

A Mixed-Effects Spatial Analysis of Violent Crime in Buffalo, NY

Intersections of Socioeconomic Inequality and Police Surveillance (2010-2023)

Stephen C. Sanders

May 17, 2025

Introduction and Literature Review

Buffalo, New York, like many Rust Belt cities, has faced significant economic decline, population loss, and elevated crime rates over recent decades (Klinenberg, 2002; Scorsone & Bateson, 2011; Kraus, 2004). Once a major industrial hub, Buffalo's population has declined dramatically from its peak in the mid-20th century, a trend strongly tied to deindustrialization and regional economic restructuring (Hollander & Cahill, 2011; Ryan, 2015). Today, nearly 30% of Buffalo's residents live below the poverty line (U.S. Census Bureau, 2020), and these concentrated socioeconomic disadvantages have been shown to be key contributors to elevated violent crime rates (Sampson et al., 1997; Krivo & Peterson, 1996).

In response to persistent crime, the Buffalo Police Department has implemented various crime prevention strategies, including the deployment of approximately 275 fixed surveillance cameras throughout the city (Open Data Buffalo, 2025). While public surveillance is increasingly used in urban policing strategies, research suggests mixed outcomes regarding their effectiveness. Some studies indicate modest reductions in violent and property crimes, while others emphasize contextual limitations (Welsh & Farrington, 2009; Piza, 2018; Lum et al., 2019).

This study endeavors to identify significant predictors of violent crime in Buffalo, with particular focus on the presence of police surveillance cameras alongside a set of socioeconomic and demographic

variables. A Generalized Linear Mixed Model (GLMM) was employed, allowing for both fixed effects (e.g., per capita income, poverty rate, educational attainment, racial composition) and random effects (e.g., spatial heterogeneity across census block groups), while a spatially lagged covariate was incorporated to account for neighborhood spillover effects (Anselin et al., 2000; Reich et al., 2006). An initial approach aggregating data over a 14-year period (2010–2023) revealed that every block group experienced at least one violent crime, thus necessitating a more nuanced model capable of isolating conditions under which violent crime is more likely to occur. By integrating spatial dynamics, police surveillance infrastructure, and neighborhood-level socioeconomic predictors, this study seeks to offer a more comprehensive view of crime risks in Buffalo, evaluate the locations of current surveillance cameras relative to block groups with high unexplained risk, and contribute to the growing body of evidence on urban crime prevention strategies.

Understanding the multifaceted nature of violent crime necessitates an examination of both socioeconomic determinants and technological interventions and their relationship to violent crime. Research consistently indicates that socioeconomic factors such as poverty, unemployment, educational attainment, income inequality, overall economic disadvantage, and racial segregation are significant predictors of crime rates (Sampson and Groves, 1989; Krivo and Peterson, 1996). For instance, Wirdze (2024) found that higher levels of income inequality often lead to increased crime rates, particularly violent crimes like homicides and robberies. Similarly, the concept of “concentrated disadvantage,” which encompasses factors like high poverty rates, unemployment, and single-parent households, has been linked to elevated crime levels in urban neighborhoods (Benson & Fox, 2004). This concept can potentially help explain the “law of crime concentration,” which accentuates a phenomenon where a small percentage of urban spaces account for a disproportionate amount of crime, emphasizing the need for place-based crime prevention strategies (Weisburd, 2015). These strategies typically involve installing surveillance cameras in an attempt to discourage crime. Systematic reviews and meta-analyses have found that CCTV surveillance is associated with a significant and modest decrease in crime, with the most consistent effects observed in parking lots (Office of Justice Programs, 2019; Welsh & Farrington, 2009). However, the effectiveness of surveillance cameras can vary based on context and implementation. For example, a study by the Urban Institute (2019) indicated that while cameras can reduce crime, particularly property crimes and vehicle crimes in parking lots, the findings are mixed and context-

dependent.

Spatial analysis has become an essential tool in crime modeling and general criminological research, allowing researchers to account for the geographic distribution of crime. Anselin et al. (2000) emphasized the importance of spatial econometric models in understanding crime dynamics, advocating for the inclusion of spatial lag variables to capture the influence of neighboring areas. Incorporating spatially lagged covariates into crime models helps address spatial autocorrelation, ensuring more accurate and reliable results (Park & Kim, 2014; Marotta, 2016). This approach acknowledges that crime in one area can be influenced by crime in neighboring areas, reflecting the interconnected nature of urban environments. The use of Generalized Linear Mixed Models (GLMMs) enables the integration of both fixed effects (e.g., socioeconomic variables) and random effects (e.g., unobserved heterogeneity across locations), providing a robust framework for analyzing complex crime and socioeconomic data (Bates et al., 2015; Wu & Li, 2024; Sariaslan et al., 2013). By combining socioeconomic data, surveillance measures, and spatial analysis, researchers can develop comprehensive models to better understand and predict crime patterns, ultimately informing more effective crime prevention strategies which include evaluating areas in which surveillance installation patterns need to be revisited. These methods will be applied in this study to attempt to determine significant predictors of violent crime and identify areas with violent crime risk that is not explained by those predictors, allowing for future discussions and research that can lead to pointed intervention strategies in these communities.

Data and Study Area

Data

The analysis looked at various demographic and socioeconomic variables along with presence of a police surveillance camera. Specifically, per capita income, poverty rate, percent of the 25 and over population with at least a bachelor degree (hereon referred to as “educational attainment”), local and neighboring percentages of the population who identified as black, percent of the block group within walking distance (0.25 miles) of a park (hereon referred to as “park proximity”), and time periods of 2010-2012, 2013-2016, 2017-2019, and 2020-2023 were used as predictors in the violent crime pre-

diction model. Additionally, the block group GEOIDs were given a random intercept to account for unobserved unique conditions of each spatial unit.

There were 62,967 violent crimes that occurred in Buffalo between 2010 and 2023, which averages out to around 4,497 violent crimes per year over the 14-year span. By 2025, Buffalo Police Department had installed 275 surveillance cameras across the city. Of the cameras, 78 did not have an attached installation date. There were also 109 parks administered by the city, county, or state within city limit, all of which are at least 0.25 acres in size.

Census data was pulled for years between 2010 and 2023. Block group-level data for 2013-2023 was gathered using the [tidycensus](#) package. Since the Census Bureau blocks access to block group data before 2013, data for 2010-2012 was pulled directly from the [IPUMS](#) data selection tool by the [NHGIS](#). The data was then downloaded and loaded using the [ipumsr](#) package. All of it originally came from American Community Survey (ACS) 5-year estimates calculated by the U.S. Census Bureau. Exact data sources can be found in the references.

```
# import necessary libraries
library(tidyverse)
library(tidycensus)
library(tigris)
library(sf)
library(ipumsr)
library(terra)
library(fasterize)
library(distanceto)
library(rgee)
library(reticulate)
library(exactextractr)
library(units)
library(performance)
library(lme4)
library(DHARMa)
library(spdep)
library(car)
library(pROC)
library(RColorBrewer)
library(cowplot)
library(gt)

source('R/functions.R')

knitr::opts_chunk$set(echo = TRUE, # cache the results for quick compiling
  fig.align = 'center',
  fig.width = 7.5,
  fig.asp = 0.618,
  fig.pos = 'H')
```

```
# set options and get census api key
options(tigris_use_cache = TRUE)
terraOptions(progress = 0)
census_token = Sys.getenv("CENSUS_TOKEN")
census_api_key(census_token)
```

Study Area

The study focuses on the 290 census block groups in Buffalo, New York as they have been delineated since 2020. However, the model was trained on a random sample consisting of ~80% (231) of the block groups.

```
# get buffalo boundary and buffalo block group boundaries
buffalo <-
  tigris::places(state = 'NY', cb = TRUE, year = 2020) %>%
  filter(NAME == 'Buffalo')

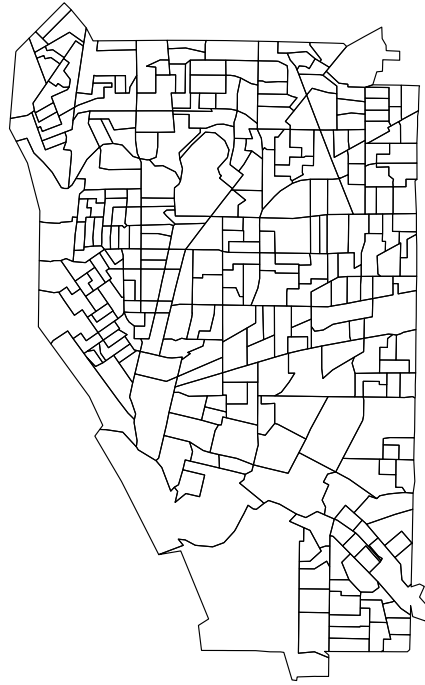
study_bgs <-
  tigris::block_groups(state = 'NY', county = 'Erie',
    cb = TRUE, year = 2020) %>%
  .[data.frame(st_intersects(st_centroid(.), buffalo))$row.id,]

# plot Buffalo block groups as of 2020
buff_map <-
  ggplot() +
    geom_sf(data = study_bgs,
      color = 'black', fill = NA, size = 2) +
    labs(title = 'Buffalo, New York Block Groups',
      subtitle = 'Since 2020') +
    theme_map()

buff_map
```

Buffalo, New York Block Groups

Since 2020



Methods

Processing Methods

The [tidyverse](#) collection of packages was the primary tool used to load, combine, and process the data used in this study. The [tigris](#) package (specifically the *places* and *block_groups* functions) was used to pull Buffalo block groups to create a map of the study area and to pull block-level population and geometry data for use as the interpolation weights. The [sf](#) package was necessary to work with spatial features, including the block group-level ACS data. Maps and plots were created using [ggplot2](#), and data tables were created using the [gt](#) package. The *interpolate_pw* function of the *tidycensus* package was employed to create uniform block group boundaries over the 14-year period through population-weighted areal interpolation via centroid assignment (Walker, 2023).

The [terra](#) package was used to convert the park proximity buffers to rasters and process them. The resulting rasters were then the basis for the calculation of a percentage walking distance park proximity

variable with the *exact_extract* function from the [exactextractr](#) package. Walking distance was defined as being at most 0.25 miles away from at least one of the 109 parks in the city.

Generalized Linear Mixed Model using lme4

A generalized linear mixed model was fitted using the [lme4](#) package to assess violent crime risk across the 290 uniform block groups. The model attempted to predict the presence of violent crime in a block group using the block group's per capita income, poverty rate, educational attainment, local and neighboring percentages of the population who identified as black (to attempt to account for spatial dependence), park proximity, and time period as predictors. The spatially lagged covariate of black population was needed given the socioeconomic nature of this study, since socioeconomic phenomena tend to be clustered. Also, incorporating the GEOID of each block group as a random effect in a Generalized Linear Mixed Model (GLMM) effectively accounts for unobserved, time-invariant characteristics unique to each spatial unit, especially when observations are repeated over multiple years (Bolker et al., 2009). This approach controls for baseline differences across block groups, allowing fixed effects to capture temporal and socioeconomic predictors of violent crime. Such modeling techniques are well-supported in statistical literature, illuminating their utility in handling repeated measures and nested data structures.

A GLMM regression model was originally fitted on a condensed dataset, where each row represented one block group and their average values over the course of the study time period. Every block group had at least one violent crime occur within the 14-year time period, so there was no second group (block groups without crime) to compare to. Another previous model attempt treated the year as a factor to observe more temporally granular violent crime risk changes, but the model faced issues with overparameterization and large standard errors in the coefficients. Creating a categorical predictor that groups the block groups into time periods was an appropriate work-around to solve issues with fitting while maintaining the ability to evaluate the change in risk since the earliest years of the study.

The *glmer* function from the *lme4* package was specifically used as it is widely utilized for fitting Generalized Linear Mixed Models (GLMMs), particularly in studies examining the influence of socioeconomic and demographic factors on various outcomes. For instance, in their comprehensive guide,

Bates et al. (2015) detail the application of *glmer* for modeling data with both fixed and random effects, emphasizing its flexibility in handling complex hierarchical structures. The bobyqa optimizer within the *glmer* function was used to enhance model convergence and computational efficiency. The bobyqa optimizer, which stands for Bound Optimization By Quadratic Approximation, is particularly effective for optimizing complex models with random effects, as it does not require gradient information and is well-suited for problems with bound constraints.

The [DHARMA](#) package was used for diagnostic tests to evaluate model assumptions after the model was fitted and before interpreting the model results. The [spdep](#) package was then used to check for spatial autocorrelation by running a Moran's I test.

Data Processing

Set Up and Load Data

The boundary and neighborhood boundaries of Buffalo, as well as the violent crimes, police surveillance cameras, and city-based parks were loaded from their respective governmental sources and then processed.

Additionally, ACS data was loaded using tidycensus and ipumsr and then processed. Population-weighted areal interpolation techniques were applied to the 2010-2019 data to match the 2020-2023 boundaries. The final variables were calculated/determined, and the dataset was finalized.

