

Lessons learned in developing a crime analytics solution



Dr. Emanuele Santos
emanuele@dc.ufc.br
Universidade Federal do Ceará

Summer School in Data Science
FGV EMAP
February 11th, 2020

Motivation

Our approach

Conclusion

1

2

3

4

5

Our story

**Lessons
Learned**

MOTIVATION

01

400 murders a day: 10 reasons why Latin America is the world's most violent place

Christopher Woody Sep 10, 2019, 10:47 AM



Latin America is home to about 8% of the world's population but has **about one-third** of its homicides — in 2016, that meant some 400 homicides a day, or roughly 146,000 a year. But the bloodshed is not evenly distributed.

“In Brazil, the total cost associated with violent crimes in 2013, including public safety, state prison maintenance and social costs, amounted to 100 billion dollars, corresponding to 5.4% of Brazil’s GDP.”

—Lima et al (2014)

8th Brazilian Yearbook of Public Safety

http://www.forumseguranca.org.br/storage/8_anuario_2014_20150309.pdf



The 50 most violent cities in the world

Christopher Woody Apr 8, 2017, 2:29 PM

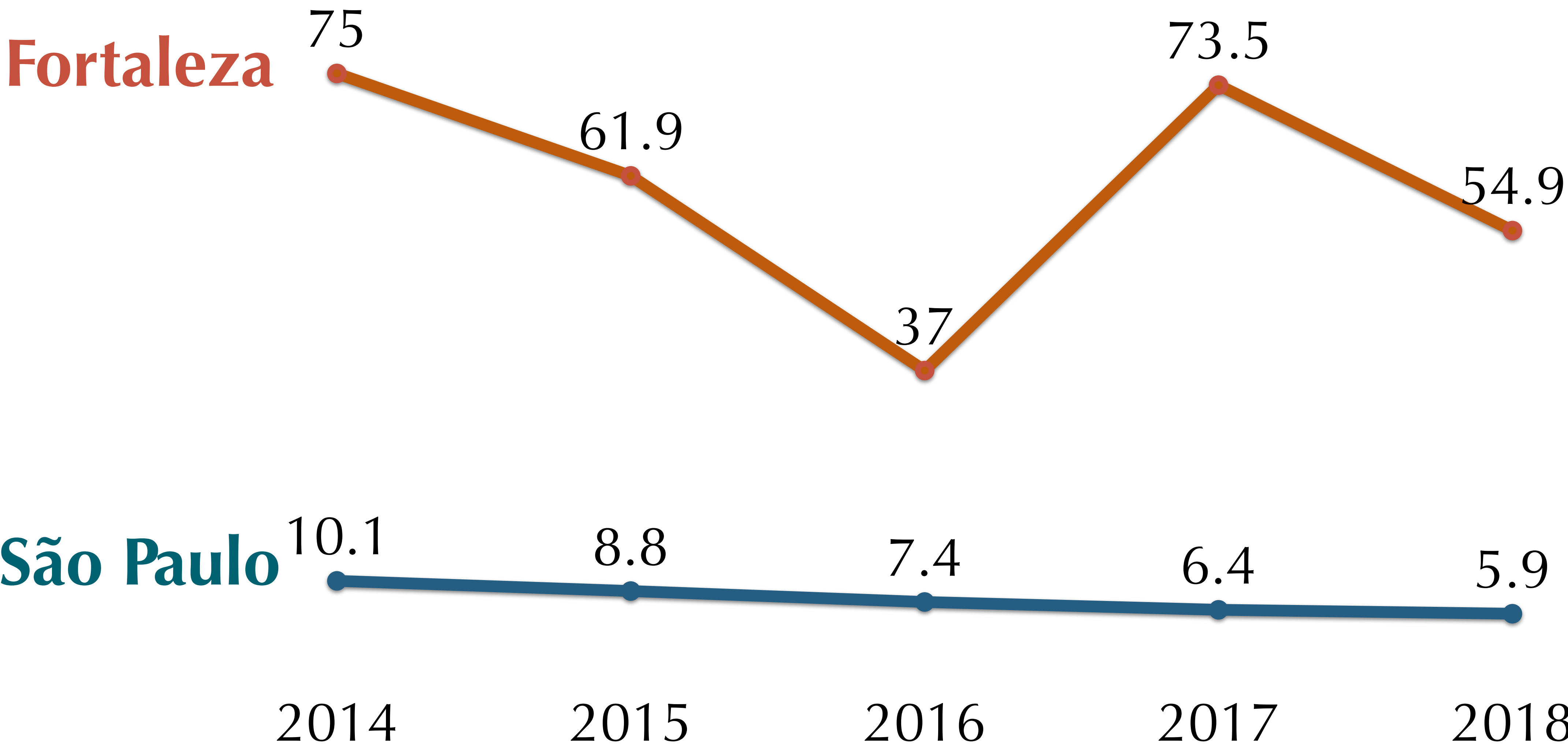


35. Fortaleza, Brazil, had 44.98 homicides per 100,000 residents.



<https://www.businessinsider.com/most-violent-cities-in-the-world-2017-4>

Number of Homicides per 100,000 Residents



Source: Brazilian Yearbook of Public Safety

OUR STORY

02

Our Story

2014

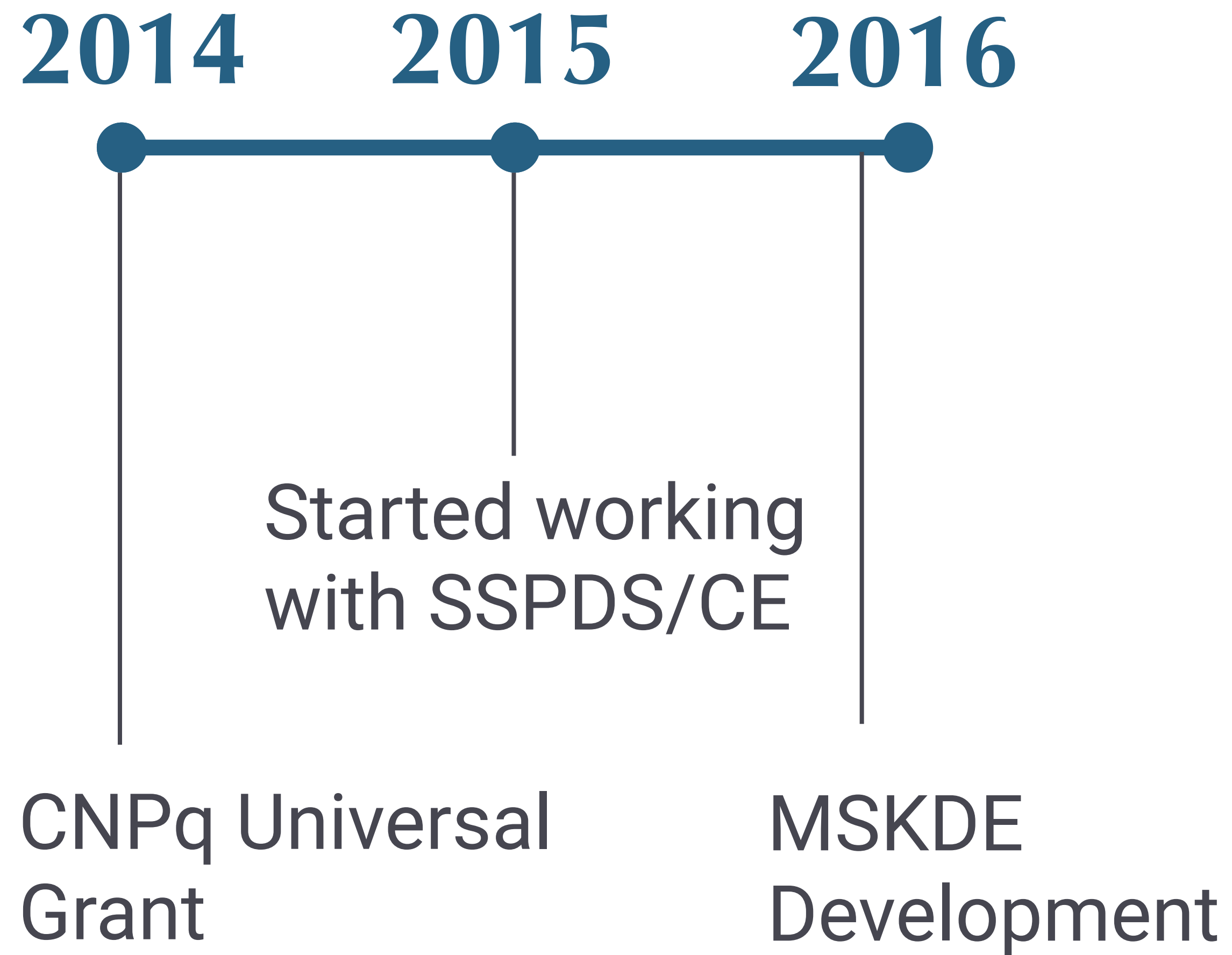


CNPq Universal
Grant

