Research Proposal

Kanika Chopra

Research Question

The research question of interest is *how does the time to exoneration differ based on demographic, geographic and crime factors?* This analysis aims to consider factors including race/ethnicity, age, gender, State and the type of crime to determine which significantly affect time-to-exoneration.

Wrongful sentences can take away year's from an individual's life and isolate them to the harsh environment of prison. This can have catastrophic mental and physical impacts on these individuals and their families by separating them from their loved ones and their comforts of home. On top of that, fighting for exoneration can take years and can be an expensive and mentally-exhausting process. Exoneration can be defined as a legal determination occurring when a "person who has been convicted of a crime is officially cleared after new evidence of innocence becomes available".

Previous literature has investigated the impact of race and ethnicity on time-to-exoneration through a lens of focal concerns (Braden 2021). Braden (2021) has show that black exonerees experienced a longer-time to exoneration than white exonerees and that Hispanic exonerees experienced the shortest time-to-exoneration. Furthermore, Itskovich, Factor, and Ohana (2023) have shown that Blacks in the United States are at greater risk of being wrongfully convicted for crimes such as murder and rape in comparison to White defendents. This motivates this research to further investigate the effects of gender, race/ethnicity, age, type of crime and State to determine the impact on time-to-exoneration.

Data

The dataset I am using is from the National Registry of Exonerations from the United States. This dataset has data on all of the 3,292 exonerations since 1989. My dependent variable is the time to exoneration (years). The independent variables are Race/Ethnicity, Gender, Age, Crime, and State.

Proposed Model

The proposed model is given below:

$$\mu_i = \beta_0 + \beta_1 x_{\text{Age}} + \beta_2 x_{\text{Gender}} + \alpha^{race/ethnicity}_{r[i]} + \alpha^{state}_{s[i]} + \alpha^{crime}_{c[i]}$$

$$log(t_i) \sim N(\mu_i, \sigma^2)$$

Let t_i represent the time-to-exoneration. We have a continuous variable for Age and an indicator variable for Gender. We fit Gender as an indicator variable since there are only two categories and although our data is imbalanced, Females still make up 12% of our data so it is not a very low proportion. We fit hierarchical variables for Race/Ethnicity, Crime and State. This is because in each of these categories, as shown in our EDA below, we see low proportions for some groups, i.e. Native Americans are 0.6% of our data and Asians are 0.9% of the data. As for States, we have some states with only one exonoree and similarly with Crime, we only have one exonoree for cases such as Harassment or Non-violent Attempt. More analysis would be conducted to determine appropriate interaction terms as well (i.e. Race and State, Race and Crime, etc.) Further analysis would also be conducted to determine suitable priors/parameters for this base model, i.e. setting Normal priors.

Exploratory Data Analysis

There are 19 columns on First and Last Name, Age, Race/Ethnicity, State, County of Crime, Tags, Official Misconduct (OM) Tags, Crime, Sentence, Convicted Year, Exonerated Year and various columns for the reason they were exonerated including DNA, Mistaken Witness Identification, etc.

Feature Engineering

Two features that are created within our dataset are Time to Exoneration and Gender. In our model, we are interested in modelling $log(Time)$. However, we have some data that has Time to Exonerated as 0 so to remedy this issue, we replace these values with 0.5, making the assumption that the minimum is half a year between Time of Conviction to Time of Exoneration. Gender is stored in the tags column us "F" so an indicator variable is created of whether "F" exists in the Tags.

Missing Data

We are missing 27 rows of Age and 1 row of Exonerated and Convicted year. The missing data does not seem to related to a certain Race/Ethnicity, Gender or Crime but is spread out proportionately between each category. Since it is only approximately 0.82% of our data, we drop these rows.

Data Cleaning

The following steps were also taken to simplify and clean our dataset:

1. **Filtering Race and Ethnicity Groups**: From our race/ethnicity category, we have mixed combinations, i.e. "Native American, White". For simplicity purposes, we remove any mixed combinations and unknown values. This leaves us with 3,240 exonerees.
2. **State Categories:** From our data, we also see that a lot of the states are written as "NC" or "F-NC" so we remove the "F-" to keep the data uniform.

Exploring Variables of Interest

We subset our dataset to the columns of interest, i.e. Race/Ethnicity, Age, Gender, Crime, ST and Time-to-Exoneration.

