The Economics of crime - A cross-sectional analysis on Toronto neighbourhoods

Gabriel Blumenstock, Felix Degenhardt and Haseeb Warsi

Humboldt-Universität zu Berlin

07/06/2018

Outline

1. Introduction
   - What is economics of crime?
   - Focus of this presentation
2. Data Structure
   - Initial Findings
3. Regression Analysis
   - Procedure
   - Results
4. Outlook

What is time economics of crime?

- Evolved in the 1970 and is a broad discipline since then
- Investigation on both behavioral (e.g. Becker 1968) as well as socio-economic aspects (e.g. Buonnano 2003)
- Major econometric investigations: time series (e.g. Cornman and Mocan 2000), panel data (e.g. Cornwell and Trumbull 1994), cross-sectional analysis (e.g. Kelly 2000)

Focus of this presentation

- Cross-sectional analysis of the effects of Toronto neighbourhood characteristics on the crime rate
- Combination of 2014 – 2017 crime statistics and 2016 Census Data

Data Structure

- More than 130,000 observations of Toronto crime occasions between 2014 – 2017
- Around 15 major characteristics of each of the 140 Toronto neighbourhoods

$\rightarrow$ Transformation to count data with both characteristics and crime per neighbourhood

Data Structure

Hood_ID	Assault	Auto Theft	Break & Enter	. . .	% youth	% low income	% immi-grants
1	0.0068	0.0083	0.0037	. . .	16.34	54.72	58.46
2	0.0066	0.0012	0.0010	. . .	15.90	56.44	65.48
⋮	⋮	⋮	⋮	. . .	⋮	⋮	⋮
139	0.0087	0.00047	0.0019	. . .	14.11	53.08	54.60
140	0.0039	0.0005	0.0017	. . .	10.78	38.87	32.43

Figure 1: Structure of final data set.

Crime by Time of Day

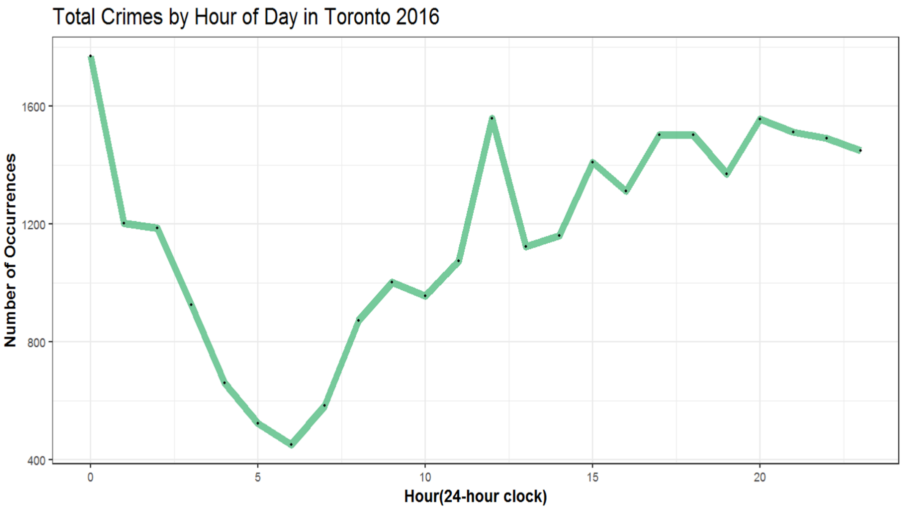


Figure 2: Crimes by Hour of Day in Toronto 2016.

Crime by Time of Day

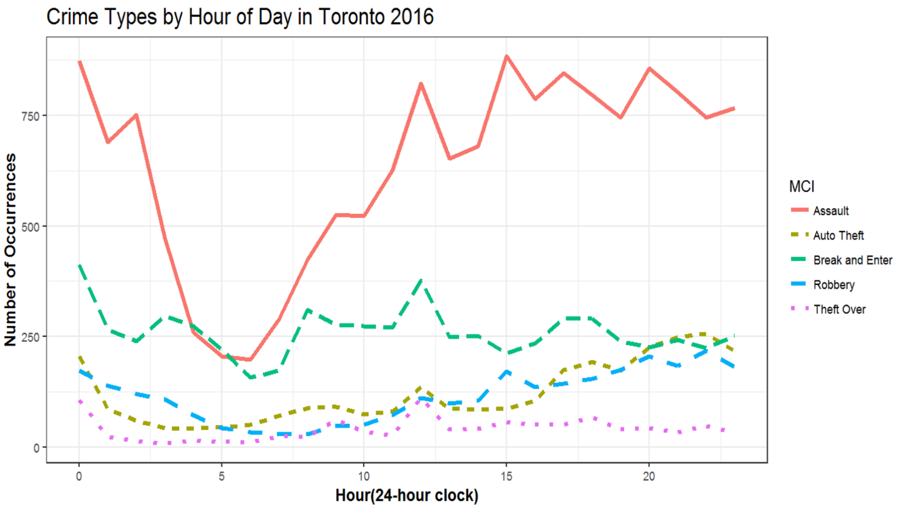


Figure 3: Crime Types by Hour of Day in Toronto 2016.

