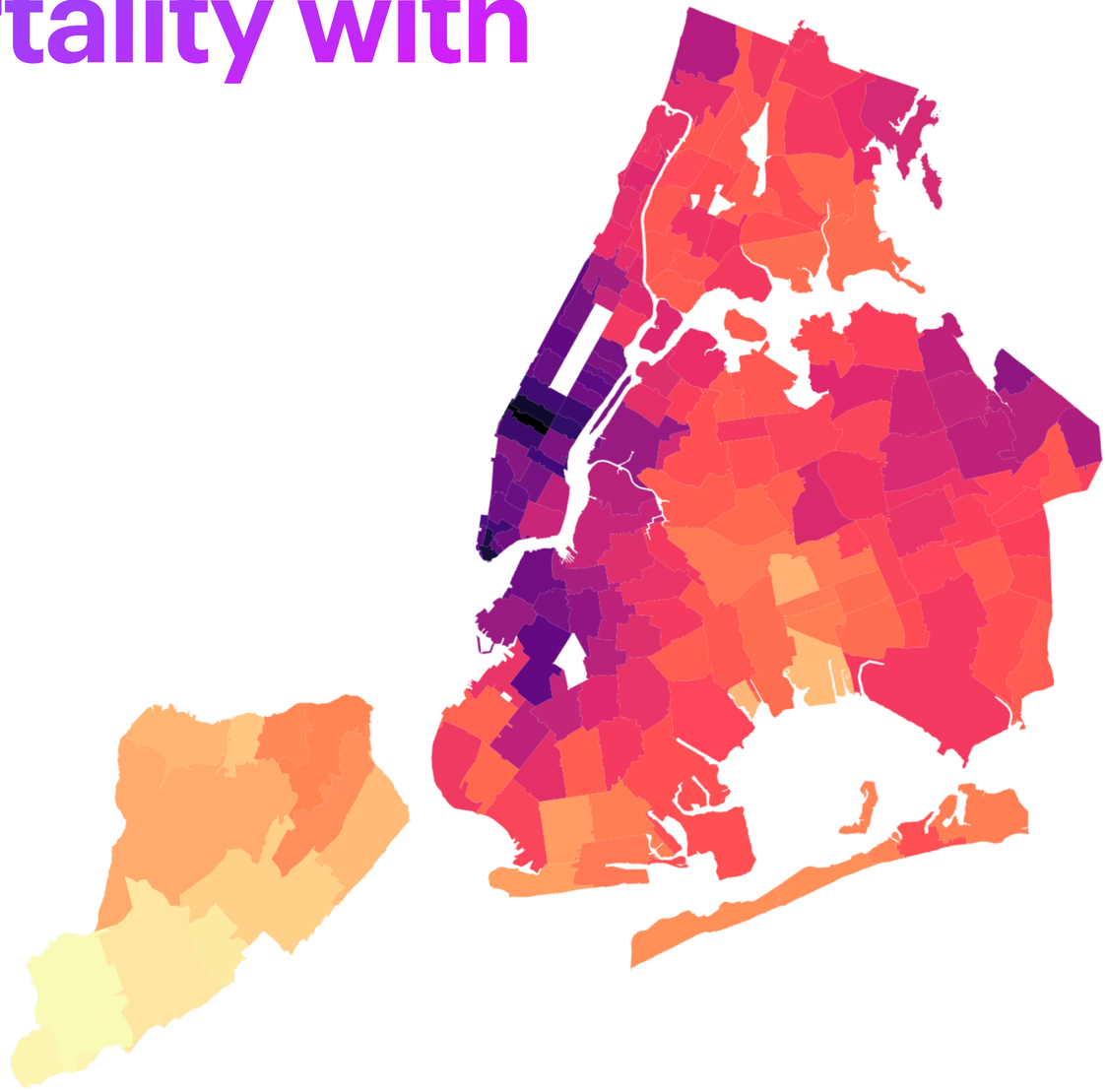
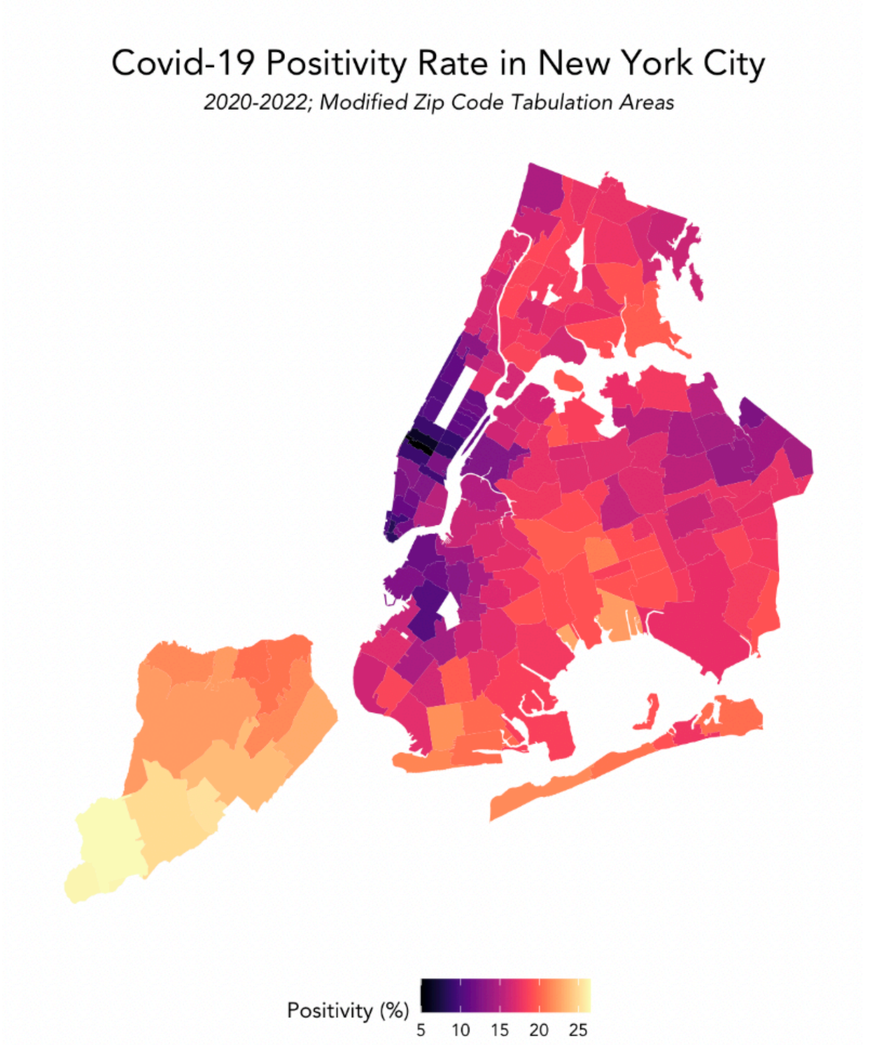
Modeling Covid-19 Mortality with Spatial Risk Factors

