Social Identity, Law Enforcement Capacity, and Criminal Justice Contact

Marty A. Davidson, II
madavidson2@wisc.edu

Kaneesha R. Johnson
kaneesha.johnson@unc.edu

Frank R. Baumgartner
frankb@unc.edu

Abstract

Does police enforcement capacity explain differences in contact rates with the criminal legal system? We use a census of over seven million defendants from the North Carolina judicial system, each person who was arrested from 2013 through 2019, to understand how increasing the per capita number of sworn officers affects arrest rates across different social and demographic groups. We use geocoded residential addresses to map each individual to their US Census place, and we create variables for race, gender, and age group that correspond to US Census categories. This allows us to compare directly to the Census and to estimate the odds of arrest for any distinct demographic group across the state. We find strong evidence for the differential policing hypothesis, showing clearly how powerfully criminal justice contact and outcomes related to state action rather than to differential rates of criminal behaviors.

Keywords: Criminal Justice, Race and Politics, Identity, Intersectionality, Disadvantage

Introduction

Public discussions on contact rates with the criminal legal system often devolve into a blame attribution problem between the defendants being processed and law enforcement agents. All sides recognize the over-representation of minority individuals in the criminal justice system. Some attribute this to differences in underlying rates of criminal behavior and others suggest that greater police attention to minority neighborhoods, and more aggressive police behavior within them, explain the differences in contact rates. It would be nice to think that being arrested is associated with violating important aspects of the social contract: people who cause harm to others should be arrested, whereas law-abiding citizens should not. It would be disturbing to think patterns of arrest are strongly related to one's surveillance context rather than to rates of involvement in criminal behavior.

To understand whether police enforcement capacity drives this relationship, we investigate here the ***differential policing hypothesis***, which stipulates that differences in contact rates across jurisdictions can be attributed to differences in policing capacity. Here we focus on the correlational relationship between contact rates and the number of sworn officers per 100,000 residents. By focusing on the per capita number instead of the raw number of sworn officers, we can approximate community surveillance rates. As this chapter reveals, levels of policing are highly correlated with levels of arresting. This should come as no surprise. But the arrest rates are anything but neutral with respect to social identity. Thus, differential rates of policing must be taken into account when considering differential rates of contact with the courts. Some of us live in communities that are lightly policed compared to others, and this makes a big difference in our odds of arrest, irrespective of our behavior. Within each community, law enforcement focus is higher on those with marginalized identity characteristics. So, we see a combination of social identity and police capacity generating large differences in criminal justice contact.

Drawing on our database of every person cited or arrested in North Carolina from January 1, 2013 through December 31, 2019, we construct a "contact score" estimating the odds of contact with the criminal justice system. For each 100 individuals with any particular demographic profile, how many are arrested? We can do this for any type of crime (e.g., traffic offenses, wildlife violations, or violent felonies), and here we focus on the overall trends and on basic differences between infractions, traffic, misdemeanors, low-level felonies, and high-level felonies. Note that these contact scores are not statistical estimates; they are observed values. For any demographic group, we can count how many such people live in the state of North Carolina and the number who are arrested for various crimes. Thus, for any crime, we can see the demographic profile of what groups of individuals have the greatest or the lowest odds of such an arrest. We begin by discussing how we construct our contact scores.

Contact Rates with the Criminal Legal System

We begin our process for estimating contact rates by identifying all relevant demographic subgroups that exist in North Carolina. By permuting U.S. Census categories on race (Asian, Black, Hispanic (Latino/a/x), Native, and White), gender (Male and Female), age cohort (18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 and over), and geographic location, we identify 50,341 unique demographic subgroups within the state. We use these subgroups as the building blocks for more aggregate demographic categories, such as those by race, gender, age-group, or location. A key element in our construction of these demographic groups is that we can correlate them precisely with the US Census. For any location, we can assess, for example, how many Asian-American men aged 25 to 34 live there. These values allow us to assess, for each group, how many are arrested and then to calculate what percent of all such people had criminal justice

contact. The idea is very simple, even if the logistics are highly demanding of computer processing power.

After identifying all relevant demographic subgroups, we generate numerator values – number of individuals who had contact with the criminal legal system – by identifying all unique NC-AOC cases for each calendar year between 2013 and 2019. Our database is a copy of the court records maintained by the North Carolina Administrative Office of the Courts (NC-AOC), which makes a five-year extract of the database available for research purposes. We originally obtained the database for 2013 through 2017 and later supplemented records going through 2019, merging them to create a seven-year snapshot of the entire state court system, with over 13 million charges against seven million individuals (North Carolina Administrative Office of the Courts. 2014).

Using this publicly available data, we geocode the residential addresses of defendants and match them to appropriate U.S. Census place designations. We then count the number of defendants for each demographic subgroup by tabulating across reported race, gender, age group, and US census place designation. Finally, we generate denominator values by calculating population totals for each subgroup using 2018 5-year ACS estimates for individuals 18 years or older. To remove the issue of zero-denominator subgroups[1] and proportions greater than 100 percent[2], we use the upper estimates of the 95 percent confidence interval as the population total (ACS + 95 MOE). Appendix A provides more technical details on our procedures and estimates.

---

1 Subgroups where the 5-year ACS estimate equals zero. This issue commonly occurs for minority group members in rural locations that the ACS does not frequently sample with at-home surveys.
2 Subgroups with high contact rates. Approximately 20 percent of demographic subgroups fall into this category when using ACS estimates.

Between 2013 and 2019, very few individuals had contact with the criminal legal system. As Figure 1 illustrates, approximately two in 100 individuals had contact with the state's system each year. This includes all offense categories, such traffic, vehicle, property, drug, violent, and sexual crimes.
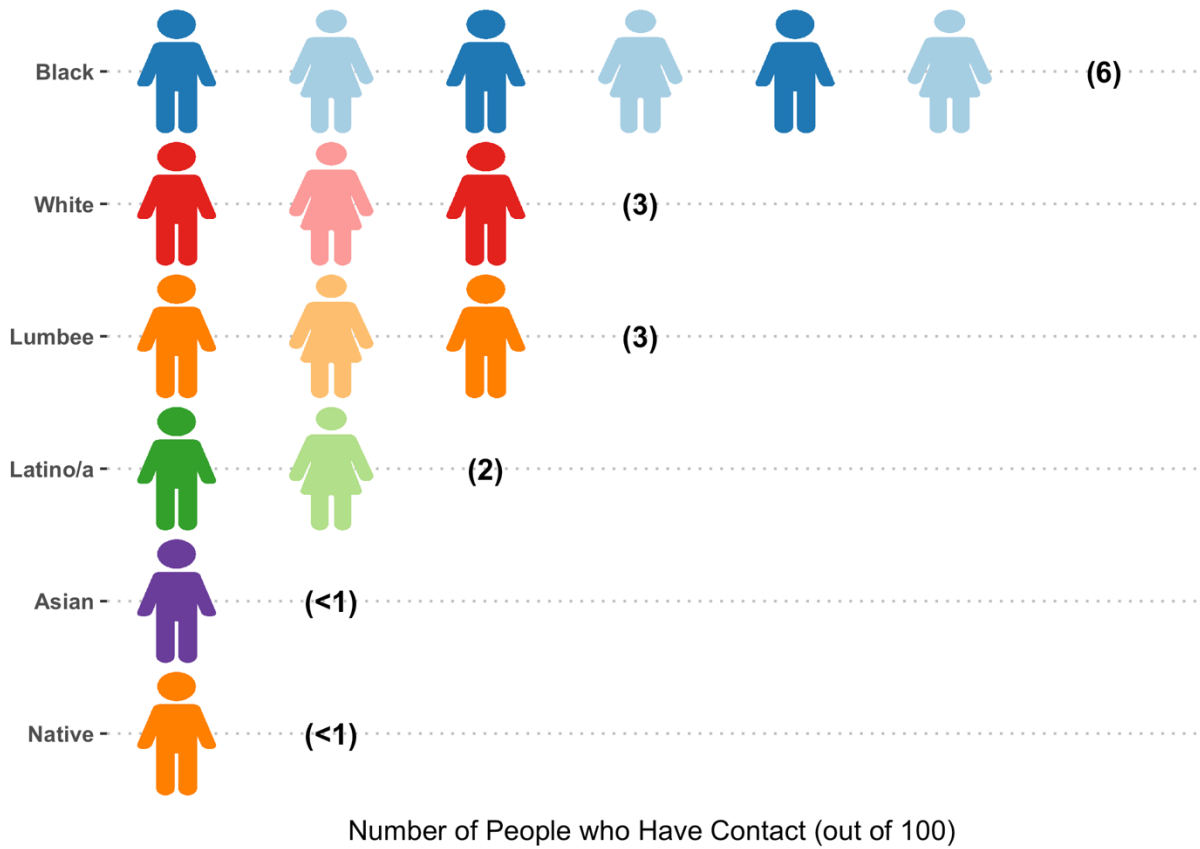
Figure 1. Overall Yearly Rates of Contact.
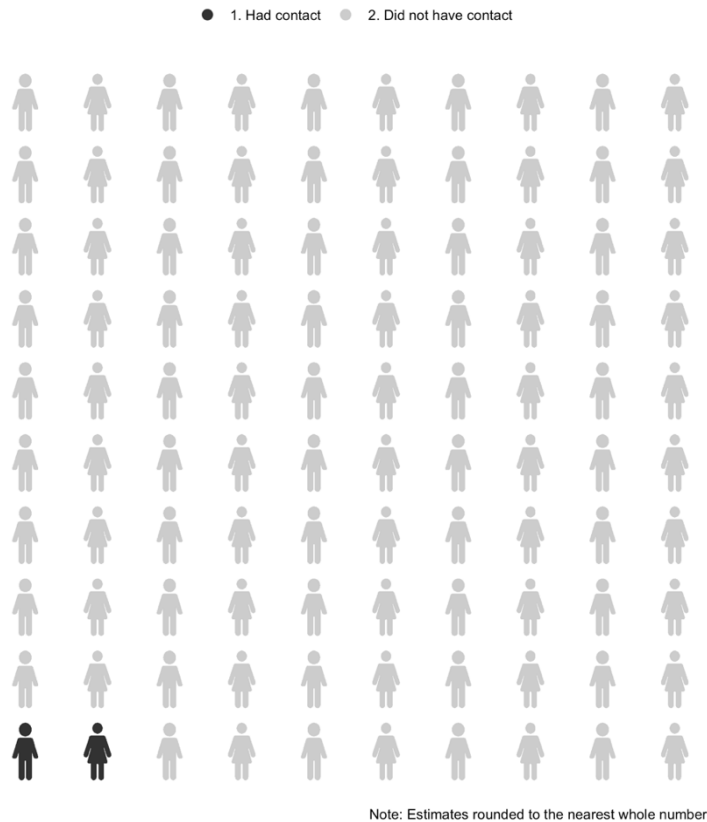
Note: each figure represents one percent of the population.

