ      race  hispa~1 region medic~2 ACA   abort~3 EPA   polic~4
##   <dbl> <fct>  <fct>     <fct> <fct>   <fct>  <fct>   <fct> <fct>   <fct> <fct>
## 1     1 Male   2-year ~  White No      North~ <NA>    <NA>  <NA>    Supp~ Support
## 2     2 Female Post-gr~  White No      South  Support <NA>  <NA>    Supp~ <NA>
## # ... with 26 more variables: police_decr <fct>, trade <fct>, minwage <fct>,
## #   work_req <fct>, residency <dbl>, pid3 <fct>, pid7 <fct>, ownhome <fct>,
## #   urbancity <fct>, unionhh <fct>, religious <fct>, ideo5 <fct>, faminc <dbl>,
## #   voted <fct>, meeting <fct>, sign <fct>, campaign <fct>, protest <fct>,
## #   contact <fct>, donate <fct>, commonweight <dbl>, age <dbl>,
## #   abortionRaw <dbl>, minwageRaw <dbl>, polengage <dbl>, faminc2 <fct>, and
## #   abbreviated variable names 1: hispanic, 2: medicare, 3: abortion, ...
```

Histogram

```r
ggplot(ces, aes(polengage, y = ..density..)) +
    geom_histogram(color = "#a8a8a8", fill = "grey", binwidth = 1) +
    labs(
        title = "Figure 1: Distribution of Political Engagment",
        ## caption = "Source: Gapminder dataset",
        x = "Political Activities",
        y = "Count"
    ) +
    theme_classic() +
    theme(
```

```
        plot.title = element_text(size = 20, face = "bold"),
        plot.subtitle = element_text(size = 100, face = "bold"),
        plot.caption = element_text(face = "italic"),
        axis.title.x = element_text(size = 20),
        axis.text.x = element_text(size = 17),
        axis.text.y = element_text(size = 17),
        axis.title.y = element_text(size = 20)
    )
```

Warning: Removed 9449 rows containing non-finite values (stat_bin).

Figure 1: Distribution of Political Engagmen
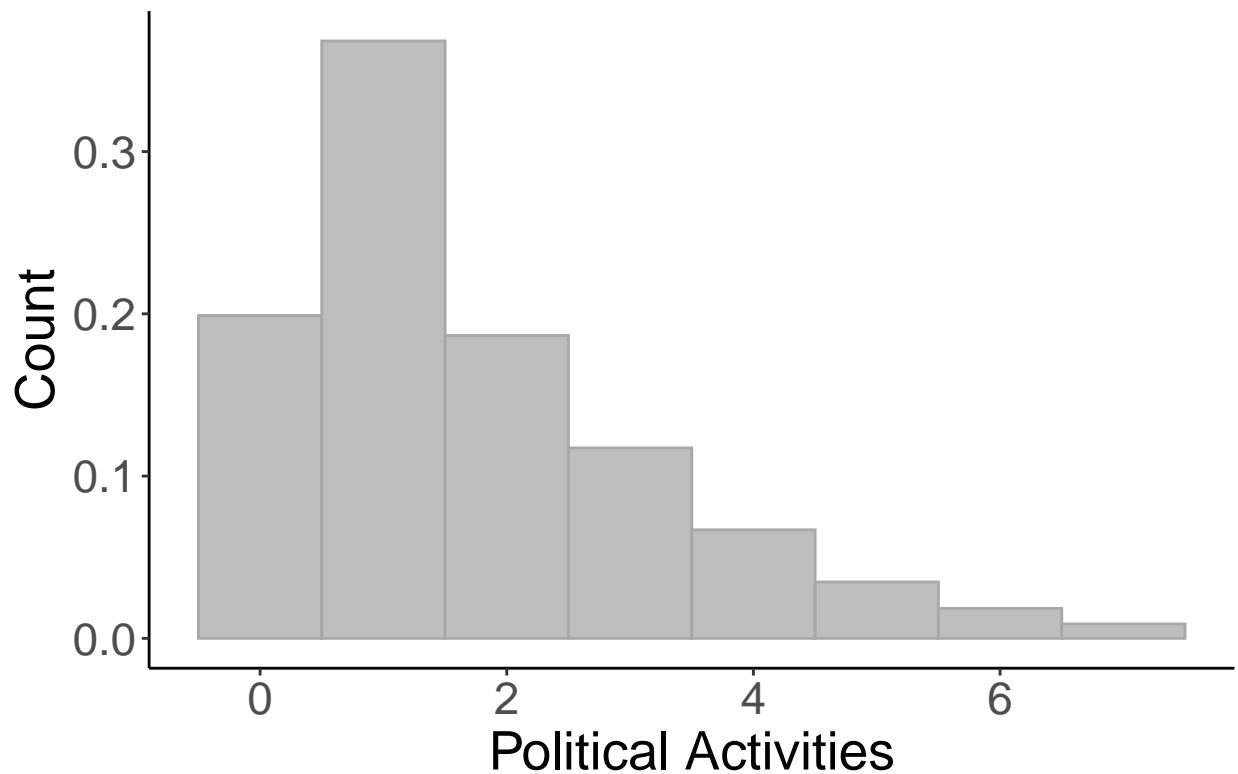