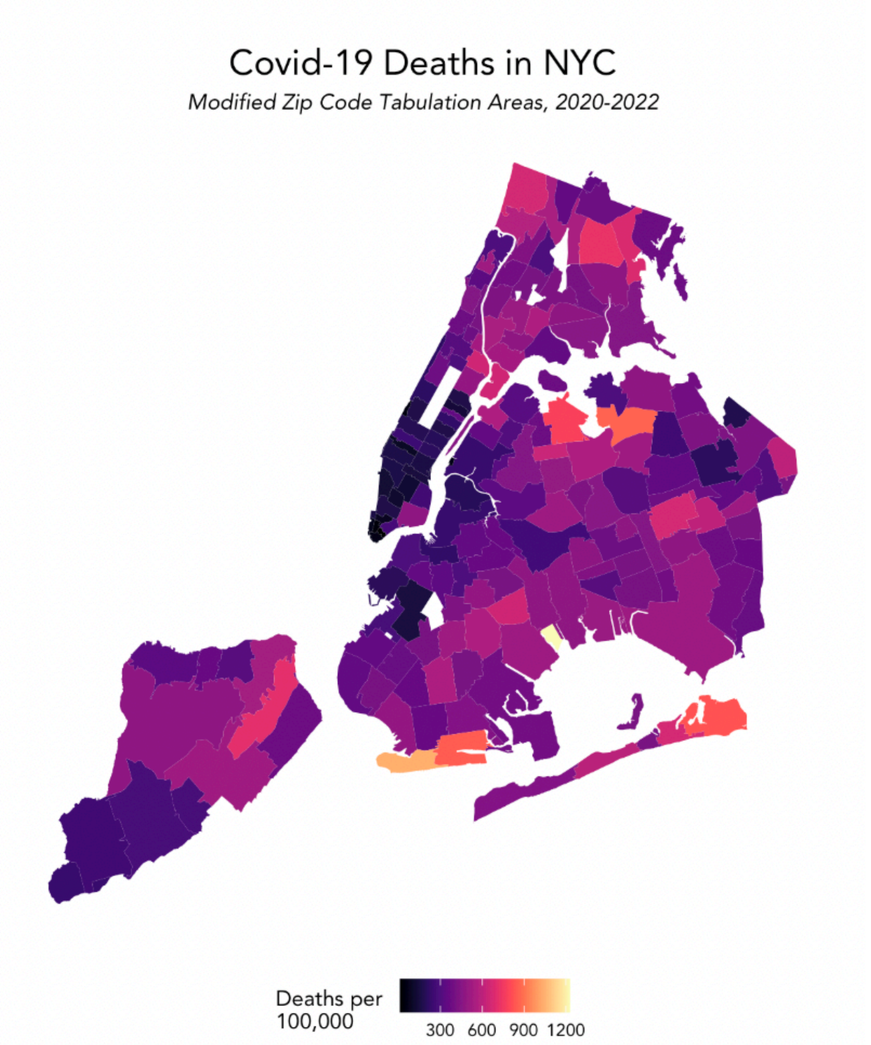
A case study of New York City

Anna Duan
CPLN 505: Planning by Numbers
Professor Megan Ryerson
Spring 2022
annaduan@sas.upenn.edu



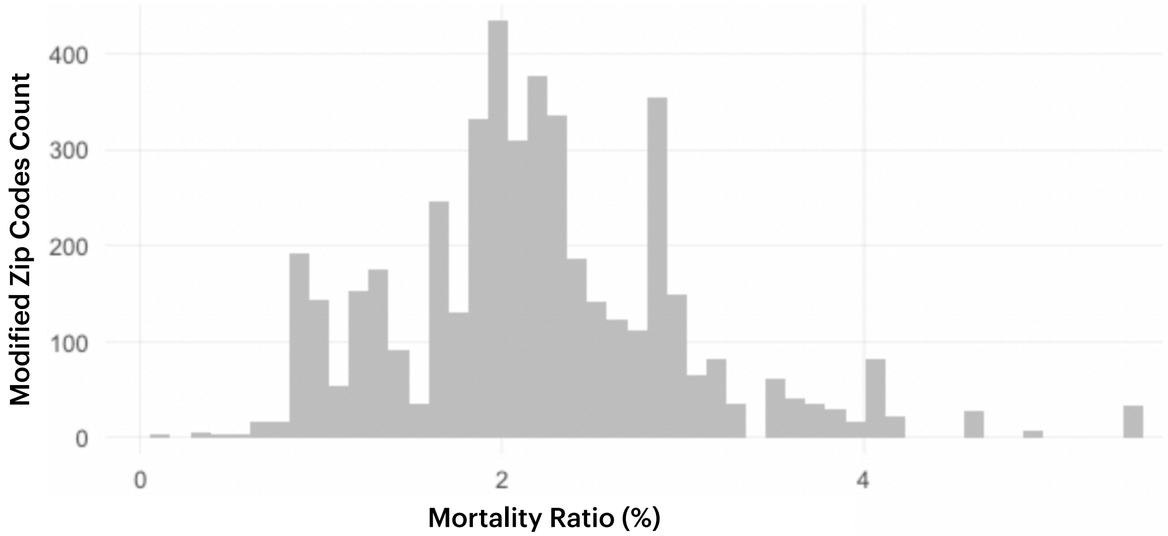
Why do these maps look different?





