

SDPD Stop Data Analysis

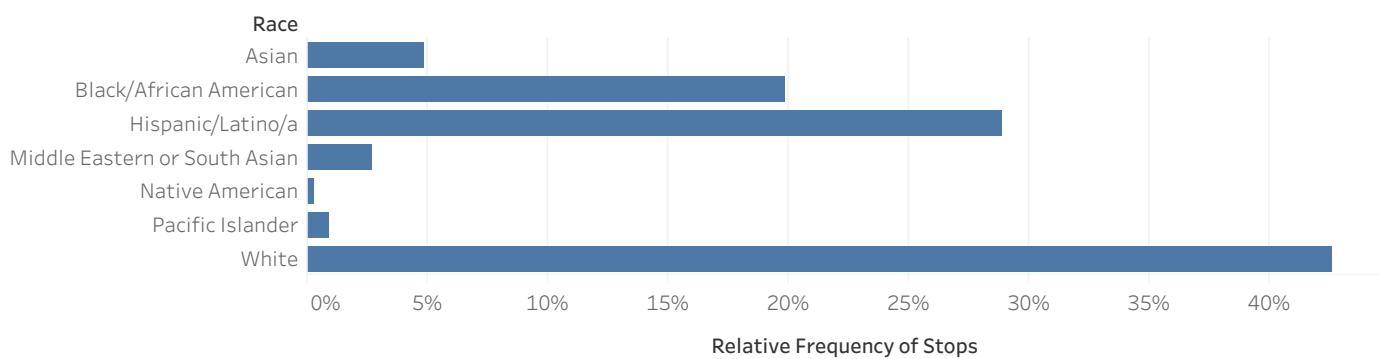
By Trey Scheid

Purpose:

My goals for this project were to learn how to use the tableau software to explore and communicate large data sets. For this data set I wanted to work on asking important questions and answering them with data visualizations that correctly and clearly showed the desired analysis such as the correlation or outlier of interest.

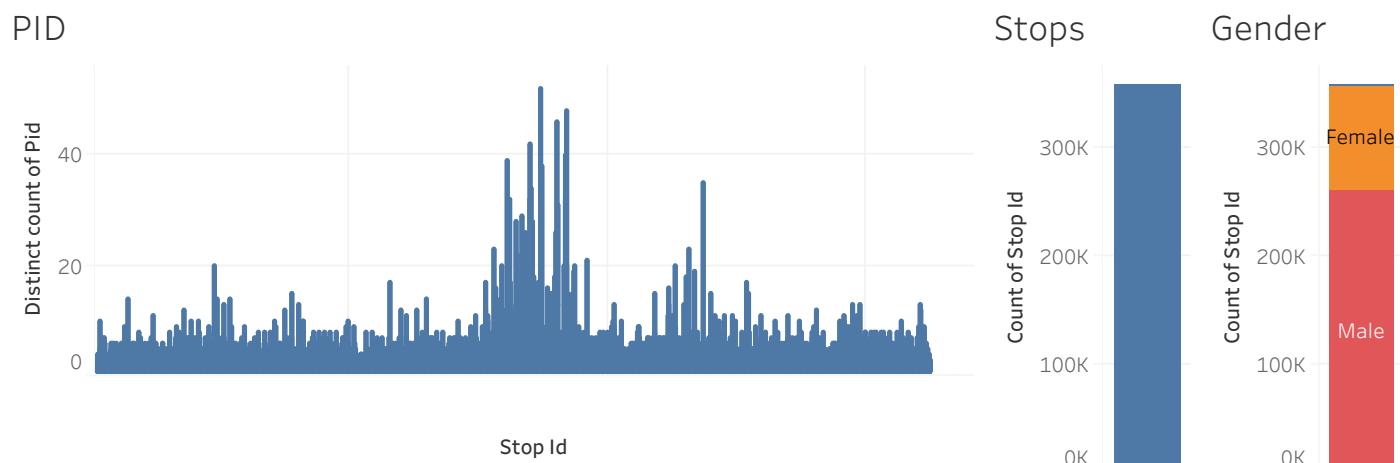
After reading the San Diego Union-Tribune article about an incorrectly done test on racial bias in the San Diego Police Department I thought exploring correlations with race specifically would be a good place to start my analysis. This meant comparing stops with race, location (and income in that area) and reason for stop (if it was suspicion or something more valid).

