GIS For Crime Analysis

Examples From Buffalo City, New York

GE5226 Group Presentation

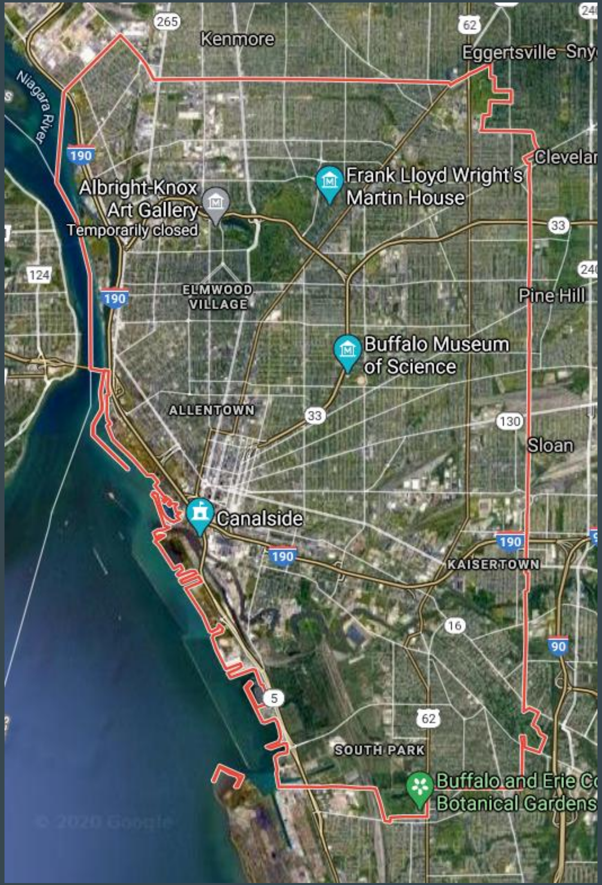
Xu Yuting
Dini Aprilla Norvyani
Lim Zhu An

Agenda

- Study Area Overview
- Data
- Types of GIS Analysis and Results
- Discussion
- Questions

City of Buffalo, New York

- Second largest city in the State of New York
- Population: ~ 250k
- High poverty rate (~29% in 2011), high unemployment rate (5.9% in 2015)
- Known for high violent crime rate



Dataset

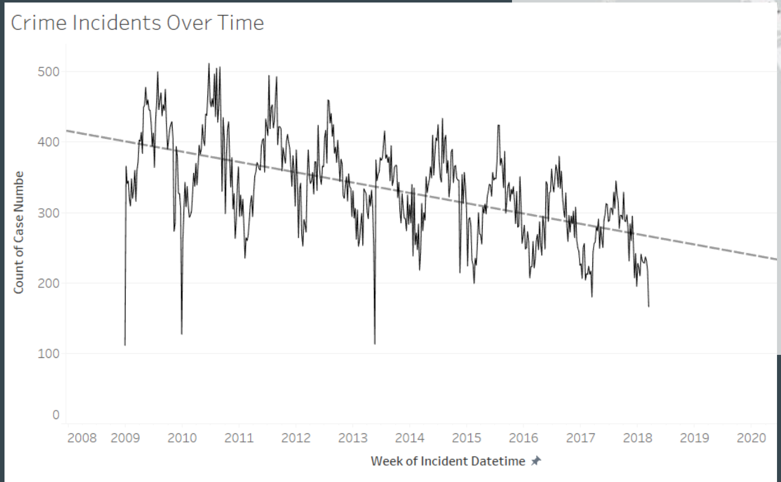
Crime Incident Dataset

Source: Department of Police, City of Buffalo, New York

Date Range: 2009 - 2018

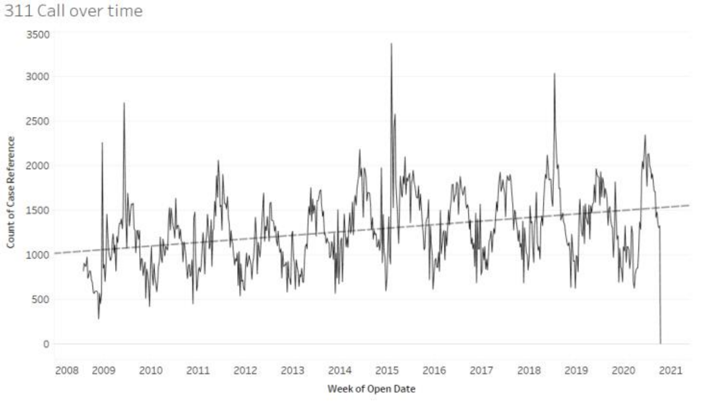
Other attributes:

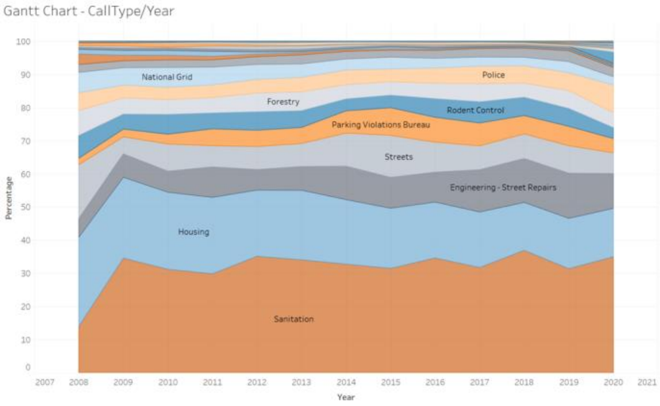
- Crime type
- Location



311 Call Data

- 311: Municipal service hotline
- Date range: 2008 - 2020
- Residents/businesses/visitors make calls to City government request for services or information/provide feedback on municipal issues
- Each call log comes with:
  - Service number/Unique ID
  - Issue description/category
  - Case open/close time
  - Location of issue