This low contact for the state's entire population, however, does not extend to all racial and ethnic groups, however. As Figure 2 illustrates, Black individuals are twice as likely to have contact with the criminal legal system as the statewide average. This is followed by members of the Lumbee tribe – Native Americans who reside in Hoke, Scotland, and Robeson counties – and White individuals. By contrast, Latinx individuals, Asian-Americans, and Native Americans outside of the Lumbee tribe were much less likely to have such contact.

Figure 2. Estimated Yearly Contact Rates across Racial and Ethnic Groups.

Black ‑ (6)
White ‑ (3)
Lumbee ‑ (3)
Latino/a ‑ (2)
Asian ‑ (<1)
Native ‑ (<1)

Number of People who Have Contact (out of 100)

Note: Estimates rounded to the nearest whole number

When we introduce gender and age, the racial differences from Figure 2 become more pronounced. As Figure 3 illustrates, these racial differences are driven, in part, by intragroup age and gender differences. Specifically, young and male-based cohorts have increased contact with the criminal legal system than older cohorts. For example, 24 out of 100 Black men between 25 and 34 have contact with the criminal legal system in a given year compared with 1 out of 100 Black men over the age of 65. Clearly, we gain a lot by looking at age groups as well as by breaking down the demographic groups by race and gender. All racial and gender groups lose contact with the criminal justice system when they enter their retirement years. Among younger individuals, however, we see great differences by race and gender, with Black men consistently seeing the highest rates of contact.

Figure 3. Estimated Yearly Contact Rates Across Racial, Ethnic, Gender, and Age Cohorts

Number of People who Have Contact (out of 100)

Note: Estimates rounded to the nearest whole number
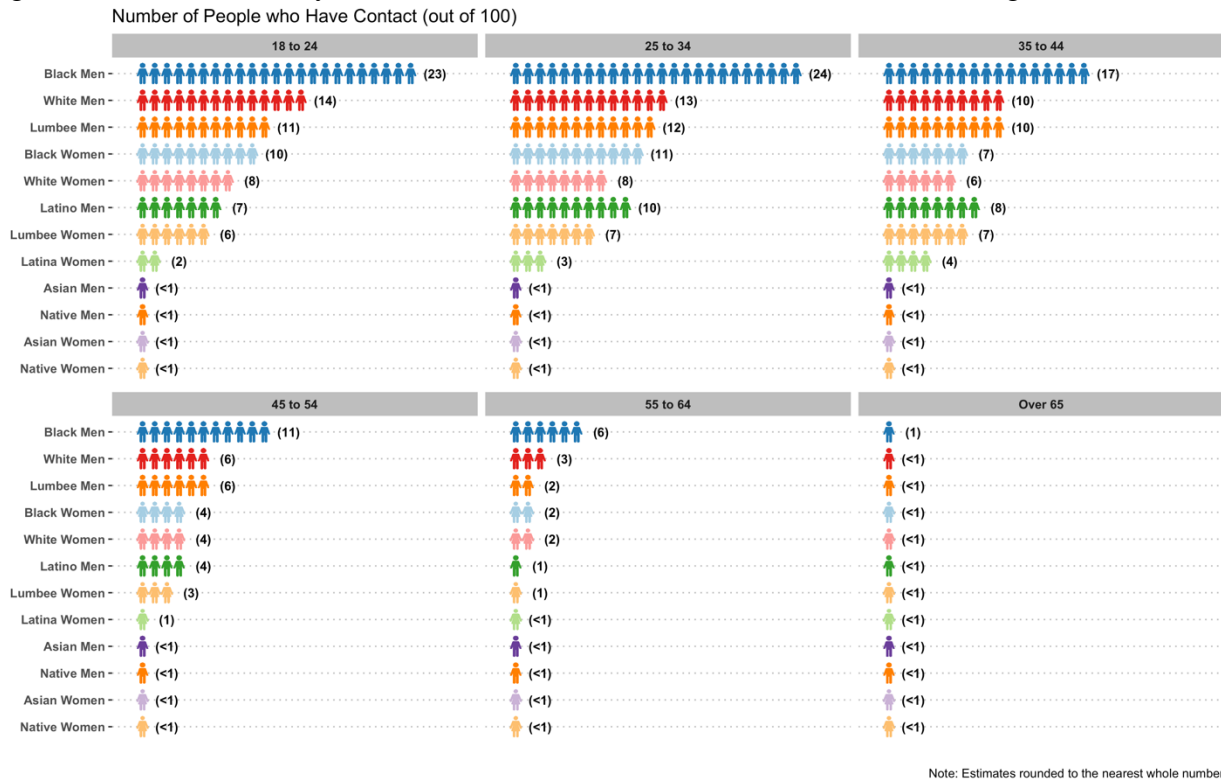
Where someone lives often has a greater impact on their chances for having contact with the criminal legal system than their descriptive attributes. Figure 4 looks at rates of contact overall, then separately for Black, Latinx, and White individuals. The Figure illustrates that contact with the criminal legal system is concentrated in certain jurisdictions, such as the state's large urban centers (i.e., Charlotte, Greensboro, Raleigh, etc.). Because the Figure shows rates of contact per 100,000 members of the population, this is not a simple artifact of population size. Rather, these geographic differences approximate the state's economic divisions where urban regions possess large enclaves of low-income neighborhoods subject to high rates of policing per capita.

Figure 4. Estimated Contact by Race and Geography
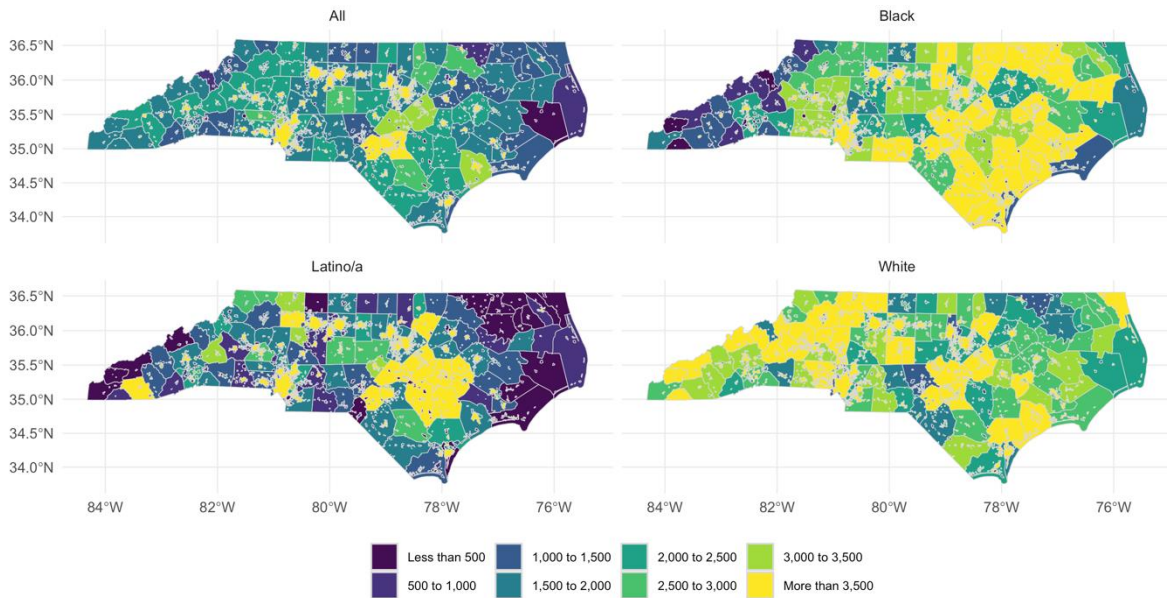
Note: Geography refers to the census place associated with the home address of the person arrested.

When broken down by race, gender, and location, a clear pattern emerges whereby Black men in urban enclaves have the highest rates of contact while white and Asian women have the lowest rates of contact. Table 1 lists the top 20 demographic who have frequent contact with the criminal legal system. Black men occupy all but two of these 20 positions, with North Carolina's top five jurisdictions all being represented. Note that if the overall rate across the state is approximately two percent (see Figure 1), these rates for black men in the towns listed are 30 times higher than this. Also recall that if we were to add age to this breakdown, we would see even higher rates of contact. If 565.6 out of every 1,000 black men in Winston-Salem can be expected to be arrested in any given year, this rate is most likely considerably higher for those at the peak age group for arrest (see Figure 3), and considerably lower for those over the age of 65.

Table 1. Demographic Groups with Highest Rates of Contact.

Location	Group	Population	Average Number who have Contact each Year	Contact per 1,000 People
Winston-Salem	Black Men	29,422	16,495	565.6
Wilson	Black Men	8,704	3,770	520.9
Wilmington	Black Men	8,040	4,498	516.2
Gastonia	Black Men	8,140	3,834	433.4
High Point	Black Men	14,793	6,338	378.5
Greenville	Black Men	12,694	5,309	369.0
Greensboro	Black Men	43,030	16,333	325.4
Goldsboro	Black Men	8,134	3,137	314.5
Burlington	Black Men	6,436	2,393	309.9
Raleigh	Black Men	48,853	16,853	308.2
Winston-Salem	Latino Men	11,644	3,963	302.9
Rocky Mount	Black Men	13,561	4,758	295.8
Kinston	Black Men	5,721	2,107	293.5
Fayetteville	Black Men	33,256	11,553	292.1
Charlotte	Black Men	100,152	33,271	281.1
Durham	Black Men	36,414	12,257	277.6
New Bern	Black Men	5,006	1,391	237.4
Kannapolis	Black Men	5,074	1,436	228.0
Jacksonville	Black Men	7,911	2,103	224.0
Henderson	Black Women	5,429	1,070	222.4

Note: List includes demographic groups with populations over 5,000

Contrast with the groups found in Table 2, which lists the 20 demographic groups who have the lowest rates of contact. Asian and White women occupy every spot with exception of one, white men leaving in the golf community of Pinehurst.

