



## **Visual Analytics for Urban Crime Analysis**

## Germain García Zanabria Advisor: Prof. Dr. Luis Gustavo Nonato

Instituto de Ciências Matemáticas e de Computação USP - São Carlos





## » Outline:

- Introduction/Motivation
- Crime Data
- Contributions
  - CrimAnalyzer
  - Mirante
  - CriPAV
- Conclusions

Crime

Crime can be defined as breaking or breaching of criminal law (penal code) that governs a particular geographical area (jurisdiction) aimed at protecting the lives, property, and the right of citizens in that jurisdiction [1].



Crime is an offense against a person, or his/her property, violation of socially accepted rules of human ethical or moral behavior.

Law enforcement agencies deploy resources in a more effective manner to:

- Prevent
- Control
- Reduce

Crime activities

Search for patterns, trends, structure, irregularities, relationships among data

**Computer** Abilities

Data Storage Numerical Computation Searching Compression & Filtering

Visual Analytics [2]

Human-Computer Interaction (Information Visualization)

> Human Cognition Perception Creativity General Knowledge Visual Intelligence Human Abilities

Keim et al. (2008)



### **Crime Mapping**

A branch of Geographic Information System (GIS) devoted to explain spatio-temporal behavior of crime activities.



#### Allows

- Demonstrate the importance of local geography for crime frequency and type.
- Identify and visualize hotspots.
- Identify the seasonality of crime types in certain locations.

[3,4]

Brazil is a dangerous place, with a high murder rate and surprisingly high disparity when compared against other large countries\*.





## » Data Set 1 - São Paulo



#### From **2000** to **2006**

- The data set contains **3 attributes**:
  - Census unit code where the crime happened.
  - Type of crime.
  - Date and time of the crime.

**Crime types** range in **121** categories:

- Passerby robbery
- Auto theft
- Larceny
- ..

20 crime records

#### **Categories** are:

- Roubo 691954
- Furto 587885
- Roubo de veículo 295 081









## » Data Set 2 - São Paulo

#### From 2006 to 2017

- Attributes:
  - ANO\_OCORR: Year of occurrence.
  - DATA\_OCORRENCIA\_BO: Date of occurrence.
  - HORA\_OCORRENCIA\_BO: Hour of occurrence (many of them nominal: Madrugada, amanhã, Noite).
  - NOME\_DELEGACIA\_CIRC: Station name
  - **RUBRICA:** Crime type (**16** types)
  - FLAG\_STATUS: Status (consumado).
  - COD\_SETOR: Code of census block
  - COORD\_X: lat
  - COORD\_Y: lng
    - Crime types range in **3** categories:
      - Passerby robbery
      - Commercial establishment robbery
      - Vehicle robbery

**30815** census blocks.

NE





## » Data Set 3 - São Carlos

Data set from **2014** to **2019** 

#### Atributes:

- ANO: Year of occurrence.
- **DATA\_OCORRENCIA\_BO:** Date of occurrence.
- HORA\_OCORRENCIA\_BO: Hour of occurrence.
- FLAGRANTE: Flagrant
- **CONDUTA:** Type of crime (**13** types)
- LATITUDE: lat
- LONGITUDE: lng







The focus of this thesis is to study and analyze crime patterns considering different factors. To do this, we have to sort out different crime analysis problems: hotspot definition and detection, space discretization, and spatio-temporal dynamics.



#### Hotspot definition and detection

Given the crime events in an urban space, we propose different methods (depending on the spatial discretization) to identify and present hotspots considering not only the intensity but also the frequency of crimes;

Space discretization

Different domain discretization, switching from grid-based to a street-based spatial discretization;

#### Spatio-temporal dynamics

Spatio-temporal crime patterns analysis, supported by visualization and machine learning mechanisms to extract and visually present different spatio-temporal patterns.