Race and Age Distributions

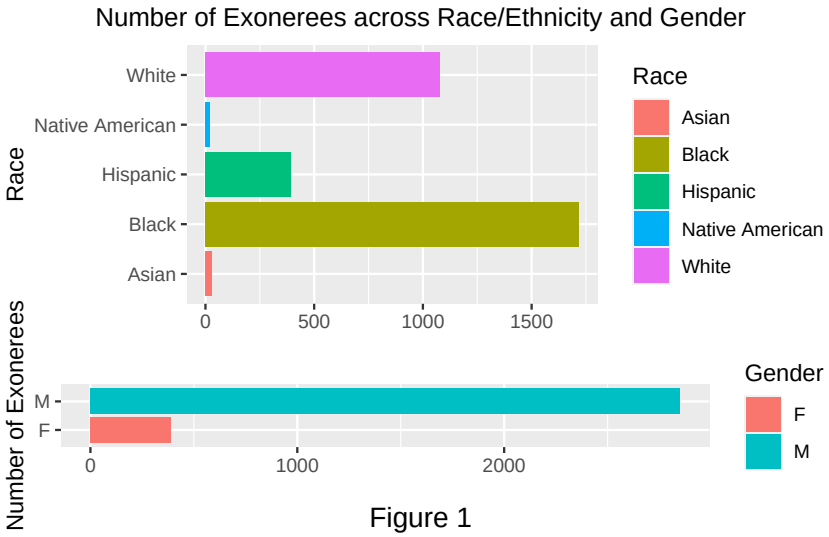


Figure 1

We can clearly see that our data is skewed towards having more Black folk, then White folk with much lower proportions for the remaining races and ethnicities. We also see that the majority of our data are male exonorees than female exonorees.

Crimes by Race/Ethnicity

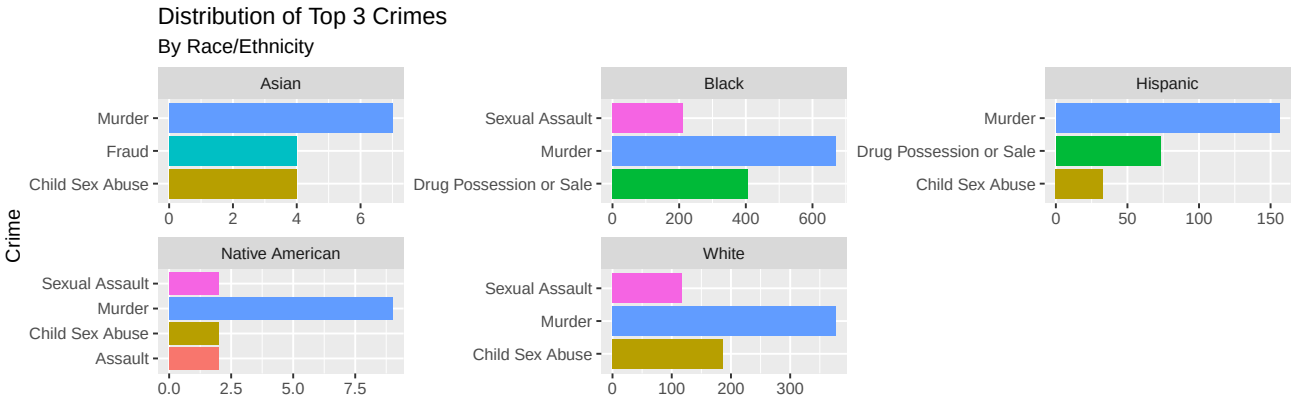
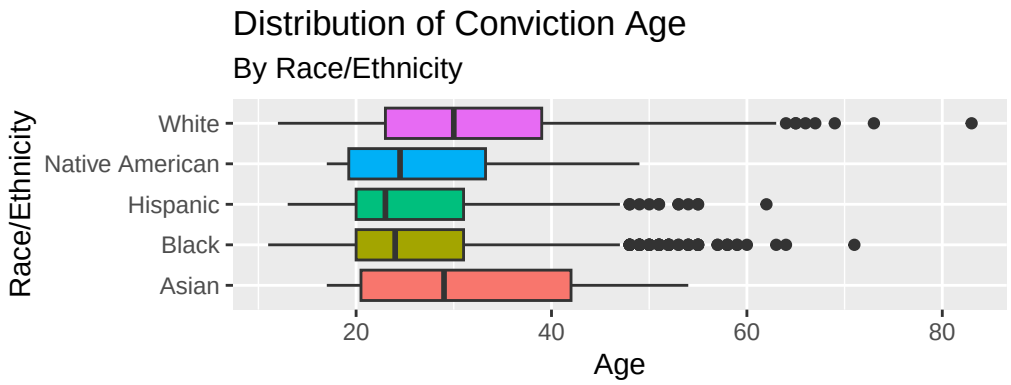


Figure 2

We look at the top 3 most common crimes for each race. Overall, the most common crimes are Murder, Drug Possession or Sale, Sexual Assault, and Child Sex Abuse. We can see some clear differences here between the types of crime most commonly accused of for the exonerees. We have that Drug Posession or Sale is much higher among Black and Hispanic exonerees while it is very low for the remaining races/ethnicities. We also observe that Assault is mainly observed by Native Americans (albeit 2 exonorees) and similarly for Fraud with Asian ethnicities. We also observe that White exonorees are the most common for Child Sex Abuse.

