NYC Stop-and-frisk: a racial perspective

This website presents the final project in DTU course Social data analysis and visualization 2021. The background code can be accessed on the link below, where the data analysis process is presented, along with a brief overview of visualization, methodology and discussion.

[Explainer notebook]

NYC Stop-and-frisk: a racial perspective

The New York Police Department is world-known for its interaction with the people of New York. During 2011 the police performed more than 600.000 stops on the street in what is known as *stop-and-frisks*. Most of the stopped people were innocent. But is there a racial bias in who is stopped? And if there is, is this bias justified by data? The NYPD log these stops in a database accessible to the general public. This project goes into the dataset to see if patterns of racial profiling can be found.

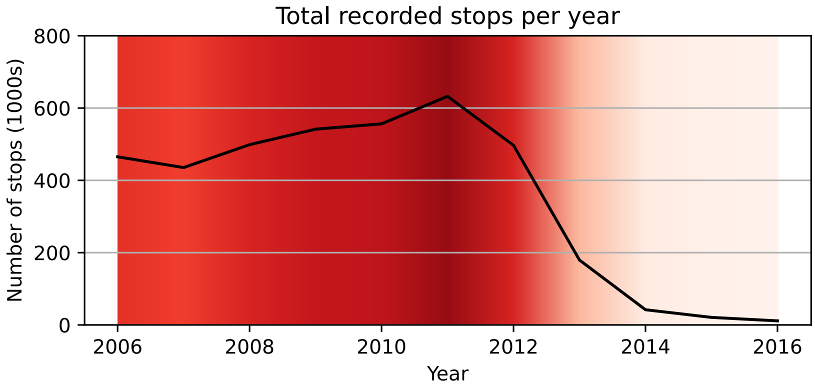
Stop-and-frisk

Terry v. Ohio, a U.S. Supreme Court decision from 1968, is the legal basis for stop-and-frisks. It held that the police stopping pedestrians for questioning and patting down for weapons or drugs without probable cause does not violate the Fourth Amendment. A suspicion based on gut feeling is all that is necessary for the police to have a legal reason to stop, question and frisk individuals on the street.

Under Michael Bloomberg's time as mayor during the early 2000s, initiatives were put in place to fight crime in New York City. A key elements in this, was the increase in stop-and-frisks in New York City. This initiative has since been criticized in media and research for being racially biased. Multiple studies find a racial imbalance in unsuccessful stops, and others have argued that the stop-and-frisk initiative has had a limited limited effect in preventing crime. The negative effects of racial profiling have been argued to have done more harm than good, with documented increases of distrust in government among marginalized groups.

