

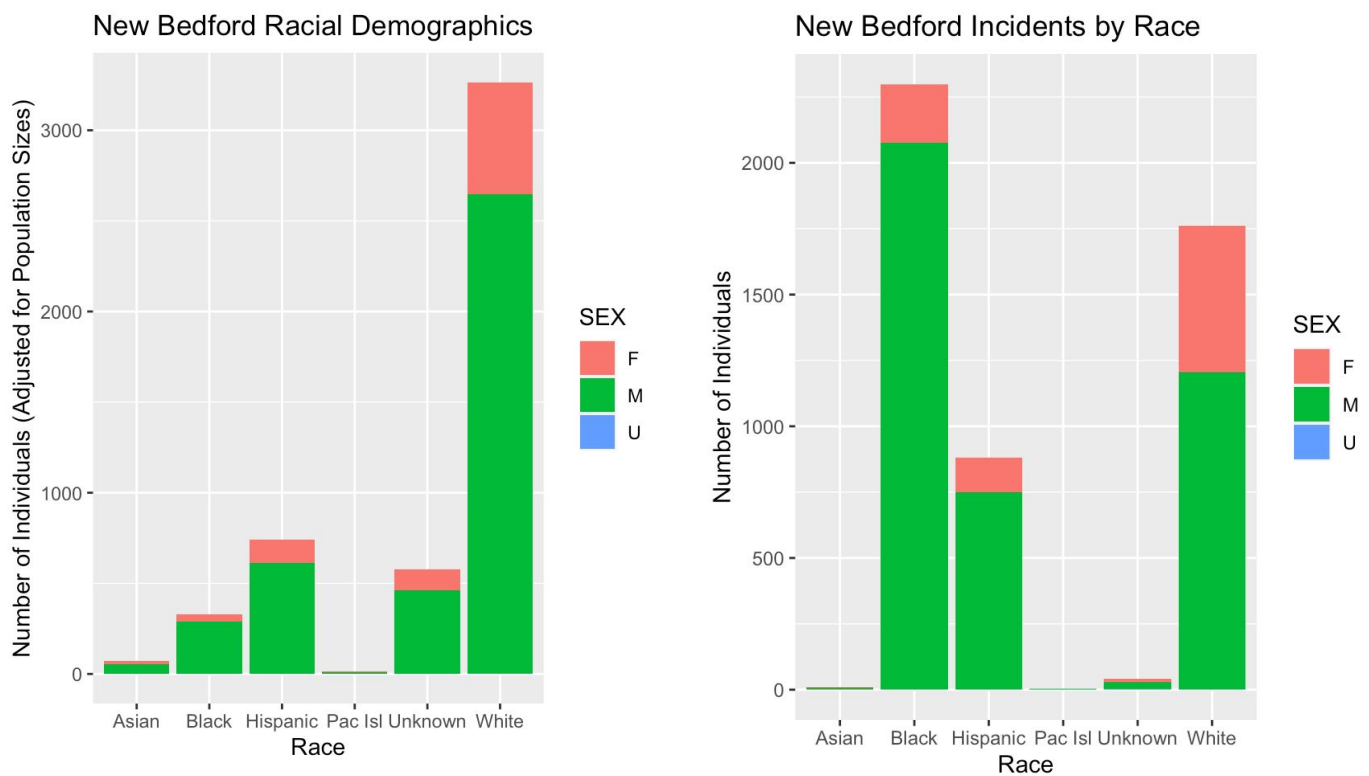
Citizens for Juvenile Justice: Deliverable 2

Written By Alex Blumenfeld, Raphael Bruni, and Dan Katz

Up to the current moment in time, all of the questions outlined in Deliverable 0 (namely, 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 city of New Bedford. Similar work will be done on the cities of Haverhill and Springfield, but the analyses that have been completed thus far with regards to New Bedford provide a good outline of any future explorations. As meetings with the lead client from the “Citizens for Justice Organization” have progressed throughout the similar, supplemental inquiries have been added to the data science undertaking. Most of the analysis that has been accomplished pertains to the features in the New Bedford dataset. Said variables have been categorized by the race and sex of the involved civilians and, moreover, certian 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.

