COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH
CENTER FOR INJURY SCIENCE AND PREVENTION

Geospatial & Applied Prevention Science Lab

The Promise and Pitfalls of Spatial Epidemiology in Firearm Violence Research

COLUMBIA CENTER FOR INJURY SCIENCE AND PREVENTION

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

1

Geospatial & Applied Prevention Science Lab

Presenters

Christopher N. Morrison, PhD MPH, Columbia University

Evan L. Eschliman, PhD MS, Columbia University

Siddhesh Zadey, MS MSc Columbia University

Christina A. Mehranbod, MPH, Columbia University

Brady Bushover, MPH, Columbia University

Leah E. Roberts, MPH, Columbia University

This work was also supported in part by grants R49CE003555 from the National Center for Injury Prevention and Control of the Centers for Disease Control and Prevention, K18HD117390 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, F31AA031193 from the National Institute on Alcohol Abuse and Alcoholism, and T32DA031099 from the National Institute on Drug Abuse.

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

2

Geospatial & Applied Prevention Science Lab

Spatial Epidemiology Is…

"…the application of theory and methods from **epidemiology**, **geography**, and **statistics** to describe spatial distributions of health outcomes and to analyze associations with possible causes to inform intervention and improve health."

(Morrison et al, 2024)

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

3

Geospatial & Applied Prevention Science Lab

Geographic Specificity

Outcome

Exposure

Time

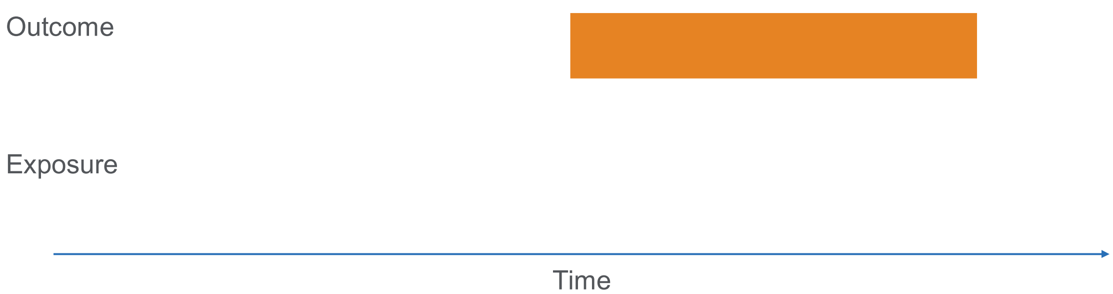
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

4

Geospatial & Applied Prevention Science Lab

Geographic Specificity



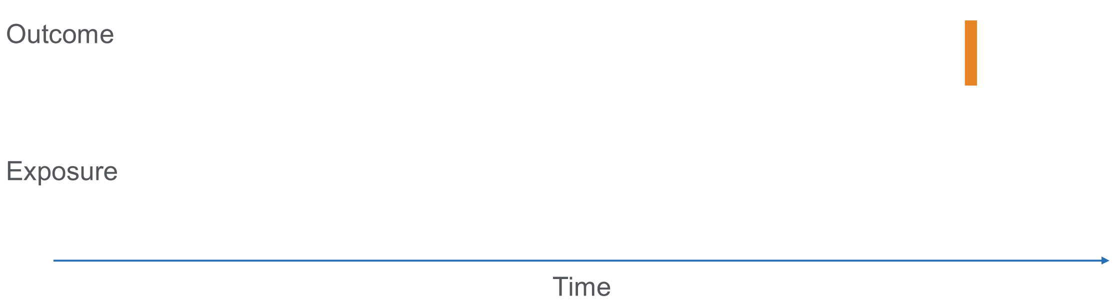
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

5

Geospatial & Applied Prevention Science Lab

Geographic Specificity



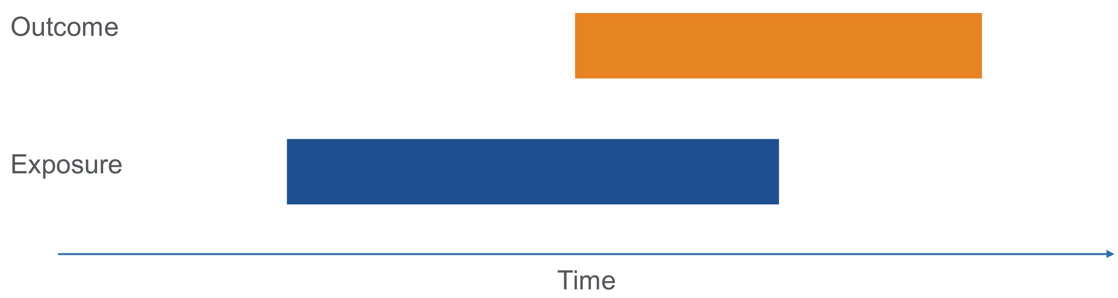
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

6

Geospatial & Applied Prevention Science Lab

Latency



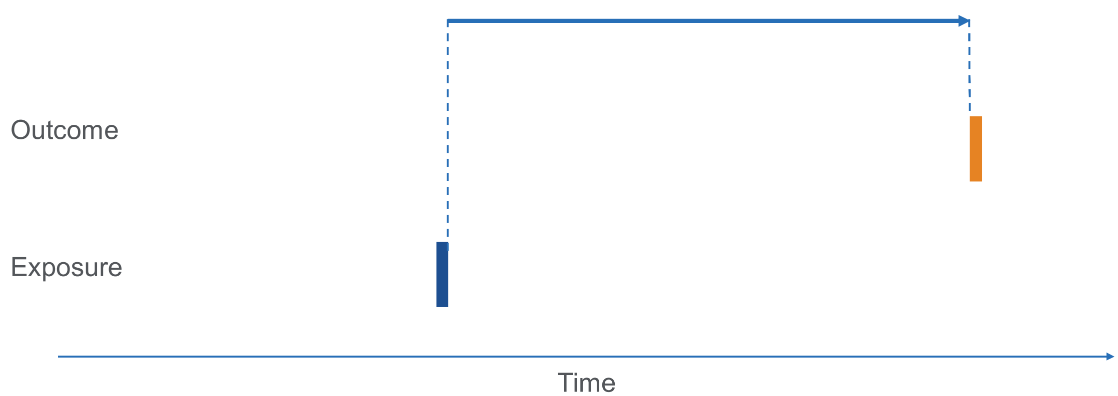
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

7

Geospatial & Applied Prevention Science Lab

Latency



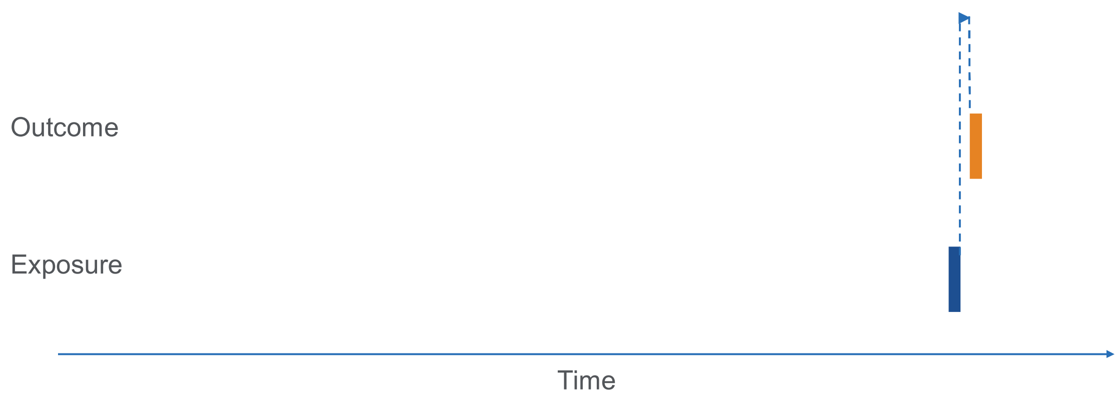
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

8

Geospatial & Applied Prevention Science Lab

Latency



COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

9

Geospatial & Applied Prevention Science Lab

Environmental Exposures

Physical

Social

Economic

Policy

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

10

Geospatial & Applied Prevention Science Lab

Today's Workshop

1.	**Spatial Theory**	Evan Eschliman, PhD MS
2.	**Spatial Data**	Siddhesh Zadey, MS MSc
3.	**Spatial Methods**	Christina Mehranbod, MPH
4.	**Spatial Analysis**	Brady Bushover, MPH
		Leah Roberts, MPH

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

11

Geospatial & Applied Prevention Science Lab

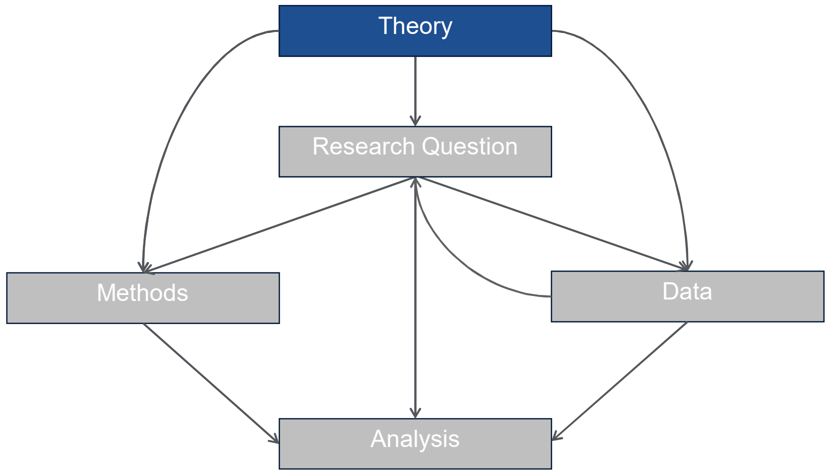
Spatial Theory

EVAN L. ESCHLIMAN, PHD, MS

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

12

Geospatial & Applied Prevention Science Lab



COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

13

13

Geospatial & Applied Prevention Science Lab

"As sources of data [maps] can '**flatten** and **simplify** complex stories.'"

