UNITED STATES MILITARY ACADEMY

FINAL PROJECT

MA478: GENERALIZED LINEAR MODELS

HOUR H2

COL NICHOLAS CLARK

By

CADET GARRETT CHRISMAN ‘24, CO E2

WEST POINT, NEW YORK

07 MAY 2024

GC I CERTIFY THAT I HAVE COMPLETELY DOCUMENTED ALL SOURCES THAT I USED TO COMPLETE THIS ASSIGNMENT AND THAT I ACKNOWLEDGED ALL ASSISTANCE I RECEIVED IN THE COMPLETION OF THIS ASSIGNMENT.

______ I CERTIFY THAT I DID NOT USE ANY SOURCES OR RECEIVE ANY ASSISTANCE REQUIRING DOCUMENTATION WHILE COMPLETING THIS ASSIGNMENT.

SIGNATURE: CHRISMAN.GARRETT.TODD.1593580586 Digitally signed by CHRISMAN.GARRETT.TODD.1593580586 Date: 2024.05.06 14:21:26 -04'00'

MA478 Final Project Report: Chicago Burglary Data

CDT Garrett Chrisman

May 6, 2024

Abstract

This report investigates the factors influencing burglary rates in Chicago using Poisson, Mixed Effects, and "Chiraq" models. The Poisson model identified unemployment, young male population, wealth, and monthly variations as significant predictors. The Mixed Effects model showed wealth and monthly variations significantly influence crime rates, while unemployment had a non-significant impact. The Chiraq model, using an INLA framework, provided the most comprehensive approach, identifying wealth as a strong positive predictor and revealing a negative trend in crime rates over time. While the Chiraq model demonstrated the best overall fit, it struggled to accurately predict higher crime rates. Future work could focus on refining the Chiraq model by exploring additional predictors, investigating alternative model structures, and incorporating policy changes and school patterns. Ethical considerations emphasize the importance of responsible communication of model results to avoid stigmatization and inform evidence-based policies that promote public safety and address the root causes of crime.

Keywords: Chicago, crime, burglaries, socioeconomic factors, statistical modeling

1 Introduction

Chicago, often referred to in hip-hop as "Chiraq," earned this nickname due to its history of crime and violence. Although controversial, the term highlights the city's ongoing struggle with these issues, particularly burglaries. Understanding the factors contributing to burglaries is crucial for developing effective crime prevention strategies and allocating resources to the most affected areas. The incidence of burglaries in Chicago has historically fluctuated, with factors like poverty, inequality, and limited access to education and job opportunities significantly contributing to crime's persistence. Identifying the root causes of these fluctuations can help policymakers and law enforcement agencies develop targeted interventions to reduce burglary rates. However, the city's challenges vary by neighborhood. Research shows a clear spatial distribution of burglaries, emphasizing the role of local socioeconomic factors. Analyzing the spatial patterns of burglaries can reveal insights into the specific needs and challenges of different neighborhoods, enabling a more tailored approach to crime prevention. Areas with higher poverty and unemployment rates typically experience more burglaries, in contrast to more affluent neighborhoods. This disparity underscores the dynamic nature of crime patterns, requiring a detailed understanding of their causes. Investigating the relationship between socioeconomic factors and burglary rates can shed light on the underlying mechanisms driving these disparities and guide efforts to address systemic issues. In this report, statistical modeling is a crucial approach to dissecting the complex factors influencing burglary rates, thereby informing targeted crime prevention and community safety measures. By leveraging data-driven insights, policymakers and law enforcement agencies can make evidence-based decisions to allocate resources effectively, improve community outreach, and ultimately reduce the incidence of burglaries in Chicago. [1]

1.1 Problem Statement

Given the noticeable disparities in burglary rates across different Chicago neighborhoods, what model can accurately capture the impact of seasonal changes, population, and wealth index on burglary counts, and account for spatial and temporal variability?

2 Sources of Variation Diagram

Table 1: Sources of Variation in Burglary Counts in Chicago

Components	**Explained Variation**	**Unexplained Variation**
Observational Units: Census Block groups in Chicago (552 Locations) **Response Variable**: Count of Burglaries	• Population • Unemployment Measure • Wealth Measure • Weather Conditions • Measure of Young Males	• Individual Criminal intent • Specific local security measures • Unpredicted temporal events • Random local events

3 Research Questions

1. What are the key socioeconomic and demographic factors that influence burglary rates in Chicago neighborhoods, and how do their effects vary across different regions of the city?

2. How do seasonal patterns, district socioeconomic profiles, and wealth indices contribute to the spatial and temporal variability of burglary incidence in Chicago over a five-year period, and which factors have the strongest influence on burglary rates?

4 Hypotheses

1. **H1:** Higher levels of poverty and unemployment in a Chicago neighborhood are positively associated with higher burglary rates, with the strength of this association varying across different regions of the city.

2. **H2:** A mixed-effects model, accounting for both fixed (socioeconomic status, unemployment, etc.) and random (seasonal variations) effects, will provide a more accurate prediction of burglary rates across Chicago neighborhoods compared to models considering only fixed effects, with the relative importance of fixed and random effects varying over time.

5 Literature Review

My literature review started with a simple Google Scholar search using the keywords "burglaries crime data temporal-spatial INLA," which were relevant to the papers I was looking to review. To my surprise, there were plenty to choose from. This literature review captures the examination of two papers that investigate prior research and systematic reviews of modeling spatio temporal crime data. As law enforcement agencies seek to better understand patterns and trends in criminal activity to inform resource allocation and prevention strategies. Variations in the methodologies and assumptions were prevalent based on the regions and data analysis done before the modeling to fit the specific data researchers were investigating. Thus, to model the complex spatial and temporal dependencies in crime data, Bayesian models with INLA were the focus to identify areas of high risk or "hotspots" [1].

To capture and highlight the effects of spatial and temporal random effects, Bayesian hierarchical models with INLA, which are easier and more efficient than prior methods, were used. In the paper, "Spatial-temporal Modelling and Mapping for Crime Data in Nairobi County," researchers used a Bayesian spatial-temporal model to analyze robbery crime rates in Nairobi, Kenya at the sub-county level over an 8-year period [2]. They found that robbery crime rates were positively associated with poverty levels and unemployment rates which we wanted to capture in our modeling.

