

NYPD Allegation Results

Exploration and Visualization

Tyler Wenslaff

COSC 5610: Data Mining

1. INTRODUCTION

The data that will be focused comes from a database comprised of over thirty-three thousand civilian complaints filed against New York Police officers. The focus of this data analysis will be to find if there are any characteristics that can be found throughout all cases that lead to a certain board disposition over other possible outcomes.

1.1 Cleaning

The first step that I had to take before properly exploring and examining my data was to clean it. My original data set started with twenty-seven columns in it. A lot of this data is extra data that for my specified research questions, I do not need and would not have any effect on the outcome. Most of the data that I removed from the table was information retaining to the police officer that the allegation was against. A couple examples of these columns include the police officer's id, first name, and last name as these columns had too specific of information that could not be used in general graphing. After the deletion of useless columns was done, I ended up with a dataset holding twenty-two columns.

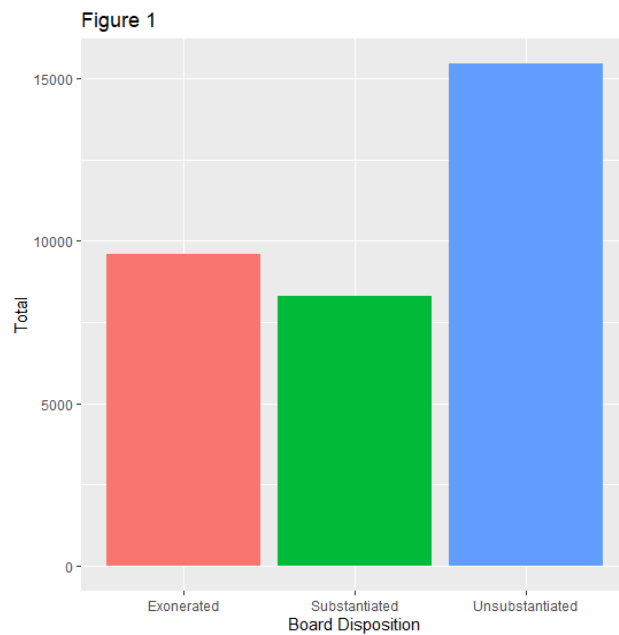
The second thing within my data that I had to clean up was the many blanks that appeared throughout my data. The main columns that had this issue were the complainant's ethnicity, gender, and age, as these columns were not filled if the complainant chose not to give their information. To solve this issue, I copied the value above each blank and copied it into the blank. This was done to all blanks, so the average between the values should not change much, especially with the amount of data that is in the dataset.

1.2 Board Disposition Data

As seen in Table 1, there are only three possible outcomes for board disposition. The first option is exonerated, meaning the stated conduct occurred, but did not violated NYPD's rules, often leading to no action being taken against the police officer. Exonerated makes up about 28.8% of all allegations. The second option, substantiated, means the conduct occurred and violated the rules. Substantiated makes up 24.9% of total reports. The final option, unsubstantiated, means an investigation occurred but nothing could be fully confirmed or denied. Unsubstantiated is the largest outcome, making up 46.3% of total outcomes. The main outcomes that needs to be focused on for this analysis is substantiated, although any patterns that can be found between any of the board dispositions will be of interest. Figure 1 gives another look at this distribution but plotted onto a bar graph to visualize the data easier.