Jacoby S. (2023) Home Owners' Loan Corporation Maps and Place-Based Injury Risks: A Complex History. *American Journal of Public Health*
Gioielli R. (2022) The tyranny of the map: rethinking redlining. The Metropole.

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

14

14

Geospatial & Applied Prevention Science Lab

Value-adds of theory

1. Help tie an analysis to a *larger story* or body of work
2. Fill in gaps in the *why* and *how*
3. Help *interpret* results and lead to *new research questions*
4. *Informs spatial scale for analysis / intervention

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

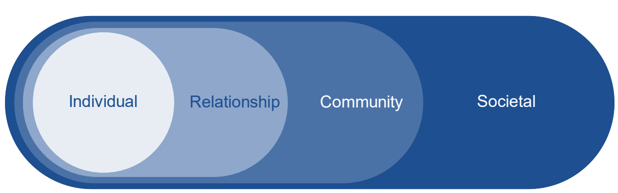
15

15

Geospatial & Applied Prevention Science Lab

Theory versus frameworks?

- Theory as the larger bucket
- Frameworks tend to be more interested in organization than explanation
- Example: social ecological models



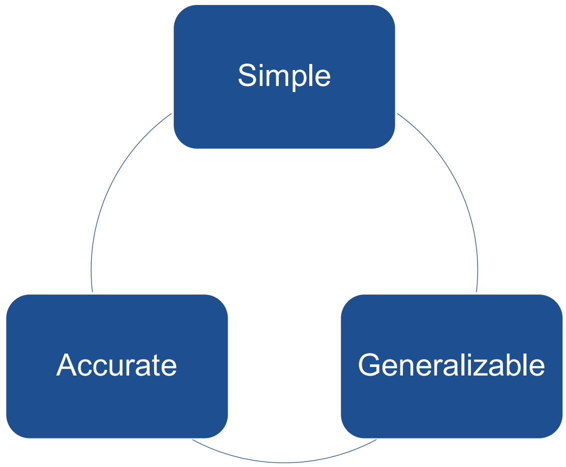
CDC (2024) https://www.cdc.gov/violence-prevention/about/index.html

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

16

16

Geospatial & Applied Prevention Science Lab



Latkin C. (2022) *Personal communication.*

Geospatial & Applied Prevention Science Lab

Some theories for motivating and interpreting spatial analyses related to firearm violence

Theoretical perspective	Scale of intervention
Busy streets	Built environment (e.g., sidewalks, green space)
Routine activities theory	Built environment and interpersonal factors (e.g., perceived safety)
Social disorganization	Built environment, interpersonal factors, sociopolitical factors
Collective efficacy	Sociopolitical factors
Fundamental cause theory	Social processes and conditions (e.g., racism, economic inequality, stigma)

Geospatial & Applied Prevention Science Lab

Spotlight on fundamental cause theory

- Focus on what puts people at risk of risks
- Fundamental causes:
  - Influence access to important resources
  - Influence multiple disease outcomes through multiple mechanisms
  - Maintain an association with disease even when intervening mechanisms change
- Focus on actions of more powerful/privileged groups
- Shared focus on *process* with justice frameworks (e.g., environmental justice)

Link & Phelan (1995); Clouston & Link (2021)

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

19

19

Geospatial & Applied Prevention Science Lab

The role of theory in indices

- Spatial analyses can often include an index as an exposure
- Indices are often “snapshots” of current conditions at a current moment
- Few if any indices are “neutral”; often have theoretical underpinnings and reasons for their existence

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

20

20

Geospatial & Applied Prevention Science Lab

Heat Vulnerability Index (HVI)

- Purpose: To show neighborhoods whose residents are more at risk for dying during and immediately following extreme heat
- Method: Uses a statistical model to summarize the **most important *social*** and ***environmental*** factors that contribute to neighborhood heat risk
  - Surface temperature
  - Green space
  - Access to home air conditioning
  - Percentage of residents who are low-income or non-Latine Black
- Differences in these risk factors across neighborhoods are rooted in past and present racism

City of New York. Interactive Heat Vulnerability Index. https://a816-dohbesp.nyc.gov/IndicatorPublic/data-features/hvi/

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

21

21

Geospatial & Applied Prevention Science Lab

Spatial Data

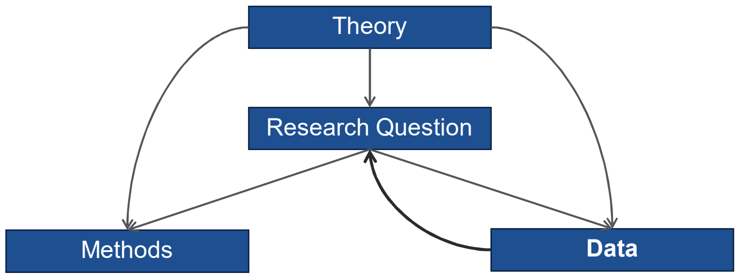
SID ZADEY, MS, MSC, PHD STUDENT

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

22

Geospatial & Applied Prevention Science Lab

From Theory To Data



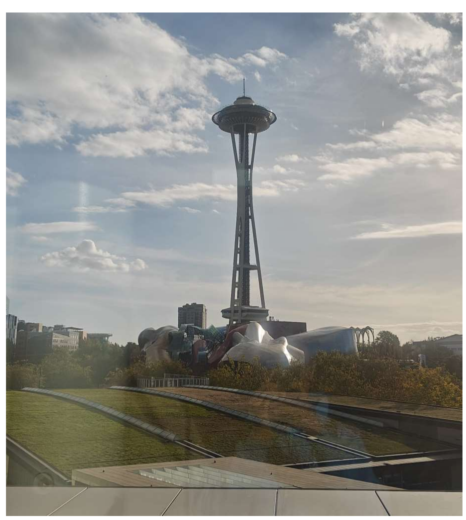
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

23

23

Geospatial & Applied Prevention Science Lab

Spatial Data



Place

Space

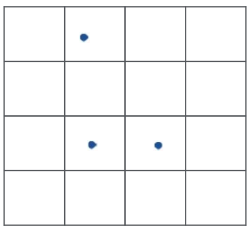
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

24

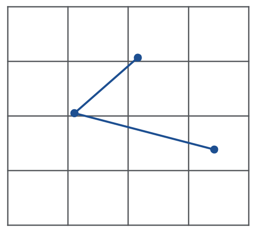
24

Geospatial & Applied Prevention Science Lab

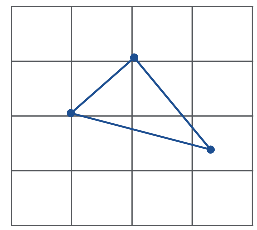
Three 'P's of Vector Data



Points



Polylines



Polygons

Geospatial & Applied Prevention Science Lab

Spatial Data Files

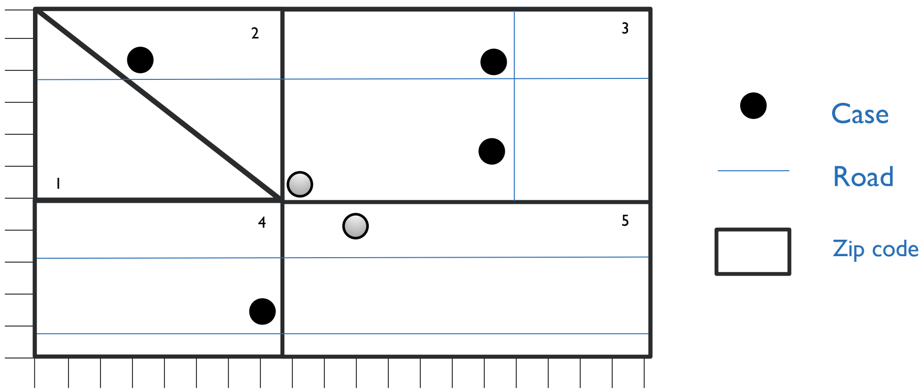
Shapefile

Index file

Attribute file

Geospatial & Applied Prevention Science Lab

Spatial Data Management



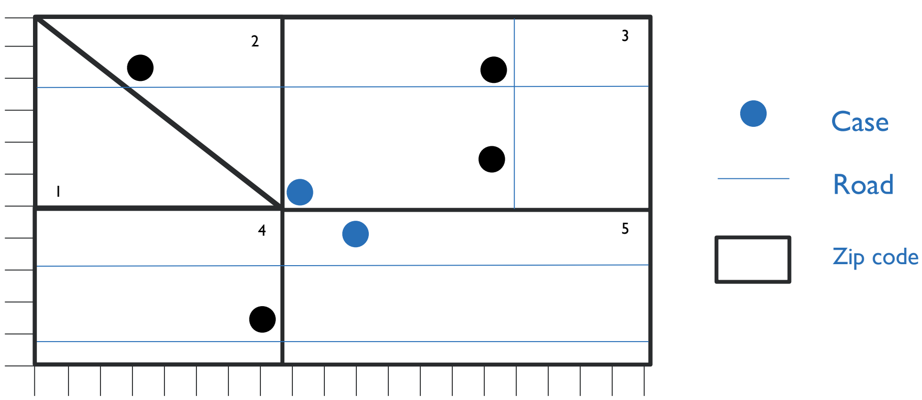
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

27

27

Geospatial & Applied Prevention Science Lab

Vector Data Operations: Selection by Attribute



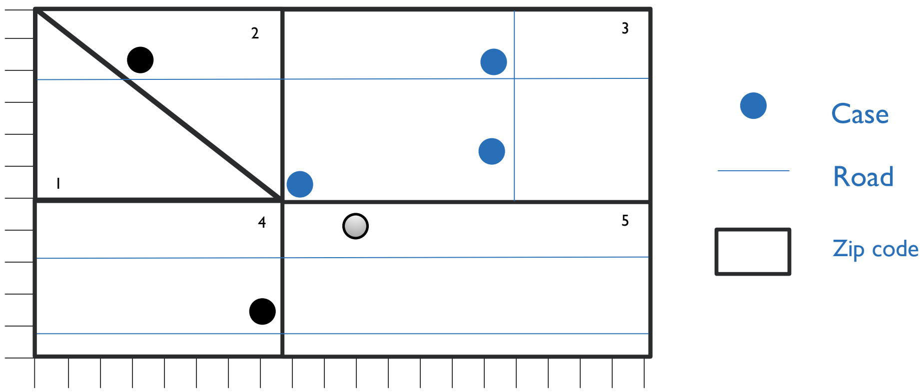
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

