The Faces of Domestic Violence: An Intersectional Analysis of Felony Assaults and Rapes in NYC*

Erping Gong

April 19, 2024

This paper presents a statistical analysis of offense classification within the criminal justice system using a multinomial logistic regression model. The research examines the influence of victim and suspect demographics—specifically race, sex, age, and socioeconomic factors—on the categorization of incidents as either felony assault, rape, or a reference category, designated 'DIR'. Our methodological approach draws on a dataset that captures a broad spectrum of reported offenses, allowing for an assessment of how these individual characteristics may bias offense classification. The study aims to illuminate patterns that could suggest systemic inequities, providing a basis for future policy discussions.

Table of contents

1 Introduction **2**
1.1 Estimand 3

2 Data **3**
2.1 Data Source 3
2.2 The Cleaned Data 4
2.3 Statistics Summary 5

3 Methology **8**

4 Model **9**
4.1 Model Justification 9

*Code and reporduction data are available at: https://github.com/ErpingGong/Final_paper

4.2 Result . . . 10
4.3 Disparities in Offense Classification by Victim Race and Sex . . . 10
4.4 Disparities in Offense Classification by Suspect Race and Sex . . . 11
4.5 Disparities in Offense Classification by Victim Age and Poverty . . . 11

5 Discussion **12**
5.1 Insight into Systemic Bias . . . 12
5.2 Limitation and weekness . . . 12
5.3 Future Steps . . . 13

References **14**

1 Introduction

Domestic violence remains a critical social issue that disproportionately affects millions of individuals across the United States, manifesting through physical, emotional, and sexual abuse within intimate partnerships. It is estimated that nearly 30% of U.S. couples (married and unmarried) will experience intimate partner violence at some point in their relationship, and between 3% and 10% of intimate relationships involve serious violence (Straus (1990)). This introduction seeks to explore the intersectional dynamics of felony assaults and rapes within New York City, examining how these crimes intersect with various socio-demographic factors including gender, race, and economic status.

Recent studies highlight the complexity of domestic violence in urban settings. Akey (Akey (2017)) emphasizes the role of socioeconomic disparities in influencing the prevalence and forms of domestic violence, noting that lower economic strata are associated with a higher risk of experiencing severe forms of physical abuse. Moreover, urban environments like New York City present unique challenges and stressors that can exacerbate domestic tensions and lead to violence (Smith et al. (2019)).

Gender plays a pivotal role in the dynamics of domestic violence. Women are disproportionately victims of intimate partner violence, with severe impacts on their physical and mental health (Johnson and Leone (2005)). This gendered pattern is particularly pronounced in cases of sexual violence, where women are overwhelmingly the victims (Clark and Burt (2010)).

Race and ethnicity also influence the experience and reporting of domestic violence. Minoritized populations often face systemic barriers in accessing support services, which can result in underreporting and a lack of adequate interventions (Martinez et al. (2020)). Furthermore, cultural norms and stigma can affect the willingness of victims to seek help (Lee (2018)).

In the context of New York City, the intersection of race, poverty, and gender illuminates the complexities of addressing domestic violence in a multicultural urban landscape (White and Gorman (2016)). The legal and social service frameworks in place often struggle to meet the diverse needs of victims, necessitating a more nuanced approach to policy and intervention (Greenberg and Schneider (2018)).

This paper will utilize an intersectional framework to analyze felony assaults and rapes in New York City, drawing upon a range of empirical studies and theoretical perspectives to understand the multifaceted nature of domestic violence in one of the world's most populous urban centers. Statistical programming language R (R Core Team (2023)) is used in this report, with packages `tidyverse` (Wickham et al. (2019)), `ggplot2` (Wickham (2016)), `testthat` (Wickham (2011)),`knitr` (Xie (2014)), `arrow` (Richardson et al. (2024)) `nnet` (Venables and Ripley (2002a)), `MASS`(Venables and Ripley (2002b)), `lmtest`(Zeileis and Hothorn (2002)), `sandwich`(Zeileis (2006)), `kableExtra` (Zhu (2024)), `reshape2`(Wickham (2007)), `caret`(Kuhn and Max (2008)), `broom`(Robinson, Hayes, and Couch (2023)) and `tinytex`(Xie (2019)).