Table 1: Board Dispositions

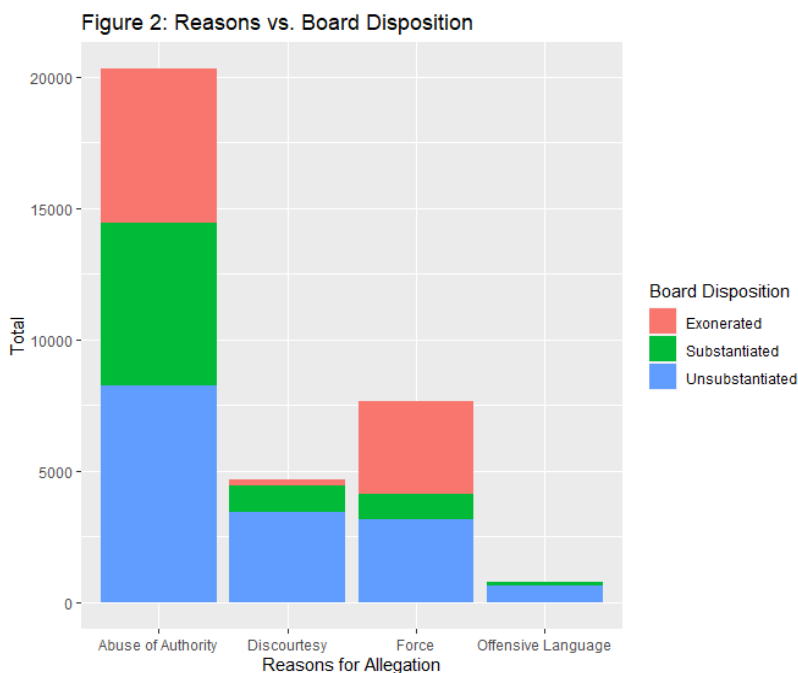
| Board Disposition | Amount | Percentage |
|-------------------|--------|------------|
| Exonerated | 9609 | 28.8% |
| Substantiated | 8299 | 24.9% |
| Unsubstantiated | 15448 | 46.3% |
| Total | 33,356 | 100% |



2. REASONS FOR ALLEGATIONS

Another metric that can be looked at is the reasons as to why an allegation happened in the first place. Allegations can happen for a multitude of reasons, but this dataset currently only obtains four reasons. Looking at Figure 2, Abuse of Authority is by far the

most common reason for an allegation. The resulting board disposition from an abuse of authority allegation is almost split perfectly into thirds, with unsubstantiated being the largest outcome.

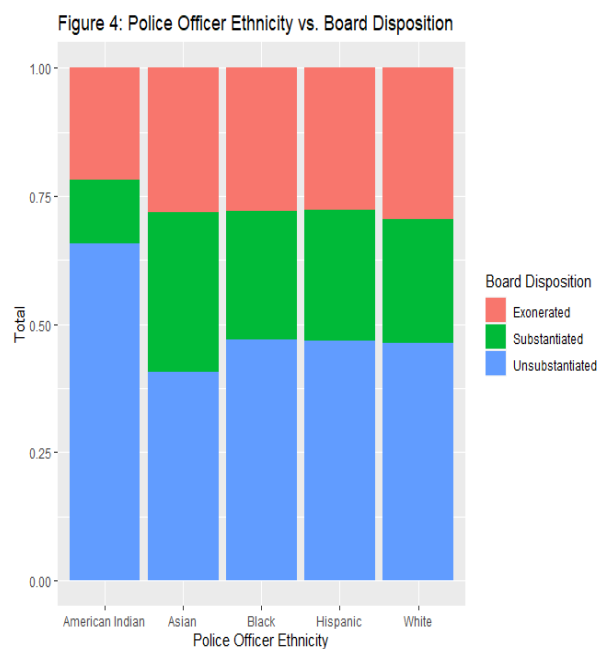
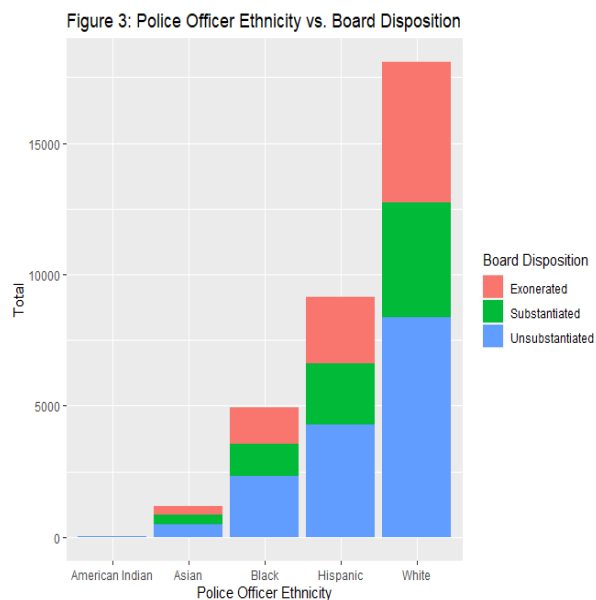


3. POLICE FACTORS

A major category of possible factors that I wanted to test were variables with the police officer that could possibly affect the resulting board disposition. These factors include the police officer's ethnicity, gender, and age.