```
ggsave(filename = "hist.jpeg")
```

Saving 6.5 x 4.5 in image

Warning: Removed 9449 rows containing non-finite values (stat_bin).

Part 2

Question 1

```
## Education
Lm_educ <- lm(polengage ~ educ, data = ces)
Lm_educ
```

```
##
## Call:
## lm(formula = polengage ~ educ, data = ces)
##
## Coefficients:
##             (Intercept)  educHigh School Graduate        educSome college
##                  0.8376                    0.2357                  0.8268
##        educ2-year degree        educ4-year degree          educPost-grad
##                  0.8284                    1.1835                  1.6016
```

```
plot(polengage ~ educ,
    data = ces,
    xlab = "Education level",
    ylab = "Number of Political Activities",
    main = "Figure 2: Impact of Education Level on Political Participation"
)

abline(lm(polengage ~ educ,
    data = ces,),col='red')
```

```
## Warning in abline(lm(polengage ~ educ, data = ces, ), col = "red"): only using
## the first two of 6 regression coefficients
```

Figure 2: Impact of Education Level on Political Participation
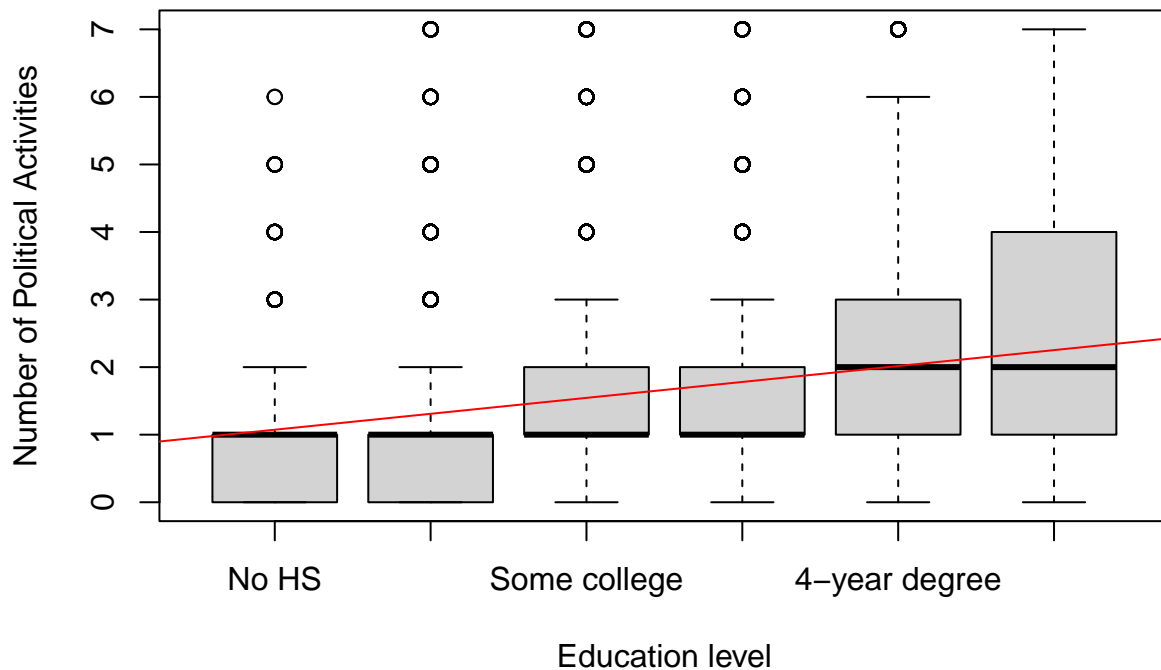
Question 2

