Clean Data - YouGov Survey

Weifang Xu, Taylor Chewning, and Qing Wang

August 6, 2024

```
## Clean the working environment and set up the working directory
rm(list = ls())
setwd("/Users/qingwang/Downloads/Data Replication")

# load the libraries
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.1
```

```
## Warning: package 'stringr' was built under R version 4.3.1
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to bec
```

```
library(haven)

# import the original dataset
fsu <- read_csv("YouGov/data_YouGov.csv")
```

```
## Rows: 1000 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (34): alliance_DV1, exp_4, sex, educ, race, birthyr, income, alliance_DV...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# re-code the DV into binary var
df_ces <- fsu %>% mutate(attack = ifelse(alliance_DV1 > 3, 100, 0), # 1 = favor
                         attack_cont = (alliance_DV1 * 25)-25, #rescale from 0-100, higher val
                         alliance = ifelse(exp_4 > 2, 1, 0), # 1 = alliance
```

```
                        hmrts = ifelse(exp_4 == 1 | exp_4 == 3, 1, 0)) %>% # 1 = violate
  mutate(male = ifelse(sex == 1, 1, 0), # 1 = man
         edu4 = as.numeric(case_when(educ == 1  ~ '1',
                                     educ == 2  ~ '1',
                                     educ == 3 ~ '2',
                                     educ == 4 ~ '2',
                                     educ == 5 ~ '3',
                                     educ == 6 ~ '4')), # 1=HS and under, 2=some college, 3=4-
         edu4 = (edu4 - 1)/3 , #rescale education
         white = ifelse(race == 1, 1, 0),
         age = (2022-birthyr),
#collapse age into categories to match PureSpectrum
         age_cat = case_when(age <= 29 ~ '1',
                             age >= 30 & age <= 39 ~ '2',
                             age >= 40 & age <= 49 ~ '3',
                             age >= 50 & age <= 59 ~ '4',
                             age >= 60 & age <= 69 ~ '5',
                             age >= 70 ~ '6'),
         age_cat = ifelse(age_cat == 1, 23.5, #average of age categories, or min
                   ifelse(age_cat == 2, 34.5,
                   ifelse(age_cat == 3, 44.5,
                   ifelse(age_cat == 4, 54.5,
                   ifelse(age_cat == 5, 64.5,
                   ifelse(age_cat == 6, 70, NA))))),
#collapse income categories to match PureSpectrum
         income_cat = case_when(income == 1 ~ '1',
                             income == 2 ~ '1',
                             income == 3 ~ '1',
                             income == 4 ~ '2',
                             income == 5 ~ '2',
                             income == 6 ~ '2',
                             income == 7 ~ '2',
                             income == 8 ~ '3',
                             income == 9 ~ '3',
                             income == 10 ~ '4',
                             income == 11 ~ '4',
                             income == 12 ~ '4',
                             income == 13 ~ '5',
                             income == 14 ~ '5',
                             income == 15 ~ '5',
                             income == 16 ~ '5',),
         inc = ifelse(income_cat == 1, 30000, #average of income categories, min and max