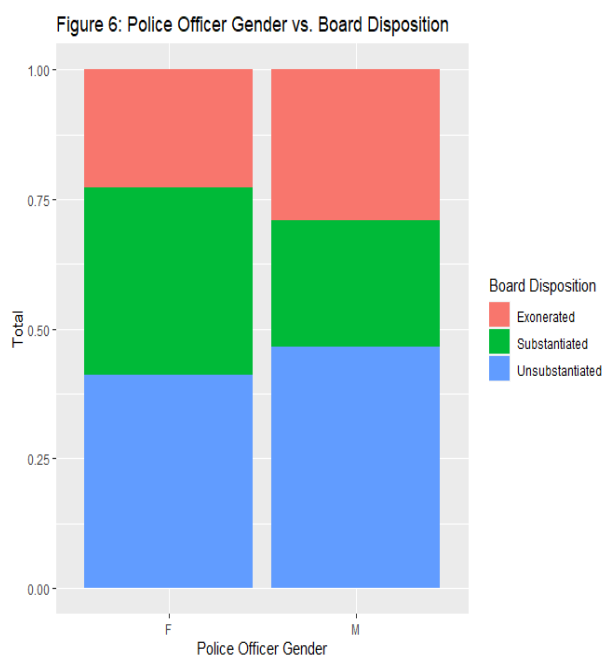
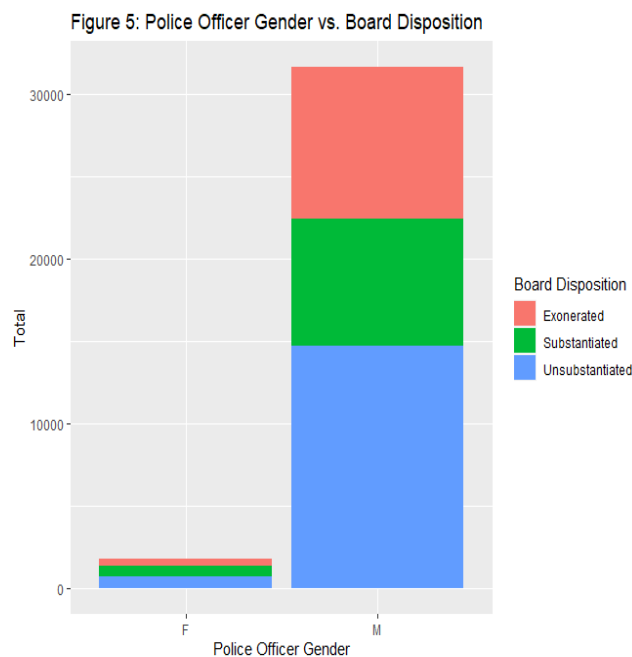
3.1 Ethnicity

Focusing on ethnicity first (See figure 3), White police officers have the most allegations against them by almost double what the second most ethnicity, Hispanic, has. Comparing the percent of each board disposition against each ethnicity (See figure 4) shows that each ethnicity has almost the same exact percent of each board disposition in their results. The furthest away ethnicity from this statement is American Indian; however, a possible reason for this is that there are far fewer allegations against American Indian officers in the first place, possibly causing a greater variance in percentages. Looking at figures 3 and 4, it can be concluded that police officer ethnicity does not have much, if any correlation to board disposition, but this will be determined for sure in section 5.