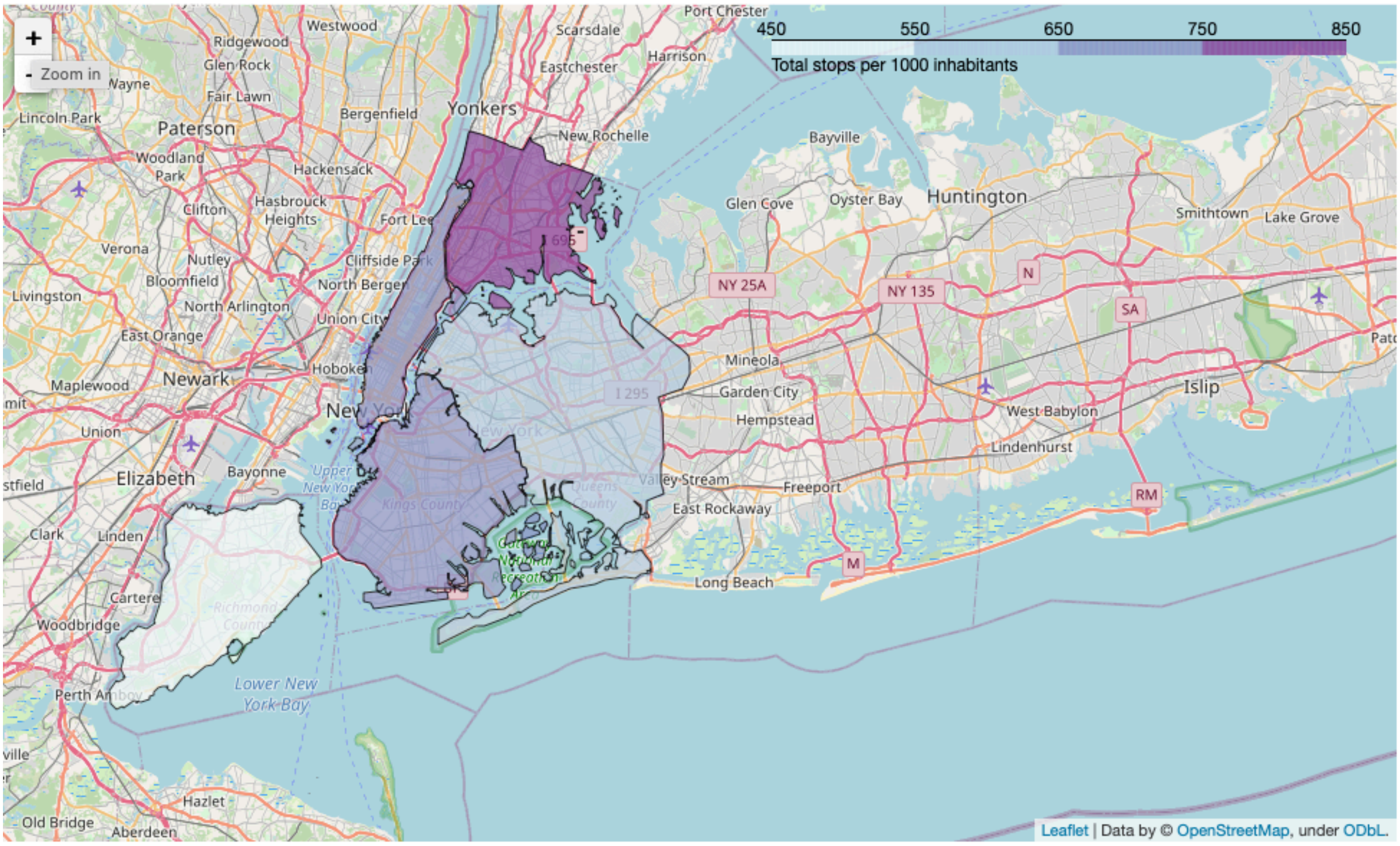
Data

This project evolves around a dataset covering all recorded stops from 2006 up to and including 2016. During these 11 years of data, the NYPD performed more than 5 million stops in total. Studies suggest that the true number is higher, with estimates up towards 15 million stops, the discrepancy caused by incomprehensive registration by officers. For each stop, an array of attributes are recorded, such as spatio-temporal data (time, year, precinct, city), personal data (age, sex, race) and criminal data (was force used, did it result in arrest, did the officer find weapons on the suspect). To simplify the analysis, only stops with the race designation Black, white and Hispanic are used, and only a subset of attributes were kept. The cleaned dataset constitutes 3.880.159 entries with 12 attributes. The kept attributes are: year, precinct, arrested, frisked, searched, was a pistol found, was force used, sex, age, race, city and criminal clothing (short for the official attribute in the data "reason for stop - wearing clothes commonly used in a crime".)



