## Quantifying Implicit Bias with the Stanford Open Policing Project

**Background**

One survey showed that US Police Officers make an average of more than 50,000 traffic stops a day. This is one of the clearest ways police officers interact with citizens, and it can leak information regarding recent developments in law and politics, as well as the implicit racial bias of the officer.

In the course of a typical traffic stop the police officer has a number of judgement calls to make using incomplete data:

* Should I pull over this vehicle?
* Should I perform a search?
* Should I write them up for a violation?
* Should I arrest them?

While there are objective measures that may make one of these courses of action more appropriate (the vehicle is going 8 miles over the speed limit, the smell of cannabis is apparent, etc), but at each decision, there are still some unknowns and some judgement is still required. It is this exercising of judgement that we want to verify is just, and in line with our values.

**The Open Policing Project**

The Stanford Open Policing Project is “collecting and standardizing data on vehicle and pedestrian stops from law enforcement departments across the country — and [they’re] making that information freely available. [They’ve] already gathered 130 million records from 31 state police agencies and have begun collecting data on stops from law enforcement agencies in major cities, as well.”

The project has published a dataset detailing all traffic stops in the state of California from the year 20013 to the year 2016. The dataset has the date and county of the stop, along with the apparent, age, gender, and race of the driver. It also details the salient events: whether or not a search was performed, whether or not contraband was found, whether or not a driver was written up for a violation, and finally whether or not the driver was arrested.

**Cleaning the Data**

The size of the original file was 2.3 GB, so it was useful to initially decide which attributes could be useful, and which ones were just taking up space. In TrimData.ipynb I explored the potential information in each attribute and then saved a new csv file with the unuseful attributes dropped.

The original list of attributes provided in the dataset follows:

['id', 'state', 'stop\_date', 'location\_raw', 'county\_name',

'county\_fips', 'fine\_grained\_location', 'police\_department',

'driver\_gender', 'driver\_age\_raw', 'driver\_age', 'driver\_race\_raw',

'driver\_race', 'violation\_raw', 'violation', 'search\_conducted',

'search\_type\_raw', 'search\_type', 'contraband\_found',

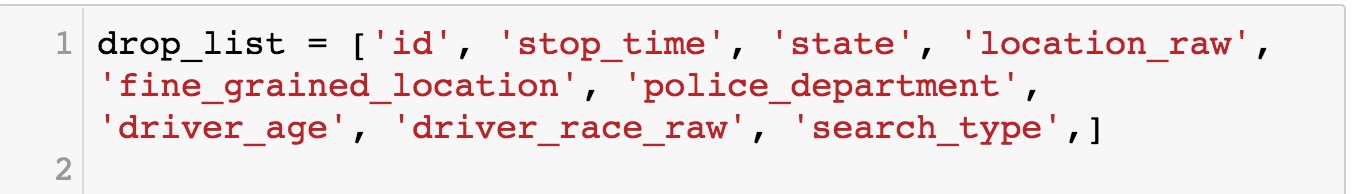
'stop\_outcome', 'is\_arrested', 'ethnicity'

]

After a subsequent investigation of each attribute, some were found to not hold much informational value. There weren’t any non ‘nan’ ‘stop\_time’ values, so I elected to drop that attribute. The same was true for the ‘state’, ‘location\_raw’, ‘fine\_grained\_location’, ‘police\_department’, ‘driver\_age’, and ‘search\_type’.

Other attributes just held redundant information, such as the ‘driver\_race\_raw’ attribute, which held a more detailed list of potential racial classifications. Because ‘driver\_race’ had a useful compression of these various racial categories, the information in ‘driver\_race\_raw’ was redundant. The relationship between ‘violation\_raw’ and ‘violation’ was similar.

After removing the following list of attributes, I then wrote the data back to a file called trimmed\_data.csv. While the original file was 2.3GB, the trimmed\_data.csv file was 1.8GB—a space savings of more than 21%.