               ifelse(income_cat == 2, 50000,
               ifelse(income_cat == 3, 85000,
               ifelse(income_cat == 4, 150000,
               ifelse(income_cat == 5, 200000, NA))))))%>%
  # convert the inc variable unit ($ to 10k$)
```

```r
  mutate(inc_10k = inc/10000)


# generate mediator vars
df_ces <- df_ces %>% mutate(threat = (alliance_DV4_1+ alliance_DV4_2 + alliance_DV4_3 + allianc
                            threat = (threat * 25)-25, #rescale from 0-100
                            success = (alliance_DV5_1 + alliance_DV5_2)/2,
                            success = (success * 25)-25, #rescale from 0-100
                            cost = (alliance_DV5_3 + alliance_DV5_4 + alliance_DV5_5 + allianc
                            cost = (cost * 25)-25, #rescale from 0-100
                            oblig = case_when(alliance_DV2 == 1 ~ 0,
                                              alliance_DV2 == 2 ~ 50,
                                              alliance_DV2 == 3 ~ 100), # 100 = US has moral o
                            immoral = case_when(alliance_DV3 == 1 ~ 100,
                                                alliance_DV3 == 2 ~ 0), # 100 = morally wrong
                            moral = (oblig - immoral + 100)/2) # 100 = moral to attack, 0 = im

glimpse(df_ces)
```

```
## Rows: 1,000
## Columns: 52
## $ alliance_DV1        <dbl> 4, 4, 4, 1, 3, 2, 3, 2, 2, 3, 1, 2, 5, 1, 1, 3, 1,~
## $ exp_4               <dbl> 4, 3, 1, 4, 3, 1, 1, 2, 4, 3, 2, 2, 2, 3, 2, 3, 3,~
## $ sex                 <dbl> 2, 2, 2, 1, 1, 2, 2, 2, 1, 1, 2, 2, 1, 1, 1, 2, 1,~
## $ educ                <dbl> 3, 2, 2, 5, 6, 5, 4, 6, 6, 2, 2, 4, 6, 5, 3, 2, 5,~
## $ race                <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 1,~
## $ birthyr             <dbl> 1952, 1999, 1963, 1953, 1983, 1999, 1997, 1961, 19~
## $ income              <dbl> 6, 6, 6, 6, 11, 3, 9, 9, 13, 5, 5, 97, 10, 1, 12, ~
## $ alliance_DV4_1      <dbl> 4, 1, 4, 3, 3, 2, 3, 3, 2, 3, 4, 1, 2, 2, 2, 4, 1,~
## $ alliance_DV4_2      <dbl> 4, 3, 4, 1, 2, 2, 3, 3, 1, 3, 3, 1, 2, 1, 1, 5, 1,~
## $ alliance_DV4_3      <dbl> 4, 2, 4, 2, 1, 1, 3, 2, 2, 4, 2, 1, 3, 1, 2, 5, 1,~
## $ alliance_DV4_4      <dbl> 4, 5, 4, 1, 1, 1, 3, 2, 1, 3, 2, 1, 2, 1, 2, 5, 1,~
## $ alliance_DV5_1      <dbl> 3, 1, 4, 5, 4, 5, 3, 2, 1, 3, 3, 3, 2, 4, 3, 3, 5,~
## $ alliance_DV5_2      <dbl> 4, 5, 4, 4, 1, 1, 3, 2, 1, 2, 3, 1, 4, 4, 1, 4, 5,~
## $ alliance_DV5_3      <dbl> 2, 1, 2, 4, 4, 2, 3, 3, 2, 2, 3, 3, 4, 4, 1, 5, 2,~
## $ alliance_DV5_4      <dbl> 3, 4, 2, 3, 2, 3, 3, 2, 3, 2, 2, 5, 3, 3, 1, 5, 3,~
## $ alliance_DV5_5      <dbl> 2, 4, 3, 4, 2, 3, 3, 3, 5, 2, 4, 3, 3, 5, 1, 5, 3,~
## $ alliance_DV5_6      <dbl> 1, 5, 2, 5, 4, 3, 3, 4, 5, 2, 3, 5, 3, 5, 3, 5, 4,~
## $ alliance_DV2        <dbl> 1, 2, 2, 1, 2, 2, 2, 2, 1, 3, 2, 1, 3, 1, 1, 1, 1,~
## $ alliance_DV3        <dbl> 1, 1, 2, 1, 2, 1, 2, 2, 1, 2, 1, 1, 2, 1, 1, 1, 1,~
## $ alliance_DV1_post   <dbl> 2, NA, 4, 1, 1, 2, 4, 2, 1, 4, 2, 1, NA, 1, 1, 1, ~
## $ sex_post            <dbl> 2, NA, 2, 1, 1, 2, 2, 2, 1, 1, 2, 2, NA, 1, 1, 2, ~
## $ birthyr_post        <dbl> 1952, NA, 1963, 1954, 1983, 1998, 1996, 1961, 1945~
## $ alliance_DV4_1_post <dbl> 2, NA, 5, 4, 3, 3, 4, 3, 3, 4, 5, 1, NA, 2, 1, 3, ~
## $ alliance_DV4_2_post <dbl> 5, NA, 5, 2, 2, 3, 3, 3, 1, 3, 5, 1, NA, 2, 1, 5, ~
## $ alliance_DV4_3_post <dbl> 5, NA, 4, 1, 1, 1, 4, 3, 2, 3, 5, 2, NA, 1, 1, 5, ~
## $ alliance_DV4_4_post <dbl> 5, NA, 3, 1, 1, 1, 3, 3, 2, 3, 5, 1, NA, 1, 1, 4, ~
```