1.1 Estimand

The aim is to estimate the difference in the proportion of domestic violence offense types (felony assault and felony rape) among different victim and suspect racial, ethnic, and gender groups in New York City for the years 2020 and 2021. The analysis will account for incidents with incomplete data through listwise deletion, focusing on cases with fully reported victim and suspect demographics and incident details. The outcome will be summarized using proportions within each demographic group and analyzed to identify statistically significant differences in offense types across these groups.

2 Data

2.1 Data Source

The dataset utilized in this study, titled "ENDGBV: The Intersection of Domestic Violence, Race/Ethnicity and Sex," is maintained by NYC Open Data and includes comprehensive incident-level data provided by the New York City Police Department (NYPD). This extensive dataset encompasses 483,791 entries, detailing domestic violence-related offenses, specifically felony assaults, felony rapes, and domestic incident reports, recorded during the calendar years 2020 and 2021. Key attributes of the dataset include: - **Offense Type):** Categories of domestic violence offenses, such as felony assaults and felony rapes. - **Date of Incident):** The specific dates on which the incidents occurred. - **Precinct of Incident):** Precinct of Incident: The NYPD precincts where the incidents were reported. - **Borough of Incident):** The boroughs within New York City where the incidents took place. - **Relationship and Demographic Details):** Information about the relationship between the suspect and the victim, along with detailed descriptions of both parties, including race, sex, and reported age. - **Community District Information):** Data related to the community district of the incident,

	Data Summary							
Offense Type	Report Date	Intimate Relationship Flag	Victim Race	Victim Sex	Victim Reported Age	Suspect Race	Suspect Sex	Poverty
FELONY ASSAULT	03//2020	NO	WHITE HISPANIC	MALE	3	BLACK	MALE	1
FELONY ASSAULT	04//2020	NO	WHITE HISPANIC	MALE	5	WHITE	MALE	0
FELONY ASSAULT	10//2020	YES	WHITE HISPANIC	MALE	5	BLACK	MALE	0
FELONY ASSAULT	10//2020	YES	WHITE HISPANIC	MALE	5	WHITE HISPANIC	MALE	0
FELONY ASSAULT	10//2021	NO	WHITE HISPANIC	MALE	8	WHITE HISPANIC	MALE	1

including socio-economic indicators such as poverty rate, median household income, and unemployment rates. Each domestic incident report (DIR) within this dataset is a standardized form filled out by police officers responding to calls of domestic violence, regardless of whether an arrest was made. This dataset is instrumental for analyzing the patterns and impacts of domestic violence across different racial, ethnic, and gender groups within the urban context of New York City.

2.2 The Cleaned Data

A Preview of the First 5 Rows of Cleaned Data

	Data Summary							
Offense Type	Report Date	Intimate Relationship Flag	Victim Race	Victim Sex	Victim Reported Age	Suspect Race	Suspect Sex	Poverty
FELONY ASSAULT	03//2020	NO	WHITE HISPANIC	MALE	3	BLACK	MALE	1
FELONY ASSAULT	04//2020	NO	WHITE HISPANIC	MALE	5	WHITE	MALE	0
FELONY ASSAULT	10//2020	YES	WHITE HISPANIC	MALE	5	BLACK	MALE	0
FELONY ASSAULT	10//2020	YES	WHITE HISPANIC	MALE	5	WHITE HISPANIC	MALE	0
FELONY ASSAULT	10//2021	NO	WHITE HISPANIC	MALE	8	WHITE HISPANIC	MALE	1

Figure 1: A Preview of the First 5 Rows of Cleaned Data

A Preview of the First 5 Rows of Cleaned Data

The cleaning of the NYPD domestic violence dataset was carried out with a focus on retaining the most relevant variables for the study. The `Report Date` was simplified to "mm/yyyy" format for a monthly analysis granularity, as exemplified by the March 2020 date in the first entry of the cleaned data. Columns not directly related to the offense details, such as geographic codes and socio-economic indicators, were removed to narrow the focus to incident specifics.

The dataset was further refined by filtering out entries with missing values in the `Intimate Relationship Flag`, ensuring the clarity of the relationship context. For the `Poverty` variable, missing values were replaced with 0 to indicate a default low poverty rate, thus maintaining uniformity. The final dataset consists solely of complete cases across all considered variables, as seen in the sample of the first 5 entries.