## CrimAnalyzer

### > CrimAnalyzer - Problem Analysis





## » CrimAnalyzer - Pipeline



# CrimAnalyzer **Hotspot Identification** Model

## » Non-Negative Matrix Factorization (NMF)

An  $m \ge n$  matrix X is said *non-negative* if all entries X are greater or equal to zero ( $X \ge 0$ ). The goal of NMF is to decompose X as a product W.H, where W and H are *non-negative* with dimensions  $m \ge k$  and  $k \ge n$ , respectively.

$$\arg\min_{W,H} \|X - WH\|^2 \quad \text{subject to} \quad W, H \ge 0$$



A set of basis vectors  $w_i$  (columns of W), and a set of coefficients  $h_j$  (columns of H), such that each column  $x_j$  of X is approximated as linear combination  $x_j \cong \sum_i h_{ij} w_i$ , (or  $x_j = W h_j$ )<sub>[22]</sub>.



## > Identifying Hotspot with NMF

NMF to identify hotspots, their frequency and "intensity". The matrix X to be decomposed as the product W.H comprises crime information in a particular region of interest.





## » Data Modelling with NMF



KIM, H.; PARK, H. Sparse non-negative matrix factorizations via alternating non-negativity constrained least squares for microarray data analysis, 2007

## CrimAnalyzer Visual Analytic Tool

## >> CrimAnalyzer



## > CrimAnalyzer





## CrimAnalyzer Case Studies

## >CrimAnalyzer - Case Study 2



**BR116** 

#### Connects São Paulo to states in the south of Brazil





**SP230** 

## >CrimAnalyzer - Case Study 2







## >CrimAnalyzer - Case Studies



USP ICMC

## > CrimAnalyzer - Publication

#### CrimAnalyzer: Understanding Crime Patterns in São Paulo

Published in: IEEE Transactions on Visualization and Computer Graphics



https://ieeexplore.ieee.org/document/8869805

JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2015

#### CrimAnalyzer: Understanding Crime Patterns in São Paulo

The final welling of the second distribution and Complete Alignment in 1100/TVCG 2010 2007211

Germain Garcia, Jaqueline Silveira, Jorge Poco Member, IEEE, Afonso Paiva, Marcelo Batista Nery, Claudio T. Silva Fellow, IEEE, Sergio Adorno, Luis Gustavo Nonato Member, IEEE

Abstract—São Paulo is the largest city in South America, with crime rates that reflect its size. The number and type of crimes vary considerably around the city, assuming different patterns depending on urban and social characteristics of each particular location Previous works have mostly focused on the analysis of crimes with the intent of uncovering patterns associated to social factors, seasonality, and urban routine activities. Therefore, those studies and tools are more global in the sense that they are not designed to investigate specific regions of the city such as particular neighborhoods, avenues, or public areas. Tools able to explore specific locations of the city are essential for domain experts to accomplish their analysis in a bottom-up fashion. Revealing how urban features related to mobility, passersby behavior, and presence of public infrastructures (e.g., terminals of public transportation and schools) can influence the quantity and type of crimes. In this paper, we present CrimAnalyzer, a visual analytic tool that allows users to study the behavior of crimes in specific regions of a city. The system allows users to identify local hotspots and the pattern of crimes associated to them, while still showing how hotspots and corresponding crime patterns change over time. CrimAnalyzer has been developed from the needs of a team of experts in criminology and deals with three major challenges; i) flexibility to explore local regions and understand their crime patterns, ii) identification of spatial crime hotspots that might not be the most prevalent ones in terms of the number of crimes but that are important enough to be investigated, and iii) understand the dynamic of crime patterns over time. The effectiveness and usefulness of the proposed system are demonstrated by qualitative and quantitative comparisons as well as by case studies run by domain experts involving real data. The experiments show the capability of CrimAnalyzer in identifying crime-related phenomena.

