Hate Crimes Analysis – Final Project

Ahmad Murtaza Sahab (2745651)

22-08-2025

Contents

Instructor: *Dr. Cem BAŞ*

STAT201: Introduction to Probability & Statistics I

This final project prepared in RStudio.

---

1 Research Question

The research question of this study is whether **income inequality** effects on **hate crime** in the United States of America.

2 Data Source

The dataset has been downloaded from Our data please ->**Click FiveThirtyEight!**

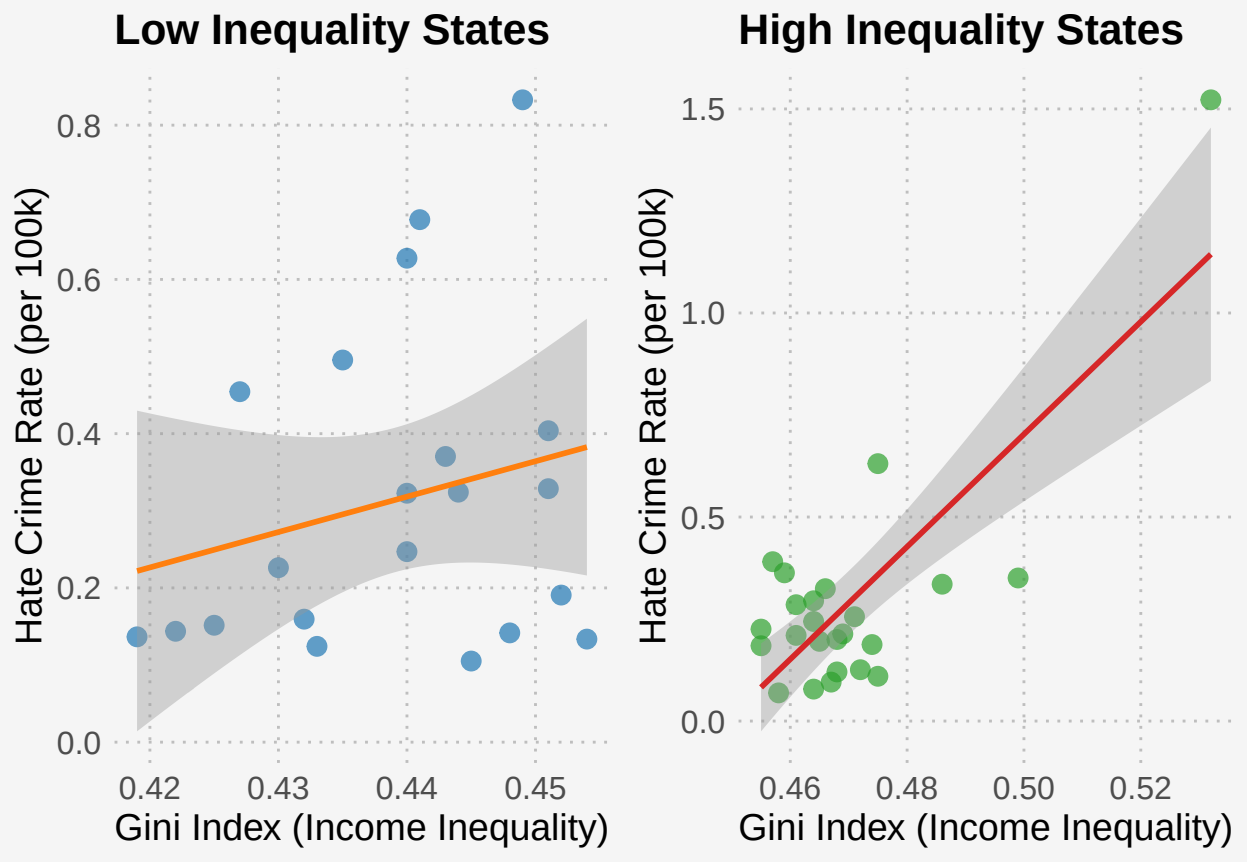
3 Summary of Dataset

Table 1: Summary Statistics of Hate Crimes

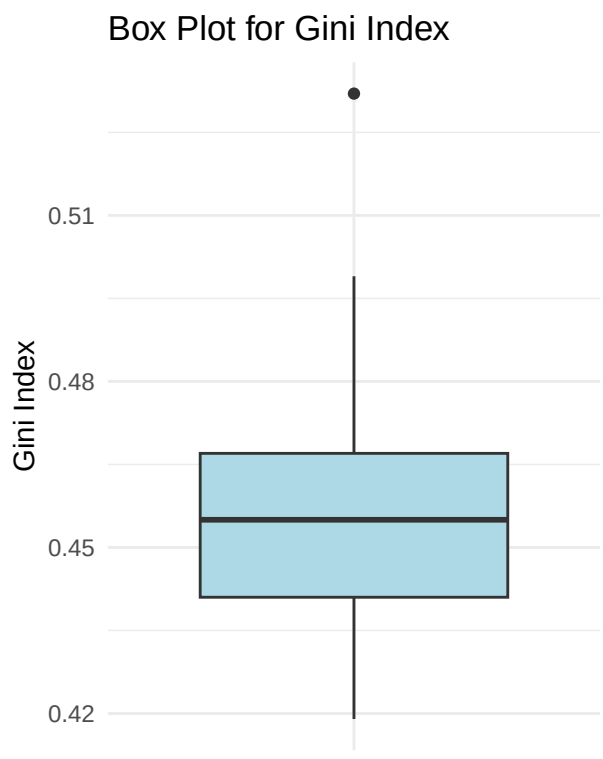
	Median	Mean	Min	Max	1stQu	3rdQu
avg_hatecrimes_per_100k_fbi	1.94	2.37	0.41	10.95	1.32	3.14
gini_index	0.46	0.46	0.42	0.53	0.44	0.47
hate_crimes_per_100k_splc	0.23	0.30	0.07	1.52	0.14	0.35
median_household_income	54916.00	55299.49	39552.00	76165.00	48060.00	60708.00
share_non_citizen	0.05	0.06	0.01	0.13	0.03	0.08
share_non_white	0.30	0.32	0.06	0.63	0.21	0.42
share_population_in_metro_areas	0.81	0.78	0.34	1.00	0.69	0.90
share_population_with_high_school_degree	0.87	0.87	0.80	0.92	0.84	0.89
