Crime, Cities and Data

Nuria Oliver, PhD
Chief Data Scientist @ DataPop Alliance
Director of Research in Data Science @ Vodafone

Work with Bogomolov, A., Lepri, B., Staiano, J., Pianesi, F., Pentland, A. De Nadai, M., Clavijo A., Lara Molina R., Letouzé E., Pestre G., Serra J., Shoup N., Ramirez A.

FONDAZIONE BRUNO KESSLER

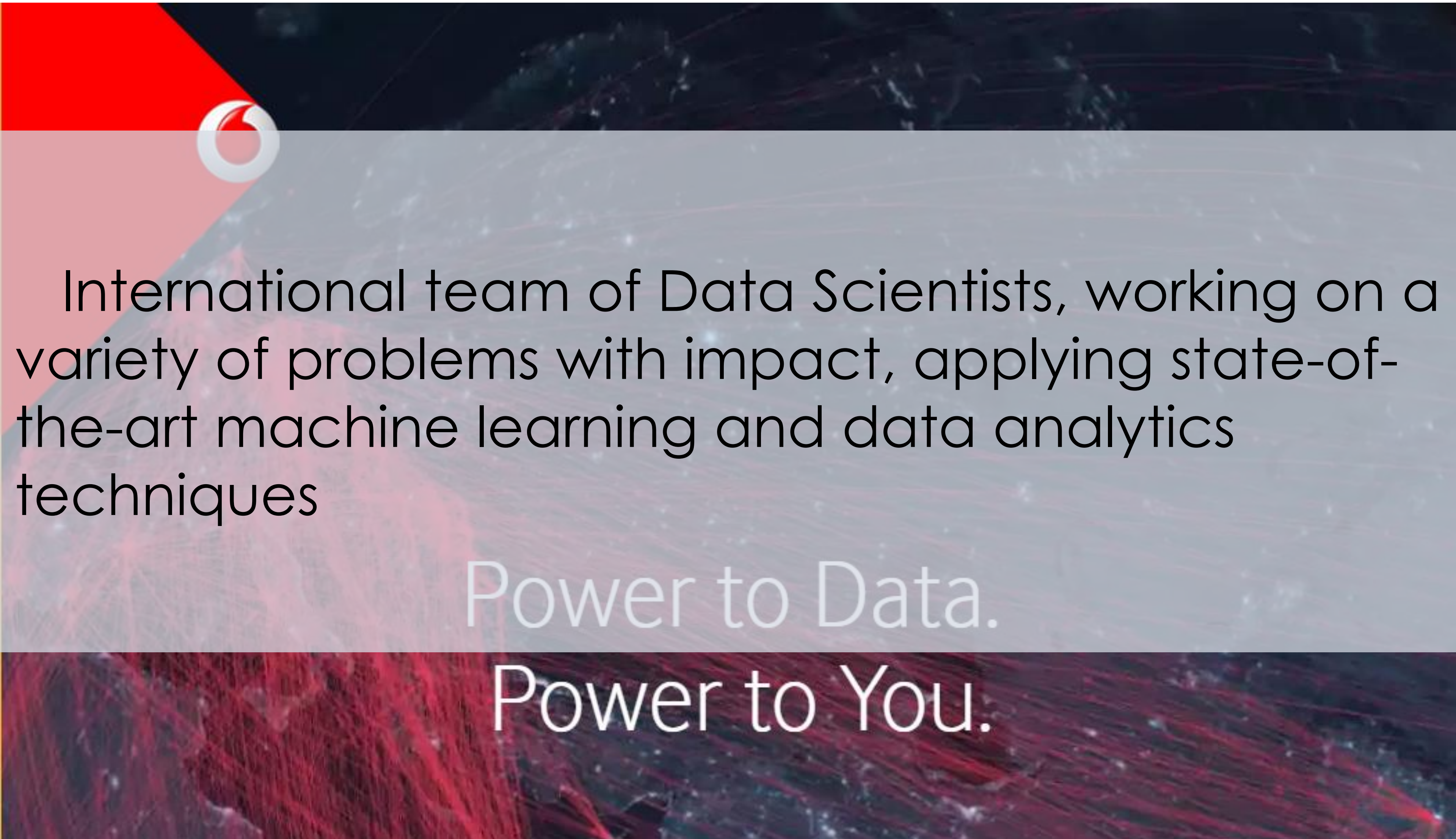
DATA POP ALLIANCE

Telefonica

MIT MEDIA LAB

Vodafone Big Data and Advanced Analytics

International team of Data Scientists, working on a variety of problems with impact, applying state-of-the-art machine learning and data analytics techniques



Data-Pop Alliance is a global coalition on **Big Data & development** created by the **Harvard Humanitarian Initiative, MIT Media Lab**, and **Overseas Development Institute** joined by **Flowminder,** bringing together researchers, experts, practitioners and activists to "**promote a people-centered Big Data revolution**" by **locally co-designing and deploying collaborative research, training,** and **engagement activities**

CYBERSECURITY

Saving Big Data from Itself

A three-step plan for using data right in an age of government overreach

By Alex "Sandy" Pentland

For the first few decades of its existence, the National Security Agency was a quiet department with one primary job: keeping an eye on the Soviet Union. Its enemy was well defined and monolithic. Its principal tools were phone taps, spy planes and hidden microphones.

After the attacks of September 11, all of that changed. The NSA's chief enemy became a diffuse network of individual terrorists. Anyone in the world could be a legitimate target for spying. The nature of spying itself changed as new digital communication channels proliferated. The exponential growth of Internet-connected mobile devices was just beginning. The NSA's old tools apparently no longer seemed sufficient.

In response, the agency adopted a new strategy: collect everything. As former NSA director Keith Alexander once put it, when you are looking for a needle in a haystack, you need the whole haystack. The NSA began collecting bulk phone call records from virtually every person in the U.S.; soon it was gathering data on bulk Internet traffic from virtually everyone outside of the U.S. Before long, the NSA was collecting an amount of data every two hours equivalent to the U.S. Census.

IN BRIEF

Data about human behavior have always been essential for both government and industry to function. But how do we enable institutions to collect and analyze data without abusing that information?

Leadership

Prof. Alex 'Sandy' Pentland
Academic Director

Dr Nuria Oliver
Chief Data Scientist

Prof. Patrick Vinck
Co-Director &
Co-Founder

HARVARD
HUMANITARIAN
INITIATIVE

Prof. Phuong Pham
Co-Director for HHI

Dr Emmanuel Letouzé
Director & Co-Founder

Dr Linus Bengtsson
FLOWMINDER.ORG

Elizabeth Stuart
Co-Director for
ODI

Dr Emma Samman

Vodafone
Institute
for Society and
Communications

Vodafone's Think and Do-Tank

Mission:

The Institute is Vodafone's European **think and do tank**.
It has the task to explore the **potential** and **responsible use of digital technologies** for **innovation**, **growth** and **sustainable social** impact.

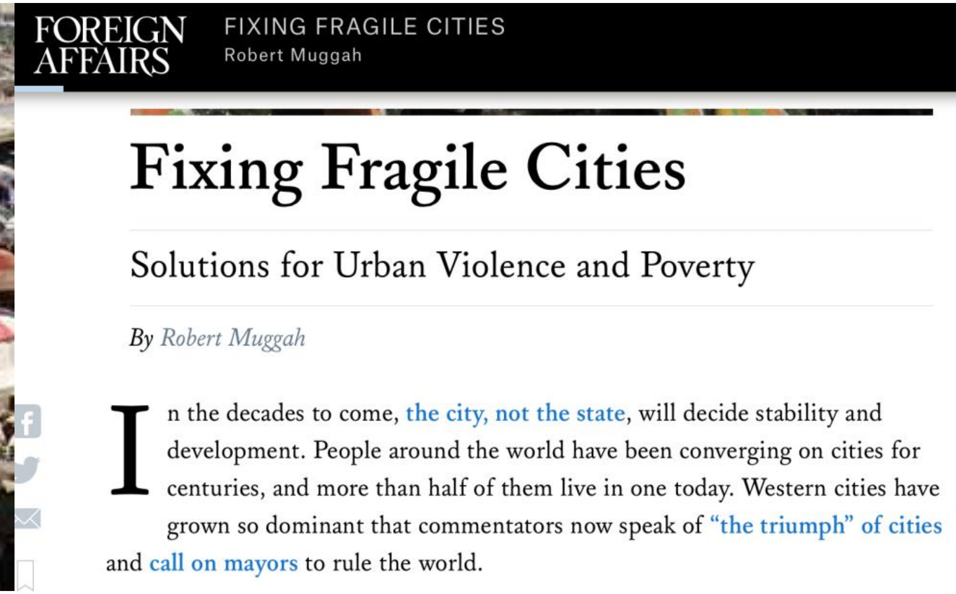
Crime, Cities and Data

Nuria Oliver, PhD
Chief Data Scientist @ DataPop Alliance
Director of Research in Data Science @ Vodafone

Work with Bogomolov, A., Lepri, B., Staiano, J., Pianesi, F., Pentland, A. De Nadai, M., Clavijo A., Lara Molina R., Letouzé E., Pestre G., Serra J., Shoup N., Ramirez A.

FONDAZIONE BRUNO KESSLER
DATA POP ALLIANCE
Telefonica
MIT MEDIA LAB

Context and Motivation



FOREIGN AFFAIRS FIXING FRAGILE CITIES Robert Muggah

Fixing Fragile Cities

Solutions for Urban Violence and Poverty

By Robert Muggah

In the decades to come, the city, not the state, will decide stability and development. People around the world have been converging on cities for centuries, and more than half of them live in one today. Western cities have grown so dominant that commentators now speak of "the triumph" of cities and call on mayors to rule the world.

*"**Managing urban areas** has become one of the most **important development challenges** of the 21st century. Our success or failure in building **sustainable cities** will be a major factor in the success of the post-2015 UN development agenda",* ***John Wilmoth, Director of UN DESA's Population Division***

Crime

- ✓ Affects quality of life and economic development both at the national and local level
- ✓ Many studies have explored the relationships between crime and socio-economic variables: education, income, unemployment, ethnicity, …
- ✓ Several studies have shown significant concentrations of crime in small geographical areas: ***crime hotspots***

Two key factors that affect crime

Routine Activity

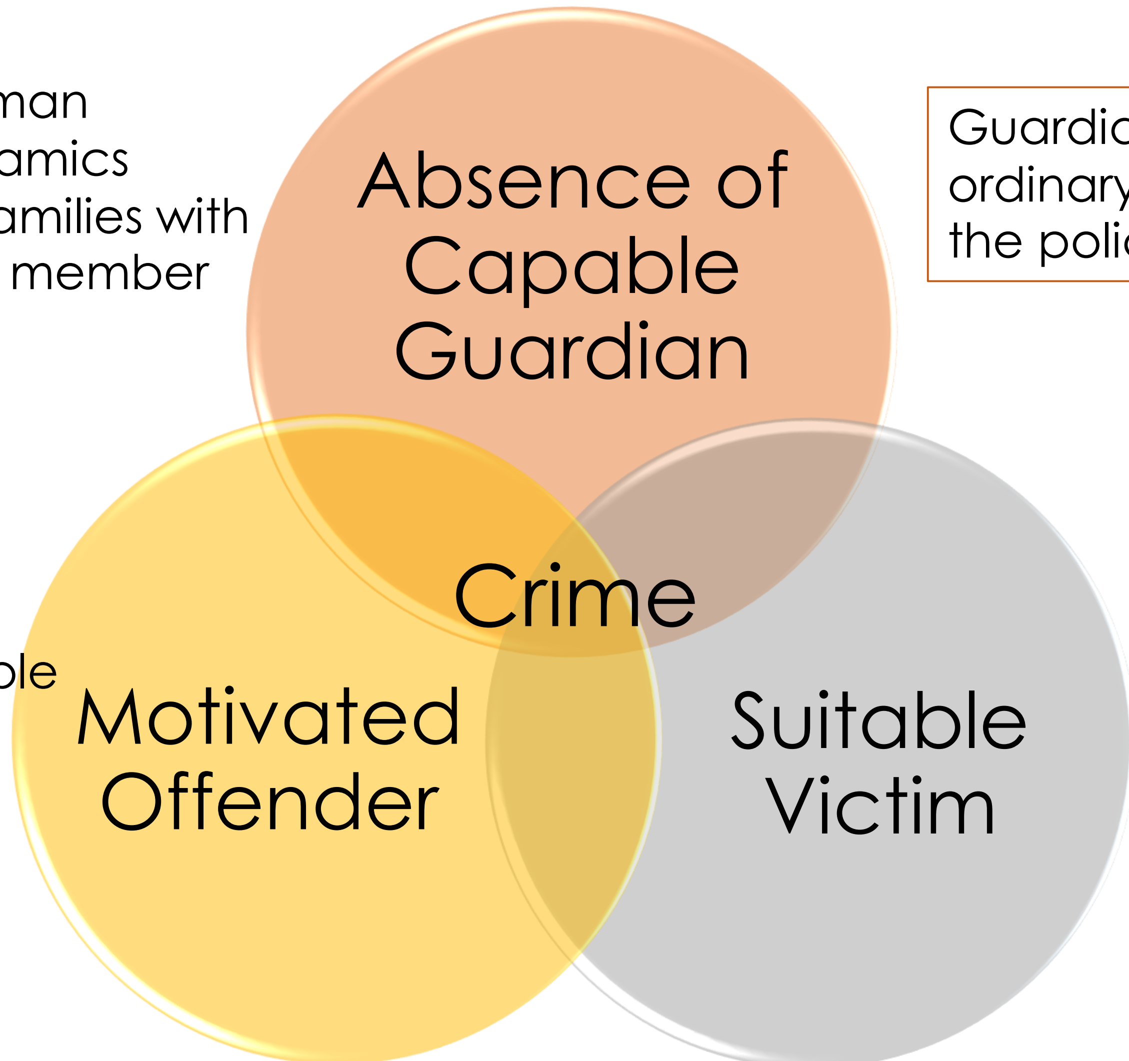
Social Disorganization

Source: Felson, M (1994) *Crime and everyday life: Insight and Implications for Society*, Thousand Oaks, CA