UTER GRAPHICS

Index Terms-Crime Data, Spatio-Temporal Data, Visual Analytics, Non-Negative Matrix Factorization

#### 1 INTRODUCTI

C INCE the mid-IEEE TRANSACTIONS ON S transition proce SUALIZATION AND With this political tr increasingly be solv has not happened. Ir an explosion of con crimes. There is stil the reasons that exand violence in Braz

Among the explanations that arise more frequently is the exhaustion to reveal particular information such as burglary in commercial and violence and the state's ability to contain them within the rule

- Germain Garcia, Jaqueline Silveira, and Afonso Paiva are with ICMC-USP, São Carlos, Brazil. E-mail: {germaingarcia,alva.jaque}@usp.br, anneto@icmc usn hr
- Marcelo Batista Nery is with RIDC -FAPESP and Institute of Advanced tudies - Global Cities Program. E-mail: mbnery@gmail.com Sergio Adorno is with NEV-USP, São Paulo, Brazil. E-mail:
- arsadorno@usp.br Jorge Poco is with Fundação Getúlio Vargas, Brazil and Universidad
- Católica San Pablo. E-mail: jorge.poco@fgv.br Claudio Silva is with New York University, USA. E-mail: csilva@nyu.edu
- . Luis Gustavo Nonato is with ICMC-USP, São Carlos, Brazil and New York University, USA. E-mail: gnonato@icmc.usp.br

under-development

formation Systems behavior of crimes. minologists in their f local geography as hich they occur in a ying and visualizing e-related attributes

of traditional security policy models. Concerning this last aspect, areas or the seasonality of auto theft in certain neighborhoods is it is undeniable that crimes have not only grown, but also become among the key components of a crime mapping approach [38]. more violent and modernized. In contrast, agencies in charge of Most existing tools developed for crime mapping focused on law and order (e.g. police and criminal justice system) have not the detection of hotspots, that is, areas with a high number of kept up with these trends. The gap between the dynamics of crime criminal incidents [14]. Although sophisticated mechanisms have been proposed to detect hotspots [15], the search for a high of law has widened. Therefore, introducing modern instruments prevalence of crimes ends up neglecting sites where certain types for the management of public order and crime containment is of crimes are frequent but not sufficiently intense to be considered imperative to make public security policies more efficient, not statistically significant [51]. Moreover, most techniques enable only rudimentary mechanisms to analyze an important component of unlawful activities, the temporal evolution of crimes and their patterns. In fact, visualization resources for temporal analysis available in the majority of crime mapping systems are very restrictive, impairing users from performing elaborated queries and data exploration [3].

There is yet another important aspect to be considered in the context of crime mapping, the specificities of urban areas under analysis. São Paulo, for example, bears one of the highest crime

1. São Paulo is both a state and a city. In this paper, any time that we do not explicitly specify, São Paulo will refer to the city.

1077-2526 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications\_standards/publications/rights/index.html for more informatio Convright (c) 2020 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing rubs, permis





### » Mirante - Motivation









## Mirante Data Modeling

## » Mirante - Data Modeling





## » Mirante - Algorithm

 $L_{crime} = \{c_0, c_1, ..., c_n\}$  G = (V, E)