Features Used in Analysis

```
# get buffalo city boundaries
buff_border_url <- paste('https://data.buffalony.gov/api/views/p4ak-r4fg/rows.csv',
                        '?date=20250308&accessType=DOWNLOAD', sep = '')

buff_border <-
  read_csv(buff_border_url) %>%
  st_as_sf(wkt = 'Geometry', crs = 'EPSG:4269')

# get buffalo neighborhoods
buff_neighborhoods <-
  read_csv('https://data.buffalony.gov/resource/ekfg-mtu8.csv') %>%
  st_as_sf(wkt = 'the_geom', crs = 'EPSG:4269') %>%
  rename(
    neighborhood_name = nbhdname,
    area_sqmi = sqmiles,
    geometry = the_geom
```



```

) %>%
select(-c(nbhdnum, calcacres, objectid_1))

# import buffalo crimes
### 319,473 ROWS IN RAW FILE
##### 312,867 HAVE COORDINATES
##### 311,222 ARE WITHIN BUFFALO BOUNDARIES
crimes <-
  read.csv('data/Crime_Incidents_20250303.csv', check.names = FALSE) %>%
  select(`Case Number`, `Incident Datetime`, `Incident Type Primary`, `Parent Incident Type`,
        `Hour of Day`, `Day of Week`, Address, City, State, zip_code,
        Latitude, Longitude, neighborhood, `Police District`) %>%
  mutate(
    Month = str_split(
      str_split(.$`Incident Datetime`, ' ', simplify = TRUE)[, 1], '/', simplify = TRUE)[, 1],
    Day = str_split(
      str_split(.$`Incident Datetime`, ' ', simplify = TRUE)[, 1], '/', simplify = TRUE)[, 2],
    Year = str_split(
      str_split(.$`Incident Datetime`, ' ', simplify = TRUE)[, 1], '/', simplify = TRUE)[, 3]
  ) %>%
  mutate_at('Year', as.numeric) %>%
  filter(Year >= 2006 & Year < 2024) %>%
  .[!((.$Latitude == 'UNKNOWN' | .$Longitude == 'UNKNOWN') |
      (.$Latitude == '' | .$Longitude == '')), ] %>%
  st_as_sf(coords = c('Longitude', 'Latitude'), crs = 'EPSG:4269')

buff_crimes <- crimes[data.frame(st_intersects(crimes, buff_border))$row.id,]

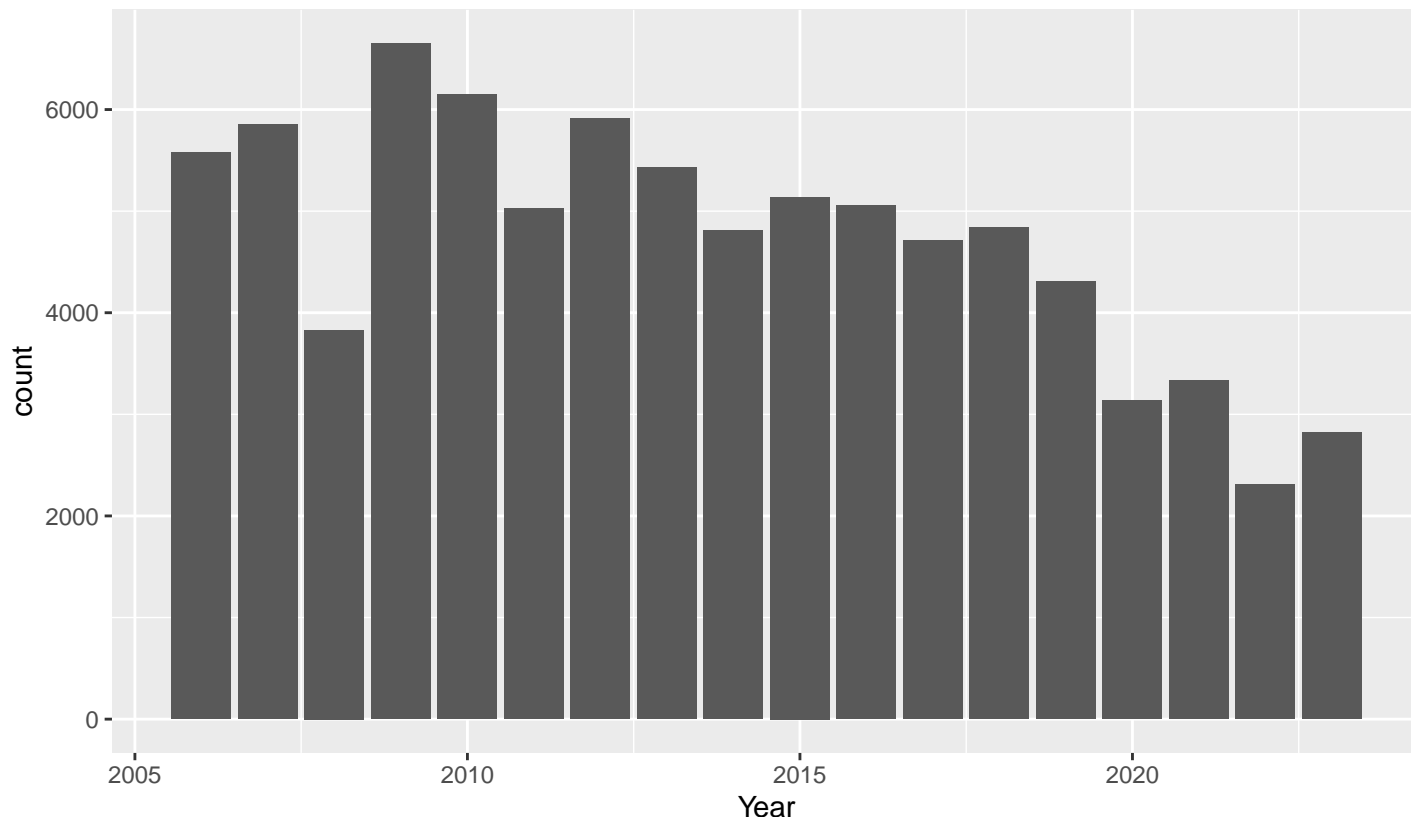
## 88,503 VIOLENT CRIMES
violent_crimes.buff <- buff_crimes %>%
  filter(`Incident Type Primary` %in%
    c('MURDER', 'CRIM NEGLIGENT HOMICIDE', 'MANSLAUGHTER', 'RAPE', 'AGGR ASSULT',
      'AGG ASSULT ON P/OFFICER', 'ROBBERY', 'Robbery', 'ASSAULT', 'Assault',
      'SEXUAL ABUSE', 'SODOMY', 'Other Sexual Offense'))

## 166,334 PROPERTY CRIMES
property_crimes.buff <- buff_crimes %>%
  filter(`Incident Type Primary` %in%
    c('UUV', 'THEFT OF SERVICES', 'LARCENY/THEFT',
      'Theft', 'Theft of Vehicle', 'Breaking & Entering'))

# show bar graph of number of violent crimes per year between 2006 & 2023
ggplot(violent_crimes.buff, aes(Year)) +
  geom_bar() +
  labs(title = 'Violent Crimes in Buffalo (2006-2023)')

```

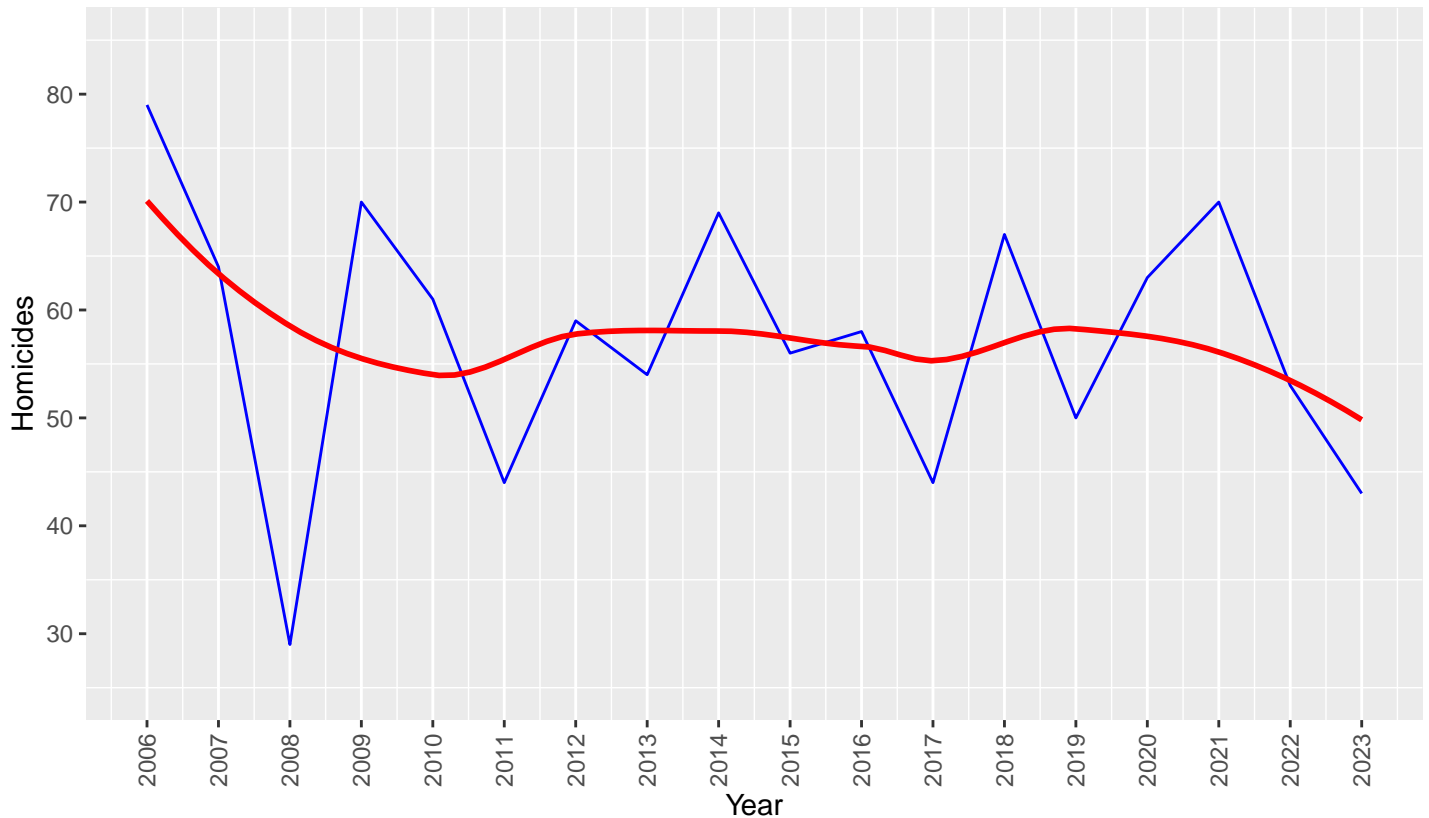
Violent Crimes in Buffalo (2006–2023)



```
# create line graph of homicides between 2006 and 2023 in Buffalo
homicide_types = c('MURDER', 'CRIM NEGLIGENT HOMICIDE', 'MANSLAUGHTER')

violent_crimes.buff %>%
  st_drop_geometry %>%
  filter(`Incident Type Primary` %in% homicide_types) %>%
  group_by(Year) %>%
  summarize(homicides = n()) %>%
  ungroup() %>%
  ggplot() +
    geom_line(aes(Year, homicides), color = 'blue') +
    geom_smooth(aes(Year, homicides), color = 'red', method = loess, se = FALSE) +
    ggtitle('Homicides in Buffalo, NY (2006 - 2023)') +
    ylab('Homicides') +
    scale_y_continuous(limits = c(25, 85),
                      breaks = c(30, 40, 50, 60, 70, 80),
                      labels = c(30, 40, 50, 60, 70, 80)) +
    scale_x_continuous(limits = c(2006, 2023),
                      breaks = c(2006:2023),
                      labels = c(2006:2023)) +
    theme(plot.title = element_text(hjust = 0.5),
          axis.text.x = element_text(angle = 90,
                                      vjust = 0.5,
                                      hjust = 1),
          panel.grid.major.y = element_blank())
```

Homicides in Buffalo, NY (2006 – 2023)