28

28

Geospatial & Applied Prevention Science Lab

Vector Data Operations: Selection by Location



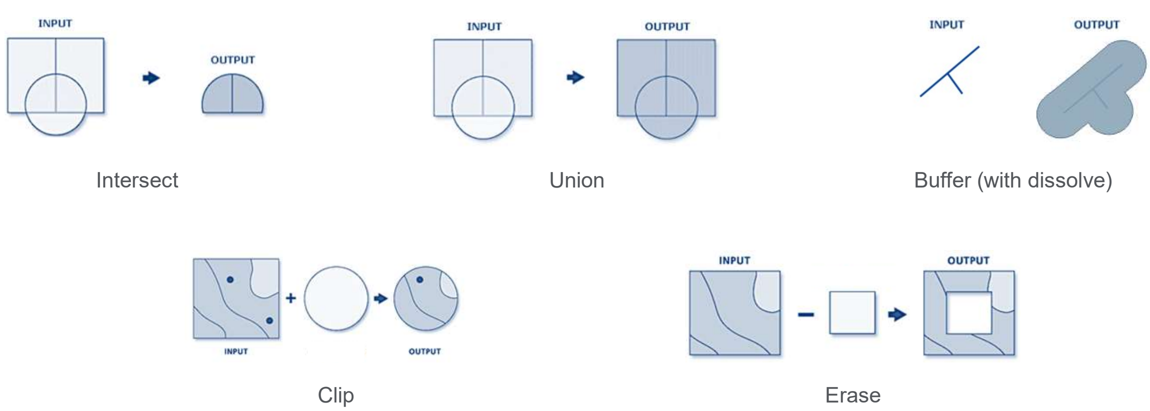
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

29

29

Geospatial & Applied Prevention Science Lab

Vector Data Operations: Vector Overlays



Source: ArcGIS

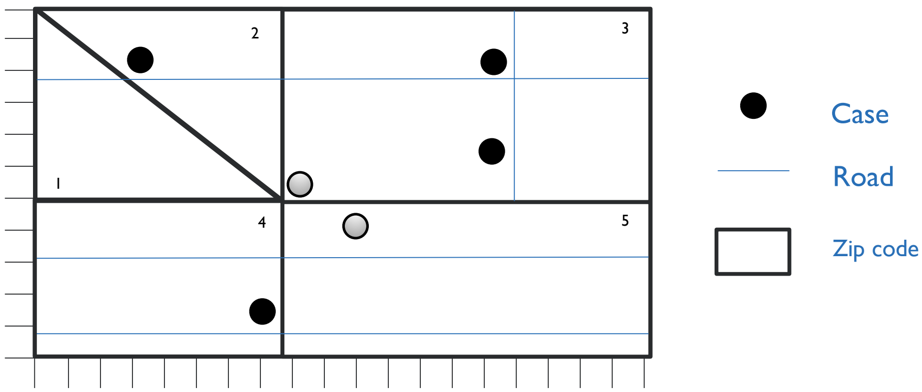
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

30

30

Geospatial & Applied Prevention Science Lab

Spatial Join



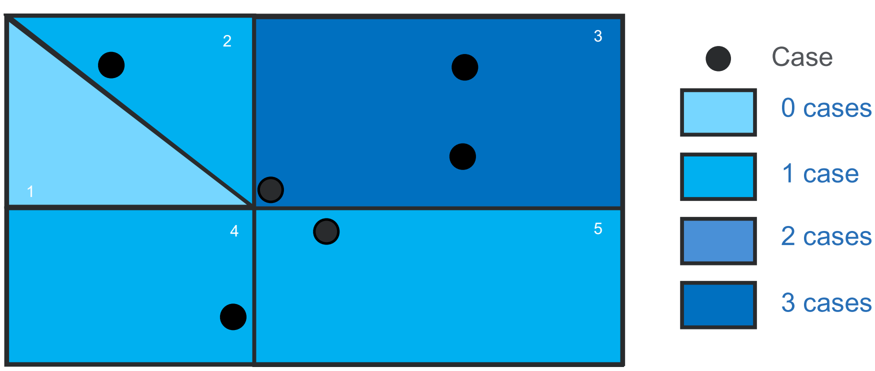
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

31

31

Geospatial & Applied Prevention Science Lab

Spatial Join



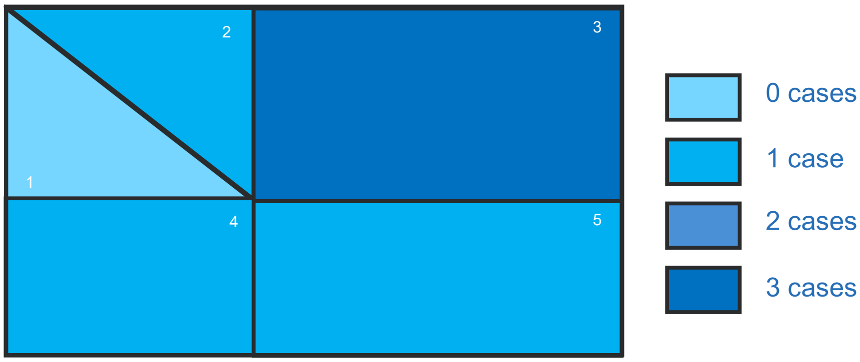
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

32

32

Geospatial & Applied Prevention Science Lab

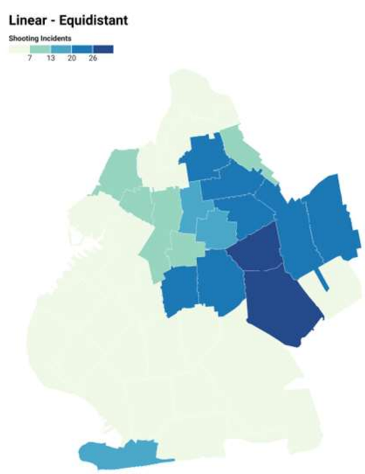
Choropleth - Polygons

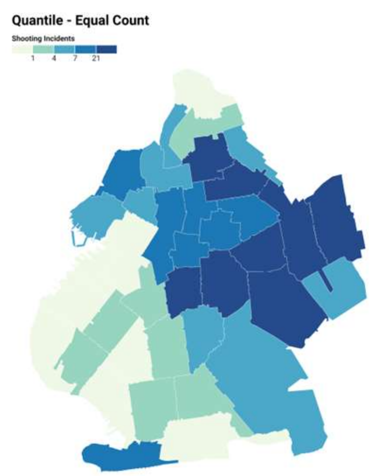


COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

Geospatial & Applied Prevention Science Lab

Spatial Data Visualization – Breaks

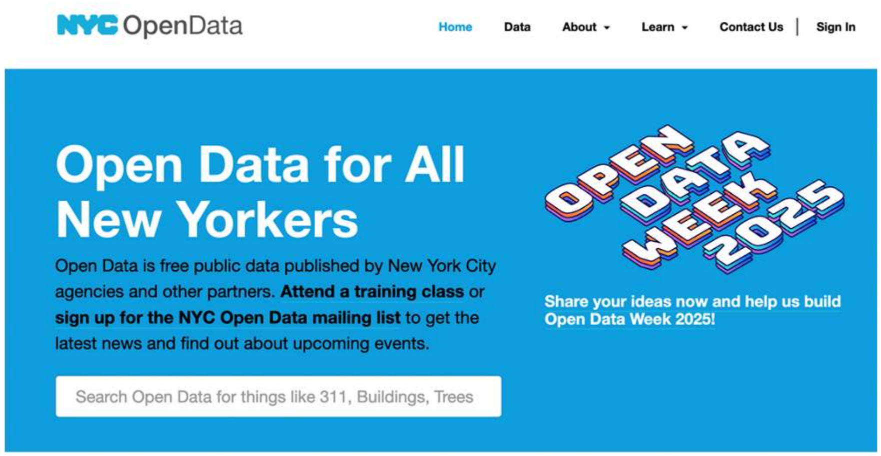




COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

Geospatial & Applied Prevention Science Lab

Acquiring Data



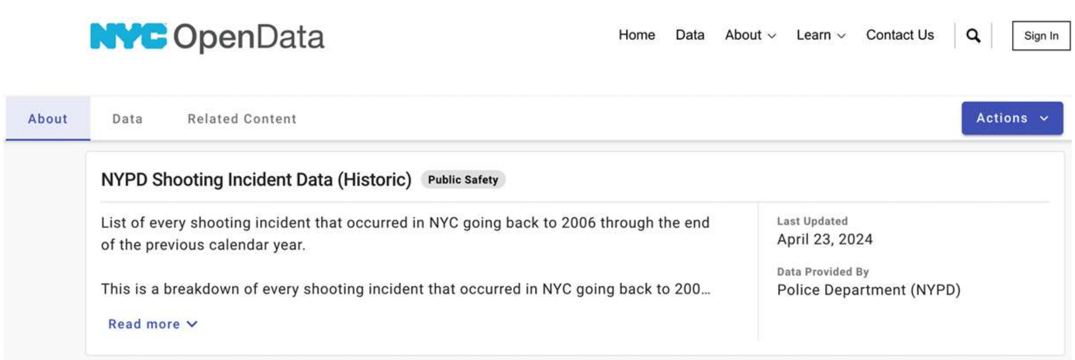
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

35

35

Geospatial & Applied Prevention Science Lab

Acquiring Data



COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

36

36

Geospatial & Applied Prevention Science Lab

Acquiring Data

What's in this Dataset?

