Citizens for Juvenile Justice: Deliverable 3

Written By Alex Blumenfeld, Raphael Bruni, and Dan Katz

Overview

The primary objective for the Citizens for Juvenile Justice (CfJJ) project this semester is to understand better the impact of police policy and practice in select cities of Massachusetts on the citizens living in those cities. Up to the current moment in time, all of the questions outlined in Deliverable 0 (namely, an analysis that pertains to gaining a more complete understanding of incident data by factors such as the race, sex, and age of civilains, as well as locations of arrest) have been answered using data from the cities of New Bedford and Springfield. Similar work will be done on the city of Haverhill. Still, the analyses that have been completed thus far regarding New Bedford and Springfield provide a good outline of any future explorations. As meetings with the lead client from CfJJ have progressed throughout the semester, supplemental inquiries have been added to the data science undertaking. Said features in the data have been categorized by the race and sex of the involved civilians and, moreover, certain age categories, such as young people aged 18 through 20, have been formed to determine if certain groups of people are disproportionately represented in the data, and to understand the impact of police practice on young people.

The first dataset that was analyzed describes all of the "field incident reports" filed by New Bedford police officers from 2015 up through a few months ago, which are records of when a police officer stops someone on the street because they're suspicious that that person may be committing a crime, more commonly referred to as a stop-and-frisk. We have been investigating connections between several important variables included in these police reports, such as the location of incidents, age of civilians involved, and proximity to points of interest which we

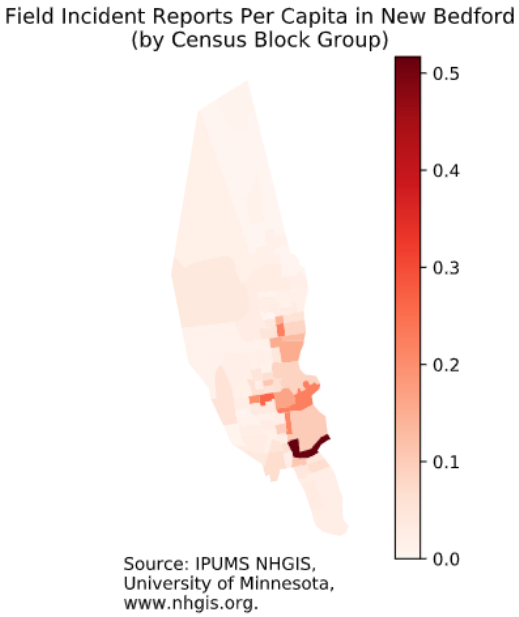
believe may be related to why officers are stopping civilians. Below is a simple map of the city (divided into census block groups) with a plotting of each of the 4,997 police incidents that occurred from 2015 through the first few months of 2020:

There are many permutations of this data which can be used to create different visualizations, and the following maps focus on age to investigate where youths are having the most interactions with police in New Bedford -- of course, this is only measuring impromptu stops initiated by officers, not arrests or police reports of actual crimes. In addition, to get an idea of whether people in the 19-20 age range are being stopped in different locations than other young people, we produced one map including them and one map without them.

Locations of Field Incident Reports Involving Persons Age 20 or Younger

Locations of Field Incident Reports Involving Persons Age 18 or Younger

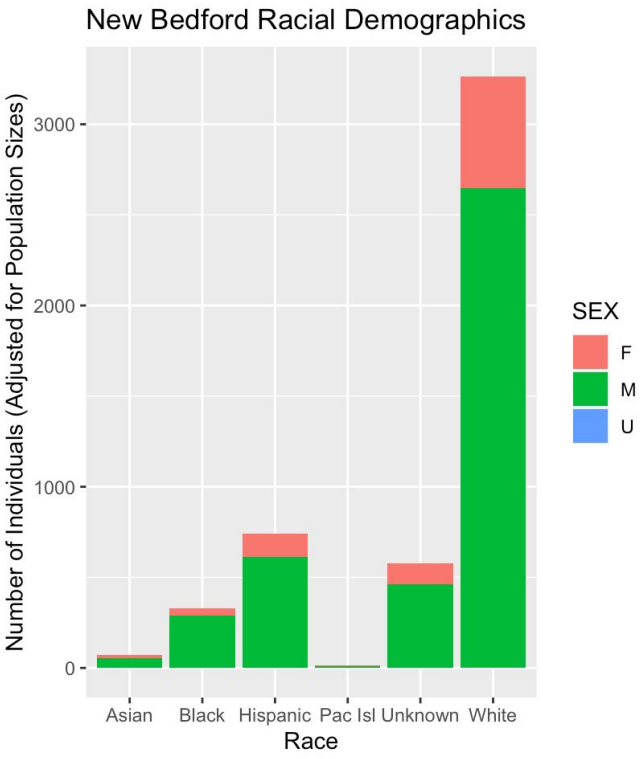
The next map resulted from an attempt to control for the populations of each census block group in New Bedford, showing how frequently field incident reports were filed relative to the total population in that area. The scale on the right represents incidents per capita: so, from 2015 through early 2020, the one neighborhood which is particularly dark red featured approximately one field incident report filed for every two people living in that area. That one ‘block group’ also had by far the most incidents in the entire city in absolute terms, showing that New Bedford police pay particular attention to people passing through that area.