а



Closest node strategy

Edge-node strategy











## » Mirante - Data Modeling





## Mirante Visualization Tool

### >> Mirante



## Mirante Case Studies



## Case Study 1 Vehicle Robbery in São Paulo

São Paulo - SP - Brazil







Leaflet | Colaboradores: CEMEAI & NEV |



Leaflet | Colaboradores: CEMEAI & NEV |







Leaflet | Colaboradores: CEMEAI & NEV |



Leaflet | Colaboradores: CEMEAL&



Leaflet | Colaboradores: CEMEAI & NEV |

## Case Study 2 Passerby Robbery in São Carlos

São Carlos - SP - Brazil



Leaflet I © OpenStreetMap © CartoDB









## » Mirante - Publication

#### Mirante: A visualization tool for analyzing urban crimes

Published in: <u>33rd Conference on Graphics, Patterns and</u> Images



https://ieeexplore.ieee.org/abstract/document/9265984

#### Mirante: A visualization tool for analyzing urban crimes Germain Garcia-Zanabria, Erick Gomez-Nieto Jorge Poco Marcelo Nery, Sergio Adorno, Jaqueline Silveira Fundação Getúlio Vargas Luis G. Nonato jorge.poco@fgv.br Universidade de São Paulo Universidade de São Paulo {germaingarcia, erick.gomez, alva.jaque}@usp.br mbnery@gmail.com, sadorno@usp.br, gnonato@icmc.usp.br Abstract-Visualization assisted crime analysis tools used by data aggregated on grid cells, each covering hundreds of public security agencies are usually designed to explore large square meters. However, recent studies point out the imporurban areas, relying on grid-based heatmaps to reveal spatial tance of analyzing micro places [13]-[16], as crime rarely crime distribution in whole districts, regions, and neighborhoods. concentrates on regions larger than a street segment or corner. Therefore, those tools can hardly identify micro-scale patterns In fact, several researchers have shown that crimes mostly occlosely related to crime opportunity, whose understanding is fundamental to the planning of preventive actions, Enabling cur near specific locations such as bars, fast-food restaurants, a combined analysis of spatial patterns and their evolution check-cashing centers, and pawnshops, since those places over time is another challenge faced by most crime analysis attract distracted and vulnerable people who carry money tools. In this paper, we present Mirante, a crime mapping and valuables [14], [17]. In other words, the environment of visualization system that allows spatiotemporal analysis of crime those places creates a crime opportunity. Therefore, relying patterns in a street-level scale. In contrast to conventional tools, Mirante builds upon street-level heatmaps and other visualization on spatial discretizations such as the regular grids renders resources that enable snatial and temporal nattern analysis. analysis a quite challenging task, since the uncovering fineidentification over time. Miran nity is not so domain experts. a cell containversatile to be im grid resolution demonstrate the SIBGRA run by domain segments, different charact hampering the were capable of ( ossible causes. of the cities while the analysis of certain types of a Patterns and Image suppose that a orner during a Understanding brner. In a grid problem due to the interplay between the spatial and temporal representation, such a temporal behavior can hardly be caught dynamics of crimes, the great variability of patterns among the if both corners lie on the same grid cell, different types of crimes, and the large amount of data involved In collaboration with domain experts, we designed Mirante, in such analysis. In this context, the branch of Geographic a scalable and versatile visualization tool tailored to explore Information Systems (GIS) called Crime Mapping focuses on crime data in a street-level of detail. Considering street corners developing tools to explore and analyze the spatio-temporal as nodes and street segments as edges, Mirante assumes city behavior of crimes, leveraging the importance of local urstreet maps as the spatial discretization. Crime data is spatially ban, social, and environmental characteristics as determinants aggregated on street corners using an edge-node strategy for crime opportunity [1], [2]. Current crime mapping tools rather than Euclidean distance, which avoids several issues combine techniques from different fields such as mathematics present in grid cell aggregation. Mirante provides a number and statistics [3]-[5], machine learning [6], [7], optimization of interactive resources to explore the spatial distribution of and visualization [8]-[10], and social sciences [11], [12]. crimes and their dynamics over time, making it possible to Examples of crime mapping systems implemented to increase identify temporal patterns such as the shift of crime hotspots transparency for the population and to support agencies in among nearby locations. Interactive filters allow users to focus charge of public security are LexisNexis1, NYC Crime Map2, their analysis on particular hours of the day, days of the CitizenRIMS3, and CrimeMapping4. week, and months of the year, making it possible to easily An important aspect of crime mapping is the spatial disscrutinize the seasonality of crimes. Using different selection cretization. Most techniques rely on regular grids with crime mechanisms, users can interactively select regions of interest in various scales, enabling the spatio-temporal analysis of large <sup>2</sup> maps.nvc.gov/crime/ <sup>3</sup> crimegraphics.com <sup>1</sup> communitycrimemap.com 4 crimemapping.com regions as well as quite specific locations of the city, a trait not



