What Shapes Bias in Hate Crimes: Social, Economic, Political, or Pandemic Influence?

1. Problem Statement

Hate crimes in the U.S. have increased over the years, with certain states experiencing noticeable spikes in specific types of bias-motivated incidents. While studies have examined some factors like socioeconomic or political influences, they often focus on these areas in isolation. There is a need for research that integrates multiple factors and applies statistical models to understand their combined impact. This study aims to fill that gap by identifying which factors most influence hate crime overall and for Racial bias hate crime. Additionally, since anti-Asian hate crimes rose sharply during the COVID-19 pandemic (Kim, 2022), I explored which states experienced the largest increases, investigate the factors behind those changes, and determine whether these shifts were due to existing trends or the pandemic itself. I hypothesize that states with higher unemployment, lower GDP, strong political leanings, and younger, male-dominated populations are more likely to experience higher hate crime rates.

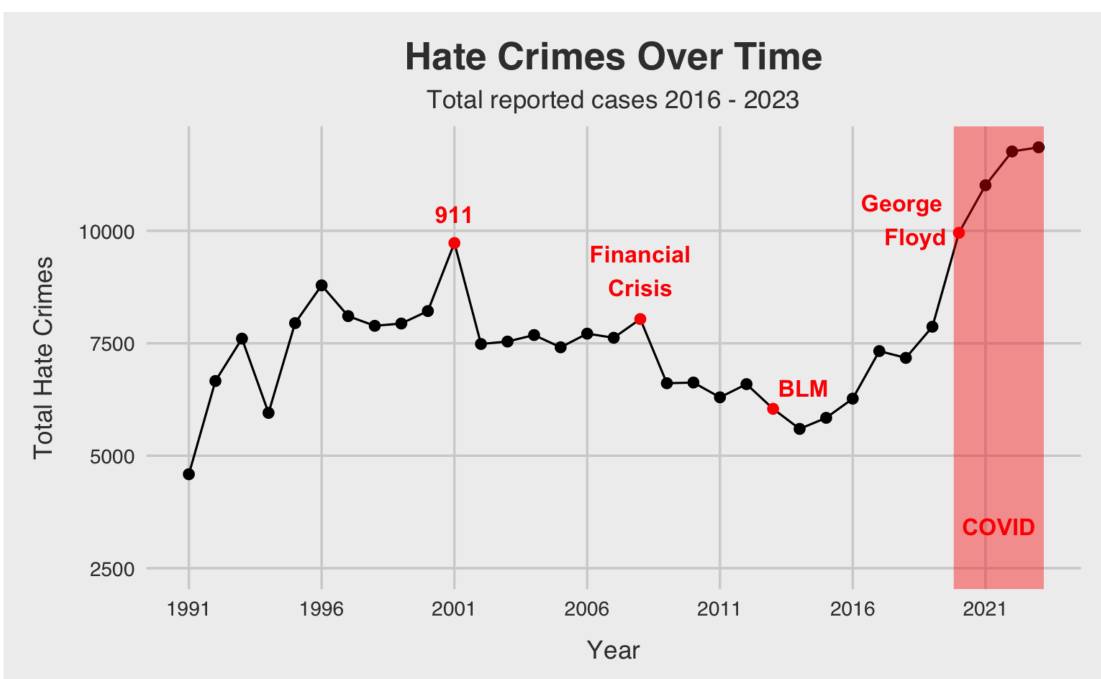
2. Data and Methods

For this study, I will use the hate crime statistics dataset provided by the FBI's Crime Data Explorer. This dataset offers detailed information about each reported hate crime incident, including bias type, offender's race, and the state where the crime occurred. However, one limitation is that the data only includes crimes officially reported to law enforcement agencies, meaning the true extent of hate crimes may be underreported (FBI, 2024). Some victims may avoid reporting incidents due to distrust or fear, leading to gaps in the data. Despite these limitations, the dataset provides valuable insights into state-level characteristics, allowing exploration of how various environmental factors shape biases. To complement the FBI data, I will integrate socioeconomic indicators, including GDP(BEA, 2023) and unemployment rate by state(NCSL, 2024) Additionally, demographic information, such as race and age distributions, will be sourced from the Kaiser Family Foundation (KFF, 2023). These datasets will help capture the economic and demographic environment surrounding offenders and incidents.