Neighbourhood Metrics

- Socioeconomic information of each neighbourhood, including:
    - Household structure
    - Education level
    - Car ownership
    - House type
    - Racial group breakdown

Census data

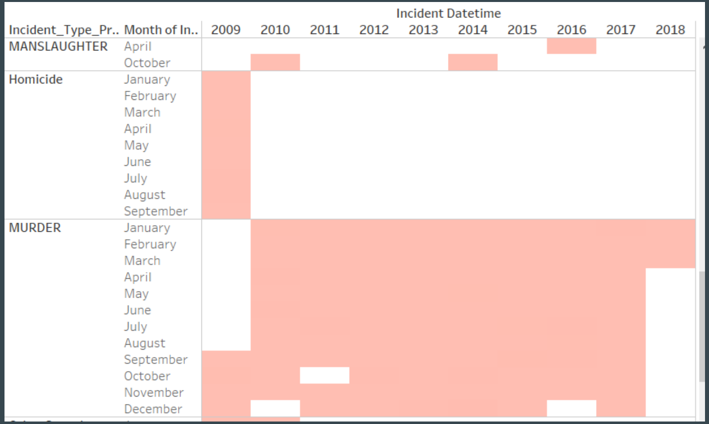
Demographic information

- Spatial unit: census block
- Population breakdown by racial groups

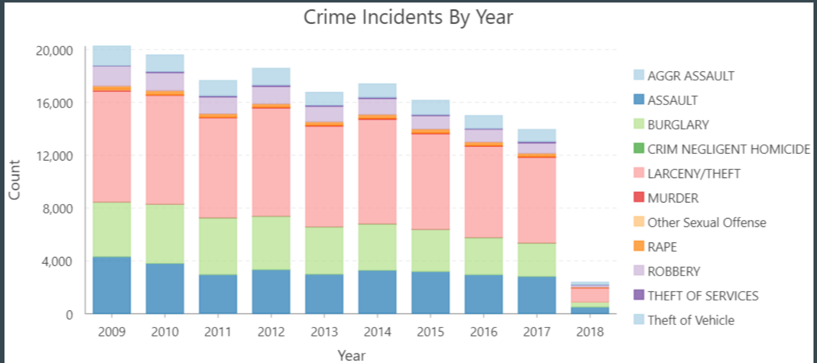
Data Preprocessing

Crime Incident Data & 311 Call Log Data

- Both are point dataset
- Digitisation + geocoding
- Clean up crime type / call reason and category
- Classify crime incident into **Part I Crimes** and **Others**
- Spatial Clip to study area
- Part I Crimes include the violent crimes of:
  - Homicide
  - Rape
  - Robbery
  - Aggravated Assault
  - and the property crimes of: Burglary, Larceny-Theft, Motor Vehicle Theft



"Manslaughter" and "Homicide" not used after 2010 → renamed all to "Homicide" for standardization



Neighbourhood and Census Block Data

- Additional indices calculated
  - Crime count within area
  - Crime rate within area (count/area)
  - Call count within area
  - Racial Diversity Index: Probability of getting two person from the same racial group

Overview of GIS Application in Crime Pattern Analysis

Mapping & Visualisation

Analysis

Spatial Analysis
- Density
- Clustering
- 80-20 Principles

Space-Time Pattern Mining
- Space-Time Hotspots
- Repeat/Near-Repeat

Correlation Modelling
- Environmental Factors
- Socioeconomic Factors

Investigative Analysis
- Incident Path/Incident Sequence

...

Technology

Police Force Deployment

Track Suspects

...

Clustering

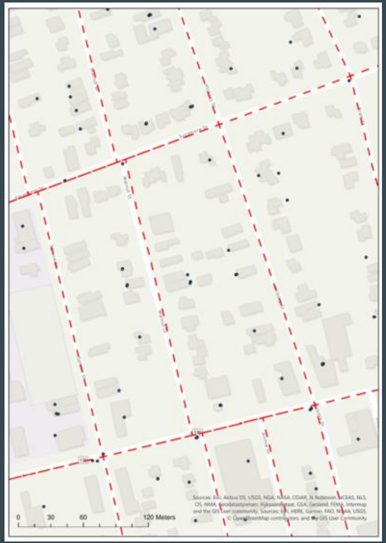
Uncovering the scale of spatial processes shaping crime locations

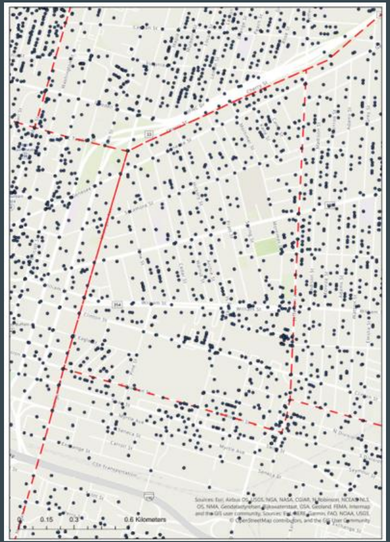
Understanding clustering of crimes at different spatial scale