The dataset features indicate the nature and circumstances of domestic violence incidents. For example, the first record in Table denotes a `FELONY ASSAULT` offense with the `Intimate Relationship Flag` indicating no intimate relationship between the victim and the suspect. The `Victim Race` and `Victim Sex` denote a Hispanic male, and the victim's reported age is 3, while the `Suspect Race` is Black, and the `Suspect Sex` is male. Notably, this row also indicates a high poverty rate with a `Poverty` value of 1. After cleaning, 327117 raws of data with 9 data features remain.

2.3 Statistics Summary

Offense_Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
DIR	311191	95.13	311191	95.13
FELONY A	14933	4.57	326124	99.70
RAPE	993	0.30	327117	100.00

Figure 2: distribuction chart

Table provides a comprehensive view of the distribution of cases across various dimensions of the dataset. The `Offense Type` variable shows that 'DIR' offenses are the most common, with 311,191 cases, followed by 'FELONY ASSAULT' with 14,933 cases, and 'RAPE' with 993 cases. The `Report Date` indicates a higher frequency of reports in October 2021, reflecting 16,138 cases.

The `Intimate Relationship Flag` variable suggests that a significant portion of the incidents, 199,630 to be precise, involved an intimate relationship, whereas 127,012 did not. The racial distribution of victims (`Victim Race`) shows that 'BLACK' victims are most frequently reported in the dataset, amounting to 163,167 cases, with 'WHITE' victims accounting for 108,305 cases.

The dataset also reveals a gender skew in domestic violence reports, with 'FEMALE' victims being overwhelmingly more common at 236,315 cases, as opposed to 'MALE' victims at 90,802 cases. All `Victim Reported Age` entries are grouped under the '0-9' category, which seems like an anomaly and may need further investigation.

Suspects' race (`Suspect Race`) and sex (`Suspect Sex`) follow similar distribution patterns to the victims', with 'BLACK' and 'MALE' being the most frequent entries at 169,278 and 225,009 cases, respectively. The `Poverty` status shows a higher frequency of incidents occurring in non-poverty-stricken areas, with 199,410 cases reported.

The data underscores the prevalence of domestic violence in certain demographics and periods, suggesting the need for targeted interventions. The summary captures the nuances of the dataset and provides a clear overview for further analysis.

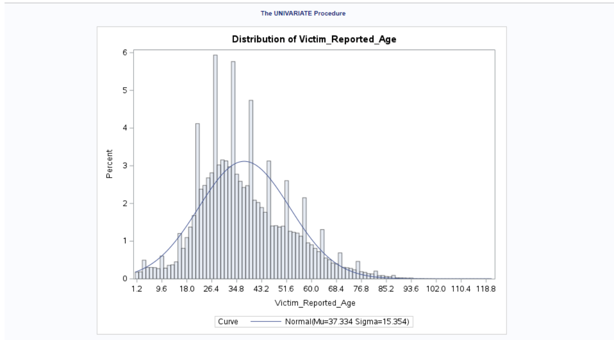

Figure 3: Age analysis

#age analysis From the histogram, we can observe the following: The distribution of victim ages is roughly bell-shaped, as indicated by the histogram bars and the overlay of the normal distribution curve. There is a concentration of victims in the middle age range, with fewer victims as the age increases or decreases from the mean. The peak of the distribution appears to be slightly below the mean age, indicating that the largest percentage of victims fall into the younger adult age range. The distribution has a tail that extends into the older ages, suggesting that there are victims across a broad age range but with diminishing frequency as age increases. However, there are some important caveats: The histogram bars do not perfectly align with the normal curve, indicating that the distribution of victim ages is not perfectly normal. This is typical in real-world data. The x-axis shows age extending beyond 100, which might include outliers or errors in data entry since such ages are relatively rare. The normal distribution is symmetric, but the histogram shows that the actual distribution of ages has a slight right-skew (longer tail on the right).