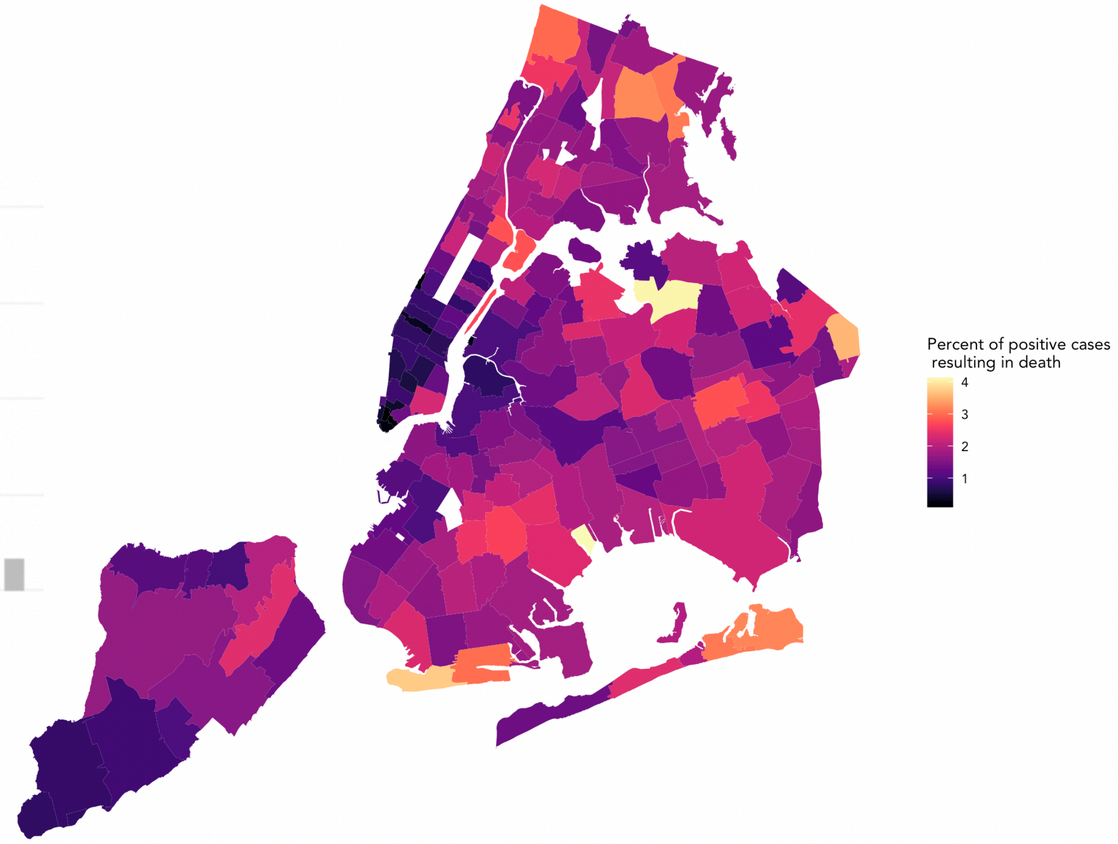
Research question: what are the spatial risk factors for fatal Covid-19 outcomes?

Mortality ratio = $\frac{\text{confirmed death count}}{\text{confirmed case count}}$



Covid-19 Cases Resulting in Death in New York City

2020-2022; Modified Zip Code Tabulation Areas

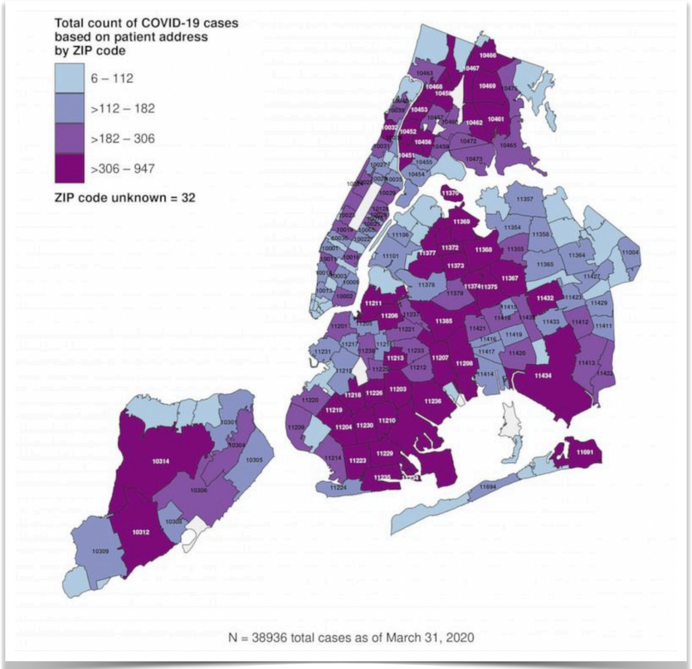


How can context-specific study of risk factors and exposure improve pandemic response?

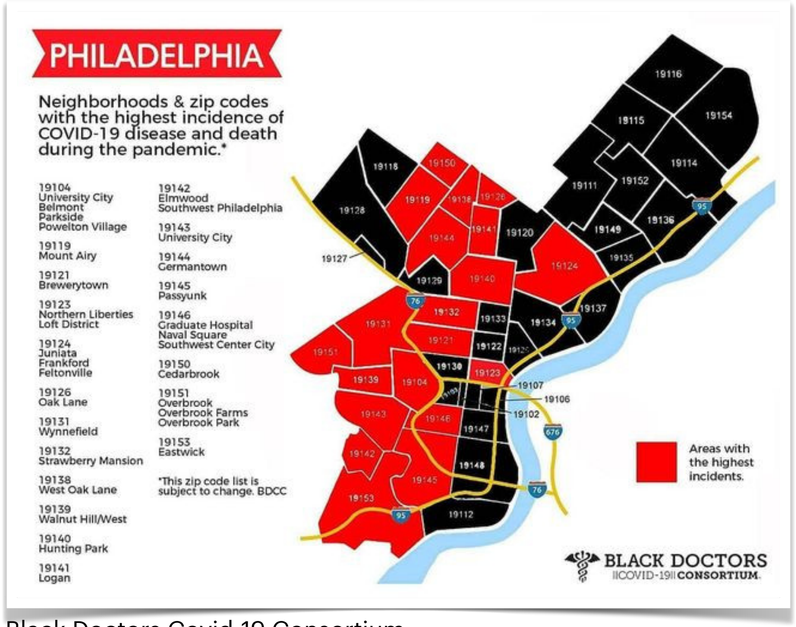
1. **Allocate resources** via school-based programs, vaccination clinics, emergency rental assistance, supplementary income