Crime by Month

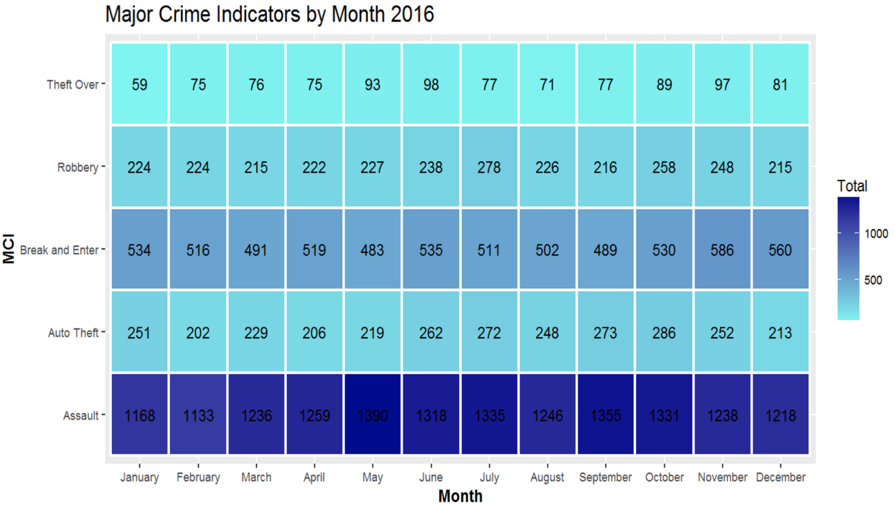


Figure 4: Major Crime Indicators by Month 2016.

Major Crime Indicators

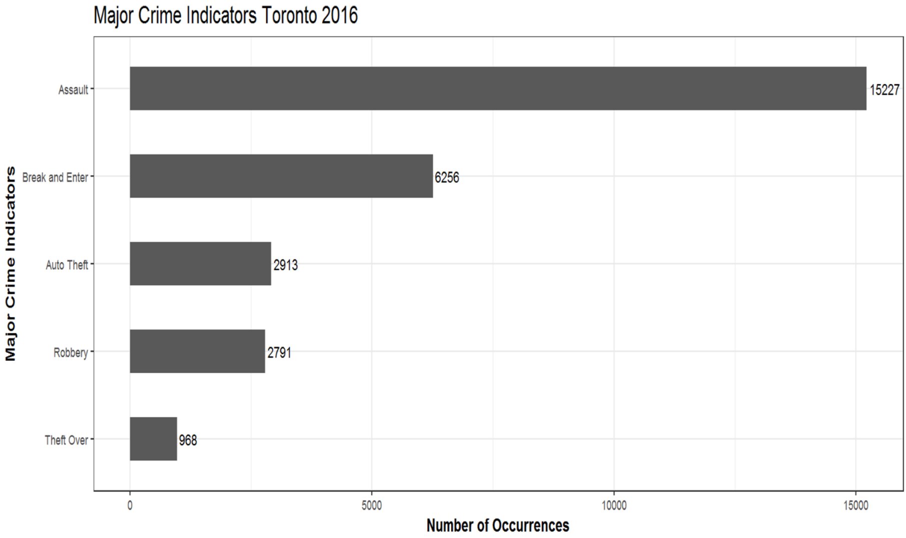


Figure 5: Major Crime Indicators Toronto 2016.

Major Crime Indicators

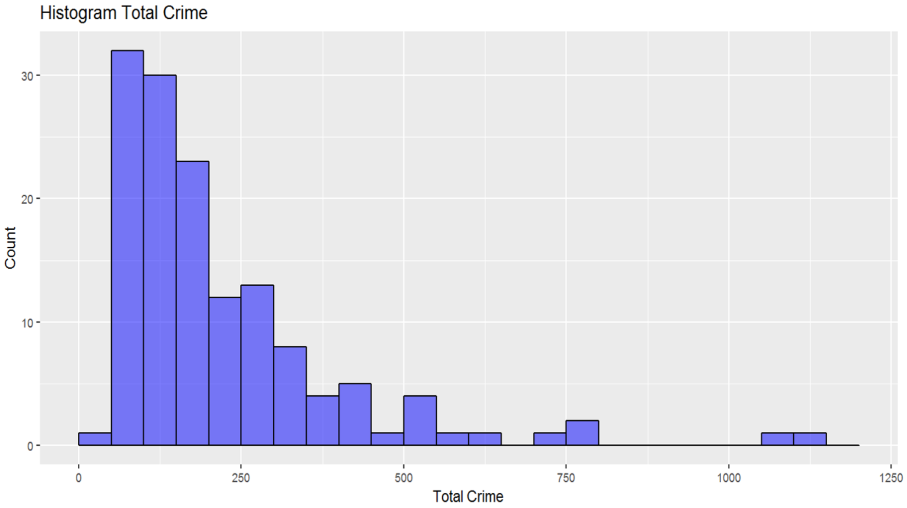
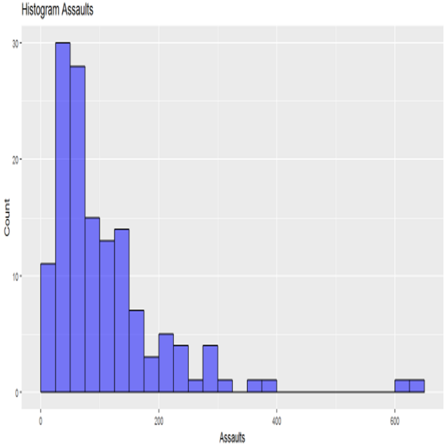
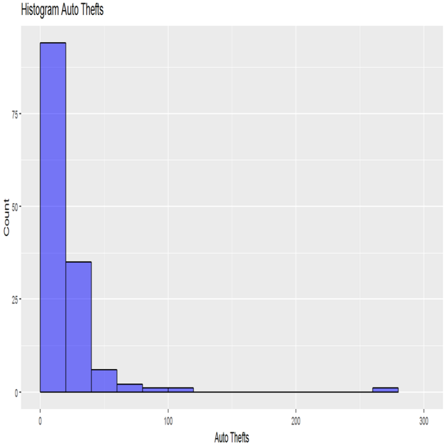


Figure 6: Histogram total crime.

