Data Exploration: Gender and World View

Yanxi Fang

October 14, 2021

```
# load packages
library(tidyverse)
library(stargazer)
library(gridExtra)

# read in dataset
suffrage <- read.csv("suffrage_data.csv")
```

Question 1

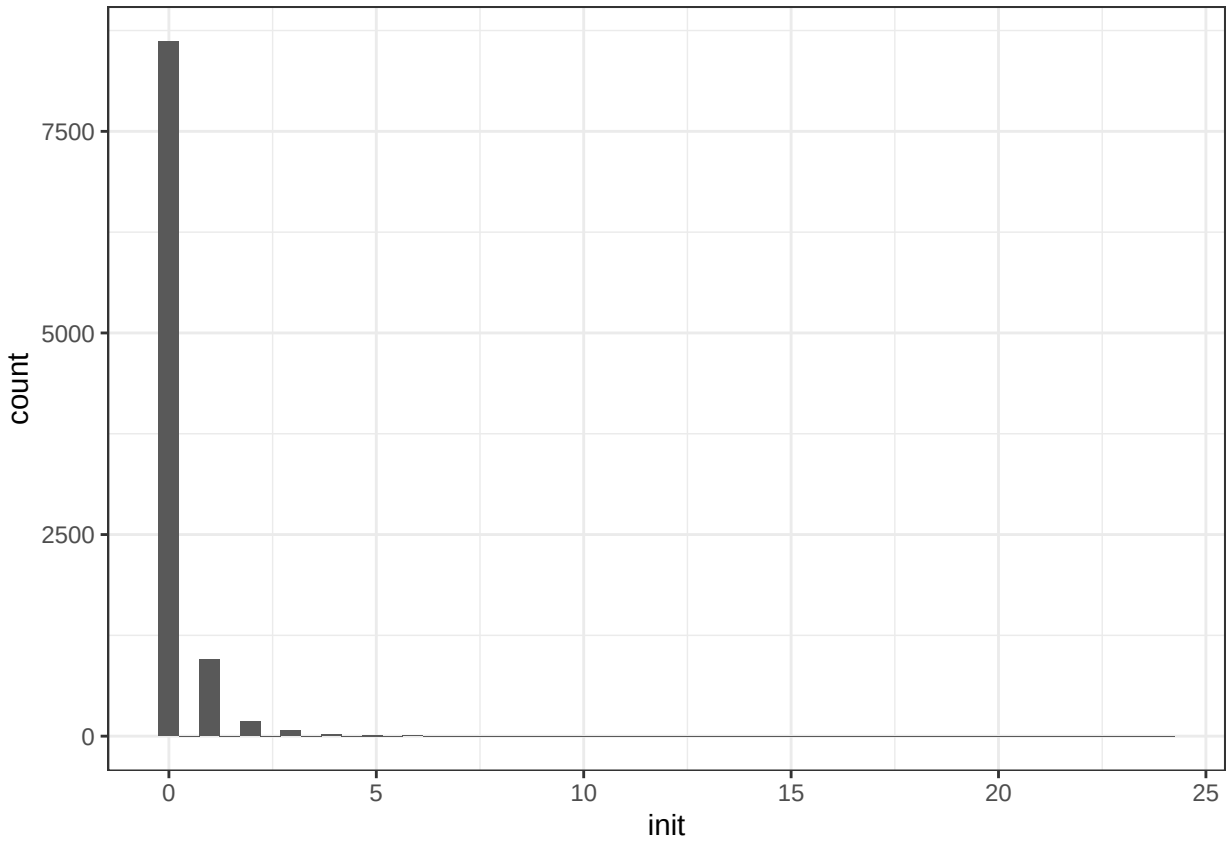
Part a

After examining the data, it appears that the unit of observation is a country i in a given year t.

Part b

As shown in the histogram and numerical summary below, conflict is somewhat uncommon (but definitely not rare) given that there are certainly many zeros, but also a somewhat large number of ones (around 1,000) across the 108 years that the dataset covers, amounting to about 10 new conflicts per year. In addition, it is important to consider that the histogram and data only reflect the number of conflicts started in each year, and does not necessarily account for the fact that previously-initiated conflicts could still be continuing in parallel.

```
# visual summary (histogram)
suffrage %>%
  ggplot(aes(x = init)) +
  geom_histogram(bins = 50) +
  theme_bw()
```




```
# numerical summary
summary(suffrage$init)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.1841  0.0000 24.0000
```

```
table(suffrage$init)
```

```
##
##    0    1    2    3    4    5    6    7    8    9   10   15   24
## 8616  951  179   67   19   15    8    2    2    2    2    1    1
```

```
summary(suffrage$year)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1900    1949    1973    1968    1991    2007
```

Question 6: Data Science Question

The regression model `init ~ suffrage + polity + polity*suffrage`, as well as the corresponding model without the interaction term, are run below.