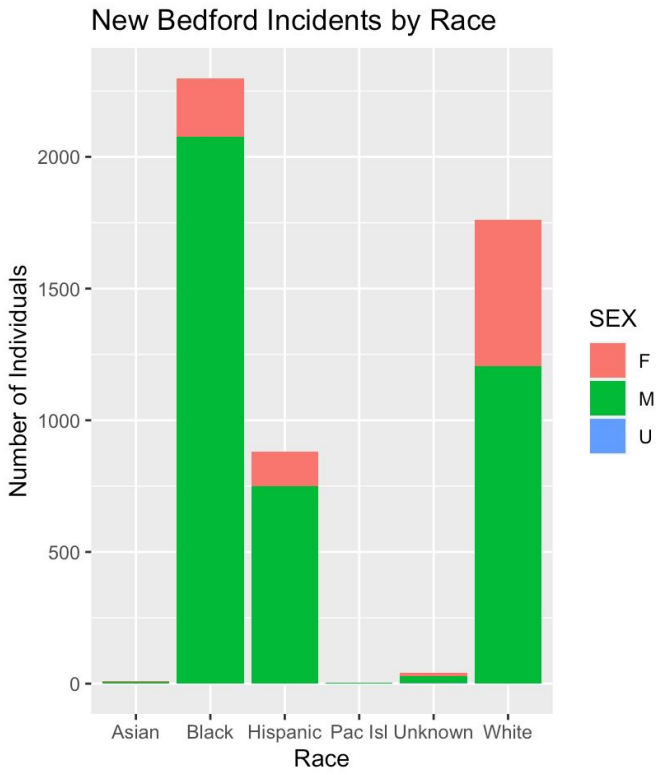


Race Equity

One of the essential requests by the client concerns the race of the individuals involved in incidents. Our analysis of the New Bedford data found that certain racial/ethnic groups (specifically, Blacks/African Americans and Hispanics) are overrepresented based on their respective population sizes. Despite making up under 7% of the New Bedford population, African Americans represented greater than 46% of the unique individuals (2,299 different people) involved in police incidents from 2015 through 2020. For context, African Americans are nearly 13 times as likely to be stopped by a police officer than their white counterparts. Hispanics are only slightly overrepresented, making up nearly 18% of incidents (881 people) despite accounting for less than 15% of the population. Hispanics are just over two times more likely to be involved in police-related incidents than whites. To better understand these differences in representation, consider the two bar charts below, one of which summarizes

demographic data in the city of New Bedford, and the other of which shows the incident data by race:





In addition to the overrepresentation of African Americans and Hispanics in the dataset, some noticeable differences exist between the two charts. For one, whites make up a substantially smaller proportion of the data based on their population of the city. In addition, races that are considered "unknown" or "other" are hardly distinguishable in the dataset. Perhaps this distinction is an implication of the bias on the part of police officers' to detect certain races of people, or it may be a reflection of data quality on the part of the police department.

Prolific Officers

To gain a deeper sense of specific police officers who engage in many of these incidents, an analysis of the most prolific officers was conducted. The table on the next page includes some statistics concerning the ten most active officers in New Bedford over the past half-decade.