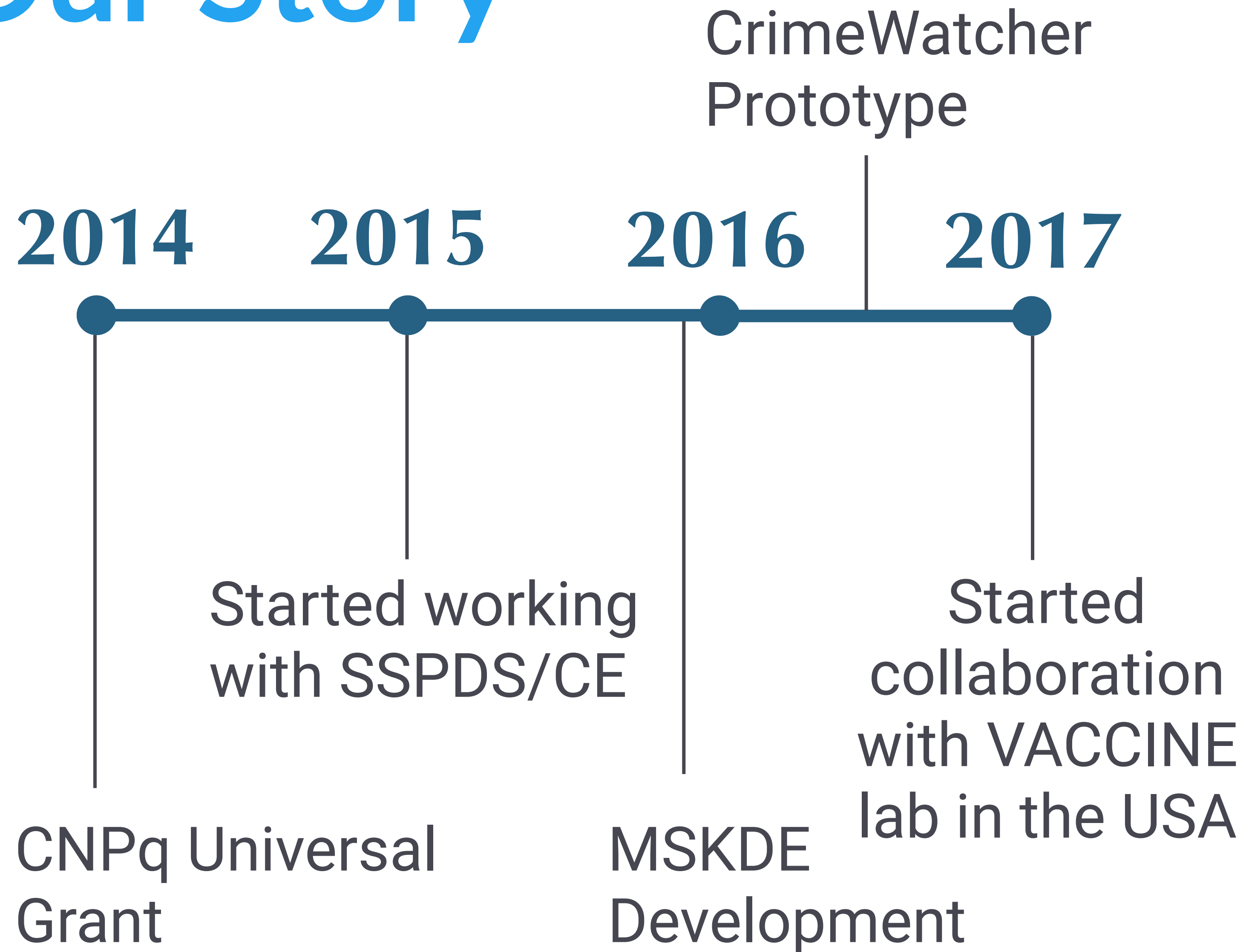
Our Story



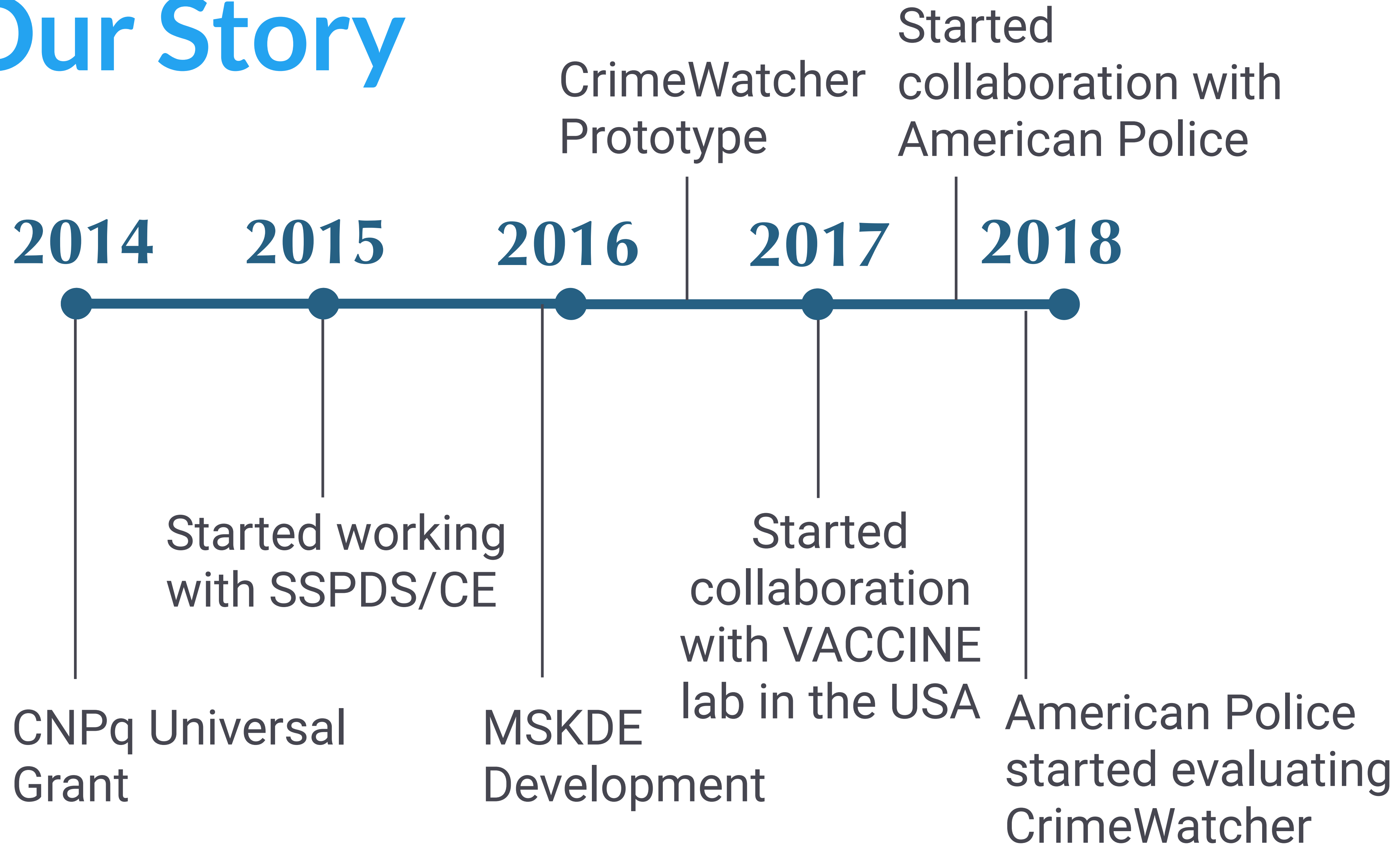
Our Story



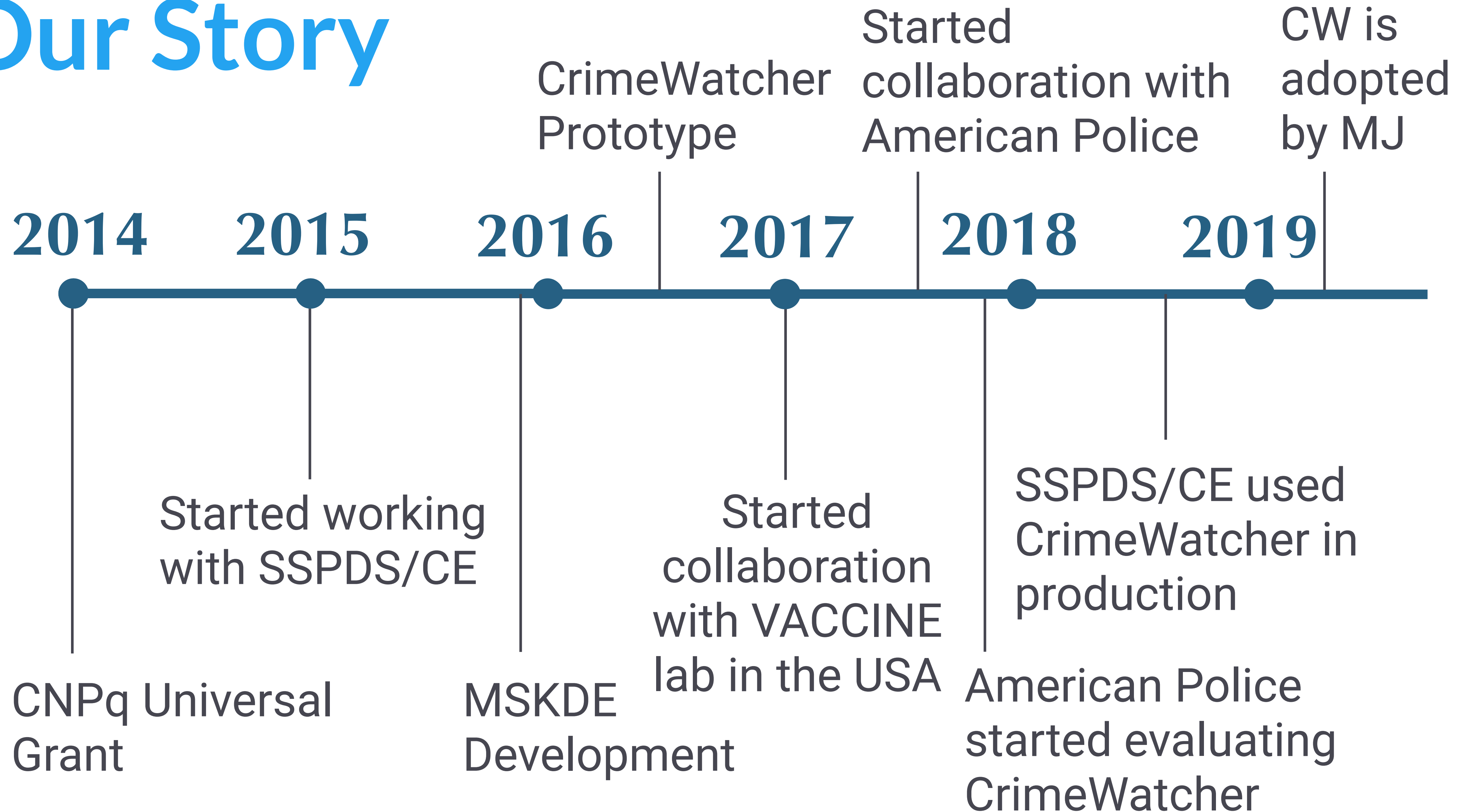
Our Story



Our Story



Our Story



OUR APPROACH

03

Problem description

Crime Analysis

Geographic location is a prime factor in crime.

The police usually adopts analytic systems that use spatial hotspot maps to predict crime and to plan patrols.

Practitioners and researchers consider KDE the most prominent technique for generating hotspot maps.

American Requirements

American Police are accustomed to use visual analytics tools that include hotspot analysis.

However, the delivered solutions did not fully meet their requirements.

Brazilian Requirements