Race by Stop



My first visualization (above) shows the distribution of race of persons stopped by the police. However a look at the dictionary shows that this is the officer's perception of their race. For a majority white area and police force that likely reflects that proportion it is possible that minorities are mis-perceived and therefore this data is likely incorrect. It does give us a general idea though.

This next visualization helped me understand what PID was as a variable. It shows the identifier for each person for each stop, so it is the total number of people stopped at each stop. The maximum shown by the longest bar is 52 while most of the stops appear to have less than 10-15. Looking at totals for Stop Id, Sex, and PID are good ways to start getting to know the data set.

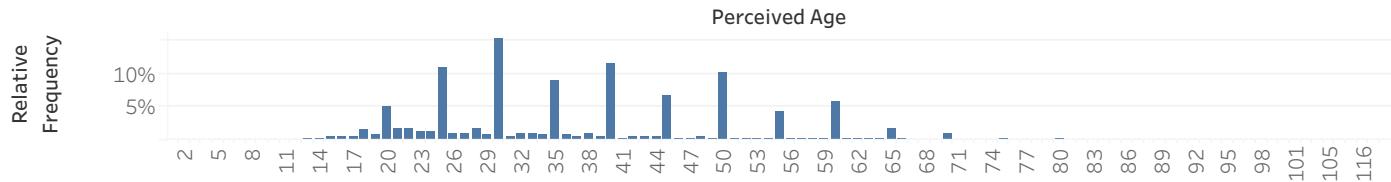


Change of Direction

Unfortunately I could not do any further analysis on race using tableau because to the best of my research you cannot create relationships/joins between data sources on tableau online. So I decided to turn towards my other goal of exploring the data and using Tableau's various features.

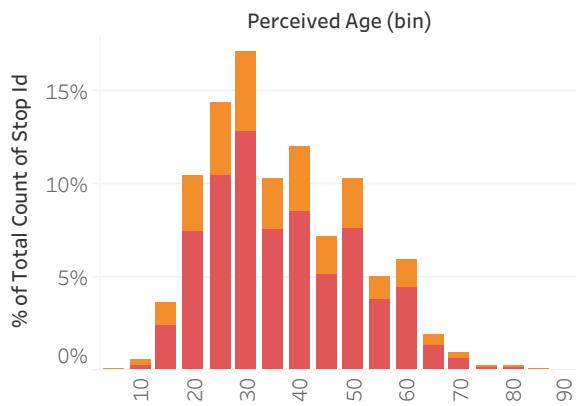
Next I decided to look at how the age data looked, some questions in mind were if this was going to be a multi-modal distribution or more unimodal with a skew. Also the maximum age would be an interesting fact pertaining to the San Diego population

Perceived Age

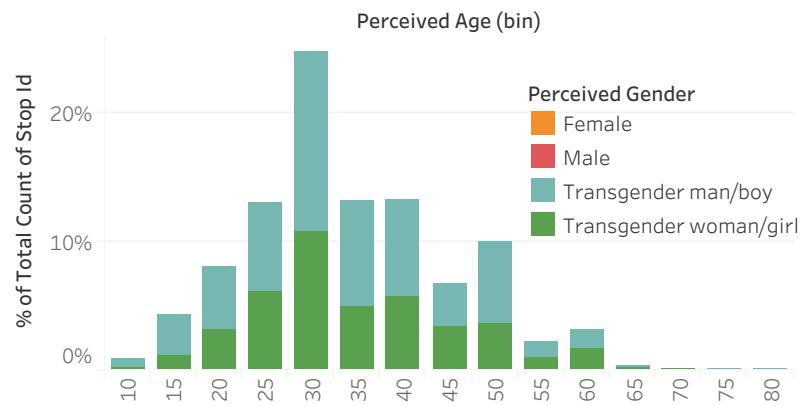


This plot has some interesting features. To start, there are many peaks at ages that are a multiple of five, likely because officers rounded their guesses since this is perceived age. On the right end we see there are counts of distinct stops all the way up to 116 years old, is this an officer's joke? We might eliminate this outlier after first reaching out to the data source to see if this was a true data point ...