This research will apply several quantitative methods to explore the relationships between variables. Multiple linear regression will be used to analyze how factors influence the general prevalence of hate crimes, while Random Forest models will identify the most important factors for each bias type. For time-based analysis, I will use Interrupted Time Series (ITS) and logistic regression to examine changes in hate crimes before and after COVID-19. Finally, ARIMA forecasting will predict future trends in hate crimes based on historical data.

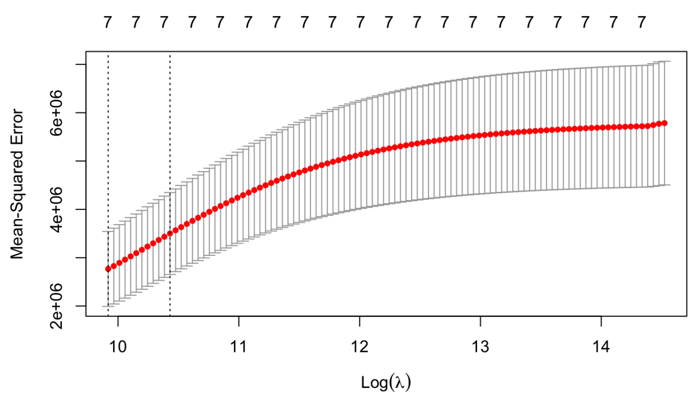
3. Early Results

3.1. Changes in Hate Crime Over Time



This graph presents the overall trend of total reported hate crimes from 1991 to 2023, highlighting key social events and their possible influence on hate crime rates. Each labeled point represents a major social change, such as the 9/11 attacks, the financial crisis, and the COVID-19 pandemic. These events correspond with noticeable surges in hate crimes, suggesting that such social changes may trigger temporary spikes in hate crime prevalence, followed by declines in subsequent years. The trend reveals that after COVID-19 began, hate crime rates surged to their highest levels since 1991. It raised the question: Are these fluctuations purely driven by social change, or do other socio-economic, demographic, and political factors play a role in shaping hate crime trends?

3.2. Correlation Between Factors and Hate Crime Prevalence

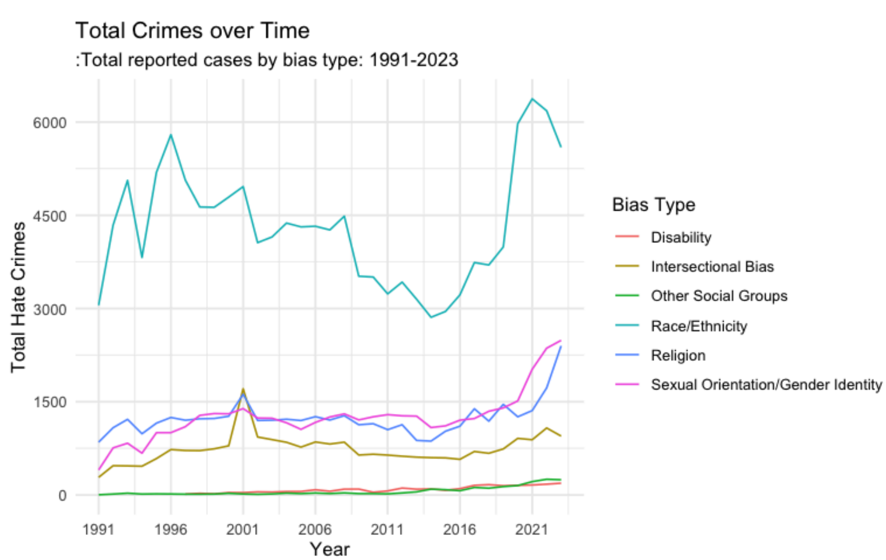



```
14 x 1 sparse Matrix of class "dgCMatrix"
                       s1
(Intercept)  1.550184e+04
0-18        -3.097522e+04
19-25       -1.834673e+04
26-34        .
35-54        .
55-64        .
65+          1.545935e+04
Asian        9.514373e+02
Black        .
Hispanic     1.852992e+04
White       -9.867858e+03
Female       .
Male         .
gdp          1.136480e-04
```