One of the most important requests by the client concerns the race of the individuals involved in incidents (to clarify, an “incident” can be informally defined as any scenario in which a police officer suspects a civilian of wrongdoing, and thus decides to interrogate that individual). Our analysis of the New Bedford data found that certain minorities (specifically, Blacks/African Americans and Hispanics) are severely overrepresented based on their respective

population sizes. Despite making up about 6.59% of the New Bedford population, African Americans represented greater than 46% of the unique individuals involved in police incidents from 2015 through 2020. Similarly, Hispanics accounted for nearly 18% of incidents despite accounting for less than 15% of the population. To better understand these differences in representation, consider the two bar charts below, one of which shows the incident data by race, and the other of which summarizes demographic data in the city of New Bedford:



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 an implication of the bias on the part of police officers’ to detect certain races of people.

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).

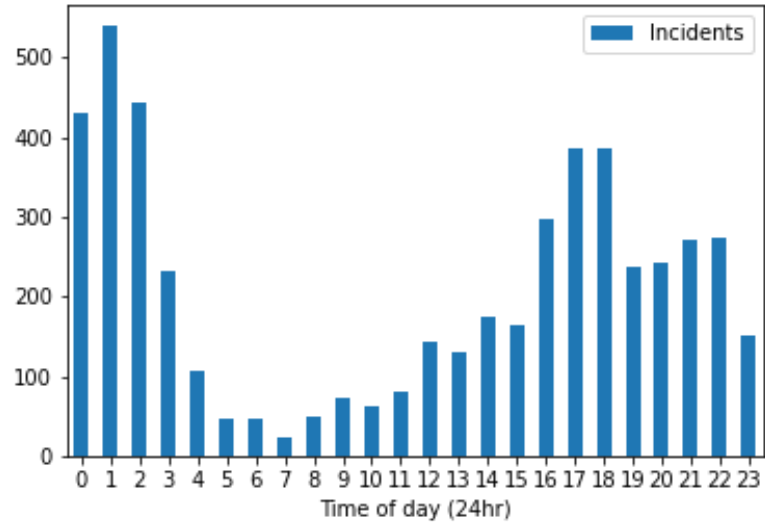
It can also be interesting to visualize where many of these officers conduct their incidents. For instance, a visualization of incidents by the most prolific officer (Officer DaCunha) is presented below:

Sheet 1

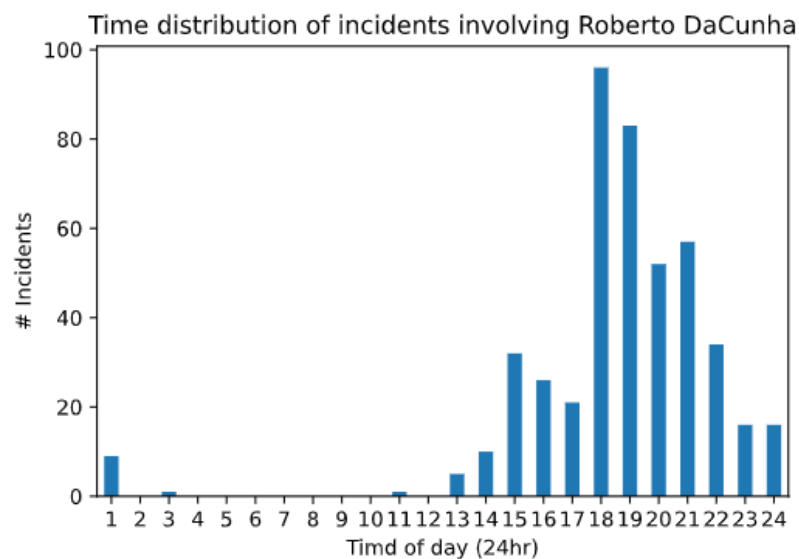


The map for Officer DaCunha is similar to that for many of the other officers in that three noticeable clusters of incidents appear: one near the southern part of the city, another near the eastern section of the city, and another in the north. These regions could be highly populated, or perhaps a disproportionate 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.

