Riley McNair
04-14-24

Conditioning the Relationship Between Polarization and Protests

Earlier parts of this project explored the relationship between US states polarization variance and their protest participation. It was determined in part B that there is not a statistically significant relationship between the polarization scores of states and the number of protests occurring within those states. Nevertheless, we aim to explore whether introducing a third variable might reveal a statistically significant relationship that was previously obscured.

Data Synthesis

While there are many variables that can have an effect on the relationship between political polarization of each state and protests per capita, the following paper has identified five potential additional variables that could influence this relationship from 2020 to 2024. These following variables have been pulled from multiple other data sets and sources. Measures of each are listed below:

1. **Diversity Index**: This index measures the probability that two randomly selected individuals from a state belong to different racial or ethnic groups, using data from the U.S. Census Bureau (2020).
2. **Education Index**: I constructed this index by integrating several educational metrics: high school dropout rates, the percentage of individuals with higher education diplomas, and adult literacy rates (for those aged 15 and older). The index is calculated as follows:

$$\text{Education Index} = \frac{\text{Dropout Rate} + \text{Bachelor Degree Rate} + (100 - \text{Literacy Rate})}{3}$$

   Data sources include the World Population Review for high school graduation rates and literacy rates (2024), and the Federal Reserve Bank of St. Louis (2022).
3. **Economic Index**: This index was developed to gauge the economic well-being of each state by combining unemployment rates, poverty levels, and median household income. It is defined as:

$$\text{Economic Index} = \left(\frac{(1 - \text{Normalized Unemployment Rate}) + (1 - \text{Normalized Poverty Level}) + \text{Normalized Median Income}}{3} \right) \times 100$$

   The data were sourced from the U.S. Bureau of Labor Statistics(2024), USDA Economic Research Service(2023), and the Federal Reserve Bank of St. Louis. (2022)

4. **Voter Turnout Rate:** Voter turnout data from each state was obtained from the United States Elections Project (2020).
5. **Party Voted:** This indicates whether a state voted Democratic or Republican in the 2020 presidential election, based on data from the CNN 2020 Presidential Election Results.

Note: The data supporting these analysis can be replicated using datasets available on my GitHub linked below.

state	voted_2020	protests_per_100k	polarization_var	diversity_index	education_index	economic_index	turnout_rate
<dbl>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Republican	7.284627	0.13333237	53.1	63.93333	78.68990	35.87
2	Republican	9.135645	0.02205297	62.8	70.33333	89.14633	47.92
4	Democrat	6.180520	0.19872212	61.5	65.83333	83.62785	44.67
5	Republican	6.740773	0.08151470	49.8	63.16667	76.76303	38.84

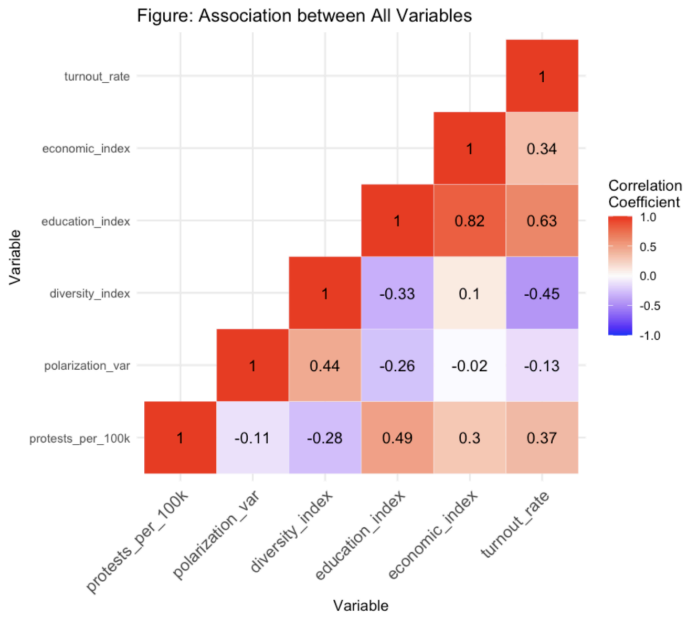
Summary Table: Numerical Variables

key	mean	sd	max	min
<fct>	<dbl>	<dbl>	<dbl>	<dbl>
protests_per_100k	11.22	5.67	32.97	4.46
polarization_var	0.15	0.12	0.53	0.01
diversity_index	49.04	14.52	76.00	18.50
education_index	68.55	3.44	74.97	60.70
economic_index	84.43	4.65	95.77	74.21
turnout_rate	44.51	7.25	59.70	31.36

Summary Table: Categorical Variable

voted_2020	n	prop	std_dev
<chr>	<int>	<dbl>	<dbl>
Democrat	25	0.5	0.07
Republican	25	0.5	0.07

Initially, these variables were pinpointed as potential influencers requiring control. The next step was developing a heatmap to explore correlations between numerical variables that might influence both polarization scores and protest scores.



