1 Executive Summary

Police practices have been the subject of intense scrutiny in the United States. In particular, many have accused the American policing system of being racially biased against people of color. To investigate this claim, this report seeks to determine two things. First, we examine the factors that may increase a person's risk of harsher treatment by the police after being arrested for possession of marijuana. Second, we look specifically at race and determine if the data support the claim that the police are racially biased.

The data used for this report contains information about 5134 individual arrestees in a major American city. In particular, the characteristics recorded include race, sex, prior traffic convictions, region, citizenship status, employment status, number of police database appearances, year of arrest, and age. A logistic model was fitted on these data, considering race, sex, citizenship status, employment status, number of database appearances, a transformed year variable, and age with the goal of predicting police detention.

Our model contains interactions and therefore it is not possible to make sweeping generalizations about the effects of certain individual characteristics on police treatment. However, we did make some important insights. For example, in this model, we found that being young (i.e. ~20 years of age) and black were associated with a remarkably high likelihood of detention. Furthermore, non-citizens were more likely to be detained than citizens, especially when they did not appear in many police databases.

Perhaps most importantly, we found evidence of racial profiling by the police. Black people were more likely to be detained than white people. Furthermore, we found that this effect was particularly strong for young age groups. As stated earlier, young black people were *far* more likely to be treated more harshly than young white people.

2 Introduction

Understanding what factors impact a police officer's decision to detain or give an arrestee a court summons gives important insight into police practices. Importantly, if there is evidence that officers are treating arrestees more harshly for arbitrary reasons such as race or sex, then a deep examination of police training and practices is needed. American citizens should live with confidence knowing that their criminal justice system does not discriminate against them based on characteristics like race.

To examine this question, we chose a bottom-up approach of analysis. We first began by analyzing and filtering out insignificant categorical predictors of police detention using contingency analyses. This was immediately followed by the same process for the numerical predictors, except with likelihood ratio tests. Transformations, changes, and other alterations to these variables were considered. Afterward, we had a list of the main effects we wanted to include in our model. This was then followed by an analysis of the possible two-way interactions between our main effects. Any insignificant interactions were filtered out. We then considered and tested 16 different models before arriving at our final model. Finally, the strengths and weaknesses of this model were assessed and are discussed in the final section.

3 Description of Subjects

Before beginning our analysis, it was important to first clean and check the dataset. Each variable was checked for anomalies. These included the outcome variable (whether an arrestee was detained or given a court summons) as well as the predictors: race (black/white), sex, number of prior traffic convictions, region of arrest in the city, employment status, number of police database appearances, year of arrest, and age.

Importantly, we found several issues within the data itself that had to be corrected. To begin, we found that one arrestee was coded as "NeverEver" for their citizenship status. We inferred that this arrestee was not likely to be a citizen and re-coded this entry as such. In contrast, several other entries had errors that were not easily fixable. One arrestee was coded as being 117 years old; another was 3 years old. Clearly, these were not plausible ages. Unfortunately, given that we could not infer their proper ages, we simply deleted these entries from the dataset. We repeated this process for every other unfixable entry. One arrestee was coded as appearing in 33 police databases (even though the maximum number was 6).

Another was coded as being arrested from the “Purple” region of the city. Furthermore, one arrestee was coded as being arrested in the 13th century (1215). Finally, one person’s race was coded as “Gr.” Given the unclear nature of the entries, we felt justified in simply deleting them from the dataset, as we had a sufficiently large number of data points.

We did not feel the need to combine any categories due to low counts nor did we find anything else that was abnormal in the dataset. Thus concluded our data cleaning. Overall, we found that 17.02% of all arrestees in the dataset were detained whereas 82.98% of them were not. Afterward, we began with a deep dive into the individual variables of the dataset. Table 1 gives a variable-by-variable breakdown of the descriptive statistics for the categorical variables. It is shown below. From it, we can observe several things. Black people appear to be detained at a higher rate than white people. Furthermore, men were far more likely than women to be detained. However, the number of prior traffic convictions appears to have little effect on police treatment given the similar rates across each category. The same appears to hold for the region of arrest in the city. That being said, we also noticed that unemployment and non-citizenship appeared to lead to higher rates of detention as well. Of course, this is just a rough initial analysis and we do not suggest any causality, but this gives us valuable clues into which variables might be significant and what effect they have on police treatment.

We also conducted chi-squared tests of association, as shown in the table, to determine which categorical predictors were associated with police treatment at the 5% significance level. We discuss these tests in-depth in the next section.

Table 1. Descriptive Statistics for Categorical Predictors of Police Detention

	Categories	Total, N (%)	Detained, N (%)	Court Summons, N (%)	χ^2 test
Race	White	3865 (75.28%)	549 (14.20%)	4585 (85.8%)	$\chi^2 = 3.089$ df = 1 p = 0.079
	Black	1269 (24.72%)	325 (25.61%)	4809 (74.39%)	
Sex	Male	4694 (91.45%)	813 (15.84%)	4321 (84.16%)	$\chi^2 = 87.185$ df = 1 $p \approx 0$
	Female	439 (8.55%)	61 (1.19%)	5073 (98.81%)	
Number of Prior Traffic Convictions	0	2038 (39.70%)	347 17.0%	1691 83.0%	$\chi^2 = 0.0051$ df = 2 p = .9902
	1	1748 (34.05%)	299 17.1%	1449 82.9%	
	≥ 2	1348 (26.26%)	228 16.9%	1120 83.1%	
Region of Arrest in the City	North	1517 (29.5%)	279 (18.4%)	1238 (81.6%)	$\chi^2 = 3.0486$ df = 3 p = 0.384
	South	1036 (20.2%)	172 (16.6%)	864 (83.4%)	
	East	1055 (20.5%)	169 (16.0%)	886 (84.0%)	
	West	1526 (29.7%)	254 (16.6%)	1272 (83.4%)	
Employment Status	Yes	4035 (78.6%)	531 (13.2%)	3504 (86.8%)	$\chi^2 = 203.19$ df = 1 $p \approx 0$
	No	1099 (21.4%)	343 (31.2%)	756 (68.8%)	
Citizenship Status	Yes	4384 (85.39%)	671 (15.3%)	3713 (84.7%)	$\chi^2 = 63.099$ df = 1 $p \approx 0$
	No	750 (14.61%)	203(27.1%)	547(72.9%)	