The ridge regression analysis(Friedman, 2010) explores the relationship between demographic, socioeconomic, and racial factors with the prevalence of hate crimes while addressing multicollinearity(Daoud, 2017). The results indicate that states with higher proportions of younger populations (0-18, 19-25) are associated with lower hate crime rates, while larger elderly populations (65+) correspond to higher rates. Additionally, the Hispanic population is positively linked to hate crime prevalence, while the coefficients for other racial groups (such as Black) were shrunk to near zero, suggesting limited predictive power or correlations with other variables. A small positive coefficient for GDP suggests that economic prosperity alone does not significantly reduce hate crimes, indicating a complex interplay between economic conditions and social behavior. Some variables, like gender, had their coefficients reduced to zero, suggesting they do not contribute meaningfully to explaining the trends. The cross-validation plot confirmed the optimal lambda value for minimizing prediction error, ensuring the model balances bias and variance effectively. This analysis highlights the nuanced influence of multiple factors on hate crime patterns, demonstrating that no single demographic or economic factor solely drives the trend.

3.3. Changes in Hate Crime Over Time By Bias Type

Hate crime is "typically one involving violence, that is motivated by prejudice on the basis of ethnicity, religion, sexual orientation, or similar grounds" (Oxford English Dictionary, n.d). There must be a related bias in order to define this crime as Hate crime. According to the hate crime report data which is being mainly used for this research contains four different categories of bias types (there are more sub-categories but in general four): racial, gender, ethnicity, and religion.



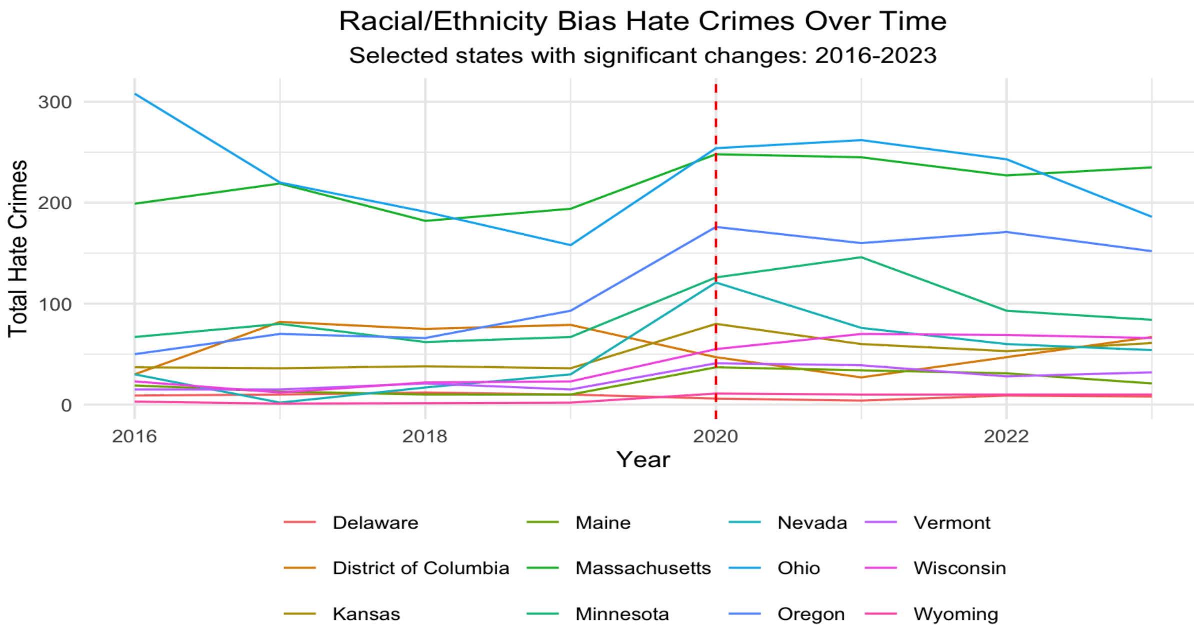
This graph reveals that hate crimes related to race and ethnicity are the most prevalent throughout the timeline, with noticeable spikes, particularly during the COVID-19 pandemic. The trends suggest that external events, such as the pandemic, can disproportionately influence certain biases, with racial and ethnic hate crimes showing the most significant increases.