Major Crime Indicators



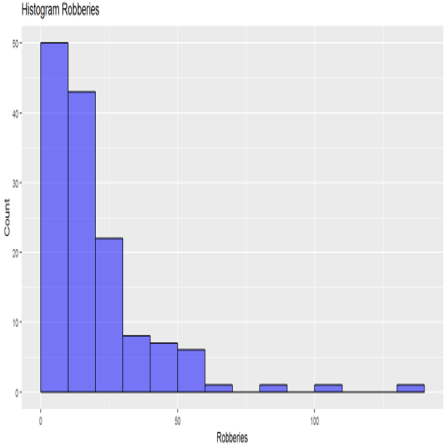
(a) Assaults



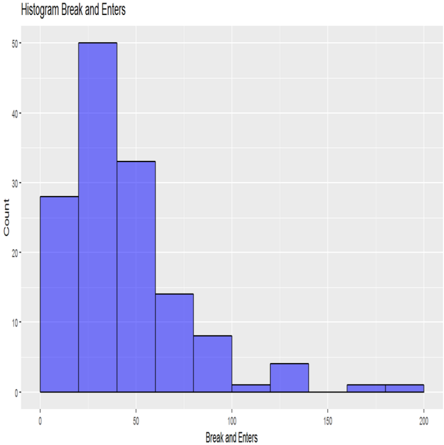
(b) Auto Thefts

Figure 7: Histogram by type of crime.

Major Crime Indicators



(a) Robberies



(b) Break and Enters

Figure 8: Histogram by type of crime.

Most Dangerous Neighbourhoods

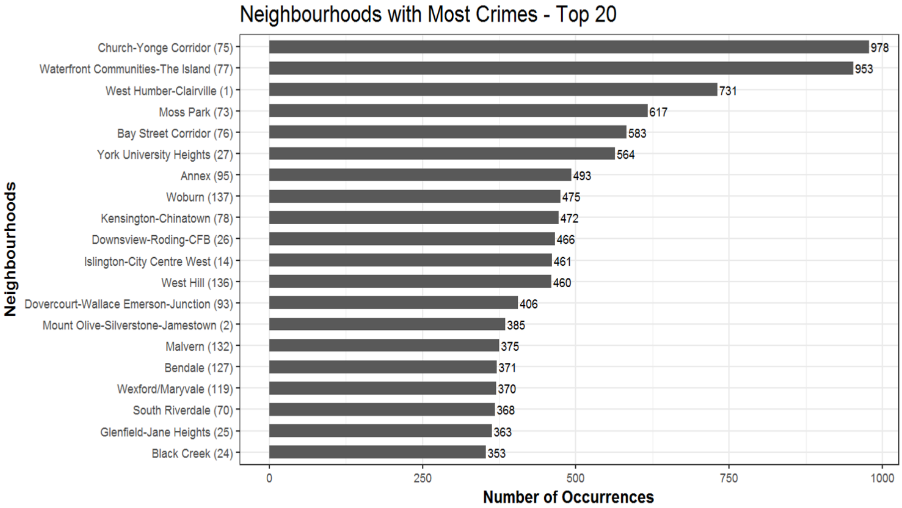


Figure 9: Neighbourhoods with Most Crimes.

Most Dangerous Neighbourhoods

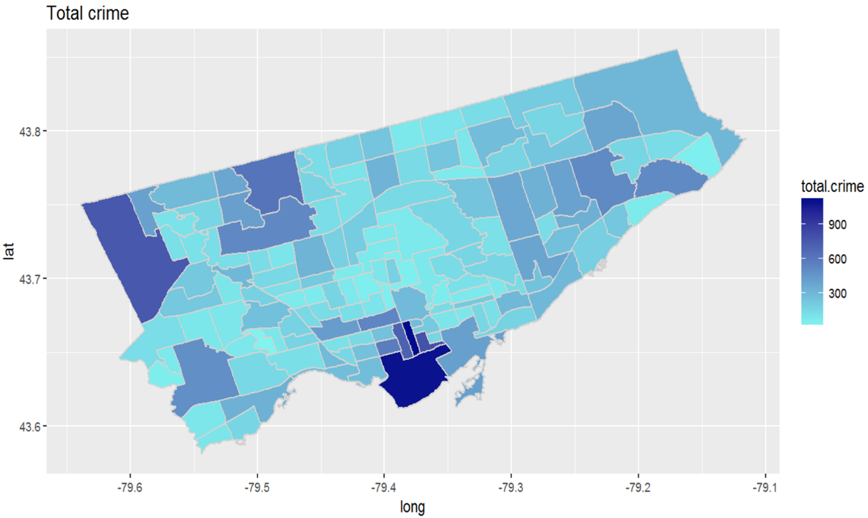


Figure 10: Heatmap of Total Crime.