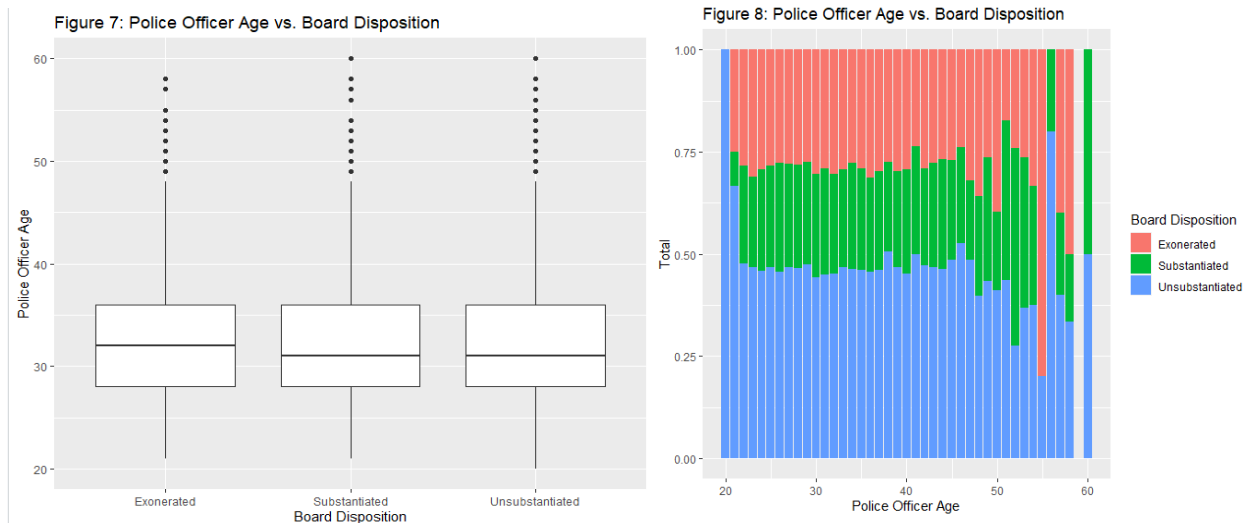
3.2 Gender

The second category related to police officers that I wanted to investigate was the gender in relation to the board disposition. Males have many more allegations than females (See figure 5); however, this is most likely due to there being many more male police officers compared to female police officers. Comparing how the board disposition makes up both groups (See figure 6), it can be seen that both males and females have about the exact same percent of all three kinds of board dispositions. Looking at figures 5 and 6, it seems as though the gender of the police officer has no correlation with board disposition.



3.3 Age

The final possible variable for police officers that could correlate with board disposition is police officer age. Looking at figures 7 and 8, both indicate that there is not much variance between all ages of police officer's that received allegations against them. There is slight variance towards the beginning and ending of figure 8, but this is most likely due to there being much less data for those specific age groups. This leads me to believe that police officer age is not correlated to the board disposition.



4. COMPLAINANT FACTORS

A second major category of possible factors that I wanted to test were variables with the complainant that could possibly affect the resulting board disposition. These factors include the complainant's ethnicity, gender, and age.

4.1 Ethnicity

The first variable I wanted to explore is the complainant's ethnicity. Looking at figure 9 and table 2, complainants with Black ethnicity have the most allegations filed against the NYPD, while American Indian have the least. This can be due to a multitude of reasons; however, that is out of the scope of this analysis. Comparing how the different board disposition options make up each ethnicity (See figure 10), each group is roughly made up of about the same of each board disposition. Those with Black ethnicity have the lowest percent of substantiated outcomes at 24%, and there are a couple of other percentages that have variance, meaning there could be some correlation between complainant's ethnicity and board disposition.

