

# Investigating Police Citations in Houston

Draft 2 - ggplot Visualizations

Grace Wang, Virginia Baskin, Siddhi Narayan

2023-02-26

## Introduction

Our team's dataset is sourced from the Stanford Open Policing Project and contains information on police stops that result in citations in Houston, TX from 2014-2020. The data contains a little over 2 million rows. The fields of the dataset include (non-exhaustively): the date, time, location, latitude, longitude of the stop, the police beat of the officer, district, the subject's race, sex, the type of violation they received, and the vehicle make, model, and color. Our team approached this dataset with a special interest in investigating trends regarding subject race, sex, and speeding amounts. Ultimately, we seek to investigate any factors that particularly dispose individuals towards receiving a citation.

We utilized various visualization techniques in order to comprehensively analyze every feature of the problem space and maximize various perspectives on the dataset. Our dataset has a lot of free text, which is difficult to visualize in the typical way. We visualized every field that was not unstructured free text in some way or another.

## Plot 1: Yearly Citations by Race

Police Citations in Houston 2014–2020 by Race

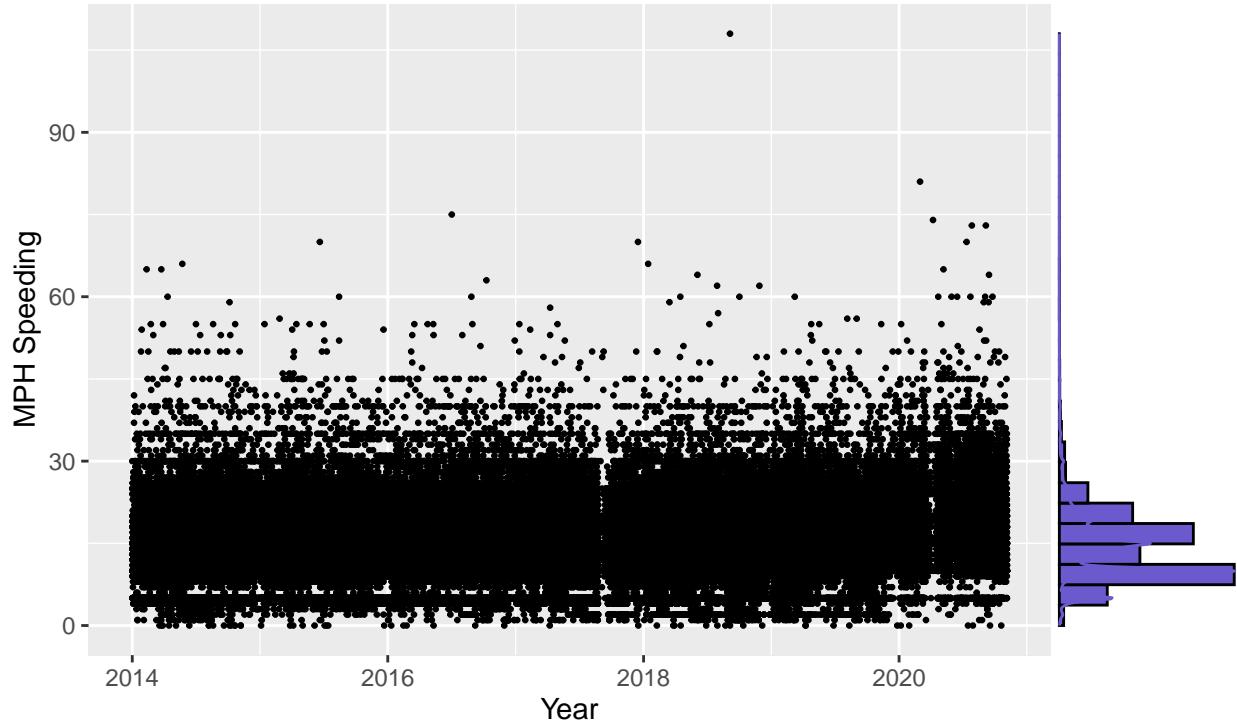


First, our team aimed to better understand the breakdown of yearly citations issued by subject race. We sought to investigate this in order to identify any patterns that might indicate racial bias or discrepancies between years in terms of citations issued to different racial groups. We observed that black and white individuals received the most amount of citations every year, while individuals of AAPI and unknown race received significantly fewer. We also observed a significant drop-off in total number of citations in 2020.