Most Dangerous Neighbourhoods

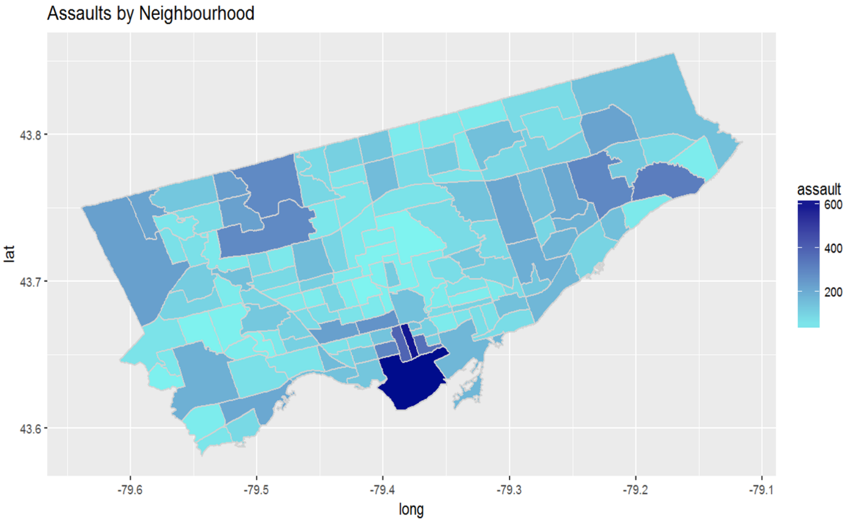
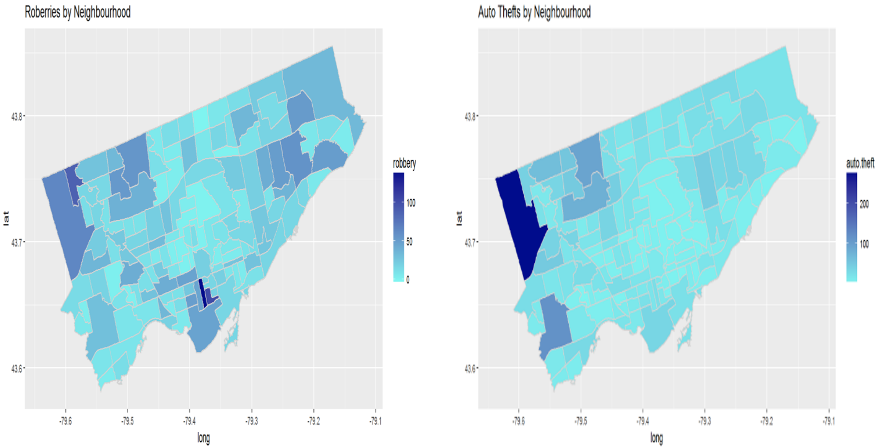


Figure 11: Heatmap of Assaults



(a) Robberies (b) Assaults

Figure 12: Heatmaps by type of crime.

Neighbourhood Characteristics

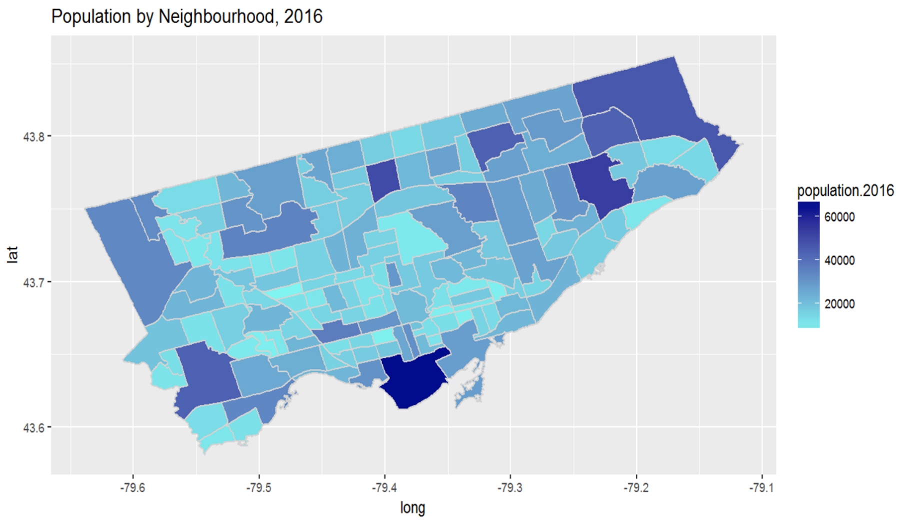


Figure 13: Heatmap Population by Neighbourhood.