Table 2. Demographic Groups with Lowest Rates of Contact

Location	Group	Population	Average Number who have Contact each year	Contact per 1,000 People
Hendersonville	White Women	7,554	422	21.2
Durham	Asian Men	6,679	166	21.1
Cary	White Women	45,685	1,525	20.5
Cary	Asian Men	12,096	289	20.2
Boone	White Women	10,024	251	19.6
Summerfield	White Women	5,182	152	19.4
Salisbury	White Women	9,154	401	19.4
Eden	White Women	5,683	294	18.2
Davidson	White Women	5,567	139	18.2
Matthews	White Women	12,121	369	16.3
Pinehurst	White Men	7,283	191	15.3
Mint Hill	White Women	10,669	310	14.2
Raleigh	Asian Women	10,687	161	12.7
Greensboro	Asian Women	6,482	110	12.4
Southern Pines	White Women	6,419	136	11.5
Charlotte	Asian Women	22,245	299	10.8
Cary	Asian Women	12,631	170	10.6
Elon	White Women	6,262	86	9.8
Durham	Asian Women	7,303	79	9.4
Pinehurst	White Women	7,898	127	9.0

Note: List includes demographic groups with populations over 5,000

If black men in Winston-Salem have a rate of contact of 565.6, and white women in Pinehurst have a rate of 9.0, the obvious question is whether the two groups occupy such divergent places because of their behavior and other characteristics, or because of differential rates of policing. Obviously, there is no reason to reach a one-or-another answer to this question; the true answer could be that behavior and surveillance both matter. Our point here is only to show that there are fast differences and that these relate not only to demographics, but also to geography. Some places have more police than others.

Enforcement Capacity and Contact with the Criminal Legal System

Does police enforcement capacity explain differences in contact rates across demographic groups? To answer this question, we investigate the correlational relationship between our estimated contact scores and police enforcement numbers provided by the Uniform Crime Reporting (UCR) database. To simplify our analysis and avoid describing outlier jurisdictions, we exclude seasonal tourist jurisdictions, such as beach towns and mountain communities. North Carolina has many communities, particularly on the beaches but as well in the mountains, where the number of seasonal visitors is very high compared to the number of year-round residents. Policing in those communities follows different patterns, and for that reason we exclude them in this analysis. Overall, these exclusions apply only to a small proportion of all those arrested throughout the state so is likely to have only a marginal impact on our analysis. Including them, on the other hand, would skew the data with a small number of communities with apparently very large numbers of police officers per 1,000 residents. Appendix B provides more details on how we identify these communities.

We distinguish among different levels of contact here. Table 3 lays out the numbers of arrests in our database by the offense level. Like other states, North Carolina distinguishes among felonies, misdemeanors, traffic charges, and infractions. Misdemeanors may range from Class A-1, the most serious, to Class 3; traffic violations may be mere infractions at the lowest level or correspond to the same levels (and punishments) as the misdemeanors. Felonies range from Class A, capital murder, to Class H and I, which typically do not involve active prison time for those found guilty of them. In our analysis below, we distinguish among: Infractions; Traffic violations; Misdemeanors; Low Felonies (Class G, H, and I); and High Felonies (Class A to F). Note that small percentages of the offenses are not classified. Also note that a single arrest incident may involve many charges, some at different levels. Here, we treat each charge as a

separate unit of analysis. In other analyses, we use only the highest level charge. For the

purposes of our analysis here, however, it is most appropriate to look at the individual charge.

Table 3. Arrests by Offense Level.

Offense Class	N	%
Felony - Class A	3,552	0.03
Felony - Class B1	16,762	0.12
Felony - Class B2	3,612	0.03
Felony - Class C	25,781	0.19
Felony - Class D	52,268	0.39
Felony - Class E	75,045	0.55
Felony - Class F	62,130	0.46
Felony - Class G	95,629	0.71
Felony - Class H	778,630	5.75
Felony - Class I	518,236	3.83
Felony - Unspecified	407,783	3.01
Misdemeanor - Class A1	375,794	2.78
Misdemeanor - Class 1	1,112,201	8.21
Misdemeanor - Class 2	540,945	4.00
Misdemeanor - Class 3	790,404	5.84
Misdemeanor - Unspecified	475,752	3.51
Traffic - Class A1	1,628	0.01
Traffic - Class 1	291,536	2.15
Traffic - Class 2	459,471	3.39
Traffic - Class 3	5,577,493	41.19
Traffic - Unspecified	469,359	3.47
Infraction	1,149,085	8.49
Class Unspecified	256,171	1.89
Total	13,539,267	100.00

Overall Contact Rate with Criminal Legal System

Figure 5 shows overall contact rates by Race and Gender groups and Figures 6 through 9 show

identical presentations for Infractions, Traffic Violations, Misdemeanors, Felonies Class G, H,

and I, and Felonies Class A to F. A later section presents data on age groups. Note that Figure 9

does not present data on women because there are few women arrested for such high-level

charges. Each figure presents the predicted number of arrests for the demographic group in

question, holding constant the age, population size, and racial composition of the municipality of residence. Such a presentation allows a direct comparison of how arrest rates vary by levels of policing depending only on the demographic characteristics shown in the graph.

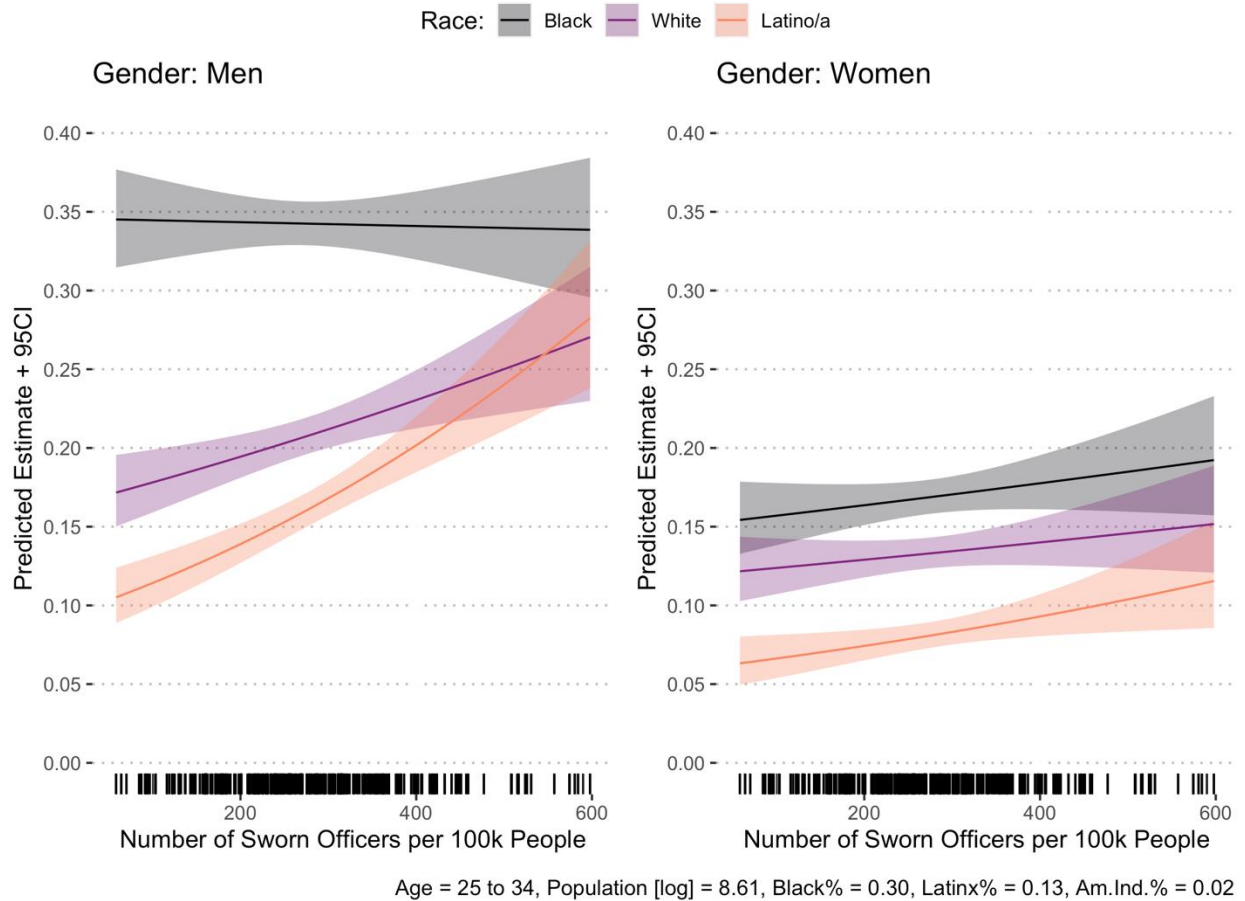Figure 5. Predicted Contact Rates by Race and Gender. All Crime Categories.

Figure 5 shows overall rates. Here, rates of arrest go sharply up for all groups except for black men, who have the highest rates overall.