Plot 2: MPH over Speed Limit for Speeding Citations 2014-2020

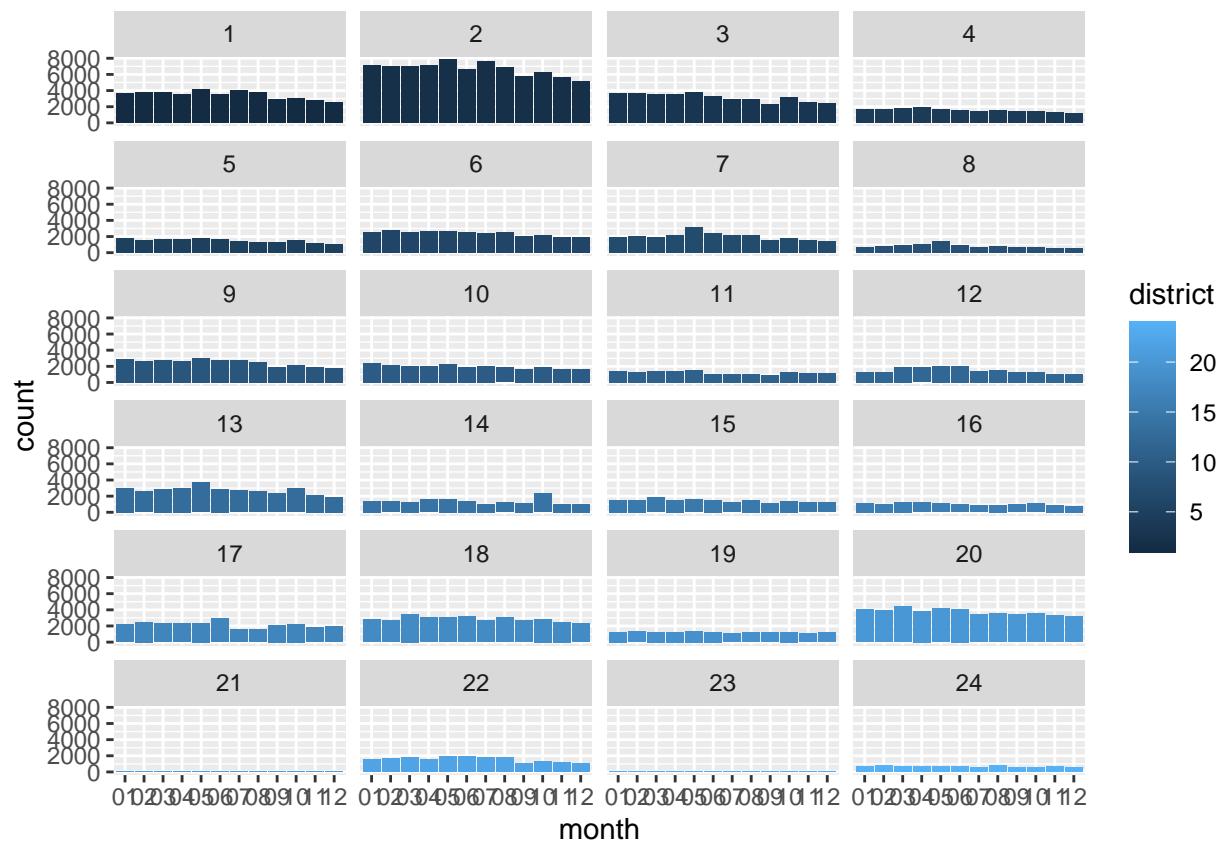
### MPH over Limit for Speeding Citations 2014–2020