Similarly, [1] applied Bayesian methods to analyze police calls-for-service data in Waterloo, Canada at the small-area scale over a 1-year period [1]. They at the overall spatial and temporal trends, as well as space-time interactions, to identify areas that depart from the overall patterns. Both papers were use cases on how to format and capture the spatial and temporal effects of their area of study. In relation to our Chicago crime data, these studies allowed for a better understanding of why the effects are placed as either fixed effects or random effects and which one is captured in relation to the research questions and area of study. An advantage of the Bayesian approaches used in these studies is the ability to quantify uncertainty in parameter estimates and predictions through the posterior distributions. This is particularly important when dealing with data at spatial and temporal scales.

Another consideration in spatial-temporal modeling of crime is the choice of spatial and temporal units. Census areas were used as the spatial unit and 2-hour time periods as the temporal unit [1], while sub-counties and years were used, similar to our defined areas and year data in Chicago [2]. Looking at the researchers' challenges and limitations, crime data are often subject to underreporting and geocoding errors, which can bias spatial patterns. There may also be important covariates that are unmeasured or unavailable, leading to residual confounding [2]. Furthermore, the complex social and environmental factors that drive crime are difficult to fully capture in a statistical model where we expect change to occur over our designated period. In our project analyzing Chicago burglary data, we aim to identify these challenges and attempt to address them in our assumptions when investigating socioeconomic factors and examining their impact on spatial and temporal effects. We hope to provide valuable insights that can support evidence-based policing and crime prevention efforts in Chicago.

6 Exploratory Data Analysis

6.1 Data Description

The dataset, obtained from the nick3703 GitHub account, contains count data for burglaries across 552 different locations in Chicago. The data is provided on a monthly basis and includes additional subsets such as wealth, unemployment, and population.

Basic data analysis on the data shows that there is some correlation between total crime and the count of male youths and having a higher wealth index score. The unemployment rate had a low score of correlation to crime showing burglaries are happening in larger populated locations that are wealthier with a high count of young males. The plots in the appendix supports these claims.

6.2 Data Analysis

Figure 2 displays the histogram of total crime rates, revealing a right-skewed distribution. This indicates that while most locations experience relatively low levels of crime, a few areas are disproportionately affected by high crime rates.

The time series plot of mean burglary rates, shown in Figure 3, exhibits a fluctuating pattern over the observed period. This plot suggests the presence of seasonal or temporal trends in burglary rates, demonstrating a noticeable decreasing pattern from 2010 to 2015. The inclusion of both month and year in the models is intended to capture this relationship.

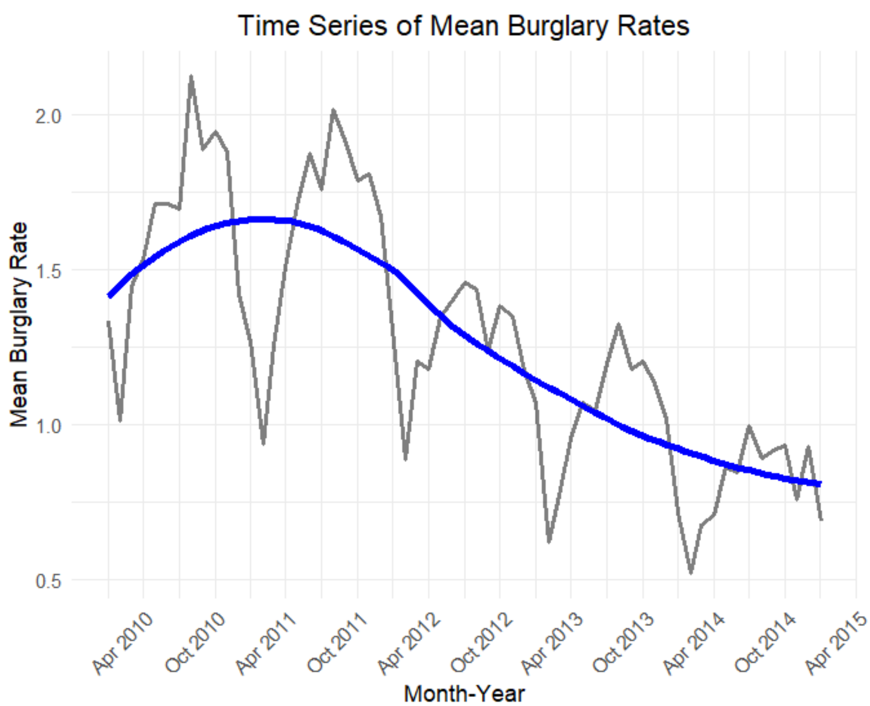

Figure 1: Time Series of Mean Burglary Rates

Figure 4 presents a scatter plot of total crime versus population, with data points colored by the unemployment rate. This plot indicates a positive correlation between crime and population, suggesting that highly populated areas tend to experience higher crime rates. The lack of a clear pattern in the coloring of data points based on the unemployment rate suggests that the relationship between unemployment and crime is not straightforward.

Figure 6 illustrates the relationship between the wealth index and total crime rates. This scatter plot suggests a positive correlation, implying that burglaries tend to occur more frequently in wealthier neighborhoods. This finding challenges the common assumption that crime is primarily associated with low-income areas and highlights the need for further investigation into the motivations and patterns of burglaries in rich communities.

Figure 5 shows the correlation between the count of young males in a population and the total crime rates. This scatter plot indicates a general trend where higher counts of young males correspond with higher crime rates. However, the considerable variability in the data points suggests that this relationship is complex and may be influenced by other factors. This could be a critical area for further analysis.

This data analysis serves as the foundation for developing a series of 3 models, increasing in complexity, to attempt to model the relationships between economic factors, space, and time to crime rates in Chicago.