Rows	Columns	Each row is a
28.6K	21	Shooting Incident

Columns (21)

Column Name	Description	API Field Name	Data Type
INCIDENT_KEY	Randomly generated persistent ID for each arrest	incident_key	Text
OCCUR_DATE	Exact date of the shooting incident	occur_date	Floating Timestamp
OCCUR_TIME	Exact time of the shooting incident	occur_time	Text
BORO	Borough where the shooting incident occurred	boro	Text
LOC_OF_OCCUR_DESC		loc_of_occur_desc	Text
PRECINCT	Precinct where the shooting incident occurred	precinct	Number
JURISDICTION_CODE	Jurisdiction where the shooting incident occurred. Jurisdiction codes 0(Patrol), 1(Transit) and 2(Housing) represent NYPD whilst codes 3 and more represent non… Read more	jurisdiction_code	Number

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

37

37

Geospatial & Applied Prevention Science Lab

Acquiring Data

NYC OpenData

Home Data About Learn Contact Us Sign In

Data Related Content Actions

Heat Vulnerability Index Rankings Health

The Heat Vulnerability Index (HVI) shows neighborhoods whose residents are more at risk for dying during and immediately following extreme heat. It uses a statistical model to summarize the most important social and environmental factors that contribute to neighborhood heat risk. The factors included in the HVI are surface temperature, green…

Read more

Last Updated
September 19, 2024

Data Provided By
Department of Health and Mental Hygiene (DOHMH)

What's in this Dataset?

Rows	Columns	Each row is a
184	2	ZIP Code Tabulation Area (ZCTA)

Columns (2)

Column Name	Description	API Field Name	Data Type
ZIP Code Tabulation Area (ZCTA) 2020	ZCTAs are a geographic product of the U.S. Census Bureau created to allow mapping, display, and geographic analyses of the United States Postal Service (USPS) ZIP Codes.	zcta20	Text
Heat Vulnerability Index (HVI)	Quintiles of heat vulnerability index, a measure of neighborhood risk of dying on extreme heat days.	hvi	Number

Rows per page 15 1-2 of 2

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

38

38

Geospatial & Applied Prevention Science Lab

Takeaways: Spatial Data Management

- Acquiring data is easy (in most cases)
- Managing different types of spatial data
- Conducting spatial data operations – based on theory and research questions
- Basic rules for visualizing data – based on the goal

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

39

39

Geospatial & Applied Prevention Science Lab

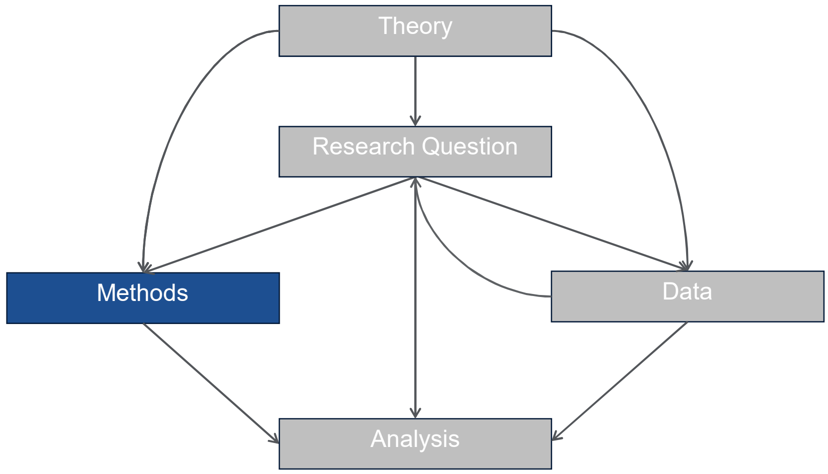
Spatial Methods

CHRISTINA MEHRANBOD, MPH, PHD(C)

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

40

Geospatial & Applied Prevention Science Lab



41

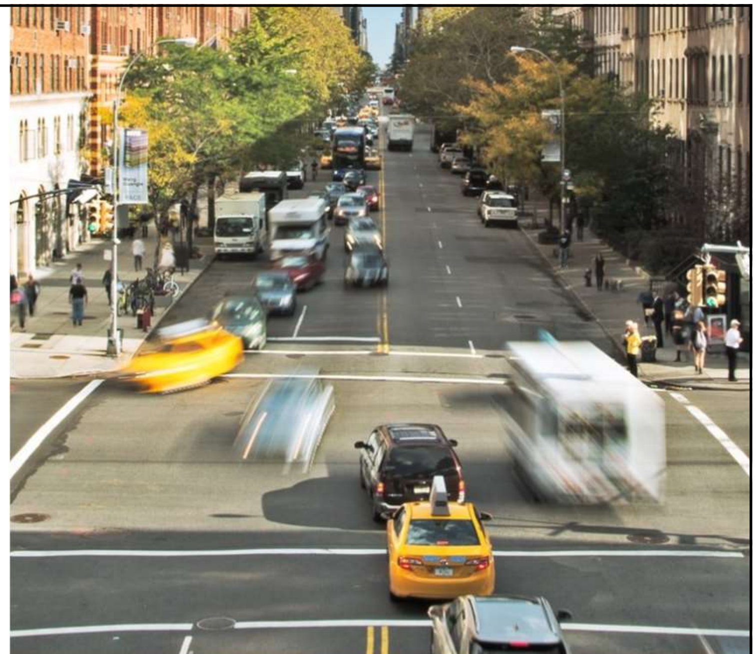
TOBLER'S FIRST LAW OF GEOGRAPHY

Everything is related to everything else, but near things are more related than distant things

42

Outline

1. What is spatial autocorrelation?
2. Case study: redlining and firearm violence
3. How to detect spatial autocorrelation
4. How to address it
    - Design
    - Spatial Regression Techniques



COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

43

43

Geospatial & Applied Prevention Science Lab

Aspatial or spatial methods

ASPATIAL

Units of analysis are independent

SPATIAL

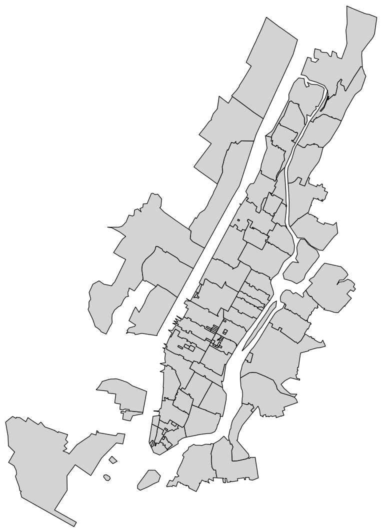
Units of analysis are not independent

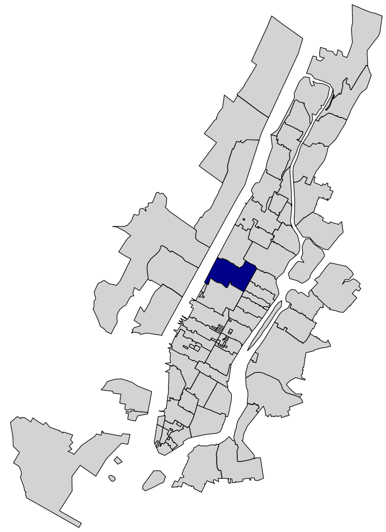
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

44

44

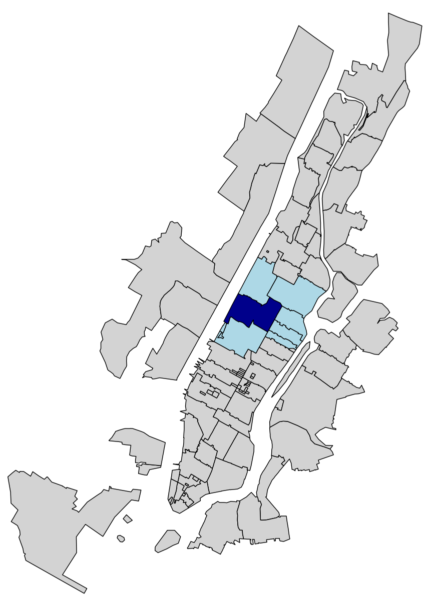
Geospatial & Applied Prevention Science Lab





45

Geospatial & Applied Prevention Science Lab

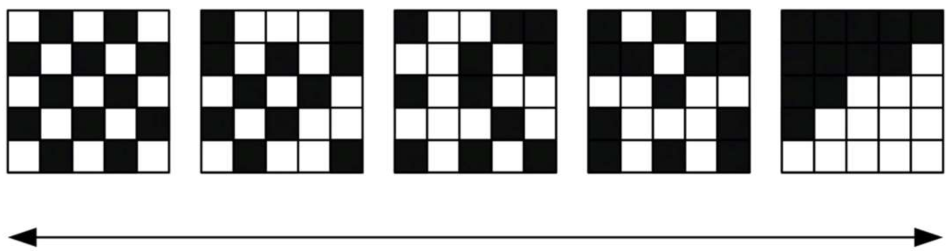


46

Geospatial & Applied Prevention Science Lab

Spatial Autocorrelation (Spatial Dependence)

Moran's *I*



Moran's $I < E(I)$ indicates tend to dispersion

Random Moran's $I = E(I)$

Moran's $I > E(I)$ indicates tend to clustering

Source: Grekousis G. (2020). *Spatial Analysis Methods and Practice: Describe – Explore – Explain through GIS*.

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

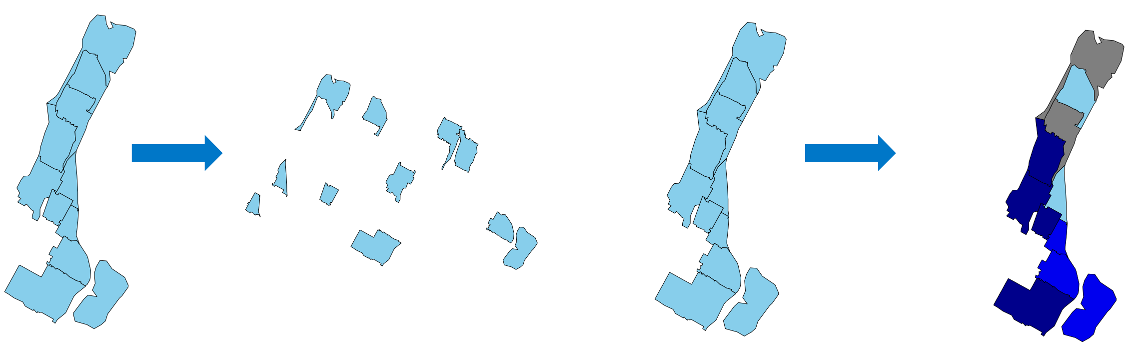
47

47

Geospatial & Applied Prevention Science Lab

Spatial dependency: a causal inference problem?

