Detecting disparities in police deployments using dashcam data

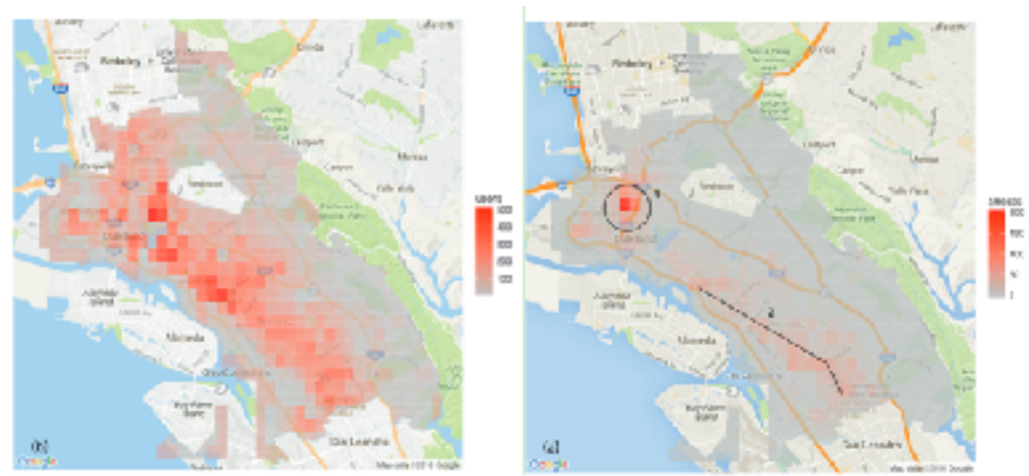
ACM FAccT, ’23

Matt Franchi, J.D. Zamfirescu-Pereira, Wendy Ju, Emma Pierson
Cornell Tech

Welcome! I’m Matt Franchi, and today I’m going to give an overview of our work in *Detecting disparities in police deployments using dashcam data.*

Background

Police deployments can affect downstream outcomes.



There's a significant body of research utilizing large-scale policing data to evaluate efficiency and equity. Interestingly, this work primarily focuses on the events unfolding after police are deployed - the so-called downstream outcomes - rather than where the deployments happen.

This downstream analysis explores topics such as disparities in police stops across different racial groups, outcomes post-stop like granted leniency, differences in police search practices, and the respect demonstrated by officers to drivers of diverse races.

Key algorithms employed in criminal justice and policing, like predictive policing algorithms and pretrial risk assessments, rely heavily on this downstream data, which includes arrests and convictions. This means that biases in police deployments can infiltrate and distort the data these algorithms use, influencing downstream outcomes. A case in point is Lum and Isaac's "To Predict and Serve?" which postulates a feedback loop where higher deployments inflate arrest rates at steady crime levels, leading to suggestions for even greater police presence.

Policing disparities quantification extends beyond the academic world, encompassing journalists, activists, and legal experts. However, a significant gap persists – the accessibility of police deployment data nationwide. The significance of this data is two-fold.

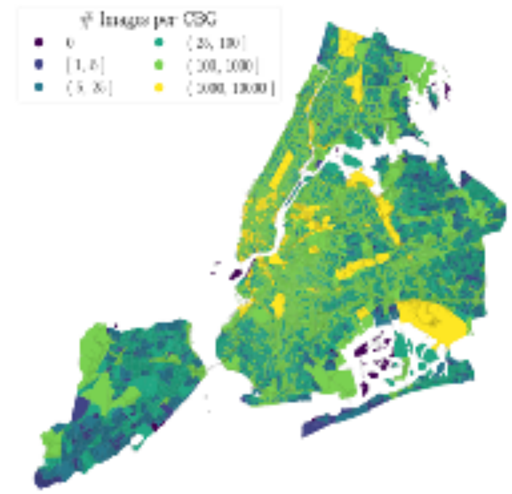
Firstly, statistical scrutiny of deployments can expose systemic inefficiencies and inequities, highlighting areas with a disproportionate police presence. Secondly, as we touched on earlier, any disparity in police deployments can lead to biases in downstream outcomes, such as arrests, impacting wider law enforcement results.