7 Methodology: Three Proposed Models

7.1 Poisson Regression

The first model implemented was a Poisson Regression, which serves as a base count regression model. The response variable, crime rate, was modeled against several covariates to capture factors influencing the variability in crime rates. The covariates included unemployment rate, count of young males, wealth index, year, and month number. A Goodness of Fit (GOF) test was conducted to ensure the Poisson distribution was suitable for the data, and overdispersion was checked.

Model equation:

$$\log(\lambda_i) = \beta_0 + \beta_1 \cdot \text{unemp}_i + \beta_2 \cdot \text{ym}_i + \beta_3 \cdot \text{wealth}_i + \beta_4 \cdot \text{year}_i + \beta_5 \cdot \text{month\_num}_i \tag{1}$$

where λ_i is the expected count of crime rates in the i-th neighborhood.

7.2 Mixed Effect Model

We propose a Mixed Effects model that allows for the inclusion of fixed covariates of interest and random effects to capture the unobserved variability among different IDs and over different years.

$$\log(\lambda_{it}) = \beta_0 + \beta_1 \cdot \text{unemp}_{it} + \beta_2 \cdot \text{ym}_{it} + \beta_3 \cdot \text{wealth}_{it} + \beta_4 \cdot \text{month\_num}_{it} + b_i + d_t \tag{2}$$

$$Y_{it} \sim \text{Poisson}(\lambda_{it}) \tag{3}$$

where λ_{it} represents the expected count of crime rates.

In the model, Y_{it} represents the count of crime rates, i indexes the ID, and t indexes the time (year). The coefficients $\beta_1, \beta_2, \beta_3, \beta_4$ correspond to the effects of unemployment rate, youth males, wealth, and month number, respectively. Random effects are included for each ID (b_i) and for each year (d_t).

Random effects b_i and d_t are assumed to be normally distributed, capturing the unobserved heterogeneity across IDs and years:

$$b_i \sim \mathcal{N}(0, \sigma_b^2)$$
$$d_t \sim \mathcal{N}(0, \sigma_d^2)$$

7.3 Chiraq Model (INLA)

The final model was a Poisson regression with a structure to account for various fixed and random effects. The model includes random walk components for year and month (with a cyclic structure for the month), and independent identically distributed random effects for each ID.

$$\begin{aligned} n_{it} &= \beta_0 + \beta_1 \cdot \text{unemployment}_i + \beta_2 \cdot \text{wealth}_i + \beta_3 \cdot \text{year}_i \\ &\quad + u_i + v_{t,\text{year}} + w_{t,\text{month}} \\ Y_{it} &\sim \text{Poisson}(\lambda_{it}) \\ n_{it} &= \log(\lambda_{it}) \end{aligned} \tag{4}$$

where λ_{it} is the expected count of crime rates, Y_{it} represents the crime rates in the i-th district at time t, β_0 is the intercept, $\beta_1, \beta_2, \beta_3$ are the coefficients for unemployment rate, wealth index, and year, respectively. The u_i represents the iid random effect for district i, $v_{t_{\text{year}}}$ is the random effect for year t, modeled with a random walk of order 1 (RW1), and $w_{t_{\text{month}}}$ is the random effect for month t, also modeled with RW1 and assuming a cyclic structure.

8 Results

describe the specifics of what you did (data exploration, data preparation, model building, model selection, model evaluation, etc.), and what you found out (statistical inference and/or predictive modeling, interpretation and discussion of the results, etc.). Be sure to have good data visualizations throughout.

8.1 Poisson Model

The Poisson regression model looked at the impact of unemployment, young male demographics, wealth, year, and month (all the variates provided) on crime rates showed several statistically significant predictors. The intercept is significant at 0.1979, indicating a baseline crime level. Unemployment positively affects crime rates with a coefficient of 0.0241. The effect of young males on crime is also significant at 0.0023, although negligible. Wealth correlates with increased crime, evidenced by a coefficient of 0.1319. Yearly trends show a decrease in crime with a coefficient of -0.1651, and month-to-month variations are significant at 0.0328, suggesting seasonal effects we looked at in the data analysis. With an AIC of 120508 and residual deviance of 61990 on 39738 degrees of freedom, the model fits well but could potentially benefit from further refinement to address possible overdispersion, indicating alternative models like negative binomial regression could also be explored. However, we will go into much more complicated models.

In figure 8, Fitted vs Residuals, plot, the pattern of residuals displays distinct, widening bands as fitted values increase, suggesting non-constant variance and overdispersion, where the mean-variance equality assumption is violated. This pattern indicates the model may be poor fit for the data, possibly underestimating the variance at higher levels of the fitted values.

Table 2: Summary of Poisson Model Coefficients

Variable	Estimate	Std. Error	p-value
Intercept	0.1979	0.0131	< 2e-16 ***
unemployment	0.0241	0.0047	2.52e-07 ***
young males	0.0023	0.0001	< 2e-16 ***
wealth	0.1319	0.0047	< 2e-16 ***
year	-0.1651	0.0027	< 2e-16 ***
month_num	0.0328	0.0013	< 2e-16 ***

The figure 7, Fitted vs Actual, plot shows the relationship between fitted and actual values, where the data points spread further apart as the fitted values increase, reinforcing the evidence of overdispersion seen in the residuals plot. The blue line, indicating a trend, deviates from the 45-degree line that would suggest a perfect fit, highlighting the model's limited ability to predict higher counts accurately.

These plots suggest that the Poisson model struggles with the data's inherent variance, particularly for higher count values, hinting that a more robust model would be needed to better capture the trends and patterns in the data.

8.2 Mixed Effects Model

The Mixed Effects Model integrates both fixed and random effects to evaluate the influence of unemployment, young males, wealth, and monthly variations on crime rates. The model's coefficients indicate that wealth has a significant positive effect on crime rates with an estimate of 0.1385, and monthly variations are also highly significant with an estimate of 0.0328. The intercept, indicating the baseline crime rate when all predictors are zero, is significantly negative at -0.3444, suggesting lower crime rates at the baseline. Unemployment shows a non-significant impact, suggesting that it may not be a strong predictor in the presence of other variables. Random effects for ID and year show variances of 0.2486 and 0.0855, respectively, indicating variability in crime rates across different IDs and over the years not explained by the model's fixed effects alone. The model's fit is denoted by an AIC of 112083.0 and a log-likelihood of -56034.5, with a deviance of 112069.0 on 39744 observations, reflecting a substantial model fit but potentially indicating the need for further model refinement, such as a mixed effects model with INLA fitting.