Table 2 gives us some descriptive statistics for our numerical predictors. From it, we can observe several things. The median arrestee appears in 1 police database, was arrested in 2004, and was 22 years old at the time of the arrest. Interestingly, we see a somewhat large standard deviation of 8.33 in age, with the youngest arrestee being only 13 years old and the oldest being 67.

More importantly, we also conducted likelihood ratio tests to determine whether any of these numerical variables were significant predictors of police treatment. We tested the individual variables separately against the null model and found that all three were significant predictors of police detention at the 5% significance level.

Table 1. Household Census Data

Predictor	Database Appearances	Year of Arrest	Age
Min	0	2001	13
Median	1	2004	22
Max	6	2006	67
StDev	1.54	1.39	8.33
p-value	$p \approx 0$	$p = 0.033$	$p = 0.002$

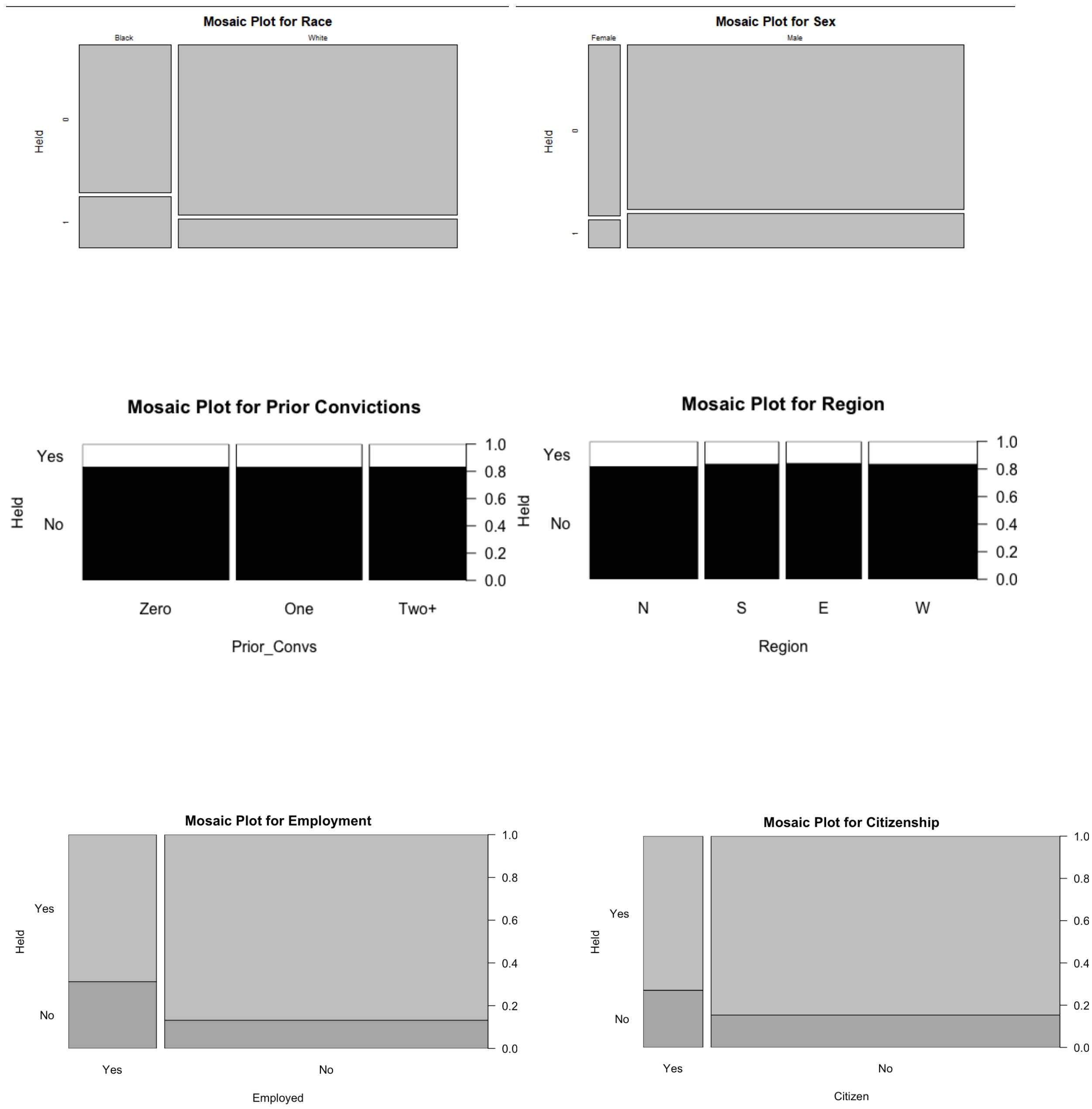
4 Results

We begin with contingency analyses of the categorical predictors in Table 1. For these analyses, we conducted chi-squared tests of association for every single categorical predictor to determine if there was a statistically significant association between the predictor and police treatment.

From Table 1, we found that two categorical predictors were not significantly associated with police treatment at the 5% significance level: number of prior traffic convictions and region of arrest in the city. Every other categorical predictor was found to be significantly associated with police treatment. These were race, sex, employment status, and citizenship status. Based on this analysis, we felt confident in excluding an arrestee's number of prior traffic convictions and region of arrest from our model. Mosaic plots were made to support our contingency analyses. They are shown collectively, separated by each categorical variable, in Figure 1. Visually, we see that the two variables we excluded seem to have very little effect on the likelihood of detention. In contrast, we see clear differences in the likelihood of detention for the other categorical variables.