Figures 6 through 9 break this down by the level of the crime. These figures clearly show that the overall rate is misleading, driven perhaps by the large numbers of traffic violations in our database. Looking separately at every type of charge except for traffic violations, black and Latinx men see sharply increasing arrest rates where the number of police officers per capita increases. The same is not true, however, for white men or for women, with few exceptions.
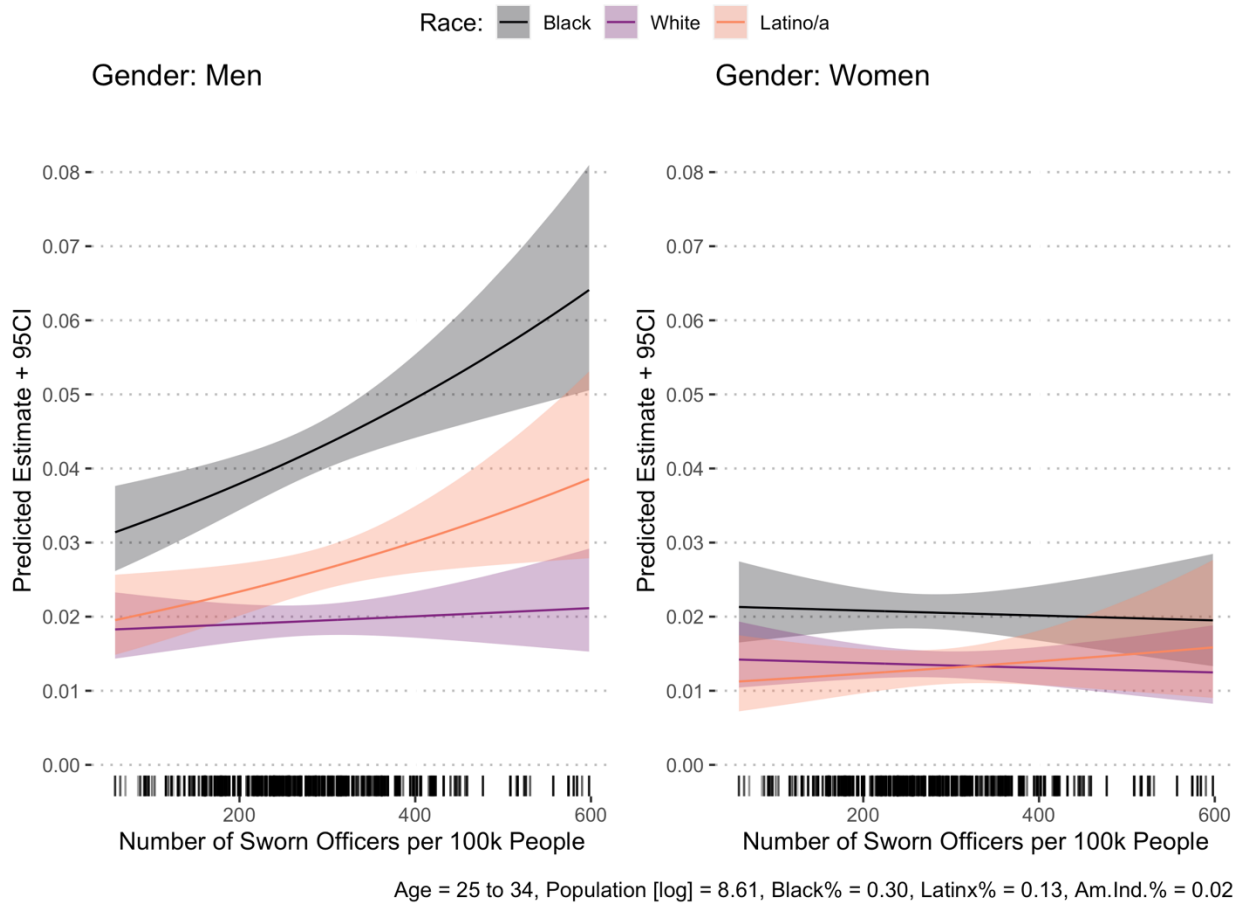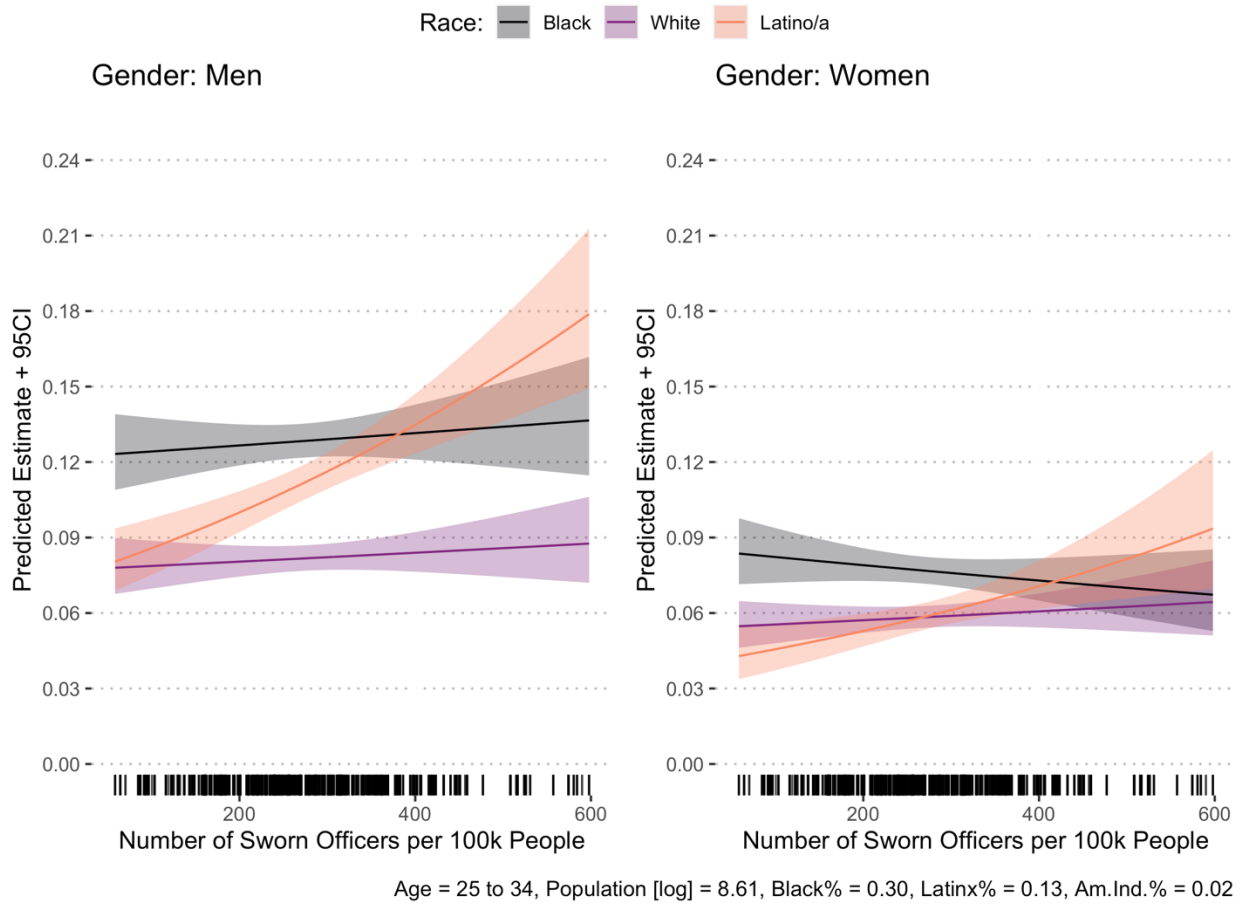
Figure 6. Infractions.

Age = 25 to 34, Population [log] = 8.61, Black% = 0.30, Latinx% = 0.13, Am.Ind.% = 0.02

Figure 7. Traffic Violations.

Race: Black ▬ White ▬ Latino/a ▬

Gender: Men

Gender: Women

Age = 25 to 34, Population [log] = 8.61, Black% = 0.30, Latinx% = 0.13, Am.Ind.% = 0.02

Figure 8. Misdemeanors.

Age = 25 to 34, Population [log] = 8.61, Black% = 0.30, Latinx% = 0.13, Am.Ind.% = 0.02

Figure 9. Felony Classes G, H, I.

