Clean Data - PureSpectrum 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)
library(dplyr)

# import the dataset
resurvey <- read_csv("PureSpectrum/data_PureSpectrum.csv")%>%
  filter(Progress == 100)%>% #remove partial responses
  glimpse()
```

```
## Rows: 4875 Columns: 34
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (34): Progress, age, sex, race, educ, US_dem_eval, pid_1, pid_2r, pid_2d...
##
## 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.
```

```
## Rows: 4,006
```

```
## Columns: 34
## $ Progress       <dbl> 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, ~
## $ age            <dbl> 2, 5, 5, 2, 5, 4, 6, 4, 5, 3, 5, 1, 4, 3, 5, 4, 4, 4, 5~
## $ sex            <dbl> 1, 2, 2, 2, 1, 2, 2, 2, 1, 1, 1, 2, 1, 2, 2, 1, 2, 1, 1~
## $ race           <dbl> 3, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 2, 1, 2, 1~
## $ educ           <dbl> 3, 3, 2, 2, 3, 3, 2, 2, 3, 3, 2, 2, 3, 3, 3, 3, 4, 3, 2~
## $ US_dem_eval    <dbl> 5, 8, 1, 10, 6, 1, 7, 1, 5, 6, 1, 10, 3, 6, 2, 5, 9, 8,~
## $ pid_1          <dbl> 3, 2, 1, 2, 1, 1, 1, 1, 1, 2, 1, 1, 3, 2, 3, 2, 2, 2, 1~
## $ pid_2r         <dbl> NA, NA, 1, NA, 1, 1, 1, 2, 1, NA, 1, 1, NA, NA, NA, NA,~
## $ pid_2d         <dbl> NA, 1, NA, 1, NA, NA, NA, NA, NA, 1, NA, NA, NA, 2, NA,~
## $ pid_2i         <dbl> 4, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 2, NA, 1~
## $ MVC_1          <dbl> 2, 2, 2, 2, 5, 2, 2, 5, 2, 2, 2, 5, 2, 2, 2, 2, 2, 2, 2~
## $ MVC_2          <dbl> 5, 1, 1, 3, 1, 1, 1, 6, 1, 1, 5, 1, 1, 1, 1, 1, 1, 1, 1~
## $ MVC_3          <dbl> 6, 3, 3, 3, 3, 3, 3, 6, 3, 3, 3, 2, 3, 3, 3, 3, 3, 3, 3~
## $ nationalism    <dbl> 3, 3, 3, 2, 2, 1, 4, 3, 4, 3, 4, 2, 3, 