Table 3: Summary of Mixed Effects Poisson Model Coefficients

Variable	Estimate	Std. Error	z-value	p-value
Intercept	-0.3444	0.1251	-2.754	0.0059**
unemp	0.0099	0.0221	0.448	0.6540
ym	0.0025	0.0006	4.613	3.97e-06***
wealth	0.1385	0.0241	5.741	9.40e-09***
month_num	0.0328	0.0013	24.723	¡2e-16***

Table 4: Random Effects and Model Statistics

Parameter	Value	Description
Random Effect for ID	0.2486 (Variance), 0.4986 (SD)	Intercept
Random Effect for Year	0.0855 (Variance), 0.2924 (SD)	Intercept
AIC	112083.0	-
Log Likelihood	-56034.5	-
Deviance	112069.0	-
Number of obs	39744	-

The figure 10, Fitted vs Residuals, plot displays a similar pattern to that observed in the basic Poisson model, with residuals fanning out as fitted values increase, suggesting heteroscedasticity and possible overdispersion. This pattern points to potential inadequacies in the mixed effects model's ability to consistently handle variance across the range of predictions.

In the figure 9, Fitted vs Actual, plot, the spread of actual values relative to fitted values widens with increasing fitted values, further indicating that the model may not accurately capture higher crime occurrences compare the the more common lower occurrences. The positive slope of the trend line in this plot shows a general underestimation of higher counts and an overestimation of lower counts, suggesting that while the model captures the general trend, its precision gets worse at higher levels of crime rates. These plots highlight the limitations of the mixed effects model in dealing with data exhibiting greater variability, more at higher fitted values.

8.3 Chiraq Model

The Chiraq Model, using a INLA framework, evaluates the impact of unemployment, wealth, and temporal changes on crime rates, incorporating both fixed and random effects. Fixed effects show the intercept at 0.349 (with 95% confidence of 0.075 to 0.622), suggesting a baseline crime rate. Unemployment's effect on crime has a mean estimate of 0.022 and is barely significant, as the confidence interval just zero. Wealth has a clearer positive influence on crime rates with an estimate of 0.186 and narrower confidence limits from 0.142 to 0.229, indicating a solid effect. The variable year shows a negative trend in crime rates over time with an estimate of -0.142. The random effects parameters indicate substantial variability captured by the year and month factors with respective mean precisions of 113.54 and 47.15, suggesting significant fluctuation in crime trends over these periods. The IID model for IDs reflects less variability among different districts or IDs with a precision of 3.86. The model's DIC of 110227.27 and marginal log-likelihood of -55760.47 suggest a good fit but also point to the complexity of the model, given the effective number of parameters at 539. This model structure allows for the exploration of time-specific effects and individual variability in crime data, providing a better understanding of the dynamics within crime data in Chicago.

Table 5: Summary of Fixed Effects for the Chiraq Model

Variable	Mean	SD	2.5% Quantile	Median	97.5% Quantile
Intercept	0.349	0.135	0.075	0.350	0.622
unemp	0.022	0.022	-0.022	0.022	0.066
wealth	0.186	0.022	0.142	0.186	0.229
year	-0.142	0.053	-0.250	-0.143	-0.034

Table 6: Random Effects and Model Hyperparameters

Effect	Model	Mean Precision	SD Precision
year_factor	RW1	113.54	84.623
month_factor	RW1	47.15	18.980
ID	IID	3.86	0.257

Table 7: Overall Model Statistics

Statistic	Value
DIC (Deviance Information Criterion)	110227.27
DIC (Saturated)	51698.67
Effective Number of Parameters	539.36
Marginal Log-Likelihood	-55760.47

The residual and fitted plots for the Chiraq model exhibit characteristics similar to earlier findings, emphasizing issues in model performance at higher crime rates. In the figure 11, Chiraq Model: Fitted vs Actual Values, plot, most data points fall along or near the red line, which shows a generally good fit for lower crime rates. However, as actual crime rates increase, the model tends to underestimate these rates. This pattern suggests that while the model performs well at lower levels of crime, it may not be good for accurately predicting higher crime occurrences, potentially calling for a reassessment of model assumptions or the incorporation of additional predictors to capture the higher-end variance.

The figure 12, Fitted vs Residual Values: Chiraq Model, plot again demonstrates a fanning pattern of residuals, which indicates heteroscedasticity—the variance of residuals increases with the fitted values. We also see a strange pattern of the values bouncing off of 0 which can be unexplained. These patterns, coupled with the presence of several outliers, points to potential model misspecification or the need for a model that can accommodate variable dispersion, such as a negative binomial model.

9 Discussion

9.1 The Models

The analysis of crime rates in Chicago using the Poisson, Mixed Effects, and "Chiraq" models provided insights into the relationships between socioeconomic factors, temporal trends, and burglary occurrences. The Poisson model identified unemployment, young male population, wealth, and monthly variations as significant predictors of crime rates. However, the model's residual and fitted plots indicated overdispersion and heteroscedasticity, suggesting the need for more advanced modeling approaches.

The Mixed Effects model, which incorporated both fixed and random effects, showed that wealth and monthly variations significantly influence crime rates, while unemployment had a non-significant impact. The model's random effects captured variability across IDs and years. Despite the improved fit compared to the Poisson model, the Mixed Effects model still had limitations in handling higher crime occurrences, as evident in the residual and fitted plots.

The Chiraq model, built with the INLA framework, provided the most comprehensive approach by incorporating fixed effects, random effects for year and month, and an IID model for IDs. The model identified wealth as a strong positive predictor of crime rates and revealed a negative trend in crime rates over time. The random effects captured significant temporal fluctuations, highlighting the importance of considering time-specific effects. While the Chiraq model demonstrated a better overall fit, it still struggled to accurately predict higher crime rates, as shown in the fitted vs. actual plot.

