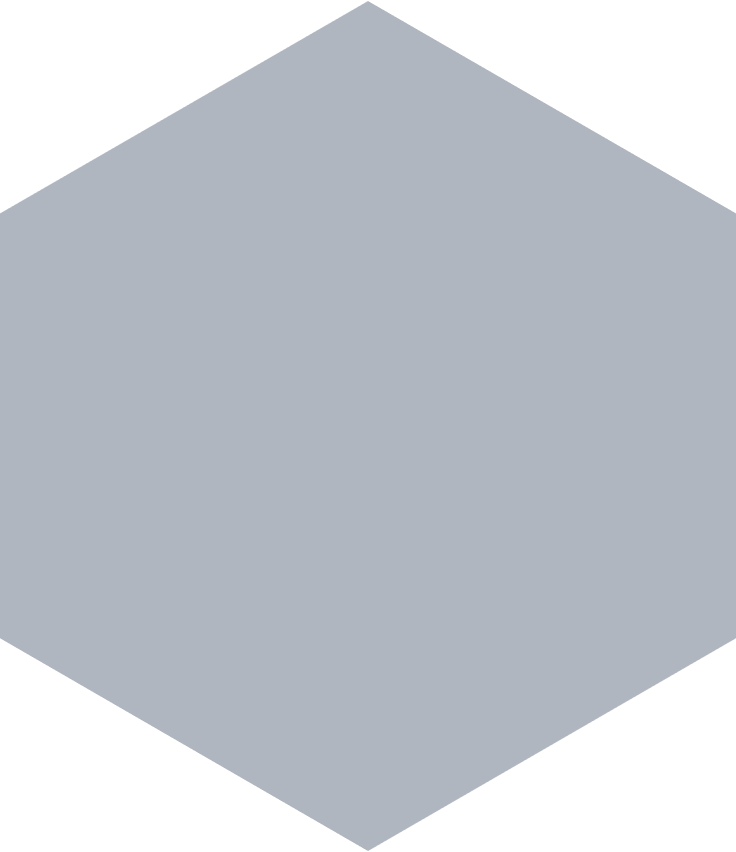
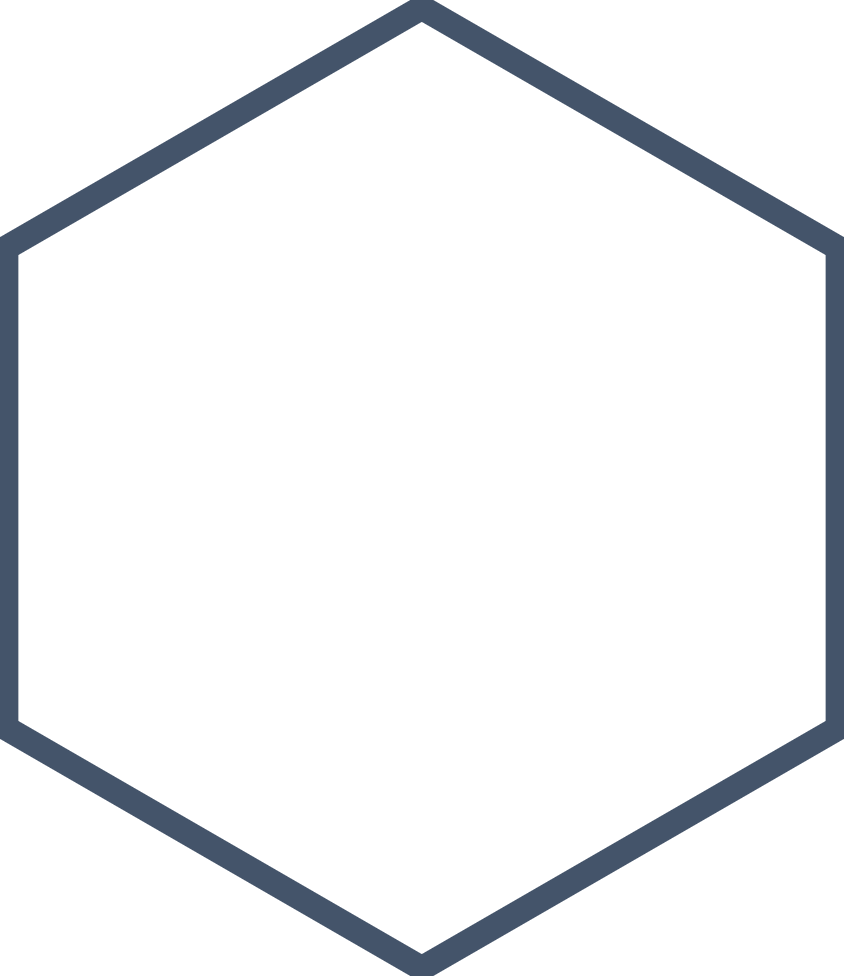