Additionally, after evaluating the empirical probabilities at each level of the categorical predictors, we considered various combinations of levels. For example, the probability of an arrestee being held in the North and West regions and the South and East regions were almost identical. We created a new region variable that had two levels: "NorthWest" and "SouthEast" and conducted a likelihood ratio test to examine these variables' predictive power. The new p-value was slightly lower (0.258), but still insignificant at the 5% level. For the number of prior traffic convictions, the empirical probabilities of an arrestee being held were nearly identical among the three levels (17%, 17.1%, and 16.9%), and thus, no logical combination of levels was apparent. We concluded that our four significant categorical predictors (race, sex, employment status, and citizenship status) were best left untouched, as our analysis did not reveal any need to change or transform them.

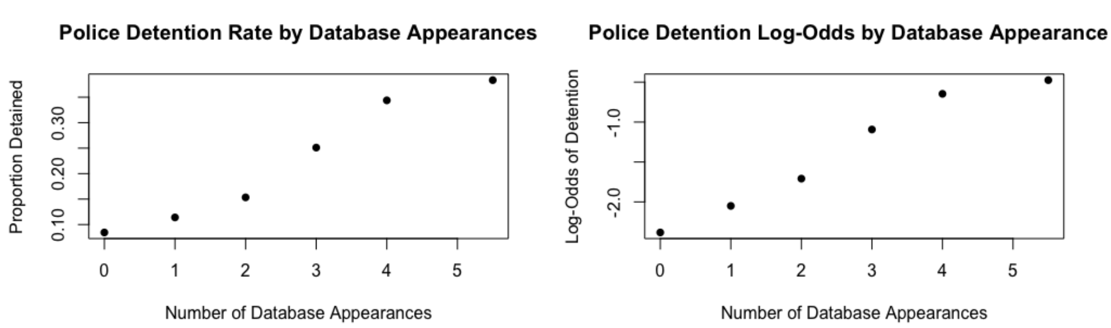
Figure 1. Mosaic Plots for Categorical Predictors

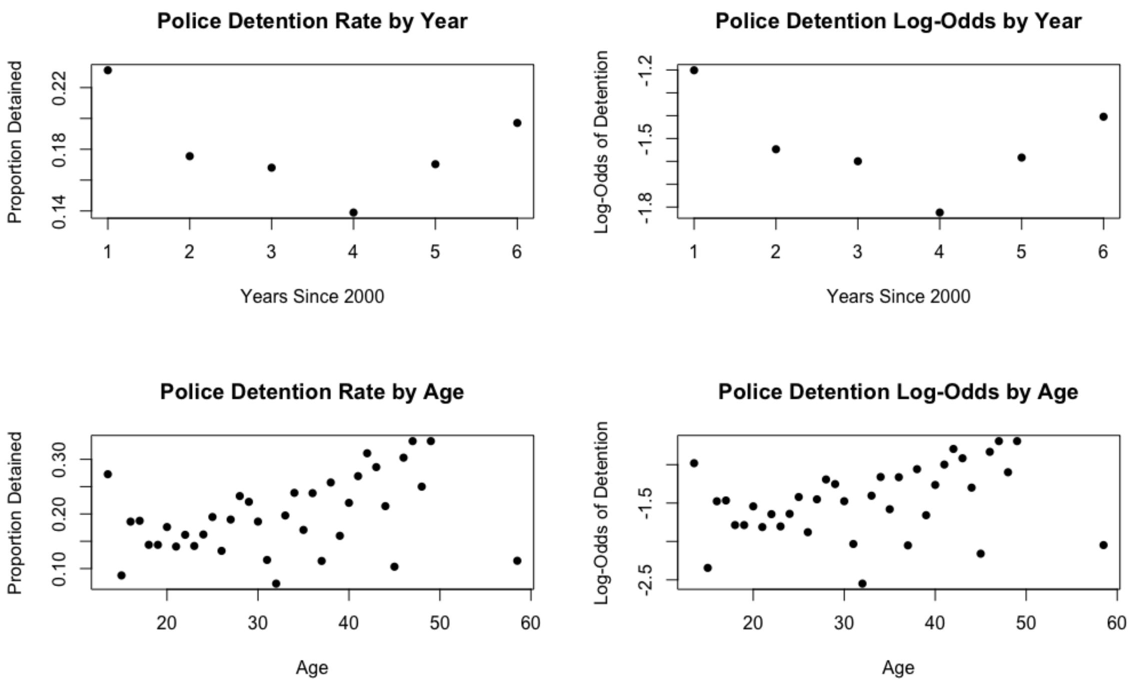


Additionally, in the previous section, we showed which numerical predictors were statistically significant in predicting police treatment at the 5% level. All three numerical variables were found to be significant. However, it was important to determine whether any transformations or changes were needed before blindly including these variables in the model.

We began by making six slicing plots for each variable: three for the empirical probabilities and three for the log odds. Unfortunately, there were less than 10 arrestees who appeared 6 times in police databases. Thus, we collapsed these entries into those arrested with 5 database appearances. As for the year variable, we made a quick linear transformation by deducting 2000 so as to avoid large numbers like 2001, 2002, etc., thus making a new variable for the number of years since 2000. Finally, for the age variable, we collapsed the 13-year-olds into the 14-year-olds category and did the same for the 51-67 age category into the 50-year-olds, due to low counts. The slicing plots are shown in Figure 2.