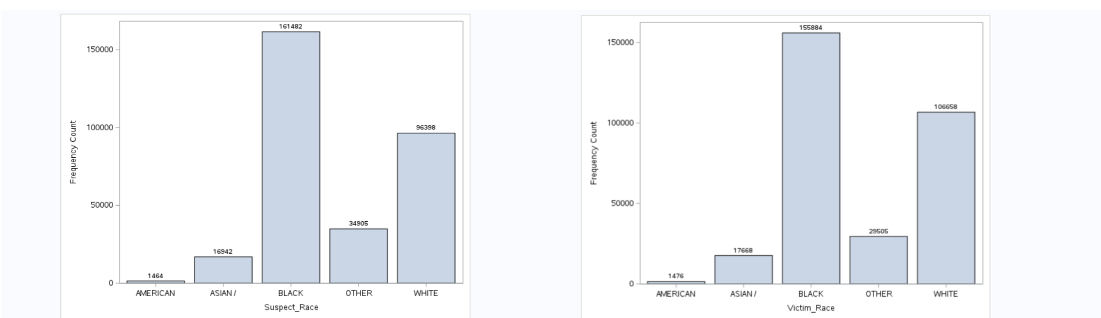

Figure 4: DIR Race barplot

#DIR race Suspect_Race (Left Chart): The categories presented are American, Asian/, Black, Other, and White. The highest frequency is observed in the Black category with 161,482 occurrences, followed by White with 96,398, and Other with 34,905. The Asian/ category has a count of 16,942, and the American category has the lowest with 1,464 occurrences. This distribution suggests a disproportionately high number of individuals identified as Black being listed as suspects in comparison to other races. Victim_Race (Right Chart): The same racial categories are present as in the suspects chart. Again, the highest frequency count is in the Black category with 155,884 occurrences. The White category follows with 106,658, and Other with 29,505. The Asian/ category has 17,768, and American has the fewest at 1,476. Similar to the suspects chart, individuals identified as Black are also represented as victims more frequently than any other racial group listed.

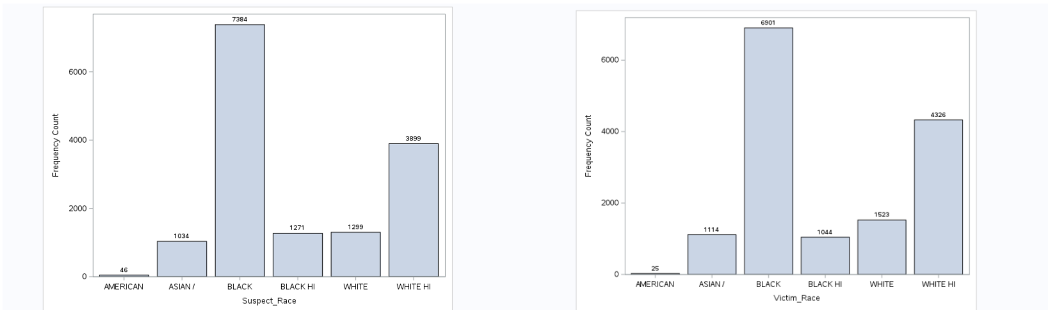

Figure 5: FA Race barplot

#FA race Suspect_Race (Left Chart): The chart displays frequencies for the following racial categories: American, Asian/, Black, Black HI (presumably Black Hispanic), White, and White HI (presumably White Hispanic). The largest frequency count is for the Black category with 7,384 occurrences. The White HI category has 3,899 occurrences, followed by White with 1,299, and Asian/ with 1,034. The Black HI category has 1,271 occurrences, and the American category has the fewest at 46. The chart indicates that Black individuals are represented as suspects in this event type more than any other racial group shown. Victim_Race (Right Chart): This chart displays frequencies for the same racial categories as victims. The largest frequency count is for the Black category with 6,901 occurrences. The White HI category has 4,326 occurrences, followed by White with 1,523, and Asian/ with 1,114. The Black HI category has 1,044 occurrences, and the American category has the fewest at 25. Similar to the suspects chart, Black individuals are represented as victims more than any other racial group listed here.