SSPDS routinely use standard GIS to create static hotspot maps by applying KDE to a historical data set of events.

The process of generating and distributing the maps was admittedly inefficient:

A small team of statisticians helps 22 regiments of 500 police officers each to patrol 20 designated areas in the state.

The process to produce hotspot analyses using GIS tools takes a long time.

Goals

During the first two years of collaboration, we realized that crime hotspot analysis would **benefit from a tool that enabled spatial comparisons.**

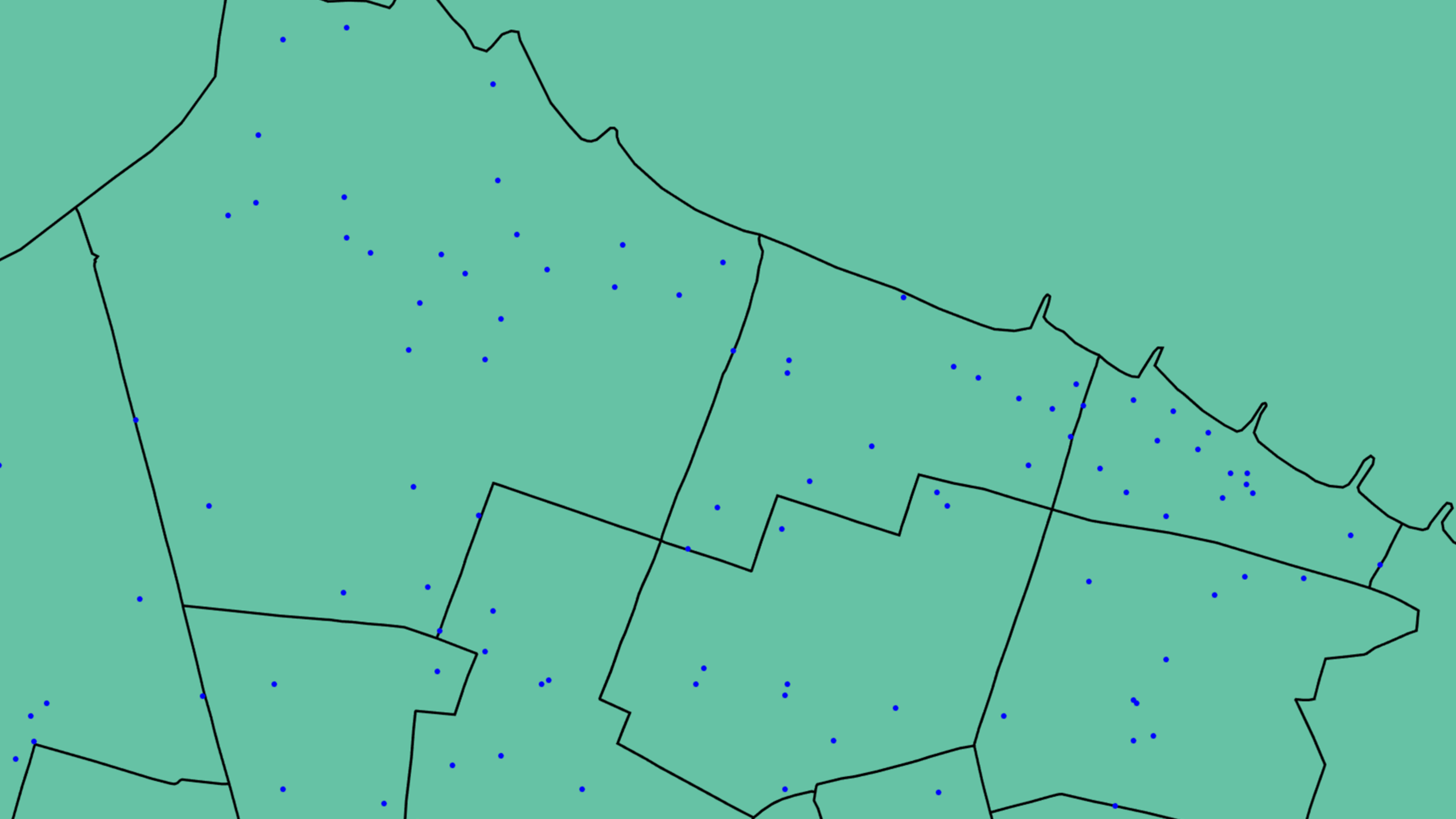
We also recognized that **such tool should be as simple as possible**, in order for non-technical users to perform hotspot analysis tasks easily.