Figure 2. Slicing Plots for Database Appearances





Importantly, we see that although the relationship is not perfect, we can be rather confident that the number of database appearances is generally linear with the log odds of police detention. Similarly, we felt the same way about age and its relationship with the log odds of detention. Thus, we felt confident that the number of database appearances and age variables required no transformations.

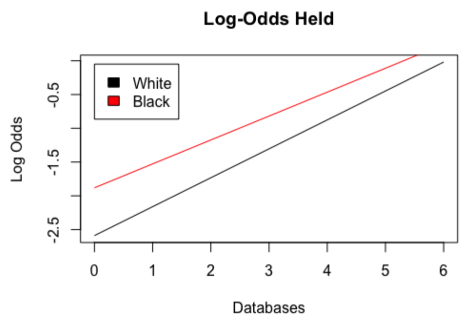
That being said, the year variable was not linear with the log odds of detention. It exhibits a quadratic behavior which prompted us to apply a square root transformation to this variable.

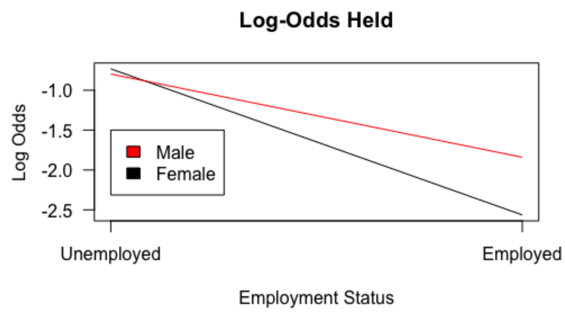
We have thus established that our model will take into consideration seven different variables: race, sex, employment status, citizenship status, number of police database appearances, the transformed year variable, and age. Clearly, logistic regression is the appropriate method for fitting these data. The outcome is binary (detention/court summons), and our assumptions for this model appear to be met. The observations are independent and our numerical predictors now have linear relationships with the log odds of the outcome variable.

Additionally, as a quality check, we also ran three logistic regressions with only these individual numerical predictors to examine the directions of the effects on police detention. We found that as either the number of database appearances increased or as one's age increased, so did (generally) the likelihood of detention. The opposite was true for the transformed year variable.

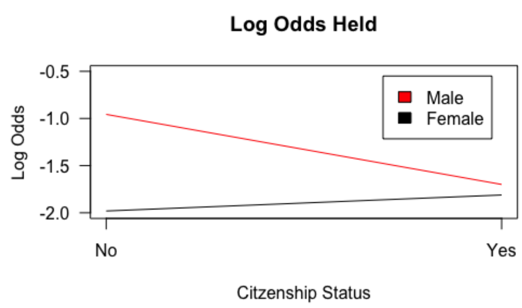
Having established which variables would be included in the final model, we then examined whether there was any evidence of any two-way interactions between these variables. With 4 categorical variables and 3 numerical variables, there are a total of 18 possible interactions. To detect which variables had genuine interactions, we made 18 different plots of the log odds for each combination of variables. Parallel lines indicate little-to-no interaction whereas non-parallel lines do. They are shown below in Figure 3. Beneath each row of plots is also a pair of p-values for the interaction term for a test of the significance of that particular interaction term, as merely a visual analysis is an insufficient way of uncovering genuine interactions. Those p-values that are less than 0.05 are colored green; those that are not are colored red.

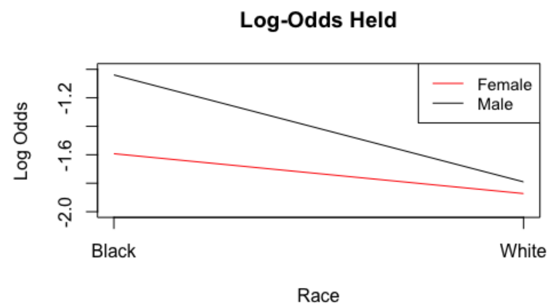
Figure 3. Log Odds Plots for Detecting Interactions