share_unemployed_seasonal	0.05	0.05	0.03	0.07	0.04	0.06
share_voters_voted_trump	0.49	0.48	0.04	0.69	0.41	0.57
share_white_poverty	0.09	0.09	0.04	0.17	0.07	0.10

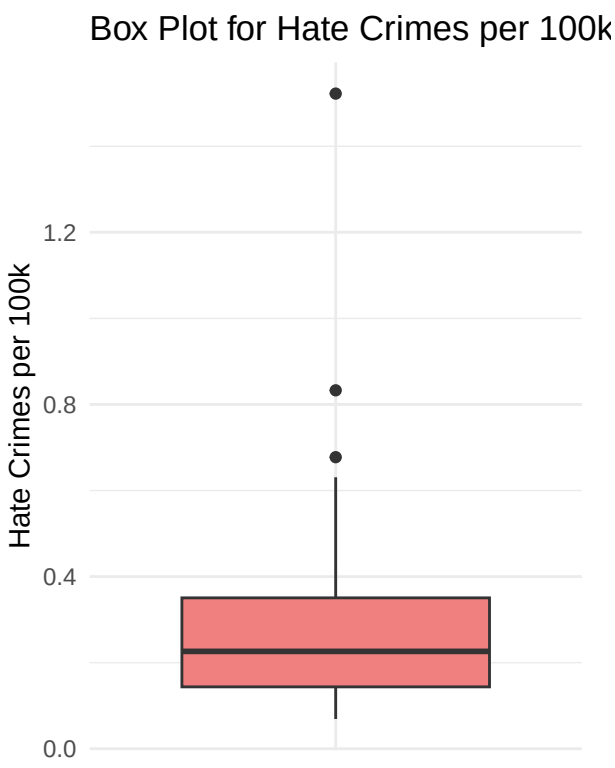
The data reveals that the majority of U.S. regions are predominantly urbanized, have a well-educated population, and demonstrate economic stability, with a median household income around $55,000 and a high school graduation rate of approximately 87%. However, there is significant variation in hate crime rates—while the FBI's average stands at 2.37 incidents per 100,000 individuals, certain areas report substantially higher levels. Additionally, demographic trends show marked differences, particularly in the percentages of non-white and non-citizen residents. Politically, the nation appears almost evenly divided, with around half of voters supporting Trump. As our understanding, the results imply that both demographic and political factors may play a crucial role in influencing regional disparities in hate crime statistics.