The stable unit treatment value assumption (SUTVA) – well-defined intervention



No interference between units

No treatment variation

Source: Kimmel K et al. (2021). *Trends in Ecology & Evolution*.

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

48

48

Geospatial & Applied Prevention Science Lab

Spatial dependency: a model misspecification problem

Major assumption of traditional linear models is that **residuals** are **INDEPENDENT**, have **constant variance** (homoscedastic), and have a mean of 0.

$$Y_i = \beta + e_i \text{ assumes } e_i \sim N(0, \sigma^2)$$

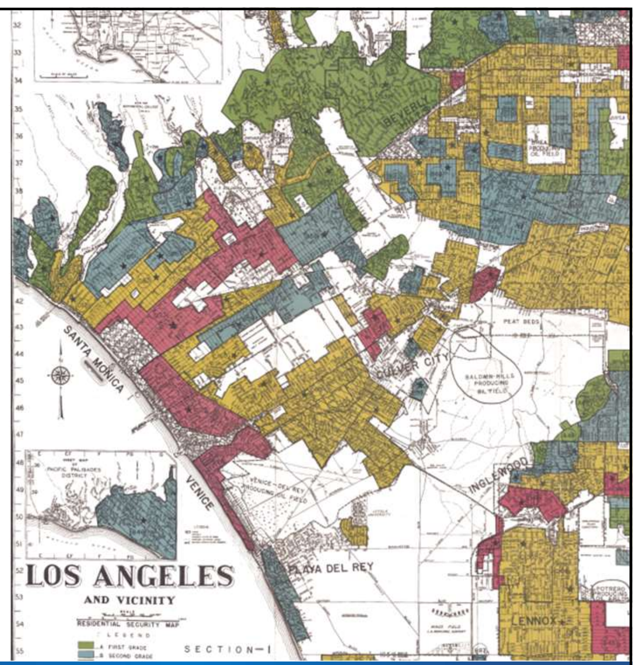
Spatial dependence violates **independence**, leading to biased estimates and small standard errors (narrow confidence intervals)

$$Y_i = \beta + \rho W Y_i + e_i$$

ρ: spatial autocorrelation estimate, WY_i: weighted average of neighboring values e_i: spatially structured residuals

Impact of redlining on present-day firearm violence

Research Aim. To examine the association between **historical redlining** and both **violent and firearm death** across the extent of the United States?



Source: Nelson, R. K., Winling, L, et al. (2023). Mapping Inequality: Redlining in New Deal America. Digital Scholarship Lab.

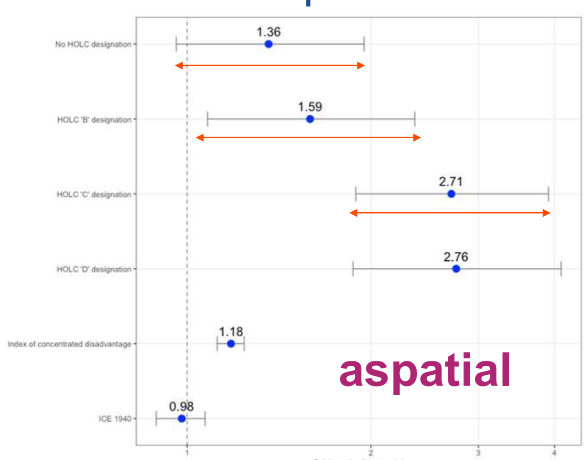
Geospatial & Applied Prevention Science Lab

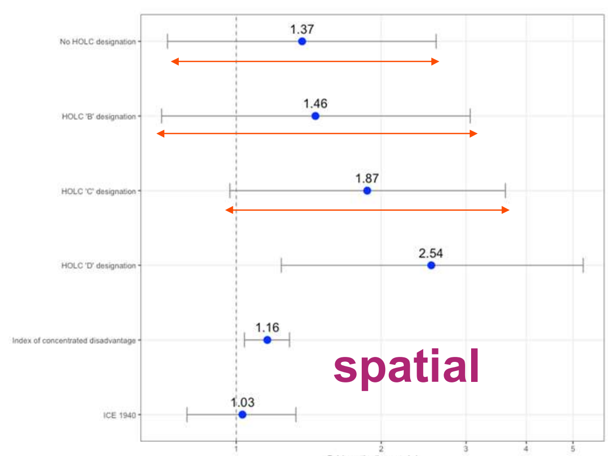
Methods

Study Design/Analysis	Data Sources	Setting	Exposure	Outcome
Multi-level panel analysis with negative binomial regression (ZIP codes nested in cities)	National Violent Death Reporting System NVDRS; MAPPING INEQUALITY REDLINING IN NEW DEAL AMERICA; IPUMS NHGIS		Redlining Grade (A, B, C, D)	Violent deaths and firearm deaths

Geospatial & Applied Prevention Science Lab

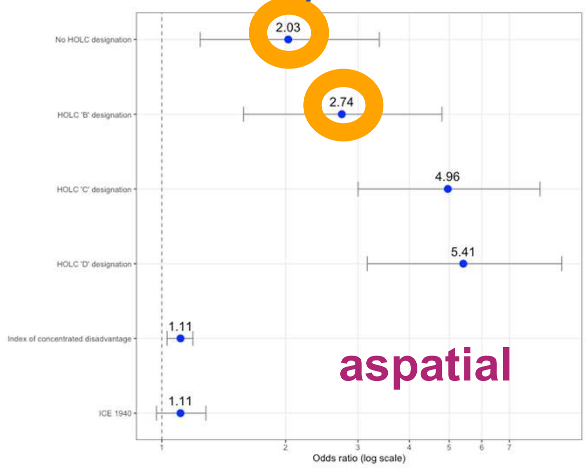
Results: Comparison

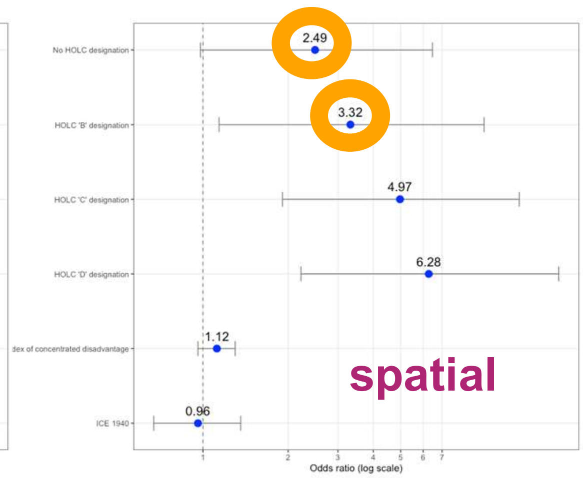




Geospatial & Applied Prevention Science Lab

Results: Comparison





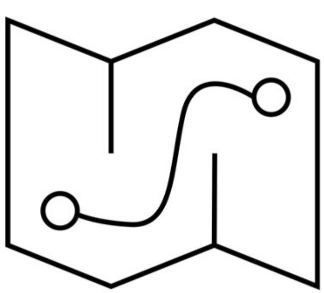
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

53

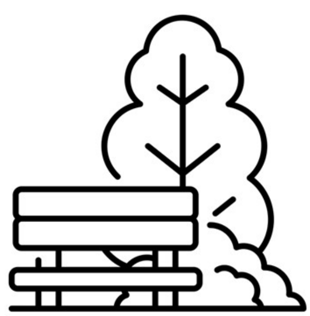
53

Geospatial & Applied Prevention Science Lab

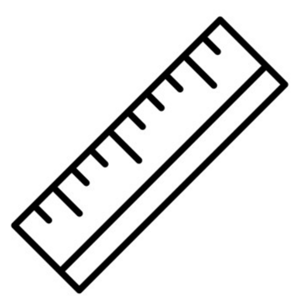
Sources of spatial autocorrelation



Spatial spillover



Spatial externalities



Measurement error

Source: Gebreab, S. Y. (2018). In A. V. Diez Roux, D. T. Duncan, & I. Kawachi (Eds.), *Neighborhoods and Health*.

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

54

54

Geospatial & Applied Prevention Science Lab

Measures of Global and Local Clustering

polygon-based measures > ecological studies *only*

A cluster is a "geographically and/or temporally bounded group of occurrences of sufficient size and concentration to be unlikely to have occurred by chance" (Knox, 1964)

Global Clustering

- Measures spatial autocorrelation across entire study area
- Common detection method: Moran's I or Geary's C
- Example: Do shootings cluster across NYC?

Local Clustering

- Identifies specific locations where autocorrelation is significant
- Common detection method: Local indicator of Spatial Association (LISA) or Scan Statistics
- Example: Identifying "hot spots" of violence in NYC

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

Geospatial & Applied Prevention Science Lab

Measures of Global and Local Clustering

conclusions (Anselin et al., 2000; Baller et al., 2001). To account for spatial dependency, we estimated global Moran's I analyses to determine the presence of clustering in health outcomes, indicating significant positive spatial autocorrelation. We identified nearest neighborhoods based on Queen contiguity, or all neighborhoods that surround a focal neighborhood. We constructed spatially lagged measures of the dependent variables in GeoDa using weighted average scores in neighboring census tracts. By including these scores in our models, we account for correlated error terms across census tracts.

Source: Semenza et al. (2021). *Journal of Behavioral Medicine.*

Methods

Violent crime data from the Miami-Dade Central Records Bureau were analyzed. The Local Indicators of Spatial Association statistics and a space-time permutation statistic were used to identify clusters of violent crimes and outliers, and Global Moran's I tool was used to assess spatial patterning in violent crime. Neighborhood disadvantage data were obtained from the American Community Survey 5-year estimates linked with arrest locations.

Source: Moise et al. (2021). *Journal of Experimental Criminology*

based on distance *vs.* contiguity). This approach has been used in prior injury research (Goldstick et al., 2015; Elliott Goldstick et al., 2015; Jay, 2020b); more broadly, splines are commonly used to address possible confounders (*e.g.*, from seasonality or long-term trends when estimating policy effects over time) (Bhaskaran et al., 2013). To test for residual spatial dependence, we calculated a Global Moran's I among residuals from each model.