```
ces$faminc <- (ifelse(ces$faminc == 2 | ces$faminc == 3, 2, ces$faminc))
ces$faminc <- (ifelse(ces$faminc == 4 | ces$faminc == 5, 3, ces$faminc))
ces$faminc <- (ifelse(ces$faminc == 6 | ces$faminc == 7, 4, ces$faminc))
ces$faminc <- (ifelse(ces$faminc == 8 | ces$faminc == 9, 5, ces$faminc))
ces$faminc <- (ifelse(ces$faminc == 10 | ces$faminc == 11, 6, ces$faminc))
ces$faminc <- (ifelse(ces$faminc == 12 | ces$faminc == 13, 7, ces$faminc))
ces$faminc <- (ifelse(ces$faminc == 14 | ces$faminc == 15, 8, ces$faminc))
ces$faminc <- (ifelse(ces$faminc == 16, 9, ces$faminc))
## Family income
ces$faminc <- factor(ces$faminc,
    levels = 1:9,
    labels = c(
        "< 10,000", "10,000 - 29,999", "30,000 - 49,999",
        "50,000 - 69,999", "70,000 - 99,999", "100,000 - 149,999", "150,000 - 249,999",
        "250,000 - 499,999", "500,000 or more"
    )
)

Lm_faminc <- lm(polengage ~ faminc, data = ces)
Lm_faminc
```

```
##
```

```
## Call:
## lm(formula = polengage ~ faminc, data = ces)
##
## Coefficients:
##            (Intercept)      faminc10,000 - 29,999    faminc30,000 - 49,999
##                 0.9226                     0.3327                   0.5624
##    faminc50,000 - 69,999    faminc70,000 - 99,999  faminc100,000 - 149,999
##                 0.7802                     0.9724                   1.2534
## faminc150,000 - 249,999  faminc250,000 - 499,999     faminc500,000 or more
##                 1.4353                     1.6558                   1.4203
```

```
plot(polengage ~ faminc,
    data = ces,
    xlab = "Family Income",
    ylab = "Number of Political Activities",
    main = "Figure 4: Impact of Family Income on Political Participation"
)