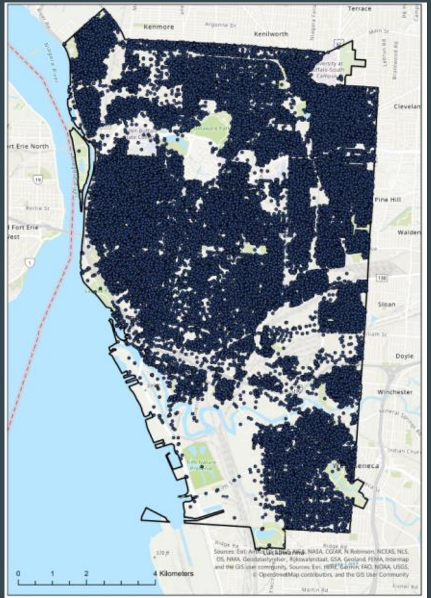
Features might display a <u>Clustered</u> or <u>Dispersed</u> pattern based on the spatial resolution that we look at

Different types of spatial scales:

E.g Census Block > Street > Neighbourhood > Municipal

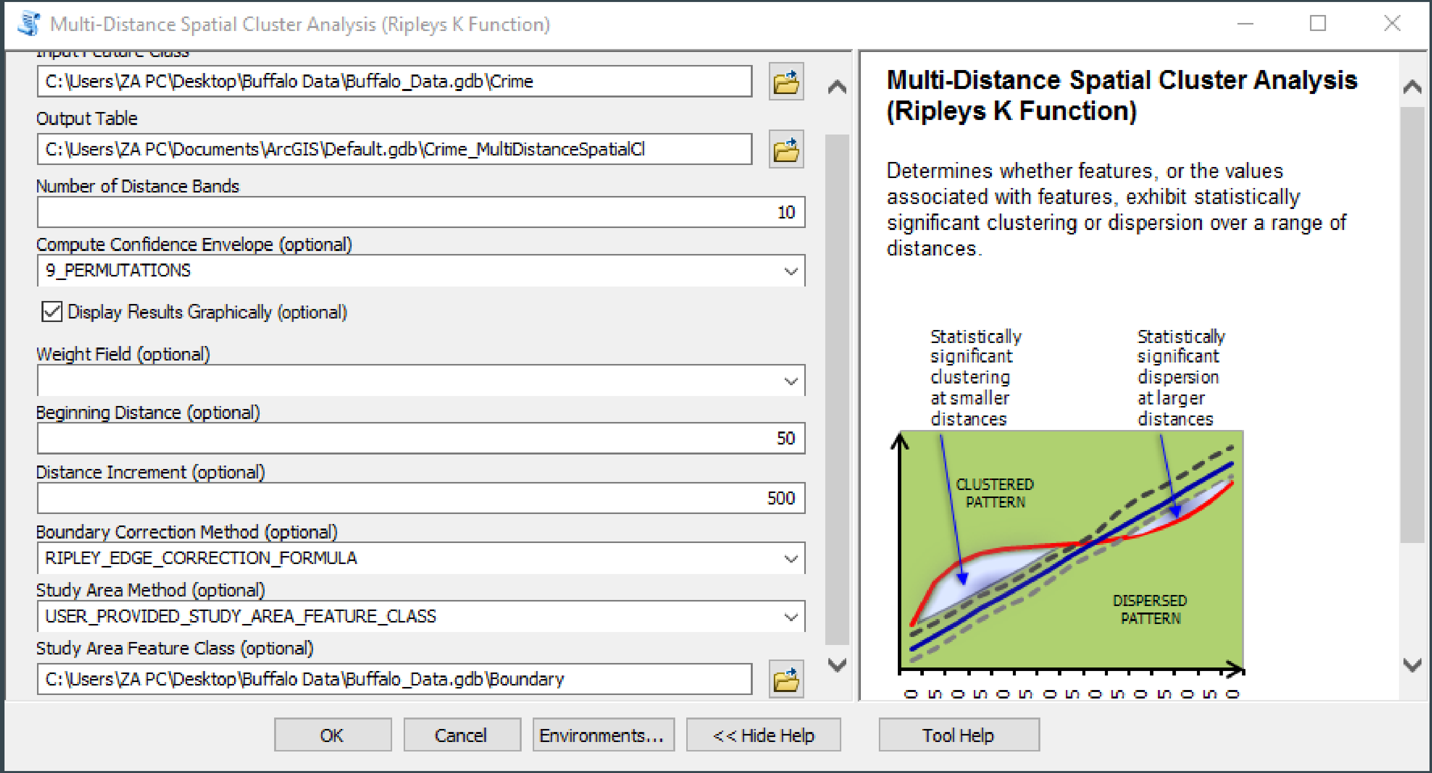






Understanding clustering of crimes at different spatial scale

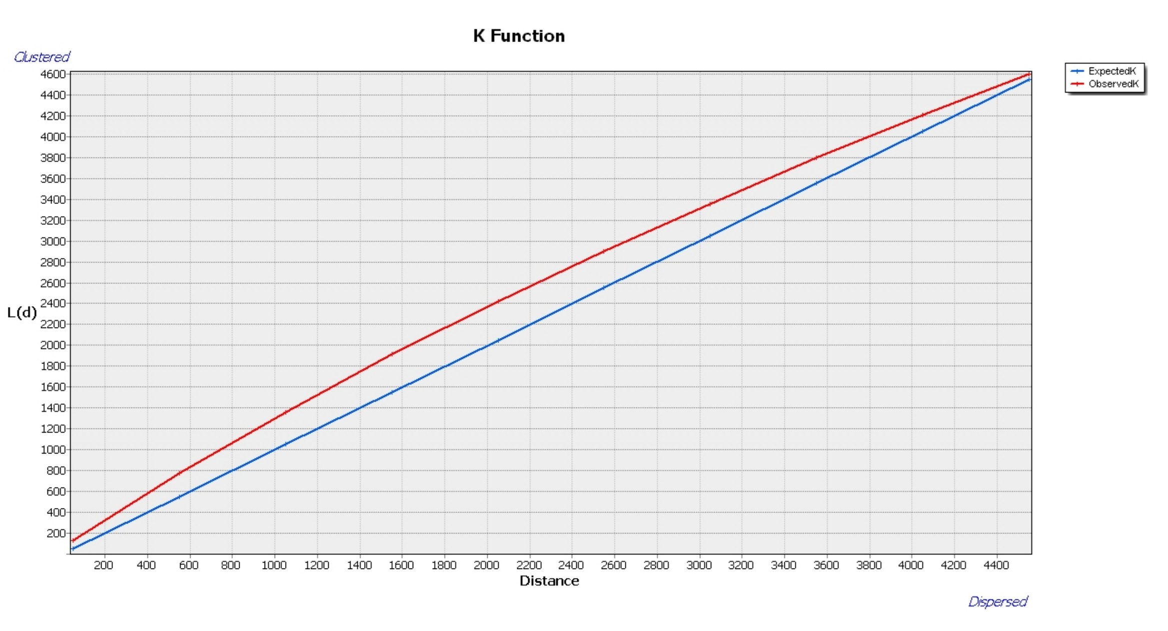
ArcGIS tool: Multi-Distance Spatial Clustering Analysis (Ripley's K Function)



Understanding clustering of crimes at different spatial scale

- Unweighted K-function results to get a baseline understanding
- Clustering observed at all distance bands
- At 2050m, most distinct clustering observed

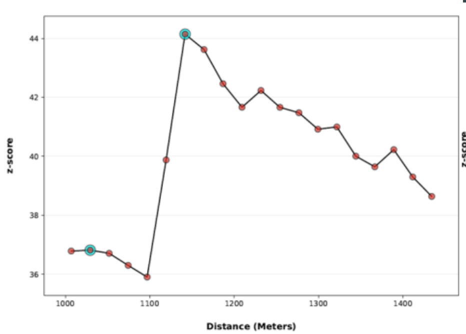
Distance Band	ExpectedK	ObservedK	DiffK