|  |
| --- |
| **Minneapolis Police Stops** |
| Project 1 Analysis Report: Washington University Data Analytics Bootcamp 2018-19 |
| Our project investigates correlation between police stops and demographic, time, and geospatial variables in Minneapolis, MN. |
|  |





**Minneapolis Police Stops**

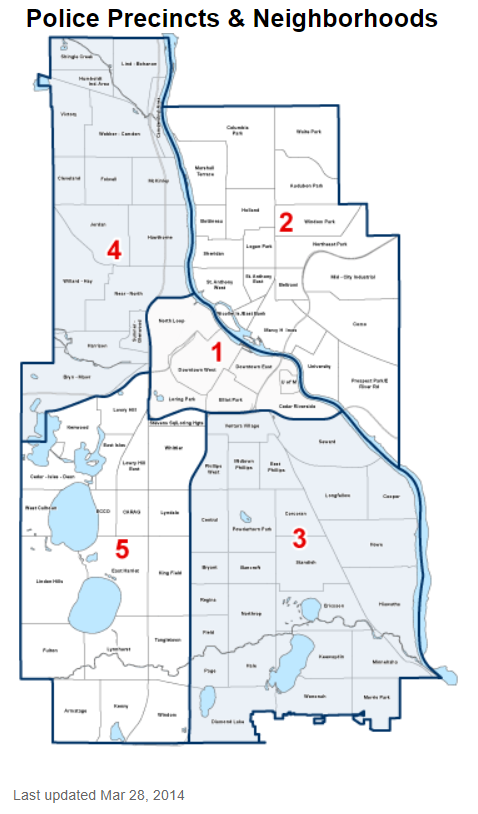
Analysis Report

Group F: *James Curtis, Raleigh Love, Emanshu Patel, Erica Unterreiner*

**Purpose**: Our project investigates correlation between police stops and demographic, time, and geospatial variables in Minneapolis, MN. Our data source is from a public-facing Minneapolis Police Department database, which collates data from MDCs (squad computers located in squad cars). To download this data, please visit the [open data geospatial portal for Minneapolis](http://opendata.minneapolismn.gov/datasets/police-stop-data?geometry=-93.984%2C44.799%2C-92.159%2C45.139). For an overview of Minneapolis police precincts, please see figure 1 below.

**Data limitations**: The data used is a subset of the main police dataset, with a few quirks due to the cleaning that occurred before it was posted. For example, categories of person and vehicular search were merged into one column, and a few columns were removed from the dataset such as call type and call reason. Data was missing for rows not entered into a squad computer, which we removed from the dataset. Also, the data reported is from the police department itself, and not an independent provider. We deemed these changes not significant enough to affect analysis of our current questions, but would revisit for future questions.

Figure 1



Research Questions:

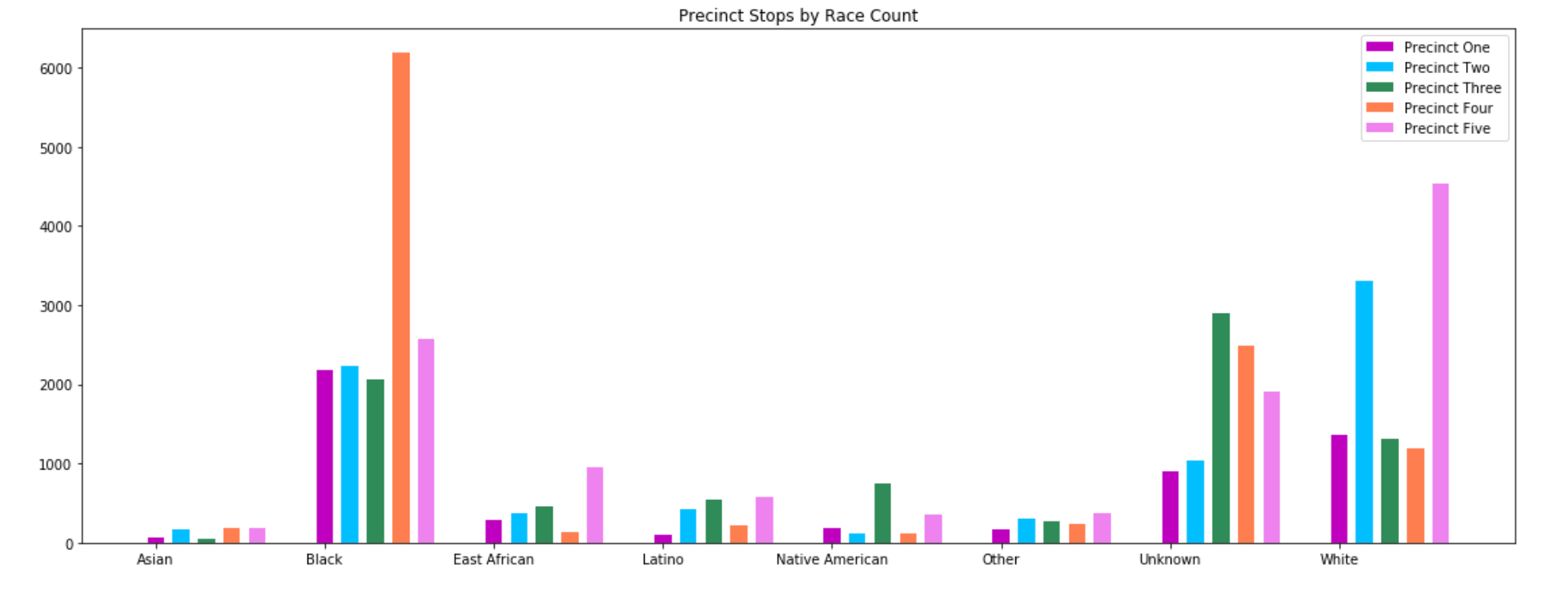
1. **Are there differences in race given the stops recorded by precinct?**
   1. *How does the chance of stop for particular races relate to the precinct?*

We wanted to observed if there were any significant differences between stops by precinct; each precinct, while still de jure bound by department rules, may display individual adaptations to their particular area, may reflect particular leadership styles, and considering the relatively large and differing areas of each precinct, may contain heterogynous population make-up. Additionally, each precinct has local road, traffic, and crime conditions.

When observing figure 1 below, what first stands out is the count of Black police stops in precinct four (> 6000), greater than twice in magnitude than the next precinct. When observing the census data for this area, it does have a significant Black population, but that doesn’t measure people actually driving in an area. Precinct five and two have markedly higher police stops for white drivers over the other precincts, by a factor of three and two, respectively. Numbers for unknown race can result from a variety of factors, such as calls that didn’t result in actual police encounter with an individual. Also, East Africans are represented twice as often in precinct five as the next nearest precinct, and Native Americans twice as often in precinct three as the next nearest precinct.

It was impossible to compare police stops proportionally to race for the driver population based on the data available, because we didn’t have a true representative sample of the driver population.

Figure 2



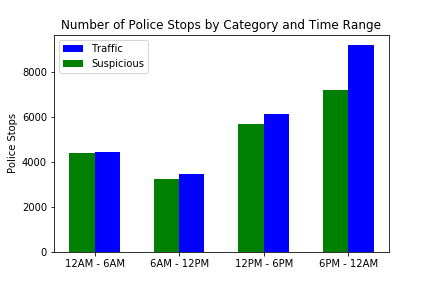
1. **Are there more stops during the morning (6AM-12PM), afternoon (12PM-6PM), evening (6PM-12PM) or nighttime (12PM-6AM) blocks?**
   1. *How does time of day influence the number of police stops?*

When considering this question, we wanted to consider potential relevant differences between various times of day in relation to city activities. We chose to divide each day into four time blocks; morning contains commute activity for much of the driving population (Source: <http://overflow.solutions/demographic-data/what-time-does-the-average-american-leave-for-work/>), afternoon includes return commute activity, evening represents leisure or commute activities (especially for shift workers), and night generally includes sleep and leisure activities.

A few interesting notes from the data for this question, represented in Fig. 2 below, include that police stops were especially prevalent from 6pm to 12AM, with traffic stops at an average of 9199 and suspicious stops at 7197. Suspicious stops nearly equaled traffic stops for periods other than 6pm to 12AM. 65% of all stops occurred during the second half of the day, from 12PM to 12AM.