The interaction model is interpreted as follows. For a country without female suffrage, an increase of 1 in the country's polity score is predicted, on average, to reduce the country's number of overall initiated conflicts by 0.002 (relative to the intercept of 0.193). For a country with female suffrage, an increase of 1 in the country's polity score is predicted, on average, to reduce the country's number of overall initiated conflicts by $0.002 + 0.011 = 0.013$ (relative to the intercept of $0.193 + 0.064 = 0.257$). In other words, the model predicts that when a country's polity score increases, there is a larger decrease in the number of overall initiated conflicts if the country has female suffrage versus if it does not.

Meanwhile, the regular (non-interaction) model is interpreted as: an increase of 1 in a country's polity score, when holding female suffrage constant, is predicted, on average, to reduce the country's number of overall initiated conflicts by 0.004 (relative to the intercept of 0.185, plus the coefficient on suffrage multiplied by the value of suffrage that we are holding constant). When holding the polity score constant instead, a country with female suffrage is predicted, on average, to increase the country's number of overall initiated conflicts by 0.0001 (relative to a country without female suffrage, and relative to the intercept of 0.185 plus the coefficient on polity multiplied by the value of polity that we are holding constant). In other words, the model implies that other things equal, a country with female suffrage is likely to initiate more conflicts than a country without.

```
# run models
model6a <- lm(init ~ suffrage + polity, data = suffrage)
model6b <- lm(init ~ suffrage + polity + polity*suffrage, data = suffrage)
```

```
stargazer(model6a, model6b, header = F, type = "text")
```

```
##
## ===============================================================
##                                 Dependent variable:
##                     ------------------------------------------
##                                        init
##                               (1)                   (2)
## ---------------------------------------------------------------
## suffrage                     0.0001                0.064*
##                              (0.026)               (0.038)
##
## polity                      -0.004**              -0.002
##                              (0.002)               (0.002)
##
## suffrage:polity                                   -0.011**
##                                                    (0.005)
##
## Constant                    0.185***              0.193***
##                              (0.012)               (0.013)
##
## ---------------------------------------------------------------
## Observations                  9,865                 9,865
## R2                            0.002                 0.002
## Adjusted R2                   0.002                 0.002
## Residual Std. Error    0.659 (df = 9862)     0.659 (df = 9861)
## F Statistic         9.682*** (df = 2; 9862) 8.132*** (df = 3; 9861)
```

```
## ========================================================================
## Note:                                  *p<0.1; **p<0.05; ***p<0.01
```