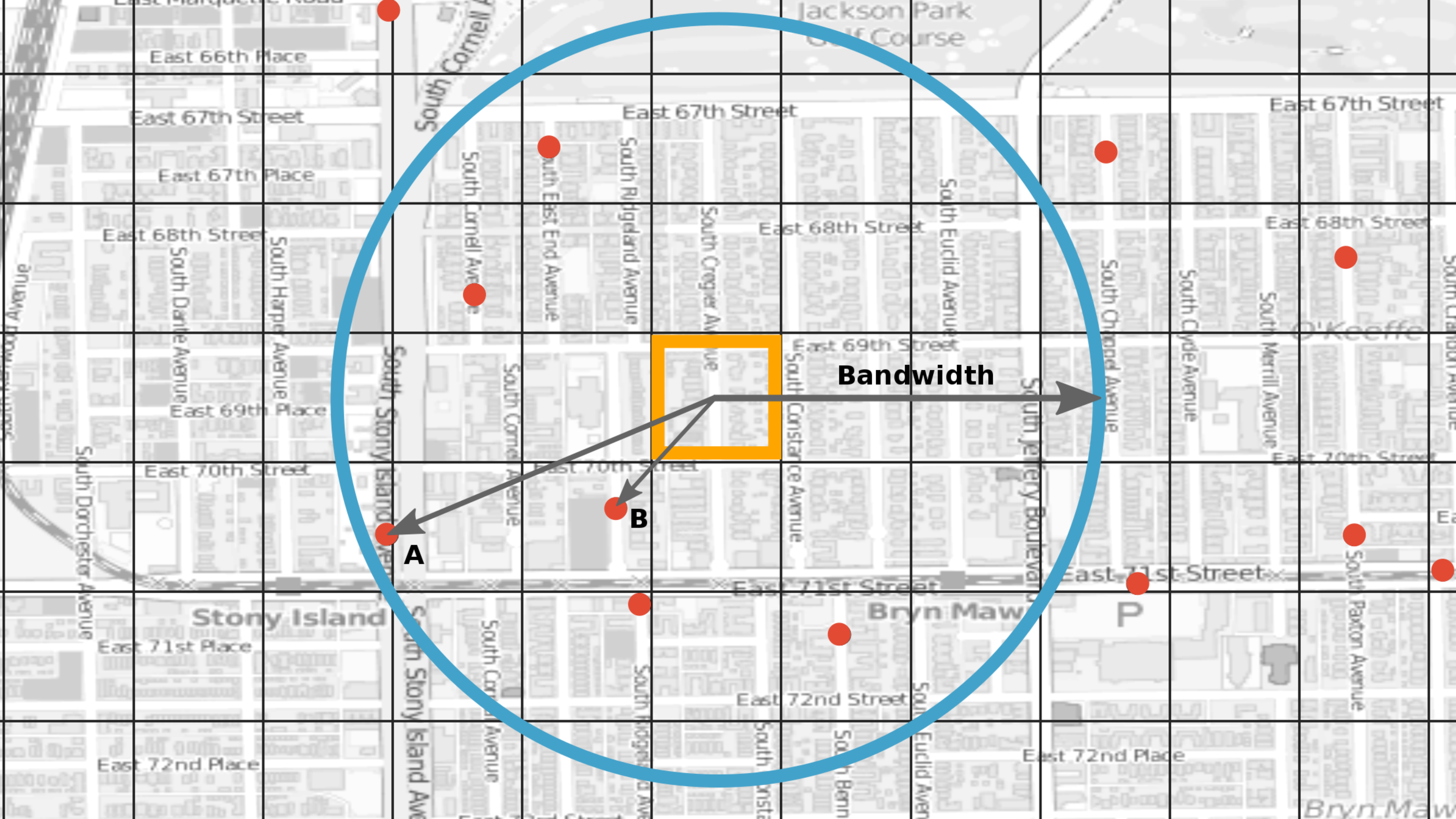
Goals

During the first two years of collaboration, we realized that crime hotspot analysis would **benefit from a tool that enabled spatial comparisons.**