Responding to this call for accessible and transparent auditing of police deployments, we propose a unique solution: employing dashcam images from public street scenes to detect police vehicles. This innovative approach allows us to monitor police deployments effectively and transparently, offering a fresh lens to understand and address policing disparities.

Dataset / ML

25 million

Street-level dashcam images throughout the 5 boroughs of NYC



To execute our methodology, we harnessed a dataset of roughly 25 million dashcam images, spanning all five boroughs of New York City, captured during the 2020 timeframe. Out of this, we handpicked 9,500 images for training our model, annotating them based on the presence or absence of police vehicles.

For machine learning, we deployed a YOLO-based deep learning model, which showed high proficiency in identifying police vehicles. The first 1,000 confirmed true positives were sourced from Scale.AI, a platform similar in function to Amazon Mechanical Turk. A prototype model was used to identify further positives, and all true and false positives were manually verified before joining the training set.

In summary, our final training set balances 4,750 images with police vehicles against an equal number of images without police vehicles, allowing us to maintain equilibrium and fairness in the training process.

Compensating for biases.

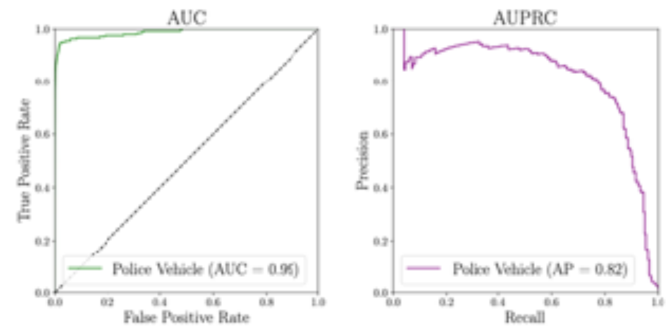
- **Non-representative locations:** reweigh the image distribution to match population distribution.
- **Imperfect classifier:** estimate + compensate for imperfect FPR/FNR. Test for disparities in performance across subpopulations.
- *Other biases are unavoidable (more later).*
  - Police vehicles are partial proxy for policing.
  - Subway policing is out-of-reach.
  - Holes around protest / other inordinate events.

Our data comes from Nexar, a firm that provides ride-sharing drivers with dashboard cameras, covering all five boroughs of New York City. However, this doesn't equate to a completely random roadside sample of the city. Consequently, we've put significant measures in place to neutralize dataset biases.

Primarily, we modified the image distribution to echo the population distribution. To put it in context, when calculating the police deployment levels Asian residents encounter, we adjust our data sample to oversample areas with larger Asian communities. We delve deeper into the mathematical intricacies of this procedure in section 4.1 of our paper.

Additionally, we scrutinize performance disparities across different subpopulations, ensuring any variations are minimal. The specifics of these results, along with classifier outcomes, will be shared later. This approach helps us create a more balanced, representative analysis of police deployments, compensating for the inherent biases in our data source.

