

PS 3

2025-11-05

Part 1 Paper Analysis

1. Research Goals The research goal of this study is to make causal explanation for what causes civil wars. The directed acylic diagram in their theory is that *economicunfavorableconditions* → *weakstate* → *civilwaronsets*. The authors have stated clearly in terms of their theory to identify what is the root cause of civil wars. However, the way they did cannot satisfy a perfect causal identification build off regression. Their regression model includes a lot of structural variables
2. Estimands There are 11 hypothesis in this paper, and each one corresponds to an implicit theoretical causal estimand.

For H1: The theoretical estimand for H1 is the average causal effect of higher ethnic or religious diversity on the probability of civil war onset — that is, the difference in civil-war risk between a world where a country has high diversity and a world where the same country has low diversity, all else held constant. The empirical estimand is the change in the log-odds of civil war onset correlated with a one-unit change in ethnic or religious fractionalization, holding other regressors constant.

For H2: The theoretical estimand for H2 is the causal interaction effect of ethnic diversity and income on civil-war onset — that is, how the causal effect of ethnic diversity on civil-war risk differs between a world with higher income and a world with lower income. The empirical estimand is the conditional change in the log-odds of civil war onset associated with diversity at different levels of income, as captured by the coefficient on diversity and any implicit interaction through the model's functional form, holding other regressors constant.

For H3: The theoretical estimand for H3 is the average causal effect of having an ethnic-majority/large-minority structure on the probability of civil war onset — comparing a world where a country has a pronounced majority-minority structure versus a world where it does not, all else held constant. The empirical estimand is the change in the log-odds of civil war onset associated with the presence of a large-minority ethnic structure, holding other regressors constant.

For H4: The theoretical estimand for H4 is the average causal effect of higher political democracy and civil liberties on the probability of civil war onset — comparing a world where a country possesses strong democratic institutions and freedoms versus a world where it does not, all else held constant. The empirical estimand is the change in the log-odds of civil war onset correlated with a one-unit increase in democracy or civil-liberty measures, holding other regressors constant.

For H5: The theoretical estimand for H5 is the average causal effect of discriminatory language or religious policies on the probability of civil war onset — contrasting a world where such discriminatory policies exist with one where they do not, all else held constant. The empirical estimand is the change in the log-odds of civil war onset associated with the presence of discriminatory language or religious policies, holding other regressors constant.

For H6: The theoretical estimand for H6 is the average causal effect of greater income inequality on the probability of civil war onset — comparing a world with high inequality to a world with low inequality, all else held constant. The empirical estimand is the change in the log-odds of civil war onset correlated with a one-unit increase in income-inequality measures (e.g., Gini), holding other regressors constant.

For H7: The theoretical estimand for H7 is the average causal effect of greater ethnic diversity on the

probability of ethnic civil war onset among countries with a minority $\geq 5\%$ — comparing a world in which diversity is high versus one where it is low, within that subgroup, all else held constant. The empirical estimand is the change in the log-odds of ethnic civil war onset associated with ethnic fractionalization in the $\geq 5\%$ minority subsample, holding other regressors constant.

For H8: The theoretical estimand for H8 is the average causal effect of rough, poorly connected terrain on the probability of civil war onset — comparing a world where a country has high rough-terrain exposure versus one where it does not, all else held constant. The empirical estimand is the change in the log-odds of civil war onset correlated with a one-unit increase in the log of mountainous terrain, holding other regressors constant.

For H9: The theoretical estimand for H9 is the average causal effect of lower GDP per capita on the probability of civil war onset — that is, the difference in civil-war risk between a world where a country has low income and a world where the same country has high income, all else held constant. The empirical estimand is the change in log-odds of civil war onset correlated with a one-unit change in GDP per capita holding other regressors constant.