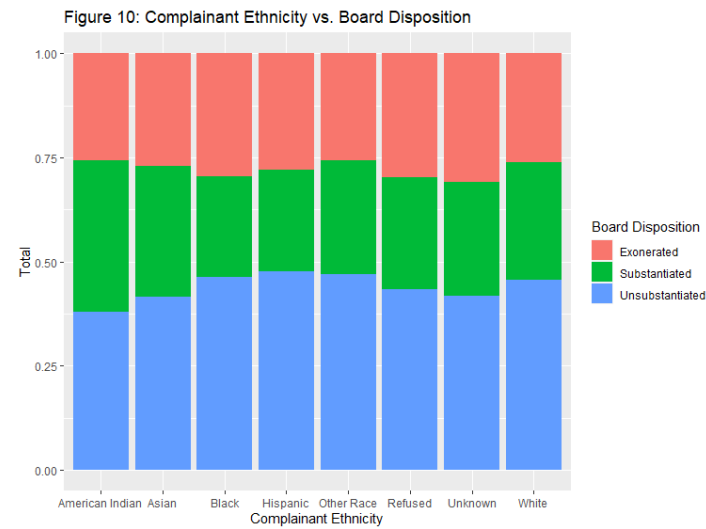
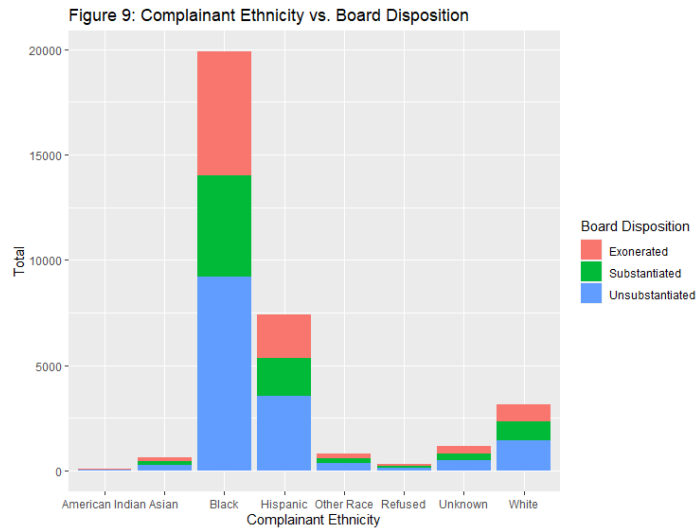
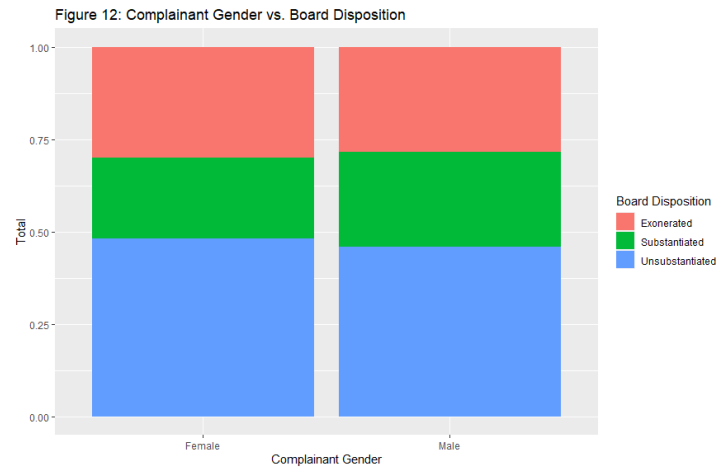
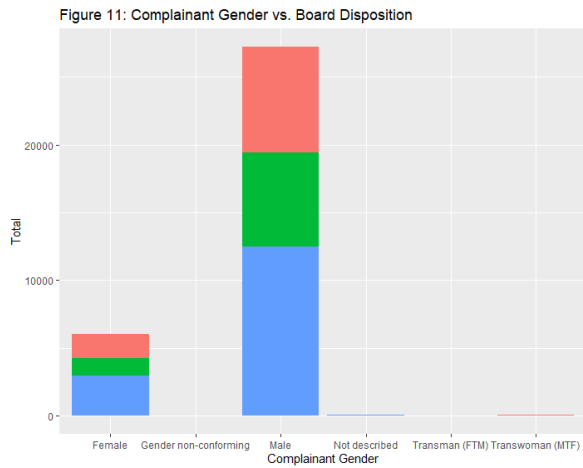


Table 2: Race vs. Board Disposition

| | Exonerated | Substantiated | Unsubstantiated | Total |
|-----------------|--------------|---------------|-----------------|--------|
| American Indian | 19 (25.7%) | 27 (36.5%) | 28 (37.8%) | 74 |
| Asian | 165 (27.1%) | 191 (31.4%) | 252 (41.5%) | 608 |
| Black | 5888 (29.6%) | 4773 (24%) | 9225 (46.4%) | 19,886 |
| Hispanic | 2070 (27.9%) | 1811 (24.4%) | 3535 (47.7%) | 7416 |
| Other | 201 (25.6%) | 214 (27.3%) | 369 (47.1%) | 784 |
| Refused | 86 (29.8%) | 78 (27%) | 125 (43.3%) | 289 |
| Unknown | 362 (30.9%) | 320 (27.3%) | 489 (41.8%) | 1171 |
| White | 818 (26.1%) | 887 (28.3%) | 1425 (45.5%) | 3130 |