Age Distribution by Race/Ethnicity



Race/Ethnicity	Min	1st Qu.	Median	3rd Qu.	Max
Asian	17	20.50	29.0	42.00	54
Black	11	20.00	24.0	31.00	71
Hispanic	13	20.00	23.0	31.00	62
Native American	17	19.25	24.5	33.25	49
White	12	23.00	30.0	39.00	83

Figure 3

We see that the age range for Asians is much larger than the remaining Race/Ethnicities. This however may also be due to the fact that we only have 31 exonerees within this group. We can more clearly see the differences in data with Black, White and Hispanic exonorees being as young as 11, 12, 13 respectively. We also see that the median age for White and Asian exonorees is closer to 30 in comparison to Black, Hispanic and Native American exonorees having median ages of 23-25.

Race/Ethnicity by State

It would also be interesting to see how the race distribution differs among states - this is shown in Figure 4. We observe that a higher proportion of Black exonerees come from Illinois, New York, and Texas. The remaining states seem to have lower amounts of data. We can also see some states with more White exonerees than Black exonerees including Florida, Wisconsin and Oregon. As for Hispanic ethnicities, the majority of their data seems to come from Illinois, New York, California and Texas.

Time to Exoneration

Now, time-to-exoneration on the other hand is a continuous variable, so below we have the summary of this value as well as the distribution plotted. We plot this on the regular time-scale for easier interpretation. We also want to see how the distribution of time differs based on Race. These two distributions are shown in Figure 5.

We see more concentration around lower time-to-exoneration values, i.e. around 1-5 years. Our distribution is right-skewed with higher time-to-exoneration values being less common. We do observe that there are some cases where the time-to-exoneration is as high as 60 years. As for our comparison with race, we see that the higher values of Time seem to be associated with Black exonorees. Furthermore, Native American and Black folk have the highest medians for this time whereas Asians have the lowest time-to-exoneration. We also observe that Hispanics and White folk have similar distributions for Time with very similar median values.