Temporal evolution of stops

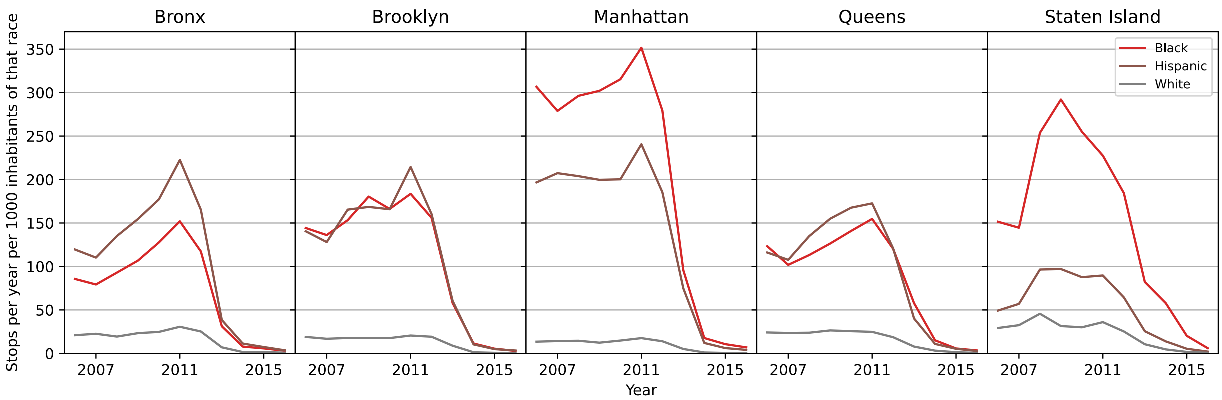
The total number of stops peaks in 2011, where more than 600.000 stops were performed. Since then, the number has been been steadily decreasing. As of 2016, the number of stops is negligible compared to earlier years, with a total of 11.338 stops. How has this significant decrease in total stops affected the performed stops? Is a higher proportion being searched and arrested? Are there different developments when dividing data by race? And are there different temporal patterns in different parts of the city? These are some of the questions this project tries to shed light on.

