Methodology

To test whether there is geospatial autocorrelation of crime rates across racial groups within NYC's police precincts, we considered the following hypotheses:

- Null: Crime rates and racial composition within an NYPD precinct are spatially independent of neighboring precincts.
- Alternate: Crime rates and racial composition within an NYPD precinct are spatially correlated with those of adjacent precincts.

We first constructed geospatial objects that reflect population race demographics across police precincts in NYC. To do this, we used areal interpolation to overlay NYPD precinct boundaries with NYC's census blocks. The census blocks represent a distribution of 2020 census data on race across NYC. To account for census blocks that crossed multiple police precincts, we use areal interpolation to allocate the proper proportion of each census blocks' race distribution to each precinct polygon it overlapped with. This produced precinct-level estimates for total population and racial composition (Black, Hispanic, Asian, and White residents), harmonized into a single geospatial layer.

We obtained the official public NYC dataset of crime activity over the past 19 years in the form of police encounters. To avoid distortions related to the COVID-19 pandemic, we limited out dataset to incidents before January 1, 2020. We assume that predicative policing algorithms would heavily prioritize violent crime, so we limited our dataset to violent felony offenses, which includes robbery, rape, assault, murder, arson, kidnapping, and other related categories.

Each incident in our dataset includes location information in the form of the NYPD precinct where it occurred, and latitude/longitude coordinates. We used point coordinate as our primary spatial reference, both to maximize geographic precision and to account for potential issues stemming from precinct boundary changes. All points were converted to sf geometry, filtered to the geographic bounds of NYC, and projected into a consistent CRS (EPSG: 32118). For temporal analysis, incidents were categorized as pre- and post-Patternizr to enable our comparison of spatial patterns before and after the algorithm was adopted.

Using spatial joins, each incident point was assigned to the precinct polygon that contained it. We generated precinct-level counts of total violent felonies, along with pre- and post-Patternizr incident counts, and calculated crime rates per 1,000 residents using precinct-level demographic estimates. We then merged these aggregates with the demographic layer to create a unified precint-level dataset for geospatial analysis.

To test whether crime and demographic patterns show spatial autocorrelation, we constructed a Queen continuity neighbors matrix and corresponding spatial weights. With this structure, we estimated a baseline OLS model of crime rate on racial composition, conducted a Moran's I test on OLS residuals to assess leftover spatial structure, and fit two regression models: SAR and SEM. We then implemented Geographically Weighted Regression to examine

how race–crime relationships vary across space rather than assuming a single global coefficient. Together, these models evaluate whether demographic–crime relationships are spatially patterned, whether racial composition remains predictive after accounting for geographic structure, and whether spatial clustering may indicate localized over-policing patterns.

Initial Findings

Preliminary analysis shows substantial variation in violent felony counts across NYC precincts, with crime rates per 1,000 residents varying accordingly. Precinct-level demographic patterns also differ widely, and simple correlations indicate that total crime is moderately positively associated with the percentage of Black residents ($r \approx 0.54$) and negatively associated with the percentage of White residents ($r \approx -0.68$). Because demographics and crime both vary across space, these early descriptive results suggest that spatial dependence may be present, motivating formal spatial autocorrelation tests and spatial regression models to determine whether demographic–crime relationships persist once geographic structure is accounted for.