4 Plots



Low inequality states indicate that there is a weak and uncertain positive correlation as shown by the regression line and extensive confidence interval, while **high inequality** states show that there is a strong and clear positive correlation pointed out by the regression line and narrow confidence interval.





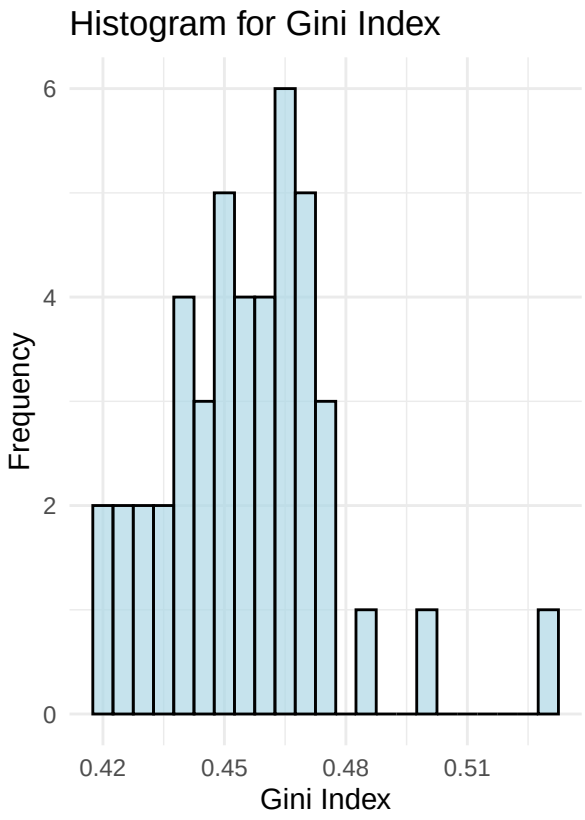
Box Plot Analysis

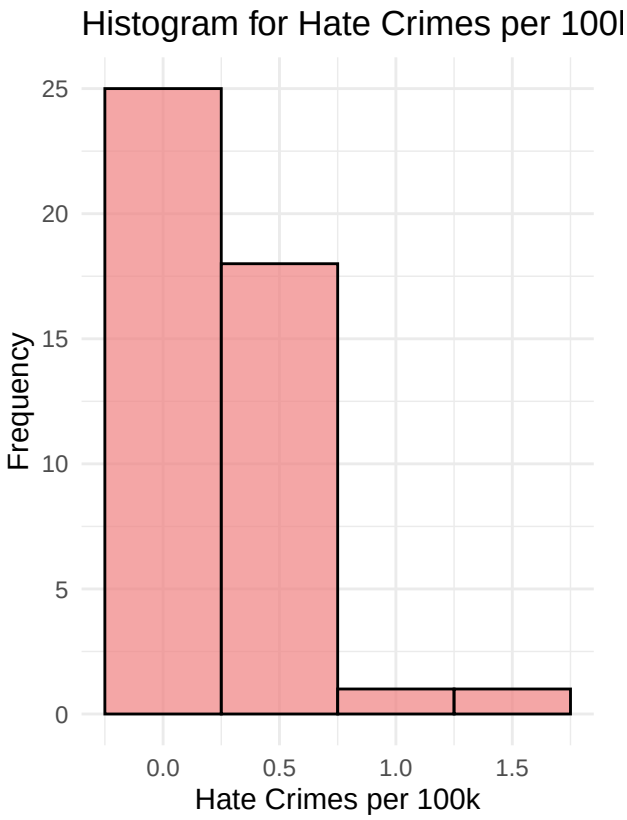
Gini Index (Income Inequality)

The distribution is relatively compact, with a majority of states centered around a median of approximately 0.45. The range is limited (minimum around 0.42, maximum about 0.53), indicating that the levels of inequality do not differ significantly from one state to another. There are very few possible outliers, suggesting that inequality tends to be quite uniform across the United States.

Hate Crimes per 100k

The distribution exhibits a right skewed (positively skewed). The median is low (around 0.23), but the highest value is significantly elevated (approximately 1.52), which indicates that although most states report low rates of hate crimes, a select few have much higher occurrences. The existence of outliers points to considerable disparities among states in terms of hate crime frequency. The stability of income inequality is consistent across various states, exhibiting minimal variation. In contrast, hate crime rates present significant differences, with most states reporting low levels, while a select few stand out as extreme outliers. This indicates that inequality by itself might not completely account for the distribution of hate crimes, although states experiencing greater inequality frequently align with elevated hate crime rates.