Most Prolific New Bedford Police Officers

Officer	Identification	Number of Incidents	Salary (2017)	Percentage of Incidents Involving Black Civilians	Percentage of Incidents involving Hispanic Civilians	Percentage of Incidents Involving Black or Hispanic Civilians
Roberto DaCunha	4B-3974	459	$89,945	65.58%	18.30%	83.88%
Brian Rei	5B-4020	305	$96,436	62.30%	16.72%	79.02%
Lorenzo Gonzalez	3B-3987	299	$117,832	58.19%	24.41%	82.60%
Gene Fortes	2C-3875	266	$86,553	60.15%	16.92%	77.07%
Clint Medas	2C-4001	241	$27,106	62.66%	21.58%	84.24%
Jorge Santos	3C-4002	194	$102,620	21.13%	17.53%	38.66%
Pedro Moco	6A-4042	161	$113,825	50.93%	21.74%	72.67%
Samuel Algarin-Mojica	2C-4045	148	$61,359	52.03%	26.35%	78.38%
Jason Orlando	1A-4011	120	$105,699	45.83%	13.33%	59.16%
Nathaniel Goncalo	2A-4073	96	$36,397	48.96%	14.58%	63.54%

The ten officers in the above data frame were involved in nearly 46% of all incidents, a tremendous percentage given that 186 different officers were involved in at least one incident during the time span. Moreover, the fraction of incidents that involved Black civilians was higher than the general African American makeup of the dataset (about 46%) for eight of these ten officers. The data frame also allows for us to compare the salaries of these prolific officers to the average salary of New Bedford police officers and city employees. Indeed, the mean salary for these officers was about $83,777 from the years 2016 through 2018, whereas the typical salary for a city police patrol officer as of September 2020 was $60,458, and the average salary for a city employee in New Bedford was $49,841. The relatively high salaries for these employees compared to other city employees may be an indication of the amount of work that they do (as represented by the number of incidents they are each involved in), their length of time on the force, and possibly the use of overtime. These are all possible areas of follow up research.

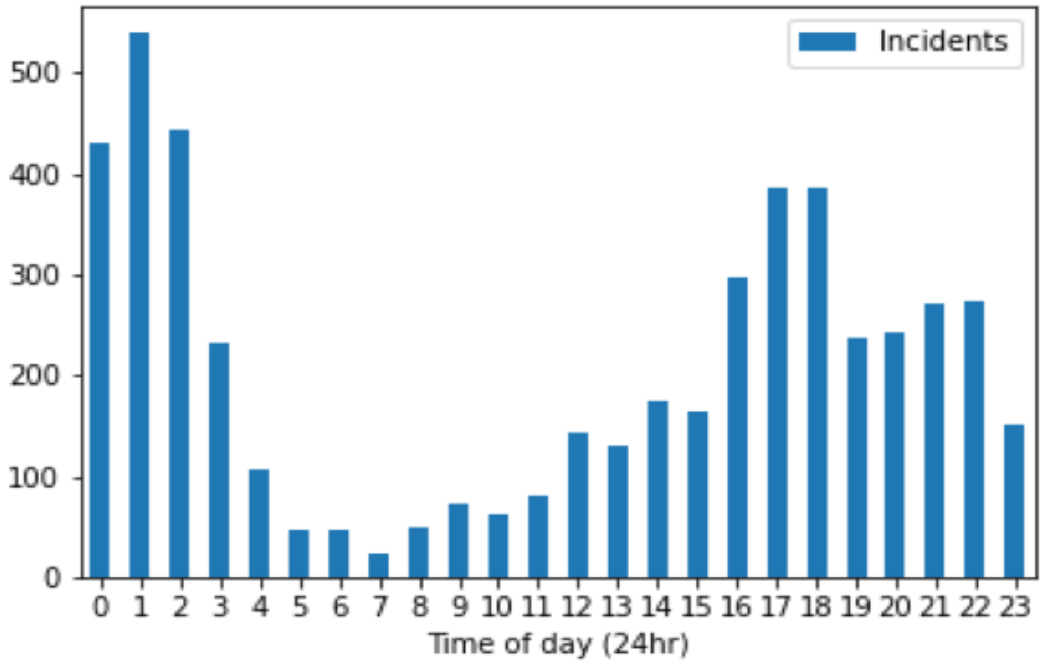
Geographic Analysis

It can also be interesting to visualize where many of these officers conduct their incidents. For instance, a visualization of incidents by the most prolific officer (Officer DaCunha) is presented below:

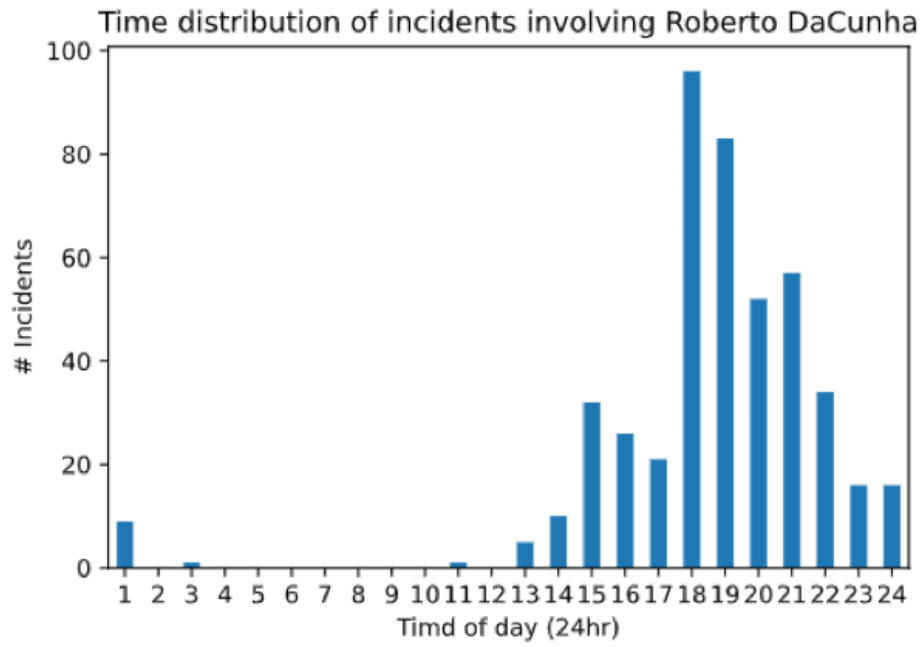
The map for Officer DaCunha is similar to that for many of the other officers in that three noticeable clusters of incidents appear: one near the southern part of the city, another near the eastern section of the city, and another in the north. These regions could be highly populated, or perhaps a disproportionate amount of crime adjusted for population occurs in these areas, and thus police officers are more prone to stop individuals who appear in these neighborhoods.

Time of Day

We were asked to analyze the New Bedford incidents by the time of day in which they occurred. Below is a distribution of the accumulated number of incidents by time of day:



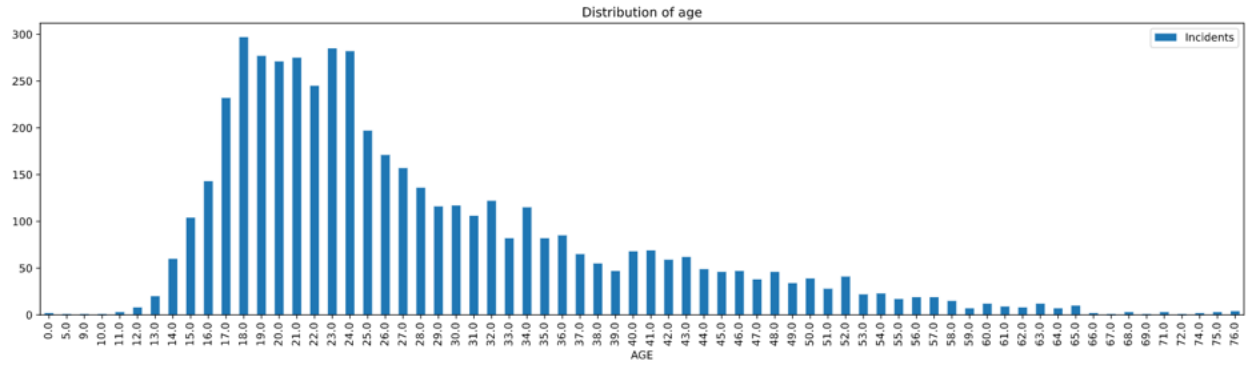
From this display it is clear that incidents are occurring with more frequency towards the night time, with a peak at around 1 AM and a low at 7 AM. This skew could indicate that more police officers are patrolling at night, it could be a sign that more suspicious behavior is occurring at those times, or of police practice stopping vehicles out at night. Another interesting observation would be to see what the incident time of day distribution looks like for the most prolific officers. Below is an analysis for Officer DaCunha, the officer with the most stops:



As can be seen in the above graph, Officer DaCunha makes many of his stops at night, with a peak occurring at 6 PM. This could be an indication of which times officer DaCunha is patrolling.