Age = 25 to 34, Population [log] = 8.61, Black% = 0.30, Latinx% = 0.13, Am.Ind.% = 0.02
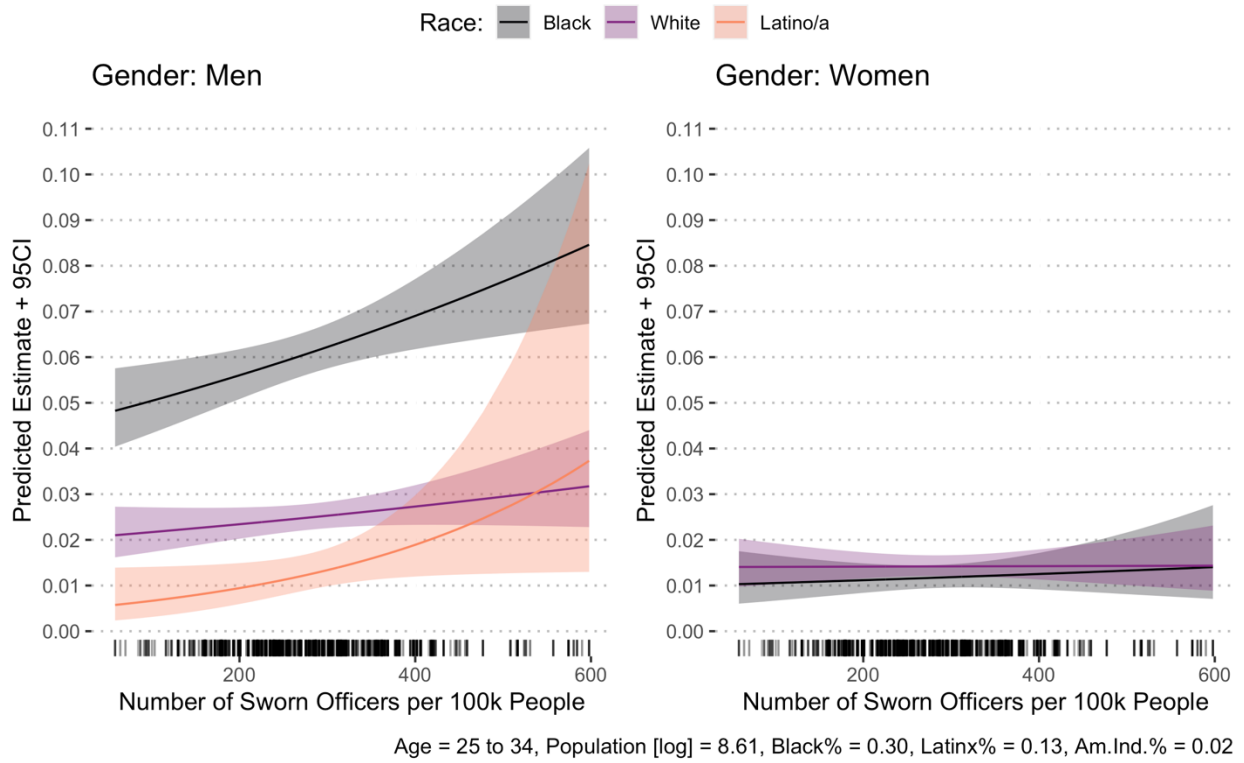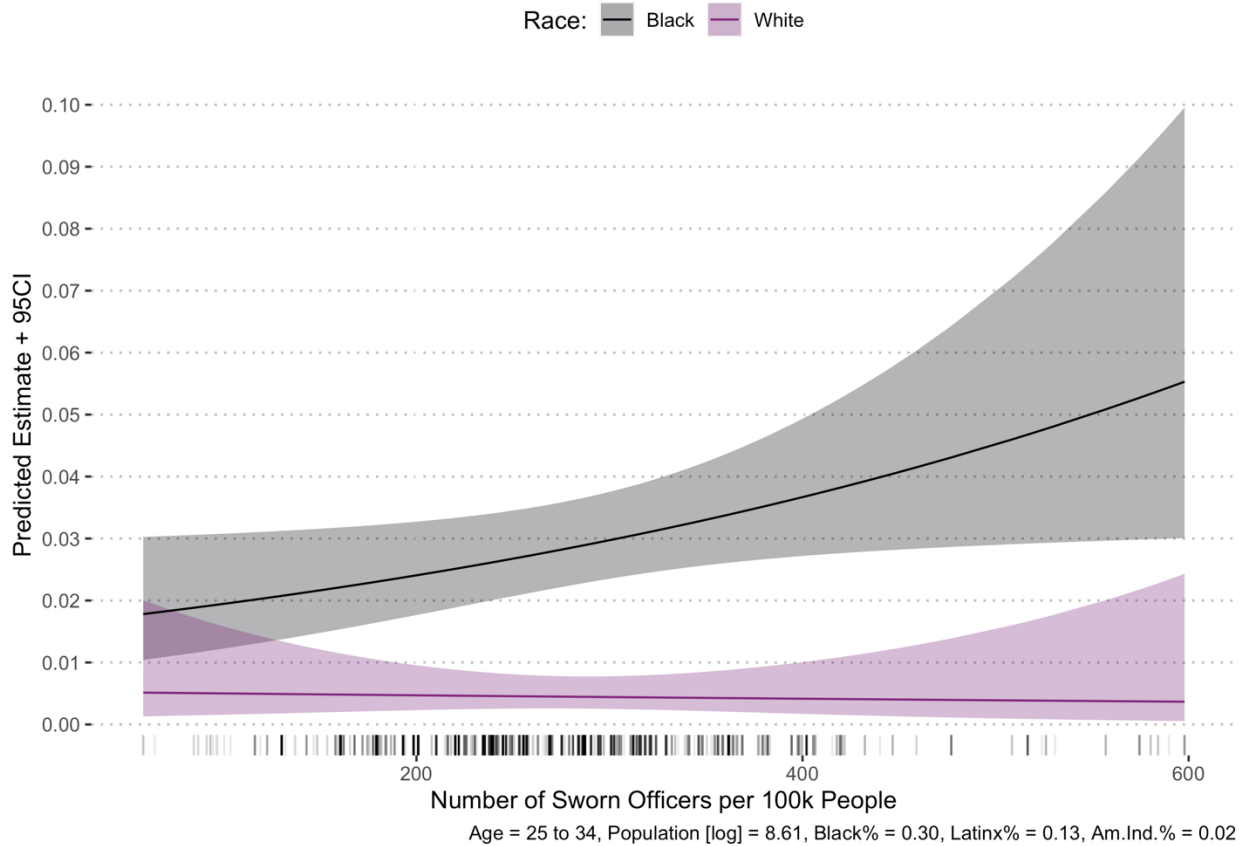
Figure 10. Felony Classes A-F.
        Gender: Men

Note: Women and Latinx men excluded because of low numbers.

By Age Group

Figures 11 through 16 replicate the analysis above for those in different age groups, limiting here

the analysis only to men. A careful review of these figures shows that the slopes (predicted rates

of arrest) generally rise as communities have more police officers per capita, but only among

those under the age of 44. For those 45 and older, as for whites and women in the previous

section, more policing has no impact on arrest rates.
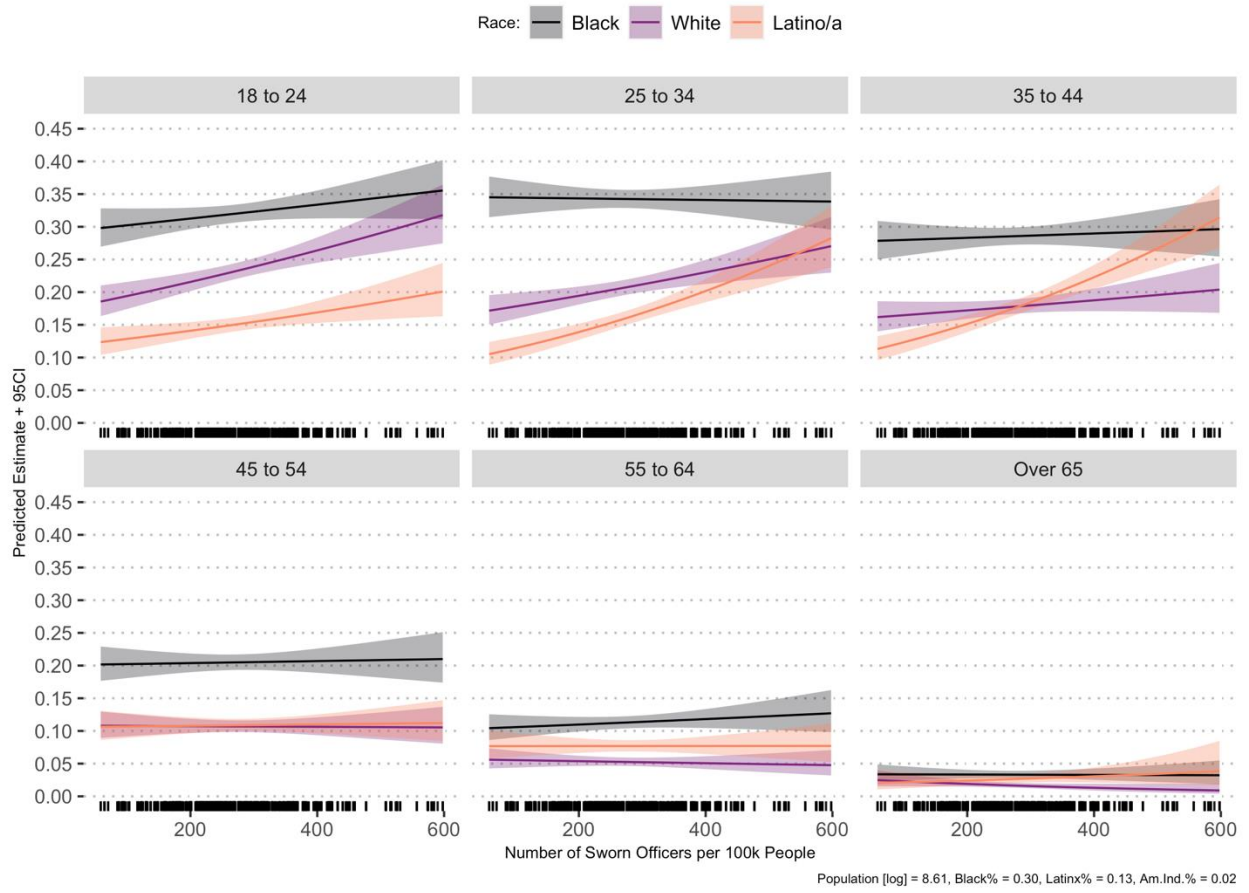
Figure 11. All Crime Categories.
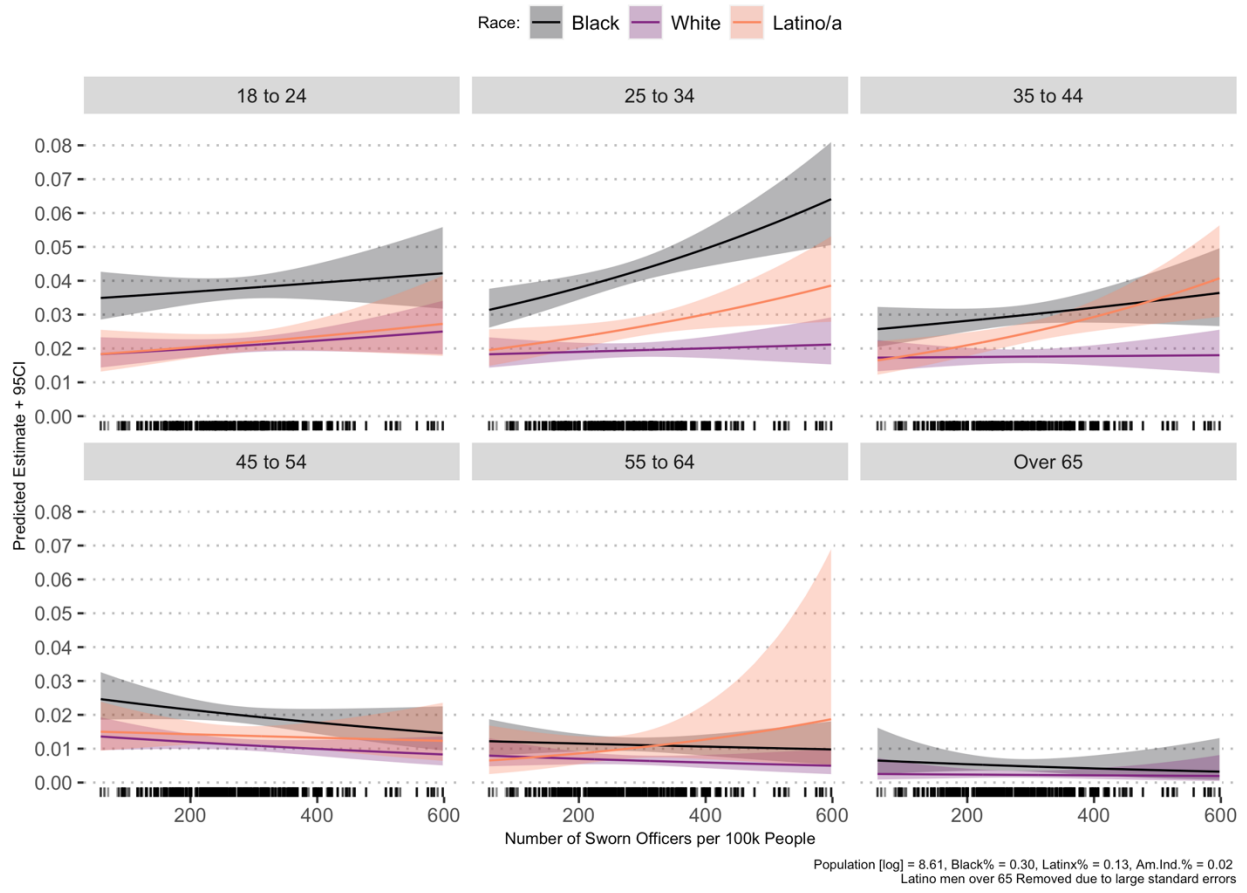
Figure 12. Infractions.