For H10: The theoretical estimand for H10 is the average causal effect of each insurgency-favoring political or military condition (new statehood, political instability, anocracy, large population, noncontiguous territory, foreign support, resource-finance potential, or oil dependence) on the probability of civil war onset — comparing worlds with and without each condition, all else held constant. The empirical estimand is the change in the log-odds of civil war onset associated with a one-unit change in each of the listed insurgency-relevant variables, holding other regressors constant.

For H11: The theoretical estimand for H11 is the average causal effect of democracy, civil liberties, income equality, and nondiscriminatory policies after conditioning on state strength on the probability of civil war onset — that is, the causal effect of these “grievance” variables in a world where GDP and other state-capacity indicators are fixed. The empirical estimand is the change in the log-odds of civil war onset associated with democracy, civil liberties, income inequality, and nondiscriminatory linguistic or religious policies conditional on GDP and other controls, holding all regressors constant.

3. Identification Strategy Their identification strategy consists of using lagged predictors and extensive controls in a country-year logit model, under the assumption that conditioning on these variables renders GDP and other regressors exogenous with respect to civil-war onset. This design treats the empirical logit coefficients as theoretical causal estimates, but it does not produce identified causal effects, as omitted-variable bias, reverse causality, and structural endogeneity remain. Therefore, the empirical estimand only approximates the theoretical estimand under strong, untested conditional-independence assumptions.
4. Assessment of Findings Fearon & Laitin have a powerful causal theory, but their identification strategy is observational and correlational. The regressions do not isolate exogenous variation in some independent variables such as GDP, and no instruments, no natural experiments, no fixed-effects identification, so the empirical coefficients cannot be interpreted as causal treatment effects. Their statistical model is a reasonable approximation of the DGP but omits key forms of unobserved heterogeneity and endogeneity. The data measure some structural features well (terrain, population, oil) but measure core theoretical constructs like grievance, ethnic salience, state capacity, and democracy with substantial noise. Thus the identification strategy supports the authors’ claims only in the sense of providing correlational evidence consistent with their causal narrative, not in the sense of establishing credible causal effects.
5. Broader Contribution Despite the lack of a modern identification strategy, Fearon and Laitin’s study still meaningfully informs our understanding of civil conflict. Their analysis provides several robust descriptive regularities—such as the strong cross-national association between low state capacity, rough terrain, political instability, and civil-war onset. These correlations are not credible causal estimates on their own, but they successfully falsify a dominant but incorrect narrative that ethnic diversity is the primary driver of civil war. Moreover, the study offers a coherent theoretical framework that organizes empirical patterns and reshapes how scholars conceptualize conflict risk. In this sense, even without

strong causal identification, the study advances knowledge by identifying stable empirical patterns, overturning misleading explanations, and offering a compelling theory that guides later causal inquiry.

Data Analysis

Q1

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

thermo <- read.csv("thermometers.csv")
survey_year <- 2017
thermo <- thermo %>%
  mutate(age = survey_year - thermo$birth_year)
```

Q2

```
library(ggplot2)
class(thermo$ft_asian)

## [1] "integer"

thermo %>%
  summarize(mean_overall_ft_asian = mean(thermo$ft_asian, na.rm = TRUE),
           median_overall_ft_asian = median(thermo$ft_asian, na.rm = TRUE),
           var_overall_ft_asian = var(thermo$ft_asian, na.rm = TRUE),
           sd_overall_ft_asian = sd(thermo$ft_asian, na.rm = TRUE),
           min_overall = min(thermo$ft_asian, na.rm = TRUE),
           max_overall = max(thermo$ft_asian, na.rm = TRUE))

##   mean_overall_ft_asian median_overall_ft_asian var_overall_ft_asian
## 1             74.18704                  79            464.8826
##   sd_overall_ft_asian min_overall max_overall
## 1            21.56114                  0            100

thermo %>%
  group_by(party_id) %>%
  summarize (mean_ft_asian = mean(ft_asian, na.rm = TRUE),
             median_ft_asian = median(ft_asian, na.rm = TRUE),
             var_ft_asian = var(ft_asian, na.rm = TRUE),
             sd_ft_asian = sd(ft_asian, na.rm = TRUE),
             min = min(ft_asian, na.rm = TRUE),
             max = max(ft_asian, na.rm = TRUE))