#### Frequency of Citations for Different Levels of Speeding



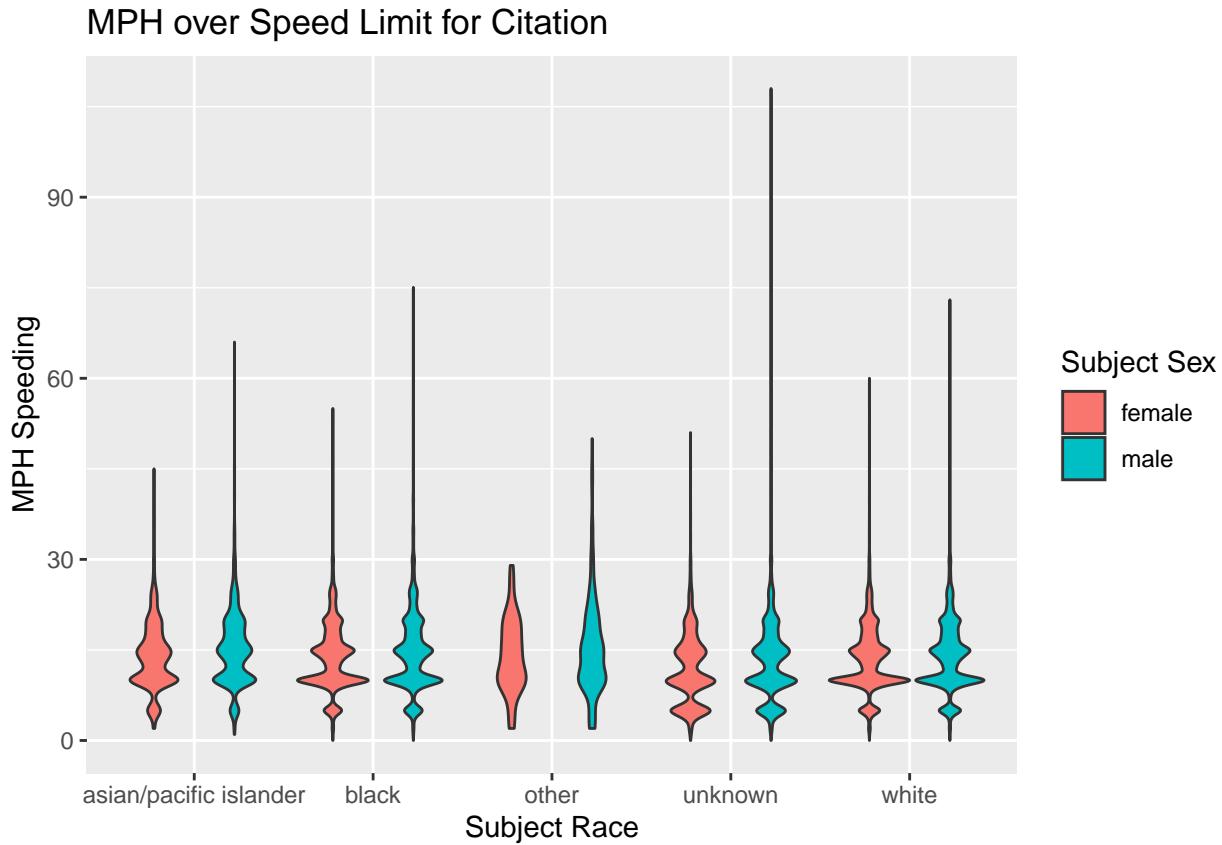
This plot investigates citations that are given for speeding. Namely, by how much were people speeding over the limit to have been issued a citation. The y axis is mph over the speed limit (as recorded on the citation). The x axis is the year/general time from the citation occurred in, and there does seem to be some evidence of temporal trends. There is a vertical line before 2018 where it seems like there were less tickets given overall, which could be a number of things that we could look into. It could be something about how the data was collected, if there was some data lost at any point, or maybe HPD deprioritized patrolling for speeding during that time period. There is no way to tell from just looking at the graph. Another interesting feature is that there seems to be less citations given for speeding less than 10mph over the speed limit 2020 and onward. It is also interesting that there are speeding citations given for going as little as 1-5 mph over the speed limit.