Population [log] = 8.61, Black% = 0.30, Latinx% = 0.13, Am.Ind.% = 0.02
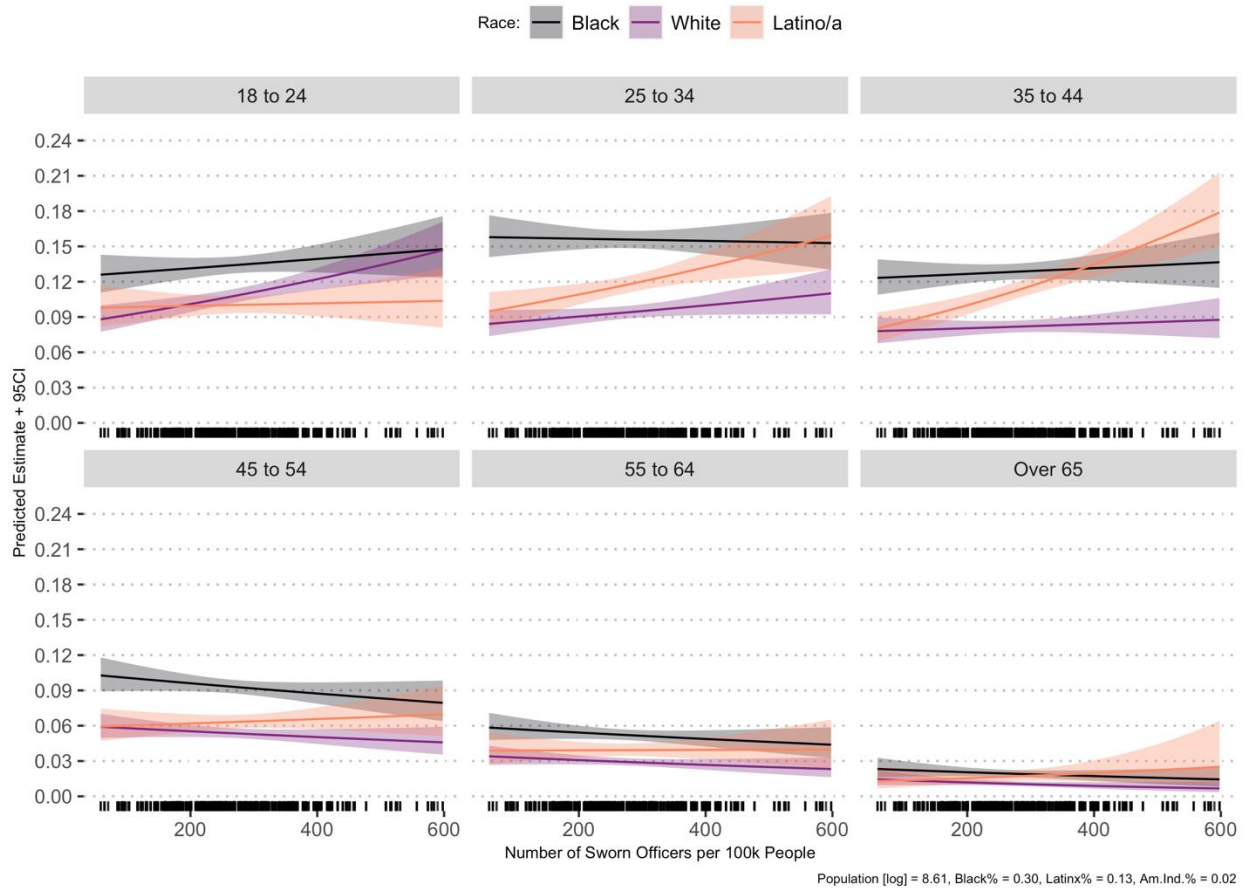Latino men over 65 Removed due to large standard errors

Figure 13. Traffic Violations.

Figure 14. Misdemeanors.
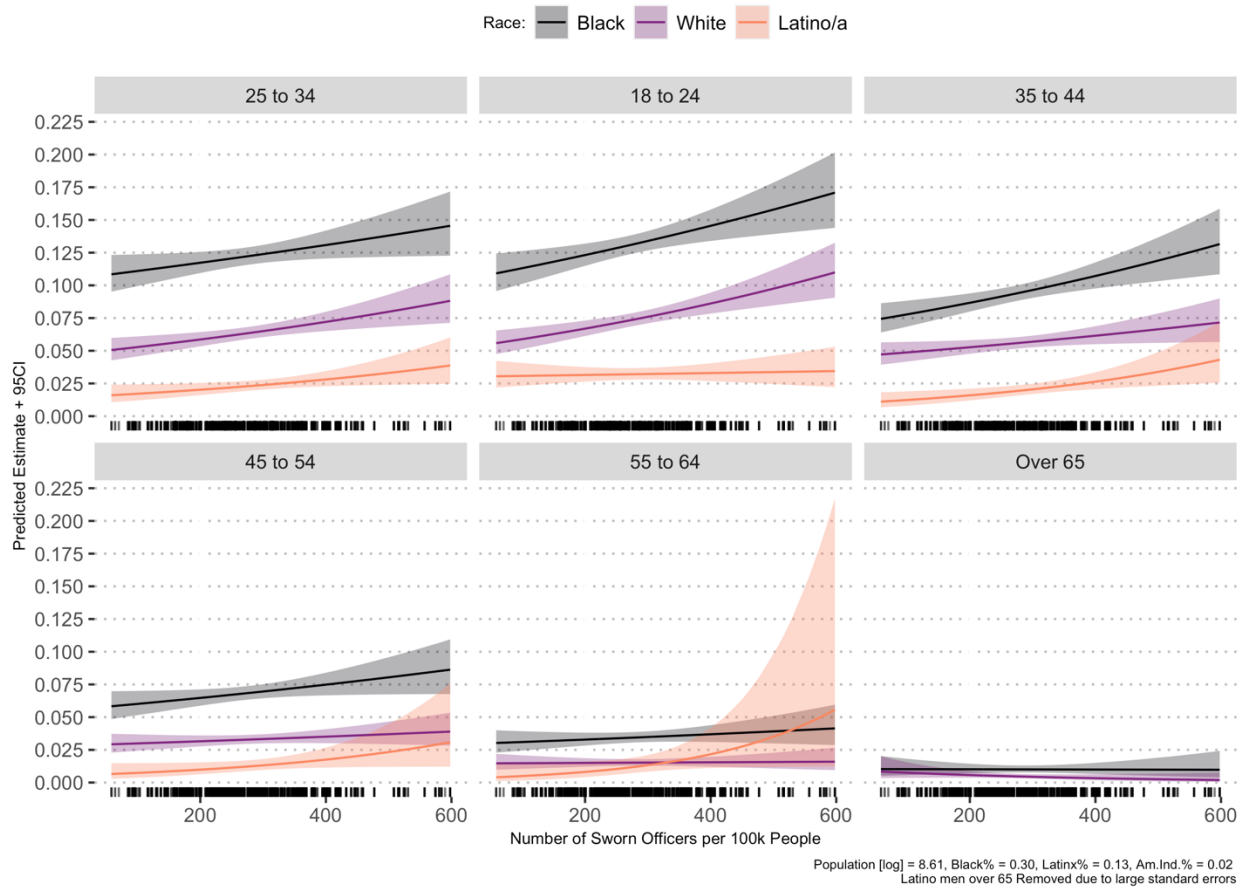
Figure 15. Felonies Class G, H, and I.