2. Advocate for accurate, data-driven analysis and communication of Covid-19's consequences

3. Inform <u>proactive</u> and <u>equitable</u> policy based on a model of risk and exposures



NYC Health Department

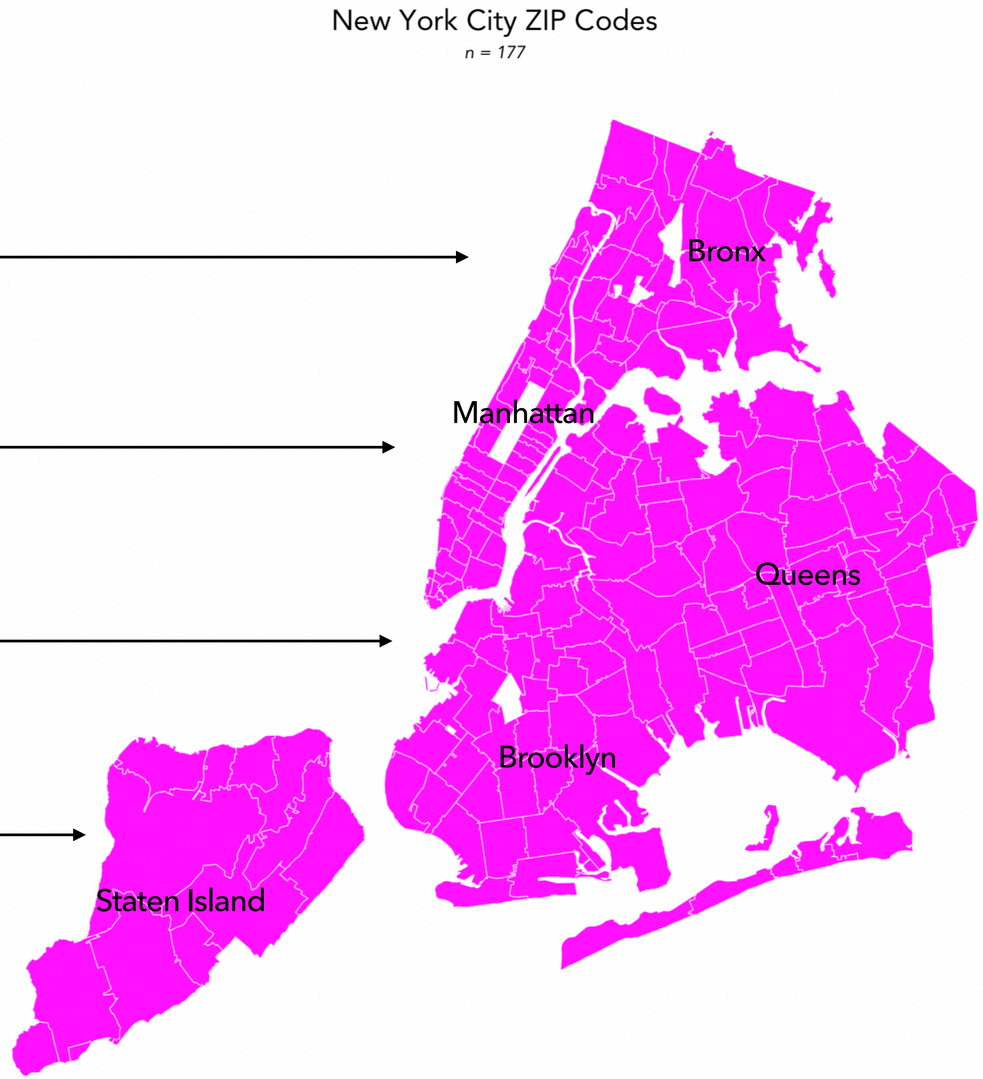


Black Doctors Covid-19 Consortium

Data & Methods

Data Sources:

- **NYC Department of Health and Mental Hygiene (DOHMH):**
  - Covid-19 cases, deaths, vaccination, positivity, hospitalizations*
  - unit of analysis: ZIP code
- **NYC Open Data:**
  - crime, schools, evictions,
  - unit of analysis: point
- **US Census Bureau American Community Survey:**
  - Socioeconomic, demographic, living conditions
  - unit of analysis: census tract
- **US Department of Housing and Urban Development**
  - Section 8 housing choice voucher rent limits
  - unit of analysis: zip code