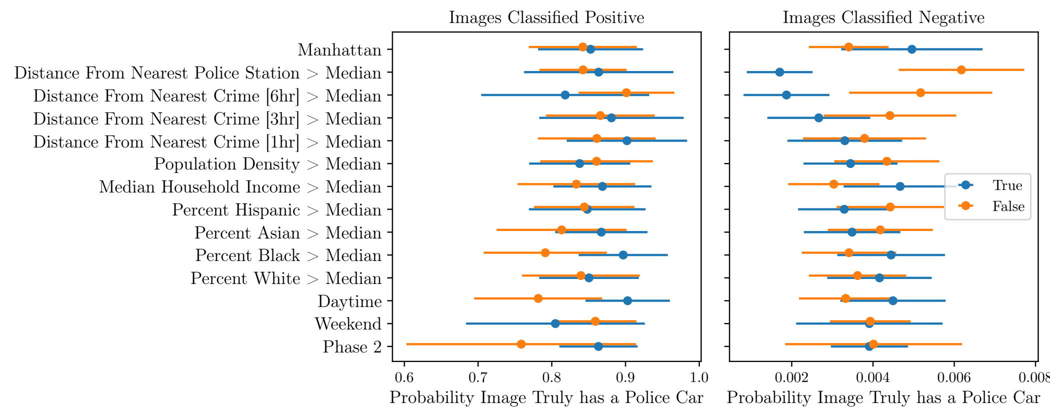
Classifier



Now, let's take a look at the classifier. We used a YOLO object detection model, which has proved reliable and effective across our tested subgroups. When we set our positive classification threshold at 0.77, it identified 233,596 images with a police vehicle from almost 25 million total images.

The figure to the right offers a glimpse at a few of these positive detections. As you can see from the top left image, our model can even handle challenging situations like occluded or small vehicles with reasonable accuracy.

Consistent performance across demographic subgroups



On this slide, you can see how our model performed across various subgroups, evaluated on the test set. The key thing to note is the gap between the orange and blue dots. If it's smaller than the line length indicating uncertainty, we can't say there's a statistically significant difference in model performance across that subgroup.

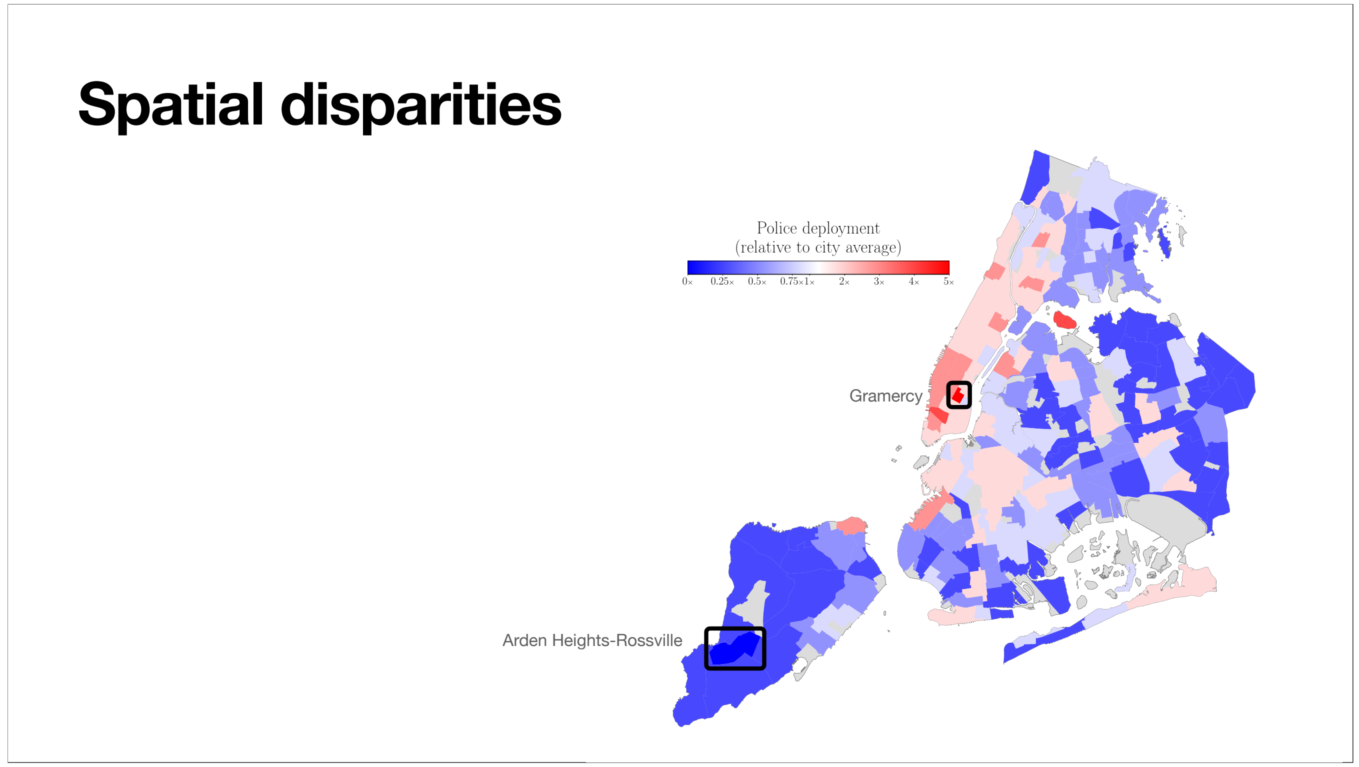
Further, we cross-referenced police detections with external data to check for anticipated correlations. We found that areas with images closer, on average, to crime incidents or police stations were indeed more likely to have police vehicles detected – just as expected. We confirmed these correlations with a Pearson correlation across Census Block Groups, giving us a p-value less than 0.001.