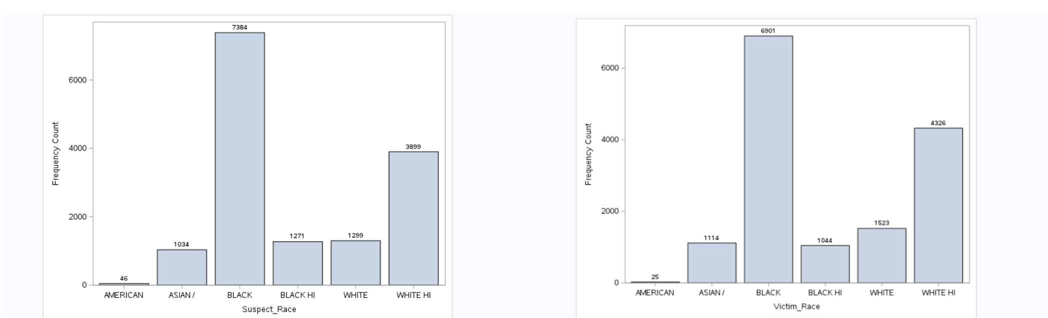

Figure 6: rape Race barplot

#rape race Suspect_Race (Left Chart): This chart shows the counts for different racial groups labeled as suspects. The categories include American, Asian/Black, Black HI (presumably Black Hispanic), White, and White HI (presumably White Hispanic). The highest frequency count is for the Asian/Black category with 412 occurrences, followed by White HI at 344. Black HI has 65 occurrences, White has 94, and American has the fewest at 4. It's important to note that the Asian/Black category is unusual and might be a labeling error or a combined category due to low counts in separate Asian and Black categories. Normally, these categories would be distinct.

Victim_Race (Right Chart): This chart shows the counts for different racial groups labeled as victims. The categories are the same as in the left chart. The highest frequency count is again in the Asian/Black category with 382 occurrences, followed closely by White HI at 338. White has 124, Black HI has 64, and American has the fewest at 1.

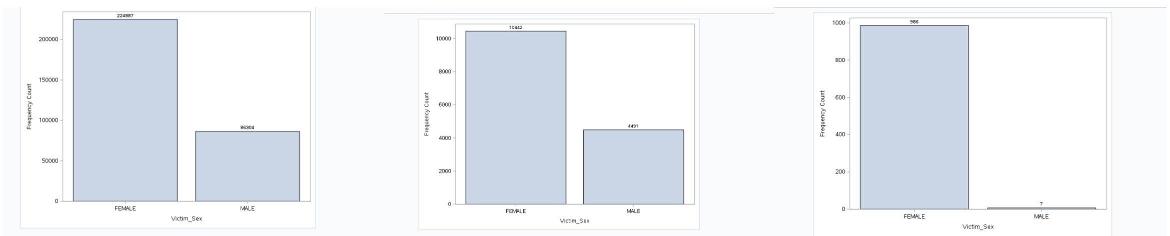

Figure 7: Sex barplot

#sex and assault It is displayed to show the follows: DIR-SEX Chart (Leftmost): The leftmost bar chart shows a significant difference in frequency counts between females and males. Female victims account for 224,897 occurrences, while male victims account for 86,304. This suggests that the event in question occurs far more frequently with females as victims compared to males. FA-SEX Chart (Center): The middle bar chart still shows a higher frequency count for females compared to males but with smaller numbers overall. Females account for 10,442 occurrences and males for 4,491. The difference between the genders is still noteworthy, but the total number of occurrences for both is much lower than in the first chart. RAPE-SEX Chart (Rightmost): The rightmost chart has significantly lower numbers, with females at 986 occurrences and males at 7. This chart displays the largest proportional difference between female and male victims, with females having a drastically higher count than males.

3 Methology

The analysis benefits from additional functionalities harnessed from various R packages such as `tidyverse` installed to gain access to other important R packages, `here` created a path to specific saved files, `readr` read and imported data, `ggplot2` made the data visualizations, `knitr` and `dplyr` manipulated and cleaned data, and `modelsummary` to create summary tables.

Further insights into the deployment of these packages will be expounded upon in the ensuing subsections.