Neighbourhood Characteristics

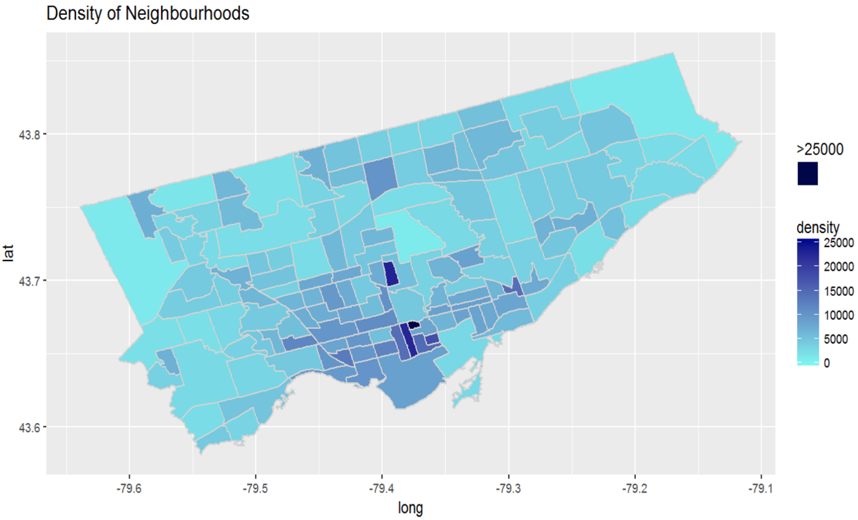
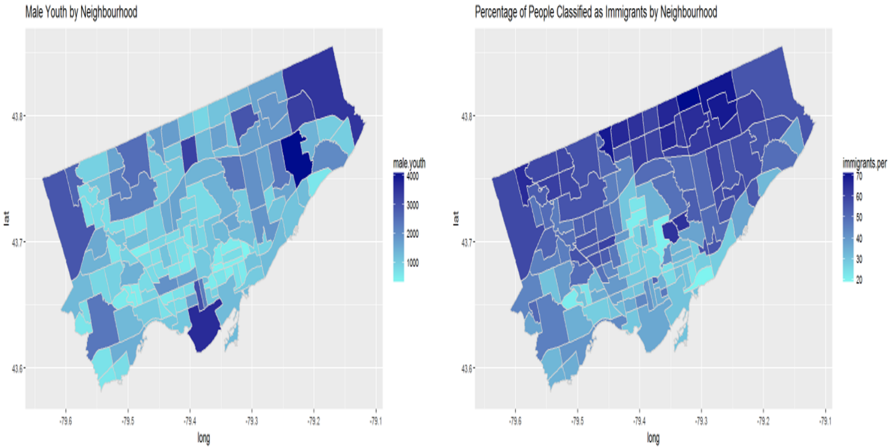


Figure 14: Heatmap Density of Neighbourhoods.

Neighbourhood Characteristics

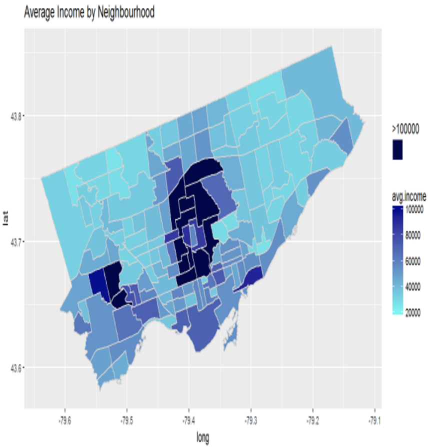


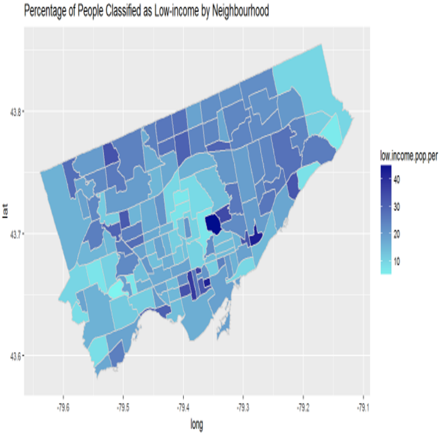
(a) Male Youth by Neighbourhood

(b) Percentage of Immigrants

Figure 15: Heatmaps of Neighbourhood Characteristics

Neighbourhood Characteristics

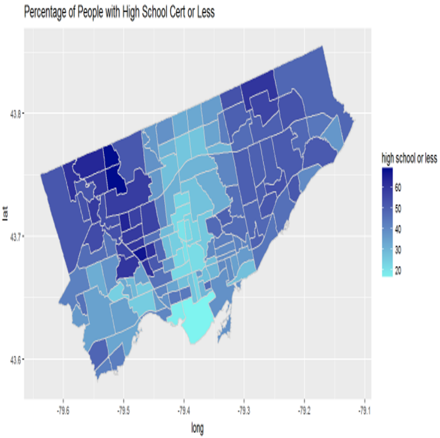




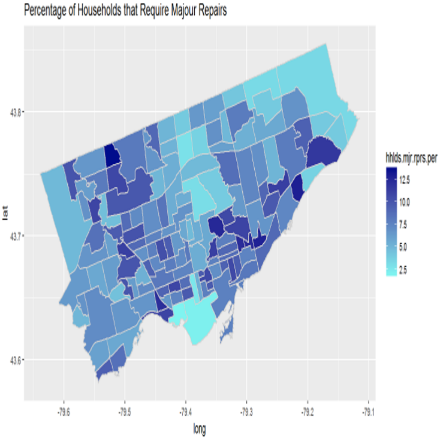
(a) Average Income by Neighbourhood (b) Percentage of Low-Income Earners

Figure 16: Heatmaps of Neighbourhood Characteristics

Neighbourhood Characteristics



(a) High School Cert. or less



(b) Households that require Major Repairs

Figure 17: Heatmaps of Neighbourhood Characteristics

Grouping Using k-means

$$J(V) = \sum_{i=1}^{C} \sum_{j=1}^{C_i} (\|x_i - v_j\|)^2$$

where

$V = \{v_1, v_2, \ldots, v_c\}$ are the set of centers,

$\|x_i - v_j\|$ is the Euclidian distance between x_i and v_j,

c_i is the number of data points in the i^{th} cluster and

c is the number of cluster centers.