This extra validation step gives us confidence that our model doesn't just perform well, but it also aligns with the correlations we'd predict between police activity, crime incidents, and proximity to police stations.

Results

Estimand: Probability residents of a given group (e.g., Asian residents) have a police vehicle visible when they are on a street in their home neighborhood.

Let's get into the results! We've based our analysis on a central metric: the likelihood that members of a certain group, say, Asian residents, will spot a police vehicle when they're out on a street in their home neighborhood. This measure syncs well with the street-level perspective offered by our Nexar dataset.

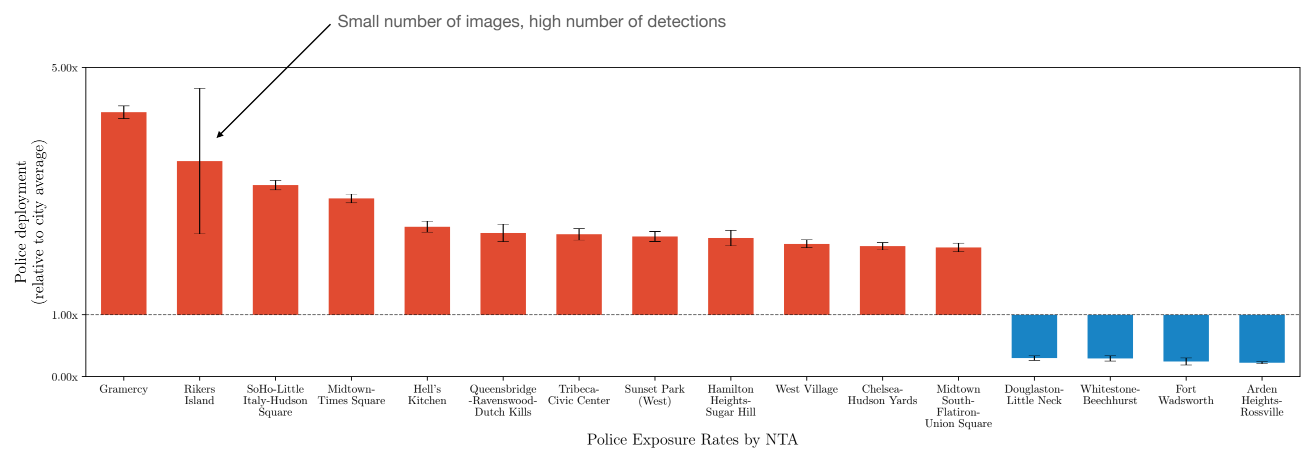


We're assessing variations in police deployments by groupings like race, type of zone, population density, median household income, borough, and neighborhood. Let's start with spatial disparities.