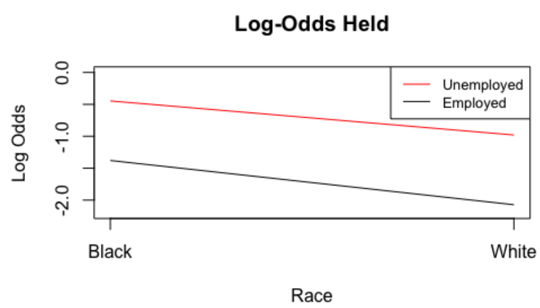


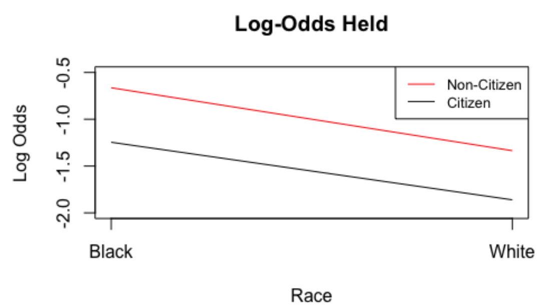
p-values (left to right): 0.172, 0.09



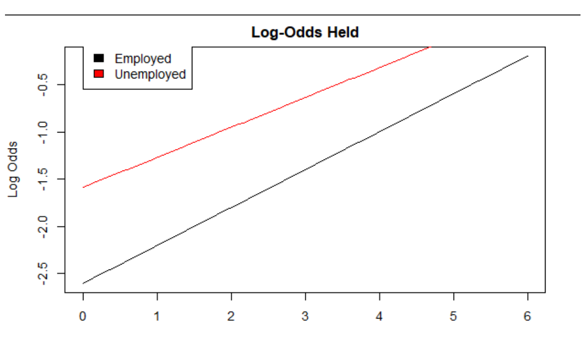


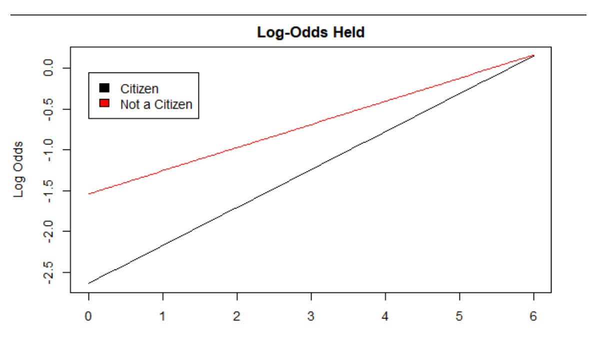
p-values (left to right): 0.103, 0.193



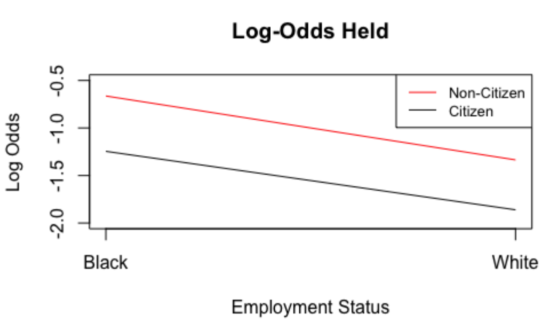


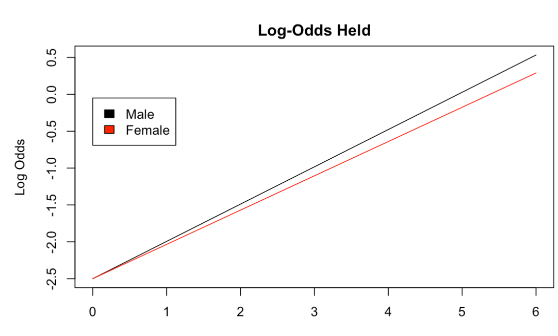
p-values (left to right): 0.344, 0.767



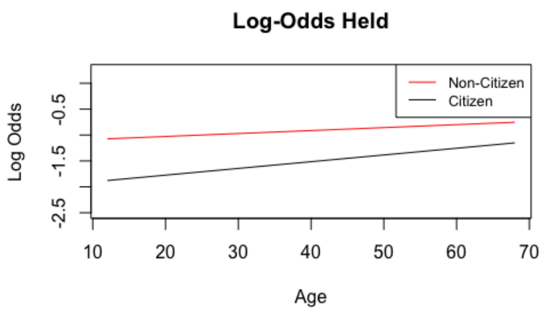


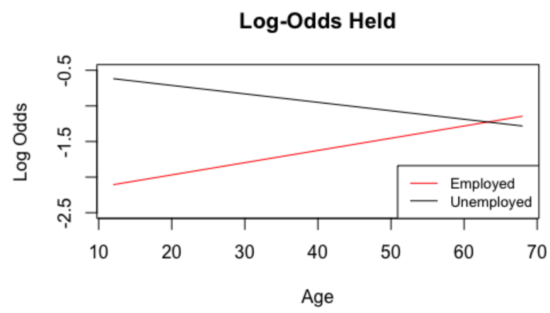
p-values (left to right): 0.130, 0.0021



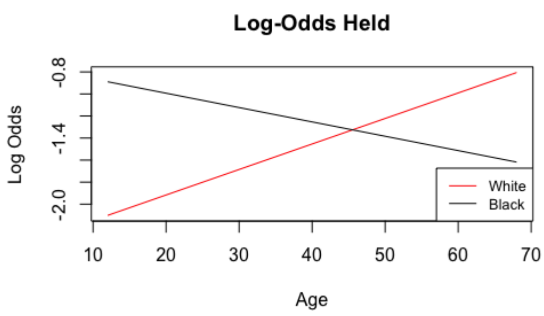


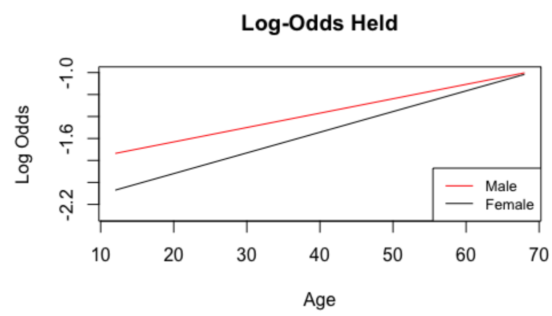
p-values (left to right): 0.317, 0.656



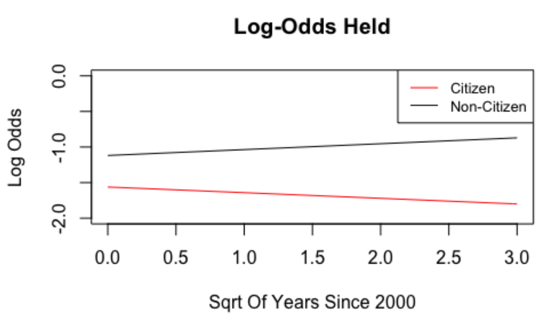


p-values (left to right): 0.4750, 0.0017





p-values (left to right): 0.000119, 0.714



p-values (left to right): 0.415, 0.08897

p-values (left to right): 0.012281, 0.00702

We have now ruled out which interactions can be wholly excluded from the model. Thus, the only significant interactions are those between:

(1) citizenship and the number of database appearances
(2) age and employment status
(3) race and age
(4) race and transformed years
(5) sex and transformed years

However, this is too many interactions to work with. If we were to consider every single possible combination of main effects and interactions, we would be forced to test dozens of possible models. Thus, we decided to restrict our significance level to 1% in order to keep only the most significant interactions. This left us with four possible interactions:

(1) citizenship and the number of database appearances
(2) age and employment status

(3) race and age
(4) sex and transformed years

This leaves us with only 16 models to test (a far more reasonable amount). With this information, we can now build and compare various models to determine which model overall is superior and will be our final model for predicting police treatment. The results are shown in Table 3. In it, we use a form of shorthand for writing out our different potential models. The letters R, S, E, C, D, Y, and A stand for race, sex, employment status, citizenship status, number of database appearances, the transformed years variable, and age respectively. An interaction is noted with a lowercase "i" followed by the two variables of interest (e.g. iRS for an interaction between race and sex).