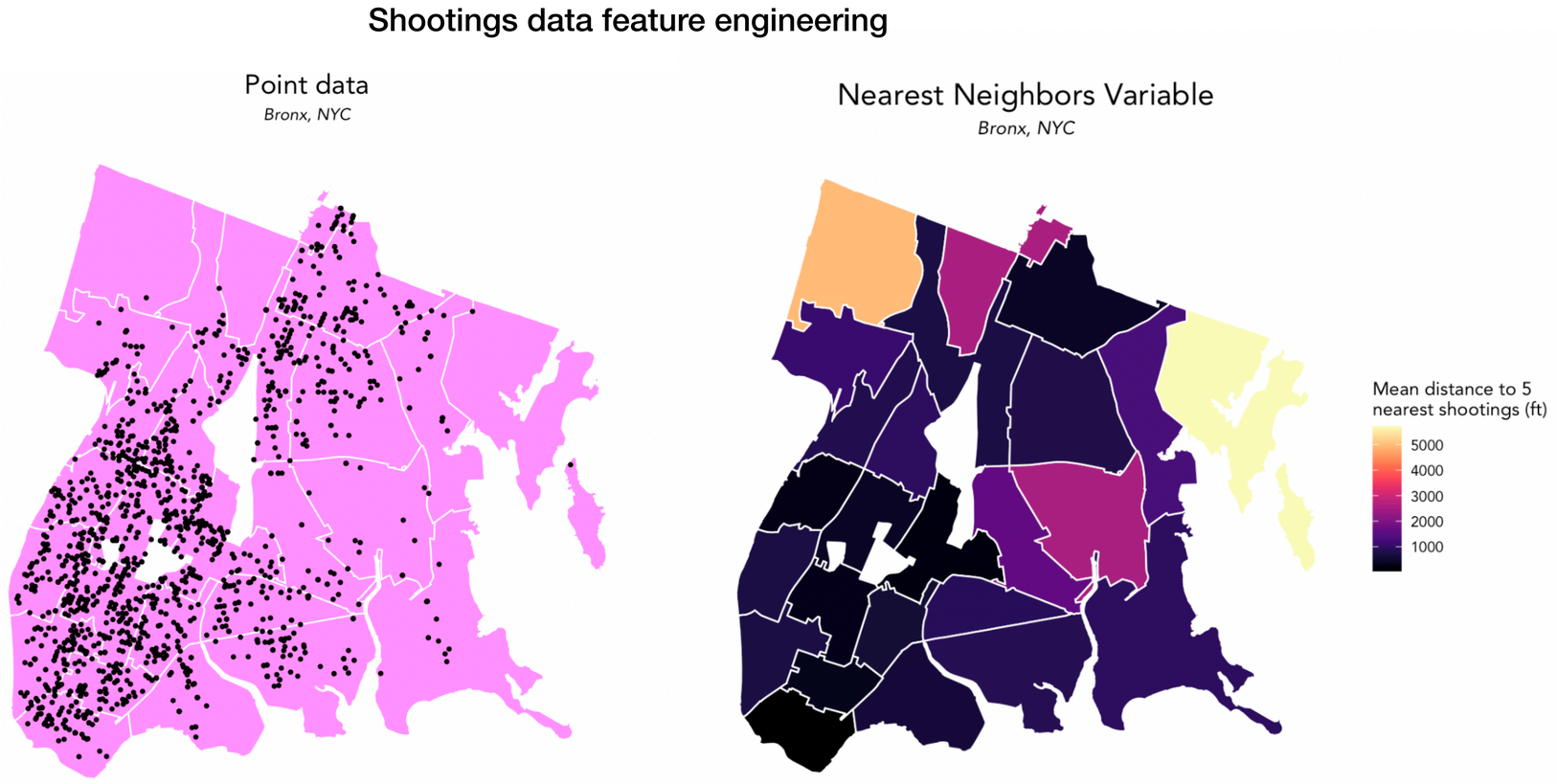


*prior to Omicron variant

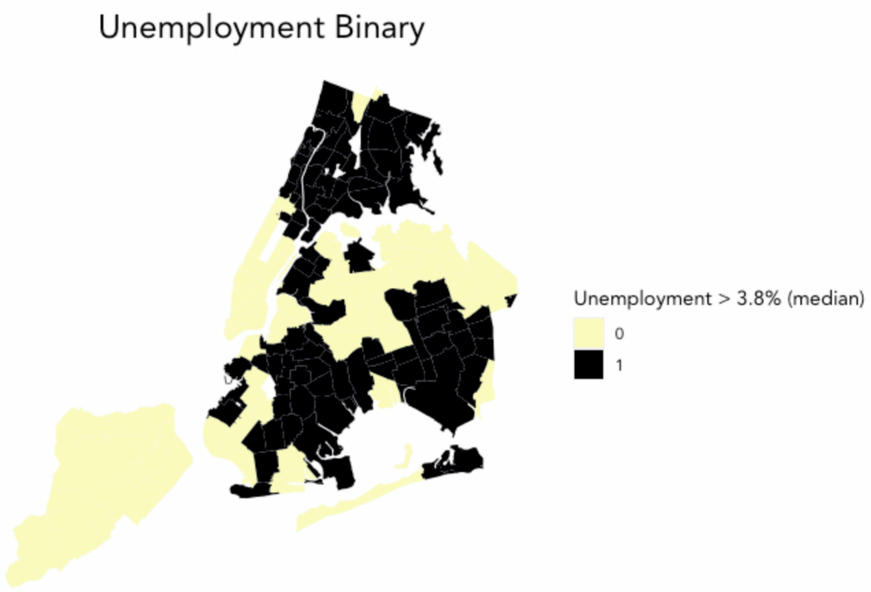
Feature engineering

Variable transformations

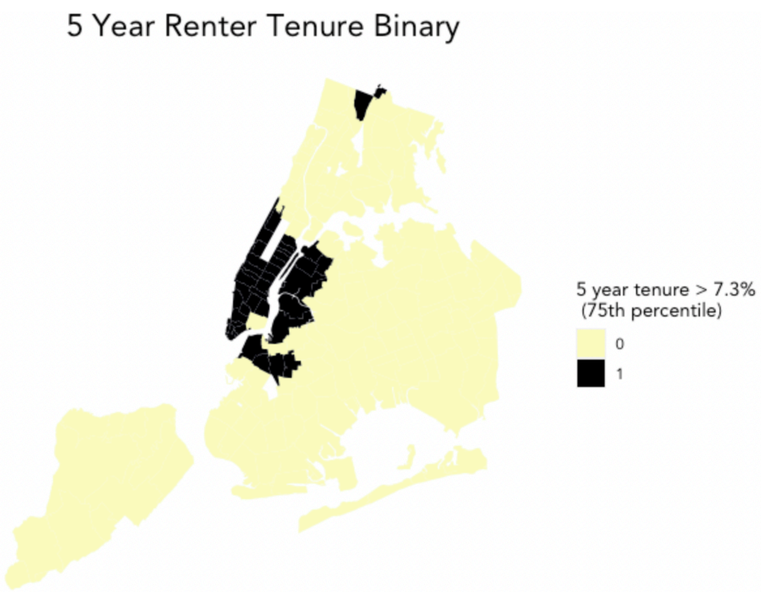
- Nearest neighbor
  - Point data, such as distance to nearest 5 shootings
  - Thresholds for continuous variables, e.g. distance to nearest 5 ZIP codes where poverty > 30%
- Distance
- Log-transformed variables
- Dummy variables
  - Boroughs
  - Above vs below citywide median, mean, or quartiles for continuous variables