Routine Activity

-Human
dynamics
-% families with
one member

Guardianship by both ordinary citizens and the police



-#young people
-economic
deprivation

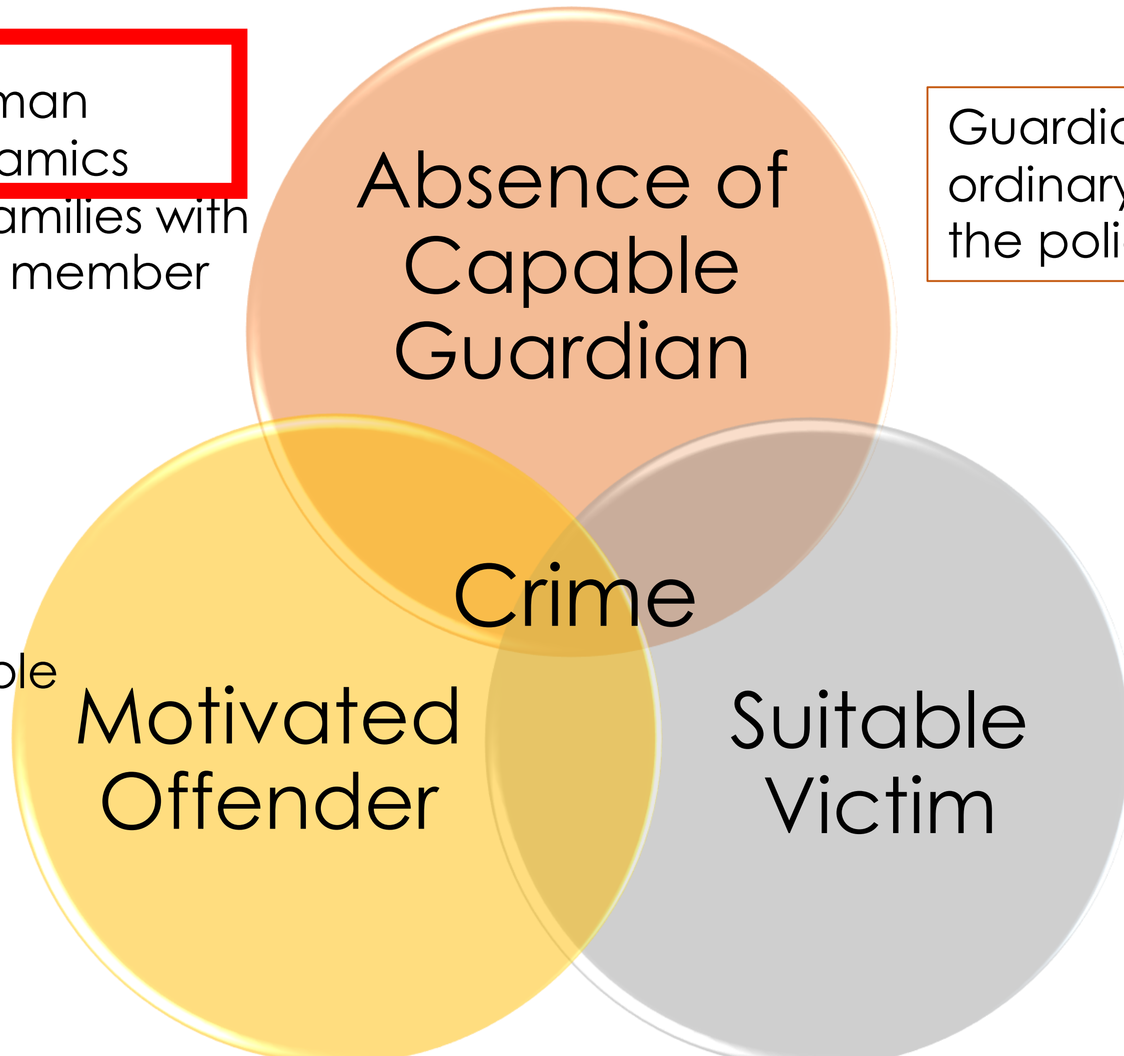
-Population
-# dwellings
-Avg family
income
-Avg dwelling
value

Source: Felson, M (1994) *Crime and everyday life: Insight and Implications for Society*, Thousand Oaks, CA

Routine Activity

-Human dynamics
-% families with one member

Guardianship by both ordinary citizens and the police



-#young people
-economic deprivation

-Population
-# dwellings
-Avg family income
-Avg dwelling value

Source: Felson, M (1994) *Crime and everyday life: Insight and Implications for Society*, Thousand Oaks, CA

Modeling Routine Activities from Mobile Data

6.8 billion subscribers
96% of world's population (ITU)

Mobile penetration of 120% to 89% of population (ITU

More time spent on our phones than watching TV or
our partner (US and UK)

Emerging and developed regions

6.8 billion subscribers
96% of world’s population (ITU)

Mobile penetration of 120% to 89% of population (ITU

More time spent on our phones than watching TV or our partner (US and UK)

Emerging and developed regions

Cell Phones as Sensors of Human Activity

Digital footprints enable large-scale analysis of human behavior

Bits

Business ▪ Innovation ▪ Technology ▪ Society

May 19, 2011, *7:06 pm* **The Sensors Are Coming!**

By NICK BILTON

ieee spectrum INSIDE TECHNOLOGY

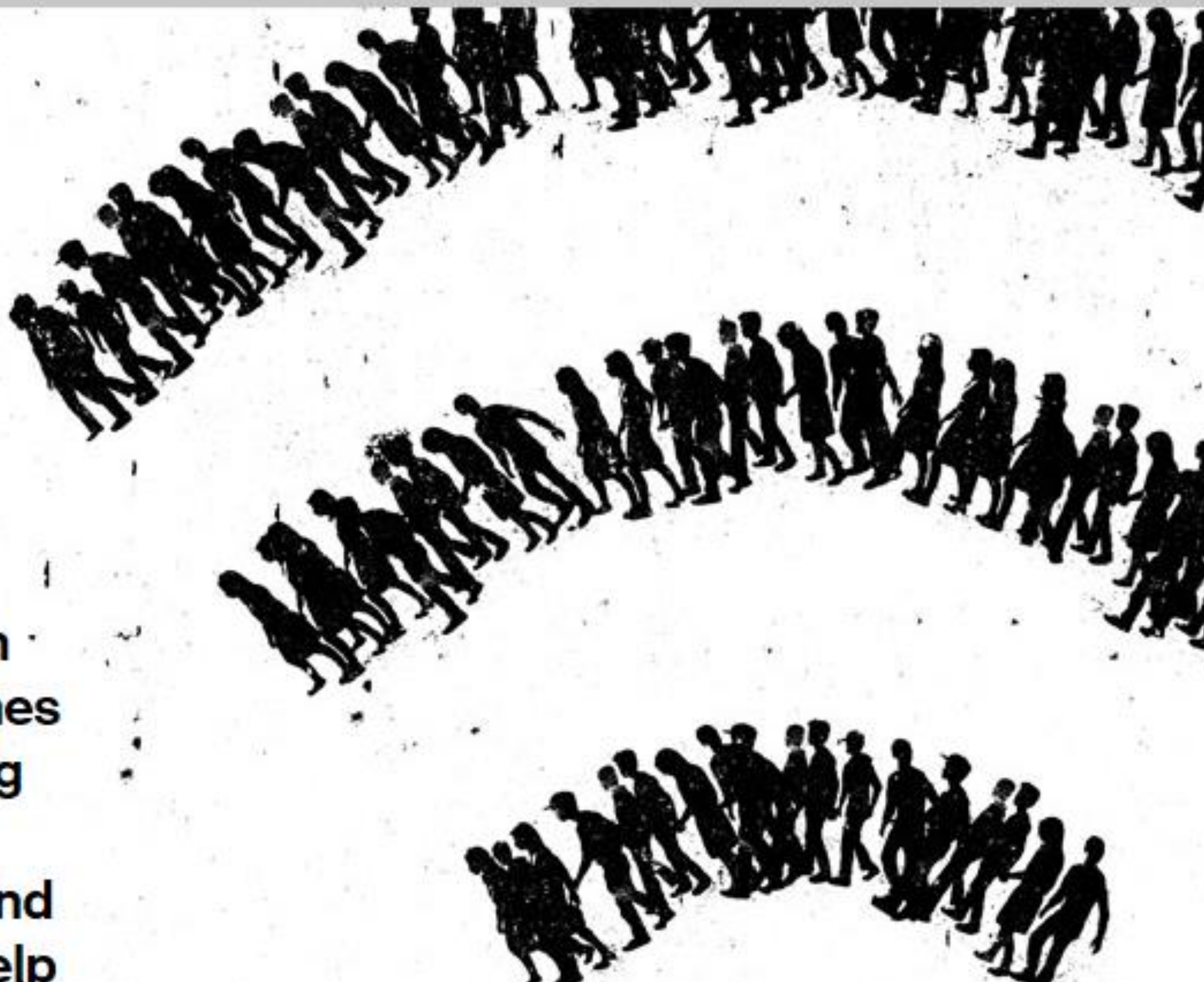
Telecom / Wireless

NEWS

Cellphones for Science

Scientists want to put sensors into everyone's hands

Big Data from Cheap Phones

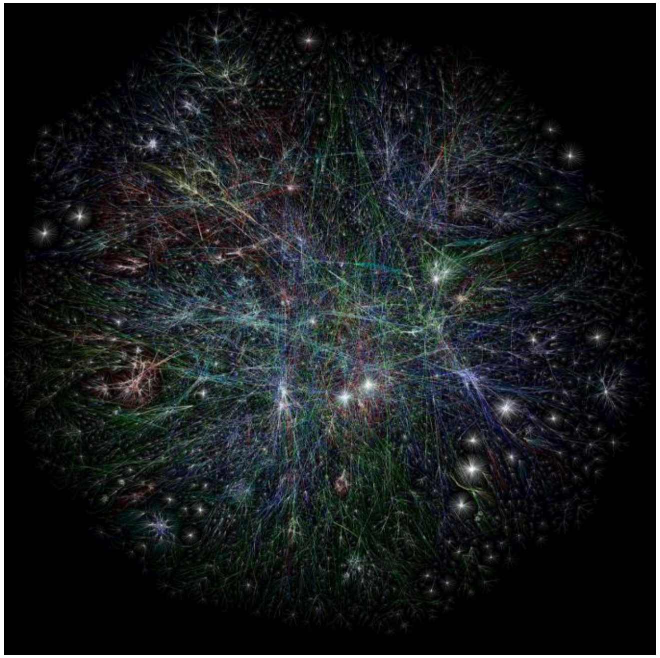


Collecting and analyzing information from simple cell phones can provide surprising insights into how people move about and behave—and even help