Source: Jay et al. (2022). *Preventive Medicine.*

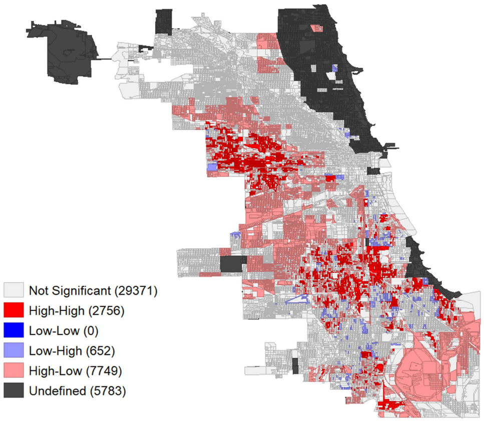
following year. Finally, we assessed model performance by combining observations of the independent variables for 1999 with the parameter estimates derived from 2000 to 2014 data (Model 3) to calculate predicted counts of firearm homicides for 1999, and calculate the global Moran's I to estimate the geographic structure of the prediction errors.

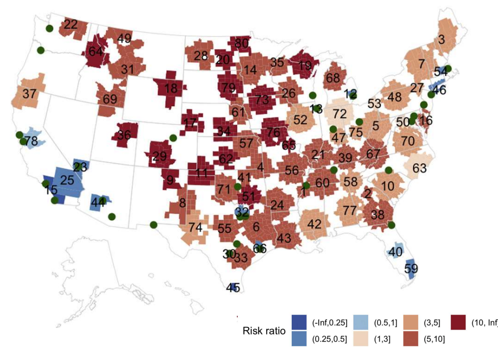
Source: Morrison et al. (2021). *Epidemiology.*

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

Geospatial & Applied Prevention Science Lab

Measures of Global and Local Clustering



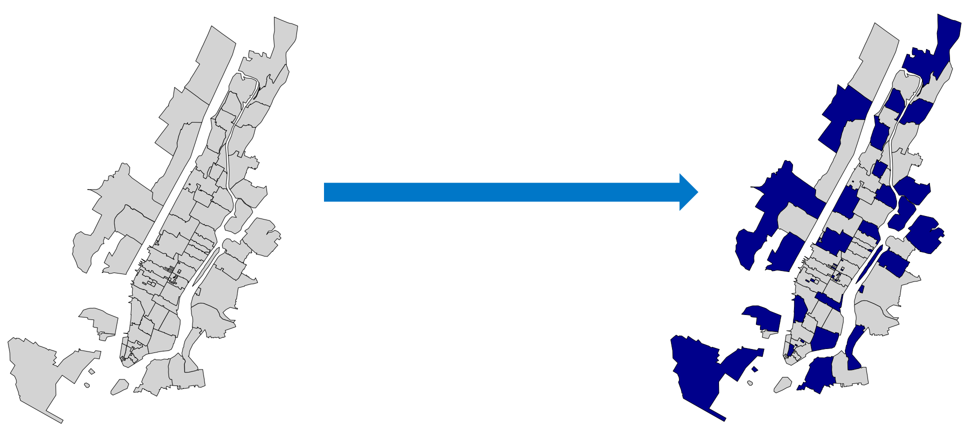


Source: Dirago, C. et al. (2024). *Journal of Surgical Research.*

Source: Platt, J. M. et al. (2022). *Journal of the American Academy of Child & Adolescent Psychiatry.*

Geospatial & Applied Prevention Science Lab

Addressing spatial autocorrelation: by **design**



Geospatial & Applied Prevention Science Lab

Addressing spatial autocorrelation: **regression**

- Geographically Weighted Regression
- Simultaneous Autoregressive Models
- Conditional Autoregressive Models

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

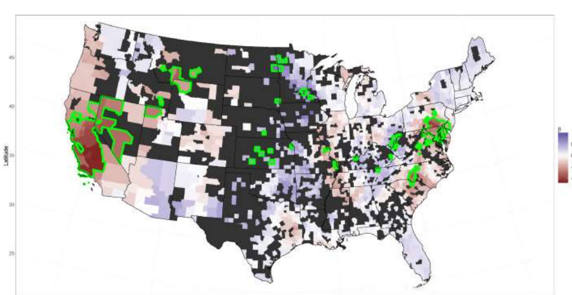
59

59

Geospatial & Applied Prevention Science Lab

Geographically Weighted Regression

- Captures spatial nonstationarity
- Uses data weighted by proximity
- Useful for exploratory analyses or hypothesis generation
- **When to use**: Theory suggests neighborhood effects are heterogeneous and vary spatially
- GWR in R: *gwr.basic( ) or gwr ( )* in *GWmodel package*



Source: Tran et al. (2020). *Spatial and Spatio-temporal Epidemiology.*

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

60

60

Geospatial & Applied Prevention Science Lab

Simultaneous Autoregressive Models

- Incorporates neighborhood effects using a spatial weights matrix
- Use theory to describe where/how spatial structure operates:
  - **lag model**: when the outcome itself is spatially influenced by neighboring outcomes
  - **error model**: when spatial dependence arises from omitted variables or contextual mismatch.
  - **lagged-mixed model**: when both the outcome and predictors are spatially structured.
- **When to use**: Theory suggests spatial dependence is substantive (lag) or nuisance (error) and spatial relationships extend across **larger** regions affecting all areas to various degrees.
- SAR in R: lagsarlm() in *spdep( ) or* errorsarlm() *spatialreg ( )* packages

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

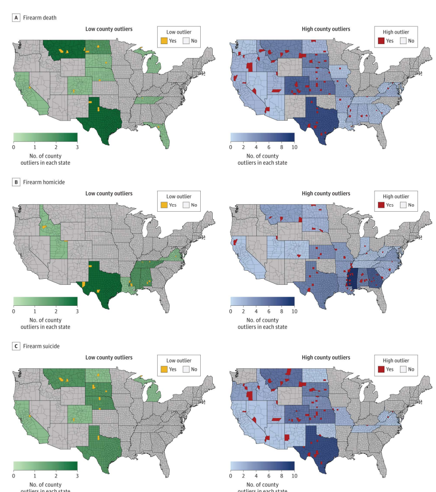
61

61

Geospatial & Applied Prevention Science Lab

Conditional Autoregressive Models

- Robust by leveraging localized information from neighboring data-rich areas to improve estimates in other regions
- Better for small-area estimation, sparse data and complex data.
- Directly conditions outcomes on immediate neighbors' values, emphasizing localized dependencies
- **When to use**: theory suggests strongest autocorrelation of among adjacent units.
- Bayesian CAR models in R: *Integrated Nested Laplace Approximation (INLA)* method



Source: Degli Esposti et al. (2022). *JAMA Network Open.*

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

62

62

Geospatial & Applied Prevention Science Lab

Methods: Hierarchical Bayesian conditional autoregressive Poisson models (R-INLA)

$$Y_{ij}|u_{ij} \sim Poisson(E_{ij}e^{u_{ij}})$$

$$u_{ij} = \beta_{00} + \beta_{01}j + \beta_n X'_{ij} + \psi_i + \theta_i$$

overall intercept

random intercept for city

matrix of independent measures X' for each ZIP code i, in city j

conditional autoregressive random (CAR) effect

spatially unstructured error term accounting for overdispersion

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

63

63

Geospatial & Applied Prevention Science Lab

Spatial Analysis

Practical Applications in R

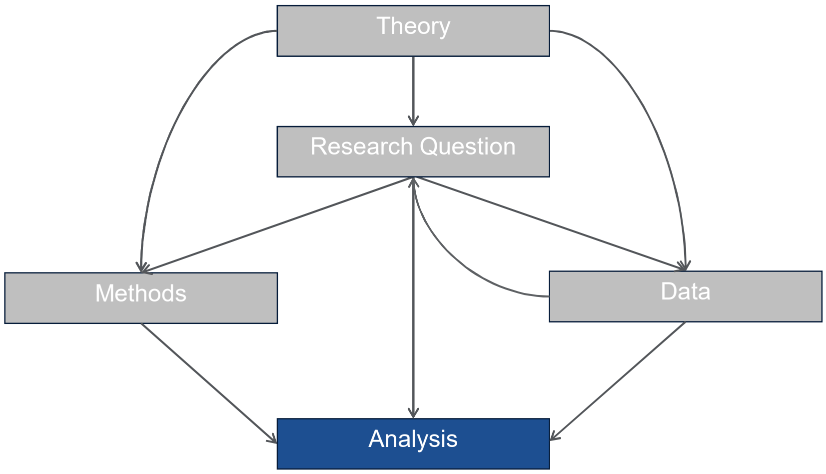
LEAH E. ROBERTS, MPH
BRADY BUSHOVER, MPH