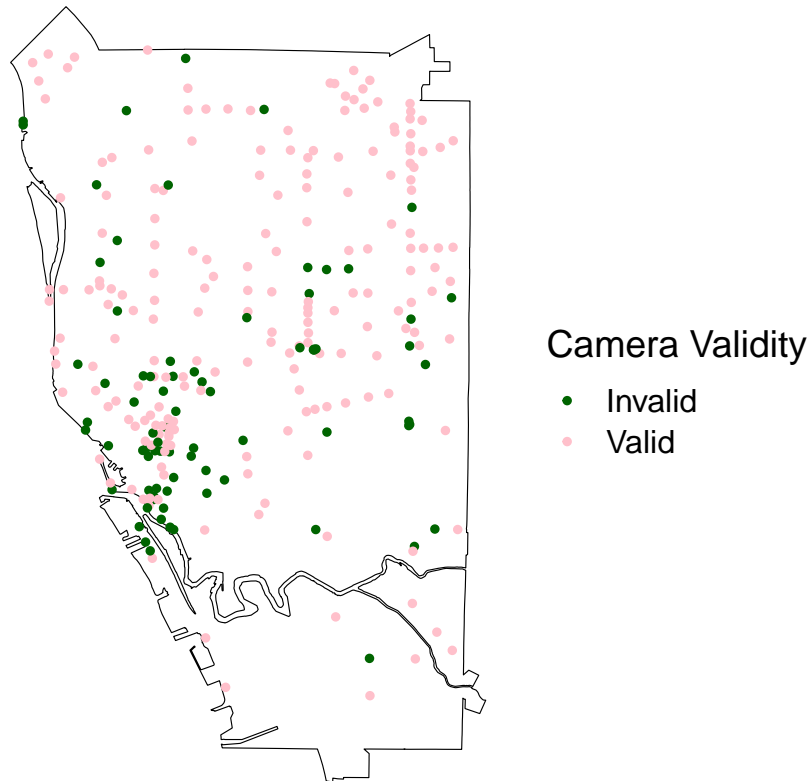
```
# import bpd cameras data
### 275 TOTAL CAMERAS
##### 83 CAMERAS WITH VALID INSTALLED DATE
bpd_cameras_url <- paste(
  'https://data.buffalony.gov/api/views/gpj7-v6rr/rows.csv',
  '?date=20250304&accessType=DOWNLOAD', sep = ''
)

bpd_cameras <-
  read.csv(bpd_cameras_url, check.names = FALSE) %>%
  mutate(year = as.numeric(str_split(.$`Installed Date`, '/')$[, 3])) %>%
  st_as_sf(coords = c('X', 'Y'), crs = 'EPSG:4269')

bpd_cameras[grepl('2010', bpd_cameras$`Installed Date`), 'Installed Date'] <- 2010
bpd_cameras[grepl('2010', bpd_cameras$`Installed Date`), 'year'] <- 2010

# map of valid and invalid police cameras
bpd_cameras %>%
  mutate(ifValid = ifelse(is.na(year), 0, 1)) %>%
  ggplot() +
    geom_sf(data = buff_border, color = 'black', fill = 'transparent') +
    geom_sf(aes(color = as.factor(ifValid)), size = 1) +
    scale_color_manual(
      name = 'Camera Validity',
      labels = c('Invalid', 'Valid'),
      values = c('darkgreen', 'pink')
    ) +
```

```
theme_map()
```



```
# separate into dataframes for valid and invalid cameras
valid_cameras <- bpd_cameras %>% filter(!is.na(year))
invalid_cameras <- bpd_cameras %>% filter(is.na(year))

# create 0.25 mile buffer around all valid cameras
#cameras_buffer <- st_buffer(valid_cameras, dist = 402.336)

# import buffalo parks data
### 218 AREAS
#####
#####
buff_parks_url <- paste(
  'https://data.buffalony.gov/api/views/tmik-tgt9/rows.csv',
  '?date=20250304&accessType=DOWNLOAD', sep = ''
)

buff_parks <-
  read.csv(buff_parks_url, check.names = FALSE) %>%
  st_as_sf(wkt = 'Geometry', crs = 'EPSG:4269') %>%
  filter(
    (`Park Class` %in% c('Major Park', 'Large Park', 'Midsize Park',
                        'Small Park', 'Parkway')) |
    (`Park Class` == 'Triangle' & Designated == 1)
```

```

) %>%
mutate('Park Category' = 'City Park',
       Area_Sqmi = units::set_units(st_area(.), 'mi2'),
       Area_Acres = units::set_units(Area_Sqmi, 'acre')) %>%
select('Park name', Address, Year, 'Park Class', 'Park Category', 'Maintained by',
       Neighborhood, 'Council District', Area_Sqmi, Area_Acres) %>%
rename(Name = 'Park name',
       geometry = Geometry)

#buff_parks

# import nys parks, then find Buffalo-based state park and fill in relevant information
nys_parks <-
  st_read('data/NYS_Park_Polygons.gpkg', quiet = TRUE) %>%
  st_transform(st_crs(buff_border))

nys_parks_in_buff.idx <-
  nys_parks %>%
  st_intersects(., buff_border) %>% data.frame() %>% .$`row.id`

nys_parks_in_buff <-
  nys_parks[nys_parks_in_buff.idx,] %>%
  filter(Category == 'State Park') %>%
  select(Category, Label, Original) %>%
  mutate(
    Address = '1111 Fuhrmann Blvd',
    'Park Class' = 'Major Park',
    'Maintained by' = 'NYSOPRHP',
    Neighborhood = 'Central',
    'Council District' = 'South',
    Area_Sqmi = units::set_units(st_area(.), 'mi2'),
    Area_Acres = units::set_units(Area_Sqmi, 'acre')
  ) %>%
  rename(
    Name = Label, Year = Original, 'Park Category' = Category, geometry = SHAPE
  ) %>%
  mutate_at(vars(Year), as.numeric) %>%
  relocate('Park Category', .after = 'Park Class') %>%
  relocate(Year, .after = Address)

#nys_parks_in_buff

# set relevant information for all 7 Buffalo parks under county jurisdiction
erie_park_addresses <- c('Aqua Ln', '3781 Main St', '20 Smith St', '1670 Seneca St',
                        '11 Fuhrmann Blvd', 'Hertel Ave', '152 Bailey Ave')
erie_park_years <- c(2013, 1902, 2012, 1994, 1987, 1925, 1996)
erie_park_park_classes <- c('Small Park', 'Large Park', 'Small Park', 'Midsize Park',
                           'Large Park', 'Small Park', 'Small Park')
erie_park_neighborhoods <- c('Black Rock', 'University Heights', 'First Ward',
                             'Seneca-Cazenovia', 'Central',
                             'Black Rock', 'Seneca-Cazenovia')
erie_park_districts <- c('North', 'University', 'Fillmore', 'Lovejoy',
                        'South', 'North', 'Lovejoy')

# read in Buffalo-based county parks and add above info
erie_parks_in_buff <-
  st_read('data/erie_county_parks_in_buffalo.gpkg', quiet = TRUE) %>%

```

```

st_make_valid() %>%
select(NAME) %>%
mutate(
  geom = st_simplify(geom, dTolerance = 0.001),
  Address = erie_park_addresses,
  Year = erie_park_years,
  'Park Class' = 'Major Park',
  'Park Category' = 'County Park',
  'Maintained by' = 'ECPD',
  Neighborhood = erie_park_neighborhoods,
  'Council District' = erie_park_districts,
  Area_Sqmi = units::set_units(st_area(.), 'mi2'),
  Area_Acres = units::set_units(Area_Sqmi, 'acre')
) %>%
rename(Name = NAME, geometry = geom)

#erie_parks_in_buff

# bind all parks data together to create df of all Buffalo-based city, county, and state parks
# filter out parks that are less than 0.25 acres and all monumental/memorial parks
all_buff_parks <-
  bind_rows(buff_parks, nys_parks_in_buff, erie_parks_in_buff) %>%
  filter(
    Area_Acres >=
      units::set_units(0.25, 'acre') &
      (!(grepl('Monument', Name) | (grepl('Memorial', Name))))
  )

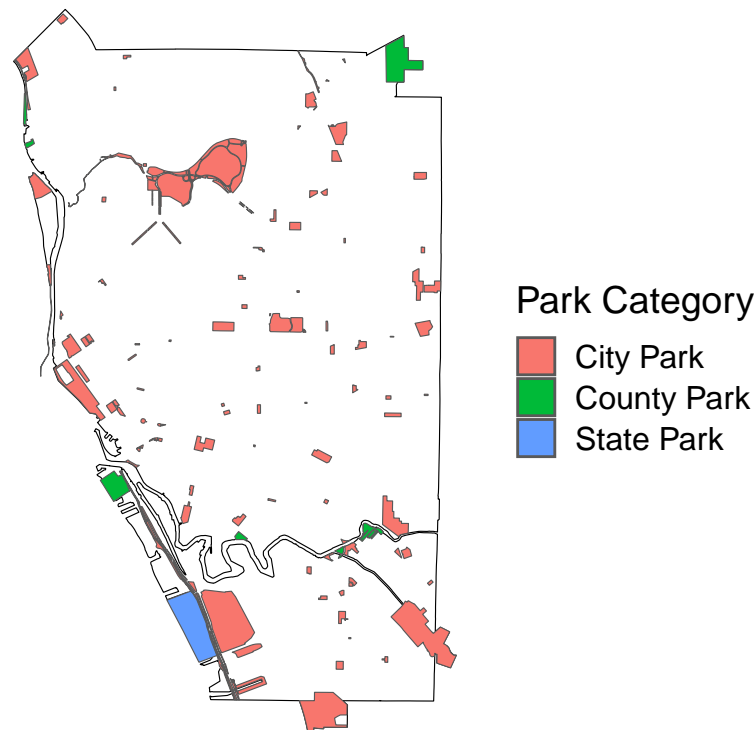
# add years to certain parks that don't currently have them
all_buff_parks[all_buff_parks$Name == 'Outer Harbor Parkway', 'Year'] <- 2013
all_buff_parks[all_buff_parks$Name == 'McCarthy Park', 'Year'] <- 1973
all_buff_parks[all_buff_parks$Name == 'Remembrance Park', 'Year'] <- 2003
all_buff_parks[all_buff_parks$Name == 'Peter St.', 'Year'] <- 2018
all_buff_parks[all_buff_parks$Name == 'Arlington Park', 'Year'] <- 1866
all_buff_parks[all_buff_parks$Name == 'Bristol Emslie Playground', 'Year'] <- 1998
all_buff_parks[all_buff_parks$Name == 'Genesee Gateway Triangle', 'Year'] <- 2008
all_buff_parks[all_buff_parks$Name == 'Unity Island', 'Year'] <- 2000

# 109 TOTAL VALID PARKS
#all_buff_parks

# map of parks by category
ggplot() +
  geom_sf(data = buff_border, color = 'black', fill = 'white') +
  geom_sf(data = all_buff_parks, aes(fill = `Park Category`)) +
  labs(title = 'Parks in Buffalo, NY by Category') +
  theme_map()

```

Parks in Buffalo, NY by Category



ACS Data

```
# load all nhgis files into lists
nhgis_data_files <- list.files('data/nhgis/vars', pattern = 'zip', full.names = TRUE)
nhgis_sf <- list.files('data/nhgis/spatial', pattern = 'zip', full.names = TRUE)

# process data for each year and store in its own df in nhgis_datasets
nhgis_datasets <- list()

for (year in 2010:2012) {
  # get data and sf files for year
  data_file <- nhgis_data_files[grepl(as.character(year), nhgis_data_files)]
  spatial_file <- nhgis_sf[grepl(as.character(year), nhgis_sf)]

  # load the data
  data <- read_nhgis(data_file, verbose = FALSE)

  # process relevant variables depending on the year
  if (year == 2010) {
    data <-
      data %>%
      rename(
        tot_pop = 42,
        white_pop = 43,
        black_pop = 44,
        native_american_pop = 45,
```

```

    asian_pop = 46,
    pacific_islander_pop = 47,
    other_race_pop = 48,
    multiracial_pop = 49,
    non_hispanic_pop = 53,
    hispanic_pop = 54,
    tot_pop_over_25 = 55,
    per_capita_income = 125,
    ppl_in_poverty = 91,
    male_no_schooling = 57,
    male_hs_dip_equiv = 65,
    male_some_college_lt_1yr = 66,
    male_some_college_gt_1yr_no_deg = 67,
    male_associates_deg = 68,
    male_bachelors_deg = 69,
    male_masters_deg = 70,
    male_professional_deg = 71,
    male_doctoral_deg = 72,
    female_no_schooling = 74,
    female_hs_dip_equiv = 82,
    female_some_college_lt_1yr = 83,
    female_some_college_gt_1yr_no_deg = 84,
    female_associates_deg = 85,
    female_bachelors_deg = 86,
    female_masters_deg = 87,
    female_professional_deg = 88,
    female_doctoral_deg = 89
  ) %>%
  select(GEOID, YEAR, COUNTY, 42:49, 53:55, 125, 91, 57, 65:72, 74, 82:89) %>%
  mutate(
    no_schooling = male_no_schooling + female_no_schooling,
    high_school_dip_equiv = male_hs_dip_equiv + female_hs_dip_equiv,
    some_college_lt_1yr = male_some_college_lt_1yr + female_some_college_lt_1yr,
    some_college_gt_1yr_no_deg = male_some_college_gt_1yr_no_deg + female_some_college_gt_1yr_no_deg,
    associates_deg = male_associates_deg + female_associates_deg,
    bachelors_deg = male_bachelors_deg + female_bachelors_deg,
    masters_deg = male_masters_deg + female_masters_deg,
    professional_deg = male_professional_deg + female_professional_deg,
    doctoral_deg = male_doctoral_deg + female_doctoral_deg
  ) %>%
  select(-c(starts_with('male'), starts_with('female')))
} else if (year == 2011) {
  data <-
  data %>%
  rename(
    tot_pop = 43,
    white_pop = 44,
    black_pop = 45,
    native_american_pop = 46,
    asian_pop = 47,
    pacific_islander_pop = 48,
    other_race_pop = 49,
    multiracial_pop = 50,
    non_hispanic_pop = 54,
    hispanic_pop = 55,
    tot_pop_over_25 = 56,
    per_capita_income = 126,