1	50	132.27	82.27
2	550	776.34	226.34
3	1050	1363.11	313.11
4	1550	1913.23	363.23
5	2050	2425.28	375.28
6	2550	2904.22	354.22
7	3050	3357.71	307.71
8	3550	3797.83	247.82
9	4050	4211.54	161.54
10	4550	4599.54	49.54



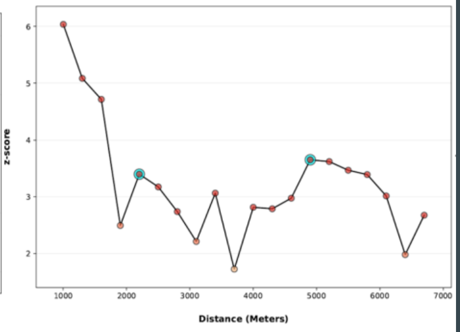
Distinct Patterns of Clustering Exhibited by Different Crime Types

Property Crimes

Burglary

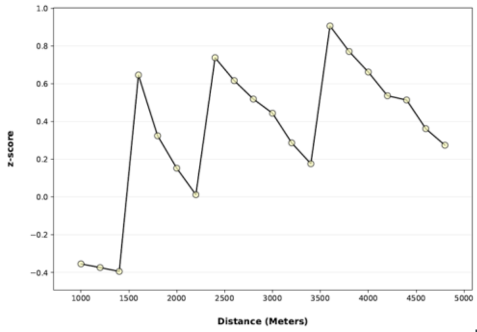


Robbery

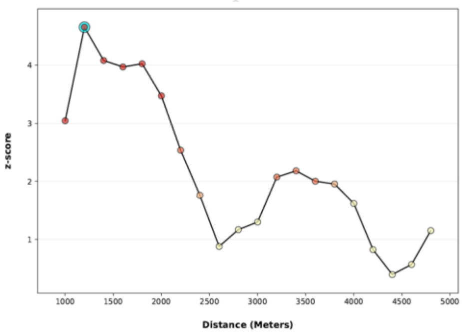


Felony Crimes

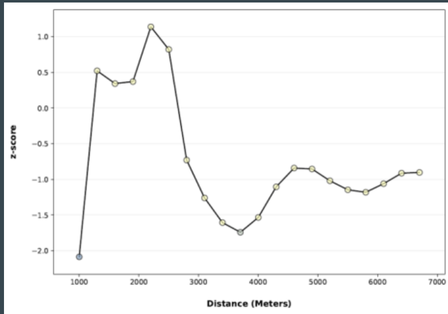
Homicide



Rape



Aggr. Assault



Possible explanations and implications

- Spatial processes operate at different scales: Different motivations and external factors that may encourage/discourage such crimes
- The need for differentiated data treatment for subsequent analysis