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

64

Geospatial & Applied Prevention Science Lab



COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

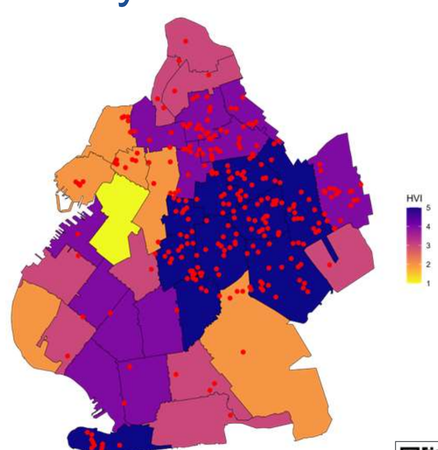
65

Geospatial & Applied Prevention Science Lab

Firearm Violence and Heat Vulnerability Index

- Demonstration investigating the association between HVI and shooting incidents

- Exposure: Heat vulnerability index (HVI)
- Outcome: Shooting incidents



COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

66

Geospatial & Applied Prevention Science Lab

Data Sources

NYC OpenData

United States Census Bureau

Heat vulnerability index: polygons

Shooting incidents: points

NYC boroughs: polygons

ZIP Code Tabulation Areas: polygons

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

67

R Packages

tidyverse

janitor

sf

INLA

lubridate

viridis

spdep

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

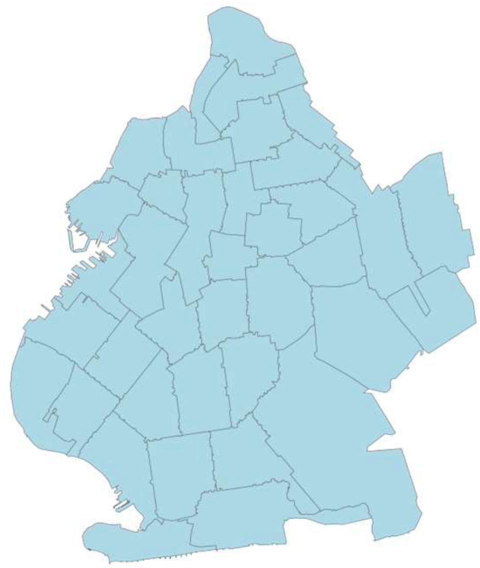
bit.ly/gapslab

68

Geospatial & Applied Prevention Science Lab

Setting

- Location: Brooklyn (ZIP Codes)
  n = 38
- Period: 2017
  Data availability



COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

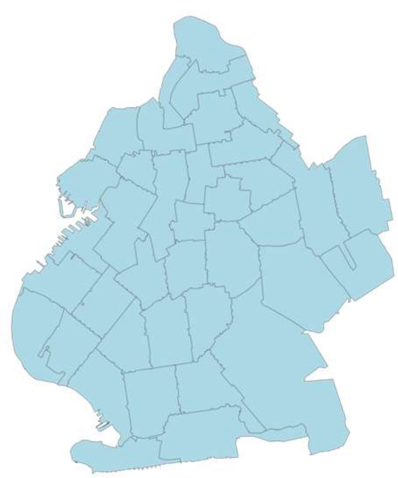
bit.ly/gapslab

69

Geospatial & Applied Prevention Science Lab

Joining Spatial Data

Shooting incidents: points



ZIP codes: polygons

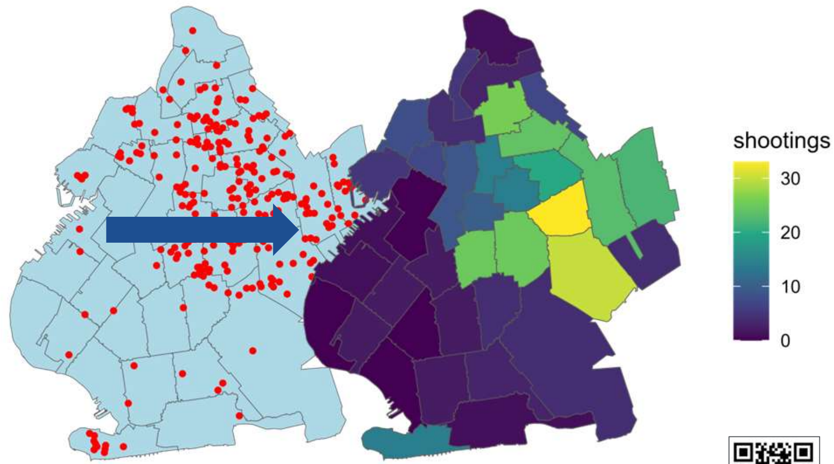
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

70

Geospatial & Applied Prevention Science Lab

Joining Spatial Data



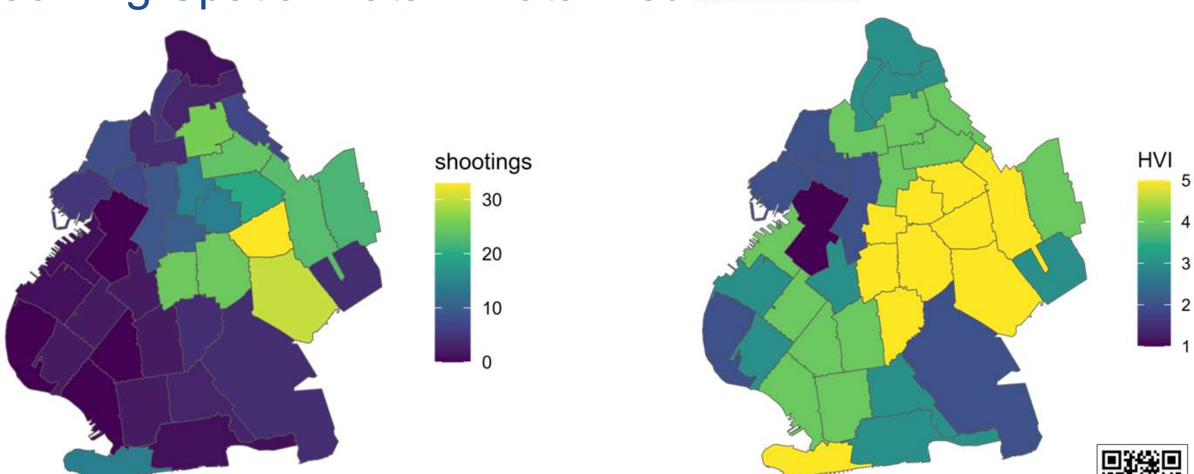
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

71

Geospatial & Applied Prevention Science Lab

Joining Spatial Data – Data Visualization



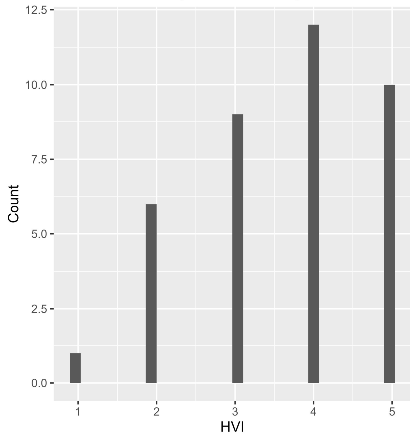
COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

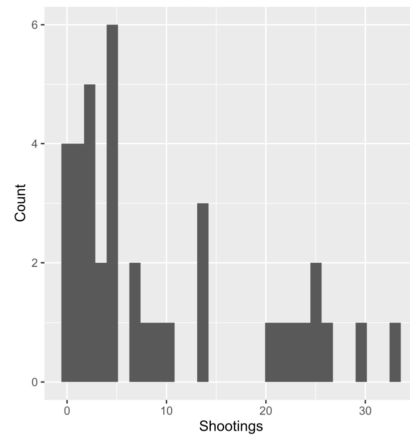
bit.ly/gapslab

72

Geospatial & Applied Prevention Science Lab

Data Visualization - Histograms





COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

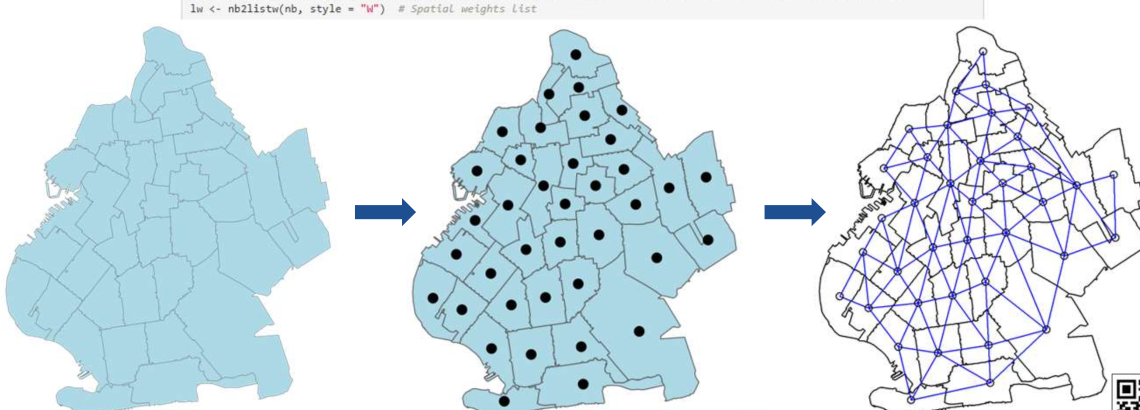
bit.ly/gapslab

73

Geospatial & Applied Prevention Science Lab

Clustering – Global Moran's I

```
# Create polygon neighbors list and their spatial weights matrix
nb <- poly2nb(spatial_data, row.names=spatial_data$id, queen=TRUE)  # Neighbors list based on ZIP code boundaries
lw <- nb2listw(nb, style = "W")  # Spatial weights list
```



COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

74

Geospatial & Applied Prevention Science Lab

Clustering – Global Moran's I

```
# Create polygon neighbors list and their spatial weights matrix
nb <- poly2nb(spatial_data, row.names=spatial_data$id, queen=TRUE)  # Neighbors list based on ZIP code boundaries
lw <- nb2listw(nb, style = "W")  # Spatial weights list

# Moran's I test for global spatial autocorrelation
moran_test <- moran.test(spatial_data$shootings, lw) # Conduct test
print(moran_test) # View results
```

```
##
##  Moran I test under randomisation
##
## data:  spatial_data$shootings
## weights: lw
##
## Moran I statistic standard deviate = 4.8218, p-value = 7.114e-07
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##        0.46545117       -0.02702703        0.01043179
```

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

75

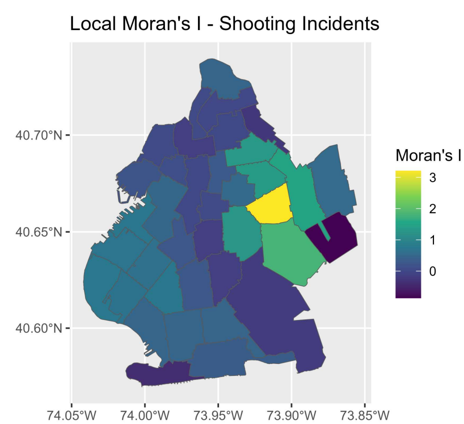
Geospatial & Applied Prevention Science Lab