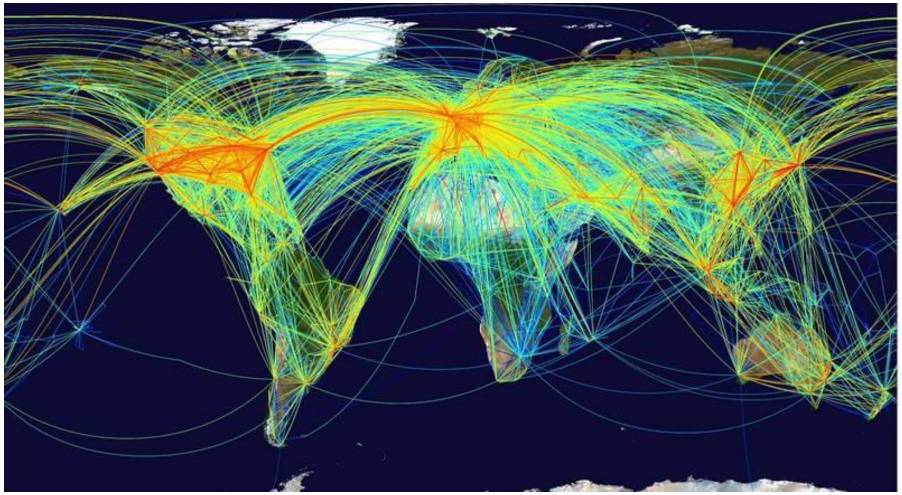
Computational Social Sciences

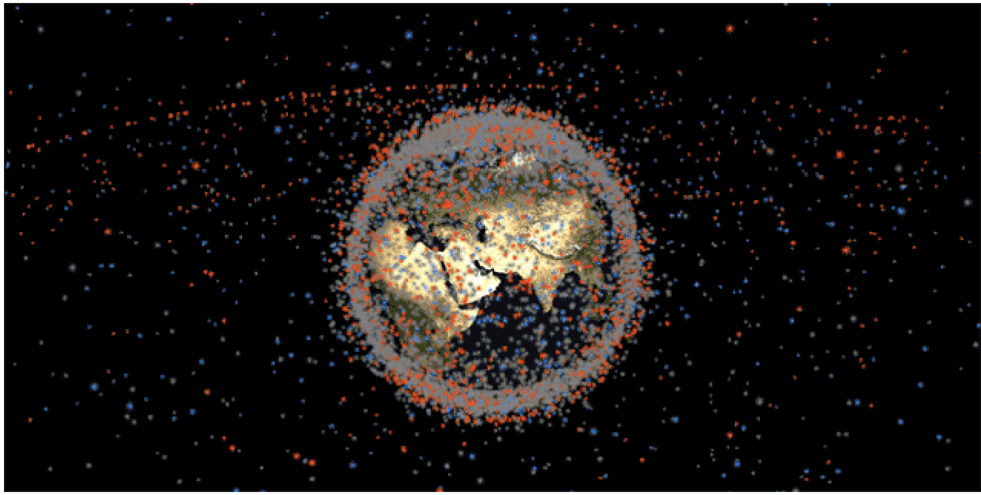
The ubiquity of mobile phones enables us to collect and analyze, for the first time in human history, **large-scale aggregated** and anonymized **human behavioral data** of entire cities, countries or even continents

The opportunity is HUGE to help decision making units (governments, UN, Red Cross…) make more informed decisions thanks to the existence of quantitative real-time information about populations



Source: Kapersky Lab





Source: Kapersky Lab

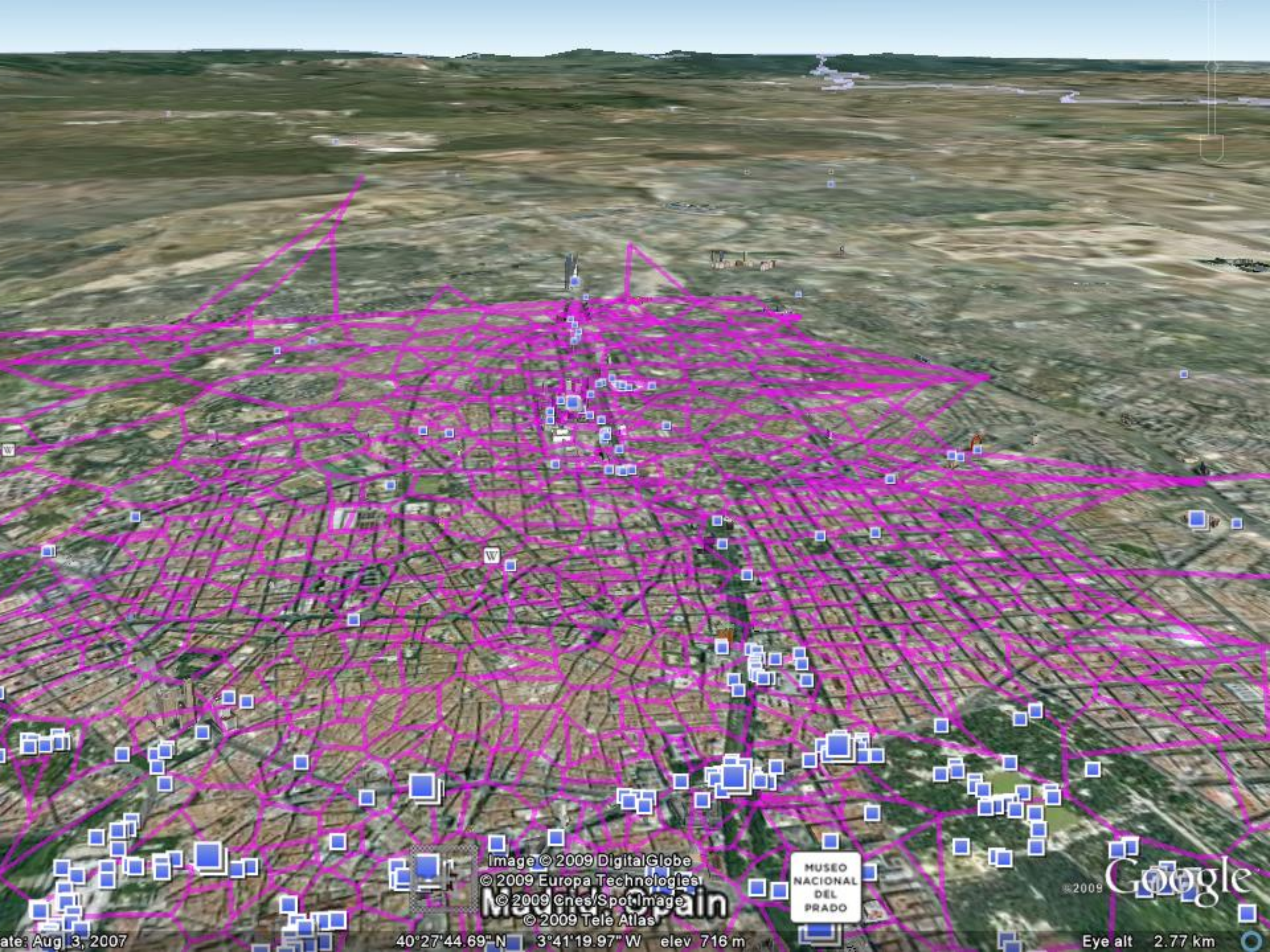


Image © 2009 DigitalGlobe
© 2009 Europa Technologies
© 2009 Cnes/Spot Image
© 2009 Tele Atlas
Madrid, Spain
MUSEO NACIONAL DEL PRADO
©2009 Google
ate: Aug 3, 2007
40°27'44.69" N
3°41'19.97" W
elev 716 m
Eye alt 2.77 km

Typical Mobile Data

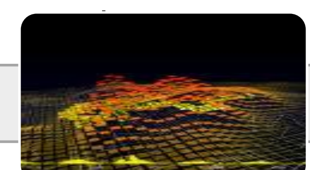
CDR (voice)

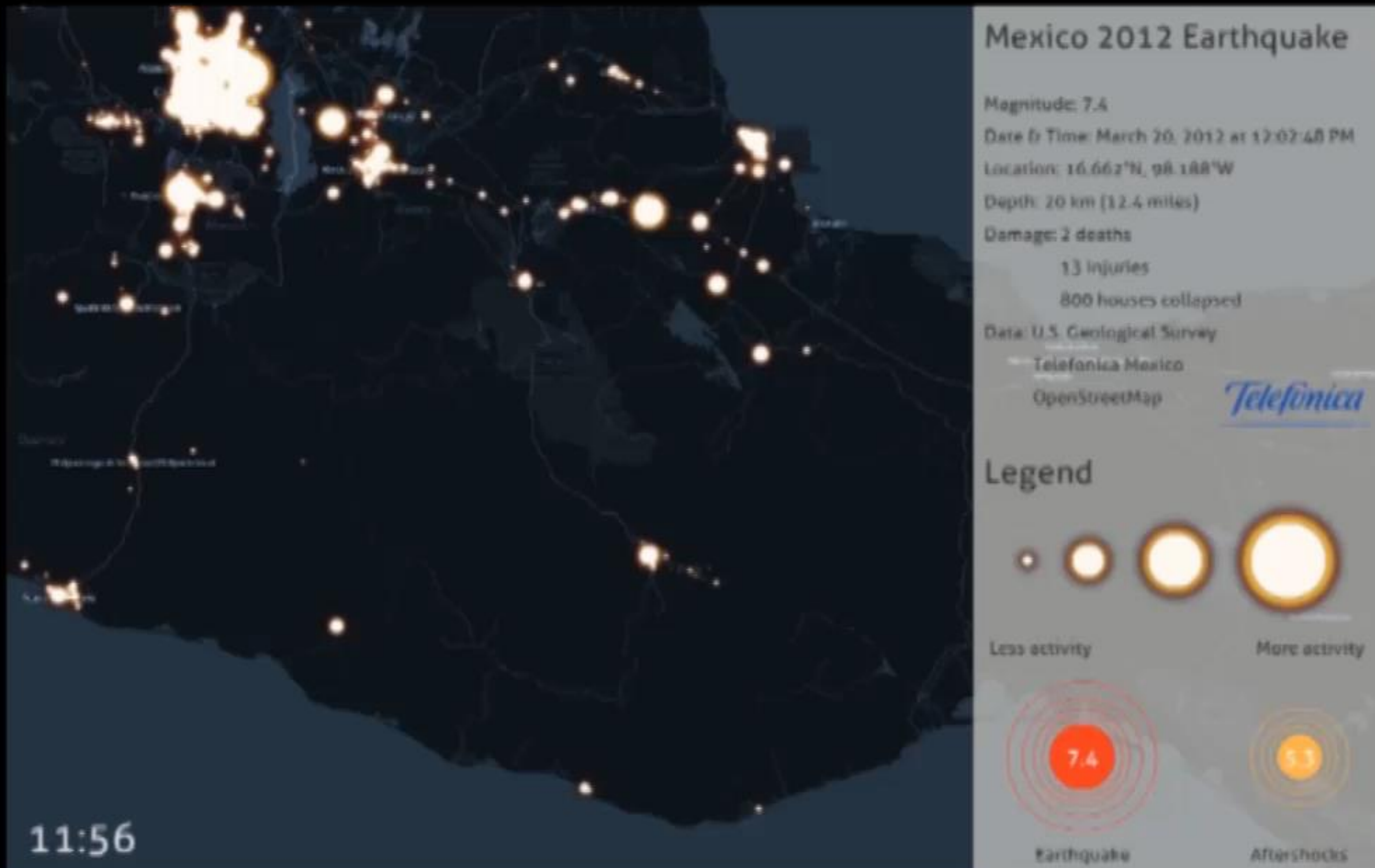
HR_ORG	TLFN_A	TLFN_B	CD_GEO_A	CD_GEO_B	DT_ORG	CD_SNTD	CD_ERB	CD_CCC	QT_DUR
20:05:31	XXX	YYY	3	11	20140519	2	1562	568	33
…	…	…	…	…	…	…	…	…	…

CDR (SMS)

HR_ORG	TLFN_A	TLFN_B	CD_GEO_A	CD_GEO_B	DT_ORG	CD_SNTD	QT_TRFG
15:53:54	XXX	ZZZ	3	25	20140506	2	1
…	…	…	…	…	…	…	…

Consumption	**Social Network**	**Mobility**
Call duration	In/Out Degree	Radius of gyration
N. Events	Delta w.r.t time window	Travelled distance
Lapse between events	Unique Calls per day	Rate of popular antennas
Reciprocated events	Unique SMS per day	Regularity of popular