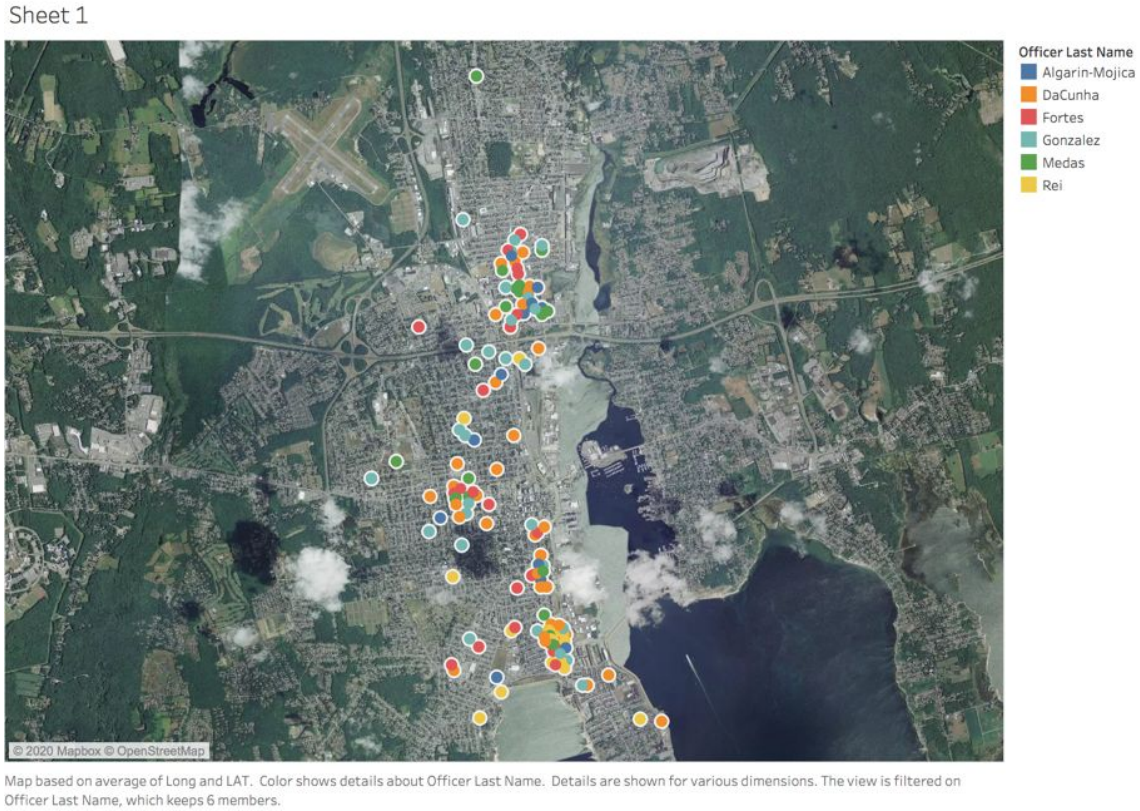
Age

In addition to an analysis of race, we are also determined to specifically analyze incidents by certain age groups. Before grouping people by age, below is presented the age distribution for incidents:



It is clear that incidents occur with a much higher frequency from ages 18 to 24. Specifically, CfJJ is interested in examining three separate categories for young people: those aged 10 through 17, those aged 18 through 20, and those aged 21 through 25. Filtering by the middle of these age groups (18 through 20) would permit one to find that people of these three ages (18, 19, and 20) make up a disproportionately high number of incidents in the dataset. Indeed, nearly 17% of total incidents involved individuals of these three ages.

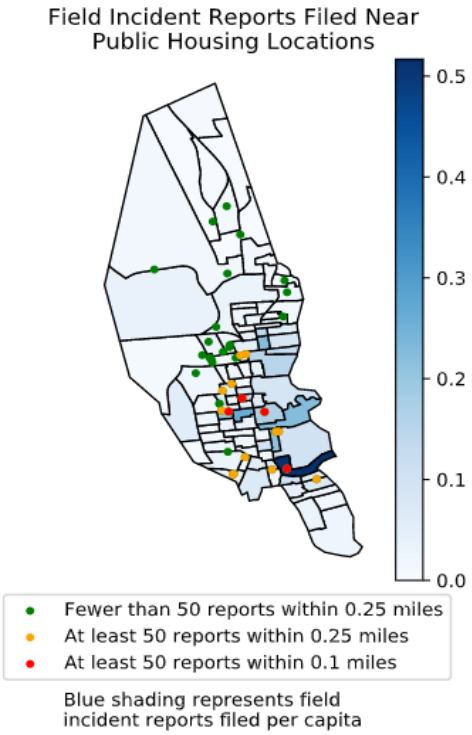
CfJJ is also motivated to understand which officers are involved in many incidents that regard young people. An analysis of the data would find that five officers (Officers Algarin-Mojica, DaCunha, Fortes, Gonzalez, Medas, and Rei) reported on more than 47% of all incidents concerning the 18-20 age group. A visualization of these six officers' work with regards to young adults is shown below:



As was the case for Officer DaCunha, three noticeable clusters appear in the southern, eastern, and northern parts of the city.

Points of Interest (Public Housing and Public Schools)

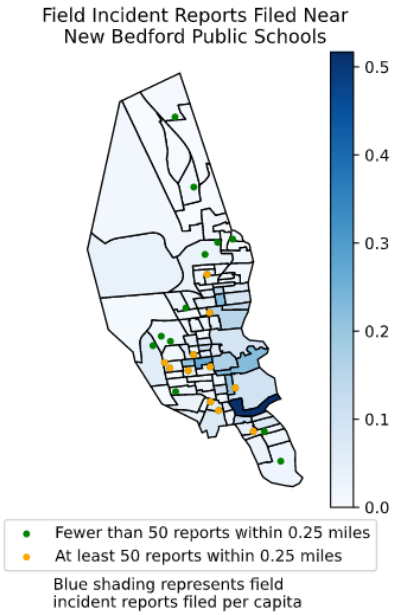
We also began investigating the relationship between locations of field incident reports and locations of public housing properties in New Bedford. Plotting the coordinates of the properties in relation to the incidents could give us a better understanding of whether interactions between the police and citizens are more likely to occur near particular public housing areas.



Based on this map, it is apparent that four properties had particularly high numbers of incidents occurring very close to them, including three locations which are all quite close to one another. This plot does certainly serve as further proof that most of the police stops performed in New Bedford are occurring in the southern region of the city, but what makes the above map so interesting is that over the past five years, four public housing properties each had at least 50 incident reports filed within 0.1 mile of them, a radius of no more than a couple of blocks.

Although none of us are well versed in the geography of New Bedford, it seems reasonable to suspect that officers might focus their patrols on public housing properties and disproportionately perform stops near public housing as a result, and the above map provides some preliminary evidence that this could be happening.

After looking at whether police might be making lots of stops near public housing locations, we repeated the same exercise with all of New Bedford's K-12 public schools. We found that none of the city's public schools had quite as many incidents within a 0.1 mile radius as the top four public housing properties, but there were still several schools which had at least 50 incidents occur less than 0.5 mile away. The public schools followed the same general trend as public housing, in which locations closer to the center of the city had more incidents transpiring nearby than those which were farther away. This result doesn't exactly specify to us whether police officers are targeting the area around public schools when looking to conduct what are essentially stop-and-frisks, but it's still useful to understand the relationship between this type of law enforcement activity and New Bedford's education system.



Springfield Police

A few weeks into the project, we found out about another dataset of interest to CfJJ: a page on the Springfield, MA, Police Department's website which contains logs describing all of the arrests performed in the city each week, all in PDF format. Our objective was to compile all of these logs into a single CSV file or some other easily searchable format, and then investigate any patterns we could find in arrests of individuals aged 18 to 20. In particular, CfJJ was interested in which offenses young people are being arrested for the most often, the frequency with which minorities are getting arrested relative to their prevalence in Springfield's population, and some maps depicting the locations of these arrests.