Remember, all police deployment figures are relative to the population-weighted average for the entire city. So, if a group's figure is 1.6x, that means its deployment levels are 60% higher than the city-wide average. We calculated error bars for all these estimates using bootstrapping, with the reported errors being 1.96 times the standard deviation across bootstrapped datasets.

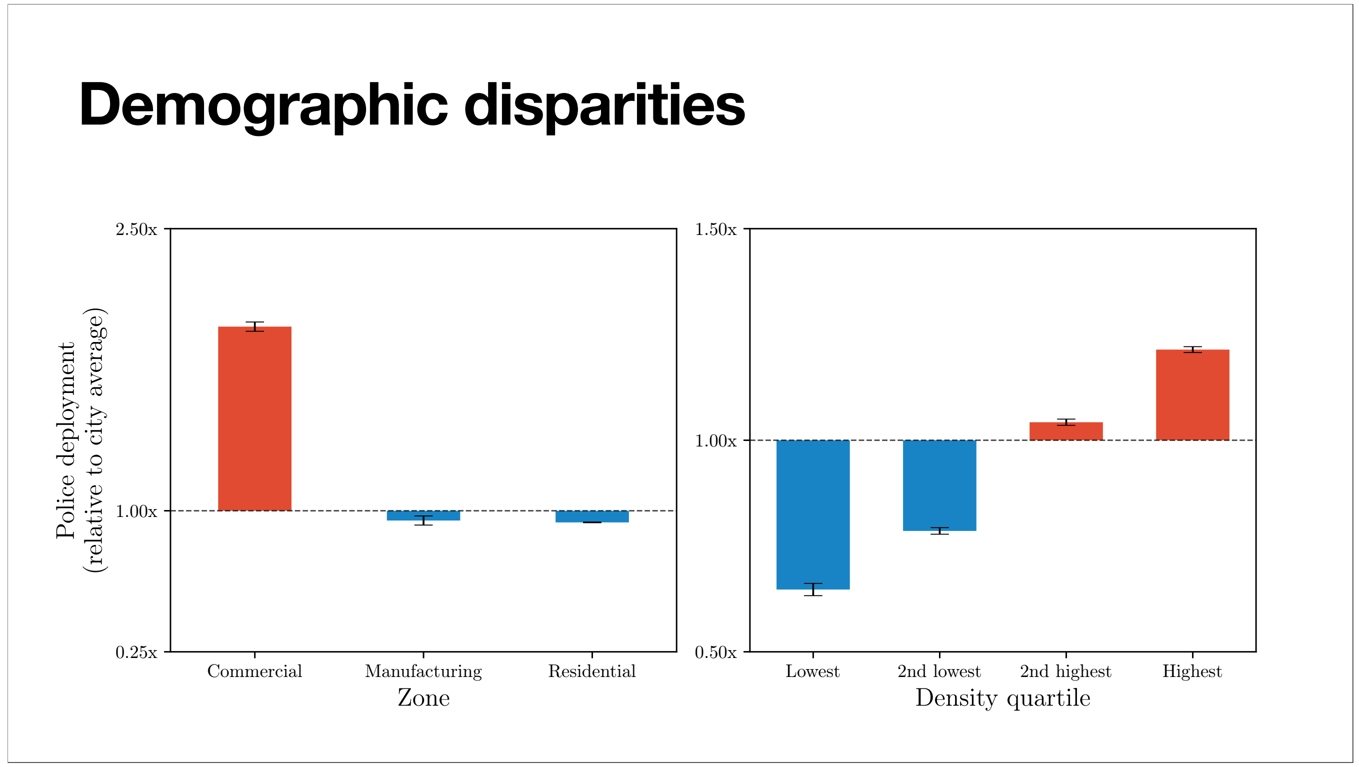
This figure gives you a city-wide overview of police deployment rates, broken down by Neighborhood Tabulation Areas or NTAs. Gray regions are areas with no population according to the Census, like airports, cemeteries, and parks. From this map, you can clearly see a wide spread in deployment rates. For instance, Gramercy, the neighborhood with the highest police deployment, is more than four times the city average, while Arden Heights-Rossville is less than a quarter of the average. That's a near 20-fold difference.