```



```

ppl_in_poverty = 92,
male_no_schooling = 58,
male_hs_dip_equiv = 66,
male_some_college_lt_1yr = 67,
male_some_college_gt_1yr_no_deg = 68,
male_associates_deg = 69,
male_bachelors_deg = 70,
male_masters_deg = 71,
male_professional_deg = 72,
male_doctoral_deg = 73,
female_no_schooling = 75,
female_hs_dip_equiv = 83,
female_some_college_lt_1yr = 84,
female_some_college_gt_1yr_no_deg = 85,
female_associates_deg = 86,
female_bachelors_deg = 87,
female_masters_deg = 88,
female_professional_deg = 89,
female_doctoral_deg = 90
) %>%
select(GEOID, YEAR, COUNTY, 43:50, 54:56, 126, 92, 58, 66:73, 75, 83:90) %>%
mutate(
  no_schooling = male_no_schooling + female_no_schooling,
  high_school_dip_equiv = male_hs_dip_equiv + female_hs_dip_equiv,
  some_college_lt_1yr = male_some_college_lt_1yr + female_some_college_lt_1yr,
  some_college_gt_1yr_no_deg = male_some_college_gt_1yr_no_deg + female_some_college_gt_1yr_no_deg,
  associates_deg = male_associates_deg + female_associates_deg,
  bachelors_deg = male_bachelors_deg + female_bachelors_deg,
  masters_deg = male_masters_deg + female_masters_deg,
  professional_deg = male_professional_deg + female_professional_deg,
  doctoral_deg = male_doctoral_deg + female_doctoral_deg
) %>%
select(-c(starts_with('male'), starts_with('female')))
} else {
data <-
data %>%
rename(
  tot_pop = 43,
  white_pop = 44,
  black_pop = 45,
  native_american_pop = 46,
  asian_pop = 47,
  pacific_islander_pop = 48,
  other_race_pop = 49,
  multiracial_pop = 50,
  non_hispanic_pop = 54,
  hispanic_pop = 55,
  tot_pop_over_25 = 56,
  per_capita_income = 116,
  ppl_in_poverty = 82,
  no_schooling = 57,
  high_school_diploma = 72,
  ged = 73,
  some_college_lt_1yr = 74,
  some_college_gt_1yr_no_deg = 75,
  associates_deg = 76,
  bachelors_deg = 77,

```

```

    masters_deg = 78,
    professional_deg = 79,
    doctoral_deg = 80
  ) %>%
  select(GEOID, YEAR, COUNTY, 43:50, 54:56, 116, 82, 57, 72:80) %>%
  mutate(high_school_dip_equiv = high_school_diploma + ged) %>%
  select(-c(high_school_diploma, ged))
}

# get erie county block groups and simplify GEOID
data <-
  data %>%
  filter(COUNTY == 'Erie County') %>%
  relocate(GEOID, .before = YEAR) %>%
  mutate(GEOID = str_split(GEOID, '15000US', simplify=T)[,2],
         YEAR = str_split(YEAR, '-', simplify = TRUE)[,2]) %>%
  select(-COUNTY)

# process sf file for year, then join to data by GEOID
if (year == 2010) {
  sf <- read_ipums_sf(spatial_file, verbose=F) %>%
    filter(COUNTYFP10 == '029') %>%
    select(GEOID10) %>%
    rename(GEOID = GEOID10) %>%
    st_transform(crs = st_crs(buff_border))

  d <- inner_join(sf, data, by = 'GEOID') %>%
    relocate(geometry, .after = last_col())
} else {
  sf <- read_ipums_sf(spatial_file, verbose=F) %>%
    filter(COUNTYFP == '029') %>%
    select(GEOID) %>%
    st_transform(crs = st_crs(buff_border))

  d <- inner_join(sf, data, by = 'GEOID') %>%
    relocate(geometry, .after = last_col())
}

# add year's dataset to list, then go to next list
nhgis_datasets[[as.character(year)]] <- d
}

# combine nhgis erie county data into single dataframe
erie_blocks.nhgis <-
  bind_rows(nhgis_datasets) %>%
  rename(year = YEAR) %>%
  mutate_at(vars(per_capita_income, year), as.numeric) %>%
  mutate(
    area_sqmi = as.numeric(st_area(.)) * 3.861E-7,
    tot_pop_density = tot_pop / area_sqmi,
    pop_25plus_density = tot_pop_over_25 / area_sqmi,
    total_income = per_capita_income * tot_pop,
    some_college_no_deg = some_college_lt_1yr + some_college_gt_1yr_no_deg,
    bachelors_deg_at_least = bachelors_deg + masters_deg + professional_deg + doctoral_deg,
    less_than_bachelors_deg = tot_pop_over_25 - bachelors_deg_at_least,
    less_than_high_school_grad = tot_pop_over_25 - (high_school_dip_equiv + some_college_no_deg +
                                                    bachelors_deg_at_least)
  )

```

```

) %>%
  relocate(geometry, .after = last_col())

erie_blocks.nhgis[is.na(erie_blocks.nhgis)] <- 0

# get ACS block group data for pre-covid and post-covid years and store in separate dfs
erie_blocks.pre_covid <-
  map(2013:2019, get_acs_block_group_data) %>%
  bind_rows()

erie_blocks.post_covid <-
  map(2020:2023, get_acs_block_group_data) %>%
  bind_rows()

# separate buffalo block groups into their own dataframes
# make sure to get block groups that aren't included in original intersect
block_buff_intersects.nhgis <-
  data.frame(
    st_intersects(st_centroid(erie_blocks.nhgis), buff_border)
  )$row.id

missing_blocks.nhgis <- which(erie_blocks.nhgis$GEOID %in% c(360290011003))

buff_blocks.nhgis <-
  erie_blocks.nhgis[c(block_buff_intersects.nhgis, missing_blocks.nhgis),] %>%
  st_make_valid()

block_buff_intersects.pre_covid <-
  data.frame(
    st_intersects(st_centroid(erie_blocks.pre_covid), buff_border)
  )$row.id

missing_blocks.pre_covid <- which(erie_blocks.pre_covid$GEOID %in% c(360290011003,
                                                                    360290072021,
                                                                    360290058022))

buff_blocks.pre_covid <-
  erie_blocks.pre_covid[c(block_buff_intersects.pre_covid, missing_blocks.pre_covid),] %>%
  st_make_valid()

block_buff_intersects.post_covid <- data.frame(
  st_intersects(st_centroid(erie_blocks.post_covid), buff_border)
)$row.id

missing_blocks.post_covid <- which(erie_blocks.post_covid$GEOID %in% c(360290011003,
                                                                    360290072021,
                                                                    360290058022))

buff_blocks.post_covid <-
  erie_blocks.post_covid[c(block_buff_intersects.post_covid, missing_blocks.post_covid),] %>%
  st_make_valid()

# combine nhgis (2010-2012) and pre-covid (2013-2019) dataframes into
# full Buffalo pre-covid dataframe, and remove empty rows
buff_blocks.pre_covid <-
  bind_rows(buff_blocks.nhgis, buff_blocks.pre_covid) %>%
  filter(!is.na(year))

```

Population-Weighted Areal Interpolation

```
# list of all variables pulled
all_vars <- c(
  'tot_pop', 'white_pop', 'black_pop', 'native_american_pop', 'asian_pop', 'pacific_islander_pop',
  'other_race_pop', 'multiracial_pop', 'non_hispanic_pop', 'hispanic_pop', 'ppl_in_poverty',
  'tot_pop_over_25', 'no_schooling', 'some_college_lt_1yr', 'some_college_gt_1yr_no_deg',
  'associates_deg', 'bachelors_deg', 'masters_deg', 'professional_deg', 'doctoral_deg',
  'high_school_dip_equiv', 'some_college_no_deg', 'bachelors_deg_at_least',
  'less_than_bachelors_deg', 'less_than_high_school_grad', 'per_capita_income'
)

# get erie county block boundaries in 2020, which was when block group boundaries changed
# then get blocks in buffalo boundary
# used to estimate population in each segment of new boundaries
erie_blocks.2020 <- tigris::blocks(state = 'NY', county = 'Erie', year = 2020)
buff_blocks.2020 <- st_intersection(erie_blocks.2020, buff_border, model = 'open')

# block group boundaries in 2020
buff_block_groups.2020 <- buff_blocks.post_covid %>% filter(year == 2020)

# list of pre-covid years
pre_covid_yrs <- unique(buff_blocks.pre_covid$year)

# empty list of interpolated block group boundary dataframes for each pre-covid year
# will bind rows after interpolation
pre_covid_data <- list()

# estimate data for years between 2013-2019 using block group boundaries in 2020-2023
# applies population-weighted areal interpolation technique (interpolation_pw)
for (i in pre_covid_yrs) {
  # get data for current year
  yr_df <-
    buff_blocks.pre_covid[buff_blocks.pre_covid$year == i,] %>%
    select(-c(year, per_capita_income, tot_pop_density,
              pop_25plus_density, area_sqmi))

  # apply population-weighted areal interpolation technique
  # to conform boundaries and estimate the data
  interpolated <-
    yr_df %>%
    interpolate_pw(
      ., buff_block_groups.2020,
      to_id = 'GEOID',
      extensive = TRUE,
      weights = buff_blocks.2020,
      weight_column = 'POP20',
      weight_placement = 'centroid',
      crs = st_crs(buff_blocks.pre_covid)
    ) %>%
    mutate(
      year = as.numeric(i),
      per_capita_income = total_income / tot_pop,
      area_sqmi = as.numeric(st_area(.) * 3.861E-7),
      tot_pop_density = tot_pop / area_sqmi,
      pop_25plus_density = tot_pop_over_25 / area_sqmi
    )
  pre_covid_data[[i]] <- interpolated
}
```

```

)

# add interpolated data for the year to pre_covid_data list
pre_covid_data[[as.character(i)]] <- interpolated
}

# combine interpolated 2010-2019 (which now follow 2020 block group boundaries)
# with post-COVID data, also round interpolated data and remove a couple columns
buff_blocks <-
  bind_rows(pre_covid_data) %>%
  bind_rows(., buff_blocks.post_covid) %>%
  select(-total_income) %>%
  mutate(
    across(all_vars, ~ round(.)),
    pct_below_poverty = ppl_in_poverty / tot_pop,
    pct_bachelors_deg_at_least = bachelors_deg_at_least / tot_pop_over_25,
    pct_less_than_high_school_grad = less_than_high_school_grad / tot_pop_over_25,
    pct_white_pop = white_pop / tot_pop,
    pct_black_pop = black_pop / tot_pop,
    pct_hispanic_pop = hispanic_pop / tot_pop
  ) %>%
  relocate(year, .after = GEOID) %>%
  relocate(geometry, .after = last_col()) %>%
  relocate(area_sqmi, .before = geometry)

```

Final Processing and Data Calculation

```

# initialize list to hold buffer rasters
buffer_rasters <- c()

# set projected coordinate system
proj_crs <- 'EPSG:26917'

# determine number of cameras in block group in year,
# as well as number of violent crimes and violent crime rate
# also process data in general in preparation for logistic regression analysis
buff_blocks <-
  lapply(2010:2023, function(yr) {
    ##print(as.character(yr))

    # blocks in year
    yr_buff_blocks <-
      buff_blocks %>%
      filter(year == yr) %>%
      st_transform(crs = proj_crs) %>%
      mutate(area_sqmi = as.numeric(units::set_units(st_area(.), 'mi2')))) %>%
      st_make_valid()

    # valid cameras in year
    cams <-
      valid_cameras[valid_cameras$year <= yr,] %>%
      st_transform(crs = proj_crs) %>%
      st_make_valid()
  })

```

```

# add remaining cameras without installation dates
# after the year 2018
if (year >= 2018) {
  cams <- bind_rows(cams, st_transform(bpd_cameras[is.na(bpd_cameras$year),],
                                     crs = proj_crs))

  cams[is.na(cams)] <- 2018
} else {
  cams <- cams
}

# get violent crimes in year
crimes <-
  violent_crimes.buff[violent_crimes.buff$Year == yr,] %>%
  st_transform(crs = proj_crs) %>%
  st_make_valid()

# number of cameras in each block
cams_per_block <-
  st_join(
    yr_buff_blocks %>% select(GEOID),
    st_join(cams, yr_buff_blocks, join = st_within) %>%
    group_by(GEOID) %>%
    summarize(Num_Cameras = n(), .groups = 'drop'),
    suffix = c('', '_y')
  ) %>%
  select(-GEOID_y)

# number of violent crimes in each block
crimes_per_block <-
  st_join(
    yr_buff_blocks %>% select(GEOID),
    st_join(crimes, yr_buff_blocks, join = st_within) %>%
    group_by(GEOID) %>%
    summarize(num_violent_crimes = n(), .groups = 'drop'),
    suffix = c('', '_y')
  ) %>%
  select(-GEOID_y)

# parks in the year
parks_yr <-
  all_buff_parks[all_buff_parks$Year <= yr,] %>%
  st_transform(crs = proj_crs) %>%
  st_make_valid()

# dissolved buffer of parks
parks_buffer <-
  st_buffer(parks_yr, dist = 402.336) %>%
  st_union() %>%
  st_make_valid() %>%
  st_transform(crs = proj_crs)

# get extent using the buffalo block groups
ext <- st_bbox(yr_buff_blocks)

# create a temporary raster of the study area
parks_temp_raster <- rast(
  extent = ext,

```

```

    resolution = 10,
    crs = proj_crs
  )

# rasterize buffer (cells inside buffer = 1, outside = 0)
# then add to buffer_rasters vector
buffer_raster <- rasterize(vect(parks_buffer), parks_temp_raster,
                           field = 1, background = 0, progress = 0)

buffer_rasters[[as.character(yr)]] <- buffer_raster

# calculate percent walking distance park proximity coverage
pct_park_coverage <- exactextractr::exact_extract(buffer_raster,
                                                    yr_buff_blocks,
                                                    'mean', progress = FALSE)

# final joining back to year data and creating remaining variables
# then return the data for the year
return(
  yr_buff_blocks %>%
  st_join(cams_per_block, st_nearest_feature, suffix = c('', '_y')) %>%
  select(-GEOID_y) %>%
  st_join(crimes_per_block, st_nearest_feature, suffix = c('', '_y')) %>%
  select(-GEOID_y) %>%
  mutate(
    Num_Cameras = replace_na(Num_Cameras, 0),
    num_violent_crimes = replace_na(num_violent_crimes, 0),
    pct_park_proximity = pct_park_coverage,
    violent_crime_rate = (num_violent_crimes / tot_pop) * 1000,
    has_crime = ifelse(num_violent_crimes > 0, 1, 0),
    has_camera = ifelse(Num_Cameras > 0, 1, 0),
    Cameras_Per_1000 = (Num_Cameras / tot_pop) * 1000
  ) %>%
  relocate(geometry, .after = last_col()) %>%
  st_transform(crs = 'EPSG:4269')
)
}) %>%
bind_rows() %>%
select(-ends_with('_y')) %>%
st_transform(crs = proj_crs) # transform to UTM Zone 17

