Enhancing the Design of Observational Studies of Community Policing

Using geospatial data mining to design non-experimental program evaluations

Christian Kreis

Netherlands Institute for the Study of Crime and Law Enforcement
VU University Amsterdam, The Netherlands

SAGEO 2012 Liège
Workshop 3: Crime mapping and modeling
Outline

- Motivation
- Methodology
- Results
- Conclusions
Impact evaluation of community policing in 5 Swiss cities: Basel, Bern, Geneva, Lausanne, and Zurich

Problem: Research carried out after implementation begun
⇒ Randomized trial or quasi-experiment not possible

Challenge of observational studies: Establish the counterfactual
⇒ Lower internal validity
Impact evaluation of community policing in 5 Swiss cities: Basel, Bern, Geneva, Lausanne, and Zurich

Problem: Research carried out after implementation begun
⇒ Randomized trial or quasi-experiment not possible

Challenge of observational studies: Establish the counterfactual
⇒ Lower internal validity
Impact evaluation of community policing in 5 Swiss cities: Basel, Bern, Geneva, Lausanne, and Zurich

Problem: Research carried out after implementation begun
⇒ Randomized trial or quasi-experiment not possible

Challenge of observational studies: Establish the counterfactual
⇒ Lower internal validity
Motivation

Burglary Crime Rates 2002-2005

Geneva

Moran's I = 0.19; p = 0.02; n = 26

Lausanne

Moran's I = 0.07; p = 0.19; n = 17

Bern

Moran's I = 0; p = 0.19; n = 13

Zurich

Moran's I = -0.07; p = 0.4; n = 12
Fear of Crime

How safe do you feel walking alone in your neighborhood after 10 pm?
1) very safe, 2) quite safe, 3) somewhat unsafe, 4) very unsafe or
5) don’t go out after 10 pm

Moran's I = -0.39 ;   p = 0.97 ;   n = 11

Basel 1987

Moran's I = 0.02 ;   p = 0.22 ;   n = 11

Basel 2000
Approach: Use spatio-temporal data mining to identify similar neighborhoods across cities

- Minimize within-cluster variance in contextual variables
- Account for significant share of between-cluster variance in outcome variables

⇒ Study impact of community policing across matched neighborhoods
Research Strategy

Approach: Use spatio-temporal data mining to identify *similar* neighborhoods across cities

- Minimize within-cluster variance in contextual variables
- Account for significant share of between-cluster variance in outcome variables

⇒ Study impact of community policing across matched neighborhoods
Research Strategy

Approach: Use spatio-temporal data mining to identify similar neighborhoods across cities

- Minimize within-cluster variance in contextual variables
- Account for significant share of between-cluster variance in outcome variables

⇒ Study impact of community policing across matched neighborhoods
Research Strategy

Approach: Use spatio-temporal data mining to identify similar neighborhoods across cities

- Minimize within-cluster variance in contextual variables
- Account for significant share of between-cluster variance in outcome variables

⇒ Study impact of community policing across matched neighborhoods
Neighborhood ecological context:

- 51 neighborhoods in 4 Swiss cities
- 89 variables (crime, demography, SES, built environment)

Outcome Indicators – Swiss Crime Survey:

- 6 items (fear of crime, disorder, satisfaction with police)

Clustering algorithm to identify similar neighborhoods:

- Self-organizing maps (Kohonen 1990; 2001): non-linear dimensionality reduction
- Random forests (Breiman 2001): variable selection
Data and Methodology

Neighborhood ecological context:

- 51 neighborhoods in 4 Swiss cities
- 89 variables (crime, demography, SES, built environment)

Outcome Indicators – Swiss Crime Survey:

- 6 items (fear of crime, disorder, satisfaction with police)

Clustering algorithm to identify similar neighborhoods:

- Self-organizing maps (Kohonen 1990; 2001): non-linear dimensionality reduction
- Random forests (Breiman 2001): variable selection
Data and Methodology

Neighborhood ecological context:

- 51 neighborhoods in 4 Swiss cities
- 89 variables (crime, demography, SES, built environment)

Outcome Indicators – Swiss Crime Survey:

- 6 items (fear of crime, disorder, satisfaction with police)

Clustering algorithm to identify similar neighborhoods:

- Self-organizing maps (Kohonen 1990; 2001): non-linear dimensionality reduction
- Random forests (Breiman 2001): variable selection
Data and Methodology

Neighborhood ecological context:

- 51 neighborhoods in 4 Swiss cities
- 89 variables (crime, demography, SES, built environment)

Outcome Indicators – Swiss Crime Survey:

- 6 items (fear of crime, disorder, satisfaction with police)

Clustering algorithm to identify similar neighborhoods:

- Self-organizing maps (Kohonen 1990; 2001): non-linear dimensionality reduction
- Random forests (Breiman 2001): variable selection
Data and Methodology

Neighborhood ecological context:

- 51 neighborhoods in 4 Swiss cities
- 89 variables (crime, demography, SES, built environment)