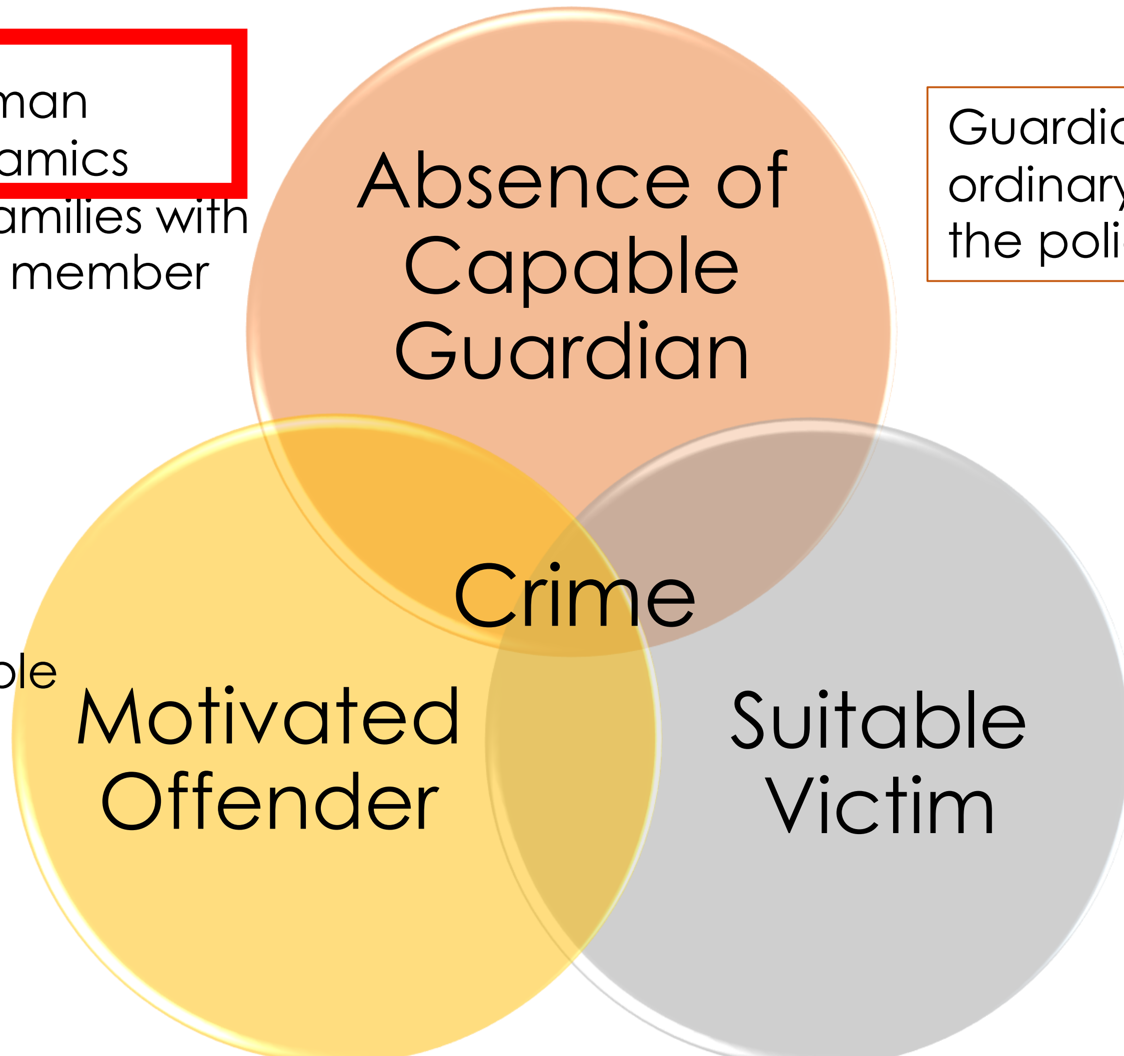

Mexico 2012 Earthquake
Magnitude: 7.4
Date & Time: March 20, 2012 at 12:02:48 PM
Location: 16.662°N, 98.188°W
Depth: 20 km (12.4 miles)
Damage: 2 deaths
13 injuries
800 houses collapsed
Data: U.S. Geological Survey
Telefonica Mexico
OpenStreetMap
Legend
Less activity
More activity
7.4
Earthquake
5.3
Aftershocks
11:56


09:00

Routine Activity

-Human dynamics
-% families with one member

Guardianship by both ordinary citizens and the police



-#young people
-economic deprivation

-Population
-# dwellings
-Avg family income
-Avg dwelling value

Source: Felson, M (1994) *Crime and everyday life: Insight and Implications for Society*, Thousand Oaks, CA

Routine Activity from CDRs

- We can estimate population density
- We can estimate population movements/flows, commuting patterns, OD matrices, …

Two key factors that affect crime

Routine Activity

Social Disorganization

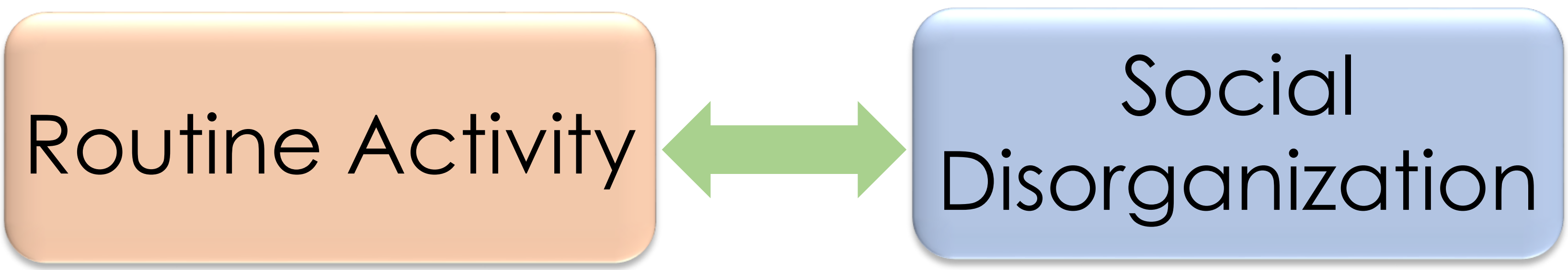
Source: Felson, M (1994) *Crime and everyday life: Insight and Implications for Society*, Thousand Oaks, CA

Social Disorganization

- Residential stability
- Ethnic heterogeneity
- Social deprivation:
  - #single family homes
  - % university degrees
- Economic deprivation:
  - Avg/std income
  - Unemployment

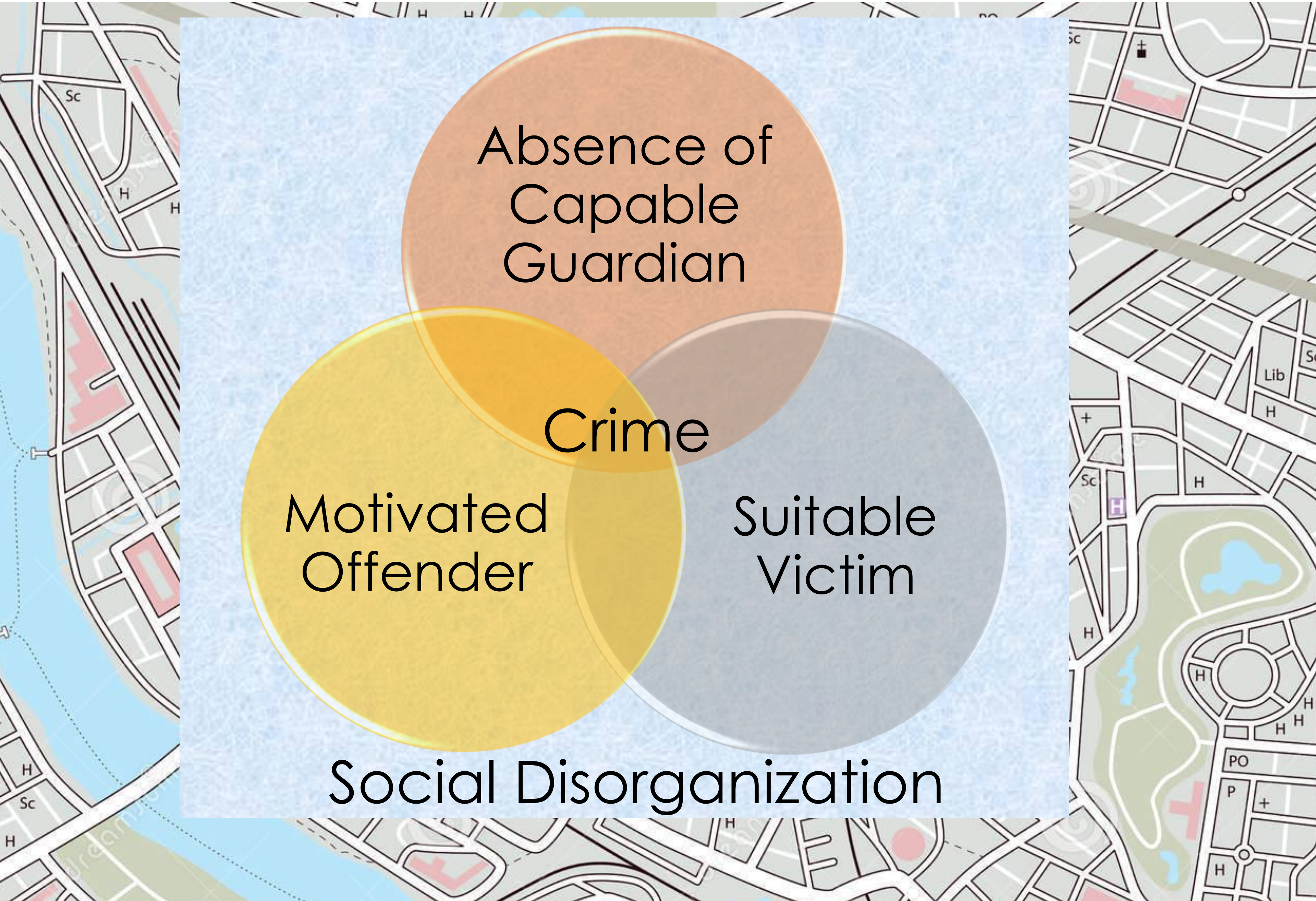
Source: Felson, M (1994) *Crime and everyday life: Insight and Implications for Society*, Thousand Oaks, CA

Two key factors that affect crime



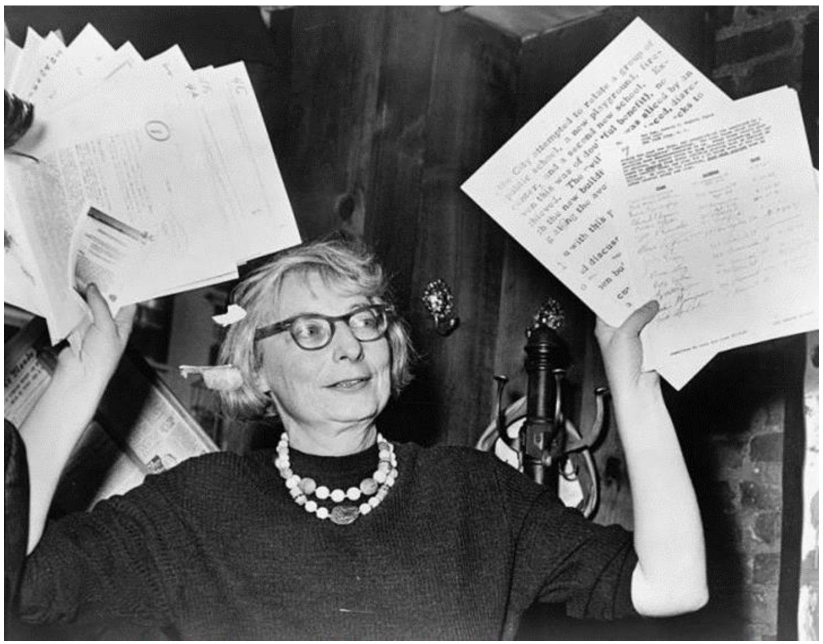
Source: Felson, M (1994) *Crime and everyday life: Insight and Implications for Society*, Thousand Oaks, CA

Crime


Absence of Capable Guardian
Crime
Motivated Offender
Suitable Victim
Social Disorganization

The Theory: Jane Jacobs

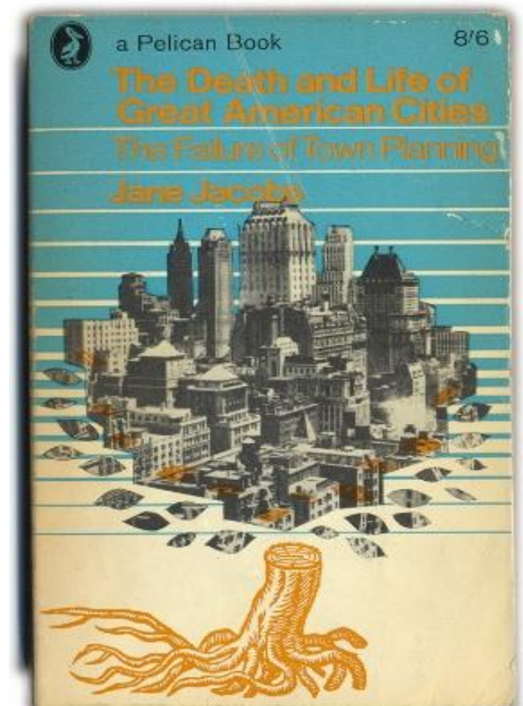
One of the most influential books in city planning



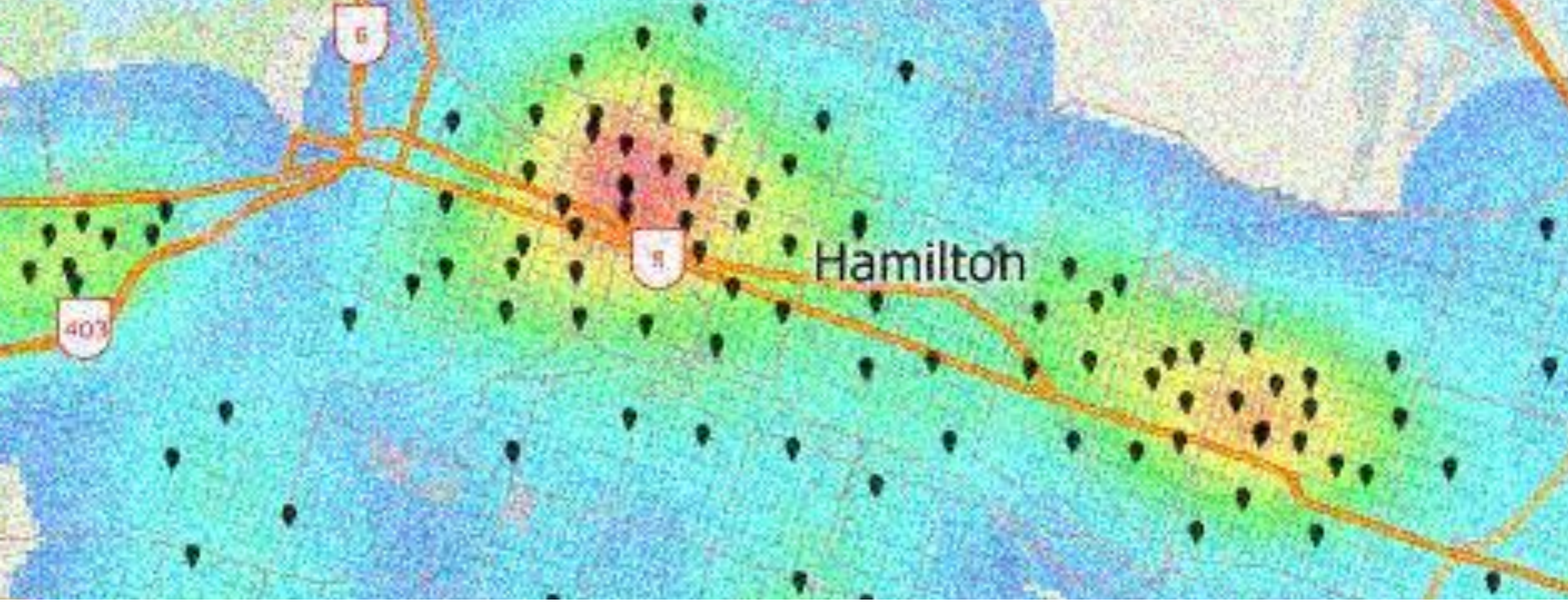
"A well-used city street is apt to be a safe street and a deserted city street is apt to be unsafe"