## # A tibble: 5 x 7
##   party_id    mean_ft_asian median_ft_asian var_ft_asian sd_ft_asian   min   max
##   <chr>          <dbl>        <dbl>        <dbl>        <dbl> <int> <int>
## 1 Democrat      76.8         80            439.       21.0      0     100
```

```

## 2 Independent      73.2      78      438.      20.9      0      100
## 3 Not sure        72.2    78.5      527.      23.0      7      100
## 4 Other            75.4      80      509.      22.6      0      100
## 5 Republican       72.1      77      510.      22.6      0      100

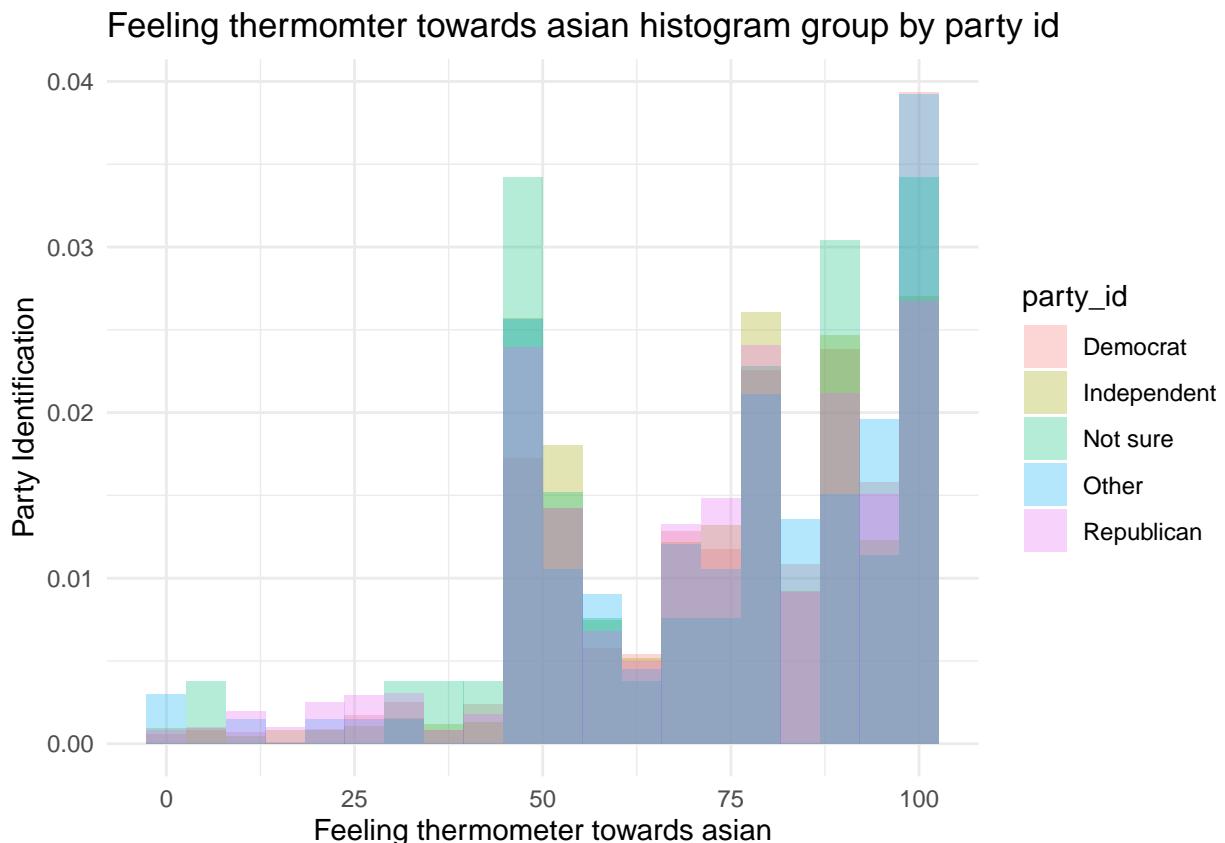