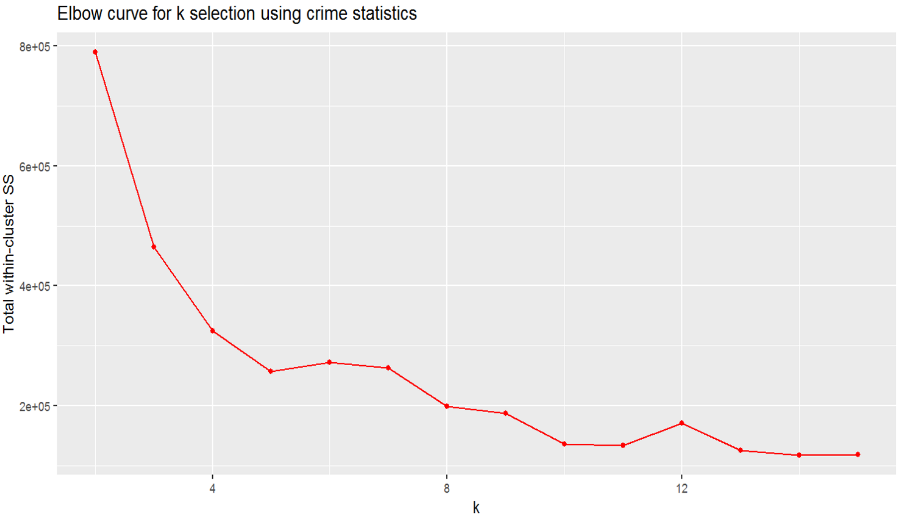


Figure 18: Elbow curve for k-selection using crime statistics.

Grouping Using k-means

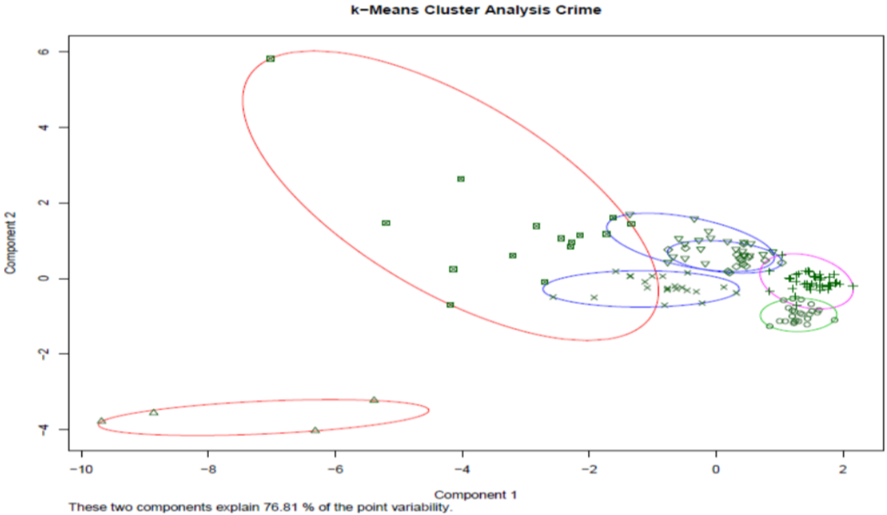
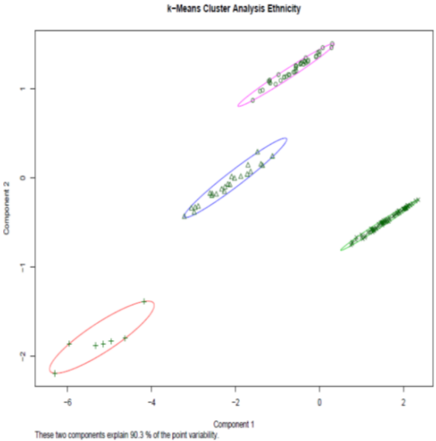
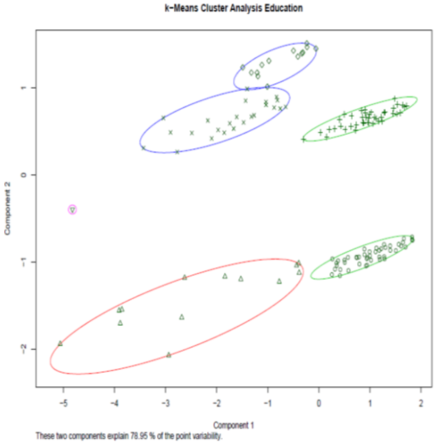


Figure 19: k-means Cluster Analysis Crime.