All Genders by Age



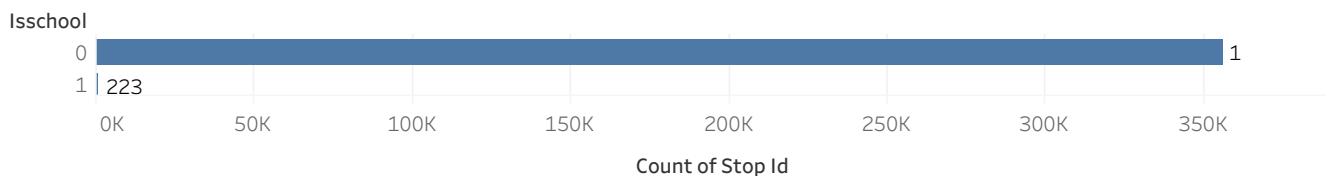
Transgender by Age



It is clear that there are more men in every age category and none by any unusual proportion. When we look at transgender people we see the same shaped distribution with no irregularities from the total. I also looked at a graph of age versus student and that did not show anything other than most students are below 20 years old as you would expect. That will bring us to a short section on students however.

I wondered how often students were stopped off of their campus, as shown below they are not at all. The next plot shows that students are actually stopped for long periods of time. It would be interesting to see what the result of these short and long stops were, such as if students mostly got off with no action or warnings.

Students at School



Student Stop times



Notice the last bar is very tall. This tells us that of all stops that go longer than 100 minutes, a larger portion than normal of them are with students. Could this be because the officer had to wait with the student for a legal guardian to arrive?

My next questions were about who was doing the most stops and why.

The data set has record of the years of experience each officer has at the time of the stop. The following visual shows how many stops are made by officers of different years of experience. This data set doesn't show the demographics for the police force nor how many are in each year of experience. So even though we see that a great majority of stops are by officers in their first year of duty, we do not know if this is because the first year officers are stopping people more or if there is just more first year officers in the force. In general it would make sense the number of officers in x year of experience would decrease over time as officers retire, maybe with a significant drop at the year where they earn their pension.