5, 1, 2, 2, 4, 5~
## $ patriotism     <dbl> 1, 2, 1, 1, 1, 1, 1, 3, 1, 2, 1, 3, 2, 2, 1, 1, 1, 2, 1~
## $ coop_int_1     <dbl> 5, 4, 1, 3, 2, 2, 1, 3, 2, 5, 2, 5, 5, 4, 3, 5, 4, 4, 1~
## $ coop_int_2     <dbl> 5, 5, 1, 3, 1, 2, 1, 3, 4, 3, 4, 4, 5, 5, 4, 5, 4, 5, 1~
## $ coop_int_3     <dbl> 5, 5, 1, 3, 2, 2, 2, 3, 4, 3, 5, 4, 5, 4, 4, 5, 4, 5, 4~
## $ coop_int_4     <dbl> 3, 5, 4, 3, 5, 4, 2, 3, 4, 3, 5, 5, 5, 5, 5, 5, 4, 5, 4~
## $ income         <dbl> 1, 1, 2, 2, 2, 2, 2, 1, 2, 3, 2, 2, 2, 3, 2, 1, 1, 1, 1~
## $ alliance_DV1   <dbl> 3, 3, 3, 3, 4, 5, 3, 3, 4, 4, 1, 1, 4, 2, 4, 2, 2, 1, 4~
## $ alliance_DV2   <dbl> 2, 3, 1, 1, 3, 3, 2, 2, 3, 3, 1, 2, 3, 2, 3, 1, 1, 1, 3~
## $ alliance_DV3   <dbl> 2, 1, 2, 2, 2, 1, 1, 1, 1, 2, 2, 2, 1, 2, 1, 2, 2, 2, 1~
## $ alliance_DV4_1 <dbl> 3, 1, 5, 3, 2, 1, 3, 1, 2, 3, 4, 3, 1, 4, 2, 5, 4, 4, 1~
## $ alliance_DV4_2 <dbl> 3, 1, 5, 3, 1, 1, 3, 1, 1, 3, 4, 3, 1, 4, 2, 4, 4, 3, 1~
## $ alliance_DV4_3 <dbl> 3, 1, 5, 3, 2, 1, 3, 1, 1, 3, 4, 4, 2, 3, 3, 5, 5, 3, 1~
## $ alliance_DV4_4 <dbl> 3, 1, 5, 3, 1, 1, 3, 1, 1, 3, 4, 3, 2, 4, 3, 3, 5, 3, 1~
## $ alliance_DV5_1 <dbl> 3, 2, 5, 3, 4, 2, 2, 1, 3, 2, 5, 4, 4, 4, 3, 4, 4, 4, 3~
## $ alliance_DV5_2 <dbl> 3, 2, 5, 3, 4, 2, 2, 1, 3, 3, 5, 4, 3, 2, 3, 4, 4, 4, 3~
## $ alliance_DV5_3 <dbl> 3, 5, 5, 3, 2, 2, 4, 1, 5, 3, 5, 4, 4, 3, 3, 3, 4, 3, 5~
## $ alliance_DV5_4 <dbl> 3, 3, 5, 3, 2, 2, 4, 4, 4, 3, 2, 2, 4, 2, 3, 3, 3, 3, 5~
## $ alliance_DV5_5 <dbl> 3, 5, 5, 3, 3, 2, 3, 2, 4, 3, 2, 4, 4, 2, 3, 3, 3, 2, 5~
## $ alliance_DV5_6 <dbl> 3, 5, 5, 3, 4, 2, 4, 1, 4, 3, 2, 2, 5, 2, 3, 3, 3, 3, 5~
## $ exp_4          <dbl> 2, 4, 1, 1, 4, 2, 3, 2, 4, 3, 1, 2, 4, 1, 2, 3, 3, 3, 2~
```