— Jane Jacobs

(May 4, 1916 – April 25, 2006)



Guardianship by ordinary citizens



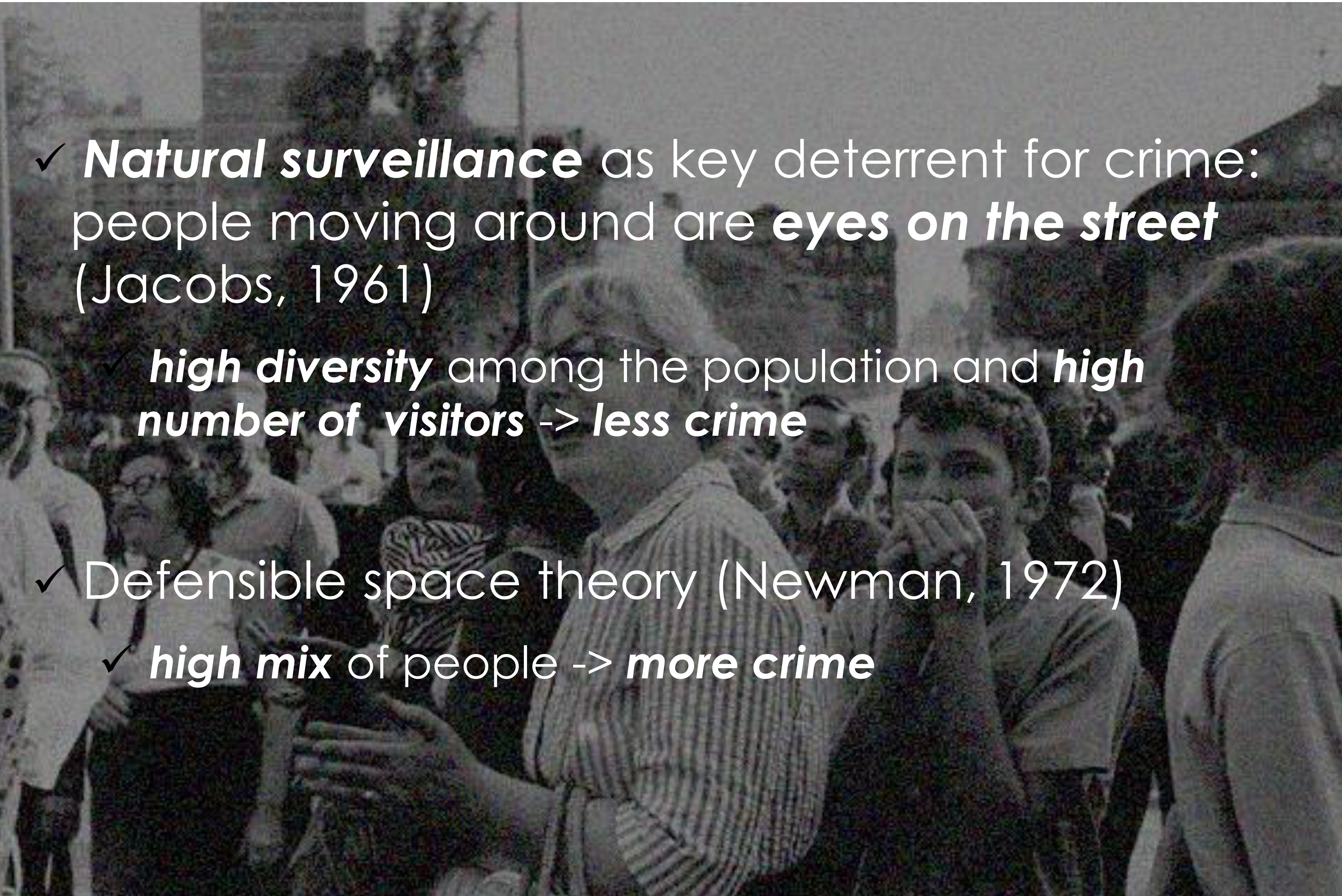
Project 1: Predicting Crime Hotspots in London

Work with Bogomolov, A., Lepri, B., Staiano, J., Pianesi, F., Pentland, A.

FONDAZIONE BRUNO KESSLER

MIT MEDIA LAB

Crime and Urban Environment



- ***Natural surveillance*** as key deterrent for crime: people moving around are ***eyes on the street*** (Jacobs, 1961)
  - ***high diversity*** among the population and ***high number of visitors*** -> ***less crime***
- Defensible space theory (Newman, 1972)
  - ***high mix*** of people -> ***more crime***

Crime Prediction

- ✓ ***People-centric*** perspective vs ***Place-centric*** perspective
  - ✓ people-centric perspective used for individual or collective criminal profiling
  - ✓ place-centric perspective used for predicting crime hotspots

Our Approach

- ✓ ***Data-driven*** and ***place-centric*** approach to crime prediction
- ✓ Multimodal approach: ***people dynamics*** derived from ***mobile network data*** and demographics
- ✓ European metropolis: London
- ✓ Prediction of ***crime hotspots*** and not criminals profiling

Data

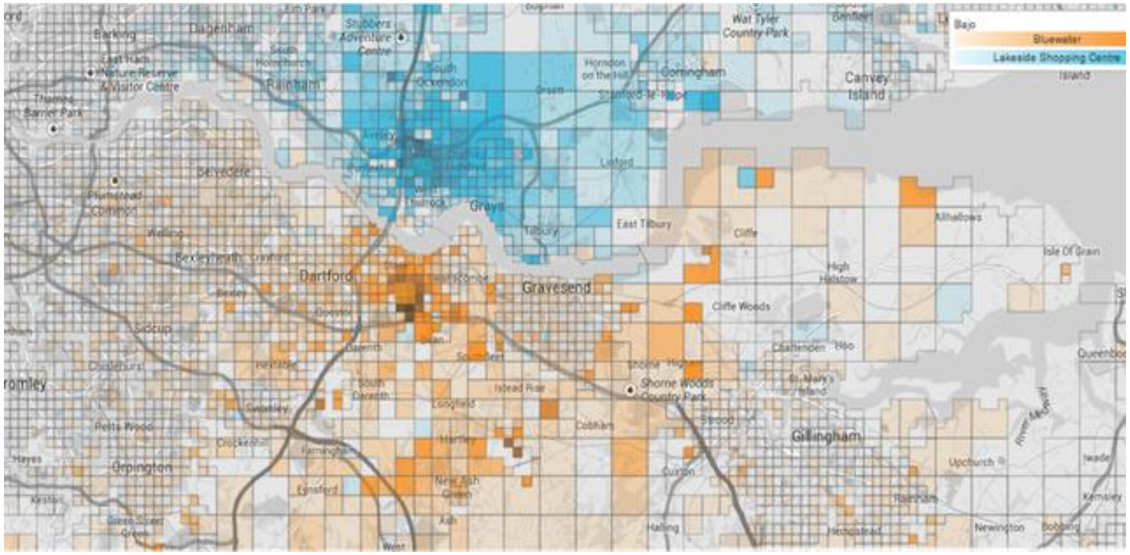
Smartsteps Dataset: for each of the Smartsteps cells a variety of demographic variables were computed every hour for 3 weeks (from December 9 to December 15, 2012 and from December 23, 2012 to January 5, 2013)

Criminal Cases Dataset: criminal cases for December 2012 and for January 2013

London Borough Profiles Dataset: open dataset containing ***68 metrics*** about the population of a particular geographic area

SmartSteps

- **Footfall count**: Shows the trend in footfall in a specified area hourly, daily, weekly and monthly. Provides a basic profile of the crowd.
- **Catchment area**: Shows which postal sectors are your customers coming from by hour, day, week and month. Shows the "battleground" for two sites.
- **Transport mode**: Shows flows of crowds from any two points, segmented by road, air, train, etc.



SmartSteps

For each **cell** and for each **hour** the dataset contains:

- ✓ an estimation of how many people are in the cell
- ✓ the percentage of these people at home, at work or just visiting the cell
- ✓ the gender splits (male vs. female)
- ✓ the age splits (0-20 years, 21-30 years, 31-40 years, ...)

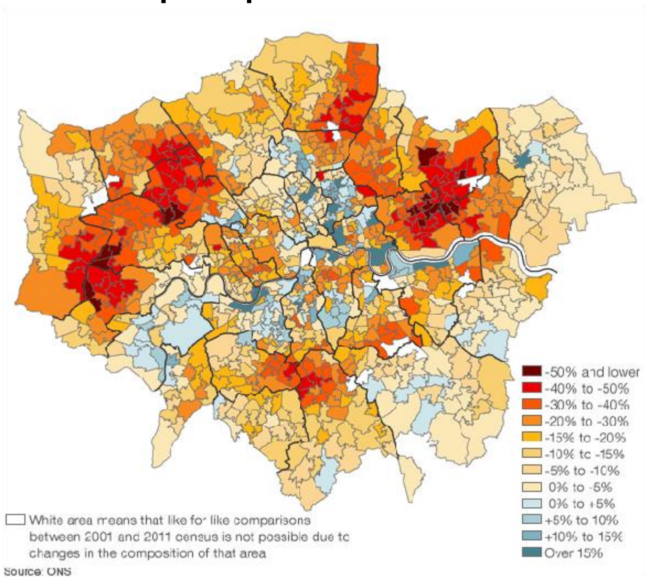
Crime data

- ✓ crime geolocation for 2 months (December 2012 – January 2013)
- ✓ all reported crimes in UK specifying month and year and not specific date
- ✓ median crime value (=5) used as threshold
- ✓ Spatial granularity of borough profiles is at LSOA levels: LSOA are small geographical areas defined by UK Office for National Statistics (mean population: 1500)

London Borough Profiles

- ✓ 68 metrics about the population of a specific geographical area: demographics, households, migrant population, employment, earnings, life expectancy, happiness levels, house prices, etc.

- ✓ Spatial granularity of borough profiles is at LSOA levels: LSOA are small geographical areas defined by UK Office for National Statistics (mean population: 1500)



Feature Extraction

From **Smartsteps** data we extract

✓ 1st order features (mean, median, min., max., entropy, etc.)