```

```

ggplot(thermo, aes(x = ft_asian, fill = party_id))+
  geom_histogram(aes(y = after_stat(density)),
                 position = "identity", alpha = 0.3, bins = 20)+
  labs(title = "Feeling thermomter towards asian histogram group by party id",
       x = "Feeling thermometer towards asian",
       y = "Party Identification") +
  theme_minimal()

```

Warning: Removed 145 rows containing non-finite outside the scale range
('stat_bin()'').



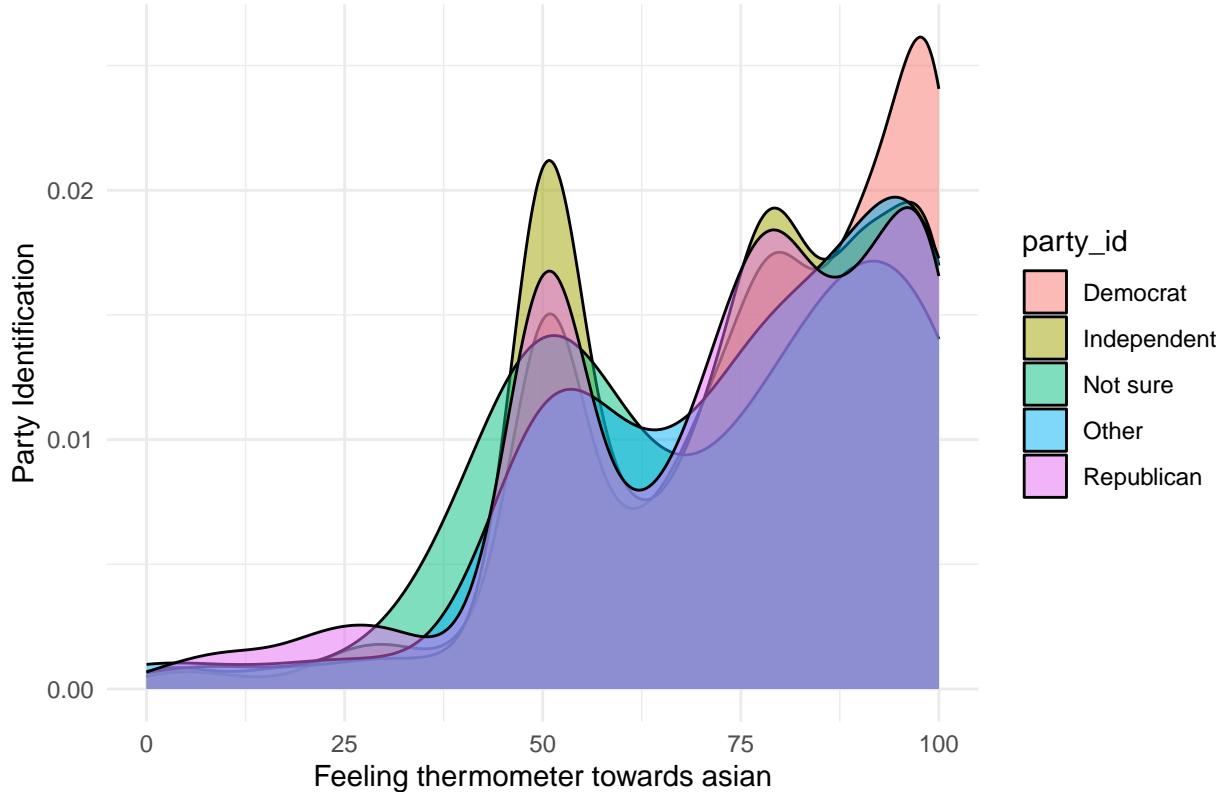
```