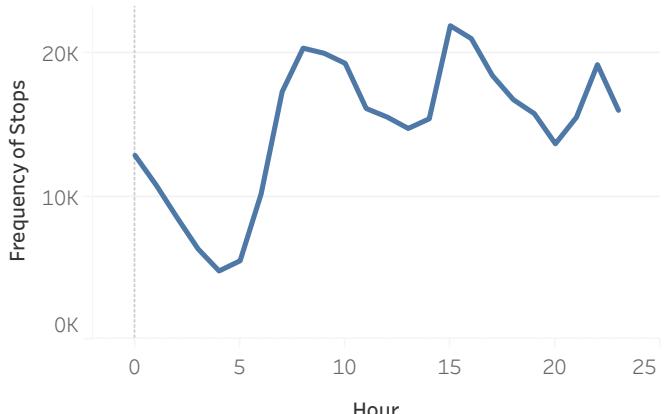
Experience of Officers



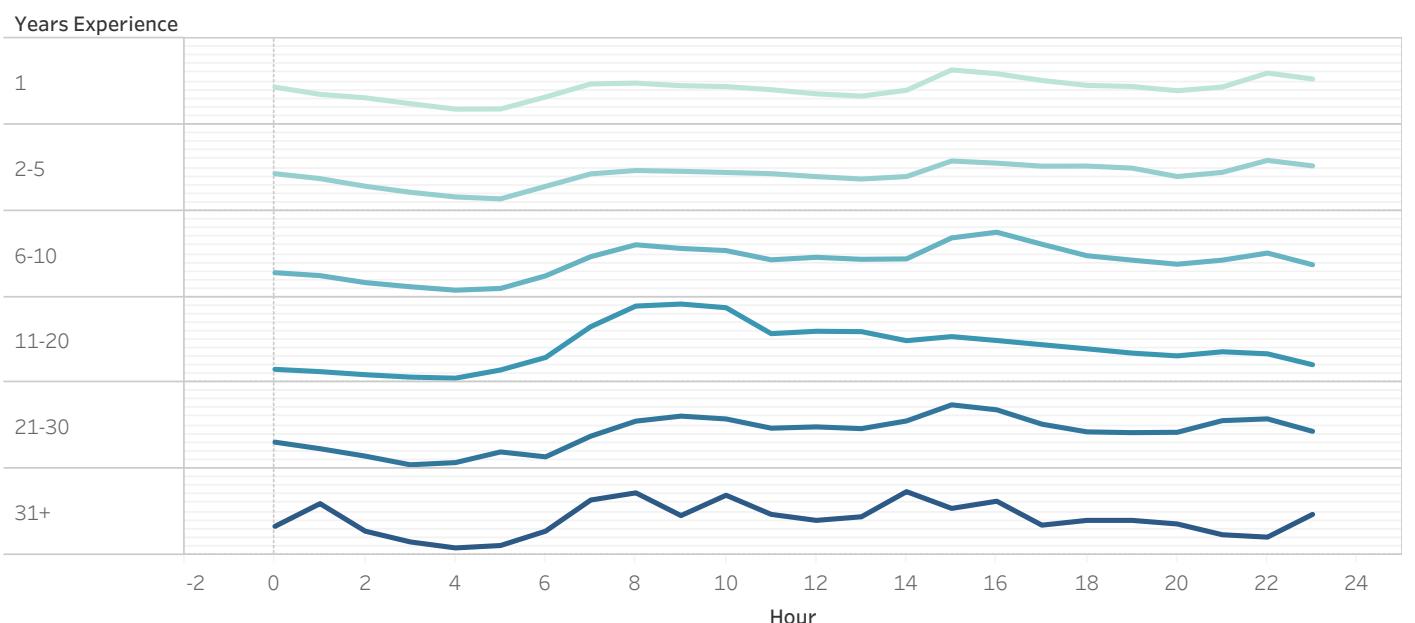
Time of Day

In the graph to the right there are spikes at 8am, 3pm, and 10pm. The 8am is during commute hour and around breakfast while the other two are after meals. There is a gentle slope down after each spike, I am unsure what this could mean.

The plot below shows a comparison of stop times for groups of officers of different levels of experience. Lining them up like this we see the most dramatic changes to the last line for the most experienced police. They seem to make fewer stops at 9am and 3pm when the other officers are peaking and then spike hard at 11pm when the rest are decreasing.

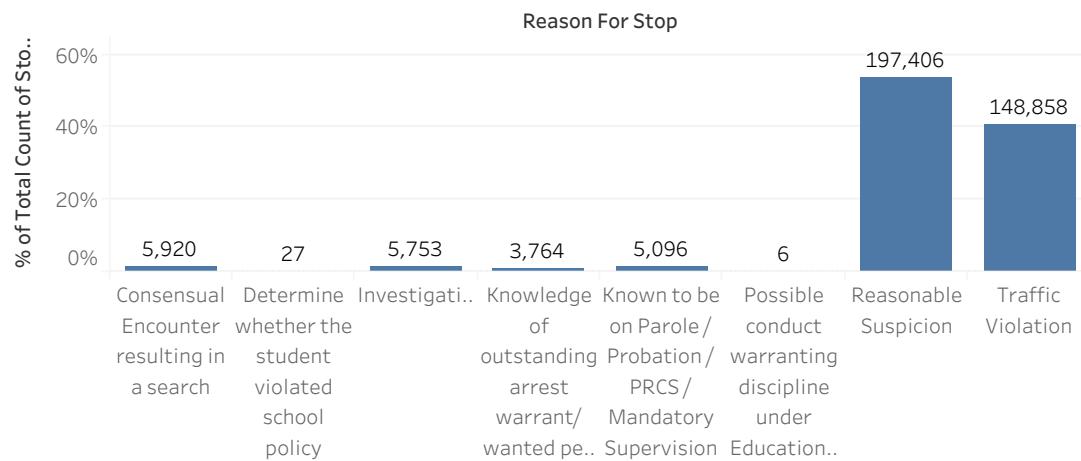


Influence of Experience



The Reason for Stop bar graph would be perfect to add segments of color based on race and compare how the proportions of minorities in each section are to their proportion of the population in San Diego to see if the police department is disproportionately stopping them on simply suspicion. However as stated before I do not know how to connect the race dataset to the stop data set although both were connected to the same workbook.