Table 3. Comparison of Various Models

#	Model	DF	Deviance	GOF Test P-Value	P	AIC	BIC	ROC AUC
1	R,S,E,C,D,Y,A	5126	4221.1	p ≈ 1	8	4237.1	4289.49	0.7233
2	R,S,E,C,D,Y,A,iCD	5125	4212.2	p ≈ 1	9	4230.2	4289.078	0.7248
3	R,S,E,C,D,Y,iAE	5125	4214.1	p ≈ 1	9	4232.1	4291.02	0.7223
4	R,S,E,C,D,Y,A,iRA	5125	4208.4	p ≈ 1	9	4226.4	4285.291	0.7247
5	R,S,E,C,D,Y,A,iSY	5125	4217.2	p ≈ 1	9	4235.2	4294.108	0.7249
6	R,S,E,C,D,Y,A,iCD,iAE	5124	4205.4	p ≈ 1	10	4225.4	4290.802	0.724
7	R,S,E,C,D,Y,A,iCD,iRA	5124	4200.2	p ≈ 1	10	4220.2	4285.603	0.7259
8	R,S,E,C,D,Y,A,iCD,iSY	5124	4207.9	p ≈ 1	10	4227.9	4293.303	0.7262
9	R,S,E,C,D,Y,A,iAE,iRA	5124	4202.8	p ≈ 1	10	4222.8	4288.23	0.7256
10	R,S,E,C,D,tY,A,iAE,iSY	5124	4209.5	p ≈ 1	10	4229.5	4294.945	0.7245
11	R,S,E,C,D,Y,A,iRA,iSY	5124	4204.5	p ≈ 1	10	4224.5	4289.917	0.7263
12	R,S,E,C,D,Y,A,iCD,iAE,iRA	5123	4194.7	p ≈ 1	11	4216.7	4288.639	0.7269
13	R,S,E,C,D,Y,A,iCD,iAE,iSY	5123	4200.3	p ≈ 1	11	4222.3	4294.312	0.7261
14	R,S,E,C,D,Y,A,iCD,iRA,iSY	5123	4195.9	p ≈ 1	11	4217.9	4289.855	0.7277
15	R,S,E,C,D,	5123	4198.2	p ≈ 1	11	4220.2	4292.179	0.727

	Y,A,iAE,iRA,iSY							
16	R,S,E,C,D, Y,A,iCD,iAE,iRA,iSY	5122	4189.7	$p \approx 1$	12	4213.7	4292.189	0.7282

According to Table 3, every model has fantastic fit according to the goodness-of-fit test. Thus, our choice of the final model comes largely down to the AIC and BIC. The model with the lowest AIC is model 16 whereas the model with the lowest BIC is model 4. Since these models do not have an excessive number of parameters, we preferred the lowest AIC model over the lowest BIC model. Thus, model 16 was selected as our final model.

As a quality check, we conducted an analysis of deviance for this model and discovered that while sex, the transformed year variable, and age were not significant predictors on their own, the fact that their interactions with other variables *were* significant allowed us to keep these variables in the model.

We now present the general model:

$$\mathrm{logit}(\Pi) = \beta_0 + \beta_W(I_{White}) + \beta_M(I_{Male}) + \beta_E(I_{Employed}) + \beta_C(I_{Citizen}) + \beta_D(\mathrm{Databases}) + \beta_{Sqrt(Years)}(\sqrt{Years}) + \beta_A(\mathrm{Age}) + \beta_{C*D}(I_{Citizen})(\mathrm{Databases}) + \beta_{E*A}(I_{Employed})(\mathrm{Age}) + \beta_{W*A}(I_{White})(\mathrm{Age}) + \beta_{M*Sqrt(Years)}(I_{Male})(\sqrt{Years})$$

where Π is the probability of being detained by the police. The parameters are defined as follows:

I_{White} = 1 if the arrestee is white, 0 if they are black.
I_{Male} = 1 if the arrestee is male, 0 if she is female.
$I_{Employed}$ = 1 if the arrestee is employed, 0 if they are unemployed.
$I_{Citizen}$ = 1 if the arrestee is a U.S. citizen, 0 if they are not.
Databases = the number of police databases the arrestee appeared in (ranges from 0 to 6).
$\sqrt{Years}$ = the square root of the years since 2000 in which the person was arrested.
Age = the age of the arrestee.

Next, we provide a table with the actual parameter estimates, odds ratios, odds ratio confidence intervals, and p-values for each predictor in the final model below in Table 4. Odds ratios are obtained from the 95% Wald confidence intervals for the coefficients. Similarly, the p-values are obtained from the Wald test. The symbol * denotes $p < 0.05$.

Table 4. Parameter Estimates, Odds Ratios, Odds Ratio Confidence Intervals, and p-values for Each Predictor in the Model.