```

Results

GLMM Model Using glmer

A generalized linear mixed model (GLMM) was fitted using the *glmer* function in the *lme4* package to estimate the probability of violent crime occurring in a block group of Buffalo, New York using several socio-environmental predictors, while accounting for unobserved heterogeneity across space via random intercepts at the block group (GEOID) level. There are 4,060 observations that represent one

of the 290 block groups for a given year between 2010-2023. For training purposes, 3,234 observations were fed into the model representing 231 block groups for each year in the study period.

```
# scale quantitative independent variables, then place each observation in a year group
buff_blocks.scaled <-
  buff_blocks %>%
  mutate(
    year_group = case_when(
      year <= 2012 ~ '2010-2012',
      year <= 2016 ~ '2013-2016',
      year <= 2019 ~ '2017-2019',
      TRUE ~ '2020-2023'
    ),
    per_capita_income = scale(per_capita_income)[,1],
    pct_below_poverty = scale(pct_below_poverty)[,1],
    pct_bachelors_deg_at_least = scale(pct_bachelors_deg_at_least)[,1],
    pct_black_pop = scale(pct_black_pop)[,1],
    pct_park_proximity = scale(pct_park_proximity)[,1]
  )

# get centroids of unique block groups
unique_bgs <-
  buff_blocks.scaled %>%
  group_by(GEOID) %>%
  slice(1) %>% # use first row of each GEOID grouping
  ungroup()
centroids <- st_centroid(unique_bgs)
coords <- st_coordinates(centroids)[, 1:2]

# create neighbor list and weights
nb <- poly2nb(unique_bgs)
lw <- nb2listw(nb, style = 'W')

# create lag of pct_black_pop
lag_black <- lag.listw(lw, unique_bgs$pct_black_pop)

# add lag, then select columns and drop geometry
buff_blocks.lagged <-
  unique_bgs %>%
  st_drop_geometry() %>%
  select(GEOID) %>%
  mutate(pct_black_pop_lag = lag_black)

# merge lag data into df for model
buff_blocks.scaled_lagged <-
  buff_blocks.scaled %>%
  left_join(buff_blocks.lagged, by = 'GEOID', suffix = c('', ''))

# GEOIDS and check for presence for any crime or camera
# assign one of 4 strata to each GEOID (e.g., has had crime AND has a camera)
geoid_strata <-
  buff_blocks.scaled_lagged %>%
  group_by(GEOID) %>%
  summarize(
    had_crime = ifelse(sum(has_crime, na.rm=T) > 0, 1, 0),
    had_camera = ifelse(sum(has_camera, na.rm=T) > 0, 1, 0)
```



```

) %>%
mutate(stratum = paste0('crime', had_crime, '_camera', had_camera))

# specify seed for consistent number generation
set.seed(1234)

# get 80% of GEOIDS that match each stratum definition
train_geoids <-
  geoid_strata %>%
  group_by(stratum) %>%
  slice_sample(prop = 0.8) %>%
  pull(GEOID)

# split data into training and testing datasets
train_data <- buff_blocks.scaled_lagged %>% filter(GEOID %in% train_geoids)
test_data <- buff_blocks.scaled_lagged %>% filter(!(GEOID %in% train_geoids))

# logistic regression with block group as random effect,
# year_group as fixed effect, spatially lagged covariate of pct_black_pop,
# and scaled data
model_glmer <- glmer(
  has_crime ~ has_camera + per_capita_income + pct_below_poverty +
    pct_bachelors_deg_at_least + pct_black_pop + pct_black_pop_lag +
    pct_park_proximity + factor(year_group) + (1 | GEOID),
  data = train_data,
  family = binomial(link = "logit"),
  control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e5))
)

# save model and training and test data
#saveRDS(model_glmer, 'model/model.rds')
#st_write(train_data, 'model/train_data.gpkg')
#st_write(test_data, 'model/test_data.gpkg')

```

In mathematical terms, the model can be expressed as such:

$$\begin{aligned}
\text{logit}(\Pr(\text{has_crime}_{it} = 1)) = & \beta_0 + \beta_1 \cdot \text{has_camera}_{it} + \beta_2 \cdot \text{per_capita_income}_{it} + \beta_3 \cdot \text{pct_below_poverty}_{it} \\
& + \beta_4 \cdot \text{pct_bachelors_deg_at_least}_{it} + \beta_5 \cdot \text{pct_black_pop}_{it} + \beta_6 \cdot \text{pct_black_pop_lag}_{it} \\
& + \beta_7 \cdot \text{pct_park_proximity}_{it} + \sum_{j=1}^3 \gamma_j \cdot \mathbb{I}(\text{year_group}_{it} = j) + u_i
\end{aligned}$$

Where:

- has_crime_{it} is a binary outcome for block group i in time t
- β_0 is the model intercept

- β_1, \dots, β_7 are fixed-effect coefficients
- γ_j are coefficients for the year group fixed effects
- $u_i \sim \mathcal{N}(0, \sigma_u^2)$ is the random intercept
- $I(\cdot)$ is an indicator function for year groups

It is imperative to conduct diagnostics tests to assess the model's fit, stability, and potential for interpretation.

Diagnostic Tests

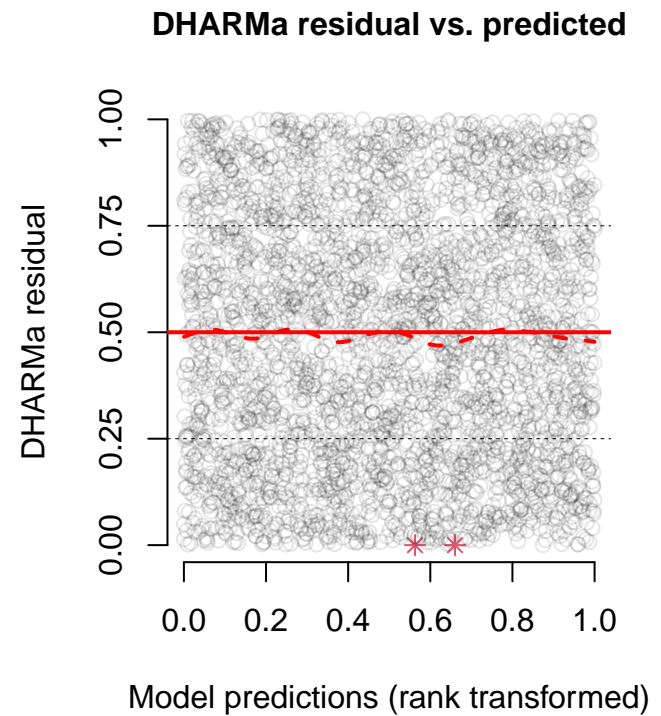
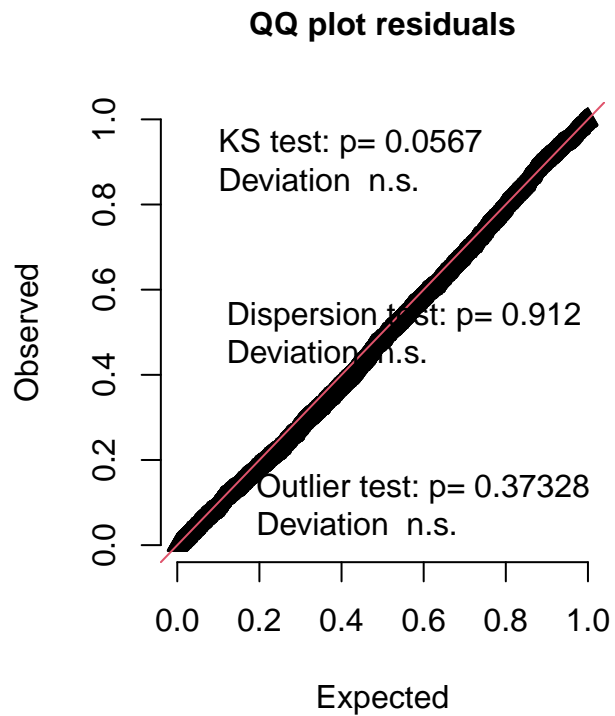
```
# load fitted model and train/test datasets
model <- readRDS('model/model.rds')
#summary(model)

train_data <- st_read('model/train_data.gpkg', quiet = TRUE)
test_data <- st_read('model/test_data.gpkg', quiet = TRUE)

train_geoids <- unique(train_data$GEOID)
test_geoids <- unique(test_data$GEOID)

# simulate residuals to help assess model fit,
# then show diagnostic plots
model.simres <- simulateResiduals(model)
plot(model.simres)
```

DHARMA residual

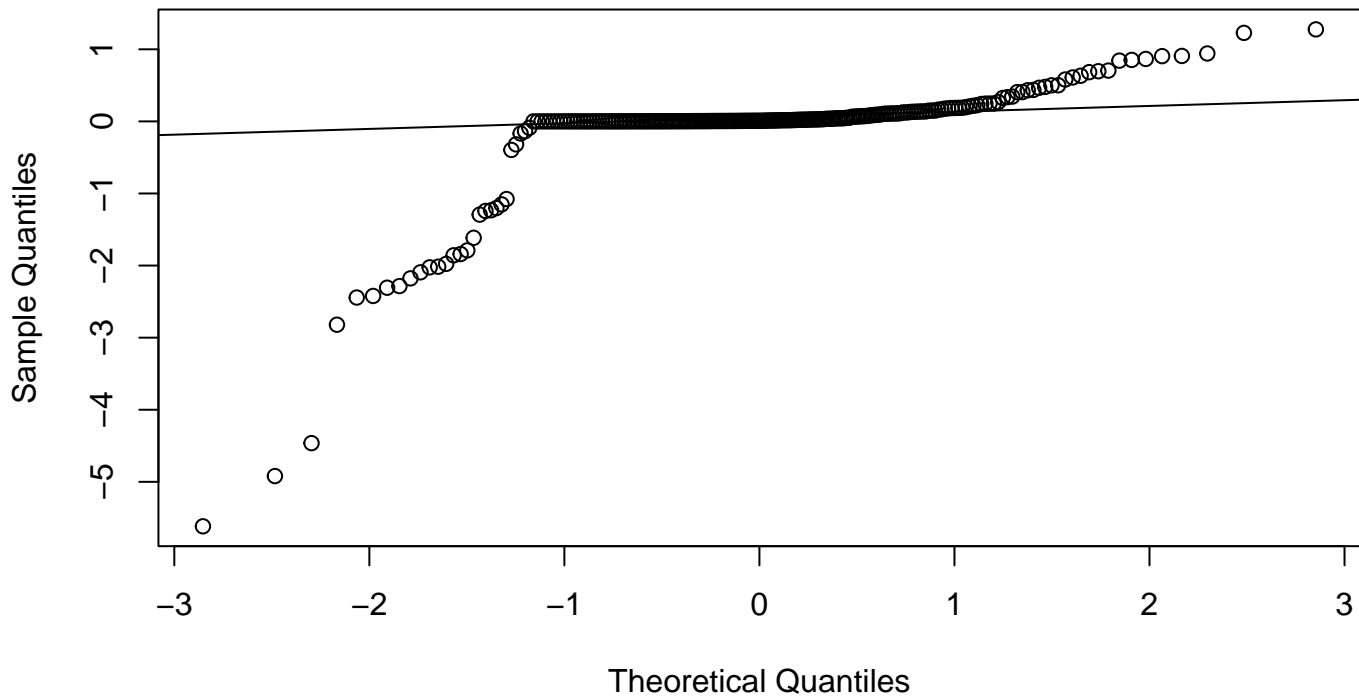


```
#####
# diagnose assumption of normally distributed estimated random effects
#####

# extract random intercepts
random_intercepts <- ranef(model)$GEOID$(Intercept)`

# qq plot of random effects to check normality
qqnorm(random_intercepts)
qqline(random_intercepts)
```

Normal Q-Q Plot



```
# check for multicollinearity
vif(model)
```

	GVIF	Df	$GVIF^{1/(2 \cdot Df)}$
has_camera	1.056929	1	1.028071
per_capita_income	2.100291	1	1.449238
pct_below_poverty	1.460462	1	1.208496
pct_bachelors_deg_at_least	1.689829	1	1.299934
pct_black_pop	1.753034	1	1.324022
pct_black_pop_lag	1.698463	1	1.303251
pct_park_proximity	1.051647	1	1.025498
factor(year_group)	1.204214	3	1.031456

```
#####
```

```
# Spatial Autocorrelation (Moran's I test)
```

```
#####
```

```
# get one row per block group
```

```
buff_blocks.ranef <-
  train_data %>%
  filter(GEOID %in% train_geoids) %>%
  distinct(GEOID, .keep_all = TRUE)
```

```
# add random intercepts
```

```
buff_blocks.ranef$random_intercept <- random_intercepts
```

```
# centroids and neighbors
centroids <- st_centroid(buff_blocks.ranef)
coords <- st_coordinates(centroids)
knn <- knearneigh(coords, k = 5) %>% knn2nb()
lw_knn <- nb2listw(knn)

# Moran's I test on random intercept
moran_test <- moran.test(buff_blocks.ranef$random_intercept, lw_knn)

moran_test
```

Moran I test under randomisation

```
data: buff_blocks.ranef$random_intercept
weights: lw_knn
```

Moran I statistic standard deviate = 0.24417, p-value = 0.4036

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
0.004815242	-0.004347826	0.001408346

Diagnostic tests indicate that the generalized linear mixed model (GLMM) fits the data well and that key assumptions are reasonably satisfied. The DHARMA residual diagnostics reveal no significant issues, with residuals closely approximating a uniform distribution. The Kolmogorov–Smirnov test returned a non-significant result ($p = 0.057$) signifying no significant misspecification, and neither the dispersion test ($p = 0.912$) nor the outlier test ($p = 0.373$) detected evidence of model misfit or overdispersion. The residuals vs predicted values plot shows a mostly uniform scatter of residuals across predicted probabilities, supporting the assumption of homoscedasticity. These findings suggest that the model’s structure adequately captures the variance in the outcome and that both fixed and random effects are appropriately specified. The adjusted Generalized Variance Inflation Factor (GVIF) values are all under 2, suggesting no evidence of problematic multicollinearity. The normal Q-Q plot of the random intercepts shows a generally acceptable distribution but reflects mild deviation in the lower and upper tails, indicating mild non-normality. This, however, is not uncommon in spatial data and not very significant given the larger sample size and the model’s conformity to other assumptions.