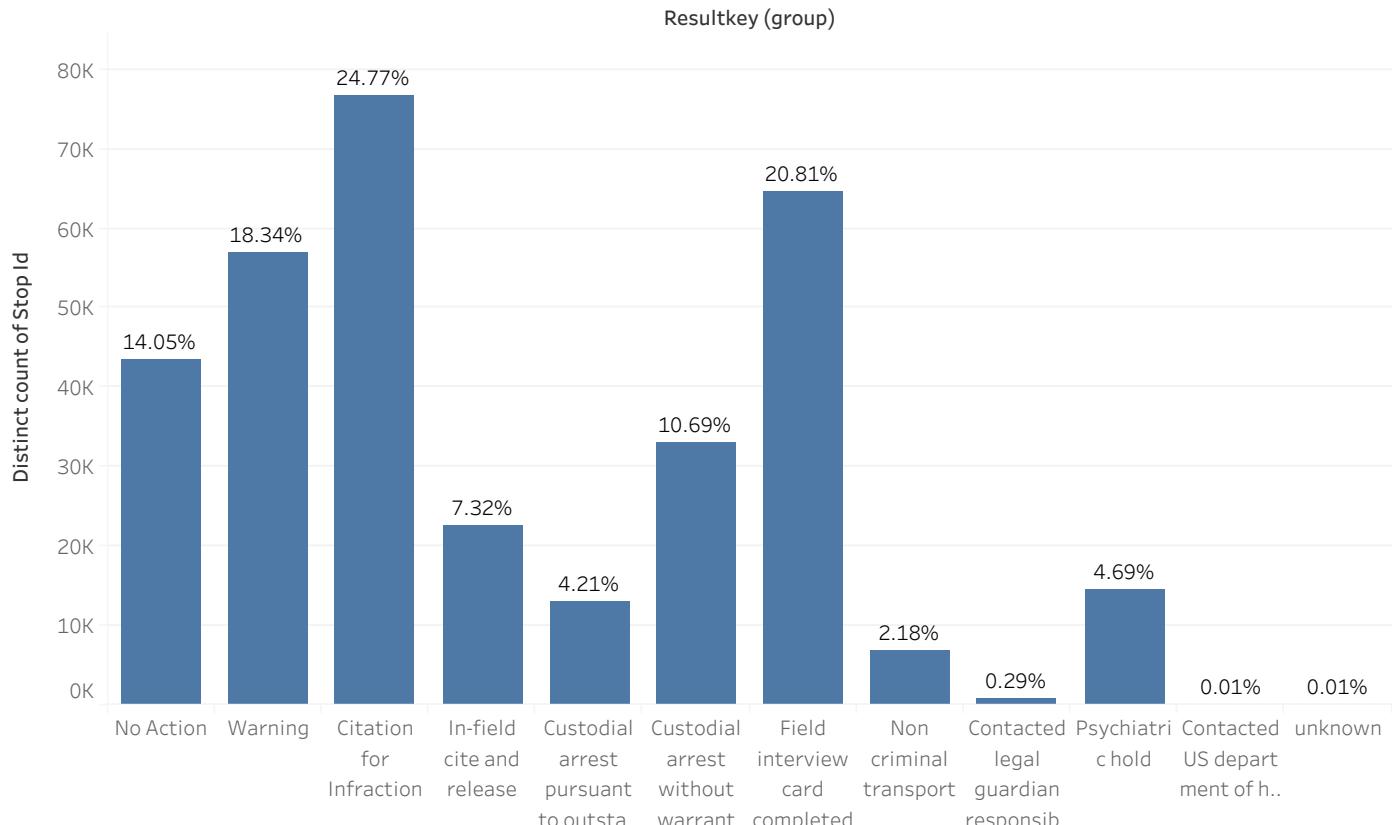
Reason for Stop



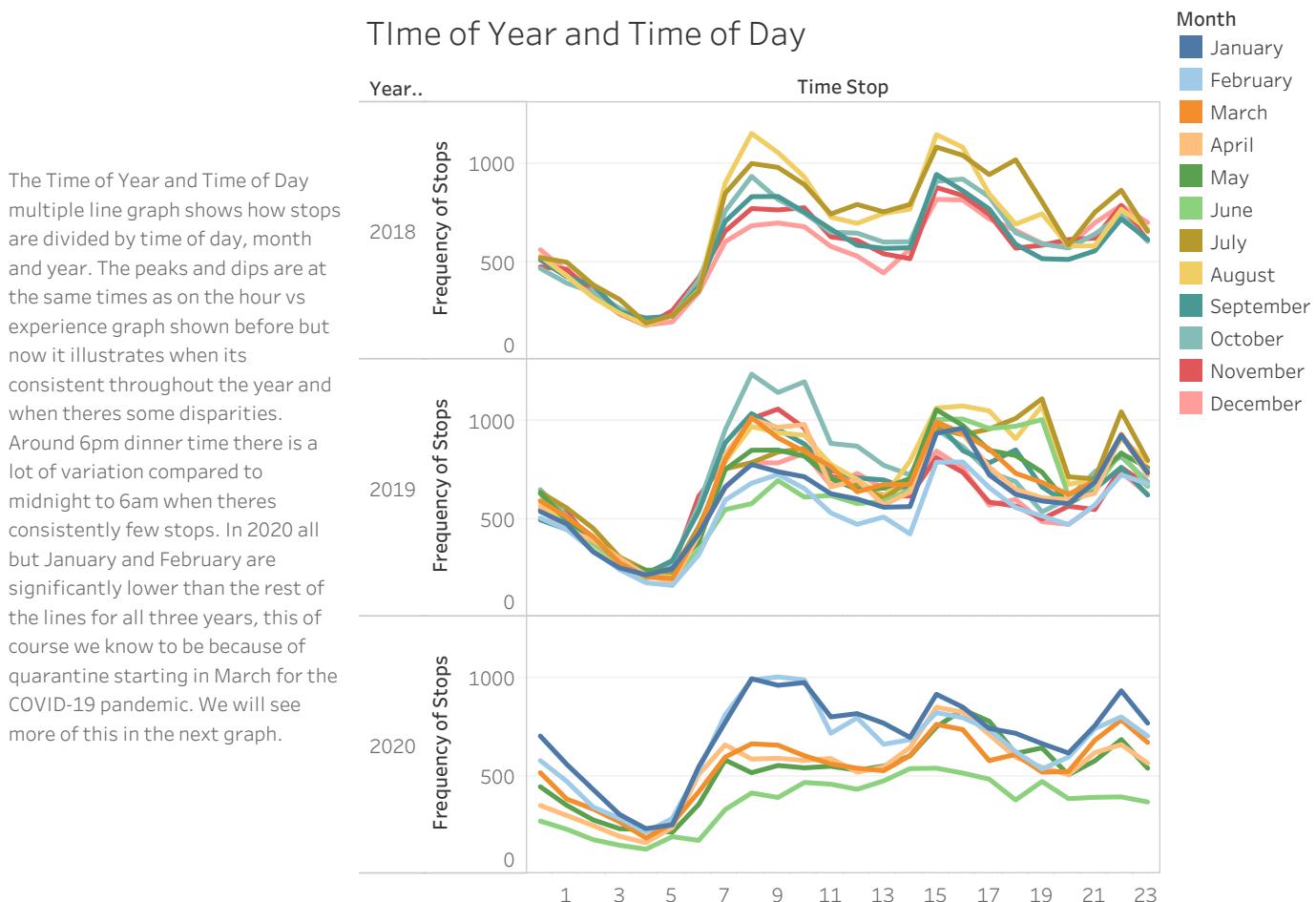
The Stop Result data here would be interesting to cross with the is student data or the officer experience data. Race could also be interesting to see if officers let off people of their own race with less severe action than others.

Without being able to cross reference those we can see that the four most common actions that sum to almost 80% of stops end with a citation at worst. Only psychiatric hold and the costodial arrests ended with taking the stopped person in.

Stop Result



The last aspect of the ripa_stops dataset I want to explore is the time and date. It is good for context and it can reveal how consistent data was collected and what kind of interval, if there are any chunks missing or any patterns.



This next timeline which shows the stops per day for the entire data set is a very granulated graph. As with the last one, this shows a general drop in stops after February of 2020. It is difficult with only two years overlapping at a time but there are dates with significant changes in count of stops which line up. For example on December 25, Christmas, both years have all time lows. One question I had was if there would be a spike towards the end of each month or every three months when officers may be trying to get enough stops to reach some sort of bonus, that is a common story I hear in neighborly talk about being extra careful driving at the end of the month. This data does not show any support for that claim.

Timeline of Stops