Variable	Estimate	Odds Ratio	Odds Ratio 95% CI	p-value
β_0	-1.253	–	–	0.129
β_w	-1.218	0.296	(0.174, 0.503)	≈ 0*
β_M	1.682	5.378	(1.116, 25.917)	0.036*
β_E	-1.379	0.252	(0.149, 0.427)	≈ 0*
β_C	-0.965	0.381	(0.275, 0.529)	≈ 0*
β_D	0.220	1.247	(1.123, 1.384)	≈ 0*
$\beta_{Sqrt(Years)}$	0.861	2.365	(1.061, 5.271)	0.035*
β_A	-0.0396	0.961	(0.942, 0.981)	0.0001*
β_{C*D}	0.178	1.195	(1.061, 1.356)	0.003*

β_{E*A}	0.0243	1.025	1.005, 1.044	0.013*
β_{W*A}	0.0324	1.033	1.013, 1.054	0.001*
$\beta_{M*Sqrt(Years)}$	-0.917	0.400	0.175, 0.912	0.029*

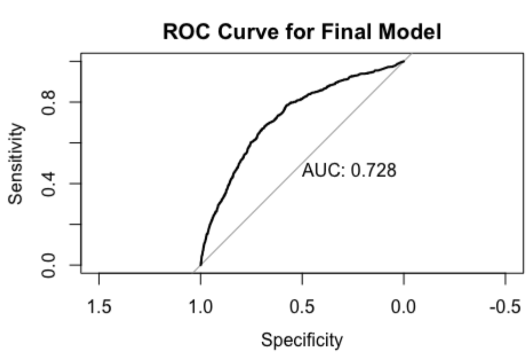
The odds ratios and confidence intervals were strictly based on the exponential of the estimate for a given β_i because all the chosen main effects were involved in at least one interaction. Therefore, these odds ratios cannot be used in interpreting the change in odds. Thus, we offer the interpretable odds ratios in Table 5.

Table 5. Interpretable Odds Ratios

Variable	Odds Ratio
Effect of Race	OR (effect of being white at median age = 22) = 0.603 OR (effect of being white at younger age = 19) = 0.547 OR (effect of being white at older age = 28) = 0.733
Effect of Sex	OR (effect of being male at median year of arrest = 2004) = 0.860 OR (effect of being male at more recent year of arrest = 2005) = 0.693 OR (effect of being male at older year of arrest = 2002) = 1.471
Effect of Employment	OR (effect of being employed at median age = 22) = 0.430 OR (effect of being employed at younger age = 19) = 0.400 OR (effect of being employed at older age = 28) = 0.497
Effect of Citizenship	OR (effect of citizenship at median number of database appearances = 1) = 0.456 OR (effect of citizenship with less database appearances= 0) = 0.381 OR (effect of citizenship with more database appearances = 3) = 0.651
Effect of Databases	OR (effect of +1 increase in database appearances when citizen) = 1.490 OR(effect of +1 increase in database appearances when non-citizen) = 1.247
Effect of Year of Arrest	OR (effect of +1 increase in sqrt(year of arrest since 2000) when male) = 0.946 OR (effect of +1 increase in sqrt(year of arrest since 2000) when female) = 2.365
Effect of Age	OR (effect of +1 increase in age when employed and white) = 1.017 OR (effect of +1 increase in age when unemployed and white) = 0.993 OR (effect of +1 increase in age when employed and black) = 0.985 OR (effect of +1 increase in age when unemployed and black) = 0.961

Next, we assessed our model overall using three tools: the ROC curve, a classification table, and the chi-squared goodness-of-fit test. We present the ROC curve first in Figure 4.

Figure 4. ROC Curve for the Final Model



We see from the ROC plot that the AUC is 0.728, which signifies that the accuracy of our predictions from our model will be somewhat satisfactory. Overall, the ROC plot gives us confidence in our model. We also present a classification table, shown in Table 6. We set a threshold of 0.5.

Table 6. Classification Table

Observed	Predicted

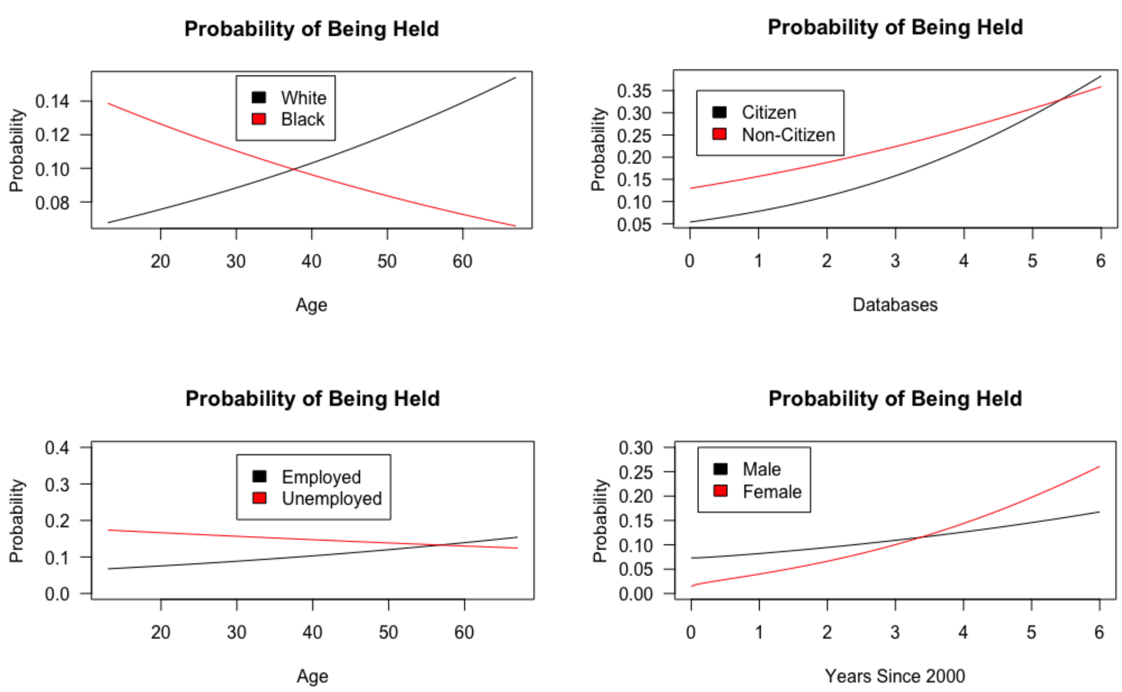
	No	Yes
No	821	53
Yes	4218	42

From Table 6, we see that our model correctly predicted the outcome of 863 observations and incorrectly predicted the outcome of 4271 observations. In particular, while our model excelled at avoiding false positives, it clearly struggled to detect true positives. We explain why we think this is so in the following section, but we suspect that police treatment of arrestees actually has a remarkable degree of randomness to it.