```
## $ alliance_DV5_1_post <dbl> 4, NA, 2, 1, 1, 5, 4, 2, 1, 3, 3, 3, NA, 5, 3, 4, ~
## $ alliance_DV5_2_post <dbl> 5, NA, 3, 1, 1, 3, 4, 2, 1, 3, 3, 3, NA, 4, 3, 4, ~
## $ alliance_DV5_3_post <dbl> 3, NA, 3, 2, 5, 4, 3, 5, 2, 3, 3, 3, NA, 5, 2, 4, ~
## $ alliance_DV5_4_post <dbl> 5, NA, 3, 2, 4, 3, 3, 3, 3, 3, 3, 4, NA, 5, 2, 5, ~
## $ alliance_DV5_5_post <dbl> 3, NA, 2, 2, 4, 3, 3, 3, 5, 3, 3, 5, NA, 5, 4, 4, ~
## $ alliance_DV5_6_post <dbl> 3, NA, 2, 2, 5, 4, 3, 4, 4, 3, 3, 5, NA, 5, 3, 4, ~
## $ alliance_DV2_post   <dbl> 1, NA, 3, 1, 2, 1, 3, 1, 1, 3, 2, 1, NA, 1, 1, 2, ~
## $ alliance_DV3_post   <dbl> 1, NA, 2, 1, 1, 1, 2, 1, 1, 1, 2, 1, NA, 1, 1, 2, ~
## $ attack              <dbl> 100, 100, 100, 0, 0, 0, 0, 0, 0, 0, 0, 0, 100, 0, ~
## $ attack_cont         <dbl> 75, 75, 75, 0, 50, 25, 50, 25, 25, 50, 0, 25, 100,~
## $ alliance            <dbl> 1, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1,~
## $ hmrts               <dbl> 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1,~
## $ male                <dbl> 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1,~
## $ edu4                <dbl> 0.3333333, 0.0000000, 0.0000000, 0.6666667, 1.0000~
## $ white               <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1,~
## $ age                 <dbl> 70, 23, 59, 69, 39, 23, 25, 61, 77, 73, 61, 70, 32~
## $ age_cat             <dbl> 70.0, 23.5, 54.5, 64.5, 34.5, 23.5, 23.5, 64.5, 70~
## $ income_cat          <chr> "2", "2", "2", "2", "4", "1", "3", "3", "5", "2", ~
## $ inc                 <dbl> 50000, 50000, 50000, 50000, 150000, 30000, 85000, ~
## $ inc_10k             <dbl> 5.0, 5.0, 5.0, 5.0, 15.0, 3.0, 8.5, 8.5, 20.0, 5.0~
## $ threat              <dbl> 75.00, 43.75, 75.00, 18.75, 18.75, 12.50, 50.00, 3~
## $ success             <dbl> 62.5, 50.0, 75.0, 87.5, 37.5, 50.0, 50.0, 25.0, 0.~
## $ cost                <dbl> 25.00, 62.50, 31.25, 75.00, 50.00, 43.75, 50.00, 5~
## $ oblig               <dbl> 0, 50, 50, 0, 50, 50, 50, 50, 0, 100, 50, 0, 100, ~
## $ immoral             <dbl> 100, 100, 0, 100, 0, 100, 0, 0, 100, 0, 100, 100, ~
## $ moral               <dbl> 0, 25, 75, 0, 75, 25, 75, 75, 0, 100, 25, 0, 100, ~
```

```
# save df_mediate dataset
# haven::write_dta(df_ces, "YouGov/YouGov_clean.dta")

### construct dataset for the post-election wave survey data
df_ces_post <- df_ces %>% mutate(attack_post = ifelse(alliance_DV1_post > 3, 100, 0), # 1 = fa
                                 attack_cont_post = (alliance_DV1_post * 25)-25) %>% #rescale
  mutate(male = ifelse(sex_post == 1, 1, 0), # 1 = man
         edu4 = as.numeric(case_when(educ == 1  ~ '1',
                                     educ == 2  ~ '1',
                                     educ == 3 ~ '2',
                                     educ == 4 ~ '2',
                                     educ == 5 ~ '3',
                                     educ == 6 ~ '4')), # 1=HS and under, 2=some college, 3=4-
         edu4 = (edu4 - 1)/3 , #rescale education
         white = ifelse(race == 1, 1, 0),