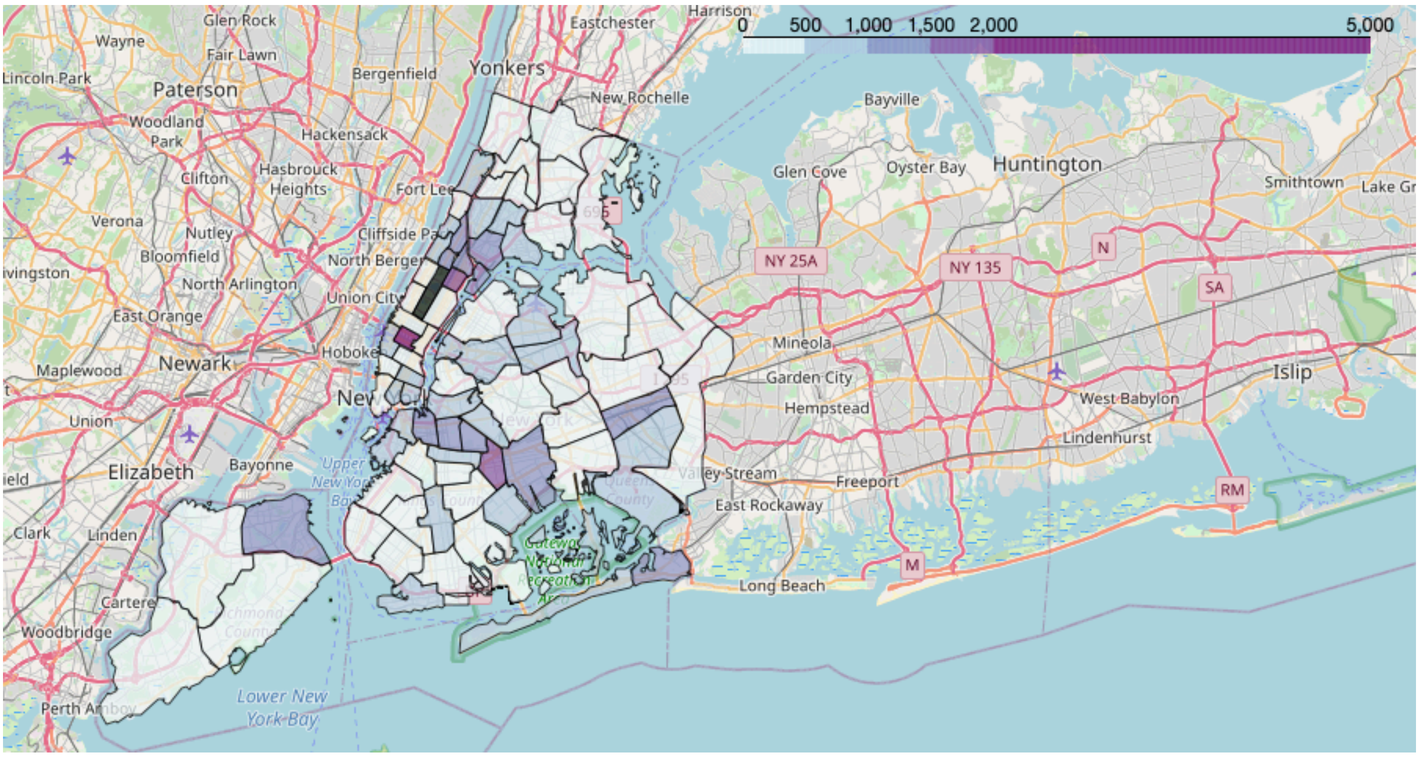


Borough distribution

New York City covers five cities: Bronx, Brooklyn, Manhattan, Queens and Staten Island. Their extent is shown on the map above, coloured by the total number of stops per 1000 inhabitants over 11 years. Clicking the image takes you to a more detailed view. In all the boroughs, there were between 450 and 850 stops per 1000 inhabitants. In particular the Bronx in the north has a higher stop rate than the other boroughs.

Below, a more detailed look into the different boroughs is presented. The graphs depict the temporal evolution of number of stops for each race, normalized to the population of each race in each city. There is a remarkable racial signal in these data. The extreme was in Manhattan in 2011 when 350 Black people were stopped for every 1000 Black people residing in Manhattan. For white people, the number is around 30, and thus the difference is of an order of magnitude. In the other boroughs, there are similar patterns, while the contrasts are smaller. The temporal development has the same general trends for all races in all neighborhoods.





Precinct distribution

On the map above, a more detailed geographical breakdown of crime prevalence is shown, here split into the different precincts of New York (precincts are a division of NYC into smaller segments used in policing). In this view, the differences are more pronounced, with a few high concentration areas. In particular, precinct 14 (the darkest colored) sticks out from its surroundings with almost 5000 stops per 1000 inhabitants over the 11 year, compared to less than 500 in the neighouring precincts. This precinct is called Midtown South Precinct, and encompasses Times Square, Madison Square Garden and Grand Central Terminal. Note that Central Park is not included, as the number of crimes registered there does not reflect the census count (25 residents in 2010). Clicking the map enables interactive exploration of the data.

Temporal development of rates

How many people are frisked and searched, how often do the police use force, and how many stops end in arrests? Below, the temporal rates for each race are depicted for NYC as a whole. Click through the tabs to see the temporal development of each rate. Common for all rates is that the trends are very similar for all three races and that from 2013 and onwards, all rates increase drastically. This coincides with the fact that the total number of stops is significantly decreasing from 2013 and onwards, presumably leading to performing stops that have more reason in actual suspicion than earlier. This explains why the rates are increasing while the amount of stops is decreasing.