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:

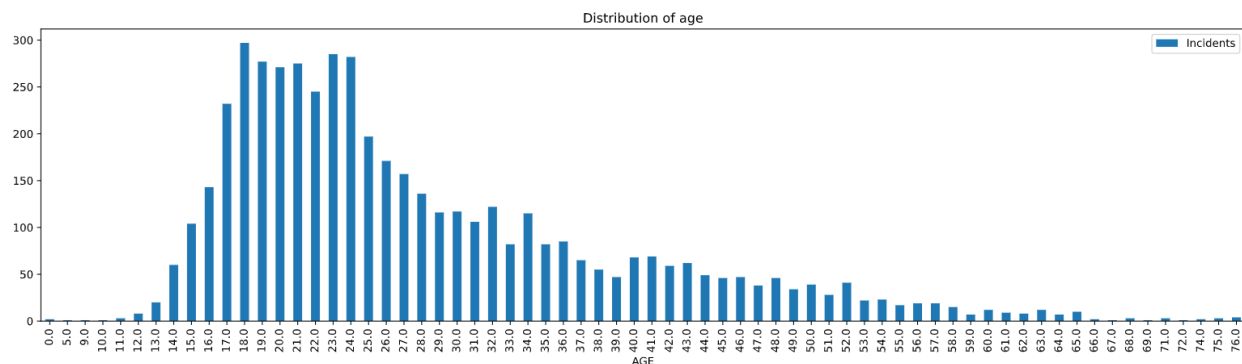


From this display it is clear that incidents are occurring with more frequency towards the night time, with a peak at around 1 AM and a low at 7 AM. This skew could indicate that more police officers are patrolling at night, or it could be a sign that more suspicious behavior is occurring at those times. 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 almost all of his stops at night, with a peak occurring at 6 PM. This could be an indication of which times officer Dacunha is patrolling.