Histogram Analysis

Gini Index

The histogram exhibits a bell shape and is symmetrical, with its center around 0.45. This indicates that the disparity among states follows a distribution similar to a normal curve, lacking any extreme values. The majority of states are clustered within the range of 0.44 to 0.47.

Hate Crimes per 100k

The histogram shows a positive skew. A large number of states report very low hate crime rates (under 0.5), whereas a handful of states have significantly higher rates (over 1.0). This skewness reflects the pattern observed in the box plot, where a few states elevate the average.

4.1 Table

Table 2: Summary Statistics for Gini Index and Hate Crimes

Statistic	Measure	Value
Gini Index	Mean	0.455688888888889
Gini Index	Median	0.455
Gini Index	Min	0.419
Gini Index	Max	0.532
Gini Index	SD	0.0209021771617459
Gini Index	Skewness	0.961223577150097
Hate Crimes	Mean	0.302429972533333
Hate Crimes	Median	0.226197105
Hate Crimes	Min	0.069060773
Hate Crimes	Max	1.52230172
Hate Crimes	SD	0.251562831268793
Hate Crimes	Skewness	2.88993748263845

4.2 Analysis

Table 3: Analysis Results Summary

Description	Result
Confidence Interval for hate_crimes_per_100k_splc	0.22685213110574, 0.378007813960926
Confidence Interval for gini_index	0.449409179668516, 0.461968598109262
Probability (Normal Distribution < 60)	0.8413
Probability (Binomial: Exactly 3 heads in 10 flips)	0.1172
Probability (Poisson: 2 events with mean 1)	0.1839

4.3 Summary of Analysis

Confidence Intervals:

Hate Crimes per 100k: The confidence interval of about (0.226, 0.378) implies that we can be 95% confident that the actual average of hate crimes per 100,000 individuals lies within this range. This interval reflects a relatively low incidence of hate crimes in the community.

Gini Index: The confidence interval of roughly (0.449, 0.461) points to a significant level of income inequality in the population being examined. This interval indicates that the Gini index consistently exceeds 0.4, a level commonly linked to considerable inequality.

Probability Calculations:

Normal Distribution Probability: A probability of 0.8413 suggests that there is an 84.13% chance for a randomly chosen observation to be below 60, indicating that this value is not considered extreme within the context of the normal distribution.

Binomial Probability: The probability of 0.1172 signifies the chance of getting exactly 3 heads in 10 coin flips with a fair coin, demonstrating the characteristics of binomial distributions in random experiments.

Poisson Probability: A probability of 0.1839 shows a moderate chance of witnessing 2 events when the mean is 1 in a specified time frame, which is important for studying infrequent occurrences.

Table 4: Linear Regression Model

term	estimate	std.error	statistic	p.value
(Intercept)	-1.527463	0.7833043	-1.950025	0.0574197
gini_index	4.020510	1.7177215	2.340606	0.0237445

The result indicate that there is positive relationship between the Gini Indix and Hate crime **0.0237445** which is less than **0.05**.Therefore, an increase in Gini Indix, the hate crime rate increases around **4.020510**.

5 Conclusion

This research examines the impact of demographic, social, and economic factors on regional dynamics in the United States. While education levels, urbanization, and economic indicators in most regions generally present a strong picture, there are significant differences in ethnicity, income levels, and political affiliations. Hate crimes, while low on

average, exhibit significant local inequalities, highlighting the need to consider social tensions alongside economic and political patterns. A linear regression analysis conducted within this framework investigated the link between income inequality (as measured by the Gini index) and hate crimes per 100,000 people. The findings indicate that the regression curve is **positive**, meaning that hate crimes are reported more frequently in states with higher income inequality. Although the Gini index varies within a certain range (0.42–0.53), even small increases are associated with significant increases in hate crimes. The explanatory power of the model indicates that income inequality explains some, but not all, differences in these crimes.

Overall, the findings demonstrate that regional variation is not only cultural but also structural, and that hate crimes are not limited to economic factors. Reducing hate crimes requires a multifaceted approach that requires not only addressing income inequality but also addressing prejudice, intolerance, and lack of awareness. This has significant implications for policy, governance, and social harmony.

6 References

GitHub: https://github.com/fivethirtyeight/data/tree/master/hate-crimes

Paul Newbold, William Carlson, Betty Thorne - Statistics for Business and Economics, Global Edition (2022, Pearson)

Montgomery, Engineering Statistics, 5th Edition-John Wiley (2010)