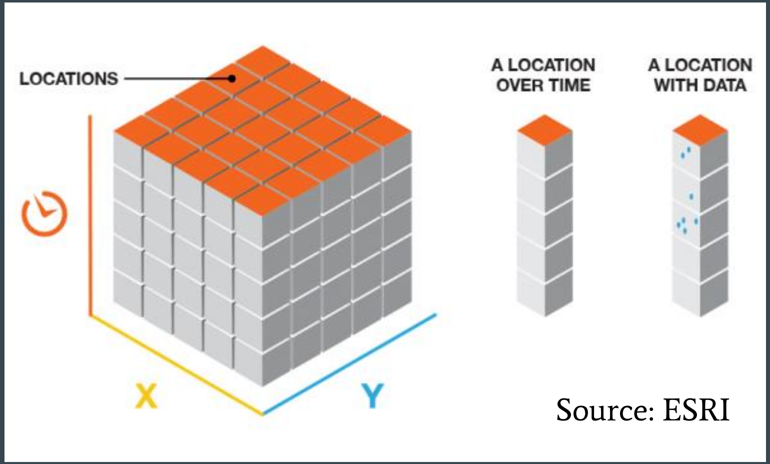
Spatiotemporal Hotspot

Assisting resource allocation and police force deployment

Everything happens somewhere and occurs at some point in time

Space-Time Cube Analysis

- Generate statistical hot and cold spots within a set study area
- Identify the change and predict where the crime pattern may appear



Source: ESRI

- Use the Mann-Kendall Trend Test to analyse trend
  - The Mann-Kendell statistic is a rank correlation analysis for the bin count or value and their time sequence.
  - Analyzes difference in signs between earlier and later data points.
    - $X1 < X2 \rightarrow +1$
    - $X1 > X2 \rightarrow -1$
    - $X1 = X2 \rightarrow 0$

(Kendall & Gibbons, 1990)

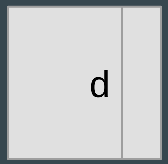
Distance and Time Interval

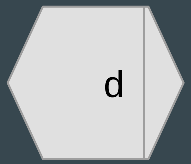
The Distance Interval

- Determines the size used to aggregate the data points.
- Fishnet/Hexagon Grid

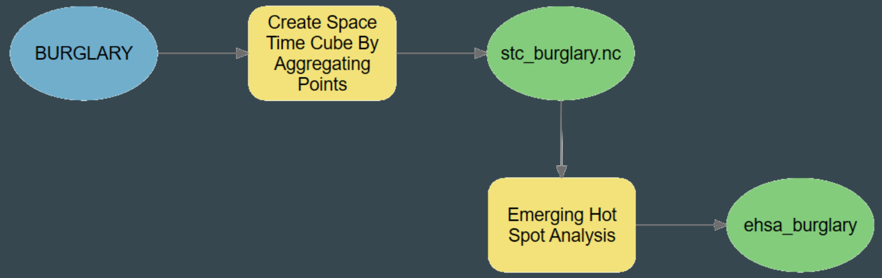
The Time Step Interval

- Specifies the time span for each bin
- Aggregate points using attribute “date”





Emerging Hot Spot Analysis



Create Space Time Cube by Aggregating Points

- Generate netCDF (Network Common Data Form) to store summarized point data in the space-time bins.
- Parameters:
  - Input: Crime Type
  - Output: Space Time Cube in netCDF
  - Time Field: Date
  - Time Step Interval: 1 Year
  - Aggregation Shape Type: Hexagon Grid
  - Distance Interval: 500m

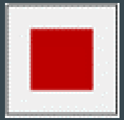
Emerging Hot Spot Analysis

- Spot trends in the clustering of point densities or values in a space-time cube
- Parameters:
  - Input: Space Time Cube of Crime Incident
  - Output: Hot Spot & Cold Spot 2D Map
  - Analysis Var: Count
  - Neighborhood Distance: 500m

Hot Spot Patterns

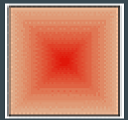
New Hot Spot

The most recent time step interval is hot for the first time.



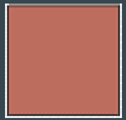
Intensifying Hot Spot

At least 90% of the time step intervals are hot, and becoming hotter over time.



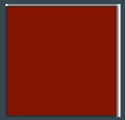
Consecutive Hot Spot

A single uninterrupted run of hot time step intervals, comprised of less than 90% of all intervals.

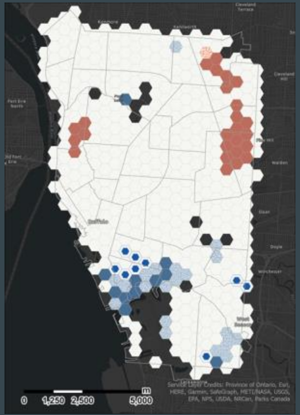


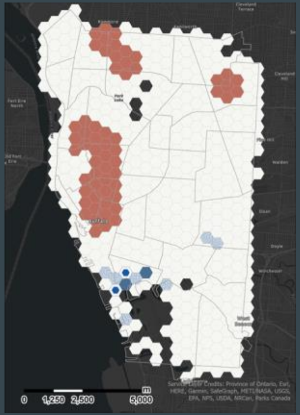
Persistent Hot Spot

At least 90% of the time step intervals are hot, with no trend up or down.