Plot 3: Citations by month stacked by district



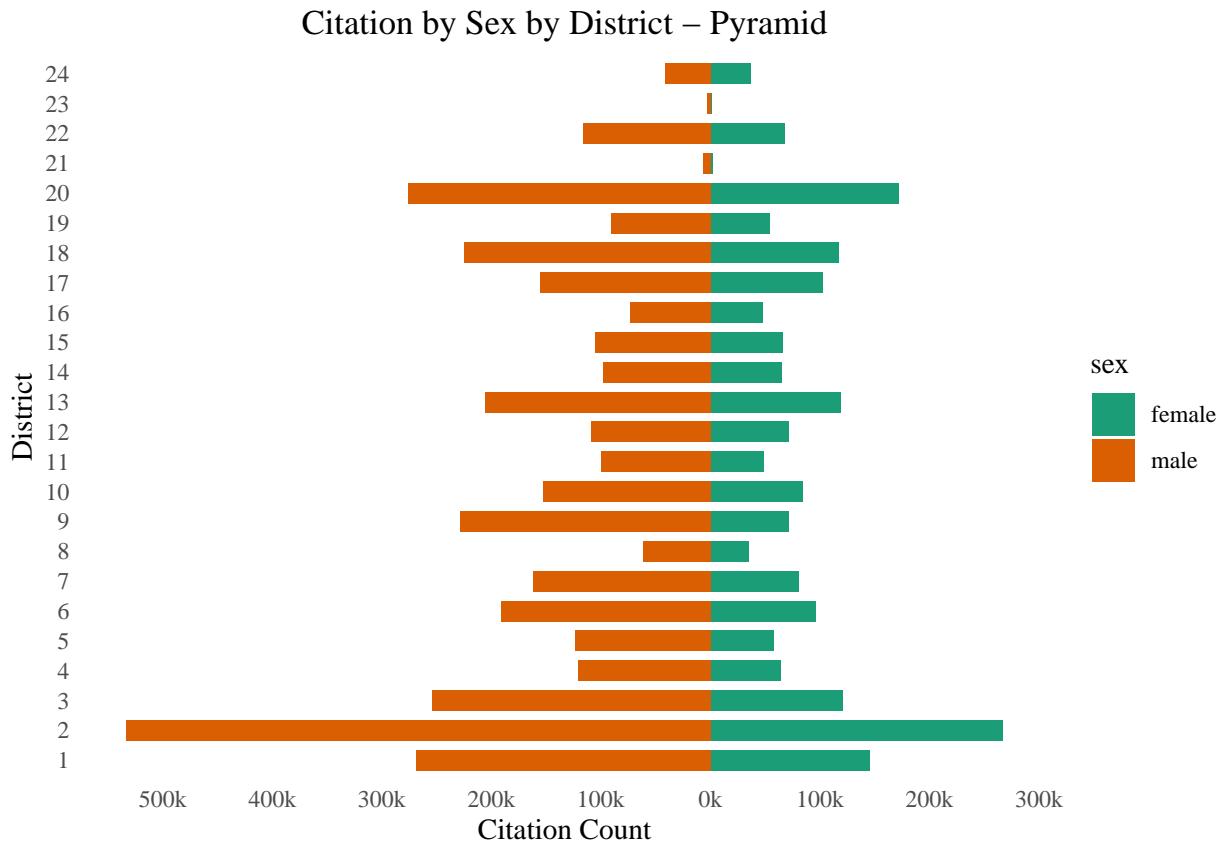
This plot describes the amount of citations per month. Each mini plot describes an individual district. District 2 has over all the highest number of citations, while District 21 and District 23 has basically no citations, (District 21 and District 23 are IAH and Hobby airports respectively, which is important to note). The distribution of citations across months is varied between districts – some of them have higher citations in the middle of the year, while others have more citations in the beginning or end of the year.

Plot 4: MPH over Speed Limit by Race and Gender



Moving forward, we chose to visualize MPH over Speed Limit for Citation by Subject race and sex. We visualized this with violin plots, which demonstrate the density and distribution of the data. We observed that each race except for “unknown” had a fairly similar distribution of speeding amounts, with the greatest peak around 10 MPH. Distributions of speeding amounts for each sex within each race were also very similar, although males of every race reached a greater maximum speeding amount than females.