         age_post = (2022-birthyr_post),
         age_cat = case_when(age <= 29 ~ '1',
                             age >= 30 & age <= 39 ~ '2',
                             age >= 40 & age <= 49 ~ '3',
                             age >= 50 & age <= 59 ~ '4',
                             age >= 60 & age <= 69 ~ '5',
```

```
                              age >= 70 ~ '6'),
         age_cat = ifelse(age_cat == 1, 23.5, #average of age categories, or min
                   ifelse(age_cat == 2, 34.5,
                   ifelse(age_cat == 3, 44.5,
                   ifelse(age_cat == 4, 54.5,
                   ifelse(age_cat == 5, 64.5,
                   ifelse(age_cat == 6, 70, NA)))))))

#Create Post-Election Survey Data

df_ces_post <- df_ces_post %>% mutate(threat = (alliance_DV4_1_post + alliance_DV4_2_post + al
                                      threat = (threat * 25)-25, #rescale from 0-100
                                      success = (alliance_DV5_1_post + alliance_DV5_2_post)/2,
                                      success = (success * 25)-25, #rescale from 0-100
                                      cost = (alliance_DV5_3_post + alliance_DV5_4_post + allia
                                      cost = (cost * 25)-25, #rescale from 0-100
                                      oblig = case_when(alliance_DV2_post == 1 ~ 0,
                                                        alliance_DV2_post == 2 ~ 50,
                                                        alliance_DV2_post == 3 ~ 100), # 100 =
                                      immoral = case_when(alliance_DV3_post == 1 ~ 100,
                                                          alliance_DV3_post == 2 ~ 0), # 100 =
                                      moral = (oblig - immoral + 100)/2) # 100 = moral to atta

glimpse(df_ces_post)
```

```
## Rows: 1,000
## Columns: 55
## $ alliance_DV1        <dbl> 4, 4, 4, 1, 3, 2, 3, 2, 2, 3, 1, 2, 5, 1, 1, 3, 1,~
## $ exp_4               <dbl> 4, 3, 1, 4, 3, 1, 1, 2, 4, 3, 2, 2, 2, 3, 2, 3, 3,~
## $ sex                 <dbl> 2, 2, 2, 1, 1, 2, 2, 2, 1, 1, 2, 2, 1, 1, 1, 2, 1,~
## $ educ                <dbl> 3, 2, 2, 5, 6, 5, 4, 6, 6, 2, 2, 4, 6, 5, 3, 2, 5,~
## $ race                <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 3, 1,~
## $ birthyr             <dbl> 1952, 1999, 1963, 1953, 1983, 1999, 1997, 1961, 19~
## $ income              <dbl> 6, 6, 6, 6, 11, 3, 9, 9, 13, 5, 5, 97, 10, 1, 12, ~
## $ alliance_DV4_1      <dbl> 4, 1, 4, 3, 3, 2, 3, 3, 2, 3, 4, 1, 2, 2, 2, 4, 1,~
## $ alliance_DV4_2      <dbl> 4, 3, 4, 1, 2, 2, 3, 3, 1, 3, 3, 1, 2, 1, 1, 5, 1,~
## $ alliance_DV4_3      <dbl> 4, 2, 4, 2, 1, 1, 3, 2, 2, 4, 2, 1, 3, 1, 2, 5, 1,~
## $ alliance_DV4_4      <dbl> 4, 5, 4, 1, 1, 1, 3, 2, 1, 3, 2, 1, 2, 1, 2, 5, 1,~
## $ alliance_DV5_1      <dbl> 3, 1, 4, 5, 4, 5, 3, 2, 1, 3, 3, 3, 2, 4, 3, 3, 5,~
## $ alliance_DV5_2      <dbl> 4, 5, 4, 4, 1, 1, 3, 2, 1, 2, 3, 1, 4, 4, 1, 4, 5,~
## $ alliance_DV5_3      <dbl> 2, 1, 2, 4, 4, 2, 3, 3, 2, 2, 3, 3, 4, 4, 1, 5, 2,~
## $ alliance_DV5_4      <dbl> 3, 4, 2, 3, 2, 3, 3, 2, 3, 2, 2, 5, 3, 3, 1, 5, 3,~
## $ alliance_DV5_5      <dbl> 2, 4, 3, 4, 2, 3, 3, 3, 5, 2, 4, 3, 3, 5, 1, 5, 3,~
## $ alliance_DV5_6      <dbl> 1, 5, 2, 5, 4, 3, 3, 4, 5, 2, 3, 5, 3, 5, 3, 5, 4,~
## $ alliance_DV2        <dbl> 1, 2, 2, 1, 2, 2, 2, 2, 1, 3, 2, 1, 3, 1, 1, 1, 1,~
## $ alliance_DV3        <dbl> 1, 1, 2, 1, 2, 1, 2, 2, 1, 2, 1, 1, 2, 1, 1, 1, 1,~
## $ alliance_DV1_post   <dbl> 2, NA, 4, 1, 1, 2, 4, 2, 1, 4, 2, 1, NA, 1, 1, 1, ~
```