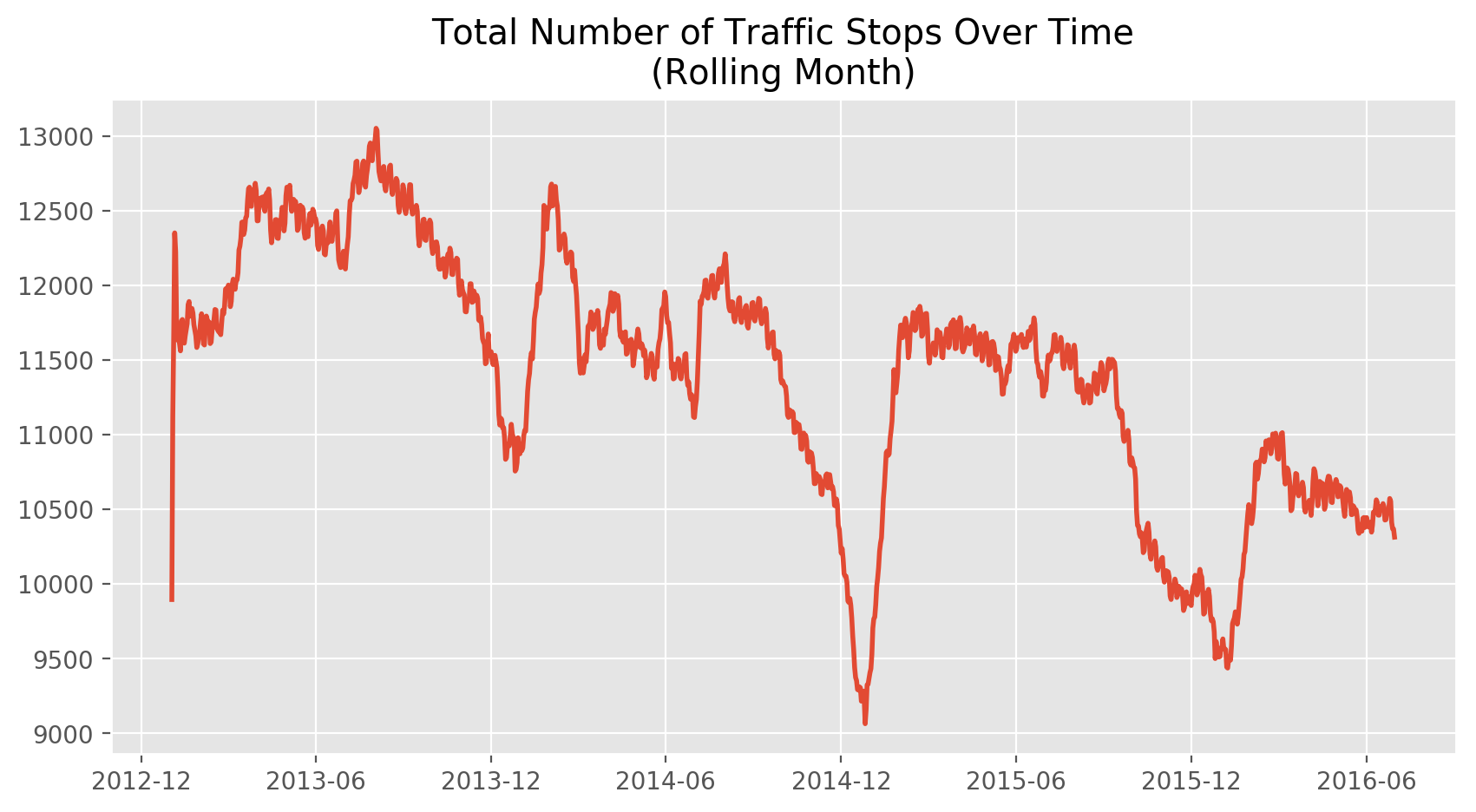
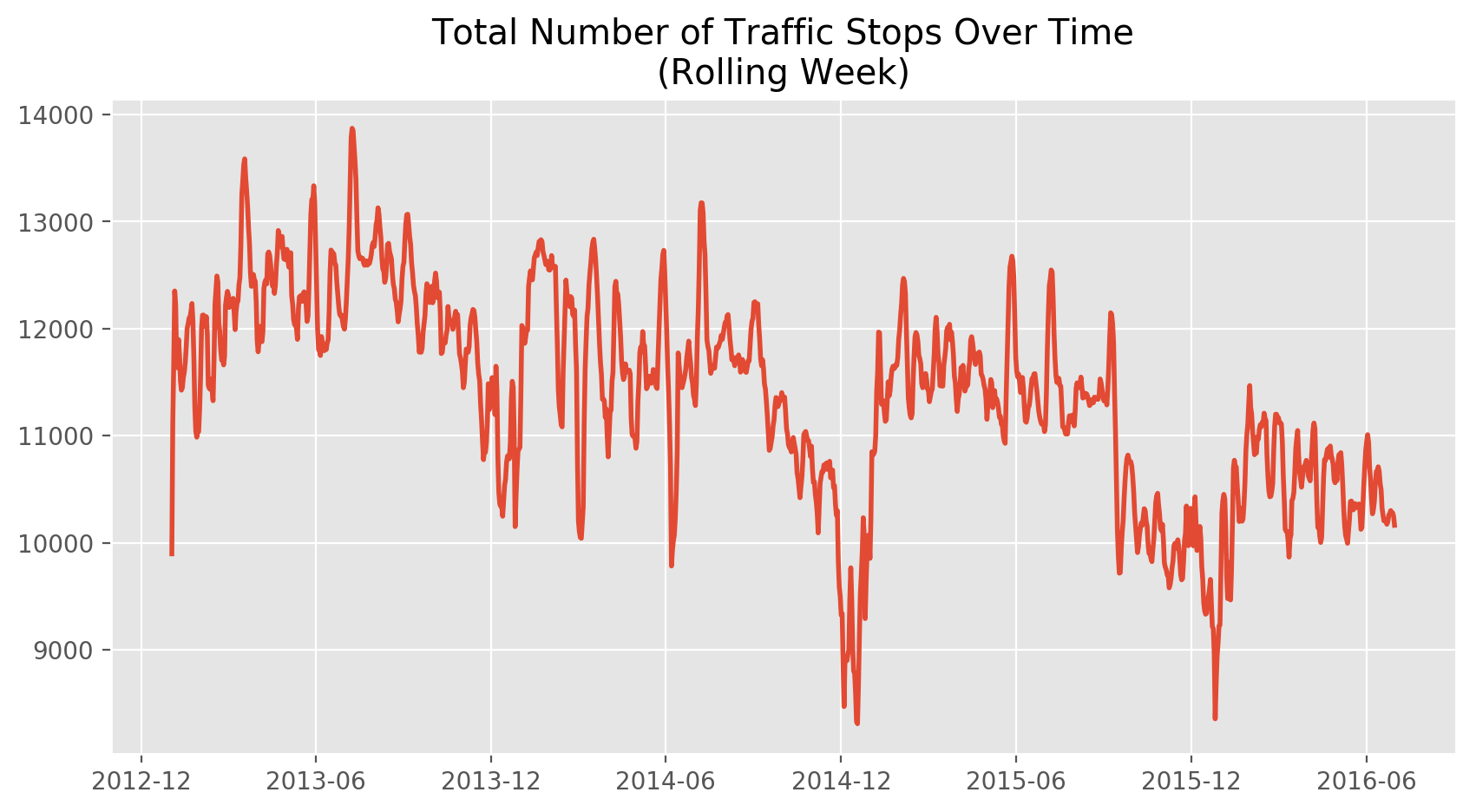
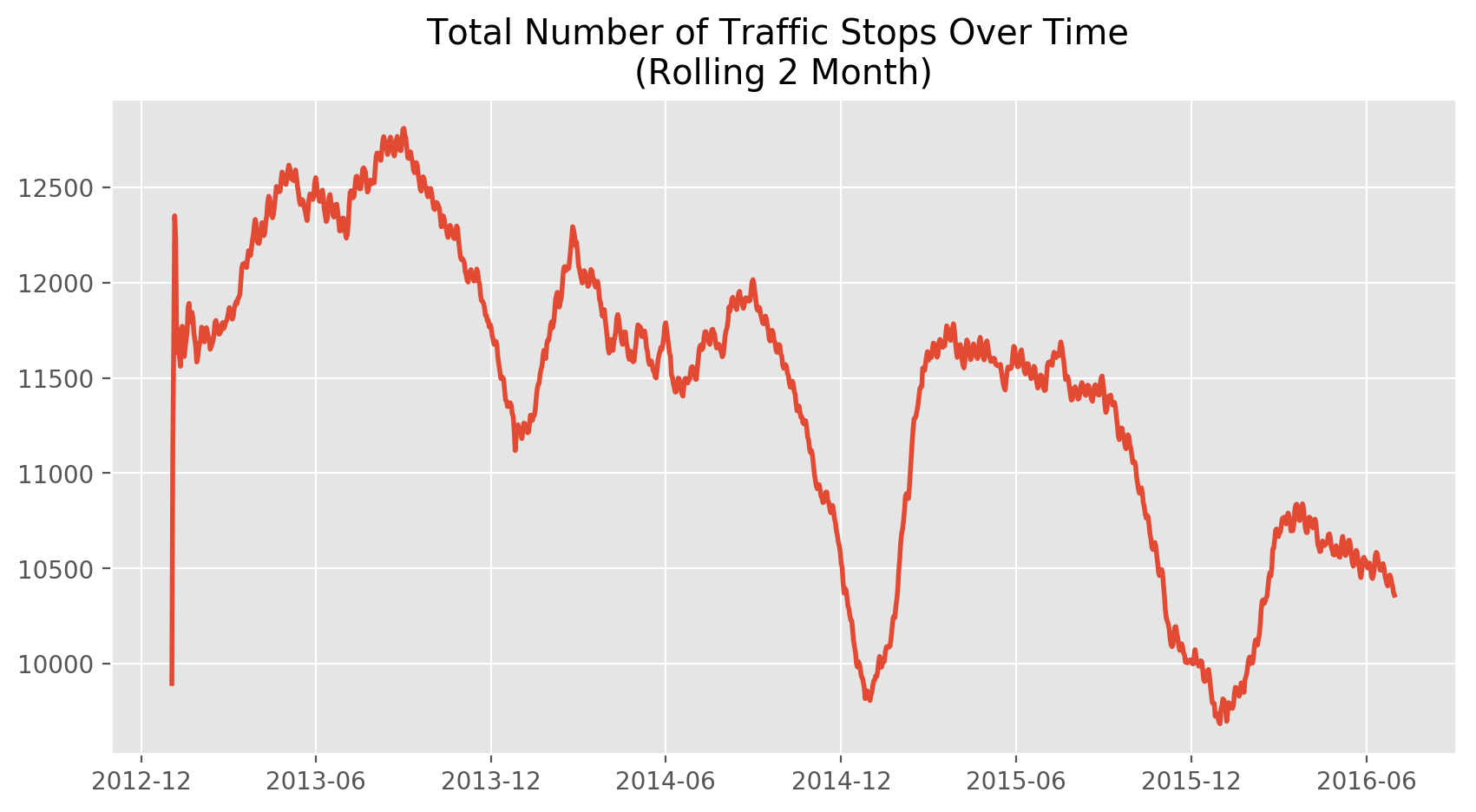
Other cleaning steps involved compressing some of the values into codes to be dereferenced (for example, compressing the text ‘violation’ field to a single numerical code that references one of the actual violation strings).