```
df <- resurvey %>%
  mutate(attack = ifelse(alliance_DV1 < 3, 100, 0), # 1 = favor
         attack_cont = (alliance_DV1 * -1) + 6, #reverse code; higher values = more support fo
         attack_cont = (attack_cont * 25)-25, #rescale from 0-100
         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),
         edu4 = as.numeric(case_when(educ == 1  ~ '1',
                                     educ == 2  ~ '1',
                                     educ == 3 ~ '2',
                                     educ == 4 ~ '3',
                                     educ == 5 ~ '4')),
```

```
         edu4 = (edu4 - 1)/3 , #rescale education
         white = ifelse(race == 1, 1, 0),
# Convert to continuous
         age_cat = ifelse(age == 1, 23.5, #average of age categories, or min
               ifelse(age == 2, 34.5,
               ifelse(age == 3, 44.5,
               ifelse(age == 4, 54.5,
               ifelse(age == 5, 64.5,
               ifelse(age == 6, 70, NA))))),
         inc = ifelse(income == 1, 30000, #average of income categories, min and max
               ifelse(income == 2, 50000,
               ifelse(income == 3, 85000,
               ifelse(income == 4, 150000,
               ifelse(income == 5, 200000, NA))))))%>%
  mutate(inc_10k = inc/10000)%>%   # convert the inc variable unit ($ to 10k$)
#create 3-cat party
  mutate(party = ifelse(pid_1 == "4", "3", pid_1),
         party = ifelse(party == 1, "Republican",
                 ifelse(party == 2, "Democrat",
                 ifelse(party == 3, "Independent", NA))))%>%
#Create 7-category party
  mutate(pid7_dem = ifelse(pid_2d == 1, "Strong Democrat",
                    ifelse(pid_2d == 2, "Not very strong Democrat", NA)),
         pid7_rep = ifelse(pid_2r == 1, "Strong Republican",
                    ifelse(pid_2r == 2, "Not very strong Republican", NA)),
         pid7_ind = ifelse(pid_2i == 1, "The Republican Party",
                    ifelse(pid_2i == 2, "The Democratic Party",
                    ifelse(pid_2i == 4, "Neither", NA))))%>%
  mutate(pid7_temp = ifelse(!is.na(pid7_dem), pid7_dem,
                ifelse(!is.na(pid7_ind), pid7_ind,
                ifelse(!is.na(pid7_rep), pid7_rep, NA))))%>%
  mutate(pid7 = ifelse(pid7_temp == "Strong Democrat", "1",
                ifelse(pid7_temp == "Not very strong Democrat", "2",
                ifelse(pid7_temp == "The Democratic Party", "3",
                ifelse(pid7_temp == "Neither" , "4",
                ifelse(pid7_temp == "The Republican Party", "5",
                ifelse(pid7_temp == "Not very strong Republican", "6",
                ifelse(pid7_temp == "Strong Republican", "7", pid7_temp)))))))%>%
#rescale dispositional variables: higher values = higher nationalism/patriotism
  mutate(nationalism_rc = (nationalism *-1) + 6,
         patriotism_rc = (patriotism *-1) + 5,
#take average of 4 cooperative internationalism measures
         coop_int = (coop_int_1 + coop_int_2 + coop_int_3 + coop_int_4)/4)

######### generate mediator vars, rescale from 0-100
df_mediate <- df %>%
  mutate(threat = (alliance_DV4_1+ alliance_DV4_2 + alliance_DV4_3 + alliance_DV4_4)/4,
```

```
        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 + alliance_DV5_6)/4,
        cost = (cost * 25)-25, #rescale from 0-100
        oblig = case_when(alliance_DV2  == 3 ~ 0,
                          alliance_DV2  == 2 ~ 50,
                          alliance_DV2  == 1 ~ 100), # 100 = US has moral obligation
      immoral = case_when(alliance_DV3  == 1 ~ 100,
                          alliance_DV3  == 2 ~ 0), # 100 = morally wrong for US to attack
      moral = (oblig - immoral + 100)/2) # 100 = moral to attack, 0 = immoral to attack

#Save data
# %>% write_rds("PureSpectrum/clean_data_PureSpectrum.rds")