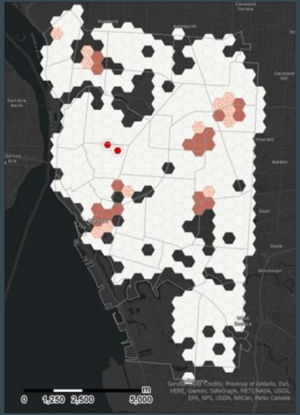


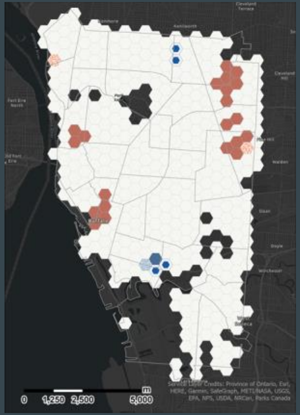
Emerging Hot Spot Analysis for Each Crime Type

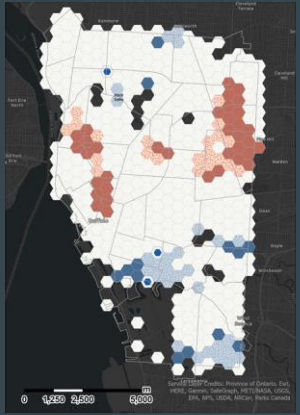

Burglary


Larceny Theft


Murder


Rape


Robbery


Theft of Vehicle

- New Hot Spot
- Consecutive Hot Spot
- Intensifying Hot Spot
- Persistent Hot Spot
- Diminishing Hot Spot
- Sporadic Hot Spot
- Oscillating Hot Spot
- Historical Hot Spot
- New Cold Spot
- Consecutive Cold Spot
- Intensifying Cold Spot
- Persistent Cold Spot
- Diminishing Cold Spot
- Sporadic Cold Spot
- Oscillating Cold Spot
- Historical Cold Spot
- No Pattern Detected

- Distance interval: **500m**
- Time Interval: **1 year**
- Not Significant: **Murder and Rape**

	% non-zero	Trend statistic	Trend p-value
Burglary	83.17%	-1.9677	0.0491
Larceny/Theft	88.53%	-1.7889	0.0736
Murder	19.15%	0.5367	0.5915
Rape	41.66%	0.3578	0.7205
Robbery	67.52%	-2.1466	0.0318
Theft of Vehicle	69.96%	-1.4311	0.1524

Modelling

Establishing correlation with environmental and socioeconomic factors that might encourage Part I Crimes

Classifying 311 Call Reason/Type

THE BROKEN WINDOW THEORY

- Correlation between disorder and incivility, and criminal activities

Konkel, Ratkowski & Tapp (2019) tested the hypothesis on crime incidents in Milwaukee, Wisconsin. A few categories of “civil disorder” behaviours were devised:

- *Social Disorder*: people loitering, drinking in public, buying/selling drugs, gambling, physical fights
- *Public Space Disorder*: graffiti, trash/litter, broken glass on the street/sidewalk, abandoned cars
- *Housing Disorder*: houses with falling/detached siding/gutters, houses with chipping/peeling paint, parcels with unkempt/overgrown lawns etc.

311 Call Reason/Type manually classified into:

- Public Space Disorder
- Housing Disorder
- Social Disorder

CallType

Reason	Type
ADA	ADA-Other (Req_Serv)
	ADA-PW Sidewalks (Req_Serv)
Adjudication - Ordinance Violation	Excess Trash (Req_Serv)
	Illegal Dumping (Req_Serv)
	Ordinance Violation (Req_Serv)
	Other Adjudication Issue (Req_Serv)
Administration	Fair Housing Issue (Req_Serv)
Animal Shelter	Animals (Req_Serv)
	Dead Animal Removal (Req_Serv)
Assessment	2020 Reassessment
	Assessment Issue (Req_Serv)
Assessment & Taxation	Assessment Issue (Req_Serv)
BFD	BFD Fire Prevention (Req_Serv)
	BFD Snow on Hydrant (Req_Serv)
	Fire (Req_Serv)
BMHA	BMHA Issue (Req_Serv)
Buffalo Sewer Authority	Basement Flooding (Req_Serv)
	Rain Barrels (Req_Serv)
	Sewer (Req_Serv)
	Street Flooding (Req_Serv)
Buffalo Water Authority	Fire Hydrant Issue (Req_Serv)
	Water (Req_Serv)
	Water Issue (Req_Serv)
	Water Tested (Req_Serv)
	Water_Billing_Meter (Req_Serv)
Buildings Division	Building Maintenance (Req_Serv)
	CityHall_CityCourt Maintenance (Req_S..
Citizen Services - Good Neighbor	Good Neighbor (Req_Serv)
Citizen Services - Graffiti	City Property (Req_Serv)
	Obscene City Property (Req_Serv)
	Obscene Other (Req_Serv)
	Obscene Parks City (Req_Serv)
	Obscene Parks Olmsted (Req_Serv)
	Obscene Private Property (Req_Serv)
	Obscene PW Engineering (Req_Serv)

Exploratory Regression

What might explain Part I Crimes at Neighbourhood Scale?

A data mining tool that:

- Receives user-specified OLS diagnostics
- Tests all combinations of explanatory variables to fit OLS regression model
- Generates a report of model suitability

Exploratory Regression is suitable when working with a large number of explanatory variables.