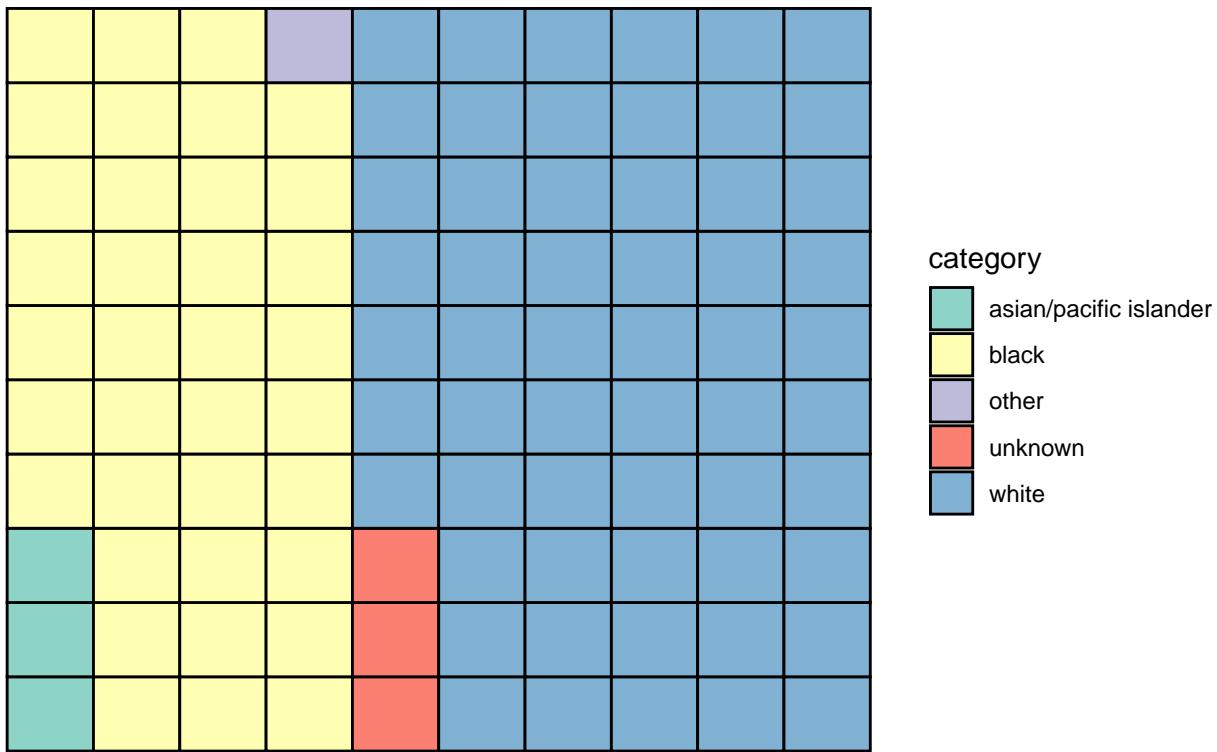
Plot 5: Citations by District and Gender



We created a “population pyramid” of citation count by district and by sex. The x-axis is the amount of citations in thousands, and the y-axis is the police district in Houston. This plot reveals interesting aspects of our dataset regarding the proportion of citations by gender. There are more citations for men than women for every single districts. This could be because men are more likely to speed, or that women speed the same but are not given citations as often. The core reason for the dependency is unknown, but the unequal count by sex is clear to see. Also, as found before in other graphs, district 21 and 23 have the least amount of citations as they are both airports (IAH and Hobby), and people are simply unable to speed in the same capacity in airport roads as they can on a highway. District 2 has the most citations, and this corresponds to the Houston’s Greater Heights district, roughly.

## Plot 6: Citations by Race

Waffle Chart of Citations per Race



Now we will look at the % of Citation by Race across all years 2014-2020 in a more geometric representation of the percentages. As we can see, white individuals have the highest percentage of citations, followed by black individuals, and then Asian/pacific islanders, then unknown races. This is somewhat proportional to the actual demographics of Houston (according to the U.S. census) – the White and Black percentages of citations are slightly higher than their population in Houston, while the proportion of citations for Asian individuals is slightly lower than the proportion of Asian people in Houston. This graph is a fun (and new, for us) way to visualize which races make up what proportions in our dataset.