Spatial disparities



This bar graph brings into sharper focus the relative police deployment levels across different neighborhoods. A couple of patterns emerge when we look at the top twelve neighborhoods. We see both some of the wealthiest areas of the city, like Hudson Yards—dubbed a "billionaire's fantasy city"—and some of the most heavily policed, such as Rikers Island, home to a large prison complex.

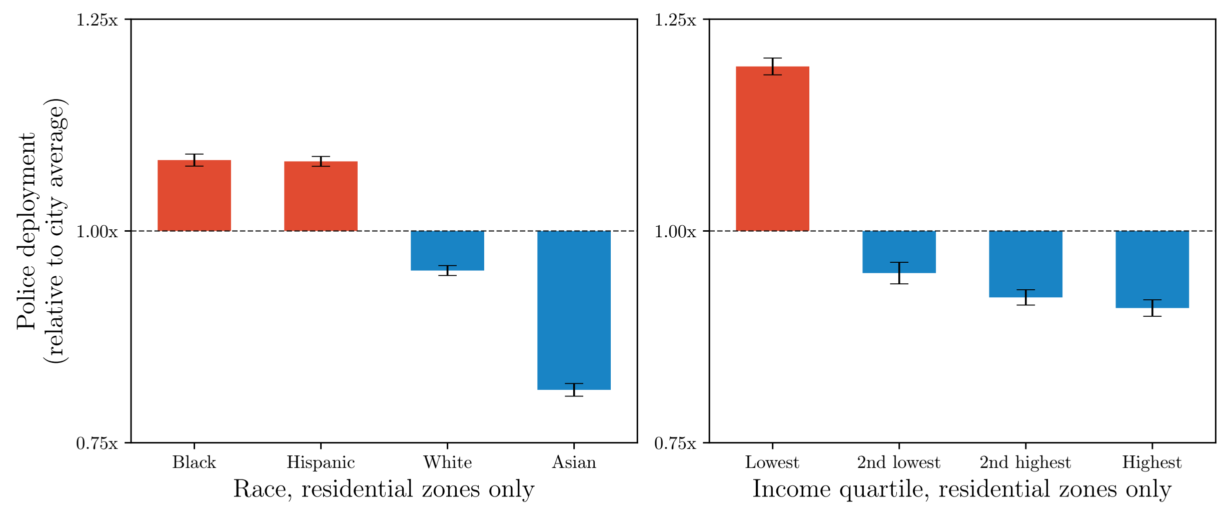
We should also consider that certain unique characteristics of a neighborhood could influence high deployment levels. For example, Gramercy has a major police training center situated on two busy roads. To dig deeper into these spatial disparities, we're going to explore correlations with zone type and population density next.



Looking at these figures, we see that commercial zones experience significantly higher police deployments—nearly double the city average—compared to residential or manufacturing zones. Adding population density into the mix reveals more insights.

These box plots show a clear trend: as population density increases, so does police deployment. In areas with the lowest density quartile, police deployment is around 65% of the city average, while in the highest density quartile, it rises to over 110% of the city average. This data suggests that areas with high police deployment levels are often densely populated, commercial regions, like downtown Manhattan.

