

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

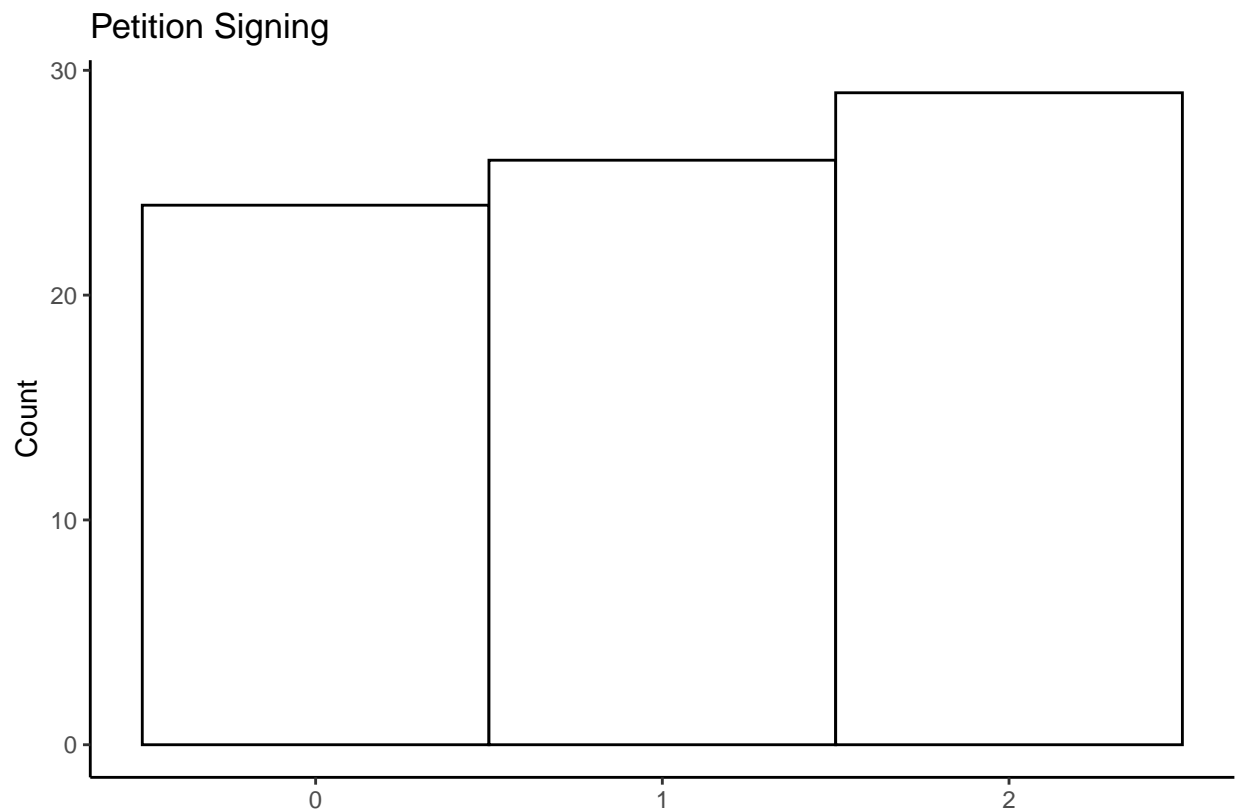
# For now I am just going to force the two datasets to have the same variable names
# colnames(ENG)
# colnames(ARA)

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.



Lets do some basic stats on the petition experiment. Here is a simple logistic regression to see if the control group is different from any of the treatments.

```

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

df2 <- df %>%
  filter(Petition_Experiment_Treatment == "ptt_treat_control_fo" | Petition_Experiment_Treatment == "ptt_treat_eco_foll")

df2$pt_econ_treatment <- 0
df2$pt_econ_treatment[which(df2$Petition_Experiment_Treatment == "ptt_treat_eco_foll")] <- 1
df2$signed <- 0
df2$signed[which(df2$Q144 > 0)] <- 1

test <- glm(signed ~ pt_econ_treatment, data = df2, family = binomial())

summary(test)

```

```

##
## Call:
## glm(formula = signed ~ pt_econ_treatment, family = binomial(),
##      data = df2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.84648   0.00008   0.00008   0.63352   0.63352
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      19.57    3584.67   0.005   0.996
## pt_econ_treatment  -18.06    3584.67  -0.005   0.996
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 13.003  on 19  degrees of freedom
## Residual deviance: 10.431  on 18  degrees of freedom
## AIC: 14.431
##
## Number of Fisher Scoring iterations: 18
## combined table
(ctable <- cbind(ctable, "p value" = p))

```

```

##
##              Value Std. Error
## Petition_Experiment_Treatmentptt_treat_bft_foll  -0.1667512  0.7582067
## Petition_Experiment_Treatmentptt_treat_control_fo  2.0974644  0.9855129
## Petition_Experiment_Treatmentptt_treat_cor_foll  -0.1171191  0.7088077
## Petition_Experiment_Treatmentptt_treat_eco_foll   0.9024628  0.8248948
## Petition_Experiment_Treatmentptt_treat_sec_foll  -0.6321240  0.7543644
## 0|1          -0.7742921  0.5908313
## 1|2           0.7744112  0.5908326
##              t value    p value
## Petition_Experiment_Treatmentptt_treat_bft_foll  -0.2199285  0.82592686
## Petition_Experiment_Treatmentptt_treat_control_fo  2.1282973  0.03331245
## Petition_Experiment_Treatmentptt_treat_cor_foll  -0.1652339  0.86875986
## Petition_Experiment_Treatmentptt_treat_eco_foll   1.0940338  0.27394015
## Petition_Experiment_Treatmentptt_treat_sec_foll  -0.8379557  0.40205559

```

```
## 0|1 -1.3105130 0.19002234
## 1|2 1.3107116 0.18995522
```

The Covid Treatment was the treatment left out for comparison by the model. Lets take a look at that.

```
## combined table
```

```
(ctable <- cbind(ctable, "p value" = p))
```

```
##                               Value Std. Error
## Petition_Experiment_Treatmentptt_treat_control_fo 2.0448068 1.0056198
## 0|1 -0.7134313 0.6389535
## 1|2 0.7136634 0.6389641
##                               t value    p value
## Petition_Experiment_Treatmentptt_treat_control_fo 2.033380 0.04201419
## 0|1 -1.116562 0.26418160
## 1|2 1.116907 0.26403415
```

Something appears to be going on here, but the N is so small that it could still be chance