Arrests Frisks Force uses Searches

Force use rates

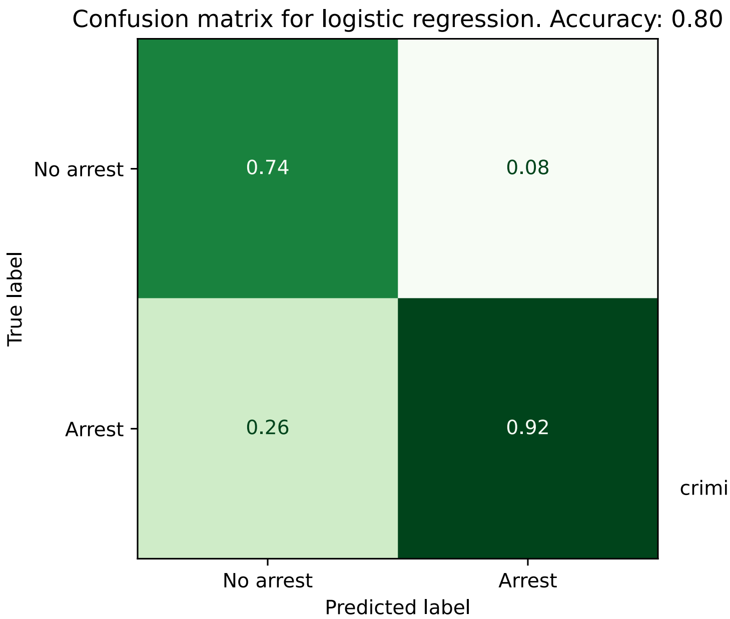
The force use rate is the proportion of stops that result in force being used by the police officer. Here, all force use has been combined into a single binary variable, and it covers a wide range of actions from "suspect against wall" to "weapon pointed". There is a notable difference in the rates, with white people being subject to force use 17% of all stops, compared

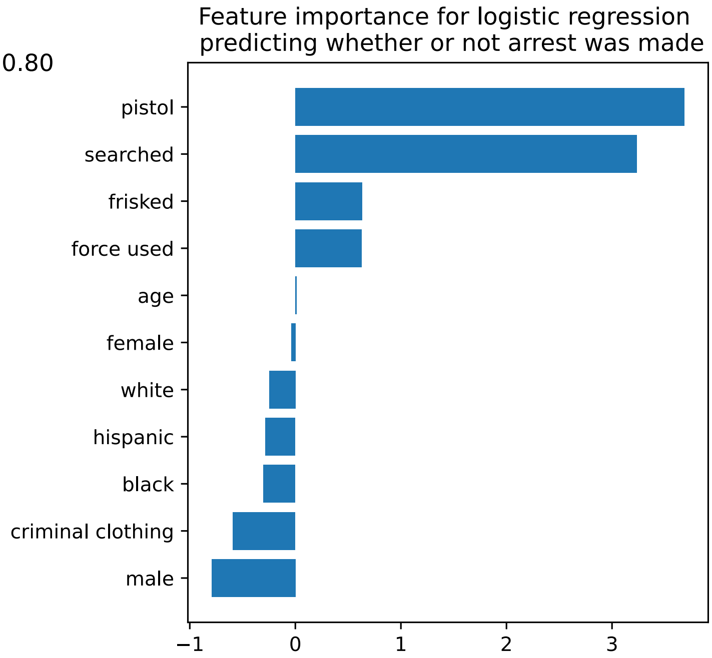
[Click here for interactive Bokeh plot of for force use rates]()

Machine learning approach

The approach until now has been centred around summing, counting and dividing to get insight into the data. Machine learning is a different approach to answer some of the same questions. The idea here is to feed the attributes to a logistic regression and train it to predict whether or not a stop resulted in an arrest. This is a crude approximation of whether or not a stopped person was innocent or not. Using logistic regression provides insight into what variables contribute to whether or not an individual was arrested.

Short explainer of machine learning, logistic regression and training a model:
A machine learning model can make predictions of some outcome, based on some input. This is the case with the logistic regression model, which based on what input it is given (such as was a pistol found, was force used, was the suspect a white male), it will output whether or not the model thinks that the suspect was arrested. Training a model is using a lot of known cases from the dataset, to tune the parameters of the model so that the model is as correct as possible. The accuracy of the model can then be quantified, by leaving some of the data out of the training data, and use them to test whether or not the model is able to predict. The parameters in the model then tell us something about the determination process of the model, and are thus an insight into the statistics of the dataset.





The **confusion matrix** for the trained model is shown above, along with the **coefficients** of the logistic regressor. The confusion matrix is a visualization and quantization of how well the model is at predicting the outcome of an arrest based on the 11 attributes fed to it. The diagonal elements are the true positives and true negatives - thus, for 74% of the predictions of no arrest, the model was correct, with the same true for 92% of the model's predictions of arrest. Based on this and the accuracy of 80%, the model is accepted for further analysis.