It would be interesting to correlate this data with information on where police are positioned in the city during various times, and which duties they are performing. As stops can be generated by police, a larger presence may result in more events that would necessitate a stop being witnessed. To our knowledge, this information is not publicly available.

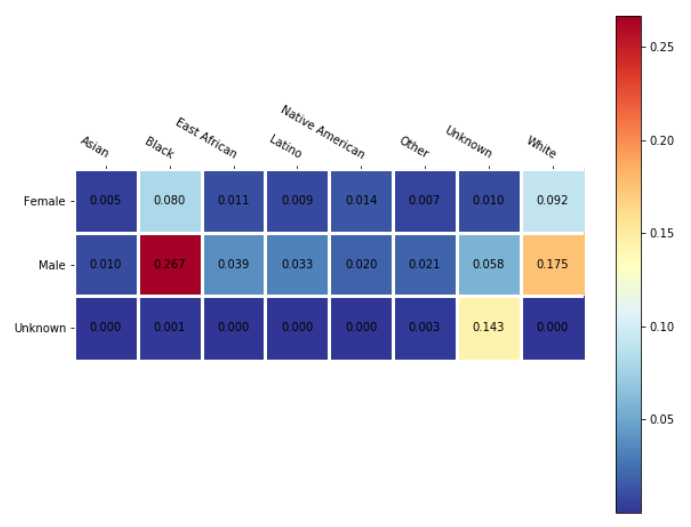
Figure 3



1. **What is the breakdown of police stops by race and gender?** 
   1. *How do race and gender predict the likelihood of a traffic stop?*

We were interested in doing a bivariate analysis of police stops using demographic indicators already in the data, such as the person’s perceived gender and perceived race by the officer responding. A few interesting notes: the highest combination was black males, at 26.7% of all police stops, followed by white males with 17.5% of all police stops. Throwing out likely noise data with unknown gender/unknown race, the next highest combination was white women at 9.2%. Again, without having a representative sample of drivers, this information is less actionable and we are cautious to draw conclusions without adding additional factors.

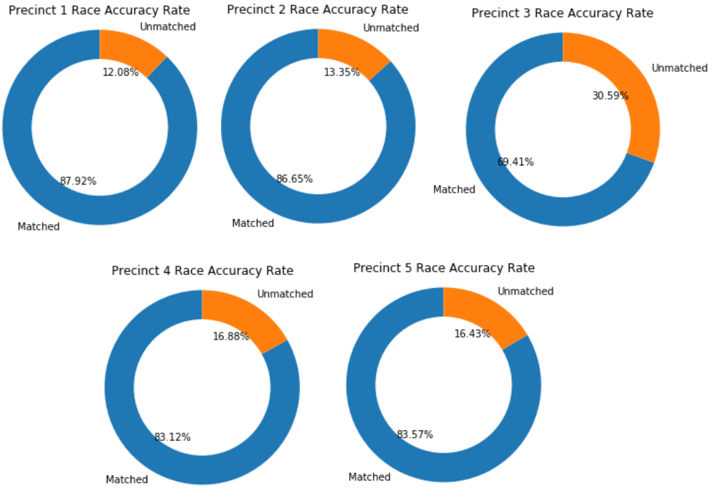
Figure 4



1. **What is the likelihood that pre-race determination matches the actual race perceived after the stop?**
   1. *Which precincts have greater accuracy in determining race prior to a stop?*

We were interested in measuring the correlation between a pre-race determination for a police stop, which represented a subset of the total police stops (as not all stops have this datapoint), and perceived race as recorded by the officer during an interaction. There is a large disparity between results in precinct 3 and the other precincts, with precinct 3 reflecting a 30% difference between pre-race and perceived race during a police stop. This is nearly greater by a factor of two, than the next nearest precinct, and precincts 1, 2, 4, and 5 cluster closely in the low to mid teens. It is possible this is due to the observed racial diversity in precinct 3, as seen on the dot plot map.

Figure 5



1. Does the type of police stop (traffic vs suspicious) impact whether a person will be searched?

I don’t think we have plots for this; up to ya’ll if we work this one, let me know if I’m missing something somewhere---

**Statistical Analysis**

While the source data itself illustrated some interesting relationships between a number of variables, we had to check to see if those relationships were statistically significant. Considering the moderate size of the dataset, Excel provided a sufficient medium for analysis. This analysis was performed in three files: The first used the main dataset to corroborate and double-check findings in the Jupyter Notebook. It was also used for exploratory trend highlighting.

