

Pilot Data Cleaner

Daniel K Baissa, Melani Cammett, and Aytug Sasmaz

December 12, 2020

Loading in data

I will start by creating a function to automatically set up the data.

All of this is a work in progress

```
# Creating a function to load datasets

data_cleaner <- function(d){

  df <- read_csv(d) %>% #reading in the data
  slice(c(-1, -2)) #qualtrics adds 2 rows of unnecessary headers. This removes them.

  # Cleaning up the conjoint data.
  # Note, this is not very efficient memory wise
  # and may need to be tuned up for larger datasets

  names <- colnames(df)

  new_names <- paste0(1:243, "_conjoint_friend1")
  new_names2 <- paste0(1:243, "_conjoint_friend2")

  df2 <- df %>%
    setnames(old = names[246:488], new = new_names) %>%
    setnames(old = names[489:731], new = new_names2)

  test2 <- df2 %>%
    pivot_longer(
      cols = ends_with("friend1"),
      names_to = "Conjoint_first_permutation",
      values_to = "Conjoint_first_permutation_answer") %>%
    filter(!is.na(Conjoint_first_permutation_answer)) %>%
    pivot_longer(
      cols = ends_with("_friend2"),
      names_to = "Conjoint_second_permutation",
      values_to = "Conjoint_second_permutation_answer") %>%
    filter(!is.na(Conjoint_second_permutation_answer)) %>%
    pivot_longer(
      cols = starts_with("ptt"),
      names_to = "Petition_Experiment_Treatment",
      values_to = "Petition_Experiment_Treatment_Answer") %>%
    filter(!is.na(Petition_Experiment_Treatment_Answer))
```

```
}
```

Now that the function is created, we can use it to make our data. There will be a warning, but it is totally ok for now.

```
csv <- c("D:/Lebanon_data/LEB Youth Civic Engagement ENG - postcut_December 11, 2020_21.29.csv",  
        "D:/Lebanon_data/LEB Youth Civic Engagement ARA - postcut_December 11, 2020_22.09.csv")
```

```
ENG <- data_cleaner(csv[1])
```

```
## Warning: Duplicated column names deduplicated: 'dem_live_1' =>  
## 'dem_live_1_1' [30], 'dem_live_2' => 'dem_live_2_1' [31]
```

```
ARA <- data_cleaner(csv[2])
```

```
## Warning: Duplicated column names deduplicated: 'dem_live_1' =>  
## 'dem_live_1_1' [31], 'dem_live_2' => 'dem_live_2_1' [32]
```