The analysis revealed that there was no substantial overlap between the variables affecting these two outcomes. Consequently, the variables selected for further analysis showed the strongest correlation with either protests or polarization. These variables were then examined to assess their impact on the significance levels in our statistical model.

The strongest correlated variables—Education Index and Diversity Index—alongside the categorical variable of State Political Affiliation, are summarized within a linear regression framework to see their influence on significance levels below.

Education Summary Table: The significant effect of the education index on protest frequency, but not on the relationship between polarization and protests, indicates that while education impacts protest levels, it does not mediate the relationship between polarization and protests. This is an important distinction that suggests education independently influences protest behaviours, but not on the x and y relations.

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -44.5779    15.2204  -2.929 0.005234 **
polarization_var   0.8310     6.2188   0.134 0.894269
education_index    0.8121     0.2177   3.731 0.000514 ***
```

Diversity Summary Table: The addition of the diversity index to the model slightly alters the coefficients, but the changes are not significant enough to affect the overall relationship between polarization and protests.

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     16.61501    2.79848   5.937 3.35e-07 ***
polarization_var  0.67403    7.34043   0.092   0.9272
diversity_index  -0.11207    0.06081  -1.843   0.0716 .
```

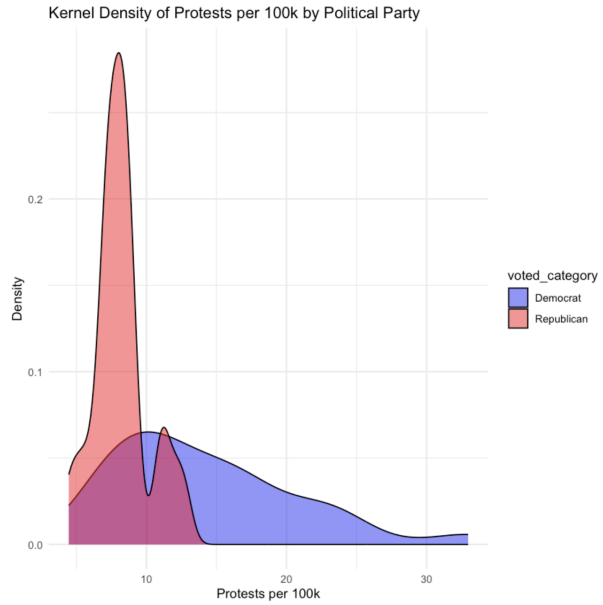
Political Affiliation Summary Table: The significant coefficient for voted_categoryRepublican suggests a strong partisan effect on protest frequency, independent of polarization. While these results are not below the significance level of .05%, it is much lower than any other previously defined variable.