Based on the model evaluations, the Chiraq model come out on top as the most suitable choice for modeling burglary rates in Chicago. Its ability to incorporate both fixed and random effects, along with its consideration of temporal trends and individual variability, makes it a more comprehensive and informative approach compared to the Poisson and Mixed Effects models. However, it is important to note the limitations of the Chiraq model in its tendency to underestimate higher crime rates which might be poor when using the model to identify high crime rate areas.

9.2 Future Work

Future work could focus on refining the Chiraq model by exploring additional predictors or interaction terms that may better capture the dynamics of high crime occurrences. Investigating other model structures, such as negative binomial or ZIP models, could help address overdispersion and improve performance at higher crime rates. Incorporating policy changes and school patterns could provide insights into factors influencing crime rates. Examining the impact of policies implemented during the study period could also help explain the decreasing pattern year over year and shed light on the effectiveness of crime prevention strategies.

9.3 Ethical Considerations

The ethical implications of using crime data for modeling purposes lies in the interpretation and communication of model results. Here the communication should be done responsibly to avoid stigmatization of specific communities or perpetuation of stereotypes. The ultimate goal of such modeling efforts should be to inform evidence-based policies and interventions that promote public safety and address the root causes of crime, rather than to reinforce existing biases.

The analysis of burglary rates in Chicago using the above statistical models has provided insights into the complex relationships between socioeconomic factors, temporal trends, and burglary occurrences. The Chiraq model, with its approach of incorporating fixed and random effects, emerges as the most suitable choice for modeling crime rates. However, the model's limitations in predicting higher crime rates underscore the need for further refinement. As researchers and policymakers continue to tackle the challenges of understanding and addressing crime, it is important to prioritize ethical considerations and strive for models that can inform effective and community-focused solutions.