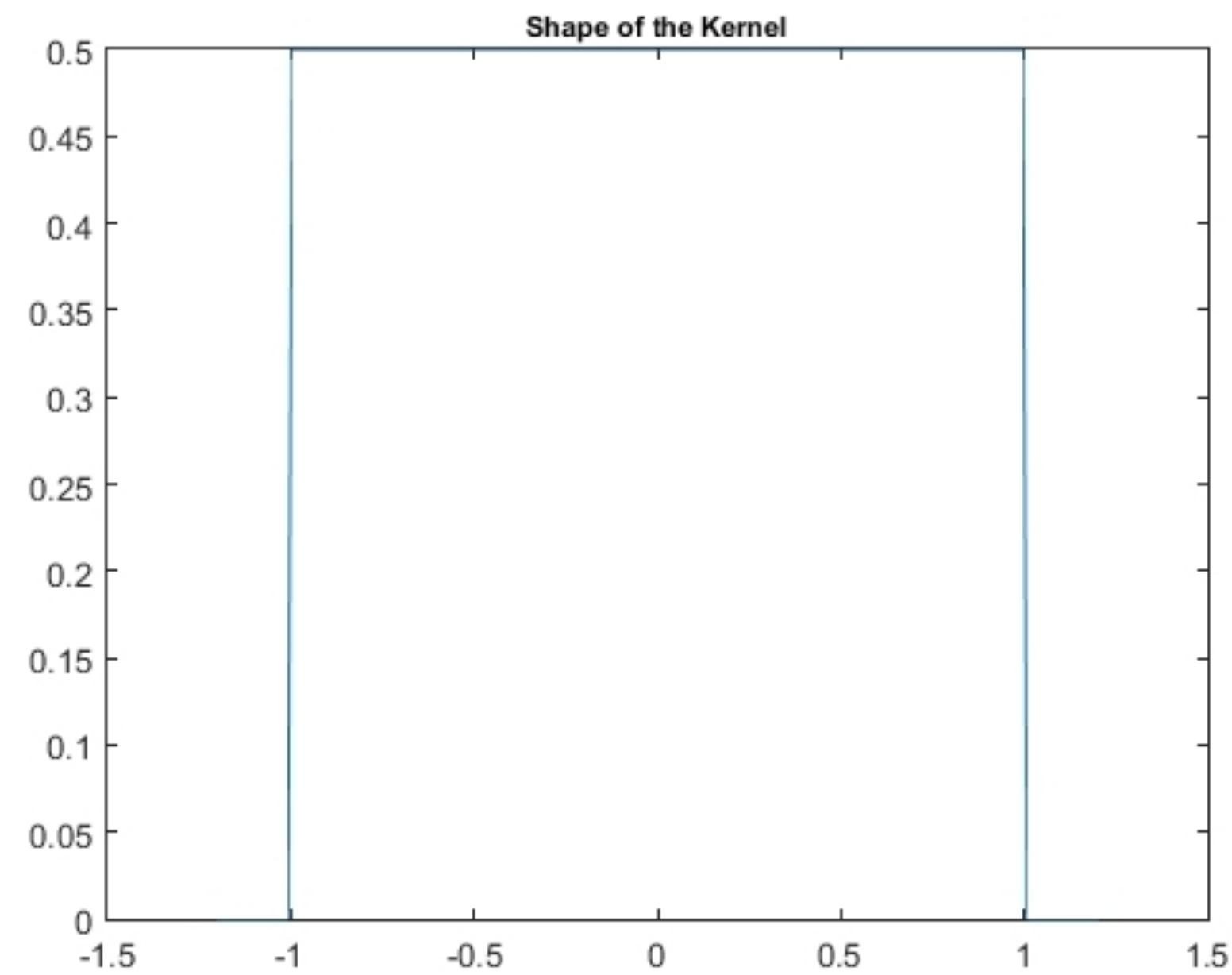
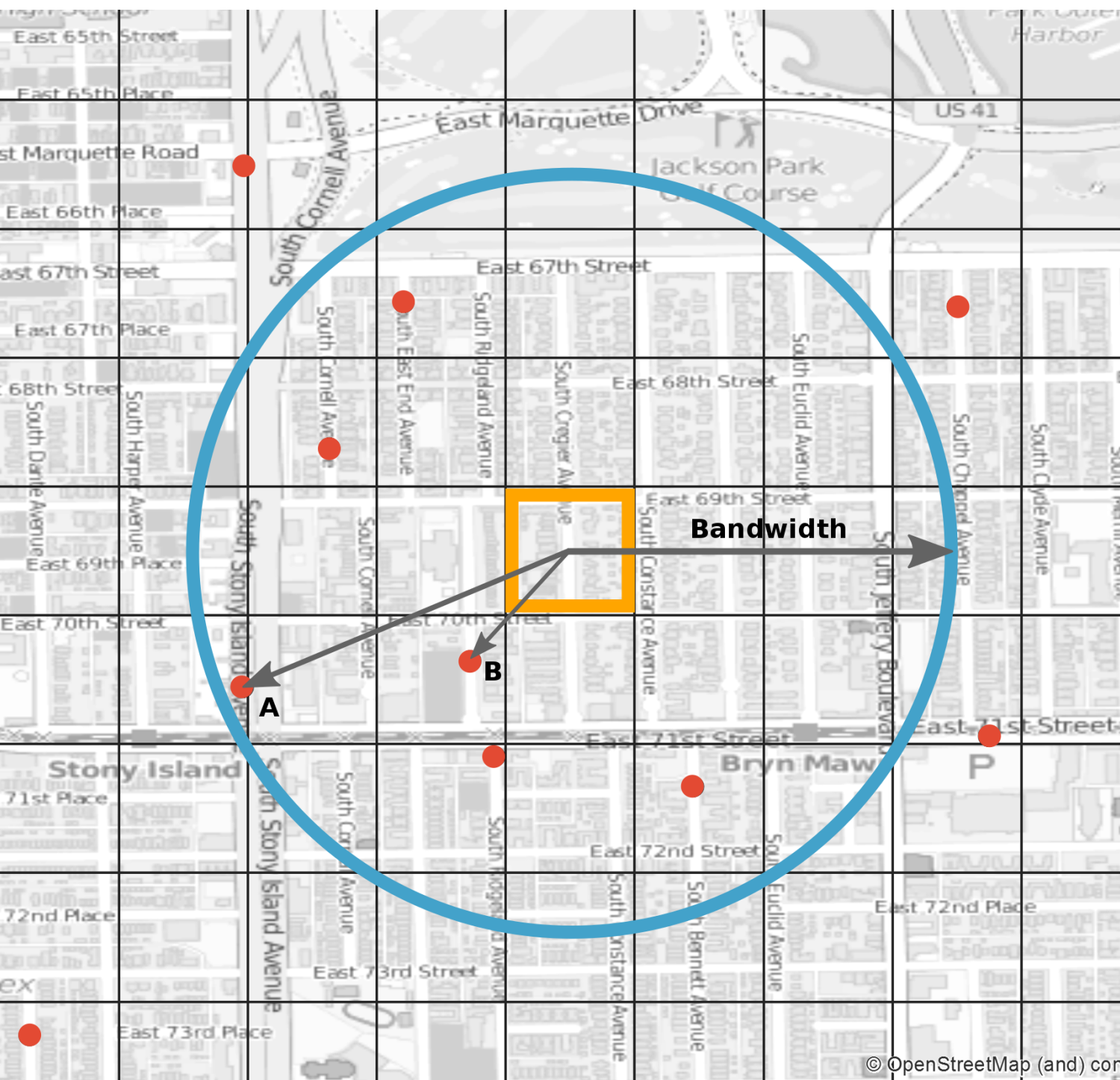
We also recognized that **such tool should be as simple as possible**, in order for non-technical users to perform hotspot analysis tasks easily.