This leads us to the learned coefficients. A high coefficient value for a given attribute means that if this attribute is true, then the model will be more likely to predict an arrest. This is the case for the attribute pistol, which indicates whether or not a pistol was found on the suspect. The same is true for the attribute searched, which is whether or not a suspect was searched. These are obvious characteristics - if a stop ends in a search, and in particular, if a pistol is found on the subject, there is a higher probability of an arrest being made, compared to a stop where no search is performed, and no pistol is found. These are *not* the interesting coefficients. They do, however, support the use of the model, as its coefficients follow along with the expected outcome.

Looking towards the bottom three attributes, we find here negative values for Black, criminal clothing, and male. The fact that these are negative means that if one or more of these attributes are true for the stopped individual, there is a higher probability that a stop will *not* end in an arrest. Thus, a Black male wearing "clothes commonly used in a crime" is more likely than others to be stopped with no resulting arrest. The interpretation of this is that there is a selection bias in police stops, which is *not* supported by statistical evidence in this dataset.

Discussion and outlook

The data cover only recorded stops and thus does not cover who was *not* stopped. Further, out of the scope of this project is a thorough statistical analysis that covers the significance of the results found. Finally, causality is not investigated in any way, and no judgements of criminal traits of races, nor judgments of structural racism among police officers are made.

With that said, the data supports that for Black people, there is a higher frequency of frisks and use of force from police officers. Further, the logistic regression shows that for stops with registered Black males with criminal looking clothing, there is a lower probability of being arrested. This suggests that there is some unjustified racial profiling is who is getting stopped by police.

In the light of political movements in 2020, this is a highly sensitive topic, and while these data do not cover the most recent years, the data suggests unjust treatment of individuals suffering from being targeted by police, for just being who they are.

Further work with this type of data could include a deeper investigation of the individual precincts to see if other patterns emerged here. Are there certain precincts where people of a particular race are more often stopped with no following arrest? Are there temporal trends in, e.g. frisk rates that are different in certain neighbourhoods? Including theory from sociology and criminology would allow going into causality and the effect of these stops, and within the framework of law, it might even be possible to find evidence of illegal structural racism and racial profiling in police work. With the use of more advanced software and coding, the dataset could be made more open for exploration by the user, which could allow new insights to be had. Further machine learning work could discover other connections among variables in the data, shedding more light on who is getting stopped.

Finally, an actionable outcome could be a framework for police officers to help them decide whether or not an individual should be stopped based on observable traits prior to stopping. Machine learning and statistical analysis would be the engine behind this framework, and the actual implementation for use on the street could take many forms. This could potentially lead to decreasing the amount of innocent people being stopped on the street. The effects of this might be an increased trust in police and government among groups subject to profiling based on their observable traits. This, however, is out of the scope of this project and will be left as an option for the reader.

Sources

https://www.britannica.com/event/Terry-v-Ohio

http://www.citymayors.com/mayors/new_york_mayor.html

https://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page

Ferrandino, Joseph. “Minority Threat Hypothesis and NYPD Stop and Frisk Policy.” Criminal Justice Review, vol. 40, no. 2, SAGE Publications Inc., 2015, pp. 209–29, doi:10.1177/0734016814564989.

Cooley, Erin, et al. “Racial Biases in Officers’ Decisions to Frisk Are Amplified for Black People Stopped Among Groups Leading to Similar Biases in Searches, Arrests, and Use of Force.” Social Psychological and Personality Science, vol. 11, no. 6, SAGE Publications Inc., 2020, pp. 761–69, doi:10.1177/1948550619876638.

Goel, Sharad, et al. “Precinct or Prejudice? Understanding Racial Disparities in New York City’s Stop-and-Frisk Policy.” Annals of Applied Statistics, vol. 10, no. 1, Institute of Mathematical Statistics, 2016, pp. 365–94, doi:10.1214/15-AOAS897.