## Conclusions

## » General - Conclusions

We presented different methodologies that allow a visual Spatio-temporal crime pattern analysis of urban areas considering different characteristics. For that, we have proposed different solutions to tackle the presented problems:

- 1. we proposed two different methods based on **NMF** and **Stochastic mechanisms** for hotspots identification considering not only the intensity of crimes but also the frequency;
- 2. we have worked with different levels of spatial discretization such as census blocks grid-based and a high-level discretization based on street-network to do more accurate analysis;
- 3. we developed different visual frameworks to represent and visualize **Spatio-temporal crime** patterns.

Each of the proposed approaches has been **designed** in close collaboration with domain experts and deal simultaneously with multiple requirements. These requirements are translated into analytical tasks that guide the development of victimization systems. Moreover, the set of case studies, experiments, and experts' feedbacks have shown the usefulness and effectiveness of the proposed methodologies.

1

Supported by the experiments, the results, and the positive feedbacks, it is safe to say the proposed methodologies have the capability and functionalities to analyze successfully different crime patterns in different scenarios.

We introduced **CrimAnalyzer**, a visual analytics tool to support the analysis of crimes in local regions. We also propose a technique based on **NMF** to identify hotspots , considering the intensity and frequency.

3

We introduced **MIRANTE**, a crime mapping visualization system that allows spatiotemporal analysis of crime patterns in a street-level of detail.

4

We introduced **CriPAV** (Crime Pattern Analysis and Visualization), a **street-level** visualizationassisted analytical tool. It is a new methodology to identify crime hotspots based not only on intensity but also on the probability of occurrence, a hotspot grouping technique based on the similarity of crime time series.





## » References (1)

**[1]** AKPINAR, E. Using geographic information systems in analysing the pattern of crime incidents and the relationship between landuse and incidents.

[2] Keim et al., Visual analytics: Definition, process, and challenges, 2008.

- [3] EMIG et al., Crime analysis–a selected bibliography. Washington, DC: US National Criminal Justice Reference Service, 1980
- [4] FELSON, M.; BOBA, R. L. Crime and everyday life. [S.l.]: Sage, 2010.
- [5] SANTOS, R. B. Crime analysis with crime mapping. [S.l.]: Sage publications, 2016
  [6] WANG, T.; RUDIN, C.; WAGNER, D.; SEVIERI, R. Learning to detect patterns of crime. In: SPRINGER
- [7] EAGLIN, T.; CHO, I.; RIBARSKY, W. **Space-time kernel density estimation for real-time interactive visual analytics.** In: Proceedings of the 50th Hawaii International Conference on System Sciences
- **[8]** NETO, J. F. de Q.; SANTOS, E. M. dos; VIDAL, C. A. **Mskde-using marching squares to quickly make high quality crime hotspot maps**. In: IEEE. Graphics, Patterns and Images.
- **[9]** ZHOU, G.; LIN, J.; ZHENG, W. A web-based geographical information system for crime mapping and decision support. In: IEEE. Computational Problem-Solving (ICCP), 2012 International Conference on

## » References (2)

**[10]** XIANG, Y.; CHAU, M.; ATABAKHSH, H.; CHEN, H. Visualizing criminal relationships: Comparison of a hyperbolic tree and a hierarchical list. Decision Support Systems.

