# Group Project

Evan McKee, Roman, etc.

12/06/2022

## Input data

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

library(ggplot2)
```

## Variables to factor

```
## input data
load("ces.RData")

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


## 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("No HS", "High School Graduate", "Some college", "2-year degree", "4-year degree", "Post

## Race
ces$race <- factor(ces$race,
    levels = 1:8,
```

```
    labels = c("White", "Black", "Hispanic", "Asian", "Native American", "Middle Eastern", "Two or more
)

## 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")
)

ces[1:4, ]
```

```
## # A tibble: 4 x 34
##    ...1 gender educ     race  hispa~1 region medic~2   ACA abort~3   EPA polic~4
##   <dbl> <fct>  <fct>    <fct> <fct>   <fct>    <dbl> <dbl>   <dbl> <dbl>   <dbl>
## 1     1 Male   2-year ~ White No      North~       0     0       0     1       1
## 2     2 Female Post-gr~ White No      South        1     0       0     1       0
## 3     3 Female 4-year ~ White No      Midwe~       1     0       1     1       0
## 4     4 Female 4-year ~ White No      North~       1     1       1     1       0
## # ... with 23 more variables: police_decr <dbl>, trade <dbl>, minwage <dbl>,
## #   work_req <dbl>, residency <dbl>, pid3 <dbl>, pid7 <dbl>, ownhome <fct>,
## #   urbancity <fct>, unionhh <fct>, religious <fct>, ideo5 <dbl>, faminc <dbl>,
## #   voted <dbl>, meeting <dbl>, sign <dbl>, campaign <dbl>, protest <dbl>,
## #   contact <dbl>, donate <dbl>, commonweight <dbl>, age <dbl>,
## #   polengage <dbl>, and abbreviated variable names 1: hispanic, 2: medicare,
## #   3: abortion, 4: police_incr
```

### Histogram

```
ggplot(ces, aes(polengage)) +
  geom_histogram(color = "#a8a8a8", fill = "#000000", binwidth = 1) +
  labs(
    title = "Distribution of Political Engagment",
    ## caption = "Source: Gapminder dataset",
    x = "Political Activities",
    y = "Count"
  ) +
  theme_classic() +
  theme(
    plot.title = element_text(color = "#0099F8", size = 16, face = "bold"),
    plot.subtitle = element_text(size = 10, face = "bold"),
    plot.caption = element_text(face = "italic")
  )
```

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