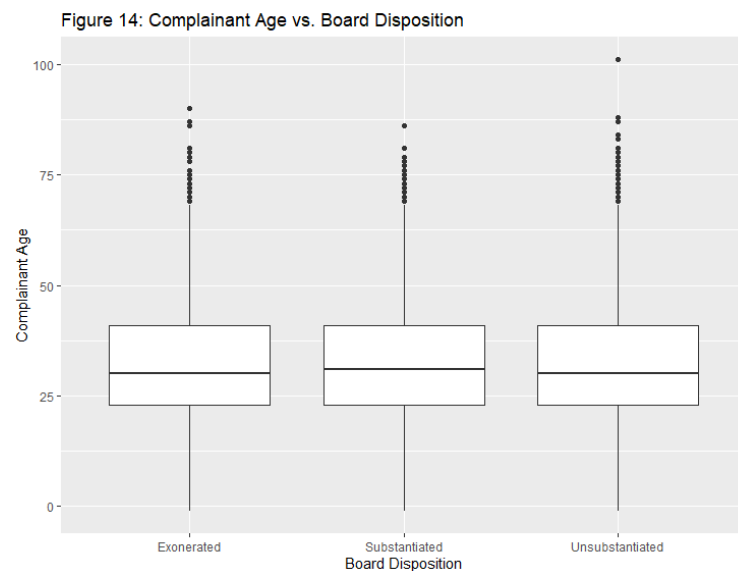
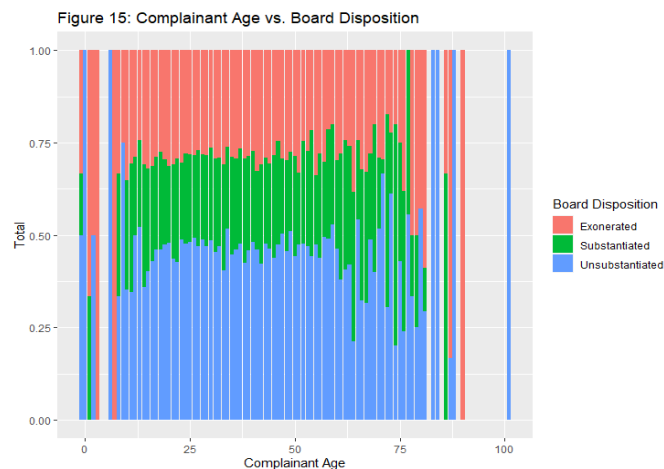
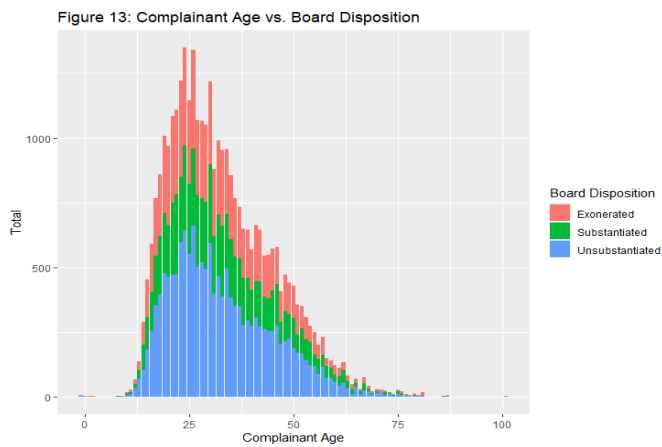
4.2 Gender

The second complainant factor I wanted to investigate was gender. Male complainants have by far the most amount of data associated with them (See figure 11), while the female group has the second most at about 6000 entries. The other four groups have little to no information entries, so my analysis will be mostly focusing on mainly male and female. Comparing how the board disposition makes up each group (See figure 12), males and females have close to the same makeup of board dispositions, but females have a lower amount of substantiated be a decent amount. Complainant gender may have a small correlation to board disposition outcome.