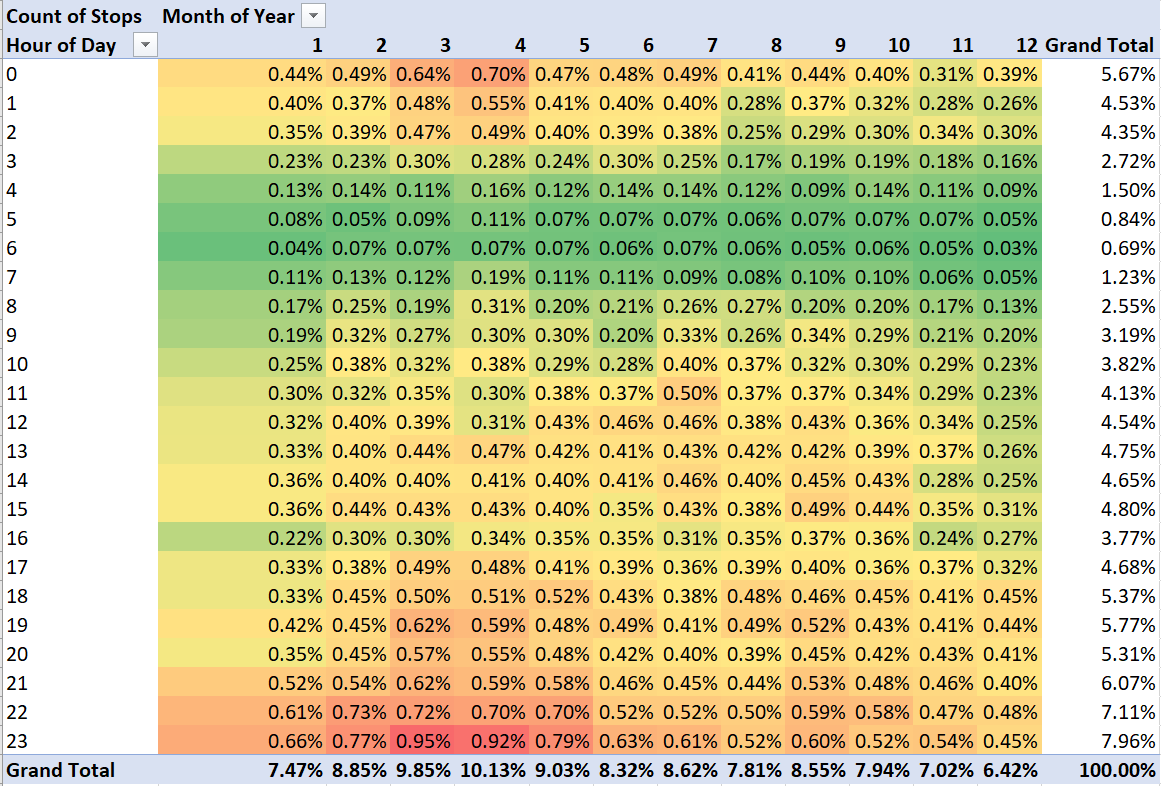


Figure 1: Why is such a high proportion of the year's police stops occurring at midnight in March & April?

The second excel file incorporated census data from Hennepin county from 2010. Unfortunately, the census data was poorly formatted for importation and recombination with the Minneapolis dataset.

We wanted to compare the racial percentage breakdown of police stops with the overall percentages of the population. This proved difficult due to the census categorizing race differently than the police data. For example, the 2010 census grouped “East African” with “Black,” while the Minneapolis police did not. Likewise, the census categorized people of two or more mixed races, while the police data did not. Each combination of census race, age, and sex categories had a number of respondents which could be aggregated for comparison. But in order to create the comparable aggregate, the census categories had to be divided into the police data categories (grouping East-African with Black) and each group of respondents was flagged with a Boolean. This was slightly complicated by the census categorizing Latino in its own category.