Demographic disparities



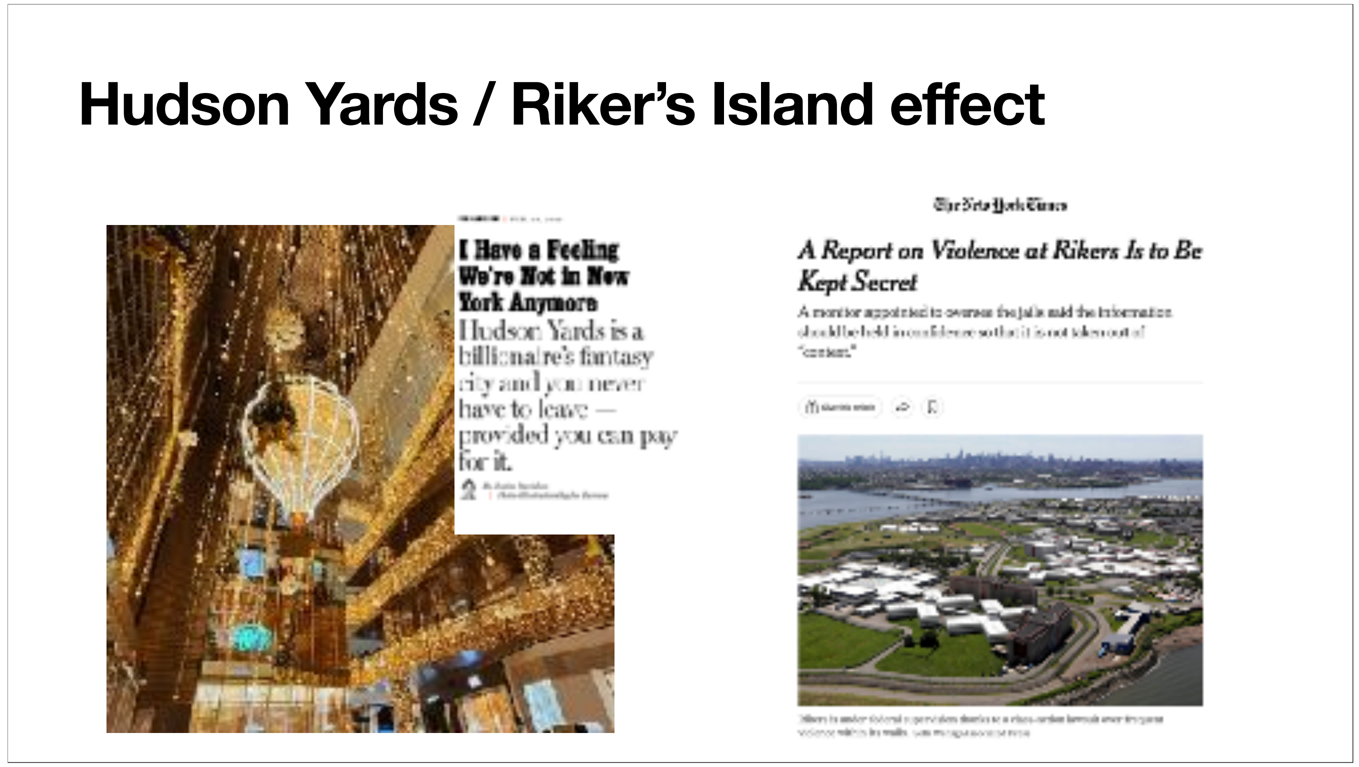
Shifting focus to demographic and economic indicators, we assess police deployment patterns in relation to racial groupings and median household income. Notably, we zero in on residential zones to best reflect where people actually live.

The first box plot shows marked racial disparities in police deployment. Black and Hispanic residents experience 1.08 times the city average police deployment, contrasted with 0.95 times for white residents and 0.81 times for Asian residents. In essence, there's a 34% discrepancy between the groups experiencing the most and least police presence. This echoes earlier observed trends in NYC where Black and Hispanic communities face heavier policing.

In the right box plot, we stratify by income quartile, revealing a trend where police deployments escalate as median household income diminishes. Residents in areas with the lowest median income encounter police deployments 1.19 times the city average, while those in the highest income quartile see 0.91 times the average.

As such, we report these results by individually stratifying each variable, acknowledging that multiple potential causal mechanisms might underlie these disparities.