```
## $ sex_post            <dbl> 2, NA, 2, 1, 1, 2, 2, 2, 1, 1, 2, 2, NA, 1, 1, 2, ~
## $ birthyr_post        <dbl> 1952, NA, 1963, 1954, 1983, 1998, 1996, 1961, 1945~
## $ alliance_DV4_1_post <dbl> 2, NA, 5, 4, 3, 3, 4, 3, 3, 4, 5, 1, NA, 2, 1, 3, ~
## $ alliance_DV4_2_post <dbl> 5, NA, 5, 2, 2, 3, 3, 3, 1, 3, 5, 1, NA, 2, 1, 5, ~
## $ alliance_DV4_3_post <dbl> 5, NA, 4, 1, 1, 1, 4, 3, 2, 3, 5, 2, NA, 1, 1, 5, ~
## $ alliance_DV4_4_post <dbl> 5, NA, 3, 1, 1, 1, 3, 3, 2, 3, 5, 1, NA, 1, 1, 4, ~
## $ alliance_DV5_1_post <dbl> 4, NA, 2, 1, 1, 5, 4, 2, 1, 3, 3, 3, NA, 5, 3, 4, ~
## $ alliance_DV5_2_post <dbl> 5, NA, 3, 1, 1, 3, 4, 2, 1, 3, 3, 3, NA, 4, 3, 4, ~
## $ alliance_DV5_3_post <dbl> 3, NA, 3, 2, 5, 4, 3, 5, 2, 3, 3, 3, NA, 5, 2, 4, ~
## $ alliance_DV5_4_post <dbl> 5, NA, 3, 2, 4, 3, 3, 3, 3, 3, 3, 4, NA, 5, 2, 5, ~
## $ alliance_DV5_5_post <dbl> 3, NA, 2, 2, 4, 3, 3, 3, 5, 3, 3, 5, NA, 5, 4, 4, ~
## $ alliance_DV5_6_post <dbl> 3, NA, 2, 2, 5, 4, 3, 4, 4, 3, 3, 5, NA, 5, 3, 4, ~
## $ alliance_DV2_post   <dbl> 1, NA, 3, 1, 2, 1, 3, 1, 1, 3, 2, 1, NA, 1, 1, 2, ~
## $ alliance_DV3_post   <dbl> 1, NA, 2, 1, 1, 1, 2, 1, 1, 1, 2, 1, NA, 1, 1, 2, ~
## $ attack              <dbl> 100, 100, 100, 0, 0, 0, 0, 0, 0, 0, 0, 0, 100, 0, ~
## $ attack_cont         <dbl> 75, 75, 75, 0, 50, 25, 50, 25, 25, 50, 0, 25, 100,~
## $ alliance            <dbl> 1, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1,~
## $ hmrts               <dbl> 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 1,~
## $ male                <dbl> 0, NA, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, NA, 1, 1, 0, ~
## $ edu4                <dbl> 0.3333333, 0.0000000, 0.0000000, 0.6666667, 1.0000~
## $ white               <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1,~
## $ age                 <dbl> 70, 23, 59, 69, 39, 23, 25, 61, 77, 73, 61, 70, 32~
## $ age_cat             <dbl> 70.0, 23.5, 54.5, 64.5, 34.5, 23.5, 23.5, 64.5, 70~
## $ income_cat          <chr> "2", "2", "2", "2", "4", "1", "3", "3", "5", "2", ~
## $ inc                 <dbl> 50000, 50000, 50000, 50000, 150000, 30000, 85000, ~
## $ inc_10k             <dbl> 5.0, 5.0, 5.0, 5.0, 15.0, 3.0, 8.5, 8.5, 20.0, 5.0~
## $ threat              <dbl> 81.25, NA, 81.25, 25.00, 18.75, 25.00, 62.50, 50.0~
## $ success             <dbl> 87.5, NA, 37.5, 0.0, 0.0, 75.0, 75.0, 25.0, 0.0, 5~
## $ cost                <dbl> 62.50, NA, 37.50, 25.00, 87.50, 62.50, 50.00, 68.7~
## $ oblig               <dbl> 0, NA, 100, 0, 50, 0, 100, 0, 0, 100, 50, 0, NA, 0~
## $ immoral             <dbl> 100, NA, 0, 100, 100, 100, 0, 100, 100, 100, 0, 10~
## $ moral               <dbl> 0, NA, 100, 0, 25, 0, 100, 0, 0, 50, 75, 0, NA, 0,~
## $ attack_post         <dbl> 0, NA, 100, 0, 0, 0, 100, 0, 0, 100, 0, 0, NA, 0, ~
## $ attack_cont_post    <dbl> 25, NA, 75, 0, 0, 25, 75, 25, 0, 75, 25, 0, NA, 0,~
## $ age_post            <dbl> 70, NA, 59, 68, 39, 24, 26, 61, 77, 73, 61, 70, NA~
```

```
# Save data
# haven::write_dta(df_ces_post, "YouGov/YouGov_post_clean.dta")
```