Race: Black White Latino/a

Population [log] = 8.61, Black% = 0.30, Latinx% = 0.13, Am.Ind.% = 0.02
Latino men over 65 Removed due to large standard errors

Figure 16. Felonies Class A to F

Race: Black White

Population [log] = 8.61, Black% = 0.30, Latinx% = 0.13, Am.Ind.% = 0.02

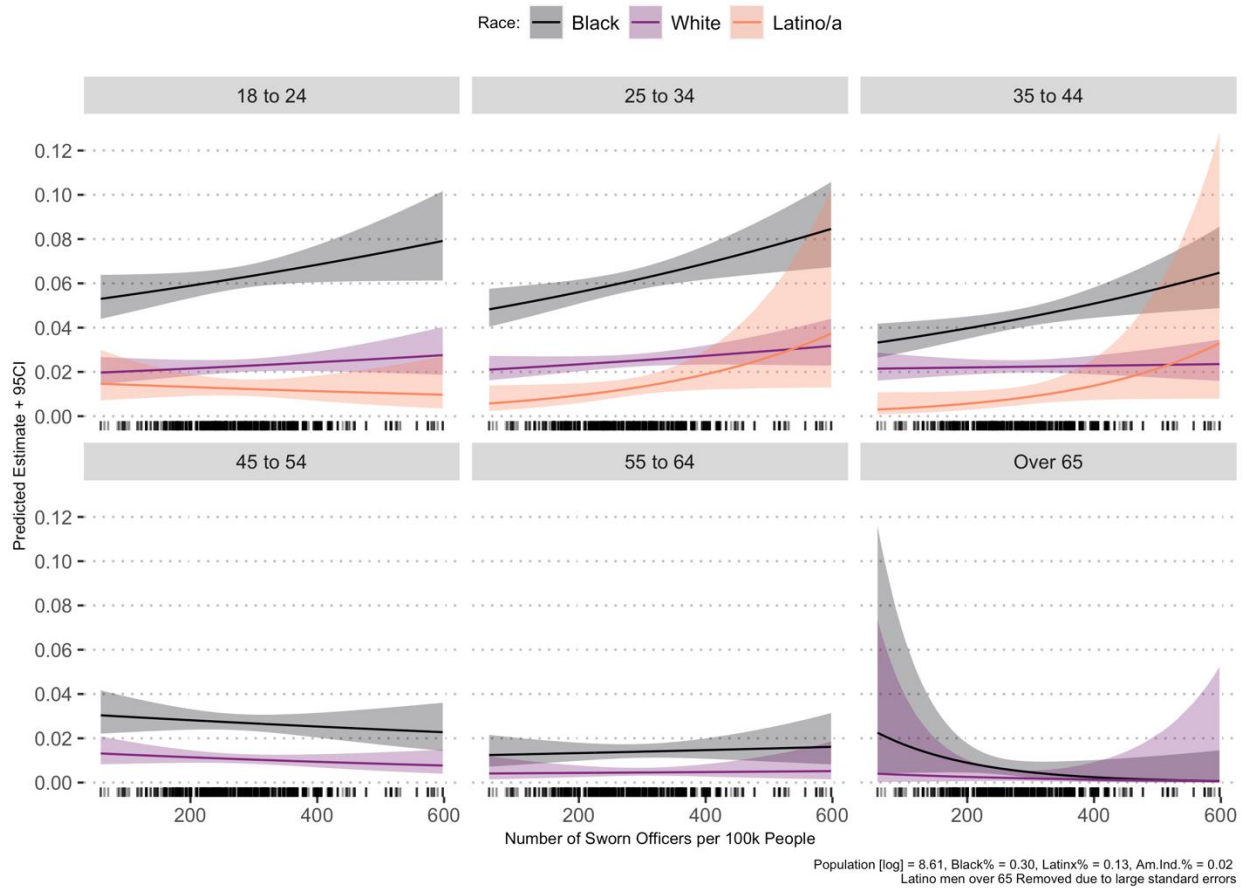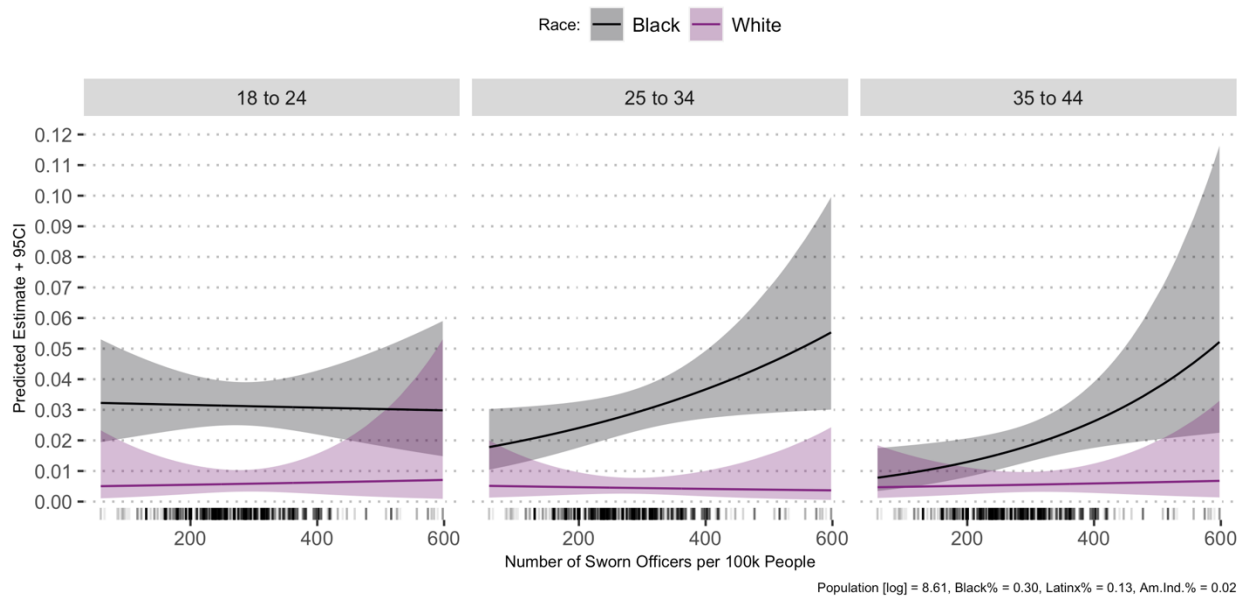## Discussion and Conclusion

We are taking a first step in this paper in answering a fundamental question in the study of social identity and the criminal justice system. Are arrest rates driven by surveillance or by criminal behaviors? We have not reached the end of this analysis, but we have taken a first step by demonstrating dramatic differences in arrest rates, and a targeting of these arrests, when communities invest in more police officers per capita.

References

Appendix A: Constructing Contact Scores for NC Demographic Groups

We explained in the main text that we generated 50,341 unique demographic groups by combining US Census categories on race, gender, age cohort, and geography (measured by US Census place designations).

Table A-1 provides a sample of the demographic subgroup permutations along with their ACS estimates and upper estimates.

Table A-1. Selected Location, Race, Gender, Age, and Census Population Estimates

Location	Place GEOID	Race	Gender	Age Cohort	ACS Estimate	ACS Estimate (+ 95 MOE)
Greensboro	3728000	Latino/a	Women	45 to 54	1,061	1,257
Raleigh	3755000	Black	Women	Over 65	6,519	7,381
Wilmington	3774440	White	Men	45 to 54	5,064	5,440
Winston-Salem	3775000	Native	Men	45 to 54	71	126
Charlotte	3712000	Latino/a	Men	55 to 64	3,209	3,392
Greensboro	3728000	White	Men	45 to 54	8,072	8,673
Charlotte	3712000	Latino/a	Men	45 to 54	6,673	6,927
Fayetteville	3722920	Black	Women	25 to 34	7,657	8,108
Raleigh	3755000	Asian	Men	35 to 44	1,669	1,979
Greensboro	3728000	Black	Women	25 to 34	10,666	11,525

Source: US Census, American Community Survey, 2018.

Using the numerator and denominator estimates for each demographic subgroup, we calculate measures of contact with the criminal legal system for each demographic subgroup as well as uncertainty measures. We achieve this by specifying a random intercept binomial model where the number of trials equals a demographic group's population (ACS estimate + 95 MOE) and the number of successes as the number of unique individuals who had contact with the criminal legal system in a given year. Figure A-1 illustrates the experimental design for the random intercept model, which includes both crossed random effects (Year) and nested random effects (race, gender, age cohort subgroup nested inside geographic location).

Figure A-1. A Random Intercept Model for Estimating Rates of Contact with the Criminal Justice System
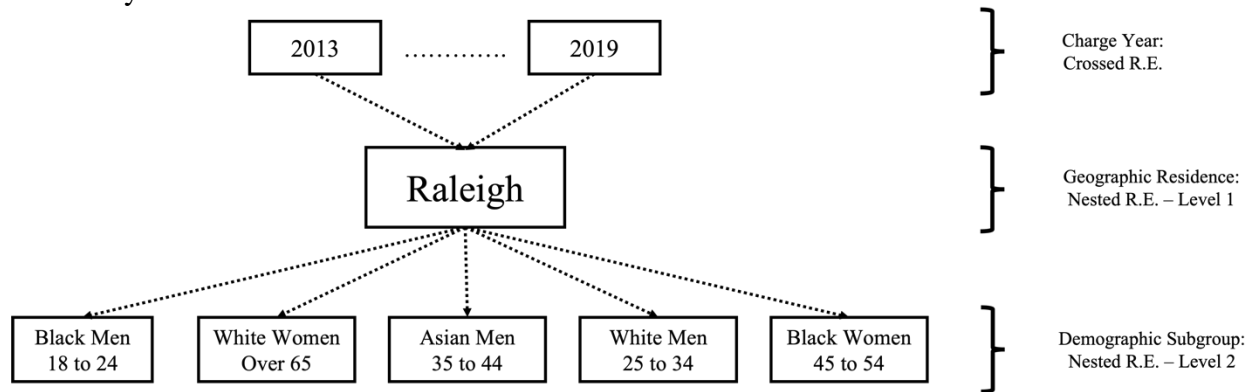
In using this experimental design, we can estimate the proportion of individuals who have contact for each demographic subgroup and year. In addition, we can difference out year-to-year random effects to better understand the degree of consistency in contact with the criminal legal system for various subgroups over time. One benefit of random intercept models over simple ratios involves the shrinkage of random effect estimates towards the mean of the distribution. This increases the precision of estimates of demographic groups with low numbers of documented contacts.

To estimate this random intercept binomial model, we adopt a Bayesian modelling approach based on INLA estimation strategy (Martino and Riebler, 2019). With this Bayesian

approach, we can provide a range of plausible estimates by summarizing the posterior

distribution for each demographic subgroup. In addition, we can use the variance of the posterior

distributions as a proxy for contact uncertainty. To estimate the proportion of individuals who

have contact, we use a Binomial likelihood (Equation 1). In this equation, *y* is the number of

individuals who had contact, *n* is the population total for a demographic subgroup, and *p* is the

proportion of individuals who have contact, which is latent.