ggplot(thermo, aes(x = ft_asian, fill = party_id))+
  geom_density(alpha = 0.5)+
  labs(title = "Feeling thermomter towards asian density group by party id",
       x = "Feeling thermometer towards asian",
       y = "Party Identification") +
  theme_minimal()

```

Warning: Removed 145 rows containing non-finite outside the scale range
('stat_density()'').

Feeling thermometer towards asian density group by party id



Interpretation: Across all respondents, the average rating on the feeling thermometer toward Asians is 74.19, with a median of 79. This indicates that most respondents express generally warm attitudes toward Asians. The standard deviation of 21.56 and a full range from 0 to 100 show a moderate spread of opinions—while the majority of respondents rate Asians positively, a small minority still give low ratings, resulting in a fairly wide distribution.

Breaking this down by party identification, Democrats display the warmest average feeling (mean = 76.76, median = 80), followed by those identifying as “Other” (75.37) and Independents (73.24). Respondents who are “Not sure” (72.24) or Republican (72.10) report slightly cooler but still positive average feelings.

Variation within each partisan group is quite similar, with standard deviations around 21–23.

Overall, the data reveal a broadly positive orientation toward Asians, with modest partisan differences in central tendency but no meaningful differences in spread.

Q3

```
m1 <- lm(ft_asian ~ as.factor(party_id), thermo)

summary(m1)

##
## Call:
## lm(formula = ft_asian ~ as.factor(party_id), data = thermo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -76.764 -20.095   4.236  17.905  27.905 
##
## Coefficients:
```

```

##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                76.7643   0.5234 146.662 < 2e-16 ***
## as.factor(party_id)Independent -3.5209   0.7485 -4.704 2.63e-06 ***
## as.factor(party_id)Not sure    -4.5243   3.0823 -1.468  0.142
## as.factor(party_id)Other      -1.3912   1.9838 -0.701  0.483
## as.factor(party_id)Republican -4.6688   0.7810 -5.978 2.42e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.48 on 4839 degrees of freedom
##   (145 observations deleted due to missingness)
## Multiple R-squared:  0.008436, Adjusted R-squared:  0.007617
## F-statistic: 10.29 on 4 and 4839 DF,  p-value: 2.682e-08

```

Q4

```

thermo_1 <- thermo %>%
  filter(party_id %in% c("Democrat", "Republican")) %>%
  mutate(party_id_binary = ifelse(party_id == "Democrat", 1, 0))

```

Q5

```

m2 <- lm(party_id_binary ~ ft_black + ft_unions + ft_gays + ft_immig + age + educ + ft_immig * educ + f
summary(m2)

```

##

Call:

```

## lm(formula = party_id_binary ~ ft_black + ft_unions + ft_gays +
##     ft_immig + age + educ + ft_immig * educ + ft_gays * age,
##     data = thermo_1)
## 
```

##

Residuals:

```

##      Min      1Q      Median      3Q      Max
## -1.13415 -0.25735  0.01805  0.24105  1.15911
## 
```

##

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|--------------------------------------|------------|------------|---------|--------------|
| ## (Intercept) | -4.974e-02 | 9.047e-02 | -0.550 | 0.58251 |
| ## ft_black | -1.149e-03 | 3.685e-04 | -3.118 | 0.00184 ** |
| ## ft_unions | 7.390e-03 | 2.606e-04 | 28.362 | < 2e-16 *** |
| ## ft_gays | 4.782e-03 | 1.096e-03 | 4.364 | 1.32e-05 *** |
| ## ft_immig | 1.424e-03 | 7.256e-04 | 1.963 | 0.04975 * |
| ## age | -2.376e-04 | 1.252e-03 | -0.190 | 0.84947 |
| ## educ4-year | -1.348e-02 | 6.101e-02 | -0.221 | 0.82519 |
| ## educHigh school graduate | -1.374e-02 | 5.634e-02 | -0.244 | 0.80740 |
| ## educNo HS | -9.320e-02 | 1.159e-01 | -0.804 | 0.42150 |
| ## educPost-grad | -1.509e-01 | 6.953e-02 | -2.170 | 0.03011 * |
| ## educSome college | 7.422e-02 | 6.570e-02 | 1.130 | 0.25873 |
| ## ft_immig:educ4-year | 1.588e-04 | 8.747e-04 | 0.182 | 0.85598 |
| ## ft_immig:educHigh school graduate | -2.745e-04 | 8.443e-04 | -0.325 | 0.74512 |
| ## ft_immig:educNo HS | 2.121e-03 | 2.041e-03 | 1.039 | 0.29882 |
| ## ft_immig:educPost-grad | 1.825e-03 | 9.810e-04 | 1.861 | 0.06289 . |
| ## ft_immig:educSome college | -8.591e-04 | 9.543e-04 | -0.900 | 0.36806 |
| ## ft_gays:age | -2.077e-05 | 1.754e-05 | -1.184 | 0.23653 |
| ## --- | | | | |

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3753 on 2821 degrees of freedom
##   (308 observations deleted due to missingness)
## Multiple R-squared:  0.4306, Adjusted R-squared:  0.4274
## F-statistic: 133.4 on 16 and 2821 DF,  p-value: < 2.2e-16