Appendix

CDT Chrisman's GitHub Code
CHAT GPT prompts and responses

A. Data Analysis Plots

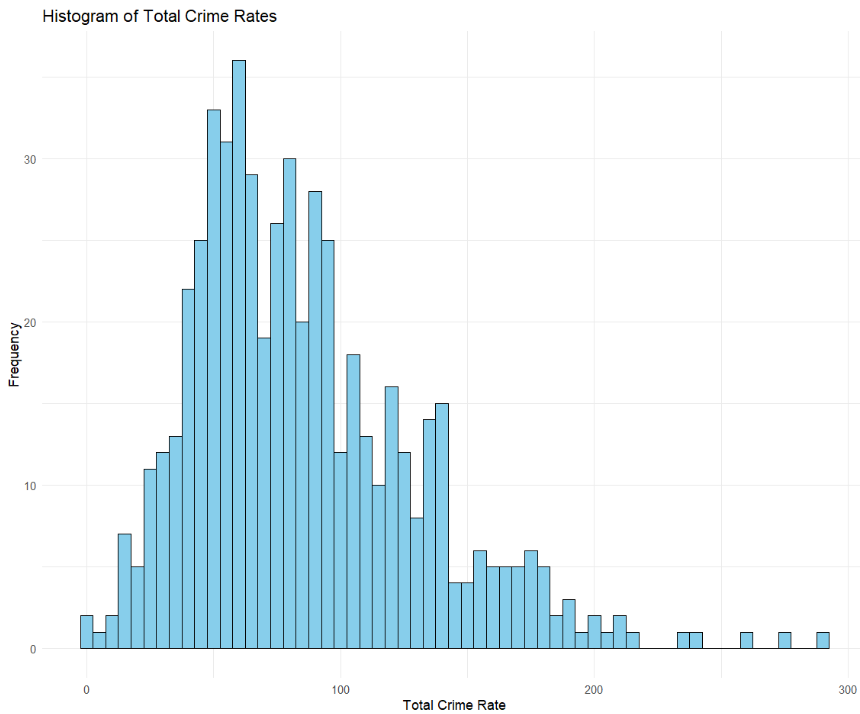

Figure 2: Histogram of Total Crime Rates

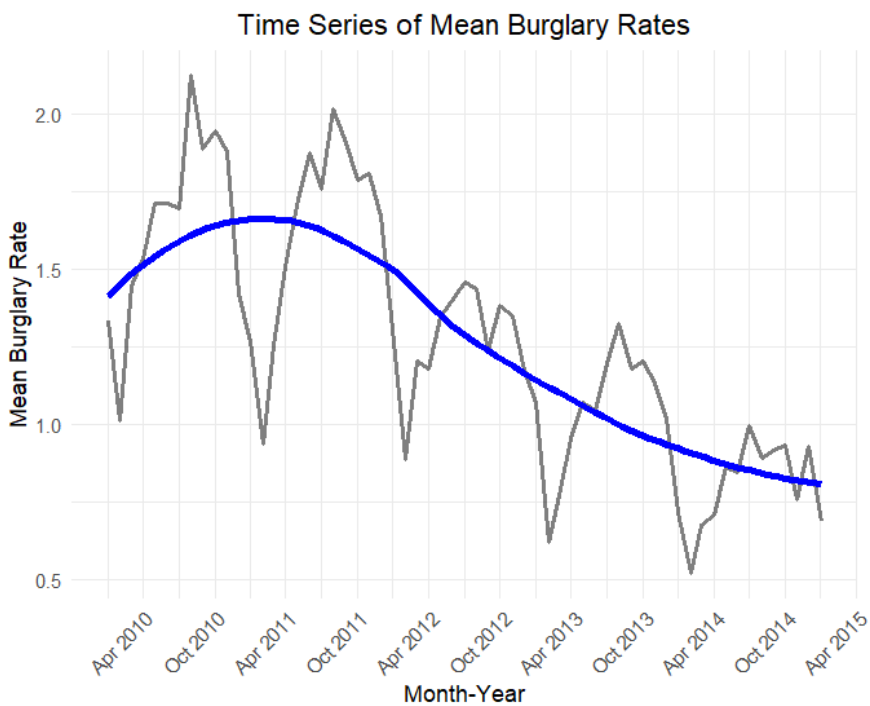

Figure 3: Time Series of Mean Burglary Rates

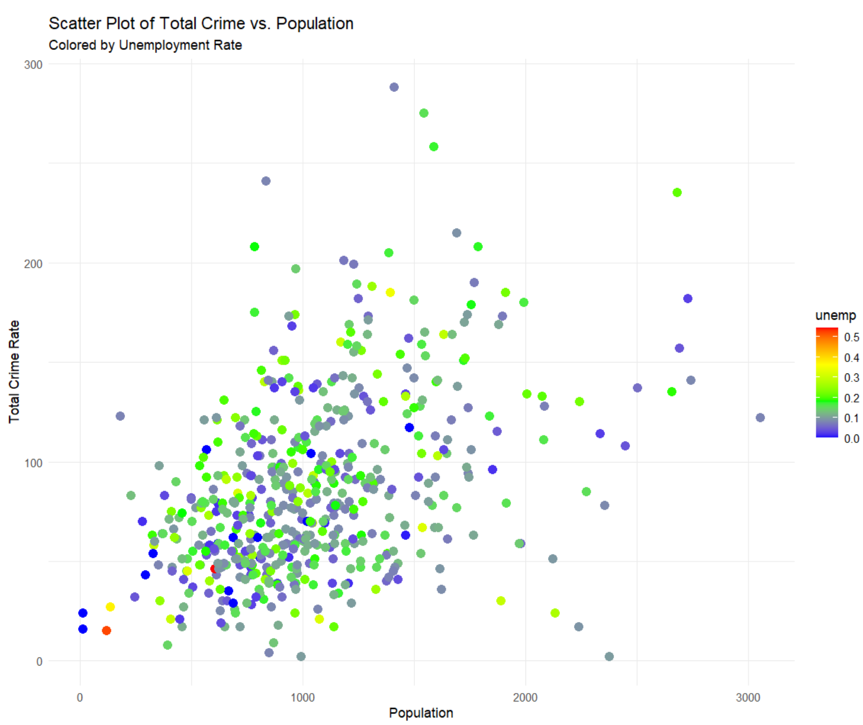

Figure 4: Scatter Plot of Total Crime vs. Population Colored by Unemployment Rate