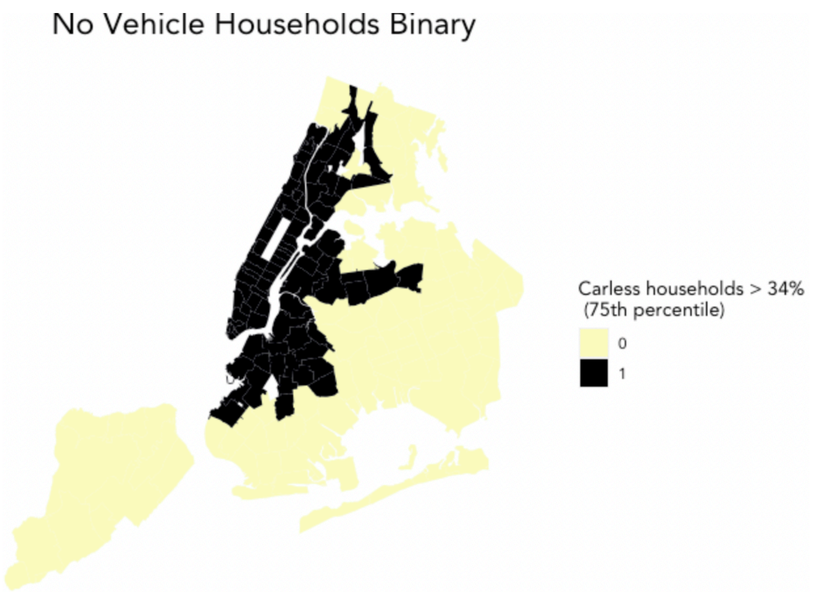
Association testing



Mortality ratio is 4 times higher in areas where unemployment > 3.8%

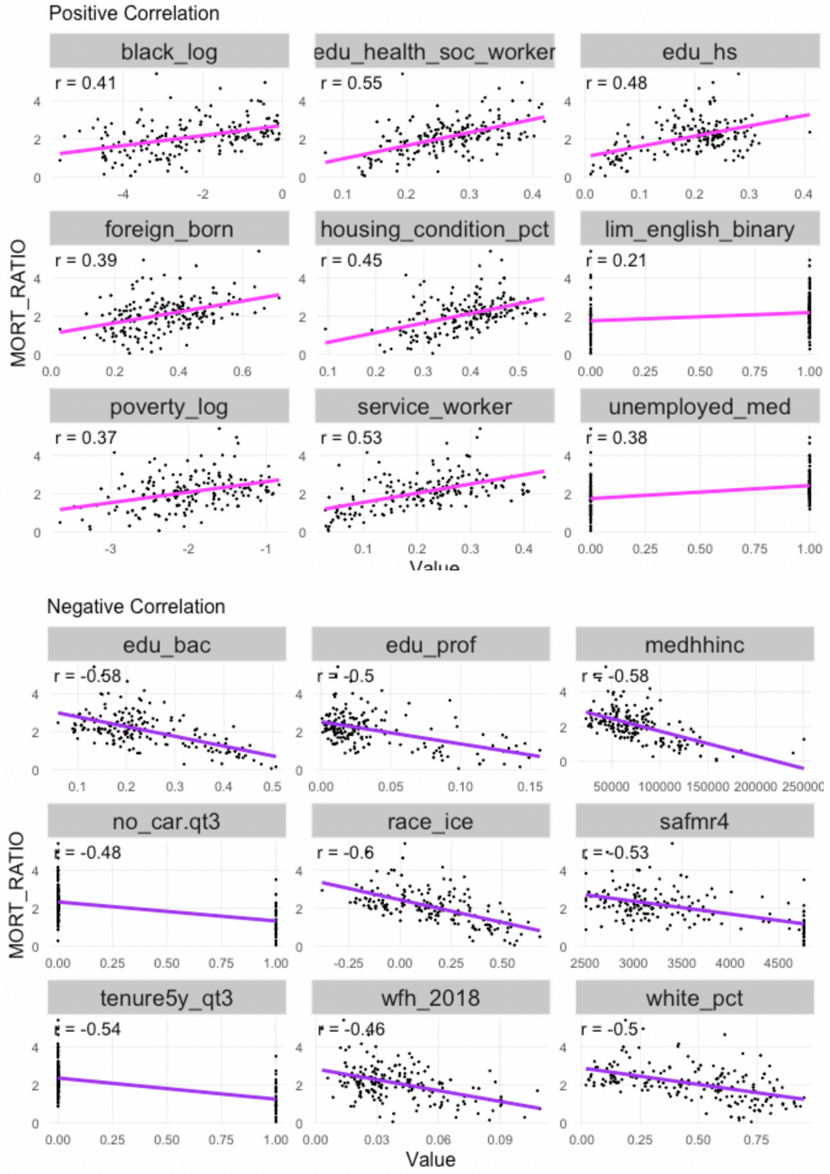


Mortality ratio is 8 times higher in neighborhoods where less than 7.3% of renters have 5 years or more of tenure



Neighborhoods with more carless households generally have a lower mortality ratio

Correlation testing



Positive: top 9

1. Black residents (%, log-transformed)
2. Education, health, or social workers (%)
3. Age 25+ with high school education level (%)
4. Foreign born residents (%)
5. Housing units with physical conditions (%)
6. Households with limited English fluency (%, binary)
7. Poverty rate (%, log-transformed)
8. Service workers (%)
9. Unemployment (%, binary)