abline(lm(polengage ~ faminc,
    data = ces,),col='red')
```
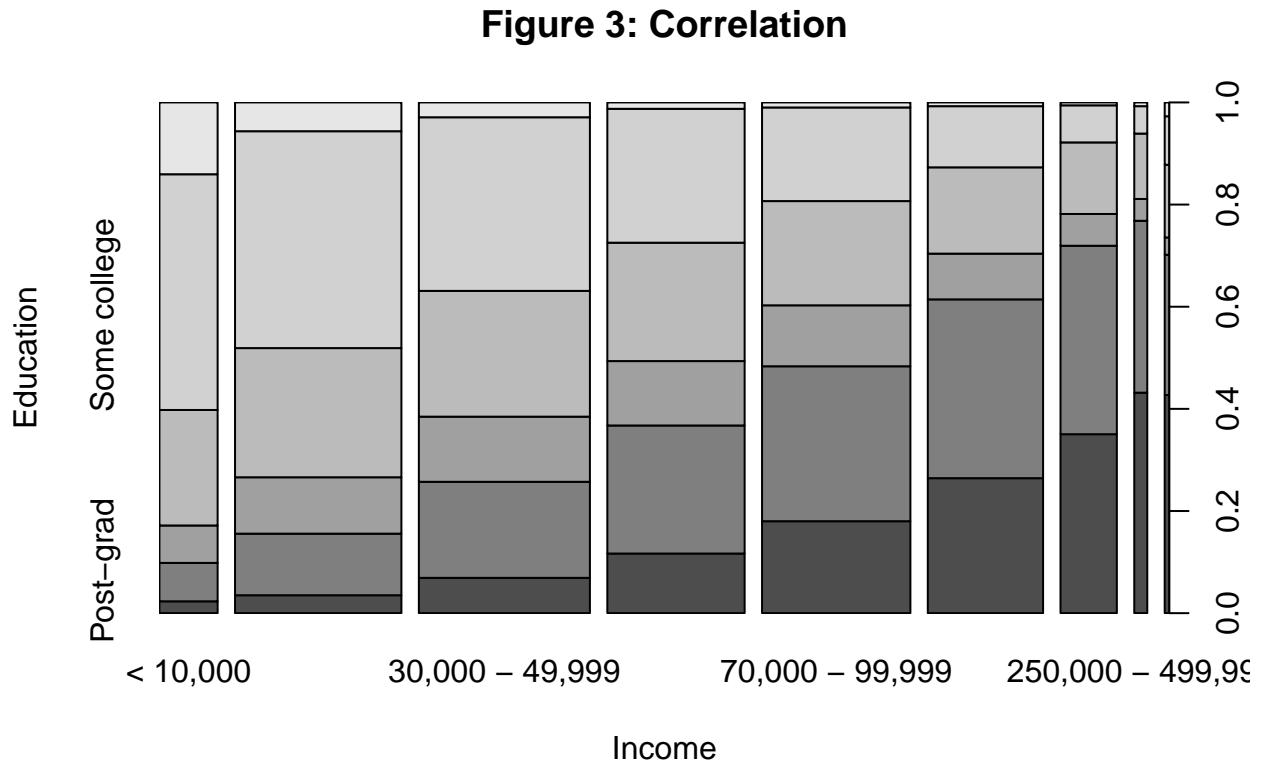
```
## Warning in abline(lm(polengage ~ faminc, data = ces, ), col = "red"): only using
## the first two of 9 regression coefficients
```

Figure 4: Impact of Family Income on Political Participation
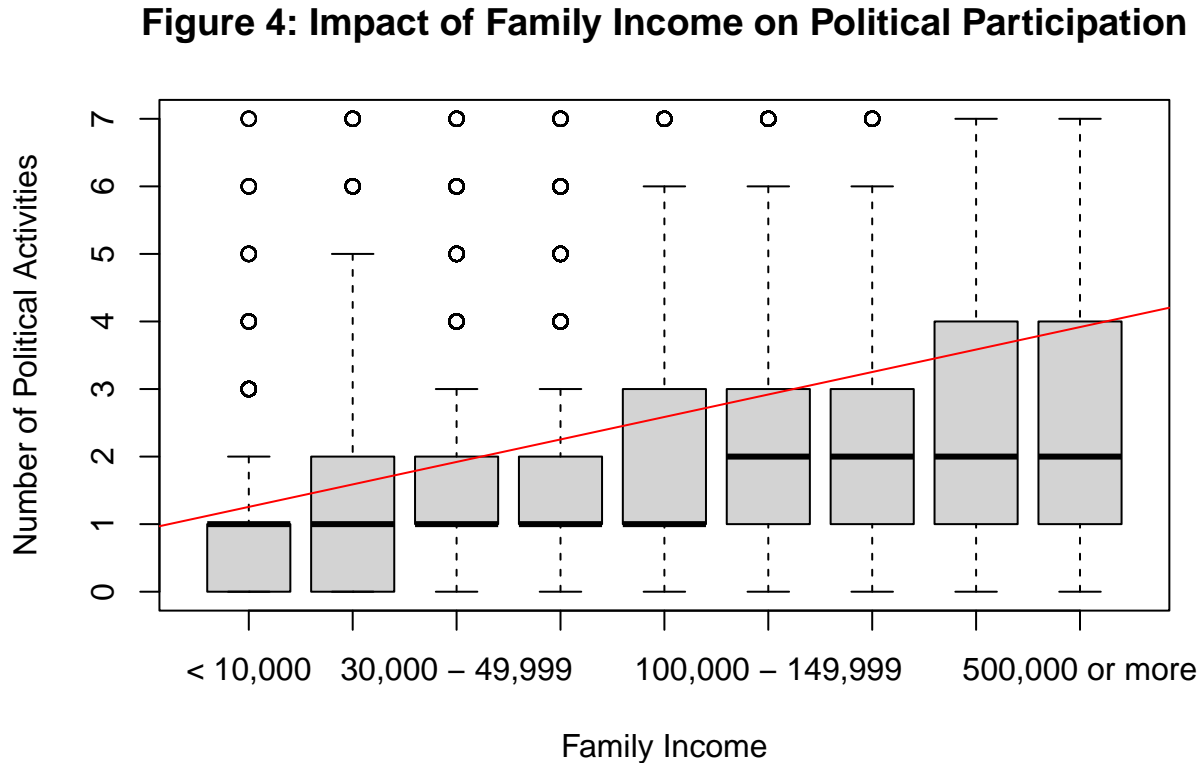
Question 3

```
plot(ces$faminc, ces$educ, xlab="Income", ylab="Education", main = "Figure 3: Correlation", cex=3)
```


Figure 3: Correlation

```
ces$educ <- as.numeric(ces$educ)
ces$faminc <- as.numeric(ces$faminc)
cor(ces$faminc, ces$educ, use = "complete.obs")
```