Equation 1.

$$y_i \sim binomial(n_i,\ p_i)$$

$$p = \binom{n}{y} p^y (1-p)^{n-y}$$

When estimating contact rates, we draw inferences about *p* using the random intercept

experimental design. Equation 2 summarizes our model specification for the random intercept

model along with chosen priors.[3] In this equation, B is the grand intercept of the model, Ψ

represents yearly cross random effects, and Φ represents the nested random effects between

geography, gender, age, and race. In this model specification, we can provide a range of

plausible estimates of contact rates for each demographic subgroup. In addition, we can

marginalize out the yearly cross random effects Ψ to generate average contact rates between

2013 and 2019.

Equation 2.

$$logit(p_{ij}) = \text{B} + \Psi_i + \Phi_j$$

$$\Psi \sim \text{Normal}(0, \tau^{-1})$$

---

[3] We assume Ψ and Φ are drawn from a normal distribution with mean of 0 and standard deviation of τ^{-1}. We assign a non-informative penalized complexity prior to τ, which allows the chosen likelihood to dominate the posterior distribution rather than our chosen prior. This penalized complexity prior makes probabilistic inferences about the precision term $\tau - p(\tau > d) = \alpha$. Here we assume that the probability τ is greater than $3 * \sigma_r$ is less than 1 percent, where σ_r is the standard deviation of the residuals from a non-hierarchical intercept model (i.e., y ~ 1).

$$\Phi \sim Normal(0, \tau^{-1})$$

$$\tau \sim pc.prior(3 * \sigma_r, 0.01)$$

Appendix B: Enforcement Capacity across North Carolina Jurisdictions

What does enforcement capacity look like across North Carolina? When discussions about size of policing emerge in the media, large jurisdictions generally occupy the center of attention. These jurisdictions often possess large squadrons of sworn officers that receive an outsized share of scrutiny. Using information from the Universal Crime Reporting (UCR) database[4], Table B-1 summarizes information on the average number of officers between 2013and 2019 for North Carolina's top 20 jurisdictions based on population.

Table B-1. Demographic Characteristics of the Largest NC Municipalities.

Jurisdiction	Total Number of Residents (18 and Older)	Total Number of Defendants	Proportion Black	Proportion White	Proportion Hispanic	Officers per 100k
Charlotte	841,727	557,076	35.4	42.2	14.2	245
Raleigh	457,569	309,259	29.8	54.1	11.5	161
Greensboro	288,802	251,771	42.4	44.0	8.1	239
Durham	264,405	189,807	40.0	39.3	14.4	193
Winston-Salem	242,215	296,470	35.4	46.6	15.2	235
Fayetteville	210,031	181,382	43.0	38.7	12.5	199
Cary	163,678	50,188	9.1	64.5	8.7	130
Wilmington	118,172	115,094	19.5	72.8	6.7	287
High Point	111,184	107,475	35.9	45.7	11.5	220
Greenville	91,267	81,948	39.7	52.7	6.1	222
Asheville	90,597	76,345	12.8	79.5	7.1	402
Concord	89,813	60,391	22.9	62.1	13.4	201
Gastonia	75,223	84,393	29.5	59.4	10.3	235

---

[4] Due to the volunteer nature of the reporting system for this database, we ignore individuals who reside in communities that did not report numbers on the total number of sworn officers. These individuals account for 6.2 percent of defendants in our database. In addition, we ignore individuals who live in unincorporated communities where local sheriffs replace the role of community police officers. These individuals account for 36.2 percent of defendants in our database.

Jacksonville	73,285	49,370	19.9	59.6	17.3	161
Chapel Hill	59,609	21,593	11.4	69.9	7.7	179
Rocky Mount	55,034	61,142	66.1	29.9	4.7	286
Huntersville	54,619	18,599	13.2	78.1	8.5	184
Burlington	52,934	51,610	29.5	50.9	20.0	239
Wilson	49,300	53,989	49.6	39.7	11.4	246
Apex	48,296	17,246	9.0	77.4	8.6	158

When comparing these sworn officer numbers to the total number of residents who live there, North Carolina's top 20 jurisdictions often possess middling rankings in the per capita number of sworn officers. As Figure B-1 illustrates, there is a negative correlational relationship between the total number of residents adjusted for seasonal population and the number of sworn officers per 100,000 residents[5]. Small jurisdictions with large tourist populations generate this negative relationship as tourist hubs tend to have higher policing needs than their year-round residential population would otherwise suggest.

Figure B-1. Officers per capita and seasonal residents.

---
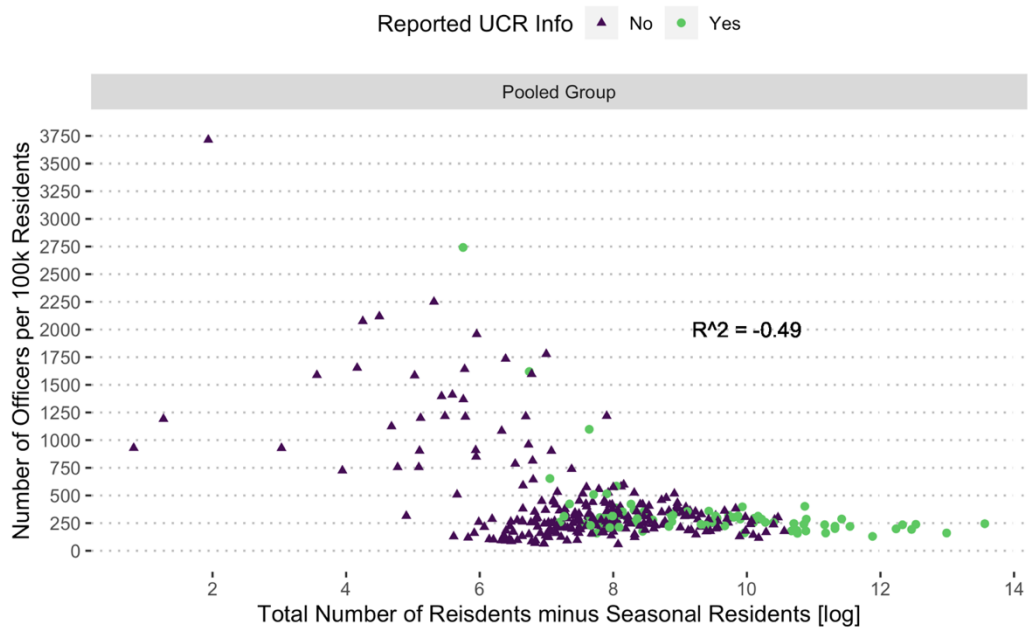
[5] Also adjusted for the seasonal population

To prevent these communities from skewing our analysis, we conduct a simple k-means clustering process that separates the scattered point locations into three distinct groups. Figure B-2 illustrates the results of this clustering process along with the correlations between the total number of residents adjusted for seasonal population and the number of sworn officers per 100,000 residents.

Figure B-2. Three Clusters Identified.
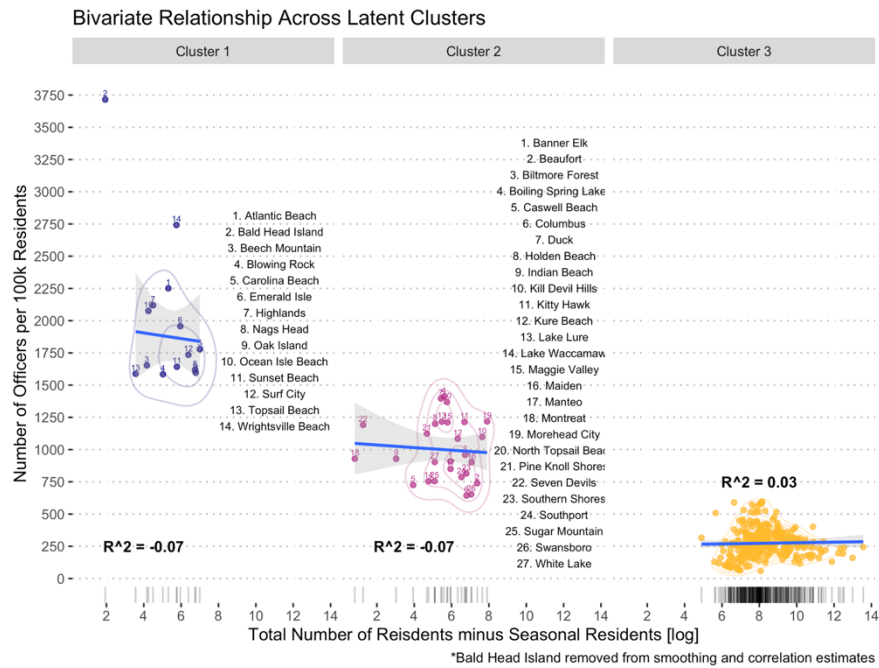
Accounting for these clusters reveals a lack of correlation between the per capita number of sworn officers and the total number of residents. As expected, the first and second clusters possess some of North Carolina's most popular summer tourist destinations, such as Atlantic Beach, Morehead City, North Topsail Beach, Surf City, and Bald Head Island. Table B-2 breaks down the demographic information for these three clusters. Although the first and second clusters possess per capita figures of 2,004 and 1,000 (for 100,000 residents) respectively, these communities account for less than 1.5 percent of defendants in the North Carolina criminal legal system. In addition, they possess an overwhelmingly white population.

Table B-2. Demographic Characteristics of Three Clusters Identified.

Cluster	Total Number of Residents (18 and Older)	Seasonal Residents (Proportion)	Proportion Black	Proportion White	Proportion Hispanic	Total Number of Defendants	Officers per 100k
1	37,813	86.6	1.5	96.4	4.8	21,547	2004
2	67,276	75.7	7.6	89.9	9.6	42,325	1000
3	5,381,697	9.4	30.8	55.3	12.3	4,216,568	273

To properly investigate the differential policing hypothesis, we focus our effort on the communities found in the third cluster, which account for over 98.5 percent of defendants in our database.