3.4. Correlation Between Factors and Racial Bias Hate crimes

3.4.1. States that has significant shift of Racial Hate crimes after covid

state <chr>	p.value <dbl>
Delaware	0.035557297
District of Columbia	0.046693165
Kansas	0.016347032
Maine	0.000299700
Massachusetts	0.014938154
Minnesota	0.027190590
Nevada	0.024263527
Ohio	0.010149885
Oregon	0.014412496
Vermont	0.014899462
Wisconsin	0.008785477
Wyoming	0.001130289

Building on these insights, I narrowed the analysis to the state to examine how socioeconomic, demographic, and political factors influence racial bias within that context. Interrupted Time Series (ITS) analysis was applied to evaluate the impact of COVID-19 on racial/ethnic bias-motivated hate crimes across U.S. states. ITS allows us to identify significant changes by comparing trends before and after the onset of the pandemic in 2020, treating it as an intervention point. This approach highlights states where the shift in hate crime trends was statistically significant, as indicated by the p-values in the results table. States such as Minnesota, Oregon, and Ohio displayed significant changes post-2020, suggesting that COVID-19 may have influenced these shifts.



While the graph shows the trends visually, the slope does not always correspond perfectly to the statistical significance identified by ITS. The p-value calculation considers more than just the visual trend—it accounts for data variability and the consistency of changes over time. This makes ITS a suitable method for identifying meaningful shifts in hate crimes during the COVID-19 period. However, there are limitations: data sparsity in some states may reduce the power of the tests. Despite these challenges, ITS provides a structured way to assess whether observed changes are linked to the pandemic or broader societal factors.

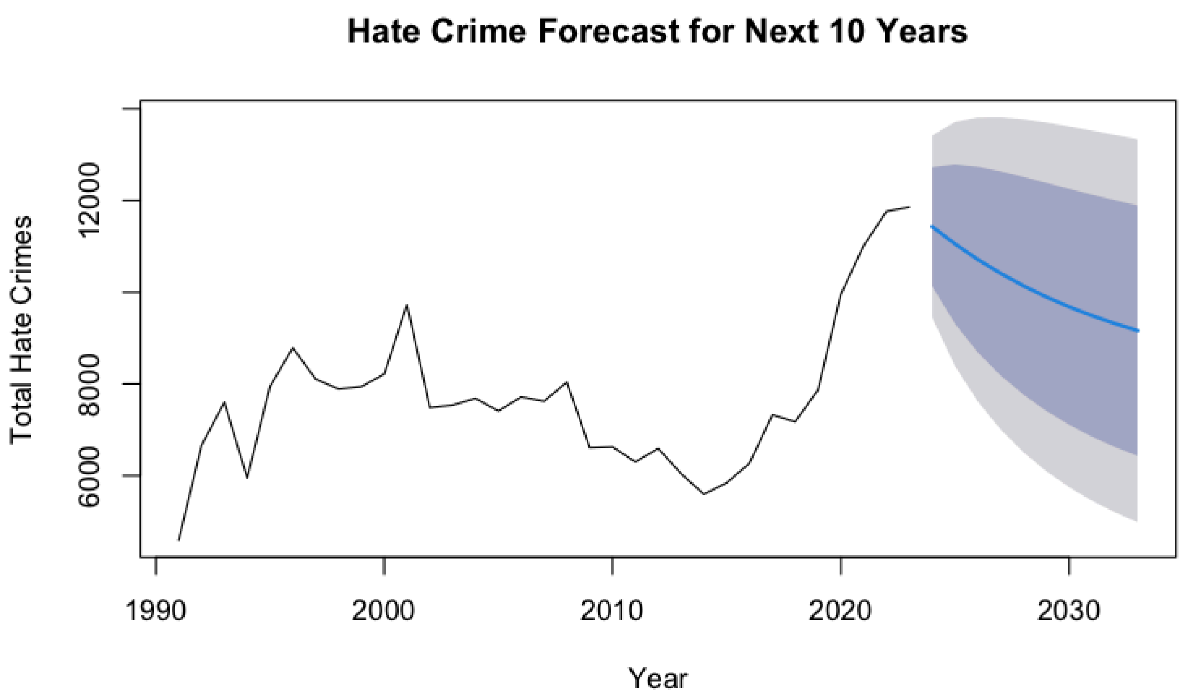
3.5. Correlation Between Factors and Racial Bias Hate crimes