Outcome Indicators – Swiss Crime Survey:

- 6 items (fear of crime, disorder, satisfaction with police)

Clustering algorithm to identify similar neighborhoods:

- Self-organizing maps (Kohonen 1990; 2001): non-linear dimensionality reduction
- Random forests (Breiman 2001): variable selection
Methodology

Data and Methodology

Neighborhood ecological context:

- 51 neighborhoods in 4 Swiss cities
- 89 variables (crime, demography, SES, built environment)

Outcome Indicators – Swiss Crime Survey:

- 6 items (fear of crime, disorder, satisfaction with police)

Clustering algorithm to identify similar neighborhoods:

- **Self-organizing maps (Kohonen 1990; 2001): non-linear dimensionality reduction**
- **Random forests (Breiman 2001): variable selection**
Data and Methodology

Neighborhood ecological context:
- 51 neighborhoods in 4 Swiss cities
- 89 variables (crime, demography, SES, built environment)

Outcome Indicators – Swiss Crime Survey:
- 6 items (fear of crime, disorder, satisfaction with police)

Clustering algorithm to identify similar neighborhoods:
- Self-organizing maps (Kohonen 1990; 2001): non-linear dimensionality reduction
- Random forests (Breiman 2001): variable selection
Self-Organizing Maps SOM

% Single households 1990 vs % Housing mixed
Self-Organizing Maps (SOM)

Input Space (high-dimensional)

% Housing mixed vs. % Single households 1990

Output Space (2-D)

Lattice

Christian Kreis (NSCR)  Observational Studies of Community Policing  SAGEO Nov 6, 2012

8 / 16
Self-Organizing Maps SOM

Input Space (high-dimensional)

Output Space (2-D)

Lattice

% Single households 1990

% Housing mixed

Training
Self-Organizing Maps SOM

Input Space (high-dimensional)
- **Training**
  - % Single households 1990 vs % Housing mixed

Output Space (2-D)
- **Lattice**
  - Clustering of the “trained” map

% Single households 1990 vs % Housing mixed
Self-Organizing Maps (SOM)

**Input Space (high-dimensional)**

- **Training**
- **% Housing mixed** vs. **% Single households 1990**

**Output Space (2-D)**

- **Lattice**

**Clustering of the “trained” map**

- **% Housing mixed** vs. **% Single households 1990**

Christian Kreis (NSCR) | Observational Studies of Community Policing | SAGEO Nov 6, 2012
Self-Organizing Maps SOM

Input Space (high-dimensional)

Output Space (2-D)

Clustering of the “trained” map

Visualization
Methodology

SOM Training and Clustering

(a) Training Progress

(b) Hits Map

(c) U–matrix

(d) Clustering Dendrogram
Methodology

SOM Training and Clustering

(a) Training Progress
(b) Hits Map
(c) U–matrix
(d) Clustering Dendrogram

Christian Kreis (NSCR)
Observational Studies of Community Policing
SAGEO Nov 6, 2012
Methodology

SOM Training and Clustering

(a) Training Progress
(b) Hits Map
(c) U–matrix
(d) Clustering Dendrogram

Christian Kreis  (NSCR)  Observational Studies of Community Policing  SAGEO Nov 6, 2012  9 / 16
Methodology

SOM Training and Clustering

(a) Training Progress

(b) Hits Map

(c) U–matrix

(d) Clustering Dendrogram

(e) Optimum Number of Clusters
Methodology

SOM Training and Clustering

(a) Training Progress

(b) Hits Map

(c) U–matrix

(d) Clustering Dendrogram

(e) Optimum Number of Clusters
Methodology

SOM Training and Clustering

Training Progress

![Training Progress Graph](image)

Hits Map

![Hits Map](image)

U–matrix

![U–matrix](image)

Clustering Dendrogram

![Clustering Dendrogram](image)

Optimum Number of Clusters

![Optimum Number of Clusters](image)

Christian Kreis (NSCR) Observational Studies of Community Policing SAGEO Nov 6, 2012 9 / 16
Methodology

SOM Best Matching Units

Christian Kreis (NSCR)  Observational Studies of Community Policing  SAGEO Nov 6, 2012
Results

SOM – Visualization in Geographic Space

Geneva

Lausanne

Basel

Bern

Zurich

Clustering of the SOM

Christian Kreis (NSCR) | Observational Studies of Community Policing | SAGEO Nov 6, 2012
### Characteristics of Neighborhood Clusters

<table>
<thead>
<tr>
<th>Neighborhood Type</th>
<th>Non-housing</th>
<th>Residential</th>
<th>Same Address 5–yrs</th>
<th>Single households 00</th>
<th>Single households 90</th>
<th>University 00</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Neighborhood Type</th>
<th>Foreigners 00</th>
<th>Foreigners 90</th>
<th>Housing mixed</th>
<th>Mandatory school 00</th>
<th>Mandatory school 90</th>
<th>Middle management</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Neighborhood Type</th>
<th>Apprenticeship 90</th>
<th>before 1900</th>
<th>Crime PC1</th>
<th>Crime PC2</th>
<th>Families 00</th>
<th>Families 90</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Neighborhood Type</th>
<th>05–09 years 90</th>
<th>10–14 years 90</th>
<th>2 stories</th>
<th>6 stories</th>
<th>7–9 stories</th>
<th>Apprenticeship 00</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Christian Kreis (NSCR)  
Observational Studies of Community Policing  
SAGEO Nov 6, 2012
Analysis of Survey Data by Neighborhood Cluster

**Fear of Crime**

How safe do you feel walking alone in your neighborhood after 10 pm?