Question 7: Data Science Question

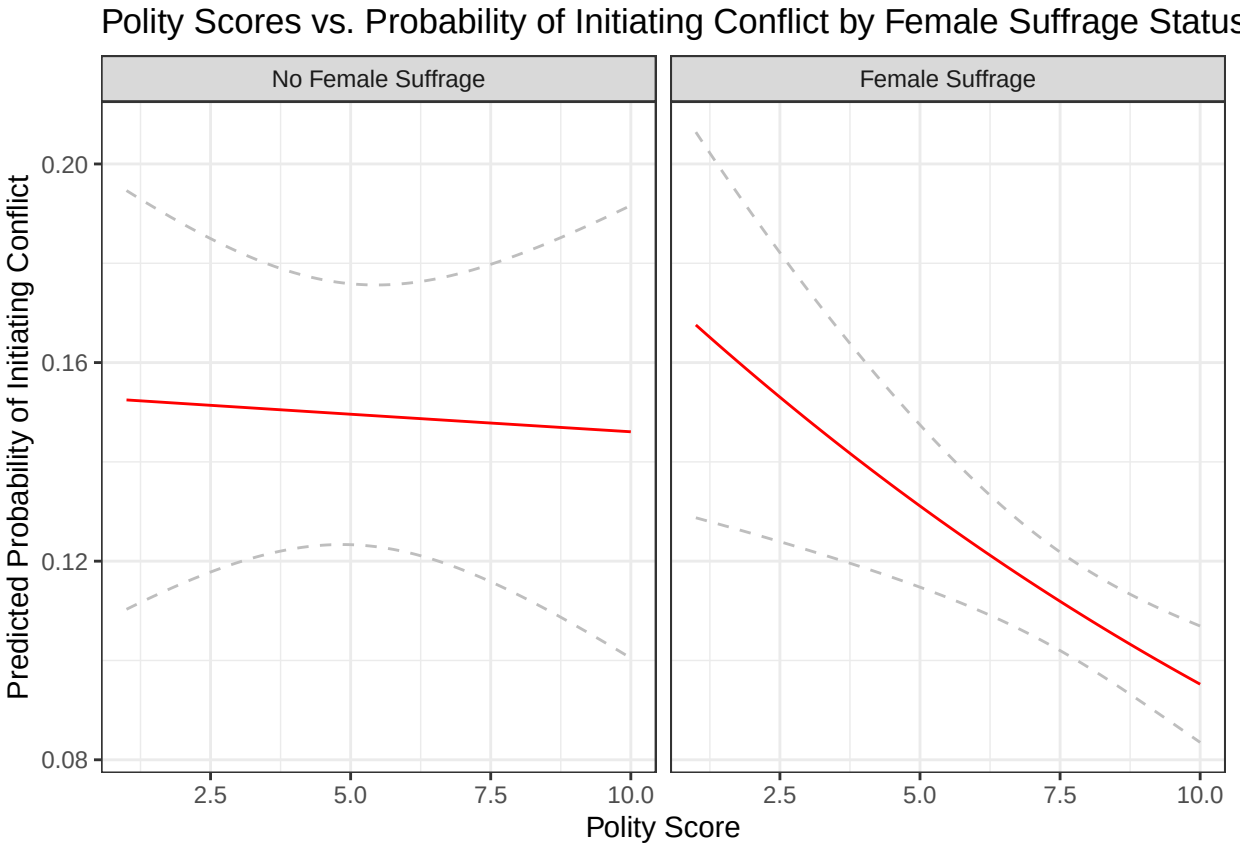
As shown on the plot on the next page, I see that the relationship between polity score and the number of initiated conflicts is negative for both countries with and without female suffrage. However, the relationship is much stronger (i.e. larger in magnitude) for countries with female suffrage than countries without female suffrage. Since polity scores are often used as indicators of how free and/or democratic a country is, one might say from the plot that more democratic countries initiate fewer conflicts, and initiate even fewer conflicts if women are allowed to vote and participate in political life. In other words, this would be consistent with the theory that women prefer peaceful policies.

```
# filter for positive polity scores only
suffrage_positivepolity <- suffrage %>%
  filter(polity >= 1)

# re-run interaction model with only positive polity scores
model7a <- lm(init ~ suffrage + polity + polity*suffrage, data = suffrage_positivepolity)

# create a new data frame for use in `predict()` function
pred_data1 <- data.frame(polity = seq(1,10,0.01), suffrage = 0)
pred_data2 <- data.frame(polity = seq(1,10,0.01), suffrage = 1)
pred_data <- bind_rows(pred_data1, pred_data2)

# run `predict()` function to generate predictions based on the model
pred_data$fit <- predict(model7a, newdata = pred_data, se.fit = T)$fit
pred_data$se <- predict(model7a, newdata = pred_data, se.fit = T)$se.fit

# new suffrage column to make graph look nicer
pred_data$suffrage_words <- ifelse(pred_data$suffrage == 0,
                                   "No Female Suffrage", "Female Suffrage")

# graph
pred_data %>%
  ggplot(aes(x = polity)) +
  geom_line(aes(y = fit + 1.96*se), col = "gray", lty = "dashed") +
  geom_line(aes(y = fit - 1.96*se), col = "gray", lty = "dashed") +
  facet_wrap(~ suffrage_words) +
  geom_line(aes(y = fit), col = "red") +
  xlab("Polity Score") + ylab("Predicted Number of Initiated Conflicts") +
  ggtitle("Polity Scores vs. Initiated Conflicts by Female Suffrage Status") +
  theme_bw()
```


```
ggsave("polity_femalesuffrage.png", height = 4, width = 8)
```

Extension: Using a Binary Dependent/Outcome Variable

```
# create a new binary variable for conflict
suffrage <- suffrage %>%
  mutate(init_binary = ifelse(init == 0, 0, 1))

# filter for positive polity scores
suffrage_positivepolity <- suffrage %>%
  filter(polity >= 1)

# run logit model (since the outcome is binary) with the same variables
model_ext <- glm(init_binary ~ suffrage + polity + polity*suffrage,
                 data = suffrage_positivepolity, family = "binomial")

# create new data frame for predictions
pred_data_e1 <- data.frame(polity = seq(1,10,0.01), suffrage = 0)
pred_data_e2 <- data.frame(polity = seq(1,10,0.01), suffrage = 1)
pred_data_e <- bind_rows(pred_data_e1, pred_data_e2)

# run `predict()` function to generate predictions based on the model
pred_data_e$fit <- predict(model_ext, newdata = pred_data_e, type = "response", se.fit = T)$fit
pred_data_e$se <- predict(model_ext, newdata = pred_data_e, type = "response", se.fit = T)$se.fit

# new suffrage column to make graph look nicer
pred_data_e$suffrage_words <- ifelse(pred_data_e$suffrage == 0,
                                     "No Female Suffrage", "Female Suffrage")

# arrange order of facet


# graph
pred_data_e %>%
  ggplot(aes(x = polity)) +
  geom_line(aes(y = fit + 1.96*se), col = "gray", lty = "dashed") +
  geom_line(aes(y = fit - 1.96*se), col = "gray", lty = "dashed") +
  facet_grid(~factor(suffrage_words, levels=c('No Female Suffrage','Female Suffrage'))) +
  geom_line(aes(y = fit), col = "red") +
  xlab("Polity Score") + ylab("Predicted Probability of Initiating Conflict") +
  ggtitle("Polity Scores vs. Probability of Initiating Conflict by Female Suffrage Status") +
  theme_bw()
```




```
ggsave("polity_femalesuffrage_binary.png", height = 4, width = 8)
```