4.3 Age

The final complainant factor I want to investigate is age. The highest amount of allegation data comes from the age groups 20-35 (See figure 13). Figures 14 and 15 show that all age groups are made up of about the same amount of board disposition options. There is variance in the makeup of age groups towards the beginning and the end of the plots, but this is most likely due to the small amount of data at those respective points. Because of these observations, complainant age should not have that big of correlation to a specific board disposition.



5. Correlation P-value

Checking the p-value between board disposition and all previously plotted variables is a great way to check for correlation.

Looking at Table 3, if we look for all p-values underneath .05, all the previously checked variables are statistically significant to board disposition. The least significant variable checked is police age, but it is still significant at .01409.

Table 3: P Values

| | Board Disposition |
|-----------------------|-------------------|
| Allegation Reason | 2.2 e-16 |
| Police Ethnicity | 3.22 e-07 |
| Police Gender | 2.2 e-16 |
| Police Age | 0.01409 |
| Complainant Ethnicity | 2.642 e-19 |
| Complainant Gender | 2.594 e-08 |
| Complainant Age | 6.884 e-11 |

6. References

- [1] Derek Willis, Eric Umansky and Moiz Syed. The NYPD Files. Search Thousands of Civilian Complaints Against New York City Police Officers. 2020. Retrieved October 7, 2020 from <https://projects.propublica.org/nypd-ccrb/>
- [2] Mark Iris. 1998. Police Discipline in Chicago: Arbitration or Arbitrary. (Fall 1998). Retrieved October 7, 2020 from <https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=6990&context=jclc>
- [3] Michael T Compton, Roger Bakeman, Beth Broussard, Dana Hankerson-Dyson, Letheshia Husbands, Shaily Krishan, Tarianna Stewart-Hutto, Barbara M. D’Orio, Janet R. Oliva, Nancy J. Thompson, and Amy C. Watson. The Police-Based Crisis Intervention Team (CIT) Model: II. Effects on Level of Force and Resolution, Referral, and Arrest. Retrieved October 7, 2020 from <https://ps.psychiatryonline.org/doi/pdf/10.1176/appi.ps.201300108>
- [4] Kyle Rozema and Max Schanzenbach. Good Cop, Bad Cop: Using Civilian Allegations to Predict Police Misconduct. 2019. Retrieved October 6, 2020 from <https://ostromworkshop.indiana.edu/pdf/seriespapers/2019spr-colloq/schanzenbach-paper.pdf>
- [5] Bryan K. Stroube. Using allegations to understand selection bias in organizations: Misconduct in the Chicago Police Department. 2020. Retrieved October 8, 2020 from <https://www.sciencedirect.com/science/article/pii/S074959781830339X>
- [6] Kimberly D. Hassel and Carol A. Archbold. Widening the scope on complaints of police misconduct. Retrieved October 7, 2020 from https://d1wqtxts1xzle7.cloudfront.net/43329603/Widening_the_scope_on_complaints_of_poli20160303-21421-1fsqi0a.pdf?1457047954=&response-content-disposition=inline%3B+filename%3DWidening_the_scope_on_complaints_of_poli.pdf&Expires=1604626913&Signature=Kso3xyQ3HIJ5L-L3AJowudUuq4gNXkdvnzo6t4WyDYpud0nJqzcJk5cRVXM19TIqUR01HnzwZuZvBBi6zt2zDxyuzrQsmUKcrXoyXvBuFwomTxxyxFQwwO3WciRA6bBXh1NjakgnU1G0qEqTc~28mEu2Vaxfy-reYamb686Mx-T7CvqJsJ5ITYQf2QmnLY2QSGRAqW7SJ23yxCTL7pi80kybZWhxeFUThLsDoz2ssgrpRlRbQgTgkVVhlRxmePwoty5~5yg73

[Y1kjFdQqMukrqL~7v52jQDMrRA000Z2zD7buwi3rsKcRWWAcCkPf9L4LUMSBXsWQofjpoP~cytog__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA](#)