- **[11]** CALHOUN, C. C.; STOBBART, C. E.; THOMAS, D. M.; VILLARRUBIA, J. A.; BROWN, D. E.; CONKLIN, J. H. **Improving crime data sharing and analysis tools for a web-based crime analysis toolkit: Webcat 2.2.** In: IEEE.
- [12] MALIK, A.; MACIEJEWSKI, R.; COLLINS, T. F.; EBERT, D. S. Visual analytics law enforcement toolkit. In: IEEE.
- **[13]** GODWIN, A.; STASKO, J. Hotsketch: Drawing police patrol routes among spatiotemporal crime hotspots. In: Proceedings of the 50th Hawaii International Conference on System Sciences.
- **[14]** SILVA, L. J. S.; GONZÁLES, S. F.; ALMEIDA, C. F.; BARBOSA, S. D.; LOPES, H. Crimevis: An interactive visualization system for analyzing crime data in the state of Rio de Janeiro. 2017
- [15] POVEDA, A. C. Violence and economic development in colombian cities: a dynamic panel data analysis. Journal of international development,
- **[16]** ALVES, L. G.; RIBEIRO, H. V.; MENDES, R. S. **Scaling laws in the dynamics of crime growth rate**. Physica A: Statistical Mechanics and its Applications,

## » References (3)

**[16]** GRUENEWALD, P. J.; FREISTHLER, B.; REMER, L.; LASCALA, E. A.; TRENO, A. Ecological models of alcohol outlets and violent assaults: crime potentials and geospatial analysis.

[17] MCCORD; RATCLIFFE. Intensity value analysis and the criminogenic effects of land use features on local crime patterns. Crime Patterns and Analysis

**[18]** CAPLAN, J. M.; KENNEDY, L. W. **Risk terrain modeling manual**. Rutgers Center on Public Security, Newark, 2010.

[19] BENNETT, T.; HOLLOWAY, K.; FARRINGTON, D. P. Does neighborhood watch reduce crime? a systematic review and meta-analysis. Journal of Experimental Criminology

[20] PATTAVINA, A.; BYRNE, J. M.; GARCIA, L. An examination of citizen involvement in crime prevention in high-risk versus low-to moderate-risk neighborhoods. Crime & Delinquency.

[21] <u>http://produtos.seade.gov.br/produtos/ivj/index.php?tip=map&mapa=12</u>,

**[22]** KIM, H.; PARK, H. **Sparse non-negative matrix factorizations via alternating non-negativityconstrained least squares for microarray data analysis.** Bioinformatics, Oxford University Press



## » References (4)

**[17] Garcia-Zanabria, G.,** Silveira, J. A., Poco, J., Paiva, A., Nery, M. B., Silva, C. T., & Nonato, L. G. (2019). CrimAnalyzer: Understanding crime patterns in São Paulo. IEEE TVCG.

**[18] Garcia-Zanabria**, **G.**, Gomez-Nieto, E., Silveira, J., Poco, J., Nery, M., Adorno, S., & Nonato, L. G. (2020, November). Mirante: A visualization tool for analyzing urban crimes. In 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) (pp. 148-155). IEEE.

**[19] Garcia-Zanabria, G.,** Raimundo, M., Poco, J., Batista Nery, M., Silva, C., Adorno, S., Nonato, G. CriPAV: Street-Level Crime Patterns Analysis and Visualization. IEEE Transactions on Visualization and Computer Graphics. **Submitted: TVCG, 2020**.

**[20]** Alvarenga Silveira, J., **Garcia-Zanabria, G.,** Paiva A., Piccirillo D., Batista Nery, M., Adorno S., Luis Gustavo Nonato. TensorAnalyzer: A New Approach to Identify Representative Patterns Based on Tensor Decomposition. Submitted: TDS, 2020.

**[21]** Alvarenga Silveira, J., **Garcia-Zanabria, G.,** Paiva A., Piccirillo D.,Batista Nery, M., Adorno S., & Luis Gustavo Nonato(2019). São Paulo City Schools and Their Surroundings: from Non-Negative Tensor Factorization to Pattern Identification. XL Congresso Nacional de Matemática Aplicada e Computacional.



Thank you Obrigado Gracias

Germain García Zanabria germaingarcia@usp.br

## **Visual Analytics for Urban Crime Analysis**







