03\_CleaningIndData

1. Set up: libraries, directory and dataset
2. Remove useless columns, rename, reorder

# Removing vars with no relevant information (columns with only NAs, same value, or web addresses)  
ind2 <- dplyr::select(ind1, 1, 5:11, 13:29, 31:35, 75:77, 79:87) # 62 obs of 42 var  
  
# Changing var names  
setnames(ind2, old = c(1:42), new = c('participant', 'surveyor', 'community', 'female', 'num\_mun', 'age', 'marital', 'kids', 'deonto\_1', 'conseq\_1', 'conseq\_2', 'deonto\_2', 'conseq\_3', 'deonto\_3', 'deonto\_4', 'conseq\_4', 'conf\_mil', 'conf\_pres', 'conf\_fedpol', 'conf\_stapol', 'conf\_munpol', 'conf\_fedcong', 'conf\_jussys', 'conf\_mungov', 'ptsd\_events', 'ptsd\_1', 'ptsd\_2', 'ptsd\_3', 'ptsd\_4', 'ptsd\_5', 'educ', 'employ\_hh', 'soc\_aid', 'fridge', 'cell', 'smart\_ph', 'wash\_m', 'computer', 'internet', 'flat\_tv', 'house', 'cars'))  
  
ind2 <- ind2[, c(1, 10, 11, 13, 16, 9, 12, 14, 15, 25:30, 17:24, 3, 5, 4, 6:8, 31:42, 2)] # reordering columns

1. Demographic variables

# Correct 'community' var  
substrRight <- function(x, n){  
 substr(x, nchar(x)-n+1, nchar(x))  
} # function to extract last letter from 'participant'  
ind2$community <- substrRight(ind2$participant, 1)  
  
# Calculate some demog variables  
ind3 <- ind2 %>% mutate(rural = ifelse(community == 'E' | community == 'M', 0, 1),  
 hs\_degree = ifelse(educ >= 4, 1, 0),  
 married = ifelse(marital == 'married', 1, 0))

1. Create indices DEONTO-MEAN: deont1 & deont2 (general) CONSEQ-MEAN: con1 & con2 (general) deont3 & deont4 cons3 & cons4 (crime or domain)

ind4 <- ind3  
  
# Replace NA code wtih NA  
ind4[ind4 == 99] <- NA   
  
# Generate indices  
ind4 <- ind4 %>% mutate(conseq\_mean = rowMeans(ind4[grep('conseq', colnames(ind4))]),  
 deonto\_mean = rowMeans(ind4[grep('deonto', colnames(ind4))]),  
 ptsd\_sum = rowSums(ind4[grep('ptsd\_', colnames(ind4))]),  
 ptsd = ifelse(ptsd\_sum >3, 1, 0),  
 conf\_mean = format(rowMeans(ind4[grep('conf\_', colnames(ind4))], na.rm = T)))

1. Write clean and complete dataset with individuals as unit of analysis

write\_xlsx(ind4, '../02\_Data/02\_Clean/individual\_clean.xlsx')