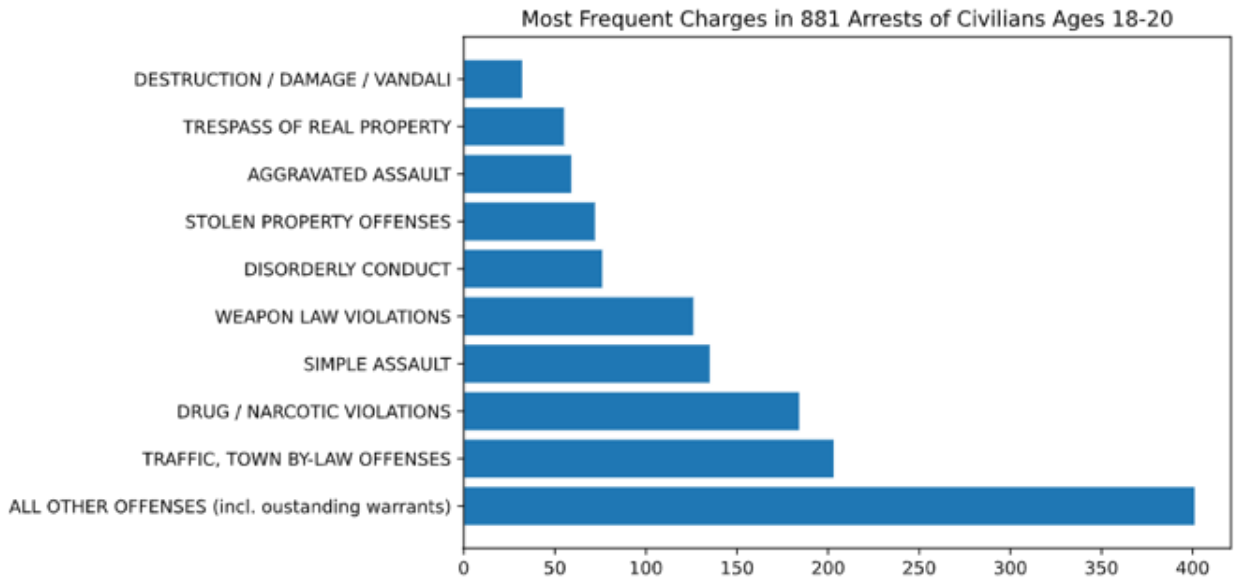
The Springfield arrest logs are easily downloadable as PDFs from the city's website, and their format is standardized enough that we could write one Python script that would download all of the files, extract the key fields from each one, and save the results to a CSV, with only a few incidents where certain blocks of text were assigned to the wrong fields. Some of the most important information in these reports includes the date of the arrest, the street address where the arrest was made, the date of birth of the suspect (from which we could calculate their age), and the list of crimes that the suspect was charged with upon being arrested. Here is an example of a portion of one arrest record after it has been retrieved from the PDFs:

Arrest Date/Time	Zone	Location	Offenses	Offense Codes	Offense Descriptions	Suspect Name	Suspect DOB	Suspect Address	Arrestee Age
2017-01-03 11:05:00	Sector H1 Forest Park	REAR OF HOME, 3 SUMNER AVE	DRUG, POSSESS TO DISTRIB CLASS A, SUBSQ.;DRUG ...	35A;35A;90Z	DRUG / NARCOTIC VIOLATIONS;DRUG / NARCOTIC VIO...	GONZALEZ HECTOR J JR	1978-06-08 00:00:00	431 RIVERSIDE RD, SPRINGFIELD, MA 01107	38

Because these PDFs do not include information about the race and ethnicity of each suspect, we need to merge the resulting CSV file with two years' worth of data that our client obtained from NIBRS (the FBI's National Incident-Based Reporting System), which is the national platform where police departments record all arrests that they make (and which lists the race of each

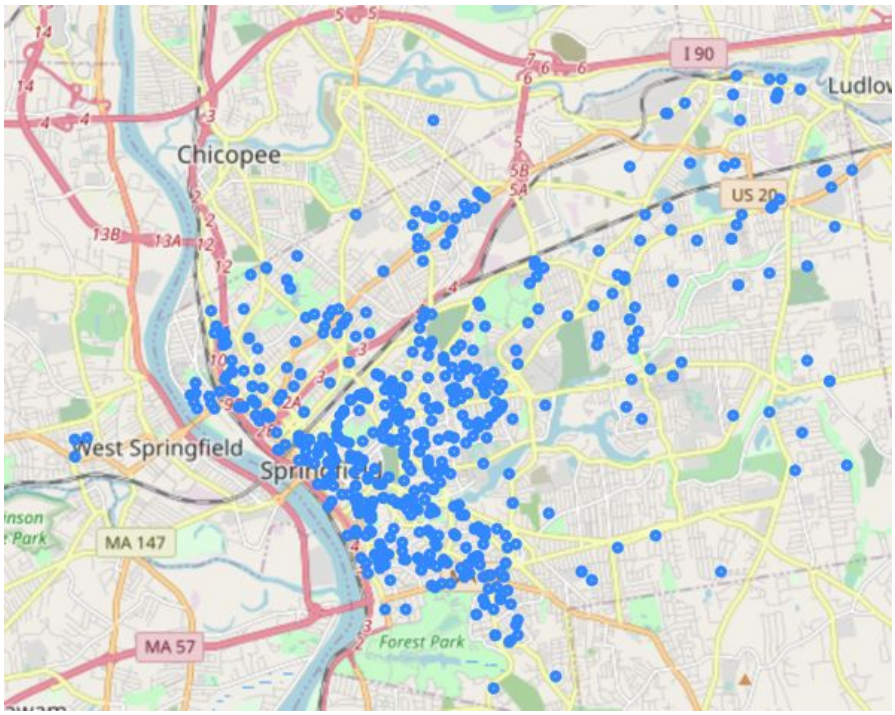
suspect). The arrest logs from the Springfield website don't match up with the NIBRS records one-to-one, and we are still working on merging them to the greatest extent possible in order to do an analysis which involves the race/ethnicity of the suspects.