4 Model

In this section, we briefly discuss Multinomial models that are being used in this analysis.

setup The model is formulated as follows:

$$\log\left(\frac{P(\text{OffenseType}_i = \text{'FELONY ASSAULT'})}{P(\text{OffenseType}_i = \text{'DIR'})}\right) = \beta_0 + \beta_1 \times X_1 + \cdots + \beta_k \times X_i$$
$$\log\left(\frac{P(\text{OffenseType}_i = \text{'RAPE'})}{P(\text{OffenseType}_i = \text{'DIR'})}\right) = \beta_0 + \beta_1 \times X_1 + \cdots + \beta_k \times X_i$$

beta0s are the intercepts for the logistic regression equations corresponding to the probabilities of the 'FELONY ASSAULT' and 'RAPE' offenses, respectively. The intercept represents the log odds of the outcome when all predictor variables are at their reference levels.

- The coefficients beta in the 'FELONY ASSAULT' equation and in the 'RAPE' equation represent the change in the log odds of the respective outcome for a one-unit change in the predictor variables, holding all other variables constant.
- Xi are the predictor variables for observation i. Each variable X corresponds to a specific attribute or characteristic that is believed to influence the probability of the offense being categorized as 'FELONY ASSAULT' or 'RAPE' as opposed to 'DIR'. The multinomial logistic regression model is chosen because it is well-suited for predicting categorical outcomes with more than two categories and can incorporate both numerical and categorical data, offering interpretable results. We use the mutinom() function from `nnet` package in R to run the model, and we perform stepwise selection for the final model based on AIC value(Akaike Information Criterion). To avoid excessive runtime, 5000 "Offense Type = DIR" data entries, 1500 "Offense Type = Felony Assualt" and 100 "Offese Type = Rape" are proportionly sampled to fit the model with random seed 999.

4.1 Model Justification

The length of the confidence intervals indicates the precision of the estimates; shorter intervals mean more precise estimates. The distance from zero indicates the strength of the association; the further away, the stronger the association. The horizontal lines crossing the points represent the confidence intervals for these estimates, usually at a 95% confidence level. It suggests that those predictors have a statistically significant association with the outcome category.

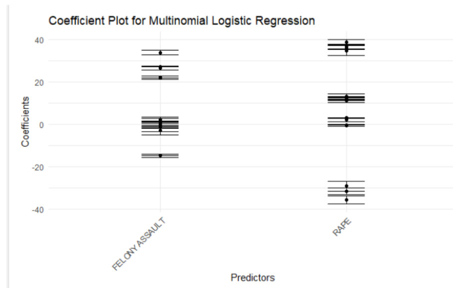

Figure 8: error plot for coefficients

4.2 Result

Based on the test result, we find that the overall accuracy of such model is 0.9714172, this is a astonishing accuracy. However, this accuracy can be misleading if we don't go into the results; Almost 100% of DIR is correctly predicted, that is because most of the sample comes from DIR type due to the originial distribution of the data. The accuracy of predicting Felcony-Assault is 0.4258373, this is a moderate value. For the accuracy of predicting rape, the acccuracy is 0 because in the original data set, the rape takes too little of the percentage of all types of assaults, so it is very likely in the random sample, we observe no rape type assault situation. The weakness of predicting rape type assault can be a potential weakness of this model. This weakness is cause by lack of data.

4.3 Disparities in Offense Classification by Victim Race and Sex

Estimated coefficients for Victim Sex and Victim Race for the outcomes of Felony Assault and Rape.

Outcome	Intercept	Asian...Pacific.Islander	Black	Black.Hispanic	Other	White	White.Hispanic	Female
FELONY ASSAULT	-2.813183	0.0731035	-0.3766749	22.19902	-14.64882	-1.110085	21.92778	-0.481088
RAPE	-31.458965	12.3362323	13.0496919	35.67085	-29.01914	12.310294	34.94341	3.109255

Figure 9: coefficient chart

This represents the relationship between a victim's race and sex and the classification of offenses. Specifically, it indicates that incidents involving Asian/Pacific Islander and Black Hispanic victims have a higher chance of being reported as rape. This suggests a particular pattern in the reporting or occurrence of such offenses against these racial groups. The negative coefficient for female victims in the context of felony assault, when considering DIR as the reference category, suggests that incidents involving female victims are less frequently reported as felony assault compared to DIR offenses. This is evident despite the overall lower occurrence of felony assault relative to DIR in the dataset. The implications of this finding could be multi-layered. It may indicate that female victims are less likely to be involved in situations classified

Table 1: Estimated coefficients for Victim Reported Age and Poverty for the outcomes of Felony Assault and Rape.