Hotspot maps using KDE

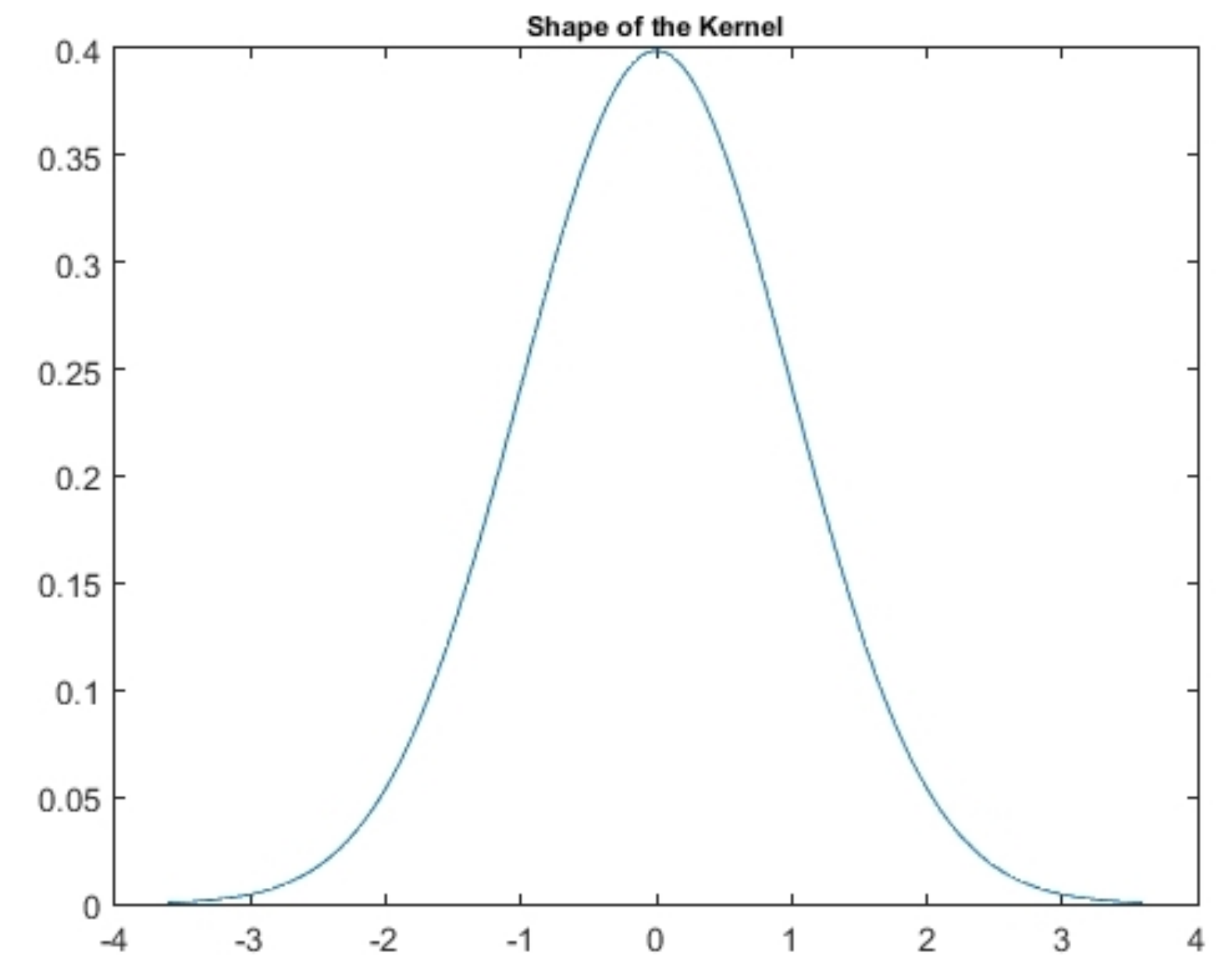




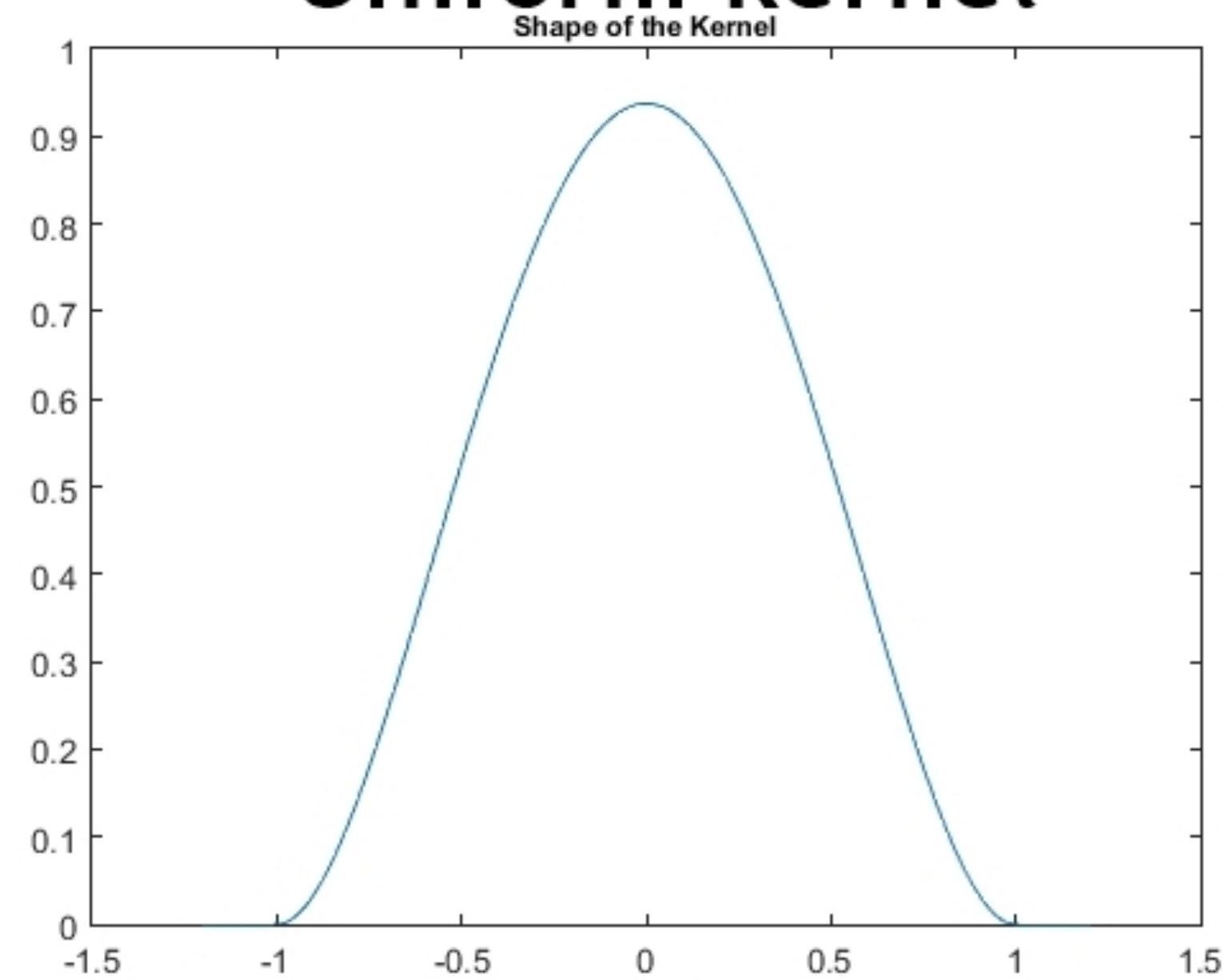