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

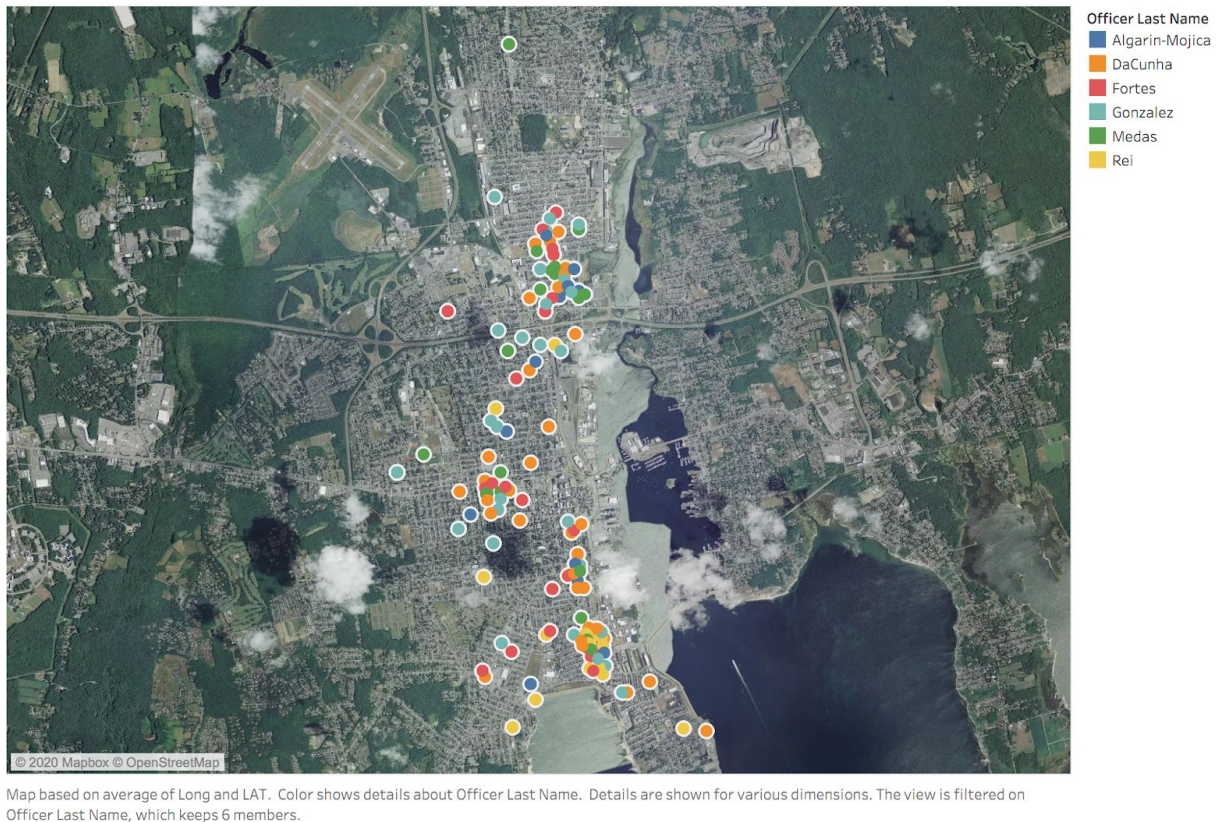


It is clear that incidents occur with a much higher frequency from ages 18 to 24. Specifically, the “Citizens for Juvenile Justice” organization is interested in examining three separate categories for young people: those aged 10 through 17, those aged 18 through 20, and those aged 21 through 25. Filtering 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.

“Citizens for Juvenile Justice” is also determined 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:

Sheet 1



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

A few weeks into the project, our client brought up another dataset that they wanted us to analyze: 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. CFJJ's request was for us 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, they are interested in which offenses young people are being arrested for the most

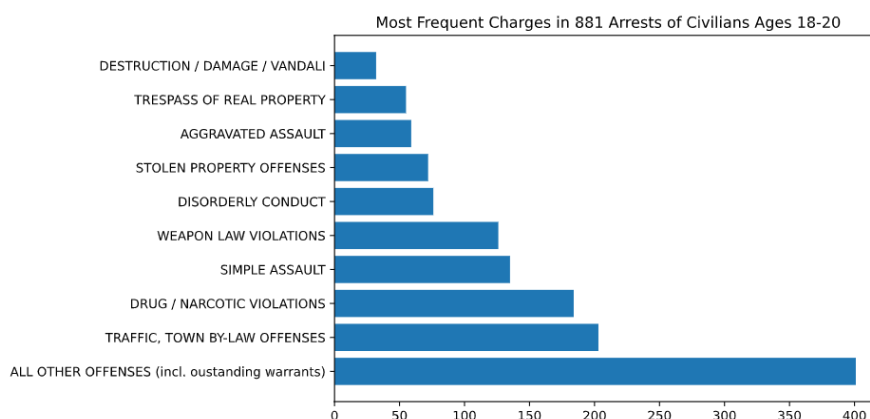
often, the frequency with which minorities are getting arrested relative to their prevalence in Springfield's population, and some maps depicting the locations of these arrests. As of right now, the data cleaning process is complete and we have access to a CSV file containing nearly four years' worth of arrest reports (over 10,000 arrests in total), and the actual analysis is still in progress.

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 got 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 are required to 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 $\frac{2}{3}$ 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 mapped below. Just like in New Bedford, most of these arrests are occurring near the downtown area of the city.