glimpse(df_mediate)
```

```
## Rows: 4,006
## Columns: 59
## $ Progress       <dbl> 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, ~
## $ age            <dbl> 2, 5, 5, 2, 5, 4, 6, 4, 5, 3, 5, 1, 4, 3, 5, 4, 4, 4, 5~
## $ sex            <dbl> 1, 2, 2, 2, 1, 2, 2, 2, 1, 1, 1, 2, 1, 2, 2, 1, 2, 1, 1~
## $ race           <dbl> 3, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 2, 1, 2, 1~
## $ educ           <dbl> 3, 3, 2, 2, 3, 3, 2, 2, 3, 3, 2, 2, 3, 3, 3, 3, 4, 3, 2~
## $ US_dem_eval    <dbl> 5, 8, 1, 10, 6, 1, 7, 1, 5, 6, 1, 10, 3, 6, 2, 5, 9, 8,~
## $ pid_1          <dbl> 3, 2, 1, 2, 1, 1, 1, 1, 1, 2, 1, 1, 3, 2, 3, 2, 2, 2, 1~
## $ pid_2r         <dbl> NA, NA, 1, NA, 1, 1, 1, 2, 1, NA, 1, 1, NA, NA, NA, NA,~
## $ pid_2d         <dbl> NA, 1, NA, 1, NA, NA, NA, NA, NA, 1, NA, NA, NA, 2, NA,~
## $ pid_2i         <dbl> 4, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 2, NA, 1~
## $ MVC_1          <dbl> 2, 2, 2, 2, 5, 2, 2, 5, 2, 2, 2, 5, 2, 2, 2, 2, 2, 2, 2~
## $ MVC_2          <dbl> 5, 1, 1, 3, 1, 1, 1, 6, 1, 1, 5, 1, 1, 1, 1, 1, 1, 1, 1~
## $ MVC_3          <dbl> 6, 3, 3, 3, 3, 3, 3, 6, 3, 3, 3, 2, 3, 3, 3, 3, 3, 3, 3~
## $ nationalism    <dbl> 3, 3, 3, 2, 2, 1, 4, 3, 4, 3, 4, 2, 3, 5, 1, 2, 2, 4, 5~
## $ patriotism     <dbl> 1, 2, 1, 1, 1, 1, 1, 3, 1, 2, 1, 3, 2, 2, 1, 1, 1, 2, 1~
## $ coop_int_1     <dbl> 5, 4, 1, 3, 2, 2, 1, 3, 2, 5, 2, 5, 5, 4, 3, 5, 4, 4, 1~
## $ coop_int_2     <dbl> 5, 5, 1, 3, 1, 2, 1, 3, 4, 3, 4, 4, 5, 5, 4, 5, 4, 5, 1~
## $ coop_int_3     <dbl> 5, 5, 1, 3, 2, 2, 2, 3, 4, 3, 5, 4, 5, 4, 4, 5, 4, 5, 4~
## $ coop_int_4     <dbl> 3, 5, 4, 3, 5, 4, 2, 3, 4, 3, 5, 5, 5, 5, 5, 5, 4, 5, 4~
## $ income         <dbl> 1, 1, 2, 2, 2, 2, 2, 1, 2, 3, 2, 2, 2, 3, 2, 1, 1, 1, 1~
## $ alliance_DV1   <dbl> 3, 3, 3, 3, 4, 5, 3, 3, 4, 4, 1, 1, 4, 2, 4, 2, 2, 1, 4~
## $ alliance_DV2   <dbl> 2, 3, 1, 1, 3, 3, 2, 2, 3, 3, 1, 2, 3, 2, 3, 1, 1, 1, 3~
## $ alliance_DV3   <dbl> 2, 1, 2, 2, 2, 1, 1, 1, 1, 2, 2, 2, 1, 2, 1, 2, 2, 2, 1~
## $ alliance_DV4_1 <dbl> 3, 1, 5, 3, 2, 1, 3, 1, 2, 3, 4, 3, 1, 4, 2, 5, 4, 4, 1~
## $ alliance_DV4_2 <dbl> 3, 1, 5, 3, 1, 1, 3, 1, 1, 3, 4, 3, 1, 4, 2, 4, 4, 3, 1~
## $ alliance_DV4_3 <dbl> 3, 1, 5, 3, 2, 1, 3, 1, 1, 3, 4, 4, 2, 3, 3, 5, 5, 3, 1~
## $ alliance_DV4_4 <dbl> 3, 1, 5, 3, 1, 1, 3, 1, 1, 3, 4, 3, 2, 4, 3, 3, 5, 3, 1~
## $ alliance_DV5_1 <dbl> 3, 2, 5, 3, 4, 2, 2, 1, 3, 2, 5, 4, 4, 4, 3, 4, 4, 4, 3~
```