Clustering – Local Moran's I

```
# With a value greater than 0, we see that similar values are clustered together in space
# There are most likley clusters in our data

# Now, we need to determine where the clusters are - use the local Moran's I
# Local Moran's I for clustering
local_moran <- localmoran(spatial_data$shootings, lw) # Conduct test
spatial_data$local_moran <- local_moran[,1]  # Add the results to our spatial data

# Once we have the local Moran's I values we can plot them and visualize our clusters
# Plot the local Moran's I clusters
ggplot(data = spatial_data) +
  geom_sf(aes(fill = local_moran)) +
  scale_fill_viridis_c() +
  labs(title = "Local Moran's I - Shooting Incidents", fill = "Moran's I")
```



COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

76

Geospatial & Applied Prevention Science Lab

Non-Spatial Model

```
# Create Poisson model
model <- glm(shootings ~ hvi,
             family="poisson",
             data=spatial_data)

# View results
summary(model)
```

```
##
## Call:
## glm(formula = shootings ~ hvi, family = "poisson", data = spatial_data)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.4774     0.2706  -1.764   0.0777 .
## hvi           0.6796     0.0616  11.033   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 377.93  on 37  degrees of freedom
## Residual deviance: 227.74  on 36  degrees of freedom
## AIC: 357.64
##
## Number of Fisher Scoring iterations: 5
```

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

77

Geospatial & Applied Prevention Science Lab

Non-Spatial Model

```
# Get the IRR and confidence interval
# Calculate IRR & CI
hvi_irr_nonsp <- round(exp(coef(model)["hvi"]),2) # IRR
hvi_conf_int <- round(exp(confint(model)["hvi", ]),2)  # Lower bound
```

```
# Create dataframe to store calculations
irr_df <- data.frame(
  Variable = "HVI",
  IRR = hvi_irr_nonsp,
  CI_Lower = hvi_conf_int[1],
  CI_Upper = hvi_conf_int[2]
)

# View results
rownames(irr_df) <- NULL
print(irr_df)
```

```
##   Variable  IRR CI_Lower CI_Upper
## 1      HVI 1.97     1.75     2.23
```

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

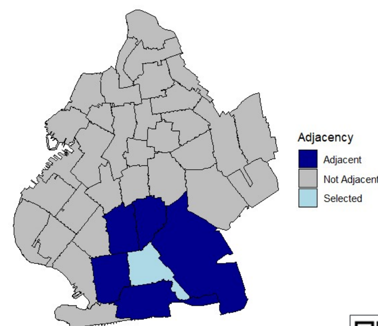
78

Geospatial & Applied Prevention Science Lab

Bayesian Inference with INLA

```
# Set up the spatial adjacency matrix
adjacency <- nb2mat(nb, # Use the previously created polygon neighbors list
                    style = "B", # Basic binary coding to create matrix (1 = neighbor; 0 = non-neighbor)
                    zero.policy = TRUE) # Permit the weights list to be formed with zero-length weights vectors
```

ID	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	1	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	1	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	1	0



COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

79

Geospatial & Applied Prevention Science Lab

Bayesian Inference with INLA

```
# Model: Shootings as outcome, HVI as predictor, controlling for spatial dependencies
# Specify formula
formula <- shootings ~ hvi + f(id, # Use the consecutive IDs created earlier
                               model = "besag", # Use the Besag model for spatial effects
                               graph = adjacency) # Use the adjacency matrix we just created

# Run the INLA model
result <- inla(formula, # Use the formula we specified above
               data = as.data.frame(spatial_data), # Use data from the `spatial_data` dataframe
               family = "poisson") # Since we have count data, use a Poisson statistical model

# Print the results
inla_result <- summary(result)
inla_result
```

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

80

Geospatial & Applied Prevention Science Lab

Bayesian Inference with INLA

```
## Time used:
##     Pre = 1.06, Running = 0.917, Post = 0.109, Total = 2.09
## Fixed effects:
##               mean    sd 0.025quant 0.5quant 0.975quant   mode kld
## (Intercept) -0.105 0.575     -1.245   -0.102      1.021 -0.102   0
## hvi          0.474 0.152      0.172    0.475      0.772  0.475   0
##
## Random effects:
##   Name     Model
##     id Besags ICAR model
##
## Model hyperparameters:
##                    mean    sd 0.025quant 0.5quant 0.975quant  mode
## Precision for id 0.556 0.199      0.263    0.524       1.03 0.467
##
## Marginal log-Likelihood:  -153.17
##  is computed
## Posterior summaries for the linear predictor and the fitted values are computed
## (Posterior marginals needs also 'control.compute=list(return.marginals.predictor=TRUE)')
```

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

81

Geospatial & Applied Prevention Science Lab

Bayesian Inference with INLA

```
# Get the IRR and credible interval
hvi_irr <- round(exp(inla_result$fixed[2, "mean"]),2) # IRR
hvi_quantiles <- inla_result$fixed[2, c("0.025quant", "0.975quant")]
hvi_ci_lower <- round(exp(hvi_quantiles[1]),2)  # Lower bound
hvi_ci_upper <- round(exp(hvi_quantiles[2]),2)  # Upper bound

# Print the IRR and credible interval
# Create a dataframe to hold our calculated values from above
irr_df_inla <- data.frame(
  Variable = "HVI",
  IRR = hvi_irr,
  CrI_Lower = hvi_ci_lower,
  CrI_Upper = hvi_ci_upper
)

# Print results
rownames(irr_df_inla) <- NULL
print(irr_df_inla)
```

```
##   Variable  IRR CrI_Lower CrI_Upper
## 1      HVI 1.61      1.19      2.16
```

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

82

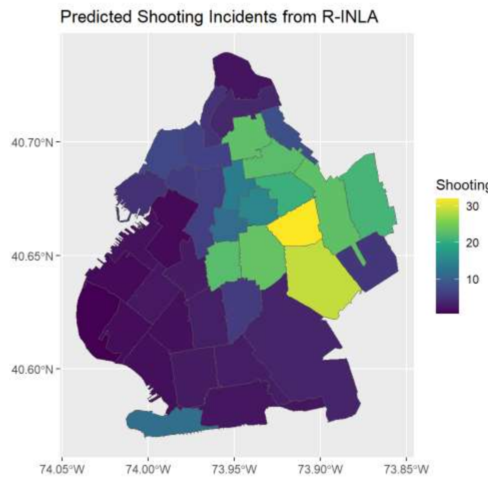
Geospatial & Applied Prevention Science Lab

Visualize the Results

```r
# Create plot of INLA results
# Create a new column in a spatial data that takes on the predicted number of incidents from our INLA model
spatial_data$pred_shootings <- result$summary.fitted.values[, "mean"]

# Plot the results
inla_plot <- ggplot(spatial_data) +
  geom_sf(aes(fill = pred_shootings)) +
  scale_fill_viridis_c() +
  labs(title = "Predicted Shooting Incidents from R-INLA", fill = "Shootings")

# View plot
inla_plot
```



COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

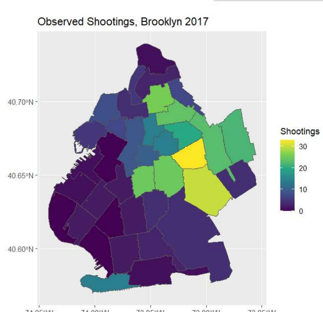
83

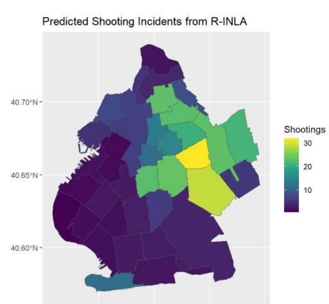
Geospatial & Applied Prevention Science Lab

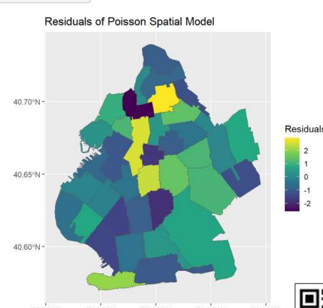
Plot the Residuals

```r
# Calculate residuals
spatial_data$residuals <- spatial_data$shootings - spatial_data$pred_shootings # Subtract predicted values from
observed

# Plot residuals
ggplot(spatial_data) +
  geom_sf(aes(fill = residuals)) +
  scale_fill_viridis_c() +
  labs(title = "Residuals of Poisson Spatial Model", fill = "Residuals")
```







COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

84

Geospatial & Applied Prevention Science Lab

Test for Spatial Autocorrelation

```
# Test for spatial autocorrelation
residual_moran <- moran.test(spatial_data$residuals, lw)
print(residual_moran)
```

```
##
##  Moran I test under randomisation
##
## data:  spatial_data$residuals
## weights: lw
##
## Moran I statistic standard deviate = -1.905, p-value = 0.9716
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic       Expectation          Variance
##       -0.22133717       -0.02702703        0.01040382
```

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

85

Geospatial & Applied Prevention Science Lab

Compare with Non-Spatial Model

Non-Spatial Model

```
##   Variable  IRR CI_Lower CI_Upper
## 1      HVI 1.97     1.75     2.23
```

Spatial Model

```
##   Variable  IRR CrI_Lower CrI_Upper
## 1      HVI 1.61      1.19      2.16
```

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

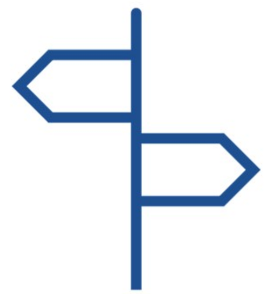
bit.ly/gapslab

86

Geospatial & Applied Prevention Science Lab

Additional Resources

- Geocomputation with R (free online book)
- R-INLA (documentation and resources for the R-INLA package)
- Bayesian inference with INLA (free online book)



COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

87

Geospatial & Applied Prevention Science Lab

Today's Workshop

1.	**Spatial Theory**	Evan Eschliman, PhD MS
2.	**Spatial Data**	Siddhesh Zadey, MS MSc
3.	**Spatial Methods**	Christina Mehranbod, MPH
4.	**Spatial Analysis**	Brady Bushover, MPH
		Leah Roberts, MPH

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

bit.ly/gapslab

88

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH
CENTER FOR INJURY SCIENCE AND PREVENTION

bit.ly/gapslab

Geospatial & Applied Prevention Science Lab

www.thegapslab.org

COLUMBIA | MAILMAN SCHOOL OF PUBLIC HEALTH

89