Importantly, the spatially lagged percent Black population was included in the model to account for potential spatial dependence in neighborhood-level demographics. This addition helped to alleviate residual spatial autocorrelation, as confirmed by the updated Moran’s I test on the random intercepts.

The Moran's I statistic was tiny ($I = 0.0048$) and statistically non-significant ($p = 0.404$), indicating that no substantial spatial clustering remains in the unexplained variation across block groups. This suggests that the inclusion of a spatially structured covariate successfully addressed previously unmodeled spatial dependence and improves confidence in the validity of the model's inferences.

Fixed Effects

```
# extract fixed effects
# then exponentiate to get odds ratios
fixef_vals <- fixef(model)
odds_ratios <- exp(fixef_vals)

# get standard errors
se_vals <- summary(model)$coefficients[, 'Std. Error']

# compute 95% confidence interval
lower_ci <- exp(fixef_vals - 1.96 * se_vals)
upper_ci <- exp(fixef_vals + 1.96 * se_vals)

# get fixed effects data
fixed_effects <-
  data.frame(
    Estimate = fixef_vals,
    Odds_Ratio = odds_ratios,
    CI_Lower = lower_ci,
    CI_Upper = upper_ci,
    p_value = summary(model)$coefficients[, 'Pr(>|z|)']
  )

# odds ratio table
odds_table <-
  fixed_effects %>%
  mutate(
    Predictor = rownames(.),
    Estimate = round(Estimate, 3),
    Odds_Ratio = round(Odds_Ratio, 2),
    CI = paste0('[', round(CI_Lower, 2), ', ', round(CI_Upper, 2), ']'),
    p_value = formatC(p_value, format = 'e', digits = 2)
  ) %>%
  select(Predictor, Estimate, Odds_Ratio, CI, p_value) %>%
  gt() %>%
  tab_header(
    title = 'Odds Ratios from GLMM Model',
    subtitle = '95% Confidence Intervals and p-values'
  ) %>%
  fmt_markdown(columns = c(CI)) %>%
  cols_label(
    Predictor = 'Predictor',
    Estimate = 'Coefficient',
    Odds_Ratio = 'Odds Ratio',
    CI = '95% CI',
```

```

    p_value = 'p-value'
  ) %>%
  tab_options(
    table.font.size = 'small',
    heading.align = 'left'
  ) %>%
  as_latex()

```

odds_table

Odds Ratios from GLMM Model

95% Confidence Intervals and p-values

Predictor	Coefficient	Odds Ratio	95% CI	p-value
(Intercept)	8.563	5235.01	[604.86, 45308.28]	7.43e-15
has_camera	3.029	20.67	[2.96, 144.35]	2.26e-03
per_capita_income	-0.675	0.51	[0.3, 0.85]	1.02e-02
pct_below_poverty	-0.017	0.98	[0.48, 2.01]	9.62e-01
pct_bachelors_deg_at_least	-0.654	0.52	[0.28, 0.96]	3.53e-02
pct_black_pop	-0.191	0.83	[0.19, 3.55]	7.97e-01
pct_black_pop_lag	1.726	5.62	[1.08, 29.1]	3.98e-02
pct_park_proximity	0.278	1.32	[0.76, 2.29]	3.22e-01
factor(year_group)2013-2016	0.195	1.22	[0.16, 9.35]	8.51e-01
factor(year_group)2017-2019	-2.197	0.11	[0.02, 0.56]	7.82e-03
factor(year_group)2020-2023	-3.088	0.05	[0.01, 0.23]	1.68e-04

```

# add signifance to highlight predictors that are significant
fixed_effects_plot_data <-
  fixed_effects %>%
  mutate(
    Predictor = rownames(.),
    Significance = ifelse(p_value < 0.05, 'Significant', 'Not Significant')
  )

# plot fixed effects with confidence intervals and significance
fixed_effects_plt <-
  ggplot(fixed_effects_plot_data,
    aes(x = reorder(Predictor, Odds_Ratio), y = Odds_Ratio)) +
  geom_point(aes(color = Significance), size = 3) +
  geom_errorbar(aes(ymin = CI_Lower, ymax = CI_Upper, color = Significance), width = 0.2) +
  geom_hline(yintercept = 1, linetype = 'dashed', color = 'red') +

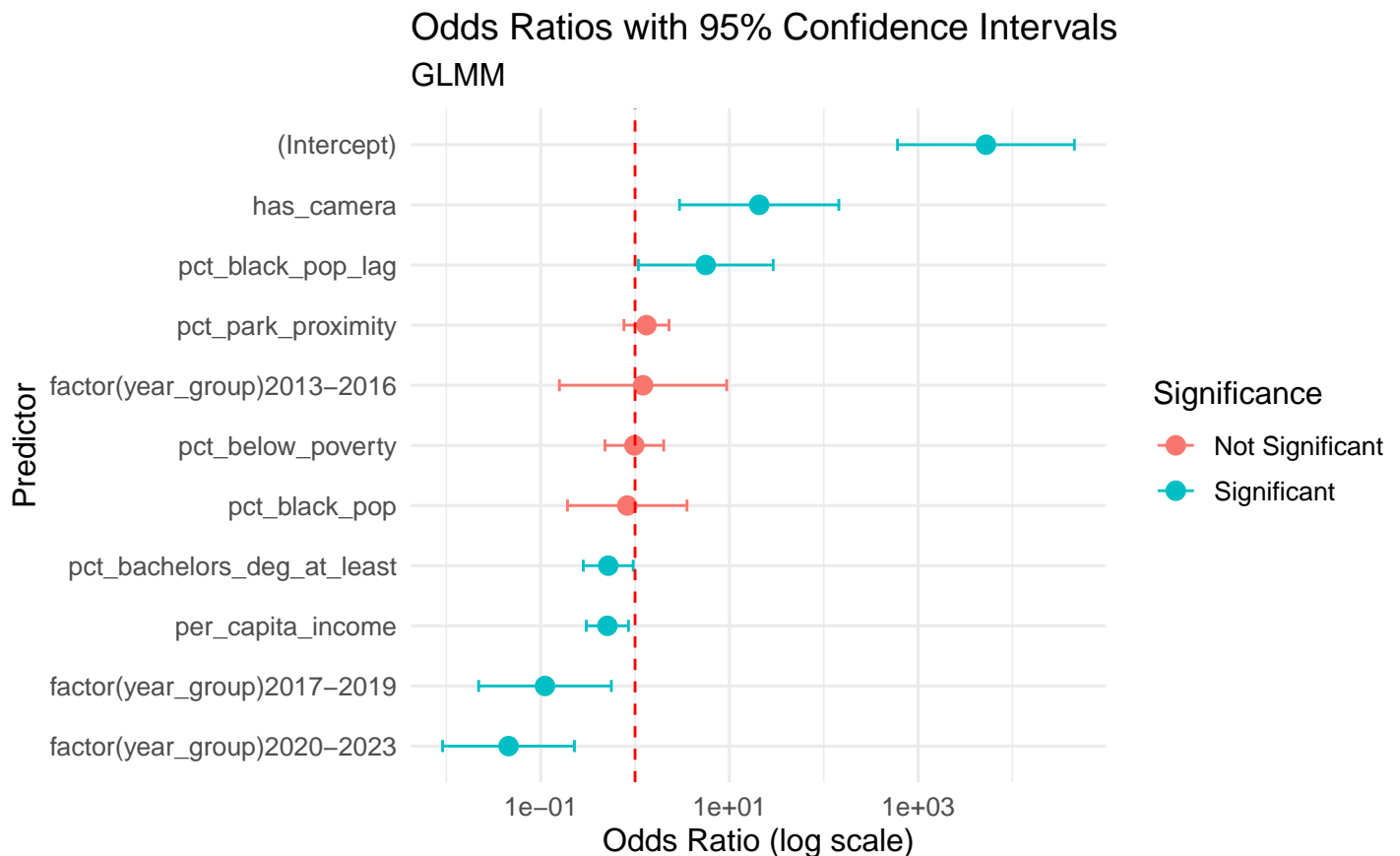
```

```

scale_y_log10() + # log scale for better visualization
coord_flip() +
labs(
  title = 'Odds Ratios with 95% Confidence Intervals',
  subtitle = 'GLMM',
  x = 'Predictor',
  y = 'Odds Ratio (log scale)',
  color = 'Significance'
) +
theme_minimal(base_size = 12)

fixed_effects_plt

```



The model performs well overall, with a relatively low AIC (386.2) and a significant random intercept variance ($\sigma = 4.097$), suggesting that baseline crime risk varies meaningfully across block groups even after adjusting for covariates.

Several fixed effects stand out as statistically significant. Most notably, block groups with a surveillance camera had 20 times greater odds of reporting violent crime (OR = 20.67, $p = 0.002$), which likely reflects reactive placement strategies or a higher probability of detection in surveilled areas rather than a causal effect. Higher per capita income was associated with reduced odds of violent crime

(OR = 0.51, $p = 0.010$), and higher educational attainment (measured by the percentage with at least a bachelor's degree) was similarly protective (OR = 0.52, $p = 0.035$), consistent with established socioeconomic theories of crime. Interestingly, the lagged percentage of Black population was associated with substantially higher odds of violent crime (OR = 5.62, $p = 0.040$), indicating that surrounding demographic context elevates the violent crime risk of a given block group, even after adjusting for its own characteristics. The direct effect of a block group's own percentage of Black population, in contrast, had no significant effect ($p = 0.797$), supporting the use of lagged variables to better capture structural or delayed effects.

Other variables, such as poverty rate, park proximity, and the block group's own Black population percentage, did not yield statistically significant associations in this model. These null findings, in combination with significant predictors, underscore the complexity of spatial crime dynamics and the need to interpret associations with caution, especially when using observational data. The temporal fixed effects (year groupings) show a clear decline in crime risk over the 14-year study period. Compared to the baseline period (2010-2012), the odds of violent crime were significantly lower during 2017-2019 (OR = 0.11, $p = 0.0078$) and further lowered during 2020-2023 (OR = 0.05, $p < 0.001$), after controlling for all other predictors. These results suggest a substantial decrease in violent crime odds in recent years, which many reflect broader shifts such as policing strategies, demographic changes, or social conditions over time. The 2013-2016 period was not significantly different from the baseline.

Overall, the model suggests that crime risk is heavily patterned by structural factors like income and educational attainment, and that camera presence strongly correlates with violent incidents, although this is likely due to endogeneity. The large and significant random effects variance confirms that important localized variation in crime remains unexplained, pointing to the value of spatial random effects in capturing latent neighborhood-level risk. The model would also suggest a significant decline in violent crime risk starting sometime after 2016.

Accuracy of Model

```
# get predicted probabilities
pred_probs <- predict(model, newdata = test_data,
                      type = 'response', allow.new.levels = TRUE)
```

```

# convert probabilities to binary predictions
pred_class <- ifelse(pred_probs > 0.5, 1, 0)

# create confusion matrix, then extract values
conf_matrix <- table(Predicted = pred_class, Actual = test_data$has_crime)

TN <- conf_matrix[1,1]
FP <- conf_matrix[2,1]
FN <- conf_matrix[1,2]
TP <- conf_matrix[2,2]

# create confusion matrix dataframe
confusion_df <- tibble(
  Outcome = c('True Negatives', 'False Positives',
              'False Negatives', 'True Positives'),
  Count = c(TN, FP, FN, TP)
)

# create gt table for confusion matrix
confusion_gt <-
  confusion_df %>%
  gt() %>%
  tab_header(title = 'Confusion Matrix: Crime Prediction Model') %>%
  fmt_number(columns = Count, decimals = 0) %>%
  as_latex()

# calculate metrics
accuracy <- (TP + TN) / sum(conf_matrix)
sensitivity <- TP / (TP + FN)
specificity <- TN / (TN + FP)

# create metrics dataframe
metrics_df <- tibble(
  Metric = c('Accuracy', 'Sensitivity (Recall)', 'Specificity'),
  Value = c(accuracy, sensitivity, specificity)
)

# create gt table for metrics
metrics_gt <-
  metrics_df %>%
  mutate(Value = round(Value, 4)) %>%
  gt() %>%
  tab_header(title = 'Model Performance Metrics') %>%
  fmt_percent(columns = Value, decimals = 2) %>%
  as_latex()

# show gt tables of confusion matrix and metrics
confusion_gt

```

Confusion Matrix: Crime Prediction Model

Outcome	Count
True Negatives	0
False Positives	16
False Negatives	4
True Positives	806

metrics_gt

Model Performance Metrics

Metric	Value
Accuracy	97.58%
Sensitivity (Recall)	99.51%
Specificity	0.00%

The accuracy results demonstrate high performance in identifying violent crime occurrences within the dataset. The model achieved an overall accuracy of 97.58%, with a sensitivity (recall) of 99.51%, indicating that it correctly classified nearly all block group–year observations where violent crime occurred. However, the model exhibited a specificity of 0.00%, meaning it did not correctly identify any of the few observations without crime. This low specificity is largely attributable to the extreme class imbalance in the data. Specifically, out of 4,060 total block group–year observations, only 83 did not report any violent crime, and in the test dataset, just 4 observations were classified as having no crime.