Having identified these states, I then applied panel regression to explore which social factors were most closely associated with the observed changes. This approach helped determine whether socio-economic conditions were more influential in driving racial hate crimes.

```
Oneway (individual) effect Within Model

Call:
plm(formula = total_hate_crimes ~ Male + young + middle_Aged +
    Asian + Black + Hispanic + gdp + density + avg.emp + political,
    data = panel_data, model = "within")

Unbalanced Panel: n = 11, T = 6-7, N = 76

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-59.67713 -16.82037   0.43688  12.20711  57.29824

Coefficients:
                      Estimate  Std. Error t-value Pr(>|t|)
Male                1.8370e+03  1.0395e+03  1.7671 0.082763 .
young              -1.0568e+03  6.3617e+02 -1.6612 0.102373
middle_Aged         1.4931e+03  7.7671e+02  1.9223 0.059756 .
Asian              -1.7324e+02  1.2034e+02 -1.4396 0.155645
Black               1.0997e+03  1.0725e+03  1.0254 0.309671
Hispanic            2.5187e+03  9.2155e+02  2.7331 0.008422 **
gdp                 8.8237e-04  3.4610e-04  2.5495 0.013606 *
density            -9.0779e-01  8.5680e-01 -1.0595 0.293997
avg.emp             1.3040e+01  6.1674e+00  2.1144 0.039032 *
politicalRepublican -6.1484e+01  2.3796e+01 -2.5838 0.012457 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    130520
Residual Sum of Squares: 46544
R-Squared:      0.64341
Adj. R-Squared: 0.51374
F-statistic: 9.9239 on 10 and 55 DF, p-value: 3.02e-09
```

The fixed effects model provides insights into the relationship between hate crimes and socio-economic, demographic, and political factors(FEC, 2020) across selected U.S. states between 2016 and 2023. Significant results include a positive relationship between higher GDP and hate crimes ($p < 0.05$) and between Hispanic population proportions and hate crimes ($p < 0.01$). The model also found that Republican-leaning states tend to have lower hate crime rates ($p < 0.05$). With an R-squared value of 0.6434, the model explains 64.34% of the variance in hate crime rates, and the F-statistic ($p < 0.001$) confirms the overall model significance. However, demographic factors like "Male," "Young," and "Black" showed less robust statistical significance, suggesting their relationships with hate crimes may be more complex.

This method is appropriate because it controls for unobserved, state-specific factors that remain constant over time, reducing the risk of omitted variable bias. By focusing on within-state variations, it effectively highlights how dynamic variables, such as economic conditions and employment, influence hate crime trends. However, the model has limitations. It does not capture between-state differences, and the need to drop variables such as "White" and "Old" due to multicollinearity limits the depth of demographic insights. Additionally, the availability of political data only for election years makes its interpretation more nuanced. Despite these challenges, the fixed effects approach offers a valid framework for understanding the social and economic factors driving changes in hate crime rates over time.