Discussion & Future Work

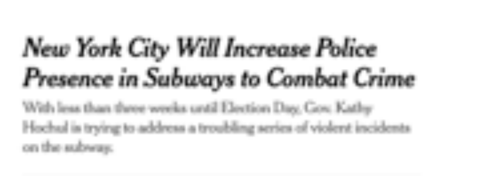


In synthesizing our analysis, we find that police deployment in New York paints a multi-layered and intricate portrait. Clear spatial disparities exist, with some neighborhoods experiencing significantly higher police vehicle deployments than others. Residential zones with higher populations of Black, Hispanic, or lower-income residents notably experience elevated levels of police presence.

Adding another layer of complexity, our findings also highlight that affluent and dense commercial districts, like Hudson Yards in downtown Manhattan, witness an increased police presence. This phenomenon has links to established research suggesting a correlation between gentrification and heightened police visibility. The dynamics of police deployment, as evidenced, extend beyond simple demographic or socioeconomic characteristics, warranting further exploration.

Innate biases

- Everybody walks in NYC, including police.
  - Police vehicles are a partial proxy
  - Subway policing
- Images *within* each Census area may be non-representative
  - e.g., possible holes in dataset around protests


New York City Will Increase Police Presence in Subways to Combat Crime

With less than three weeks until Election Day, Gov. Kathy Hochul is trying to address a troubling series of violent incidents on the subway.

(Oct 2022)

While our study offers valuable insights, it also carries inevitable limitations. Despite our best efforts to minimize bias, some inherent obstacles persist.

First, the sampling within each Census area may have its own biases. For instance, areas with protest activity during the George Floyd demonstrations might have been less accessible, affecting our dashcam image captures. We've tried to mitigate this by focusing on data from October 2020 onwards when protest activities lessened.

Also, it's essential to consider that the social dynamics during the pandemic period were unique and could have influenced our findings. Moreover, the sample collection process for the Nexar dataset remains largely unclear, leaving room for unknown biases.

Secondly, we must acknowledge that using police vehicles as a proxy for all policing activity has its shortcomings. Aspects like officers on foot, unmarked cars, and subway policing remain out of our dashcam data reach.

While police vehicles are a significant indicator of policing activity, they do not provide a complete picture. For example, our model could not effectively distinguish on-foot officers from civilians due to low-quality annotations.

Nevertheless, we believe our study provides a strong foundation and invaluable insights for future research in this domain.

Better (aggregated) policing data.

In conclusion, our research underscores the urgent need for more extensive and comprehensive deployment data from police departments. Past resistance to releasing this data, citing potential public safety compromises, is debatable.

Even aggregated deployment data can serve as a powerful tool for identifying disparities and biases, as evidenced by our study and previous research. Importantly, it's worth noting that granular data on various police activities, such as stops, searches, arrests, citations, and even shootings, is already publicly available. This fact challenges the argument that the release of aggregated deployment data could uniquely endanger police operations.

By releasing more comprehensive data, police departments could democratize the audit of policing practices, allowing for greater transparency, accountability, and ultimately, fostering a stronger relationship with the communities they serve.

Discussion & Future Work

- Our approach works, but not for the ordinary citizen.
- The police should consider releasing (aggregated) deployment data.
- **What's Next?**
  - Study variation over time?
  - Other audits with dashcam images?
    - Garbage & *Clean Curbs* NYC, older adults & accessibility

While our methodology has been effective, it's not a practical solution for everyday citizens wanting to monitor police activity. This underlines the importance of police departments releasing aggregated deployment data to the public.

The potential of dashcam data extends far beyond police audits; it holds great promise for auditing other government services and agencies for efficiency and equality. It also adds to the growing body of work in computational social science.

Looking ahead, we're keen to expand our scope. Given more data, we could examine variation in police deployment over longer and more recent periods. Beyond policing, we could use dashcam images to audit other aspects of urban life, like tracking streetside garbage for environmental assessments or evaluating street accessibility for older adults.

I'd like to end by expressing my gratitude for your attention. This presentation is part of the 2023 ACM Conference on Fairness, Accountability, and Transparency. We're looking forward to taking these insights and using them to drive further research and, ultimately, create more equitable cities.