This distributional imbalance skews the specificity metric and makes it less meaningful in this context. However, the aim of the study is not to build a balanced binary classifier, but to identify significant predictors of violent crime and uncover spatial patterns in unexplained risk through random effects modeling. Therefore, the model’s high sensitivity and its ability to detect where crime is likely to occur are far more relevant to the study’s objectives. The model identifies key social and spatial risk fac-

tors while providing insight into how existing surveillance infrastructure aligns with areas of elevated, unexplained violent crime risk.

Random Effects

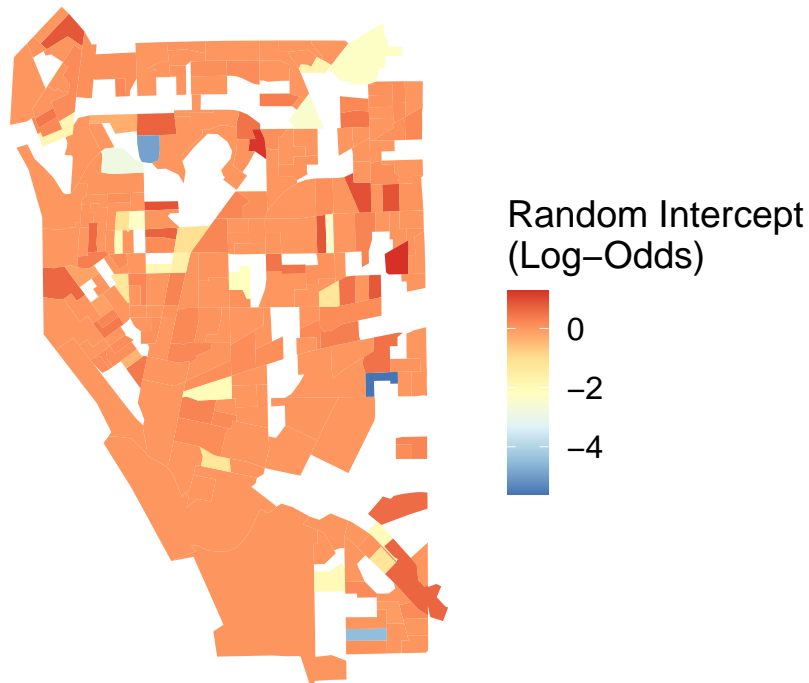
Investigating Persistent Positive Random Effects by Block Group

```
# make map of random effects to see which areas have a higher risk of crime
# than expected possibly due to unmeasured factors, and which areas have lower
# than expected crime risk
ranef_map <-
  ggplot(buff_blocks.ranef) +
  geom_sf(aes(fill = random_intercept), color = NA) +
  scale_fill_distiller(
    palette = 'RdYlBu',
    name = 'Random Intercept\n(Log-Odds)',
    na.value = 'grey80',
    direction = -1
  ) +
  labs(
    title = 'Random Effects by Block Group',
    subtitle = 'Baseline Crime Risk After Controlling for Predictors'
  ) +
  theme_map()

ranef_map
```

Random Effects by Block Group

Baseline Crime Risk After Controlling for Predictors



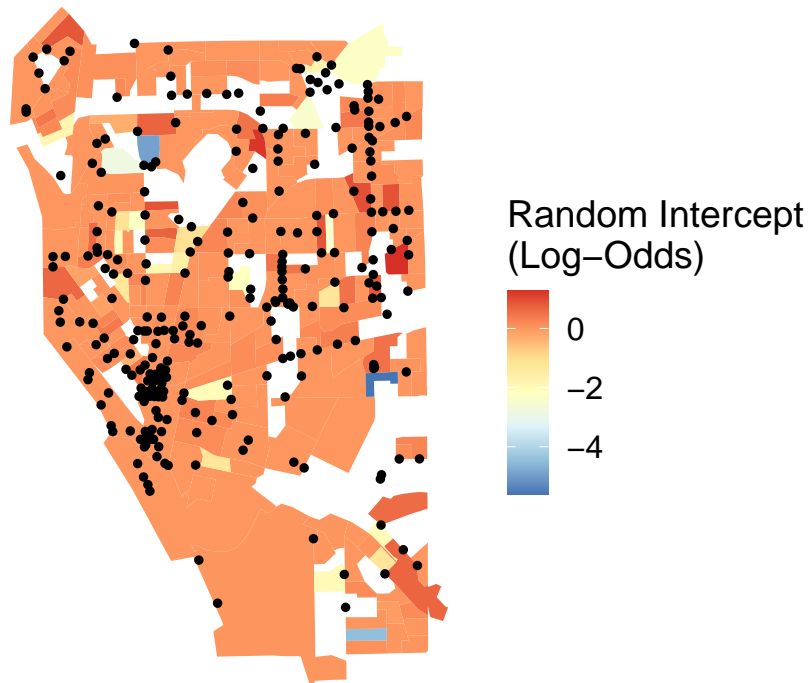
High Unexplained Risk Block Groups vs. Current Police Cameras

```
# create map of random effects with full bpd cameras dataset overlaid on top
ranef_cams_map <-
  ggplot() +
  geom_sf(data = buff_blocks.ranef, aes(fill = random_intercept), color = NA) +
  geom_sf(data = bpd_cameras %>%
    st_transform(crs = st_crs(buff_blocks.ranef)),
    color = 'black', size = 1) +
  scale_fill_distiller(
    palette = 'RdYlBu',
    name = 'Random Intercept\n(Log-Odds)',
    na.value = 'grey80',
    direction = -1
  ) +
  labs(
    title = 'Crime Risk by Block Group',
    subtitle = 'BPD Surveillance Camera Locations'
  ) +
  theme_map()

ranef_cams_map
```

Crime Risk by Block Group

BPD Surveillance Camera Locations



Top 30 Block Groups with Most Unexplained Violent Crime Risk

```
# get top-30 block groups in terms of highest random intercept
top_30_bgs <-
  buff_blocks.ranef %>%
  select(GEOID, has_camera, random_intercept) %>%
  arrange(desc(random_intercept)) %>%
  head(n = 30)

# map of top-30 block groups in terms of highest random intercept
top_30_bgs_map <-
  ggplot() +
  geom_sf(data = buff_border %>%
    st_transform(crs = st_crs(top_30_bgs)),
    color = 'black', fill = 'white') +
  geom_sf(data = top_30_bgs, aes(fill = random_intercept)) +
  geom_sf(data = bpd_cameras %>%
    st_transform(crs = st_crs(top_30_bgs)),
    color = 'black', size = 1) +
  scale_fill_distiller(
    palette = 'RdYlGn',
    name = 'Random Intercept\n(Log-Odds)',
    na.value = 'grey80',
    direction = -1
  ) +
  labs(
```

```

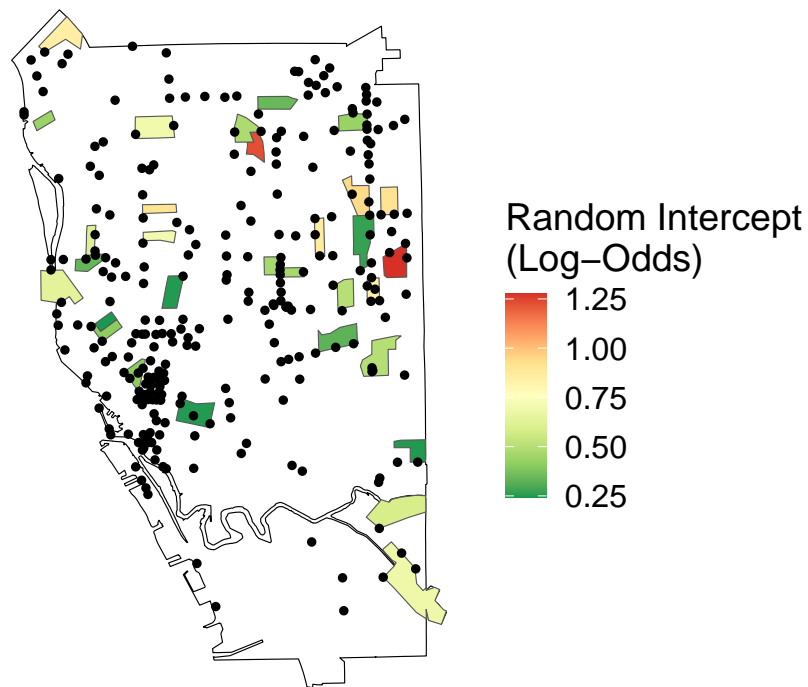
    title = 'Most Unexplained Violent Crime Risk',
    subtitle = 'Top 30 with BPD Surveillance Camera Locations'
) +
theme_map()

top_30_bgs_map

```

Most Unexplained Violent Crime Risk

Top 30 with BPD Surveillance Camera Locations



The sequence of maps provides valuable spatial insights into block group-level variation in violent crime risk across Buffalo, New York after adjusting for a range of predictors. The first map, titled “Random Effects by Block Group,” shows the estimated random intercepts from the mixed-effects model. These random effects capture the baseline crime risk in each block group that remains unexplained by fixed effects such as income, educational attainment, racial composition, and the presence of police cameras. Warmer colors (closer to red) indicate higher unexplained log-odds of violent crime, suggesting areas where violence is more prevalent than the included predictors would suggest. In contrast, cooler tones (blues) mark block groups with substantially lower-than-expected crime risk. The fact that these deviations persist even after controlling for known predictors underscores the presence of unmeasured local influences or structural factors.

The second map, “Crime Risk by Block Group: BPD Surveillance Camera Locations,” overlays the same random effects with the locations of Buffalo Police Department surveillance cameras (shown as black dots). This comparison facilitates a visual assessment of whether surveillance cameras are systematically deployed in high-risk areas. While some high-risk block groups have a visible camera presence, there are also notable high-risk areas without nearby cameras, high-risk areas with cameras only on the periphery, and areas saturated with surveillance in relatively low-risk neighborhoods. This discrepancy suggests potential inefficiencies or misalignments in camera deployment strategy, raising questions about the criteria used for surveillance infrastructure decisions.

The third map, “Most Unexplained Violent Crime Risk,” isolates the 30 block groups with the highest positive random effects, representing the most disproportionately at-risk areas after accounting for all model covariates. These block groups stand out as locations where crime is significantly elevated beyond what would be expected from measurable social, economic, or demographic variables. There are specific areas of the city that contain more block groups with more unexplained risk, which include the east side and areas closer to downtown. Overlaying camera locations again highlights that several of these highest-risk block groups lack any police surveillance presence or are only covered by surveillance cameras on their peripheries, revealing important gaps in monitoring coverage. This insight may inform future place-based interventions, resource allocations, or deeper community-level analyses to uncover latent risk factors not captured in the current model.

Together, these maps imply that while known predictors explain much of the variation in violent crime across Buffalo, persistent pockets of unexplained high risk remain (especially on the east side and close to downtown). These areas merit further investigation and could represent strategic priorities for resource allocation, targeted interventions, or community engagement to address gaps in surveillance or crime prevention infrastructure.

Discussion and Conclusions

This study employed a generalized linear mixed model (GLMM) to investigate the spatial and socioeconomic determinants of violent crime in Buffalo, New York, from 2010 to 2023. Drawing from a dataset that incorporated both fixed effects (e.g., income, educational attainment, racial composition,

park proximity) and random intercepts for block groups, the model attempted to explain variation in violent crime while accounting for spatial and temporal dependencies. The findings reveal complex interactions between socioeconomic factors, law enforcement infrastructure, and persistent spatial disparities in crime risk.

Police surveillance cameras were significantly associated with increased odds of violent crime (OR = 20.67, $p = 0.002$). While this might appear counterintuitive, it likely reflects increased detection and reporting rather than increased incidence, as heavily surveilled areas may experience heightened police presence and administrative attention (Gómez et al., 2021; Priks, 2015). Higher per capita income (OR = 0.51, $p = 0.010$) and educational attainment (OR = 0.52, $p = 0.035$) were associated with lower crime odds, supporting findings from prior research emphasizing the protective role of socioeconomic advantage (Lochner, 2010; Sampson et al., 1997). Conversely, spatial lags of racial composition emerged as significant (OR = 5.62, $p = 0.040$), underscoring the importance of spatially dependent contextual influences on crime, a finding supported by studies showcasing the utility of spatially lagged covariates in capturing neighborhood spillover effects (Anselin et al., 2000).

Random effects mapping revealed that areas of elevated unexplained crime risk were disproportionately concentrated in Buffalo's East Side and around downtown, a region historically associated with redlining and systemic disinvestment (Partnership for the Public Good, 2018). These findings suggest that historical and structural inequalities continue to shape crime patterns, even after controlling for contemporary socioeconomic variables. This aligns with evidence that spatially clustered deprivation and past discriminatory policies can produce long-lasting effects on neighborhood safety and wellbeing (Sharkey, 2013).