The ARIMA model forecasts the trend in hate crimes over the next ten years. The prediction suggests that while the total number of hate crimes may stabilize or slightly decline in the immediate years post-2023, the confidence intervals widen over time, reflecting increasing uncertainty in the forecast. The blue line in the shaded area represents the model’s mean forecast, while the gray bands capture the 80% and 95% confidence intervals. This indicates that, although the central forecast predicts a slight decrease, there is still a range of possible outcomes where the total hate crimes could increase or remain stable, depending on external factors such as societal or political events. While the ARIMA model is effective for capturing past trends and projecting them into the future, it is limited in accounting for sudden, unanticipated events like another pandemic or significant socio-political movements.

4. Conclusion

The preliminary findings suggest that COVID-19 influenced racial hate crime trends, particularly in selected states identified through ITS analysis, such as Minnesota, Oregon, and Ohio. However, the fixed-effects panel regression indicates that other social and economic factors, such as GDP and the proportion of Hispanic populations, also played significant roles in shaping these patterns. This implies that the pandemic was not the sole driver of the increase in hate crimes, highlighting the interplay between socio-economic conditions and bias. For the next steps, I will refine the predictive model to explore future hate crime trends, accounting for additional factors such as economic and demographic shifts. I also plan to revisit the panel regression to include other relevant predictors or revise the methodology to ensure robust insights. Furthermore, limitations related to missing data and variable multicollinearity will be addressed to strengthen the reliability of the final results.

References

Jueun Park. (2023). Hate-Crime-Pattern-Analysis [Software]. GitHub. https://github.com/jueun7954/Hate-Crime-Pattern-Analysis

BEA. (2023). Regional Data GDP and Personal Income [Data set]. https://apps.bea.gov/itable/?ReqID=70&step=1#eyJhcHBpZCI6NzAsInN0ZXBzIjpbMSwyOSwyNSwzMSwyNiwyNywzMF0sImRhdGEiOltbIlRhYmxlSWQiLCI1MTIiXSxbIk1ham9yX0FyZWEiLCIwIl0sWyJTdGF0ZSIsWyIwIl1dLFsiQXJlYSIsWyJYWCJdXSxbIlN0YXRpc3RpYyIsWyIxIl1dLFsiVW5pdF9vZl9tZWFzdXJlIiwiTGV2ZWxzIl0sWyJZZWFyIixbIjIwMjMiLCIyMDIyIiwiMjAyMSIsIjIwMjAiLCIyMDE5IiwiMjAxOCIsIjIwMTciLCIyMDE2Il1dLFsiWWVhckJlZ2luIiwiLTEiXSxbIllllYXJfRW5kIiwiLTEiXV19

Daoud, J. I. (2017). Multicollinearity and regression analysis. Journal of Physics: Conference Series, 949, 012009. https://doi.org/10.1088/1742-6596/949/1/012009

FBI Crime Data explorer. (2024). Hate_Crime.csv [Data set]. https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/downloads

Federal Election Commission. (2020). Federal election results [Data set]. https://www.fec.gov/introduction-campaign-finance/election-results-and-voting-information/

FRED Economic Data. (2023). populationState [Data set]. https://fredaccount.stlouisfed.org/datalists/330027/download

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1–22. https://doi.org/10.18637/jss.v033.i01

KFF. (2023). Population Distribution by Age [Data set]. https://www.kff.org/other/state-indicator/distribution-by-age/?currentTimeframe=6&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

KFF. (2023). Population Distribution by Race/Ethnicity [Data set]. https://www.kff.org/other/state-indicator/distribution-by-raceethnicity/?currentTimeframe=6&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

KFF. (2023). Population Distribution by Sex [Data set]. https://www.kff.org/other/state-indicator/distribution-by-sex/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

Kim, C., Lee, C. S., & Lim, H. (2023). Hate-motivated crime/incidents against Asians in the United States of America: A systematic review. Race and Justice, 13(1), 9–31. https://doi.org/10.1177/21533687221117280

NCSL. (2024). State Unemployment Rates [Data set]. https://www.ncsl.org/labor-and-employment/state-unemployment-rates

Oxford English Dictionary. (n.d). Oxford English Dictionary. Retrieved 2023, from https://doi.org/10.1093/OED/1735003731.

World Population Review. U.S. States by Size in Square Miles [Data set]. https://worldpopulationreview.com/state-rankings/states-by-area