```
## [1] 0.4352379
```

Question 4

```
## Trying to set party identification to part of a party or not##
ces$gender <- (ifelse(ces$gender == 1, 1, 0)) # 1 if you are male, 0 if female
ces$race <- (ifelse(ces$race == 1, 1, 0)) # 1 if you are white, 0 if non-white
ces$pid3 <- (ifelse(ces$pid3 == 1 | ces$pid3 == 2, 1, 0)) # 1 if you identify as a member of a politica
Lm_our <- lm(polengage ~ gender + educ + race + faminc + pid3, data = ces)
summary(Lm_our)
```

```
##
## Call:
```

```
## lm(formula = polengage ~ gender + educ + race + faminc + pid3,
##     data = ces)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9276 -0.9993 -0.2614  0.7386  6.1480
##
## Coefficients: (3 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.223591   0.020054   11.15   <2e-16 ***
## gender            NA         NA      NA       NA
## educ        0.245989   0.004862   50.60   <2e-16 ***
## race              NA         NA      NA       NA
## faminc      0.136453   0.004118   33.14   <2e-16 ***
## pid3              NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.425 on 46615 degrees of freedom
##   (14382 observations deleted due to missingness)
## Multiple R-squared:  0.1164, Adjusted R-squared:  0.1164
## F-statistic:  3071 on 2 and 46615 DF,  p-value: < 2.2e-16
```

```
Lm_educr <- lm(polengage ~ educ + race, data = ces)
Lm_educr
```

```
##
## Call:
## lm(formula = polengage ~ educ + race, data = ces)
##
## Coefficients:
## (Intercept)         educ         race
##      0.5420       0.3084           NA
```

```
Lm_educg <- lm(polengage ~ educ + gender, data = ces)
Lm_educg
```

```
##
## Call:
## lm(formula = polengage ~ educ + gender, data = ces)
##
## Coefficients:
## (Intercept)         educ       gender
##      0.5420       0.3084           NA
```

```
Lm_educpid <- lm(polengage ~ educ + pid3, data = ces)
Lm_educpid
```

```
##
## Call:
## lm(formula = polengage ~ educ + pid3, data = ces)
##
```

```
## Coefficients:
## (Intercept)         educ         pid3
##      0.5420       0.3084          NA
```

```
## Part 3


tests <- list()



tests[[1]] <- t.test(ces$abortionRaw[ces$polengage >= 2],
       ces$abortionRaw[ces$polengage < 2])



tests[[2]] <- t.test(ces$minwageRaw[ces$polengage >= 2],
       ces$minwageRaw[ces$polengage < 2])


   table <-    sapply(tests, function(x) {
     c(x$estimate[1],
       x$estimate[2],
       ci.lower = x$conf.int[1],
       ci.upper = x$conf.int[2],
       p.value = x$p.value,
       x$statistic,
       x$parameter)
})



colnames(table) <- c("Abortion", "Min Wage")
```

```
##
## Please cite as:

##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Tue, Dec 06, 2022 - 8:25:12 PM

Table 1: Education and Income

	Dependent variable:	
	polengage	
	(1)	(2)
educHigh School Graduate	0.236***	
	(0.043)	
educSome college	0.827***	
	(0.044)	
educ2-year degree	0.828***	
	(0.046)	
educ4-year degree	1.184***	
	(0.043)	
educPost-grad	1.602***	
	(0.044)	
faminc10,000 - 29,999		0.333***
		(0.034)
faminc30,000 - 49,999		0.562***
		(0.034)
faminc50,000 - 69,999		0.780***
		(0.035)
faminc70,000 - 99,999		0.972***
		(0.034)
faminc100,000 - 149,999		1.253***
		(0.035)
faminc150,000 - 249,999		1.435***
		(0.040)
faminc250,000 - 499,999		1.656***
		(0.062)
faminc500,000 or more		1.420***
		(0.098)
Constant	0.838***	0.923***
	(0.041)	(0.030)
Observations	51,551	46,618
R^2	0.098	0.069
Adjusted R^2	0.098	0.069
Residual Std. Error	1.438 (df = 51545)	1.463 (df = 46609)
F Statistic	1,123.923*** (df = 5; 51545)	430.134*** (df = 8; 46609)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2: T-Tests

	Abortion	Min Wage
mean of x	0.640	0.661
mean of y	0.573	0.674
ci.lower	0.059	-0.021
ci.upper	0.076	-0.004
p.value	0	0.003
t	15.506	-3.000
df	$48,707.200$	$47,773.880$