```
## $ alliance_DV5_2 <dbl> 3, 2, 5, 3, 4, 2, 2, 1, 3, 3, 5, 4, 3, 2, 3, 4, 4, 4, 3~
## $ alliance_DV5_3 <dbl> 3, 5, 5, 3, 2, 2, 4, 1, 5, 3, 5, 4, 4, 3, 3, 3, 4, 3, 5~
## $ alliance_DV5_4 <dbl> 3, 3, 5, 3, 2, 2, 4, 4, 4, 3, 2, 2, 4, 2, 3, 3, 3, 3, 5~
## $ alliance_DV5_5 <dbl> 3, 5, 5, 3, 3, 2, 3, 2, 4, 3, 2, 4, 4, 2, 3, 3, 3, 2, 5~
## $ alliance_DV5_6 <dbl> 3, 5, 5, 3, 4, 2, 4, 1, 4, 3, 2, 2, 5, 2, 3, 3, 3, 3, 5~
## $ exp_4          <dbl> 2, 4, 1, 1, 4, 2, 3, 2, 4, 3, 1, 2, 4, 1, 2, 3, 3, 3, 2~
## $ attack         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 100, 100, 0, 100, 0, 100,~
## $ attack_cont    <dbl> 50, 50, 50, 50, 25, 0, 50, 50, 25, 25, 100, 100, 25, 75~
## $ alliance       <dbl> 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 0~
## $ hmrts          <dbl> 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0~
## $ male           <dbl> 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 1~
## $ edu4           <dbl> 0.3333333, 0.3333333, 0.0000000, 0.0000000, 0.3333333, ~
## $ white          <dbl> 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1~
## $ age_cat        <dbl> 34.5, 64.5, 64.5, 34.5, 64.5, 54.5, 70.0, 54.5, 64.5, 4~
## $ inc            <dbl> 30000, 30000, 50000, 50000, 50000, 50000, 50000, 30000,~
## $ inc_10k        <dbl> 3.0, 3.0, 5.0, 5.0, 5.0, 5.0, 5.0, 3.0, 5.0, 8.5, 5.0, ~
## $ party          <chr> "Independent", "Democrat", "Republican", "Democrat", "R~
## $ pid7_dem       <chr> NA, "Strong Democrat", NA, "Strong Democrat", NA, NA, N~
## $ pid7_rep       <chr> NA, NA, "Strong Republican", NA, "Strong Republican", "~
## $ pid7_ind       <chr> "Neither", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ pid7_temp      <chr> "Neither", "Strong Democrat", "Strong Republican", "Str~
## $ pid7           <chr> "4", "1", "7", "1", "7", "7", "7", "6", "7", "1", "7", ~
## $ nationalism_rc <dbl> 3, 3, 3, 4, 4, 5, 2, 3, 2, 3, 2, 4, 3, 1, 5, 4, 4, 2, 1~
## $ patriotism_rc  <dbl> 4, 3, 4, 4, 4, 4, 4, 2, 4, 3, 4, 2, 3, 3, 4, 4, 4, 3, 4~
## $ coop_int       <dbl> 4.50, 4.75, 1.75, 3.00, 2.50, 2.50, 1.50, 3.00, 3.50, 3~
## $ threat         <dbl> 50.00, 0.00, 100.00, 50.00, 12.50, 0.00, 50.00, 0.00, 6~
## $ success        <dbl> 50.0, 25.0, 100.0, 50.0, 75.0, 25.0, 25.0, 0.0, 50.0, 3~
## $ cost           <dbl> 50.00, 87.50, 100.00, 50.00, 43.75, 25.00, 68.75, 25.00~
## $ oblig          <dbl> 50, 0, 100, 100, 0, 0, 50, 50, 0, 0, 100, 50, 0, 50, 0,~
## $ immoral        <dbl> 0, 100, 0, 0, 0, 100, 100, 100, 100, 0, 0, 0, 100, 0, 1~
## $ moral          <dbl> 75, 0, 100, 100, 50, 0, 25, 25, 0, 50, 100, 75, 0, 75, ~
```