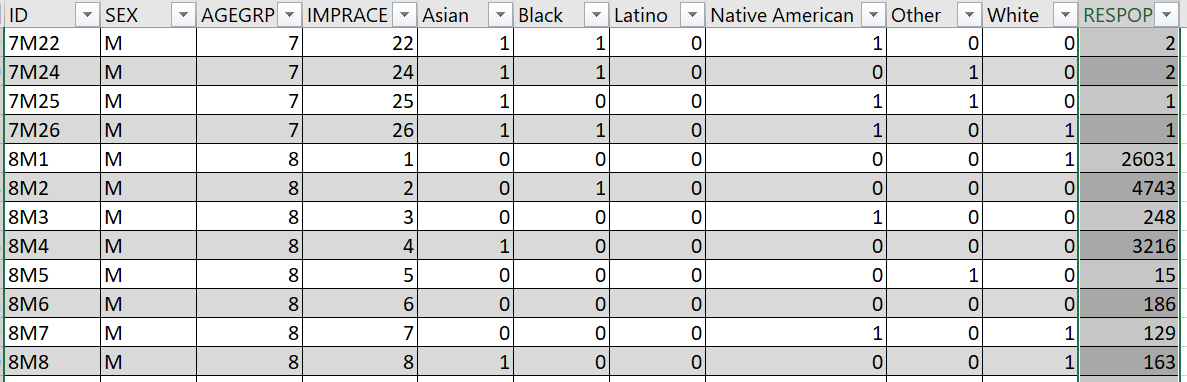


Figure 2: Sample of Census data cleaned to boolean police data race categories. For example: Census Racial category 22 is classified as a mix of Asian AND Black AND Native American. There are two men of age category 7 (age 30 -34 years) in Hennepin county matching that description.

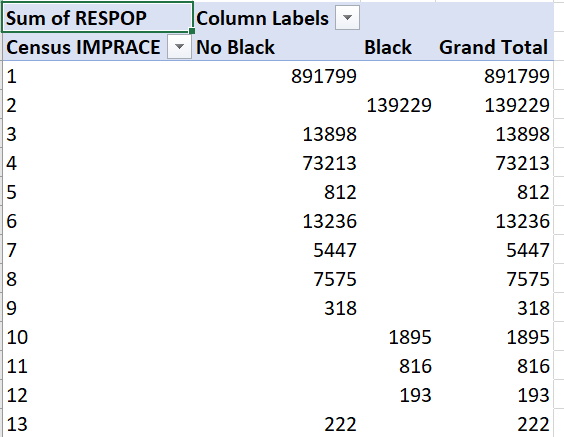


Figure 3: Census Race categories to the left, one of the selectable Police data categories at the top. The table illustrating which respondent groups match the police data designation.

The third excel data file used the data in the previous two to create correlation tables and run linear regressions. While there were numerous correlations, most were statistically weak. An example of some stronger correlations occurred for the question of whether certain races were searched more frequently during traffic stops. While correlation is not causation, it was significantly less likely that a racial determination would NOT be made during a traffic stop. This is in logical contrast to the possibility that people whose race could not be determined were significantly less likely to get stopped by police for traffic violations.

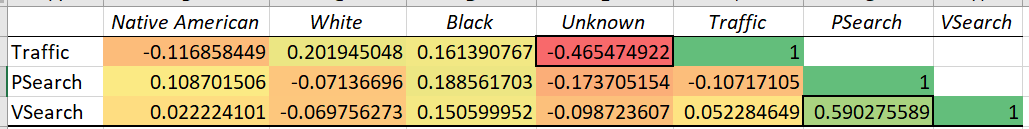


Figure 4: Native Americans were slightly less likely to be involved in traffic stops. Whites were more likely to be involved in traffic stops, but less likely to be searched. Blacks were more likely to be in traffic stops and more likely to be searched.

The regressions performed were descriptive, but not in the manner we had hoped. While the T-stats and p-values were all good, the R² values were low, indicating that the variables very accurately accounted for a very small portion of the variation in the data; the rest was random noise. Regressions were run on precinct predictability of race mismatches, precinct predictability of issuing citations, racial prediction of traffic stops, personal searches and vehicle searches; none of them were statistically significant enough to reject the null hypothesis.

**Conclusions:** After performing our analysis of the data using tables, visualizations, and statistical analysis, we cannot confidently draw actionable information from these initial analyses based on the data we have. While we have new ideas for interesting future analyses, such as finding other data sources for a representative sample of driver demographics, or seeking to correlate more fine-grained geospatial information with points of interest data which might be related to types of police stops, we know enough about the limitations of the current dataset to be cautious.