Outcome	Victim.Reported.Age	Poverty
FELONY ASSAULT	-0.0045945	-0.1067098
RAPE	-0.0695155	-0.4474157

as felony assault or that such incidents are underreported or often recorded under a different categorization.

4.4 Disparities in Offense Classification by Suspect Race and Sex

A Preview of the First 5 Rows of Cleaned Data

	Data Summary							
Offense Type	Report Date	Intimate Relationship Flag	Victim Race	Victim Sex	Victim Reported Age	Suspect Race	Suspect Sex	Poverty
FELONY ASSAULT	03//2020	NO	WHITE HISPANIC	MALE	3	BLACK	MALE	1
FELONY ASSAULT	04//2020	NO	WHITE HISPANIC	MALE	5	WHITE	MALE	0
FELONY ASSAULT	10//2020	YES	WHITE HISPANIC	MALE	5	BLACK	MALE	0
FELONY ASSAULT	10//2020	YES	WHITE HISPANIC	MALE	5	WHITE HISPANIC	MALE	0
FELONY ASSAULT	10//2021	NO	WHITE HISPANIC	MALE	8	WHITE HISPANIC	MALE	1

Figure 10: coefficient chart 2

Demographic variables such as suspect race and sex are linked to the occurrence of felony assault and rape charges. Notably, suspects identified as Black Hispanic are disproportionately associated with these charges, a fact that may point towards deeper systemic issues, such as socioeconomic challenges or biases within the criminal justice system. For sex, the data aligns with established trends, showing males more frequently charged with rape. A significant finding is the high association of charges with the 'Unknown' sex category, suggesting that there may be specific scenarios or complexities in these cases that obscure the identification of the suspect's sex at the time of reporting.

On the other end of the spectrum, suspects identified as belonging to the 'Other' racial category are less likely to be charged with either offense compared to the baseline group. This could reflect various factors, from actual lower rates of offending to different patterns in reporting or law enforcement prioritization.

4.5 Disparities in Offense Classification by Victim Age and Poverty

The negative coefficients for victim reported age in both felony assault and rape categories suggest a trend where younger victims are more likely to be associated with these offense types compared to 'DIR', or it might reflect a tendency for these incidents to be reported or classified differently as the age of the victim increases. As for poverty, the negative coefficients in both categories imply that offenses reported in areas with higher poverty rates are less likely to be classified as felony assault or rape, compared to 'DIR'. This might highlight socioeconomic factors at play in the reporting and classification of crimes, suggesting that poverty might be inversely related to the likelihood of an incident being recorded under these more severe offense categories. The model results could point to systemic issues such as underreporting in impoverished areas, possibly due to distrust in authorities or a lack of resources to pursue justice.

5 Discussion

In this paper, we've constructed a multinomial logistic regression model to examine how the characteristics of victims and suspects, such as race, sex, and age, influence the classification of reported offenses as 'FELONY ASSAULT', 'RAPE', or 'DIR'. By analyzing a dataset comprised of various reported incidents, we've quantified the strength of associations between these characteristics and offense classifications.

5.1 Insight into Systemic Bias

One key finding from this research is the potential evidence of systemic bias in the classification of offenses based on race and sex. The higher coefficients associated with certain racial groups for 'FELONY ASSAULT' and 'RAPE' suggest that suspects and victims of specific ethnicities are disproportionately represented in these offense categories. This raises questions about the equity of the criminal justice process and suggests that individuals from these groups may face biases that influence how incidents involving them are reported and recorded.

5.2 Limitation and weekness

The study's findings are subject to several limitations that warrant a cautious interpretation. A significant limitation is the small number of rape cases compared to other offenses in the dataset, which challenges the robustness and reliability of the statistical inferences made for this category. With such a small sample, the model's estimates for rape may be less stable, potentially misrepresenting the true associations. Furthermore, reporting biases are a critical concern. The data only includes reported incidents, which means it may not accurately represent the prevalence of offenses across different demographics. Factors such as cultural norms, trust in law enforcement, and perceived repercussions could influence whether and how offenses are reported. This limitation is particularly relevant for sensitive crimes such as rape, where underreporting is a well-documented issue.