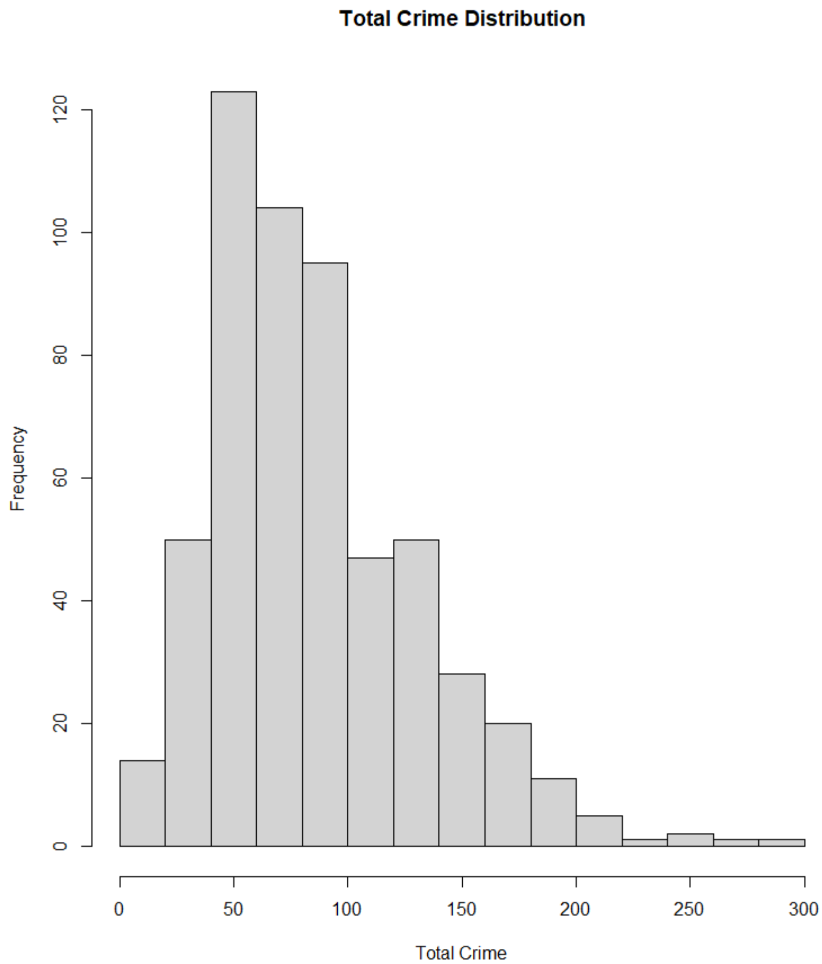

Figure 5: Scatter Plot of Total Crime vs. Count of Young Males

Figure 6: Scatter Plot of Total Crime vs. Wealth Index

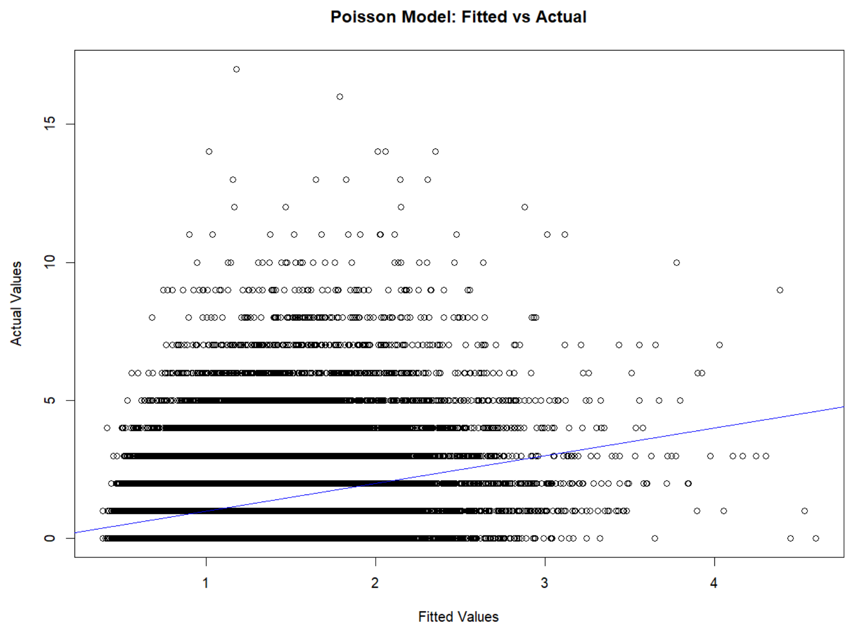

Figure 7: Poisson Model: Fitted vs Actual

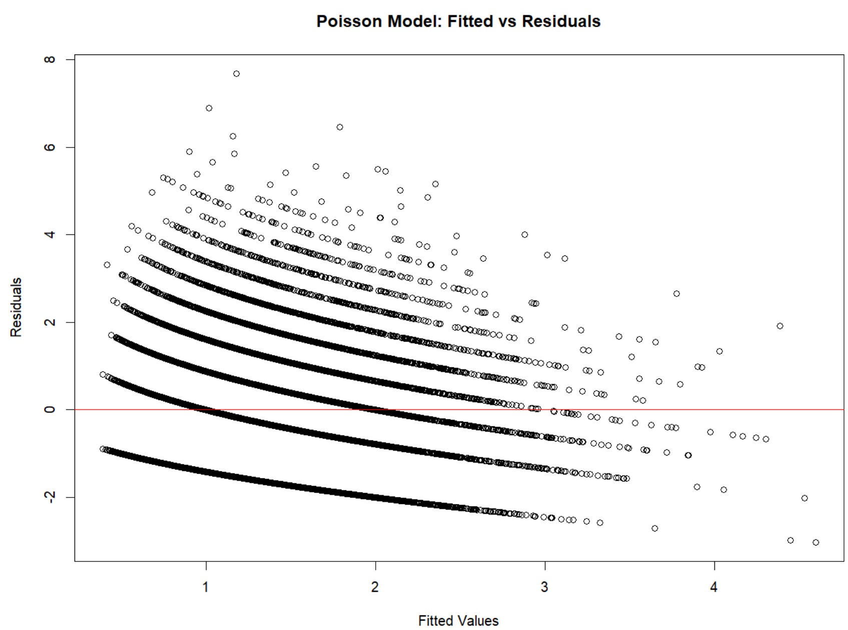

Figure 8: Poisson Model: Fitted vs Residuals

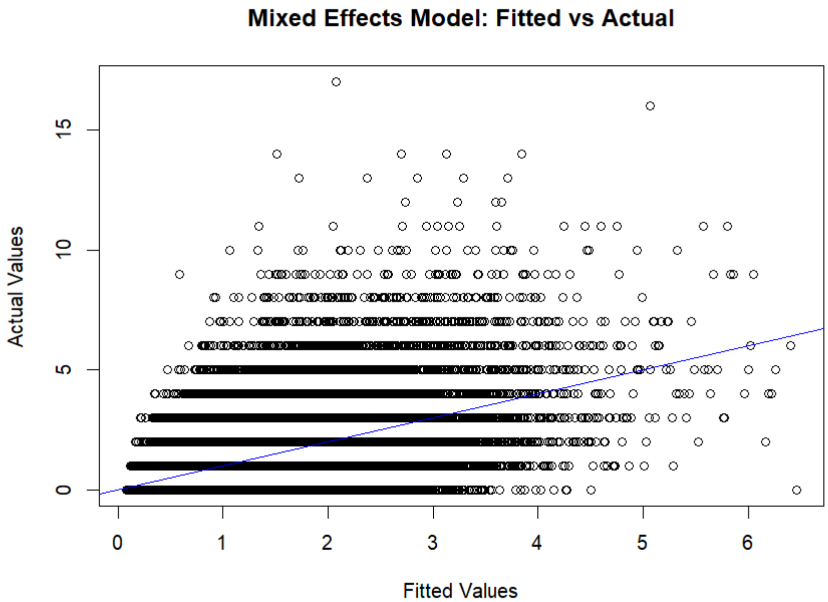

Figure 9: Mixed effects Model: Fitted vs Actual

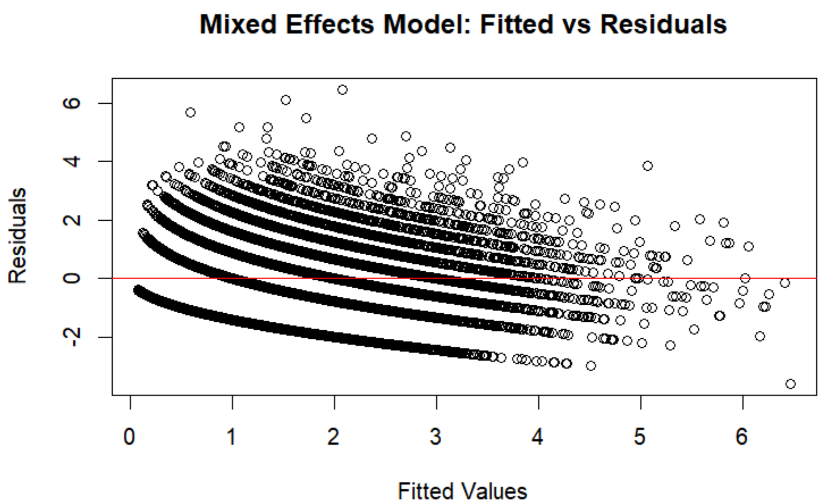

Figure 10: Mixed effects Model: Fitted vs Residuals

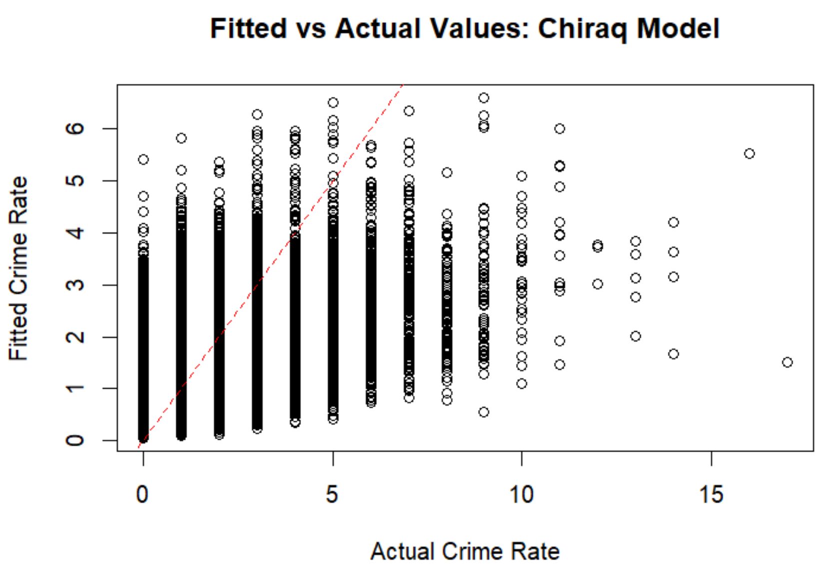

Figure 11: Chiraq Model: Fitted vs Actual Values

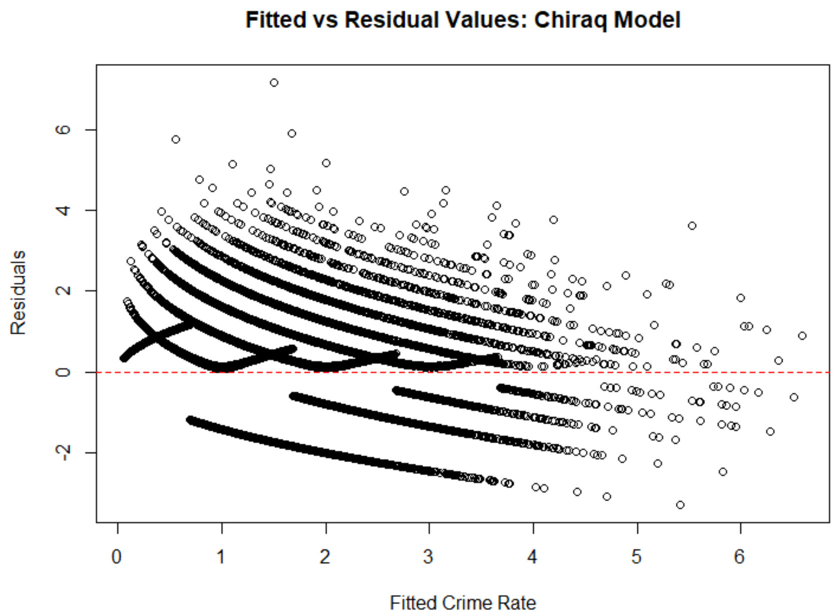

Figure 12: Chiraq Model: Fitted vs Residual Values

References

[1] Hui Luan, Matthew Quick, and Jane Law. Analyzing local spatio-temporal patterns of police calls-for-service using bayesian integrated nested laplace approximation. *ISPRS International Journal of Geo-Information*, 5(9):162, 2016. https://www.mdpi.com/2220-9964/5/9/162.

[2] George Chege, Samuel Mwalili, Anthony Wanjoya, and Anthony Bayesian. Spatial-temporal modelling and mapping for crime data in nairobi county. *International Journal of Data Science and Analysis*, 5, 2019. https://doi.org/10.11648/j.ijdsa.20190506.11.

1. Monroe, Sydney. "Chi-Raq and the Myth of Chicago Gang Wars." *The New York Times Magazine*, 7 Dec. 2015, https://www.nytimes.com/2015/12/07/magazine/chi-raq-and-the-myth-of-chica html.

2. Sander, Nikola, Susan Bijak, and John Bryant. "Demography, foreclosure, and crime: Assessing spatial heterogeneity in contemporary models of neighborhood crime rates " *Demographic Research*, vol. 26, no. 18, 2012, pp. 451-486, https://www.demographic-research.org/volumes/vol26/18/26-18.pdf.

3. Guido Vivaldi. "Using Mixed Effects Models for Linear Regression.", https://towardsdatascience.com/using-mixed-effects-models-for-linear-regression-7b7941d249b.