Grouping Using k-means



(a) Ethnicity



(b) Education

Figure 20: More k-means Cluster Analyses

Regression Analysis

- Seven types of crime $\rightarrow$ seven regressions with independent variables Assault, Auto theft, Break and Enter, Robbery, Theft Over, Drug Arrests, Total Crime
- 30 possible independent variables
- Selection of dependent variables male youth population, less than high school population, low income population, immigrants

Procedure after each regression:

- check for influential observations
- variable selection
- check for OLS-assumptions

- Example: Regression for robbery data

$$robbery = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 + \beta_4 * x_4 + u$$

where $x_1 = male.youth$,
$x_2 = less.than.high.school$,
$x_3 = low.income$,
$x_4 = immigrants$,
$u = error\ term$

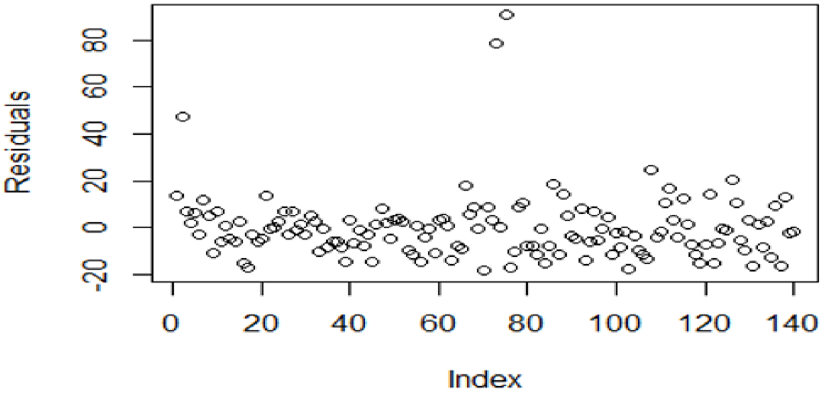
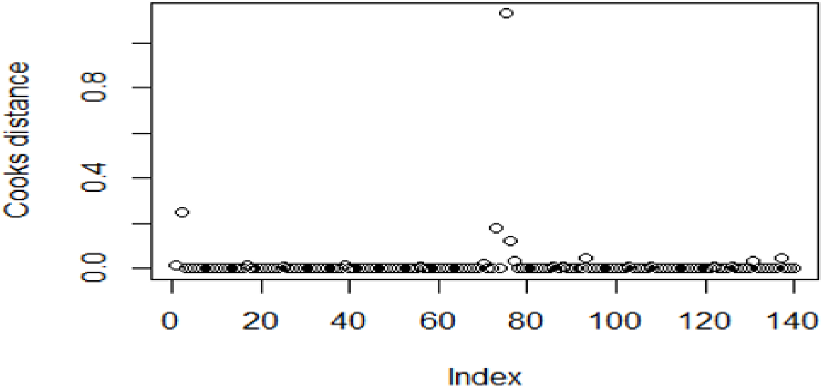


Figure 21: Check for Influential Observations.

Regression Analysis



```
> step(model_robbery)
Start:  AIC=752.1
robbery ~ male.youth + less.than.high.school + low.income + immigrants

                        Df Sum of Sq   RSS    AIC
<none>                               28068 752.10
- less.than.high.school  1     671.6 28740 753.42
- low.income             1    1711.6 29780 758.39
- male.youth             1    3606.8 31675 767.03
- immigrants             1    5571.3 33639 775.45

Call:
lm(formula = robbery ~ male.youth + less.than.high.school + low.income + 
    immigrants, data = r)

Coefficients:
          (Intercept)             male.youth  less.than.high.school             low.income  
            -5.576685               0.020819               0.001985               0.003358  
           immigrants  
            -0.004072  
```

Figure 23: Variable selection using Akaike Information Criterion.

Tests for normality	p-value
Shapiro-Wilk	0.00%
Anderson-Darling	0.00%
Lilliefors (Kolmogorov-Smirnov)	0.00%

Tests for heteroscedasticity	p-value
Studentized Breusch-Pagan test	0.00%

Figure 24: Check for OLS-Assumtions.

Regression Analysis

Assault		Auto theft		Break and enter		Robbery	
-44.389225	***	-1.6744493		2.586451		-5.576685	**
0.08942	***	0.0243493	***	0.023912	***	0.0208186	***
-0.009738	**	0.0057757	***	-0.00767	***	0.0019854	*
0.034263	***	-0.0018859		0.011376	***	0.0033584	***
-0.027793	***	-0.0005785		-0.008091	***	-0.004072	***
66.17%		24.69%		58.91%		45.30%	
Theft over		**Drug arrests**		**Total crime**			
-3.1328562	***	-7.124131	*	-59.310895	***		
0.0089558	***	0.022357	***	0.189813	***		
-0.0018057	***	-0.003351	**	-0.014803	*		
0.0020808	***	0.008842	***	0.058035	***	***	<0.01
-0.001721	***	-0.008113	***	-0.050367	***	**	<0.05
52.76%		39.89%		66.61%		*	<0.1

Figure 25: Preliminary Results.