5.3 Future Steps

To address the limitations identified in the current study and to enhance the validity of future research in crime classification patterns, several steps can be undertaken. Firstly, future research should aim to collect more comprehensive datasets, with a particular focus on increasing the sample size of less frequently reported offenses like rape. A larger dataset would provide more robust statistical power to detect true associations and would allow for more nuanced analysis of the factors at play. Efforts could include creating collaborations with multiple law enforcement agencies to gather a wider array of reports and ensure that a diversity of incidents is represented. Secondly, it is crucial to account for potential reporting biases that may affect the data. Future studies could incorporate qualitative research methods, such as interviews or focus groups with victims, law enforcement officers, and legal professionals, to understand the barriers to reporting and the decision-making processes behind crime classification. This qualitative insight can complement quantitative models and provide a deeper understanding of the data.

References

Akey, Thomas. 2017. "Socioeconomic Status and Domestic Violence: A Meta-Analysis." *Journal of Family Violence* 32 (6): 637–49.

Clark, Helen, and Michael Burt. 2010. "Understanding Sexual Violence in Intimate Relationships." *Domestic Violence Quarterly* 28 (1): 52–68.

Greenberg, Michael, and Daniel Schneider. 2018. "Challenges in Addressing Gender Violence in Urban Areas." *Urban Studies Journal* 55 (14): 3061–77.

Johnson, Michael P, and Janel M Leone. 2005. "The Differential Effects of Intimate Terrorism and Situational Couple Violence." *Journal of Family Issues* 26 (3): 322–49.

Kuhn, and Max. 2008. "Building Predictive Models in r Using the Caret Package." *Journal of Statistical Software* 28 (5): 1–26. https://doi.org/10.18637/jss.v028.i05.

Lee, David. 2018. "Cultural Factors Influencing Domestic Violence: The Role of Racial and Ethnic Identity." *Journal of Community Health* 43 (5): 833–41.

Martinez, Omar et al. 2020. "Evaluating the Impact of Immigration Policies on Health Status Among Undocumented Immigrants: A Systematic Review." *Journal of Immigrant and Minority Health* 22 (3): 548–59.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Robinson, David, Alex Hayes, and Simon Couch. 2023. *Broom: Convert Statistical Objects into Tidy Tibbles*. https://CRAN.R-project.org/package=broom.

Smith, David J, David E Duffee, Claire M Steinke, Han Huang, and Gregory L Larkin. 2019. "Urban Stressors and Mental Health: Understanding the Impact of Urban Living." *Urban Health Bulletin* 96 (4): 572–84.

Straus, Murray A. 1990. "Physical Assault by Wives: A Major Social Problem." In *Physical Violence in American Families*, edited by Richard J Gelles and Murray A Straus, 67–87. New Brunswick, NJ: Transaction.

Venables, W. N., and B. D. Ripley. 2002a. *Modern Applied Statistics with s*. Fourth. New York: Springer. https://www.stats.ox.ac.uk/pub/MASS4/.

———. 2002b. *Modern Applied Statistics with s*. Fourth. New York: Springer. https://www.stats.ox.ac.uk/pub/MASS4/.

White, Melody E, and Bridget K Gorman. 2016. "Racial and Ethnic Differences in the Desire to Seek Treatment for Intimate Partner Violence." *Rural Sociology* 81 (3): 359–83.

Wickham, Hadley. 2007. "Reshaping Data with the reshape Package." *Journal of Statistical Software* 21 (12): 1–20. http://www.jstatsoft.org/v21/i12/.

———. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

———. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

———. 2019. "TinyTeX: A Lightweight, Cross-Platform, and Easy-to-Maintain LaTeX Distribution Based on TeX Live." *TUGboat* 40 (1): 30–32. https://tug.org/TUGboat/Contents/contents40-1.html.

Zeileis, Achim. 2006. "Object-Oriented Computation of Sandwich Estimators." *Journal of Statistical Software* 16 (9): 1–16. https://doi.org/10.18637/jss.v016.i09.

Zeileis, Achim, and Torsten Hothorn. 2002. "Diagnostic Checking in Regression Relationships." *R News* 2 (3): 7–10. https://CRAN.R-project.org/doc/Rnews/.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.