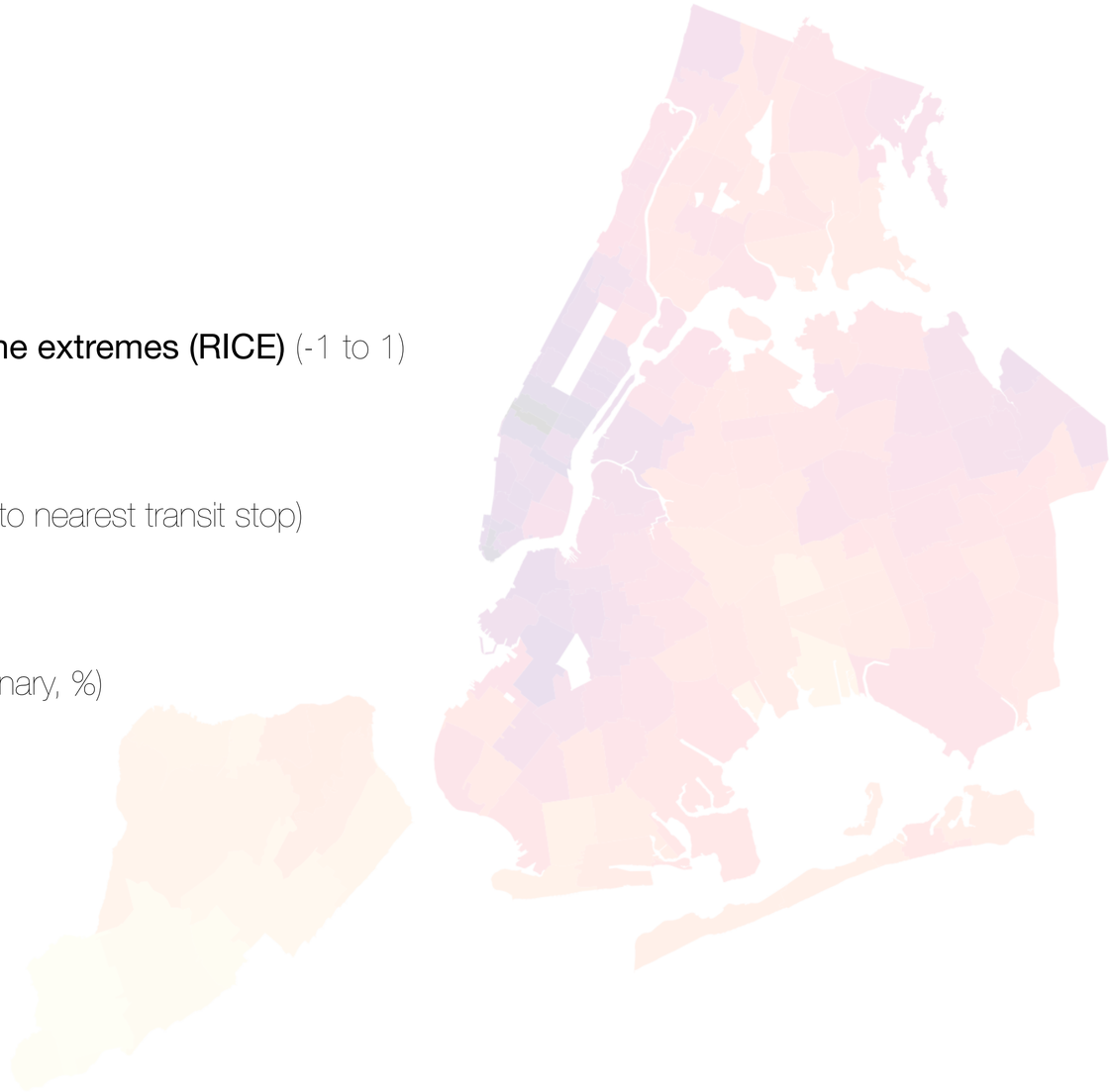
Negative: top 9

1. Age 25+ with Bachelors degree (%)
2. Age 25+ with professional degree (%)
3. Median household income ($)
4. Carless households (%, binary)
5. Racialized Index of Concentration at the Extremes (range: -1 to 1)
6. HUD Housing Choice Voucher limit for 4 bedroom units ($)
7. Renters with more than 5 years tenure (%, binary)
8. Workers working from home in 2018 (%)
9. White residents (%)

Final Model: Covid-19 Mortality Ratio

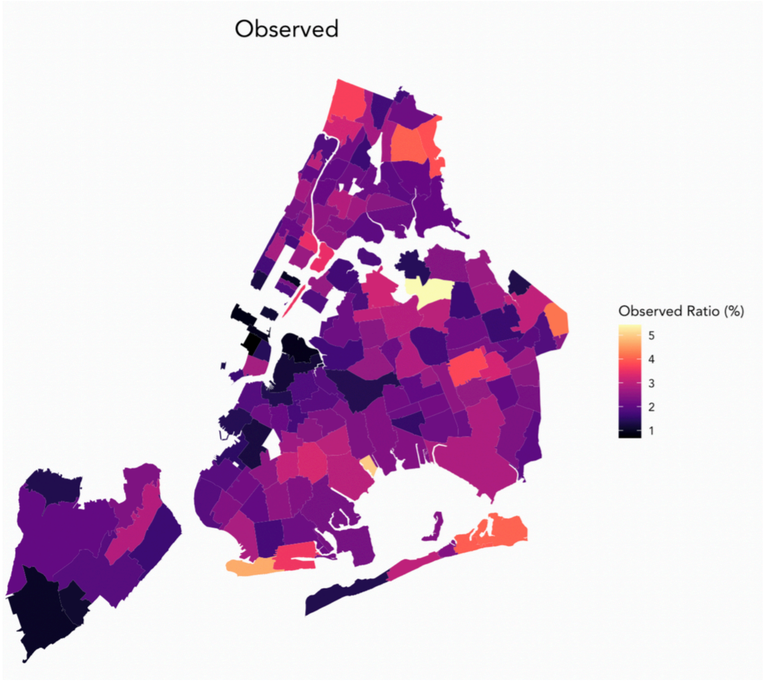
Variables:

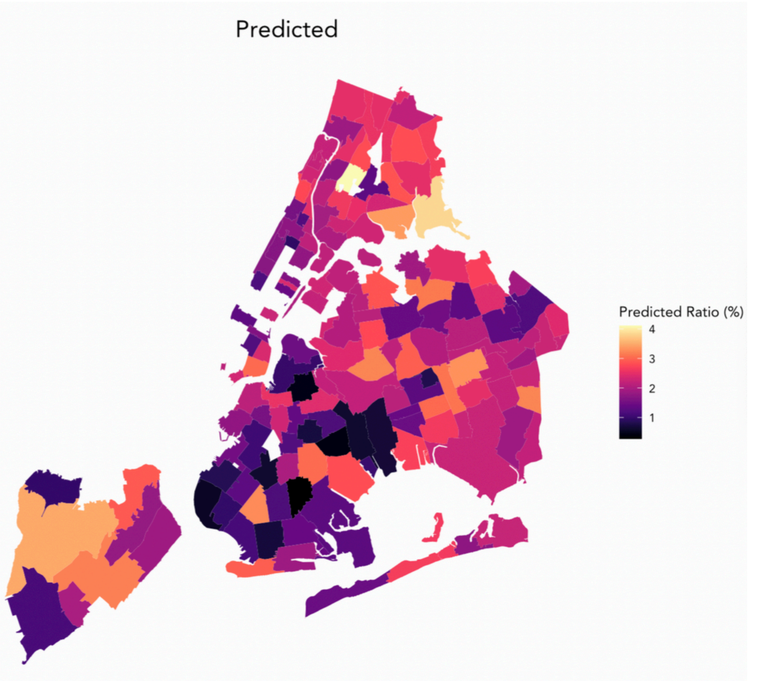
- Median age
- Racialized index of concentration at the extremes (RICE) (-1 to 1)
- Apartments (% share of all housing units)
- Transit density (binary, average distance to nearest transit stop)
- Carless households (binary, %)
- Borough : middle school education (binary, %)