Despite its strengths, the study has certain limitations that are worth mentioning. First, violent crime was treated as a binary outcome, which obscures distinctions between types of offenses such as homicide, assault, and robbery. Future research should disaggregate crime categories for more targeted insights. Second, although population-weighted areal interpolation through centroid assignment was sufficient for this analysis, alternative interpolation methods may yield greater accuracy, especially in areas with irregular geometries or sparse population data (Reibel & Bufalino, 2005). Third, the model did not incorporate potential predictors such as indicators of social cohesion or disorder, including 311 service request data, which other researchers have found useful in predicting crime (Kontokosta, 2018).

Unfortunately, Buffalo's 311 dataset was too incomplete to include.

Attempts to perform a difference-in-differences analysis to evaluate the causal effect of camera installation were unsuccessful due to multicollinearity and incomplete installation date data. Many cameras lacked installation dates, forcing the assumption that they were installed after 2017, which likely introduced bias and undermined model stability. Without reliable treatment timing, DiD estimates cannot robustly differentiate pre- and post-treatment effects (Angrist & Pischke, 2009). Nevertheless, difference-in-differences methods remain a powerful tool for evaluating the causal impact of interventions like surveillance camera installation, provided accurate timing and treatment group classification can be established (Gómez et al., 2021; Priks, 2015).

Simulating the model multiple times (e.g., 1,000 iterations) could enhance stability and reduce the loss of block groups in cross-validation splits. This approach has been recommended to ensure better generalizability and mitigate issues related to data sparsity in multilevel models (Arel-Bundock, 2020). Despite efforts to stratify by crime and surveillance presence, the small number of block groups without crime (only 83 of 4,060 observations) presented challenges for evaluating specificity. Future studies might oversample or simulate these low-crime areas to better understand their distinguishing characteristics.

The model's fixed effects provide evidence that socioeconomic conditions and surveillance presence significantly influence violent crime risk, but the random effects show that these predictors cannot explain all spatial variance. Many high-risk areas lack surveillance infrastructure, raising questions about equity and resource allocation in policing. Overall, this study demonstrates the value of spatially informed mixed-effects models in urban crime analysis and points toward important avenues for methodological and policy development.

The protective effects of income and education, identified in this study, align with extensive criminological literature. For example, neighborhoods with greater economic resources tend to experience fewer violent crimes due to lower levels of stress, greater institutional support, and enhanced social cohesion (Hipp, 2007). Similarly, higher educational attainment is associated with increased employment opportunities and reduced criminal behavior, particularly among youth (Lochner & Moretti, 2004). Policies that expand access to quality education and workforce development could contribute significantly to violence prevention, especially in historically marginalized areas.

From a policy perspective, investing in built environment improvements may also reduce crime. Research has shown that enhancing street lighting, maintaining vacant lots, and creating more public gathering spaces can increase informal surveillance and reduce criminal activity (Branas et al., 2011; Mazerolle et al., 2014). Additionally, promoting community-led safety initiatives and encouraging civic engagement may bolster collective efficacy, which has been repeatedly linked to reductions in violence (Sampson et al., 1997). Given the spatial disparities observed in this study, particularly in Buffalo's East Side and around the city's downtown district, a multifaceted strategy that combines place-based investment, equitable policing, and targeted surveillance deployment is warranted.

References

Works Cited

1. Angrist, J. D., & Pischke, J.-S. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.
2. Anselin, L., Cohen, J., Cook, D., Gorr, W., & Tita, G. (2000). Spatial analyses of crime. *Criminal Justice*, 4(2), 213-262. https://www.ojp.gov/criminal_justice2000/vol_4/o4e.pdf
3. Arel-Bundock, V. (2020). Simulate! Simulate! - Part 4: A binomial generalized linear mixed model. Retrieved from <https://aosmith.rbind.io/2020/08/20/simulate-binomial-glmm/>
4. Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
5. Benson, M.L. and Fox, G.L. (2004). Concentrated Disadvantage, Economic Distress, and Violence Against Women in Intimate Relationships. *Office of Justice Programs*. Retrieved from <https://www.ojp.gov/pdffiles1/nij/199709.pdf>.
6. Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127–135. <https://doi.org/10.1016/j.tree.2008.10.008>
7. Branas, C. C., South, E., Kondo, M. C., Hohl, B. C., Bourgois, P., Wiebe, D. J., & MacDonald, J. M. (2011). Citywide cluster randomized trial to restore blighted vacant land and its effects on violence, crime, and fear. *Proceedings of the National Academy of Sciences*, 115(12), 2946–2951. <https://doi.org/10.1073/pnas.1718503115>
8. Gómez, M., Mejía, D., & Ramos, J. (2021). The effects of surveillance cameras on crime: Evidence from Bogotá. *Journal of Public Economics*, 194. <https://doi.org/10.1002/pam.22280>
9. Hartig, F. (2024). DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.4.7, <https://github.com/florianhartig/dharma>.

10. Hipp, J. R. (2007). Income inequality, race, and place: Does the distribution of race and class within neighborhoods affect crime rates? *Criminology*, 45(3), 665–697. <https://doi.org/10.1111/j.1745-9125.2007.00088.x>
11. Hollander, J.B. & Cahill, B. (2011). Confronting Population Decline in the Buffalo, New York, Region: A Close Reading of the “Erie-Niagara Framework for Regional Growth”. *Journal of Architectural and Planning Research*, 28(3), 252-267. <https://www.jstor.org/stable/43030944>
12. Klinenberg, E. (2002). Heat Wave: A Social Autopsy of Disaster in Chicago. University of Chicago Press. <https://press.uchicago.edu/ucp/books/book/chicago/H/bo20809880.html>
13. Kontokosta, C. E. (2016). The Quantified Community and Neighborhood Labs: A Framework for Computational Urban Science and Civic Technology Innovation. *Journal of Urban Technology*, 23(4), 67-84. <https://doi.org/10.1080/10630732.2016.1177260>
14. Kraus, N. (2004). Local policymaking and concentrated poverty: the case of Buffalo, New York. *Cities*, 21(6), 481-490. <https://doi.org/10.1016/j.cities.2004.08.004>
15. Krivo, L. J. & Peterson, R. D. (1996). Extremely disadvantaged neighborhoods and urban crime. *Social Forces*, 75(2), 619-648. <https://doi.org/10.1093/sf/75.2.619>
16. Lochner, L. & Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review*, 94(1), 155–189. Retrieved from <https://www.jstor.org/stable/3592774>
17. Lochner, L. (2010). Education policy and crime. NBER Working Paper No. 15894. National Bureau of Economic Research. Retrieved from https://www.researchgate.net/publication/228310577_Education_Policy_and_Crime
18. Lum, C., Koper, C. S., & Willis, J. J. (2019). Understanding the limits of technology’s impact on police effectiveness. *Police Quarterly*, 22(2), 189–213. <https://doi.org/10.1177/1098611116667279>
19. Manson, S., Schroeder, J., Van Riper, D., et al. (2024). IPUMS National Historical Geographic Information System: Version 19.0 [dataset]. Minneapolis, MN: IPUMS. <http://doi.org/10.18128/Do50.V19.0>

20. Marotta, P. (2016). Assessing Spatial Relationships Between Rates of Crime and Rates of Gonorrhea and Chlamydia in Chicago, 2012. *Journal of Urban Health*, 94, 276-288. <https://doi.org/10.1007/s11524-016-0080-7>
21. Mazerolle, L., Sargeant, E., Cherney, A., et al. (2014). *Procedural Justice and Legitimacy in Policing*. Springer. <http://dx.doi.org/10.1007/978-3-319-04543-6>
22. Office of Justice Programs. (2019). CCTV Surveillance for Crime Prevention: A 40-Year Systematic Review with Meta-Analysis. Retrieved from <https://www.ojp.gov/library/publications/cctv-surveillance-crime-prevention-40-year-systematic-review-meta-analysis>.
23. Park, Y.M. & Kim, Y. (2014). A spatially filtered multilevel model to account for spatial dependency: application to self-rated health status in South Korea. *International Journal of Health Geographies*, 13(6). <https://doi.org/10.1186/1476-072X-13-6>
24. Partnership for the Public Good. (2018). A City Divided: A Brief History of Segregation in the City of Buffalo. Retrieved from https://ppgbuffalo.org/files/documents/data-demographics-history/a_city_divided__a_brief_history_of_segregation_in_the_city_of_buffalo.pdf
25. Pebesma E (2018). "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal*, 10(1), <https://doi.org/10.32614/RJ-2018-009>.
26. Piza, E. L. (2018). The crime prevention effect of CCTV in public places: A propensity score analysis. *Journal of Criminal Justice*, 57, 42–50. Retrieved from https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1188&context=jj_pubs.
27. Priks, M. (2015). The effect of surveillance cameras on crime: Evidence from the Stockholm subway. *Journal of Public Economics*, 115(588), F289-F305. <https://doi.org/10.1111/eoj.12327>
28. Reibel, M., & Bufalino, M. E. (2005). Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A*, 37(1), 127-139. <https://doi.org/10.1068/a36202>

29. Reich, B. J., Hodges, J. S., & Zadnik, V. (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*, 62(4), 1197–1206. <https://doi.org/10.1111/j.1541-0420.2006.00617.x>
30. Ryan, B. D. (2015). *Design after Decline: How America Rebuilds Shrinking Cities*. University of Pennsylvania Press. <https://www.pennpress.org/9780812223040/design-after-decline/>
31. Sampson, R. J. & Groves, W. B. (1989). Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology*, 94(4), 774-802. <https://doi.org/10.1086/229068>
32. Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277(5328), 918–924. <https://doi.org/10.1126/science.277.5328.918>
33. Sariaslan, A., Långström, N., D’Onofrio, B., et al. (2013). The impact of neighbourhood deprivation on adolescent violent criminality and substance misuse: A longitudinal, quasi-experimental study of the total Swedish population. *International Journal of Epidemiology*, 42(4), 1057-1066. <https://doi.org/10.1093/ije/dyt066>
34. Scorsone, E. and Bateson, N. (2011). *Long-Term Crisis and Systemic Failure: Taking the Fiscal Stress of American Cities Seriously (Case Study: City of Flint, Michigan)*. East Lansing, MI: Michigan State University. Retrieved from https://www.cityofflint.com/wp-content/uploads/Reports/MSUE_FlintStudy2011.pdf.
35. Sharkey, P. (2013). *Stuck in Place: Urban Neighborhoods and the End of Progress toward Racial Equality*. University of Chicago Press.
36. Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology*, 53(2), 133-157. <https://doi.org/10.1111/1745-9125.12070>
37. Welsh, B. C., & Farrington, D. P. (2009). Public area CCTV and crime prevention: An updated systematic review and meta-analysis. *Justice Quarterly*, 26(4), 716-745. <https://doi.org/10.1080/07418820802506206>

38. Wu, L., Li, N. (2024). Neighborhood effects and consequences of criminal justice contact: a research framework. *Computational Urban Science*, 4(27). <https://doi.org/10.1007/s43762-024-00138-w>
39. U.S. Census Bureau. (2020). American Community Survey 5-Year Estimates. Retrieved from <https://www.census.gov>.
40. Urban Institute. (2019). Public Surveillance Cameras and Crime. Retrieved from https://www.urban.org/sites/default/files/publication/101649/public_surveillance_cameras_and_crime.pdf.
41. Walker, K. (2023). “7.3 Small area time-series analysis,” in *Analyzing US Census Data: Methods, Maps, and Models in R*. CRC Press: Boca Raton, Florida. Retrieved from <https://walker-data.com/census-r/spatial-analysis-with-us-census-data.html#small-area-time-series-analysis>.
42. Walker, K. & Herman, M. (2025). tidycensus: Load US Census Boundary and Attribute Data as ‘tidyverse’ and ‘sf’-Ready Data Frames. R package version 1.7.1, <https://walker-data.com/tidycensus/>.
43. Weisburd, D. (2015). The Law of Crime Concentration and the Criminology of Place. *Criminology*, 53(2), 133-157. <https://doi.org/10.1111/1745-9125.12070>
44. Wirdze, L. (2024). Impact of Income Inequality on Crime Rates in Urban Areas. *European Journal of Sociology*, 7(1), 43-53. <https://doi.org/10.47672/ejs.2373>

Data Sources

- City of Buffalo Open Data. (2025). City boundary. Retrieved from <https://data.buffalony.gov/Government/City-Boundary/kkz7-hj9w>
- City of Buffalo Open Data. (2025). Crime incidents. Retrieved from https://data.buffalony.gov/Public-Safety/Crime-Incidents/d6g9-xbgu/about_data
- City of Buffalo Open Data. (2025). BPD cameras. Retrieved from <https://data.buffalony.gov/Public-Safety/BPD-Cameras/egev-8jf4>

- City of Buffalo Open Data. (2025). BPD stations. Retrieved from https://data.buffalony.gov/Public-Safety/BPD-Stations/yjub-v64n/about_data
- City of Buffalo Open Data. (2025). Parks. Retrieved from <https://data.buffalony.gov/Quality-of-Life/Parks/b6ah-eygb>
- Erie County GIS Office. (n.d.). Map gallery: County parks. Retrieved from <https://www3.erie.gov/gis/map-gallery>
- New York State GIS Clearinghouse. (2025). NY State parks property. Retrieved from <https://data.gis.ny.gov/maps/nysparks::ny-state-parks-property/about>
- U.S. Census Bureau. (2010-2023). American Community Survey 5-Year Estimates: Comparison Profiles 5-Year. <http://api.census.gov/data/2010/acs/acs5>