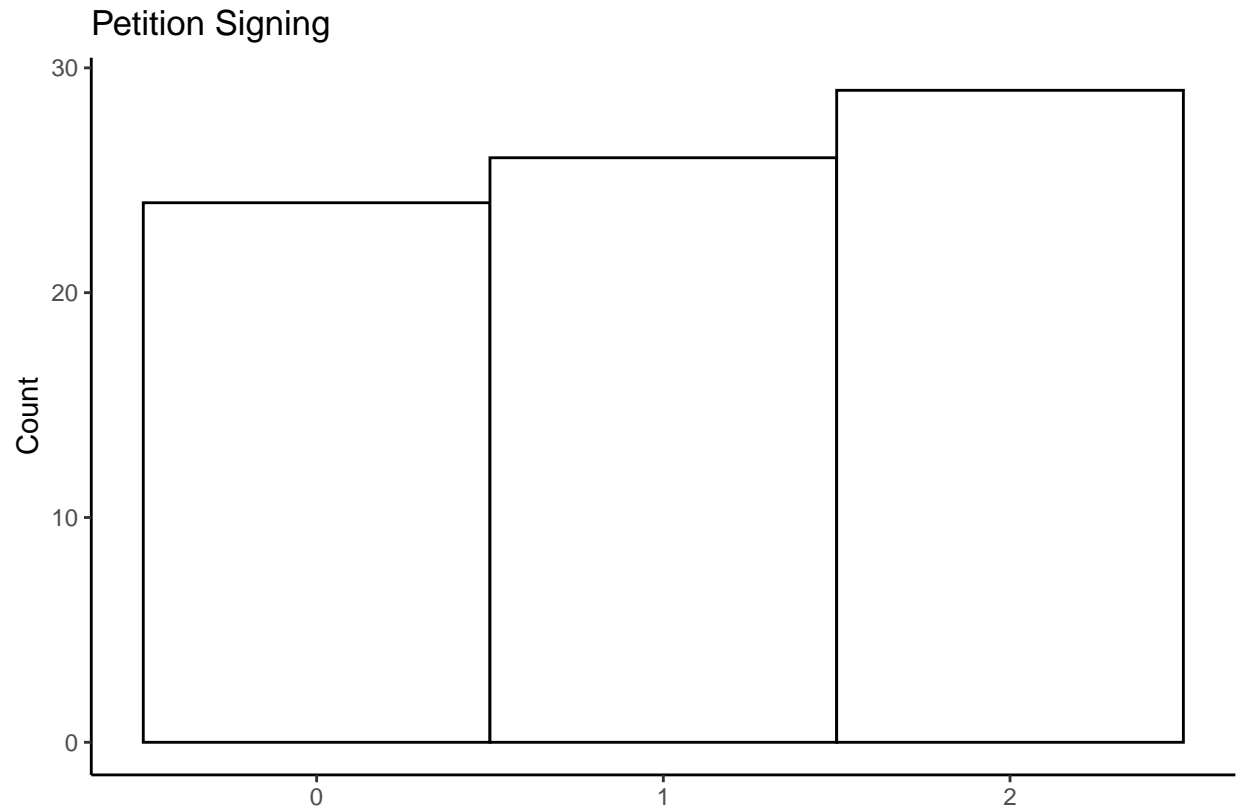
```
# For now I am just going to force the two datasets to have the same variable names  
# colnames(ENG)  
# colnames(ARA)
```

```
ARA <- ARA %>%  
  setnames(old = colnames(ARA), new = colnames(ENG))
```

```
df <- rbind(ENG, ARA)
```

Petition Signing distribution

How many people actually signed the petition? Lets find out.



```
# colnames(df)
# unique(df$Petition_Experiment_Treatment)

df$signed <- 0
df$signed[which(as.double(df$Q144) > 0)] <- 1
df$pt_control <- 1
df$pt_control[which(df$Petition_Experiment_Treatment == "ptt_cont_foll")] <- 0

test <- glm(signed ~ pt_control + UserLanguage, data = df, family = binomial())

summary(test)
```

```
##
## Call:
## glm(formula = signed ~ pt_control + UserLanguage, family = binomial(),
##      data = df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8562  -1.4006   0.6272   0.9695   0.9695
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.55962    0.62678   0.893  0.3719
## pt_control     -0.04879    0.70677  -0.069  0.9450
```

```
## UserLanguageEN 1.01523 0.59173 1.716 0.0862 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 97.020 on 78 degrees of freedom
## Residual deviance: 93.622 on 76 degrees of freedom
## AIC: 99.622
##
## Number of Fisher Scoring iterations: 4
m <- polr(as.factor(Q144) ~ pt_control + UserLanguage, data = df, Hess=TRUE)
#summary(m)
## store table
(ctable <- coef(summary(m)))

##              Value Std. Error  t value
## pt_control    -0.2261537  0.6395171 -0.353632
## UserLanguageEN  0.9689714  0.4657642  2.080391
## 0|1            -0.7196066  0.5832066 -1.233879
## 1|2             0.7196256  0.5832068  1.233912
p <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2

## combined table
(ctable <- cbind(ctable, "p value" = p))

##              Value Std. Error  t value    p value
## pt_control    -0.2261537  0.6395171 -0.353632 0.72361472
## UserLanguageEN  0.9689714  0.4657642  2.080391 0.03748971
## 0|1            -0.7196066  0.5832066 -1.233879 0.21724788
## 1|2             0.7196256  0.5832068  1.233912 0.21723587
```