We also conducted a goodness-of-fit test on the model. While the results are in Table 3, we go into the test in more detail here. The residual deviance was found to be equal to 4189.7 on 5122 degrees of freedom. The p-value was found to be very close to 1. We fail to reject the null hypothesis at the 5% significance level. Thus, we are satisfied that the model has very good fit overall.

Lastly, we provide probability plots, focusing on the significant interactions. We provide plots for race and age, citizenship status and the number of database appearances, age and employment, and sex and transformed years. In each plot, the numerical predictors vary along the x-axis and a curve is presented for each level of the categorical variable. All other variables are held at a median level. For the numerical predictors, we used the median values provided in Table 2 and for the categorical variables, the "median" value we chose was the level that had a higher proportion of observations (presented in Table 1).

Figure 5. Probability Plots



From these probability plots, we notice one clear, striking observation: young black people are *far* more likely to be detained than young white people. We know from the real world that young black men often feel targeted by the police, and these data seem to support that sentiment numerically.

We also see that non-citizens are generally more likely to be detained by police than citizens at lower levels of database appearances. Furthermore, we also see that those who are young and unemployed have a higher likelihood of being detained than those who are young and employed, though the difference is not striking. Lastly, we note that closer to the year 2000, men tend to have a higher probability of detention than women, although again, the difference is not striking.

5 Discussion

Our model contains interactions with every main effect, so it is not possible to make sweeping conclusions about the individual characteristics of an arrestee on police treatment without considering interaction terms. However, we found that *in general*, the following characteristics were associated with a higher risk of police detention: being black, male, unemployed, or a non-citizen (based purely on the main effects). Furthermore, being young, appearing in many police databases, or being arrested in years further from 2000 were also *in general* more likely to lead to a higher risk of police detention (also based purely on the main effects alone). However, we stress that these effects are not uniform and urge the reader to refer to our probability plots, as the interactions between these variables led to non-uniform effects on the likelihood of detention based on increasing age, time since the year 2000, and number of database appearances.

Perhaps the most important question we must address is whether the data support the notion that there is evidence of racial profiling by the police. In our view, that answer generally leans toward yes. Recall that the chi-squared test of association found a significant difference in the effect on police treatment by race. Furthermore, we found that the main effect of being white (as compared to black) reduced the log-likelihood of being detained since the coefficient was negative (-1.218) as well as significant.

Additionally, and perhaps most strikingly, the interaction between race and age is hard to ignore. The probability plots *clearly* show that young black people have a far higher likelihood of being detained than young white people. One might object that the other end of this plot suggests that old white people have a higher likelihood of being detained than old black people. We respond in two ways. First, this is still racial profiling. Second, we do not reject this argument, but our data had a much higher concentration of observations for younger age groups than older ones. Perhaps this argument could be supported by more data on older arrestees, but we felt more comfortable drawing conclusions about young black people than we did for old white people. Thus, not only do we find evidence of racial profiling among police, but the effect is far more pronounced for younger ages.

However, we admit our analysis was not perfect. Importantly, we take the time now to discuss flaws with our model. We do not note it above, but the null deviance was actually found to be equal to 4684.9. With a residual deviance of 4189.7, even if our model has good fit and a relatively low AIC, the reality is that there is a lot of unexplained deviance going on in the background. This is supported by our model assessment analysis which found that our model struggled to detect true positives (i.e. had poor sensitivity). We suspect either two things. The first possibility is that our data is not telling us the full story. Perhaps there is a characteristic that officers are clearly weighing above all else when judging whether or not to detain an arrestee that this data did not note. The second (and perhaps more likely) possibility is that predicting a policeman's judgment is rather difficult. After all, the police are not a uniform body, but rather, a group of individuals who all have different judgments. Their actions as a whole may be somewhat random and difficult to predict. This does not detract from our conclusion about racial profiling, but rather, seeks to explain some flaws with our model.

We did not encounter any serious flaws during our analysis, but we did raise some important questions. First, we were admittedly puzzled by the employment variable, even if it was a significant predictor. Unlike race or sex, one's employment status is not a particularly visible trait, nor do we imagine that policemen usually ask an arrestee for their occupation when judging whether to detain them. We suspect collinearity. Being willing to commit crime may lead to challenging employment prospects, but we strongly doubt whether employment itself actually has any causal relationship with police detention.

Furthermore, we struggled a little bit with our interaction terms. The interactions between race and age and citizenship status and database appearances made heuristic sense, but we struggled to completely understand the interactions between age and employment status and especially, transformed years and sex. These interactions were kept in the model due to their significance and good fit, but it was difficult for us to completely reconcile with the existence of these interactions.

We conclude this report by exploring avenues of further research. We are particularly interested in seeing whether *individual* characteristics of the police officer are predictive of their treatment of arrestees. Would black officers treat arrestees differently than white officers? What about officers who were relatively new vs. veteran policemen? What if the police officer was wearing a body camera (which was likely non-existent in the early 2000s, but still something to consider)? Furthermore, we would have loved to explore more recent data, especially in the wake of high-profile events of police brutality and killings against people of color to see if we can find similar trends of racial profiling in modern-day data. Overall, the avenues of exploration are many, and we are excited to see future developments in this line of research.