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)            15.876      1.370  11.587 2.24e-15 ***
polarization_var       -9.426      5.704  -1.653    0.105
voted_categoryRepublican   -6.387      1.358  -4.702 2.29e-05 ***
```

While the coefficients of these variables do not reach statistical significance, their values suggest a trend that may suggest an emerging variable capable of mediating or moderating the relationship between political variance and protests. The direction of their estimates and the p-values close to .05%, hint at an underlying influence that warrants further exploration. Moreover, the limited sample size of our dataset could be a contributing factor to the current lack of significance. As such, expanding the dataset may enhance the robustness of these findings and potentially reveal the significant effects these variables may exert on the dynamics of political activism.

Conditioning a Correlation Variable

Focusing on the variable that most closest approaches influence the relationship, we now explore whether holding the party affiliation of each state constant affects this relationship.



From this graph it can be seen that Republican states have a taller peak at a lower Protests per 100k value, suggesting a higher concentration of Republican states with a lower number of protests per 100k. While Democratic states have a lower peak and appear less steep, suggesting that Democratic states might have a more even distribution of Protests per 100k values, with less concentration around a single mode, and have higher protest numbers

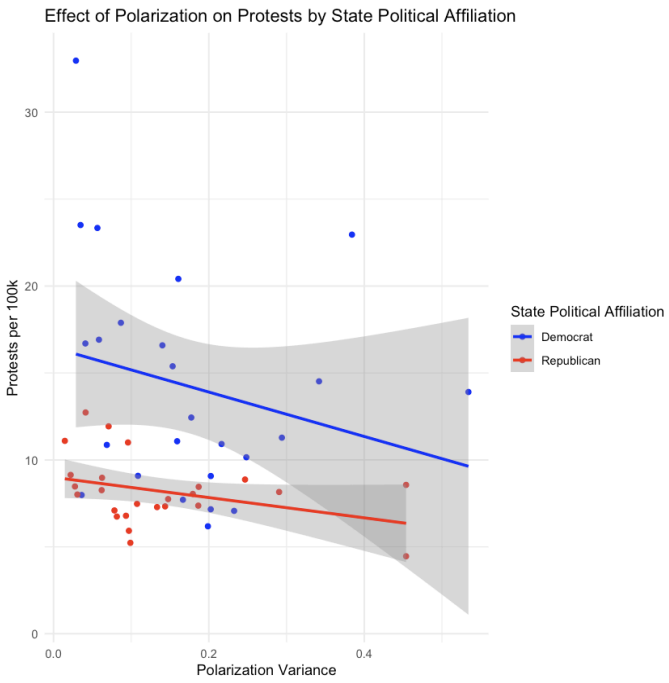
Examining the variable of political party affiliation, reveals a potential impact on the relationship between the number of protests and the political polarization of a state. This influence could be because of a variety of reasons including, party response to varying policy priorities and social issues, which can influence the likelihood and nature of protests. For example, states with a government of one party might adopt policies that are met with resistance from groups with opposing views, leading to increased protest activity. At the same time a state's dominant party affiliation can reflect its overall level of political polarization. In highly polarized states, where there is a division between the dominant party and the opposition, there may be an increase in protests as a form of political expression.

Further Analysis: Multiple Linear Regression using a categorical variable

In our extended analysis incorporating a categorical variable into the multiple linear regression model, we observe that the interaction terms are insignificant, which suggests that political polarization's impact on protest frequency does not differ significantly between states identified as predominantly Republican or Democrat. Here, voted_categoryRepublican serves as a

dummy variable, providing a means to assess the unique impact on states that voted Republican as compared to the reference group of Democrat-voting states.

The graphical representation of our model shows two lines representing Democrats and Republicans and their individual effects on the relationship between protests and polarization. Marked by shaded areas, the confidence intervals are portrayed for each group. The broader confidence intervals for states that voted Democrat may imply a higher level of uncertainty in their protest rate estimates or a more variable response to changes in political polarization, in contrast to the narrower intervals observed for Republican states.



A takeaway from this model is that political affiliation (Republican vs. Democrat) significantly influences the number of protests per 100k and it does to a certain extent effect change the slope of the line representing the relationship between protest and polarization, although it is still not statistically significant.

Conclusion

The analysis presented in this study aimed to unravel the intricate relationship between political polarization and protest frequency within U.S. states. While searching for a significant statistical link in Part B did not yield definitive results, the introduction of additional variables in Part C offered insights into the factors that may influence these civic behaviours.

The findings suggest that political party affiliation does influence protest activities. But with the inclusion of polarization in the model, its coefficient did not emerge as a significant predictor of protest frequency, suggesting that other, unaccounted-for variables might interact with its effect. Future research could benefit from a broader data collection, incorporating longitudinal data to track changes over time and employing different approaches to capture other variables that influence the nature of political expression.

Reference

Bureau of Labor Statistics, U.S. Department of Labor. (2024). *Local Area Unemployment Statistics*.https://www.bls.gov/web/laus/laumstrk.htm

CNN. (2020). *Election 2020: Presidential election results.* https://www.cnn.com/election/2020/results/president

Federal Reserve Bank of St. Louis. (2022). *Educational Attainment, Annual: Bachelor's Degree or Higher by State*. https://fred.stlouisfed.org/release/tables?eid=391444&rid=330

Federal Reserve Bank of St. Louis. (2022). *Real Median Household Income by State, Annual.* https://fred.stlouisfed.org/release/tables?eid=259515&rid=249

McDonald, M. (2022). 2022 General Election Turnout. *United States Elections Project.* https://election.lab.ufl.edu/voter-turnout/2022-general-election-turnout/

United States Department of Agriculture, Economic Research Service. (2023). *Percent of total population in Poverty.* https://data.ers.usda.gov/reports.aspx?ID=17826

U.S. Census Bureau. (2020). *Racial and Ethnic Diversity in the United States: 2010 and 2020 Census.* https://www.census.gov/library/visualizations/interactive/racial-and-ethnic-diversity-in-the-united-states-2010-and-2020-census.html

World Population Review. (2024). *High School Graduation Rates by State 2024.* https://worldpopulationreview.com/state-rankings/us-literacy-rates-by-state

World Population Review. (2024). *U.S. Literacy Rates by State.* https://worldpopulationreview.com/state-rankings/high-school-graduation-rates-by-state

Github Links to data set and replicate:

https://github.com/rpmcnair/Poli_381_Project