<table>
<thead>
<tr>
<th></th>
<th>Very safe</th>
<th>Quite safe</th>
<th>Rather unsafe</th>
<th>Very unsafe</th>
<th>Don’t go out</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n = 193  n = 281  n = 87  n = 266  n = 30

X−squared = 80.5     p−value = 0

**Risk of Victimization**

How likely do you rate a burglary of your home over the next 12 months?

<table>
<thead>
<tr>
<th></th>
<th>Very unlikely</th>
<th>Rather unlikely</th>
<th>Rather likely</th>
<th>Very likely</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n = 186  n = 242  n = 78  n = 260  n = 27

X−squared = 31.5     p−value = 0.007

**Behavioral Response**

Walking alone in your neighborhood after 10 pm, do you stay away from certain streets, areas, or people to avoid crime?

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
<th>Don’t go out</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n = 215  n = 188  n = 253  n = 269  n = 30

X−squared = 60.9     p−value = 0

**Physical Disorder**

Close to your home are there any graffiti on the walls or lots of rubbish lying around on the streets?

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n = 223  n = 213  n = 266  n = 87  n = 281  n = 30

X−squared = 22.5     p−value = 0

**Social Disorder**

Close to your home are there disreputable people loitering on the streets?

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n = 223  n = 213  n = 266  n = 87  n = 281  n = 30

X−squared = 36.9     p−value = 0

**Satisfaction with Police**

How good do you think the police in your area are in controlling crime?

<table>
<thead>
<tr>
<th></th>
<th>Very good</th>
<th>Quite good</th>
<th>Not so good</th>
<th>Not good at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n = 189  n = 172  n = 228  n = 71  n = 251  n = 24

X−squared = 25.3     p−value = 0.047
Outcome Indicators by Neighborhood Cluster

Behavorial Response 2000

Police Effectiveness 2000
Conclusions

The resulting neighborhood typology

- shows similar patterns from center to periphery across cities
- reduces within-cluster variance in contextual variables
- accounts for a significant share of the between-cluster variance of the outcome indicators

Research design:

⇒ Evaluate community policing impact by neighborhood type across cities
⇒ Data mining enhances the internal validity of an observational study
Conclusions

The resulting neighborhood typology

- shows similar patterns from center to periphery across cities
- reduces within-cluster variance in contextual variables
- accounts for a significant share of the between-cluster variance of the outcome indicators

Research design:

⇒ Evaluate community policing impact by neighborhood type across cities
⇒ Data mining enhances the internal validity of an observational study
Conclusions

The resulting neighborhood typology

- shows similar patterns from center to periphery across cities
- reduces within-cluster variance in contextual variables
- accounts for a significant share of the between-cluster variance of the outcome indicators

Research design:

⇒ Evaluate community policing impact by neighborhood type across cities
⇒ Data mining enhances the internal validity of an observational study
Conclusions

The resulting neighborhood typology

- shows similar patterns from center to periphery across cities
- reduces within-cluster variance in contextual variables
- accounts for a significant share of the between-cluster variance of the outcome indicators

Research design:

⇒ Evaluate community policing impact by neighborhood type across cities
⇒ Data mining enhances the internal validity of an observational study
Conclusions

The resulting neighborhood typology

- shows similar patterns from center to periphery across cities
- reduces within-cluster variance in contextual variables
- accounts for a significant share of the between-cluster variance of the outcome indicators

Research design:

⇒ Evaluate community policing impact by neighborhood type across cities
⇒ Data mining enhances the internal validity of an observational study
Conclusions

The resulting neighborhood typology

- shows similar patterns from center to periphery across cities
- reduces within-cluster variance in contextual variables
- accounts for a significant share of the between-cluster variance of the outcome indicators

Research design:

⇒ Evaluate community policing impact by neighborhood type across cities
⇒ Data mining enhances the internal validity of an observational study
Conclusions

The resulting neighborhood typology

- shows similar patterns from center to periphery across cities
- reduces within-cluster variance in contextual variables
- accounts for a significant share of the between-cluster variance of the outcome indicators

Research design:

⇒ Evaluate community policing impact by neighborhood type across cities
⇒ Data mining enhances the internal validity of an observational study
Thank you for your attention!

ckreis@nsr.nl

Acknowledgements
This research was supported by funding from the Swiss National Science Foundation (Award No. 100017-132675).