```

Interpretation: I include four feeling thermometer variables — ft_black, ft_union, ft_gays, and ft_immig — representing race, labor, gender, and immigration respects. I also include two interaction terms to examine how demographic characteristics condition these effects. The first, ft_immig * educ, tests whether education moderates the relationship between attitudes toward immigrants and partisanship. I expect that higher levels of education may weaken the link between anti-immigration sentiment and Republican identification, as more educated individuals tend to hold more cosmopolitan or inclusive views. The second interaction, ft_gays * age, examines whether the effect of attitudes toward gay and lesbian people on partisanship varies across age groups. I expect that younger respondents will show stronger positive associations between warm feelings toward LGBTQ individuals and Democratic identification, reflecting generational differences in social liberalism.

Q6 Interpretation: In this multiple linear regression model, the dependent variable party_id_binary is coded as 1 for Democrats and 0 for Republicans. Therefore, the coefficients represent the change in the predicted probability of identifying as a Democrat associated with a one-unit increase in the independent variable, holding all other variables constant.

For example, the coefficient for ft_black (0.00115) means that for each one-point increase on the feeling thermometer toward Black people, the probability of identifying as a Democrat increases by approximately 0.00115, or 0.115 percentage points, on average. Similarly, the coefficient for ft_unions (0.00739) indicates that a one-point increase in warmth toward labor unions increases the probability of being a Democrat by about 0.74 percentage points, holding all other variables constant.

For the interaction terms (ft_immig * educ, ft_gays * age), the coefficients represent how the relationship between the feeling thermometer and party identification changes depending on the level of the interacting variable. For instance, a negative coefficient on ft_gays × age would mean that the effect of attitudes toward gay and lesbian people on Democratic identification weakens as age increases.