"Best" OLS model to explain Part I Crime in neighbourhoods

- Employment Rate (+)
- Poverty Rate (+)
- % Age >65 (-)
- % Renter Tenure (+)
- Median value for Rent-burdened renters (+)
- % Single-person household (+)
- Diversity Index (+)
- Social Disorder (-)
- Public Space Disorder (+)

Assess Explanatory Variable Multicollinearity

Search Criteria

Maximum Number of Explanatory Variables	9
Minimum Number of Explanatory Variables	7
Minimum Acceptable Adj R Squared	0.7
Maximum Coefficient p value Cutoff	0.05
Maximum VIF Value Cutoff	7.5
Minimum Acceptable Jarque Bera p value	0.1
Minimum Acceptable Spatial Autocorrelation p value	0.1

********** Exploratory Regression Global Summary (PARTICRIMEBYAREA) **********

Percentage of Search Criteria Passed

Search Criterion	Cutoff	Trials	# Passed	% Passed
Min Adjusted R-Squared	> 0.70	371450	112	0.03
Max Coefficient p-value	< 0.05	371450	32	0.01
Max VIF Value	< 7.50	371450	320467	86.27
Min Jarque-Bera p-value	> 0.10	371450	194580	52.38
Min Spatial Autocorrelation p-value	> 0.10	12	12	100.00

Choose 9 of 20 Summary

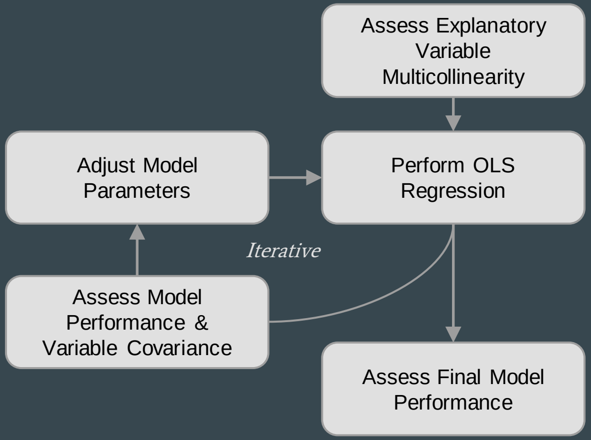
AdjR2	AICc	JB	K(BP)	VIF	SA	Model
0.72	-429.33	0.91	0.29	3.31	0.86	+EMPLOYMENT_RATE*** +POVERTY_RATE***
0.72	-429.15	0.56	0.22	4.09	0.57	+EMPLOYMENT_RATE*** +POVERTY_RATE***
0.72	-429.02	0.86	0.40	2.72	0.60	+EMPLOYMENT_RATE*** +POVERTY_RATE**

Summary of Variable Significance

Variable	% Significant	% Negative	% Positive
PERCENT_RENTER_TENURE	92.16	0.00	100.00
EMPLOYMENT_RATE	50.57	0.20	99.80
POVERTY_RATE	48.71	4.03	95.97
PERCENT_BLACK	43.38	8.28	91.72
DIVERSITYINDEX	39.73	9.03	90.97
MEDIAN_VALUE_FOR_RENT_BURDENED_RENTERS	35.76	1.18	98.82
PERCENT_SINGLE_PERSON_HOUSEHOLDS	28.74	5.23	94.77
PERCENT_AGE_65__	22.41	99.30	0.70
COUNTGRAFFITI	20.71	3.27	96.73
PERCENT_HIGH_SCHOOL_EDUCATION	16.66	83.25	16.75
AGE___24	16.06	9.40	90.60
PERCENT_20_OR_MORE_UNIT_STRUCTURE	15.27	47.75	52.25
PERCENT_FEMALE_HOUSEHOLDER_W__CHILDREN_UNDER_18	11.49	57.85	42.15
COUNTHSDISORDER	10.45	26.98	73.02

Ordinary Least Square (OLS) Regression

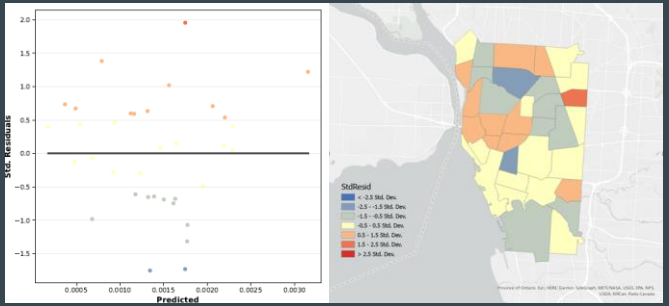
What might explain Part I Crimes at Neighbourhood Scale?



Establish a global OLS regression model with explanatory variables recommended by Explanatory Regression:

- Employment Rate (*)
- Poverty Rate (*)
- % Age > 65 (*)
- % Rental property (*)
- Median value of rent-burdened renters (*)
- % Single person household
- Diversity Index (*)
- Social Disorder (*)
- Public Space Disorder (*)

Number of Observations:	35	Akaike's Information Criterion (AICc) [d]:	-429.331248
Multiple R-Squared [d]:	0.795686	Adjusted R-Squared [d]:	0.722133
Joint F-Statistic [e]:	10.817857	Prob(>F), (9,25) degrees of freedom:	0.000001*
Joint Wald Statistic [e]:	247.842448	Prob(>chi-squared), (9) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	10.837608	Prob(>chi-squared), (9) degrees of freedom:	0.287005
Jarque-Bera Statistic [g]:	0.183326	Prob(>chi-squared), (2) degrees of freedom:	0.912413



Results from OLS regression suggest random distribution of residue

Results suggest:

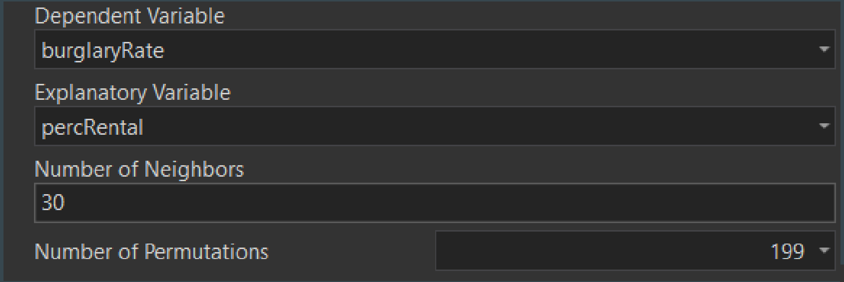
- Low multicollinearity among explanatory variables
- Model performance is relatively good (R2 and adjusted R2 > 0.7)
- Stationary relationship
- Residues are randomly distributed - non-bias

Local Bivariate Relationship

*What if the relationship changes **form** spatially?*

Bivariate relationship determined by:

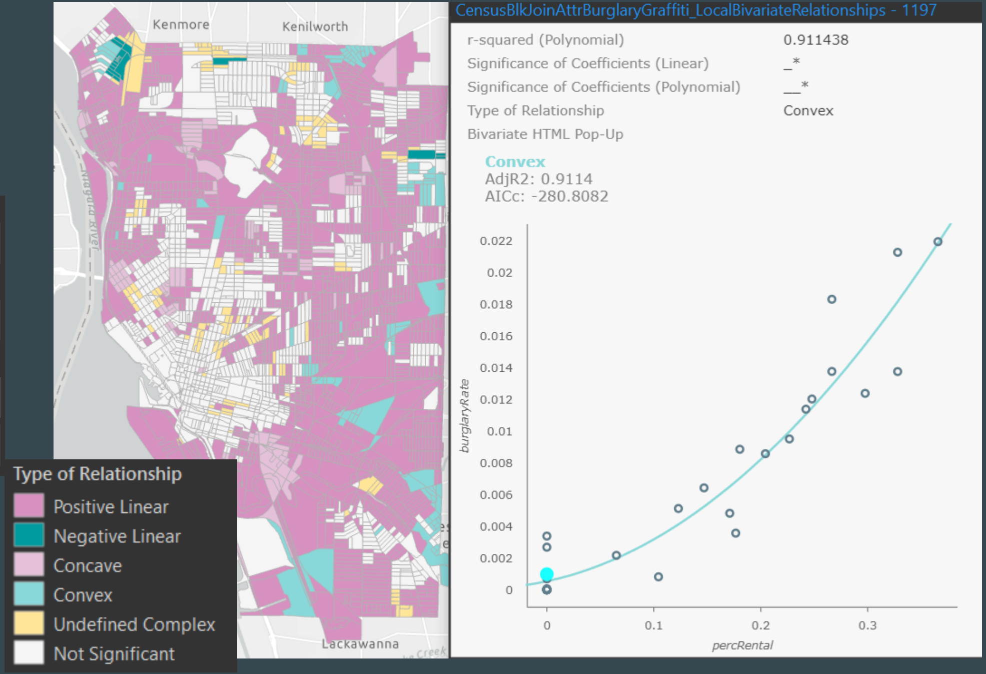
- Assessing the statistical significance of the null hypothesis that the two variables are independent, based on comparison of *joint entropy* and sum of individual entropies
- Construct random permutations of x & y and and test for local spatial relationships
- Classify the local relationships



Entropy (Information Theory): a measure of uncertainty in a variable

- High uncertainty → high entropy
- High dependency between two variables → low joint entropy
- Assessed using power-weighted minimum spanning trees

- What is the relationship between Burglary Rate and Percentage Rental Properties?
- Is the relationship consistent across the study area?



For discussion

Choosing suitable scales (with suitable segmentation of data)

- Spatial processes operate in different scales for different phenomenon
- Aggregating the phenomenon - unclear patterns of spatial autocorrelation (assumption: 1st law of geography)

Correlation =/= Causation: some cautions with Exploratory Regression

- Data mining approach in Exploratory Regression disregards any meaning of the relationship: an exercise of numbers; overfitting?

- Plausible mechanisms between explanatory variables and dependent variables should be discussed based on *domain knowledge and theories*

How do we account for time in spatial analysis, especially in modelling?

- Step-wise / time intervals?
- How to account for long-term or lagged dependency between variables?

Thank You

References

Kendall, M. G. and Gibbons, J. D. (1990). Rank Correlation Methods, fifth ed. Griffin, London.

Konkel, R. H., Ratkowski, D., & Tapp, S. N. (2019). The Effects of Physical, Social, and Housing Disorder on Neighbourhood Crime: A Contemporary Test of Broken Windows Theory. International Journal of Geo-Information, 2019, 8, 573, doi:10.3390/ijgi8120583.

Leitner, M. (2013). Crime Modeling and Mapping Using Geospatial Technologies. Dordrecht: Springer.

Schnell, C., Braga, A.A. & Piza, E.L (2017) The Influence of Community Areas, Neighborhood Clusters, and Street Segments on the Spatial Variability of Violent Crime in Chicago. *J Quant Criminol* 33, 469–496

Elmes, G. A., Roedl, G., & Conley, J. F. (2016). Forensic GIS: The role of geospatial technologies for investigating crime and providing evidence. Dordrecht: Springer.