✓ 2nd order features on sliding windows of variable length (1 hour, 4 hours, 1 day, etc.) to account for temporal patterns

Feature Selection

- ✓ mean decrease in Gini coefficient of inequality
  - ✓ the feature with maximum mean decrease in Gini coefficient is expected to have the maximum influence in minimizing the out-of-the-bag error

- ✓ the feature selection process produced a reduced subset of **68** features (from an initial pool of about 6000 features)

Classification Approach

✓ Binary classification task: high crime area vs low crime area

✓ 10-fold cross-validation

✓ Classifier: Random Forest (RF)

✓ RF overcomes logistic regression, support vector machines, neural networks, decision trees

Experimental Results

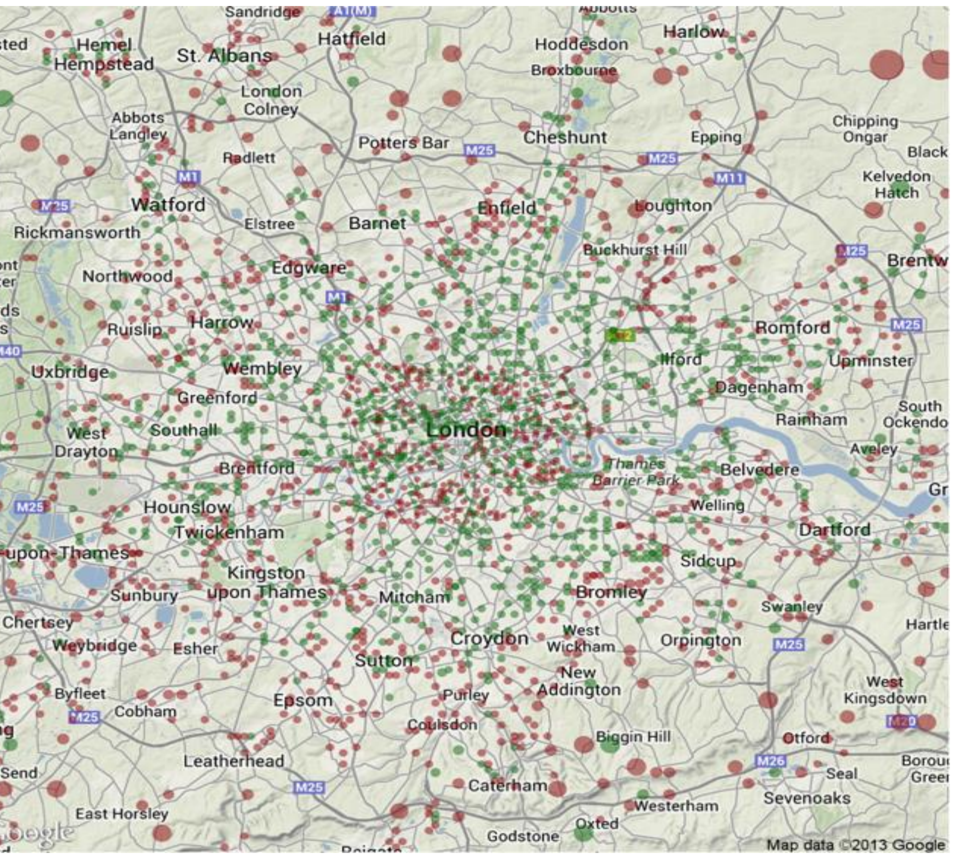
Table 3: Metrics Comparison

Model	Acc.,%	Acc. CI, 95%	F1,%	AUC
Baseline Majority Classifier	53.15	(0.53, 0.53)	0	0.50
Borough Profiles Model (BPM)	62.18	(0.61, 0.64)	57.52	0.58
Smartsteps	68.37	(0.67, 0.70)	65.43	0.63
Smartsteps + BPM	69.54	(0.68, 0.71)	67.23	0.64

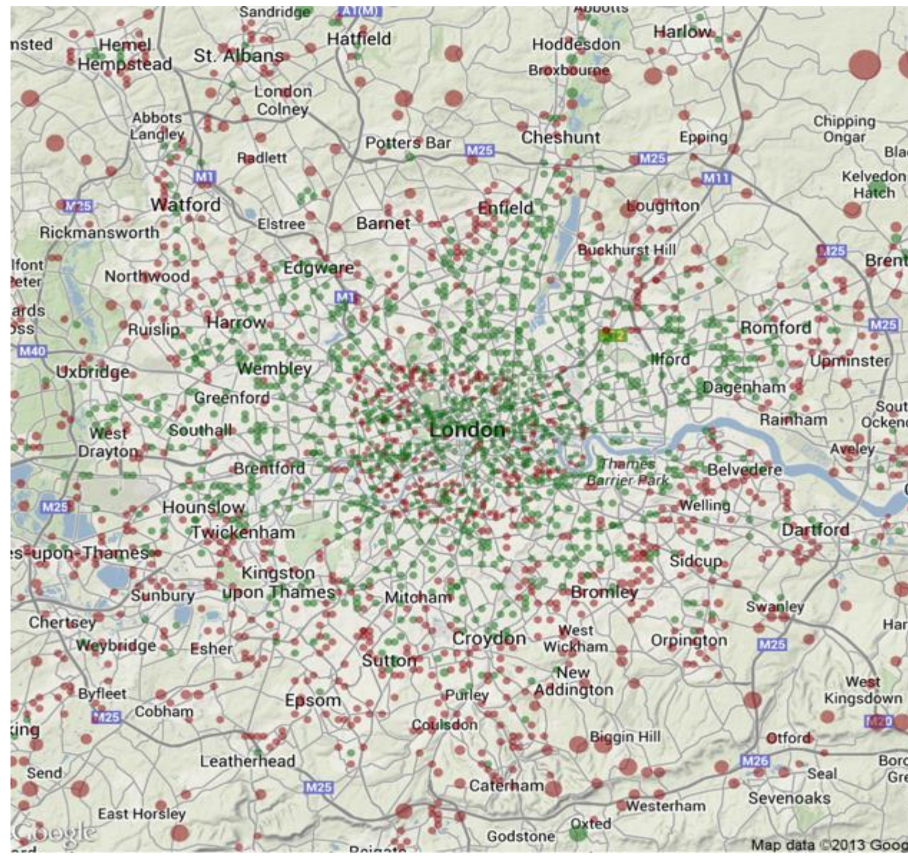
Smartsteps-based classifier significantly outperforms baseline majority and borough profiles-based classifiers

Experimental Results

~70% accuracy in predicting crime hotspots



ground-truth



predictions

Relevant Features

- ✓ Features encoding ***daily dynamics*** have more predictive power than features extracted on a monthly basis

- ✓ Relevance of ***high number of residents*** to predict crime areas

  - ✓ increased ratio of residents -> more crime (in contrast with Newman's thesis)

- ✓ ***Entropy-based features*** are useful for predicting the crime hotspots

  - ✓ high diversity of functions (home vs work) and high diversity of people (gender and age) act as ***eyes on street*** decreasing crime (in line with Jacobs' thesis)

Relevant Features

- ✓ Only 6 out of 68 features in the joint model are London Borough features, namely
  - ✓ %working population claiming out of work benefits
  - ✓ Largest migrant population
  - ✓ % overseas nationals entering the UK
  - ✓ % resident population born abroad

Implications

- ✓ Our method captures the **dynamics** of a **place** rather than making extrapolations from previous crime histories. We can use it in areas where people are less inclined to report crimes
- ✓ Our method provides new ways of describing geographical areas: novel risk-inducing or risk-reducing features of geographical areas

Project 2:
Testing Jane Jacobs in Bogotá

Marco De Nadai*, Andrey Bogomolov*, Andrés Clavijo**, Rodrigo Lara Molina**, Bruno Lepri*, Emmanuel Letouzé**, Nuria Oliver**, Gabriel Pestre**. Joan Serra***, Natalie Shoup**, Alvaro Ramirez Suarez***

*

FONDAZIONE BRUNO KESSLER

**

DATA POP ALLIANCE

HARVARD HUMANITARIAN INITIATIVE

mit media lab

ODI

FLOWMINDER.ORG

***

Telefonica

Why and how Bogotá & Colombia?

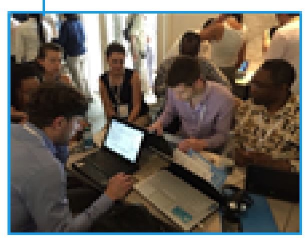
1

60th Anniversary of the National Department of Statistics of Colombia
Official Statistics in the Big Data Era
Emmanuel Letouzé

Data-Pop alliance is hired to help Colombia's National Statistical Office (DANE) to develop its **Big Data Strategy**

Delivery of **Big Data strategy** analysis to DANE

Data-Pop Alliance gets **$230k from the World Bank** to develop a research program in Colombia



Cartagena Data Festival, April 20-22

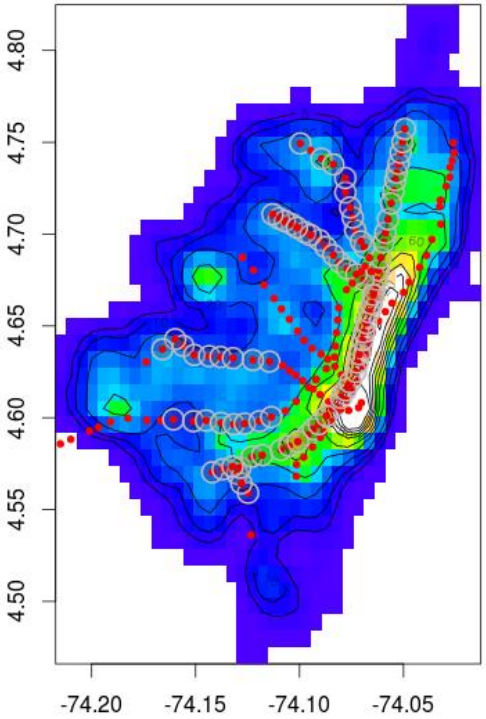
Data-Pop Alliance gets **$500k from Hewlett Foundation** to develop professional training workshops in Senegal, Rwanda, Kenya and Colombia

March 15th 2016, Kick-off of **Ciudad Laboratorio** in Bogotá

Agreements with
Alcaldía Mayor de Bogotá D.C.
Bogotá Mejor para Todos
Cámara de Comercio de Bogotá
FIP Fundación Ideas para la Paz
Universidad de los Andes Colombia

DATA POP ALLIANCE

2013 | 2014 | 2015 | 2016 | 2017



Bogotá, a big complex city

- 9.8 million people (1.2 times NYC)
- > 13,000 people/sq km (top 10)

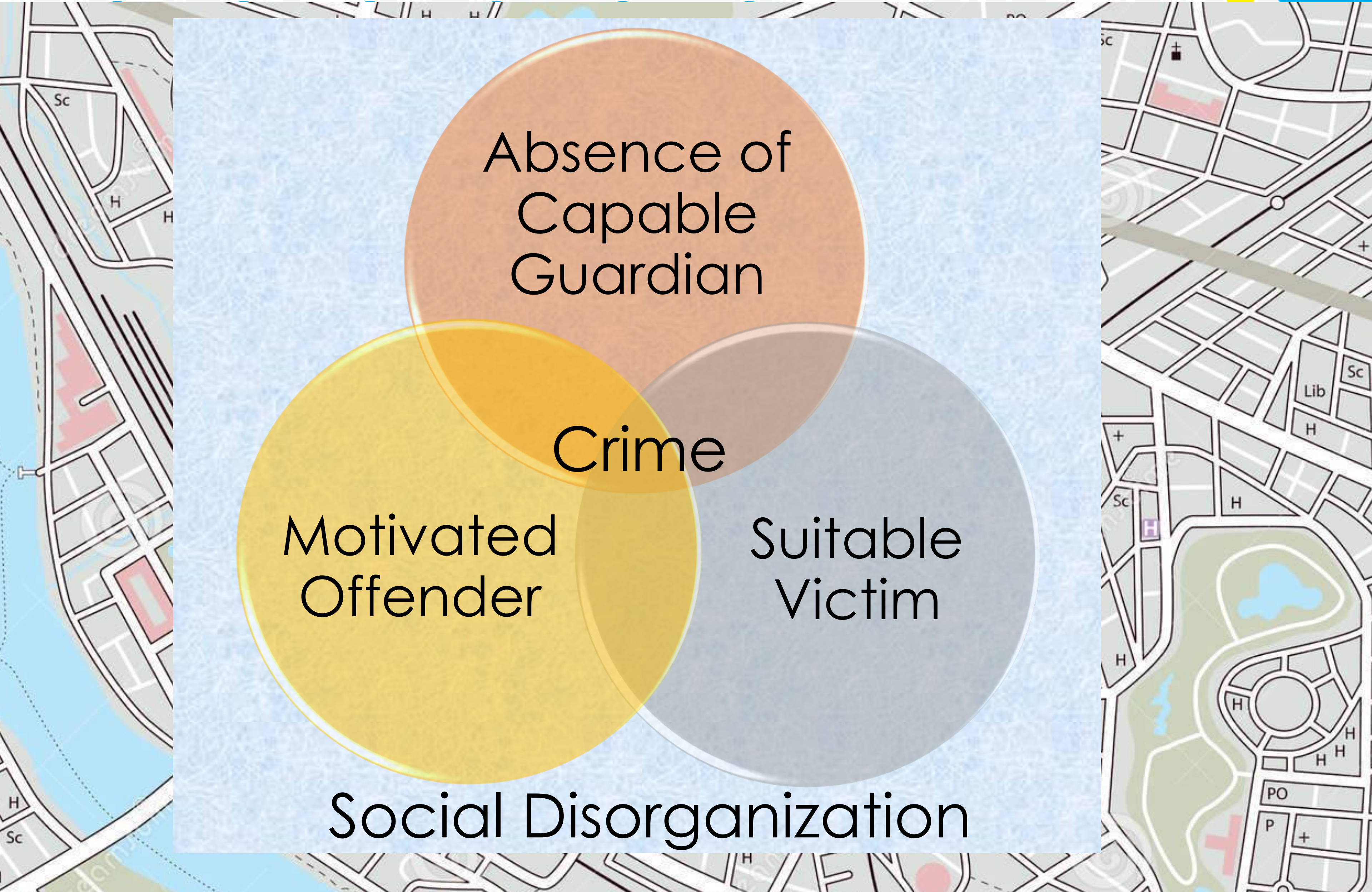
Crime: a top issue in Bogotá

- High crime rate
- A lot of crimes not reported to the police

➢ **Questions:**

- Have **structural** features an incidence on crime?
- Is it possible to predict crime rates using structural features?
- Do structural features have more relevance than socio-economic conditions and / or behavioral outcomes in the prediction of crime

Crime


Absence of Capable Guardian
Crime
Motivated Offender
Suitable Victim
Social Disorganization

Jane Jacobs 101:
vitality & safety == diversity

1

Jacobs' 4 pillars for urban vitality and safety

(i) ***mixed land uses*** to attract people who have different purposes;

(ii) ***small blocks*** that promote contact opportunities among people

(iii) ***building diversity***: mix high-rent and low-rent tenants

(iv) ***people concentration***: promote high density levels

1961

Call Detail Records (CDRs), public transportation data, OpenStreet Map...