**Initial Exploration**

To get a general sense of the data I wanted to do some preliminary exploration. On the highest level it was useful to see how many traffic stops there were, and with what frequency they occurred. After transforming the stop\_date into a useable format, I plotted the total number of stops on a day with respect to time. Plotting just the number of stops on a single day was incredibly noisy to visualize, so I tried plotting the rolling 7-day mean over this period. Seeing this it was immediately clear that there were interesting trends and potentially some distinct periodicity, but there was still a substantial amount of noise.

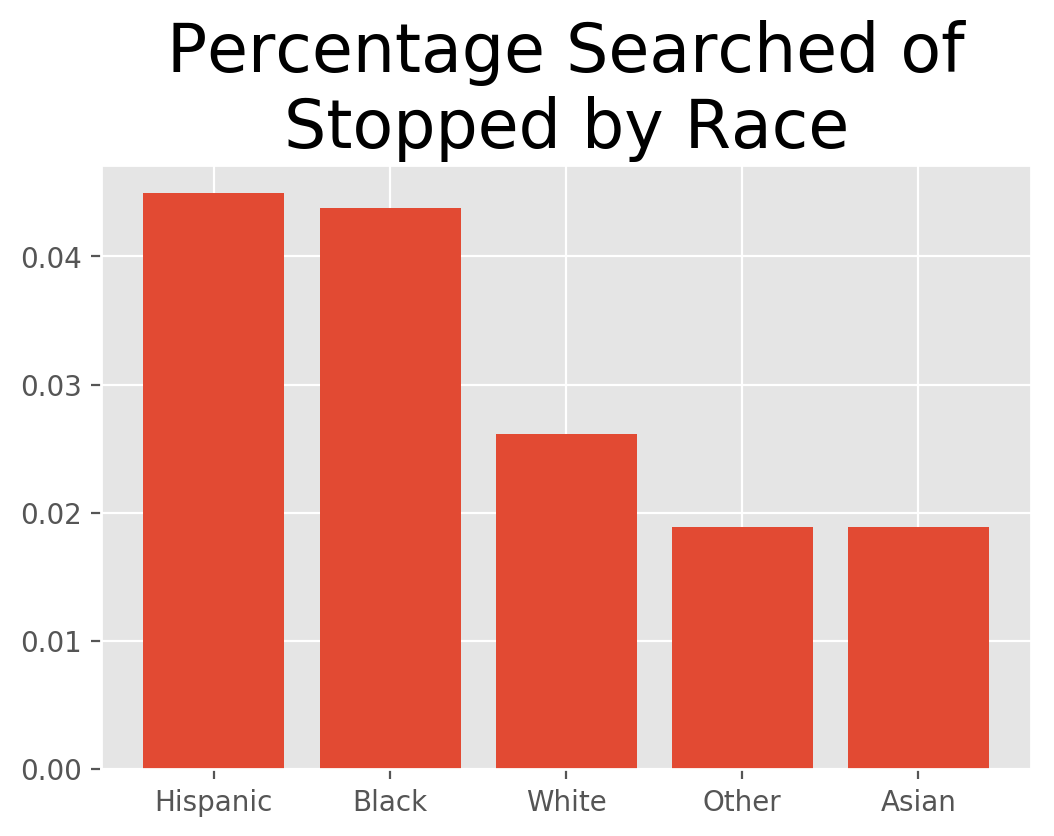
Calculating and visualizing both the rolling month average and the rolling 2 month average were helpful in seeing this periodicity with greater clarity. There is overall, a decreasing trend in the frequency of stops, but there also seems to be a distinct drop in stops leading up to and right around the new year.

It would be interesting to search for explanatory causes of this behavior, and match empirical effects in the data with more quantitative events from politics and public policy.

I also wanted a high-level look at the potential for implicit bias. It can be problematic to just look at the total number of those stopped/searched/etc, due to the uneven prior probabilities; for example, there are much more white drivers in California than there are black drivers, and its possible that those of certain ethnic backgrounds are more likely to exhibit certain types of behavior. These variants are clearly outside the scope of implicit racial bias.

One way of getting around this is to look at the percentage of people who advance from being stopped to being searched. Because we know the exact racial breakdown of those who are stopped and those who are searched, we can isolate the probability for each race that a traffic stop leads to a search (a more severe interaction with the office, based on the officers own judgements).



From this visualization it is clear that the rate of a traffic stop leading to a search is much greater for Hispanic and Black people than it is for White and Asian people. This provides a very rough proof of concept that there is a statistically significant phenomenon to explore in more depth.

**Further Work**

Further work will include looking at the full funnel of potential police action (living in California, driving, getting stopped, getting searched, getting arrested), with a particular focus on whether or not any difference in racial behavior is justified by later results (for example, if a search of a Hispanic driver was much more likely to lead to discovering contraband than that of an Asian driver, it might be seen as justified to have a lower threshold for searching a Hispanic driver).

I’ll also take into account a geographic component. If it’s possible to succinctly describe a “racial bias metric”, that metric can be compared over different counties to see if certain areas have more racial bias than others. If possible I’d like to add more data from other currently available states, but of course this will lead to its own challenges.