Model Results

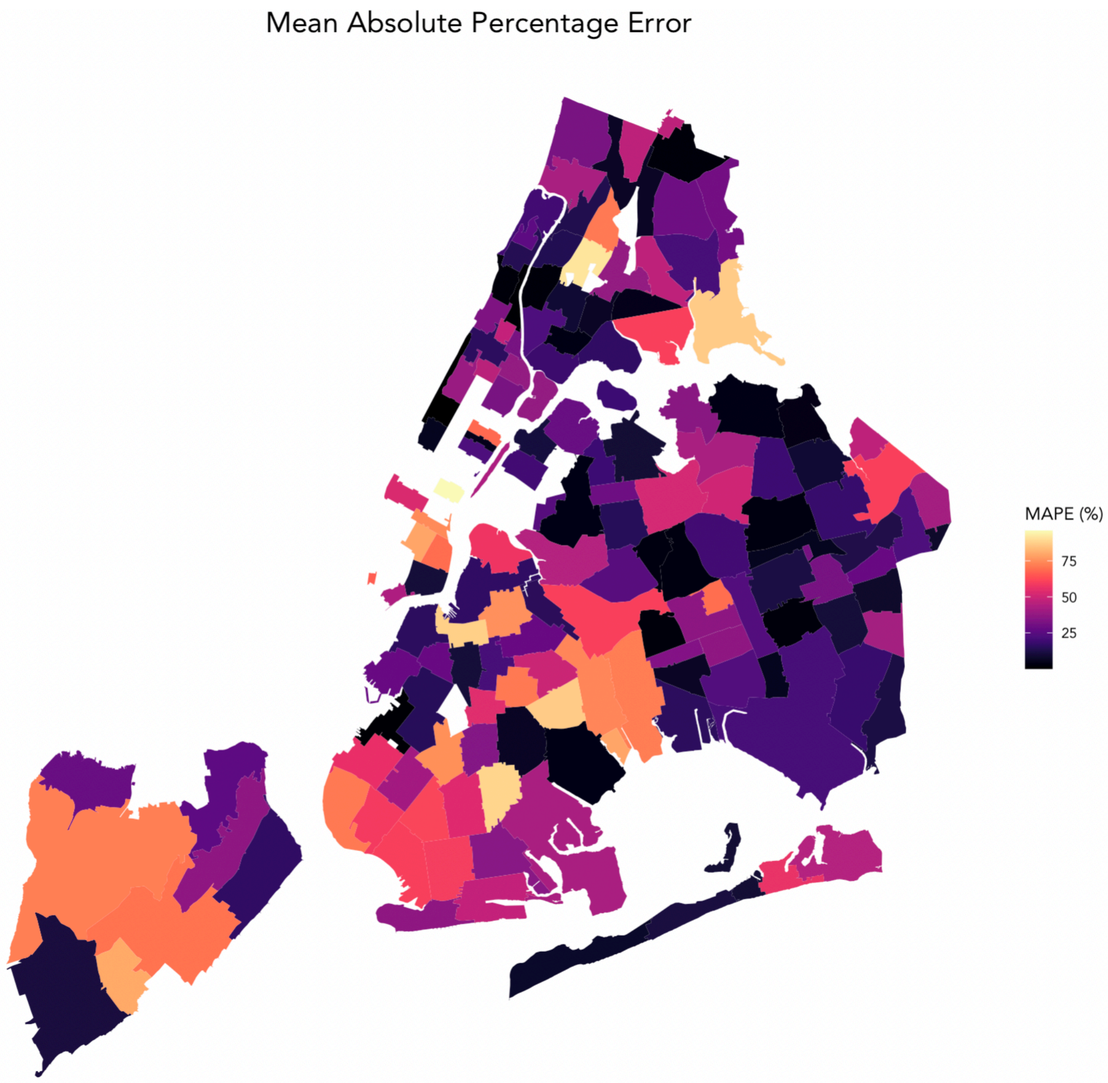
1. Strongest predictors: RICE, apartments, and interaction term between middle school degrees and Queens
2. A higher share of apartments is correlated with a higher mortality ratio
3. White, wealthy neighborhoods have lower ratios
4. Share of adults with only middle school education is associated with higher ratios
5. Higher transit density correlates with lower ratios
6. Proximity to institutions such as hospitals and schools and incidence of shootings and evictions do not predict significantly for mortality ratio





Generalizability: spatial

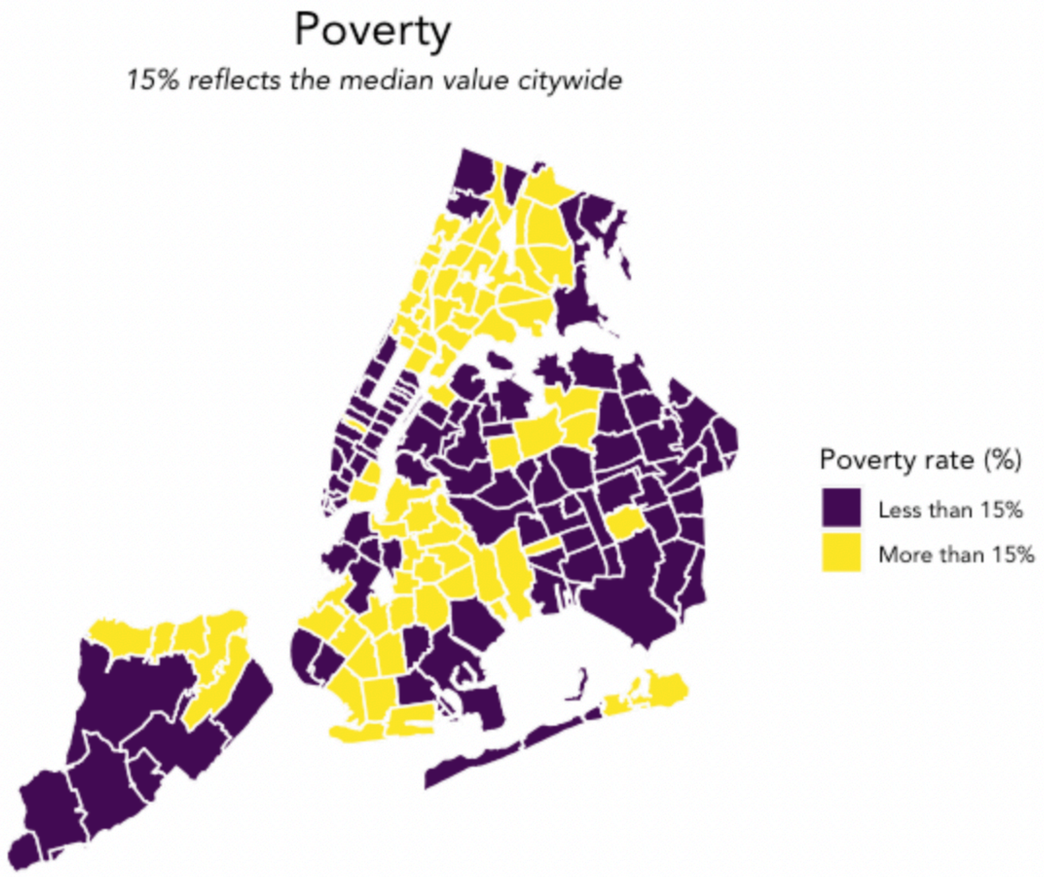
The model predicts with lower accuracy in Staten Island and Manhattan, the boroughs with the lowest mortality ratios.



Generalizability: demographic

The model predicts with higher accuracy in ZIP codes where:

- majority non-White residents
- poverty rate > 15% (citywide median)
- foreign born population > 35% (citywide median)



Mean Error by Poverty Context

Less than 15%	More than 15%
0.8192159	0.9692555

Conclusion

1. Case outcomes need to be a central metric in the ongoing pandemic response
2. Commonly known risk factors for Covid-19 cases and deaths don’t predict well for how Covid-19 affects individuals and communities
3. Race is relevant, but doesn’t tell the whole story: living conditions are very relevant
4. Future interventions need to account for housing conditions, environmental stressors, and transit access
5. More research is needed on the impact of schools, hospitals, and other institutions, as well as the disparate impacts of the pandemic on workers in different industries