Other Regressions

- Again for Robbery data: log-transformed dependent variable

$$log.robbery = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \beta_3 * x_3 + \beta_4 * x_4 + u$$

where $x_1 = male.youth$,
$x_2 = less.than.high.school$,
$x_3 = low.income$,
$x_4 = immigrants$,
$u = error\ term$

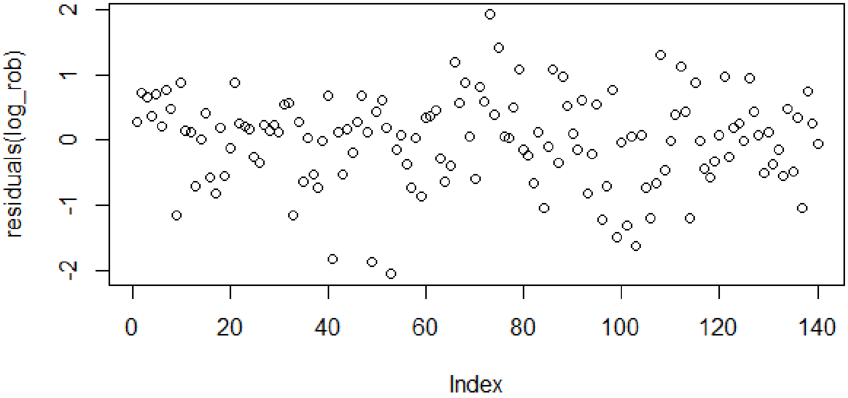


Figure 26: Residuals log-transformed model.

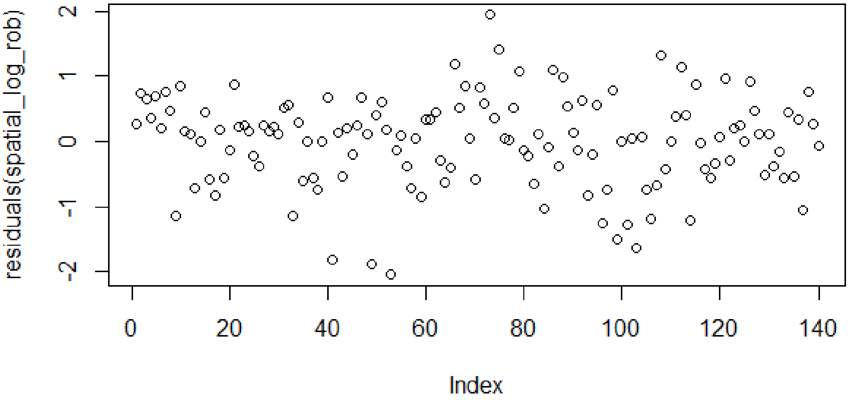


Figure 27: Residuals of spatial regression.

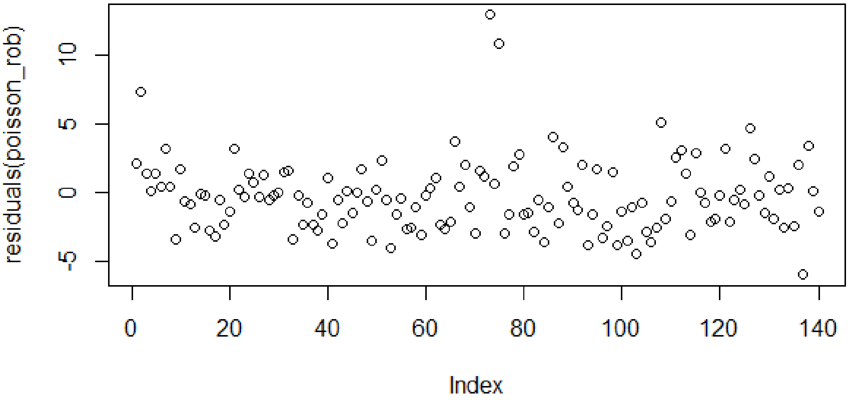


Figure 28: Residuals of Poisson Regression.

Test for normality	**p-value**
Shapiro-Wilk	0.09619
Lilliefors (Kolmogorov-Smirnov)	0.006859
Test for Heteroscedasticity	
Studentized Breusch-Pagan test	0.06004

Figure 29: Check for OLS-Assumptions.

Regression Analysis

Table 1: Preliminary Results of Different Regression Approaches

	Dependent variable:			
	Robbery	Log.Robbery		Robbery
	OLS	*OLS*	*spatial autoregressive*	*Poisson*
	(1)	(2)	(3)	(4)
Male youth	0.021***	0.001**	0.001**	0.001***
	(0.005)	(0.0002)	(0.0002)	(0.0001)
Less than High School	0.002*	0.0002***	0.0002***	0.0001***
	(0.001)	(0.0001)	(0.0001)	(0.00001)
Low Income Households	0.003***	0.0002***	0.0002***	0.0001***
	(0.001)	(0.0001)	(0.0001)	(0.00002)
Visible immigrants	−0.004***	−0.0002***	−0.0002***	−0.0002***
	(0.001)	(0.00004)	(0.00004)	(0.00001)
Constant	−5.553**	1.384***	1.497***	1.914***
	(2.760)	(0.135)	(0.338)	(0.043)
Observations	140	140	140	140
R^2	0.468	0.447		
Adjusted R^2	0.453	0.431		
Log Likelihood			−147.010	−817.725
σ^2			0.478	
Akaike Inf. Crit.			308.021	1,645.450
Residual Std. Error (df = 135)	14.417	0.705		
F Statistic (df = 4; 135)	29.748***	27.328***		
Wald Test			0.131 (df = 1)	
LR Test			0.129 (df = 1)	
Note:			*p<0.1; **p<0.05; ***p<0.01	

Outlook

- Taking into account the spatial correlations
- Finding a way to deal with count data appropriately
- Implementing additional tests for verification

End of Presentation

Thank you for your attention