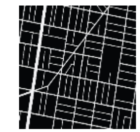
2016

Jacobs' Structural Features

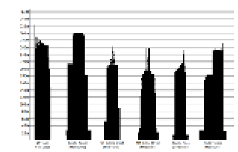
3

Mixed land uses
Land Use Mix
Residential non-res. Mix
Closeness to small parks
Average number of floors
Closeness to Daily 3rd places density

Small blocks
Mean block area
Intersection density



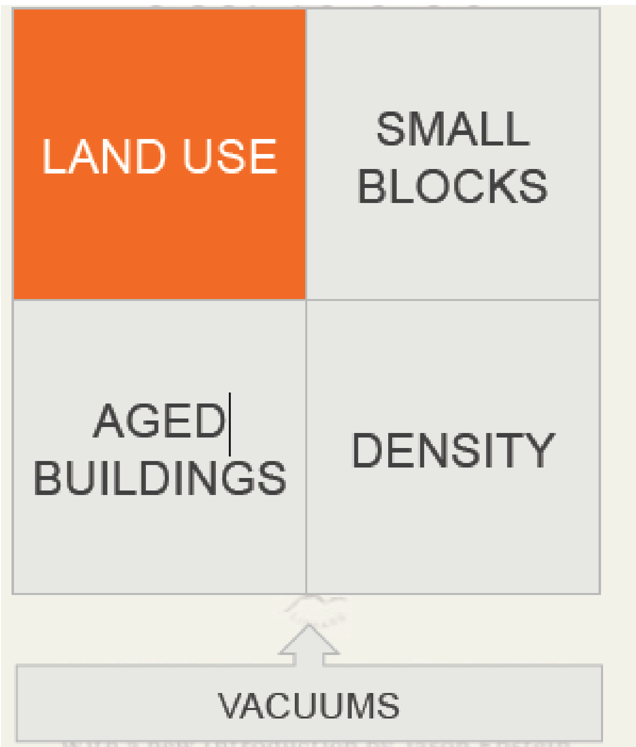
Building diversity
Strata mean
Strata Variation



Density:
Population density
Employment density
% Employed population
Buildings density

Vacuums
% Border vacuum
Closeness to large parks
Closeness highways
Closeness water

Jacobs’ Structural Features



2+ primary uses (contemporarily)

For district i:

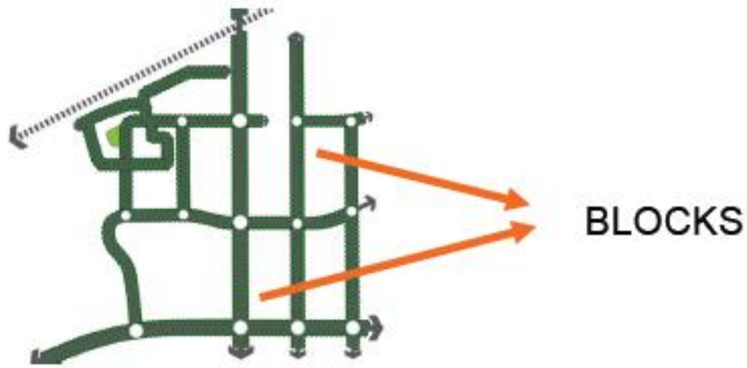
$$LUM_i = -\sum_{j \in N} \frac{P_{i,j} \log(P_{i,j})}{\log |N|}$$

$P_{i,j}$: % square footage of land use j

N: {residential, commercial, recreation}

Jacobs' Structural Features

City blocks should be small/short



For district i:

$$\frac{\sum_{i=0}^{n} blocksize(i)}{n}$$

Jacobs' Structural Features

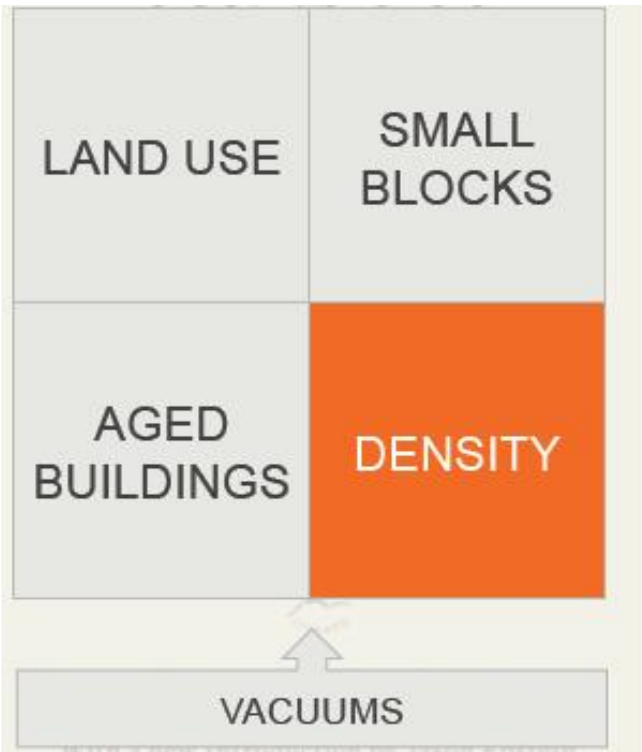
3

Buildings mixed (age and types)

Standard deviation of building ages

Jacobs' Structural Features

3



Concentration of people and enterprises

$$\text{Population density}_i = \frac{|\text{Population}_i|}{\text{area}_i}$$

$$\text{Buildings density}_i = \frac{|\text{Buildings}_i|}{\text{area}_i}$$

Jacobs' Structural Features

3

Places that act as physical obstacles to pedestrian activity, e.g. large parks, large highways, large bodies of water

Structural Features: Bogota

3

1. Land Use:
    a) Type: residential, comercial, parks, leisure, others…
    b) Points of Interest (POI)
    c) Boundaries: wasteland, highways, rivers, etc…
2. Blocks:
    a) Shape or Geometry
    b) Intersections
3. Building Diversity
    a) Stratum
    b) Form (heigh, built ground area)
4. Concentration
    a) Residents and employees
    b) N° of appartments by building
    c) Daily and non-daily POI

Social Disorganization

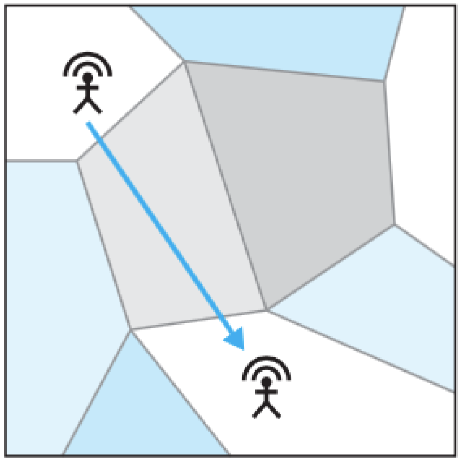
3

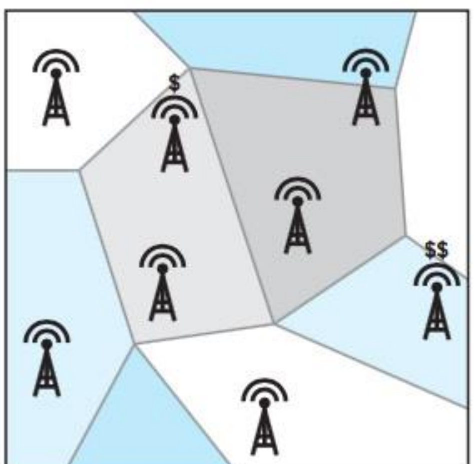
Subjective poverty prevalence

Monetary poverty prevalence

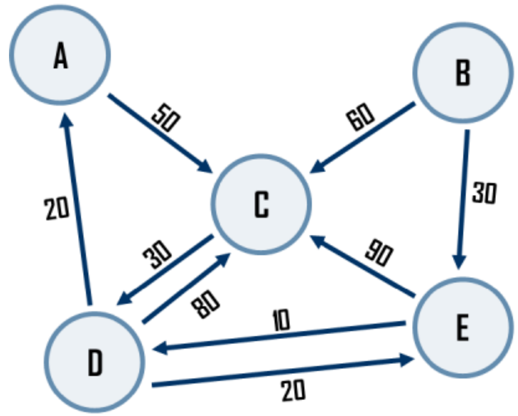
Multidimensional Poverty Index

Daily Routine from CDRs





Spatial Interactions



O/D Matrix

	A	B	C	D	E	Ti
A	0	0	50	0	0	50
B	0	0	60	0	30	90
C	0	0	0	30	0	30
D	20	0	80	0	20	120
E	0	0	90	10	0	100
Tj	20	0	280	40	50	**390**

- Compute OD matrix of a typical day from the CDRs

Ground truth: Crime Data

3

27,863 cases of **homicide** and **theft** (burglaries of commercial property, burglaries of houses, and robberies) for 2014.

Specifically, the dataset includes the category and subcategory of the crime, the longitude, latitude, and address of where the crime was reported to have occurred, and the responsible police department.