The first important result from the Springfield dataset is the graph below, which depicts the offenses that 18-20 year-olds are being arrested for in the city. Each arrest can have several charges associated with it, so this visualization represents the number of arrests which included a particular charge, not the number of arrests which were exclusively for that one charge. The 'All Other Offenses' designation, which is by far the most common, is usually (about ⅔ of the time) meant to denote that a suspect already has a warrant outstanding for their arrest, as the existence of a warrant is listed as a separate charge in the Springfield police department's public logs.

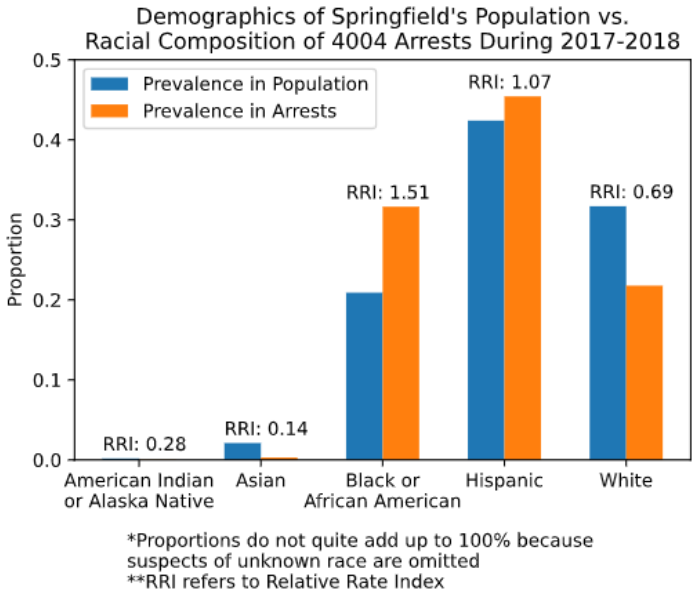


As for some basic mapping of the Springfield arrests which involved youths, the maps on the next page plots the locations of arrests in Springfield which involved a suspect who was between 18 and 20 years old. Because the street addresses recorded in the reports are imperfect, we were only able to get latitude and longitude information for about 85% of the 881 arrests in question,

and it is this portion of them that are displayed below. Just like in New Bedford, most of these arrests are occurring near the downtown area of the city.



Once we were able to merge the Springfield arrest logs with the NIBRS data, we found that only about 63% of the arrests from the Springfield PDFs (from the years 2017-2018 in particular) could be uniquely matched to a NIBRS record, but this still gave us a pool of about 4,000 arrests for which we knew the race and ethnicity of the suspect. At this point, the most useful direction is to calculate a "relative rate index" for each main racial group, which captures the extent to which a particular group is over- or under-represented in the arrest figures relative to its frequency in the overall population of Springfield. The result of this is shown below:



So, we can see that Black/African American residents are the most over-represented group in this collection of arrests, as they made up only 21% of Springfield's population in 2018, but accounted for roughly 32% of this subset of the arrests performed in 2017 and 2018. As it turned out, Hispanics, who are the single largest ethnic group in Springfield, appeared in the arrest logs at a rate roughly equivalent to their frequency in the population, telling us that most of the over-representation of Black suspects is compensated for by a relative under-representation of white suspects. This shows a much lower level of over-representation of Black residents than what occurred in the field incident reports in New Bedford, where the proportion of people being stopped who were Black was nearly 8 times the proportion of Black residents in the city's population.