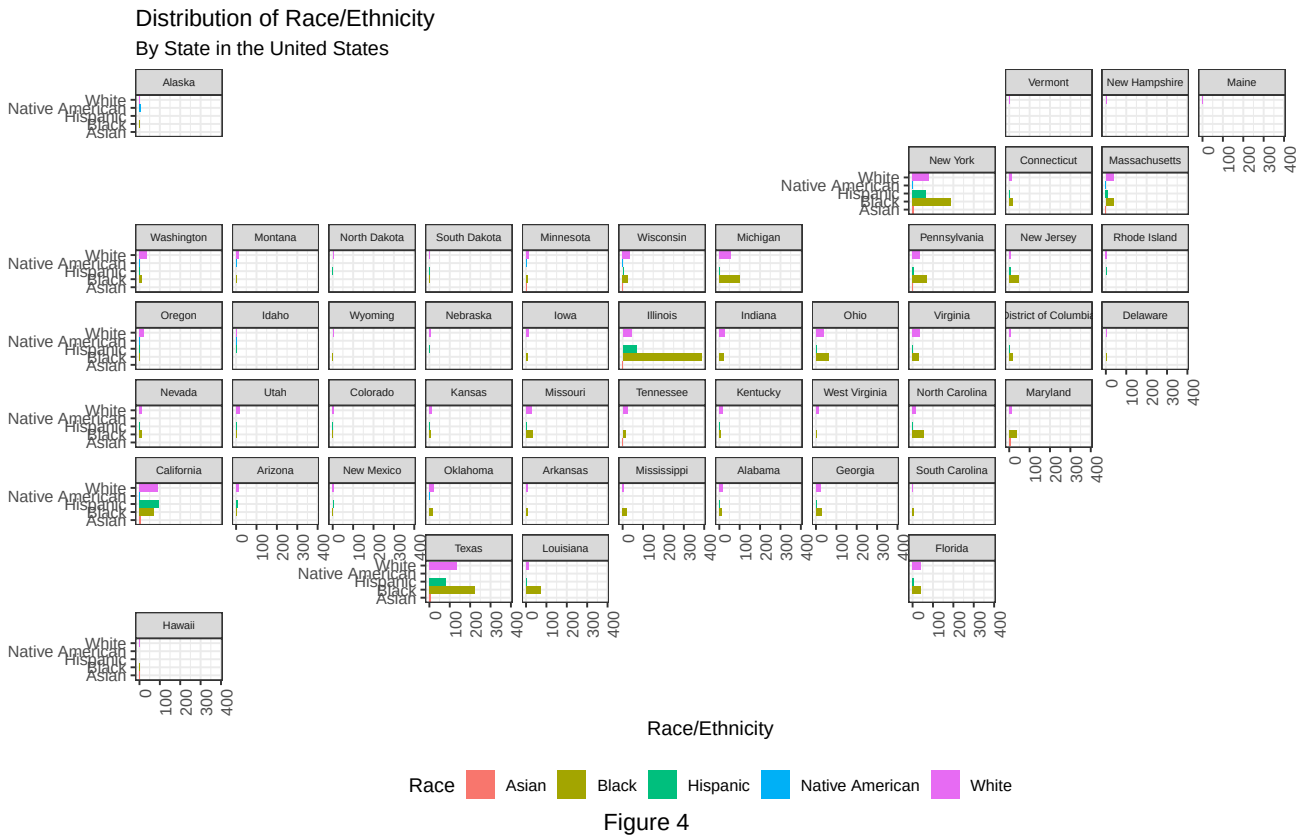
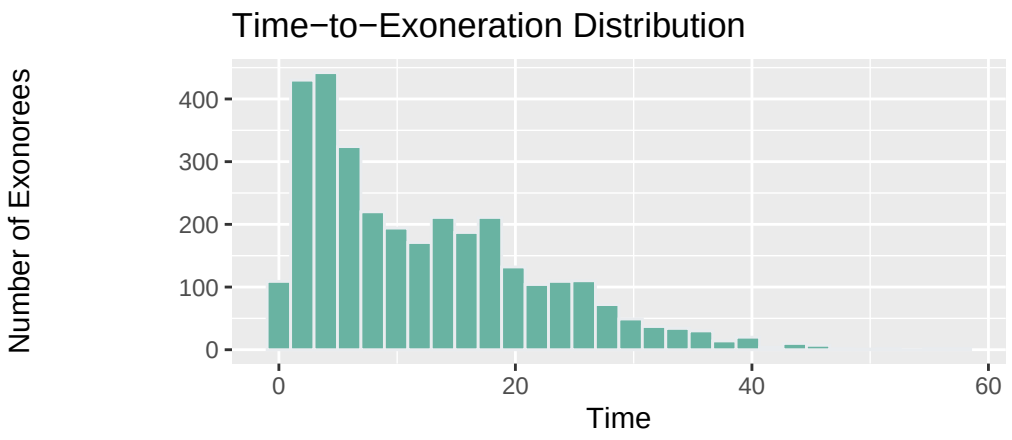


Figure 4



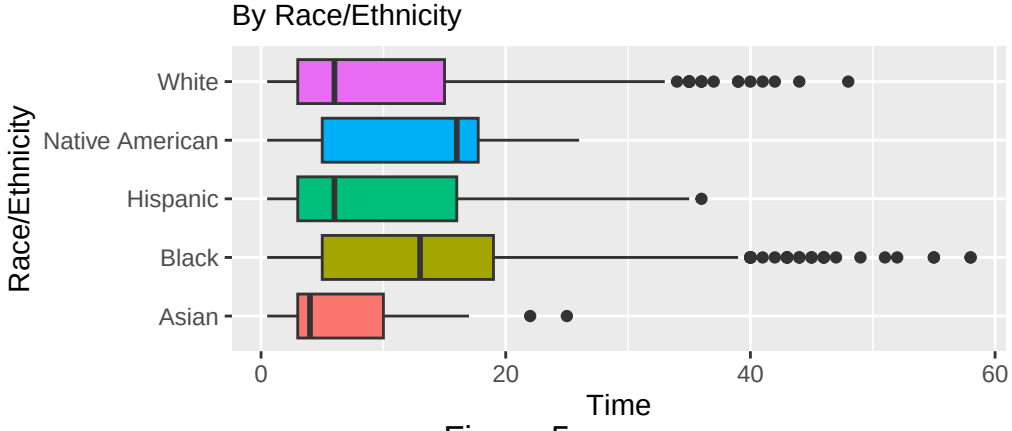


Figure 5

References

Braden, Virginia E. 2021. "In Pursuit of Innocence: A Study of Race and Ethnicity Differences in Time to Exoneration."

Itskovich, Eran, Roni Factor, and Daniel Ohana. 2023. "Haven't They Suffered Enough? Time to Exoneration Following Wrongful Conviction of Racially Marginalized Minority-Vs. Majority-Group Members." *Punishment & Society*, 14624745221148318.