4. Virgillio Gomez-Rubio. "Spatial Models in INLA." In *Bayesian Inference with INLA*, https://becarioprecario.bitbucket.io/inla-gitbook/ch-spatial.html.

5. Kounadi, Ourania, Alina Ristea, Adelson Araujo, and Michael Leitner. "A systematic review on spatial crime forecasting." *Crime Science*, vol. 9, no. 1, 2020, https://doi.org/10.1186/s40163-020-00116-7.

6. OpenAI. (2024). ChatGPT Retrieved MAY 05, 2024, from OpenAI. Assisted author with generative AI for LaTeX equations and tables. The AI was requested to format the entered model and coefficients into LaTeX formatted equations. It was also requested to provide a LaTeX table for the summary model output for coefficients and associated P values. The prompt for latex equations was "take the below summary output from the [X] model and create the LaTex code to present the equations in a report. Use B0, B1...Bn in place of the coefficient values for the equation. Summary(model)..." The LaTeX tables were inputted into a larger report. The AI was also used for grammar and spell checking following writing process to maintain a clear and technical tone. West Point, NY.

7. OpenAI. (2024). ChatGPT. Assisted author with generative AI for grammar and spell checking following the writing process to maintain a clear and technical tone. The prompt was "fix grammar, sentence structure, and spelling: (written paragraph) Note what changes you made at the bottom." After re-reading and double-checking that no information was changed regarding what I was saying to answer the prompt, the paragraph was then used in the report with some changes not taken from the output. This was part of the larger writing process when addressing the prompt's sections and editing for clarity and syntax, ensuring the prompt was addressed and answered for each section. West Point, NY. Accessed 05 MAY. 2024.