Q7 Use ft_gay as prediction indicator

```

library(tidyr)
prediction_data <- thermo_1 %>%
  summarise(across(c(ft_black, ft_unions, ft_immig, age), median, na.rm=TRUE)) %>%
  mutate(educ = "4-year")

## Warning: There was 1 warning in `summarise()` .
## i In argument: `across(c(ft_black, ft_unions, ft_immig, age), median, na.rm =
##   TRUE)` .
## Caused by warning:
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
## Supply arguments directly to `.fns` through an anonymous function instead.
##
##   # Previously
##   across(a:b, mean, na.rm = TRUE)
##
##   # Now
##   across(a:b, \((x) mean(x, na.rm = TRUE))`)

prediction_data <- crossing(prediction_data, ft_gays = 0:100)

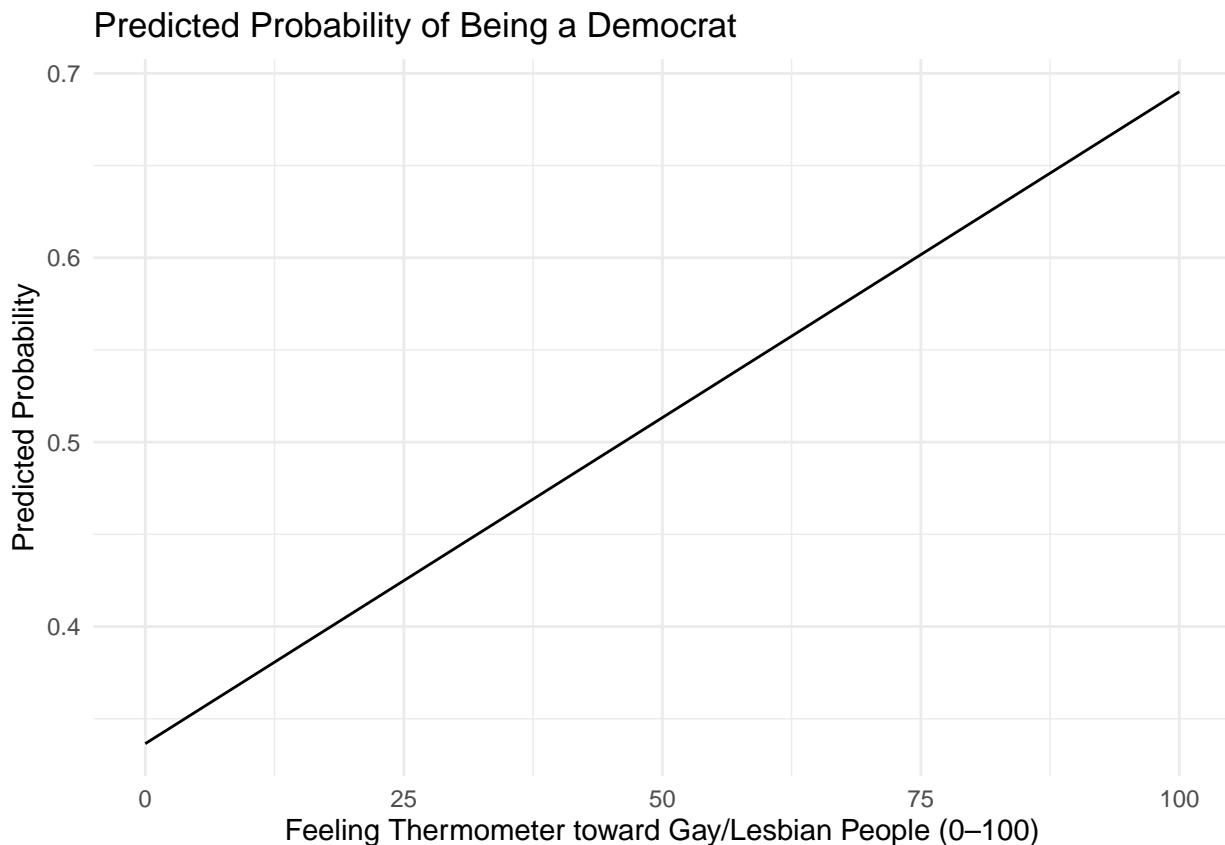
prediction_data$prediction <- predict(m2, prediction_data)

```

```

ggplot(prediction_data, aes(x = ft_gays, y = prediction))+
  geom_line()+
  labs(title = "Predicted Probability of Being a Democrat",
       x = "Feeling Thermometer toward Gay/Lesbian People (0-100)",
       y = "Predicted Probability") +
  theme_minimal()

```



Interpretation: The results indicate a strong positive association between warmth toward gay and lesbian people and the probability of identifying as a Democrat. Respondents who express more positive attitudes toward LGBTQ individuals are substantially more likely to identify with the Democratic Party.

However, because this analysis relies on observational survey data, the relationship should be interpreted as correlation, not causal. Differences in underlying ideology, religiosity, or other factors could jointly influence both LGBTQ attitudes and party identification.