Ground truth: Crime Data

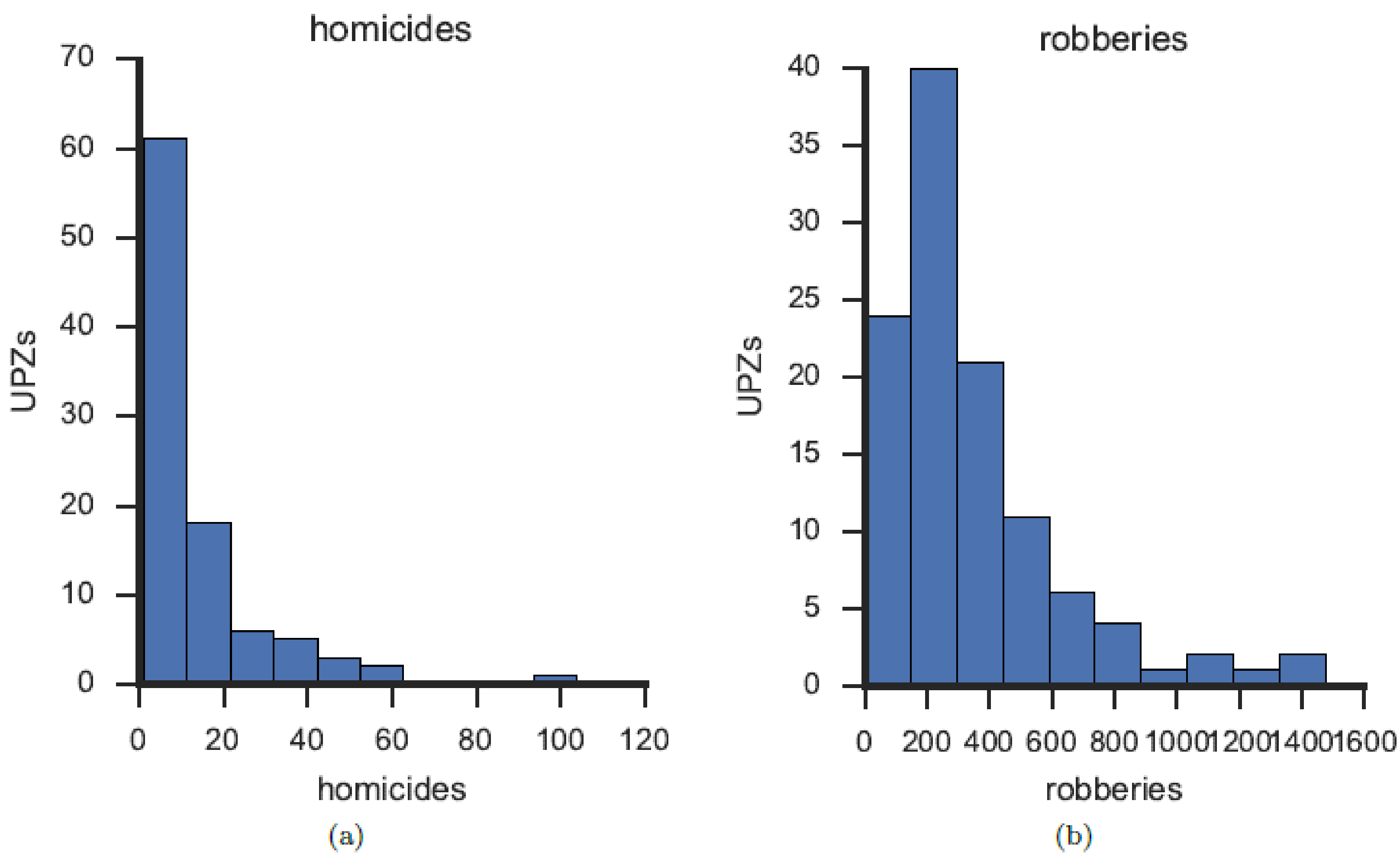
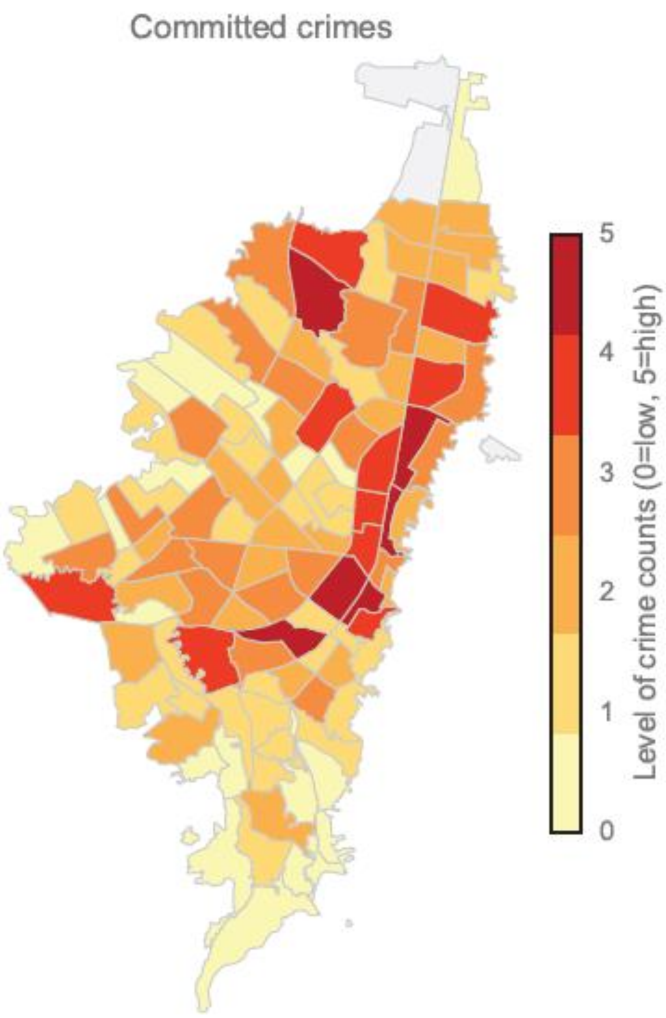
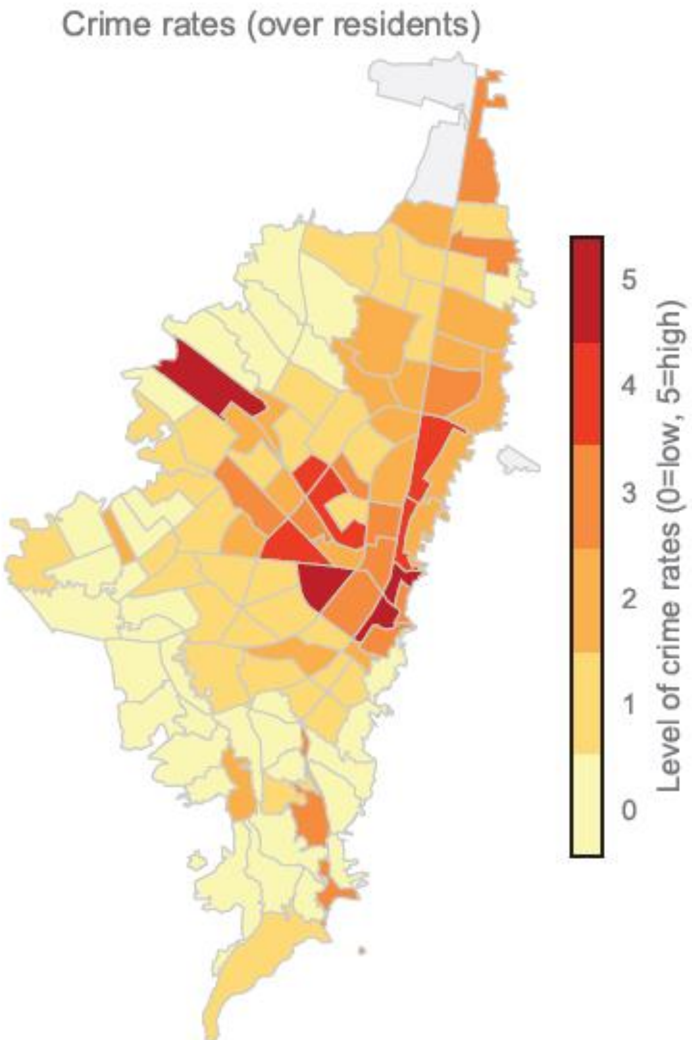


Figure 1: a) Violent crime distribution per UPZ; b) Robberies distribution per UPZ.

Ground truth: Crime Cases

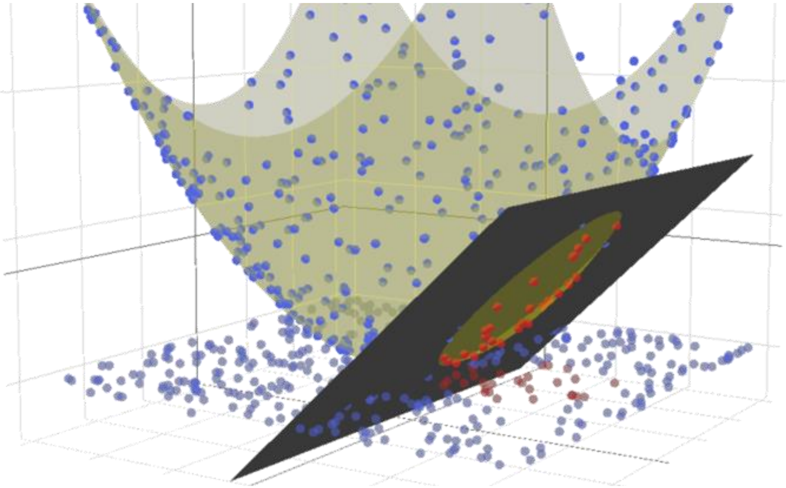




Regression Models

3

- ~~**Linear Regression:**~~
  - Negative values must not be allowed
- ~~**Poisson Regression:**~~
  - It is adapted for predicting a count variable
  - But, it enforces mean be equal to variance, leading to overdispersion
- **Negative Binomial Regression:**
  - Also adapted for count variable
  - Take overdispersion into account



Negative Binomial Regression

3

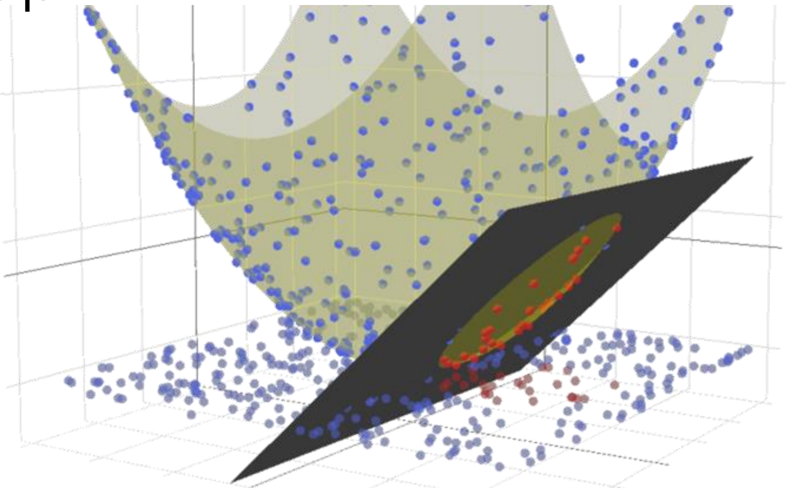
$$\log(E(Y_i)) = \beta_0 + \sum_{k=1}^{n} X_k B_k + \sum_{j=1}^{p} E_j B_{n+j} + \sum_{t=1}^{p} R_t B_{n+j+t}$$

Crime	Features	Auto-correlation	Human routine
(ground truth)	(e.g. land use mix, deprivation)	Spatial Eigenvectors	O/Ds Eigenvectors

Regression Models: Quality

3

- McFadder Pseudo R^2 which compares the log likelihood of the full model with that of the intercept model
- Also RMSE and MAE



$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{x}_i)^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - \widehat{x}_i|$$

Results

	Deprivation	City's diversity	People's dynamics	Full
Land use				
Land use mix (1)	-	×	-	-0.120**
Residential *vs.* non-res. mix^s(2)	-	0.302*	-	0.470***
Closeness small parks^l(3)	-	0.303***	-	0.248*
Housing type avg. (4)	-	-0.402**	-	-0.400***
Housing type std. (4)	-	-0.385*	-	-0.503***
Closeness daily buildings^l(5)	-	-0.537***	-	-0.292***
3^{rd}Places^s(6)	-	×	-	-0.155
Small blocks				
Block area (7)	-	-0.457***	-	-0.340***
Intersection density (8)	-	×	-	-0.439***
4-ways intersections density (9)	-	×	-	0.158***
Buildings				
Strata value^s(10)	-	-0.400*	-	-0.322*
Strata variation^s(11)	-	-0.190**	-	-0.176*
Concentration				
Population density (12)	-	×	-	-0.777*
Employment density (13)	-	0.327***	-	0.928**
Vacuums				
Closeness parks^l(16)	-	-0.221***	-	-0.130*
Closeness water^l(18)	-	×	-	×
Deprivation				
Multidimensional poverty (19)	0.452**	-	-	-0.408**
Subjective poverty	0.287	-	-	0.408***
OD network				
Spatial Diversity (20)	-	-	×	×
Centrality (21)	-	-	-0.805***	×
Total flow (23)	-	-	×	-0.187**
Attractiveness (24)	-	-	-0.314***	×
Negative Residual (27)	-	-	-0.765***	-0.214*
RMSE	12.27	7.71	11.60	6.24
MAE	7.92	4.88	8.07	1.98
McFadden Pseudo-R^2	0.08	0.19	0.09	0.24

Homicides

	Deprivation	City's diversity	People's dynamics	Full
Land use				
Land use mix (1)	-	-0.108*	-	-0.149***
Residential *vs.* non-res. mix^s(2)	-	0.316***	-	0.307***
Closeness small parks^l(3)	-	0.303***	-	-0.127*
Housing type avg. (4)	-	-0.096	-	×
Housing type std. (4)	-	-0.564***	-	-0.394***
Closeness daily buildings^l(5)	-	-0.266***	-	-0.232***
3^{rd}Places^s(6)	-	×	-	0.068
Small blocks				
Block area (7)	-	-0.290***	-	-0.205***
Intersection density (8)	-	×	-	-0.532***
4-ways intersections density (9)	-	×	-	0.271***
Buildings				
Strata value^s(10)	-	×	-	×
Strata variation^s(11)	-	-0.044	-	-0.128*
Concentration				
Population density (12)	-	-0.470*	-	-0.328
Employment density (13)	-	0.744***	-	0.766***
Vacuums				
Closeness parks^l(16)	-	-0.169***	-	-0.136***
Closeness water^l(18)	-	0.037	-	×
Deprivation				
Multidimensional poverty (19)	-0.115	-	-	0.189**
Subjective poverty	0.048	-	-	×
OD network				
Spatial Diversity (20)	-	-	×	×
Centrality (21)	-	-	0.436***	×
Total flow (23)	-	-	-0.051	-0.068***
Attractiveness (24)	-	-	-0.314***	-0.175***
Negative Residual (27)	-	-	-0.104	×
RMSE	229.49	104.18	184.30	93.35
MAE	158.95	72.88	128.45	60.64
McFadden Pseudo-R^2	0.04	0.15	0.07	0.18

Robberies

Results

- **Structural characteristics** are a ***better*** predictor of the target variables (robberies and homicides) than *socio-economic* variables
- Structural + socio-economic improves predictions
- Employment density is strongly correlated with crime
- Low crime is correlated with:
  - High housing and buildings diversity
  - Low deprivation
  - Large blocks (!) –as opposed to slums
- The larger the **flow of people** in a neighborhood, the **lower the crime**

Results: Confirming Jacobs' Theory

Population diversity, activity and a high mix of functions lead to less crime

Relevant Publications

"Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data" - A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi, A. Pentland 16th **ACM International Conference on Multimodal Interaction (ICMI 2014)**

- "Moves on the street: classifying crime hotspots using aggregated and anonymized data on people dynamics"
  Bogomolov, A., Lepri, B., Staiano, J., Letouze, E., **Oliver, N.**, Pianesi, F. and Pentland, A.
  Big Data Journal, Mary Ann Liebert, Inc. Vol 3, Issue 3, Sept 2015

"What makes a city vital and safe: Bogota case study" – De Nadai, M. et al, DataPop Alliance Report

Limitations and challenges

- Multi-disciplinary, multi-party projects which adds complexity
- Quality in the ground truth and input data
- What are the optimal actions to take after the insights of the analysis?
  - Creating knowledge that can inform policies but how?
  - Which features can a city/country act upon?
  - How to engage with the civil society?

Next Steps

- New project with **IDB** to expand to 3-5 cities in Colombia (Bogota, Medellin, Cartagena, Cali and Santa Marta)
- Deepen parts of the analysis
- Study how people feel about the police, trust issues
- Assess the level of **bias** in reporting
- Do **interventions** to create a more enabling environment for a better public policing and reduced criminality

THANK YOU!

nuria@alum.mit.edu
@nuriaoliver
http://www.nuriaoliver.com