Kernel functions



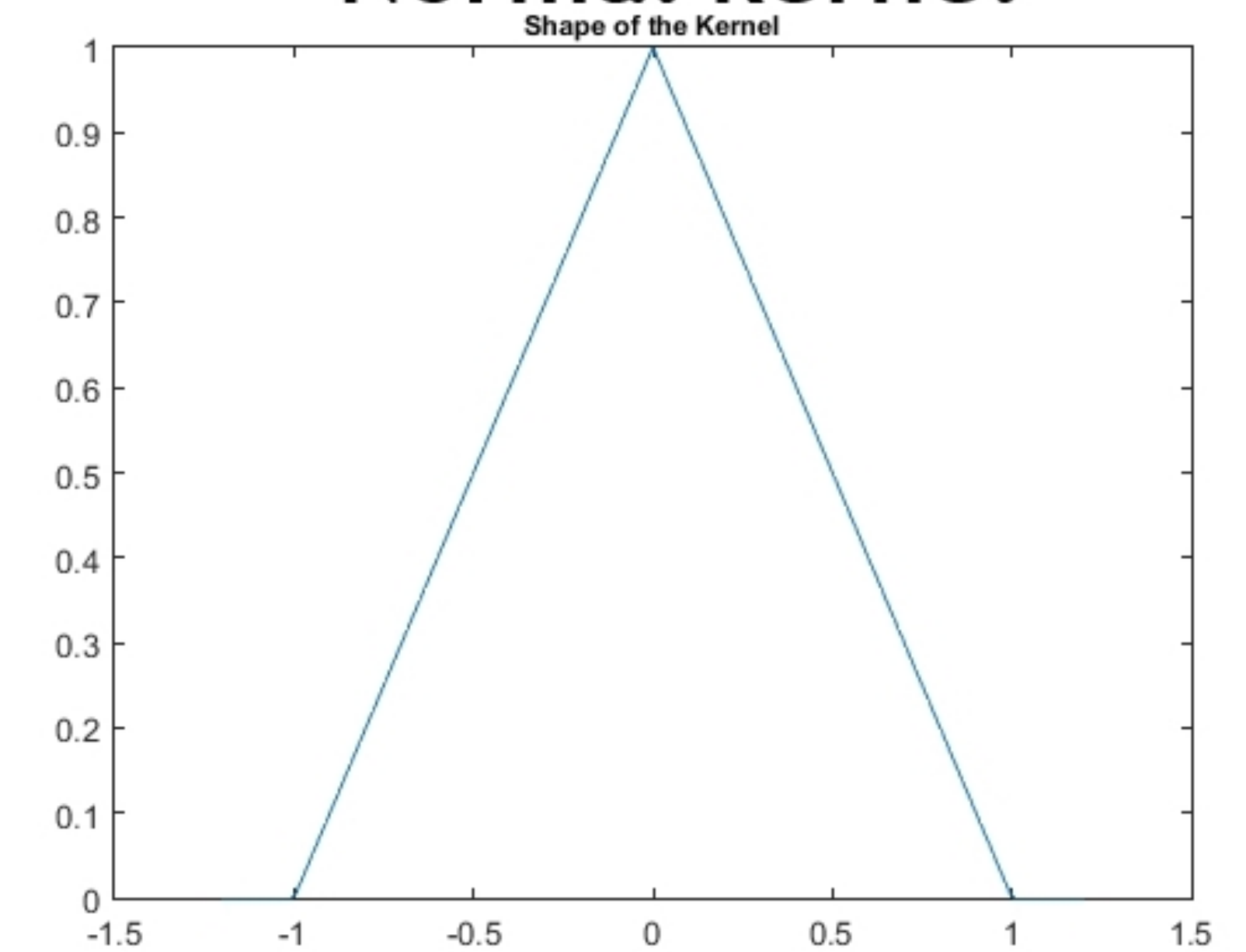
Uniform kernel



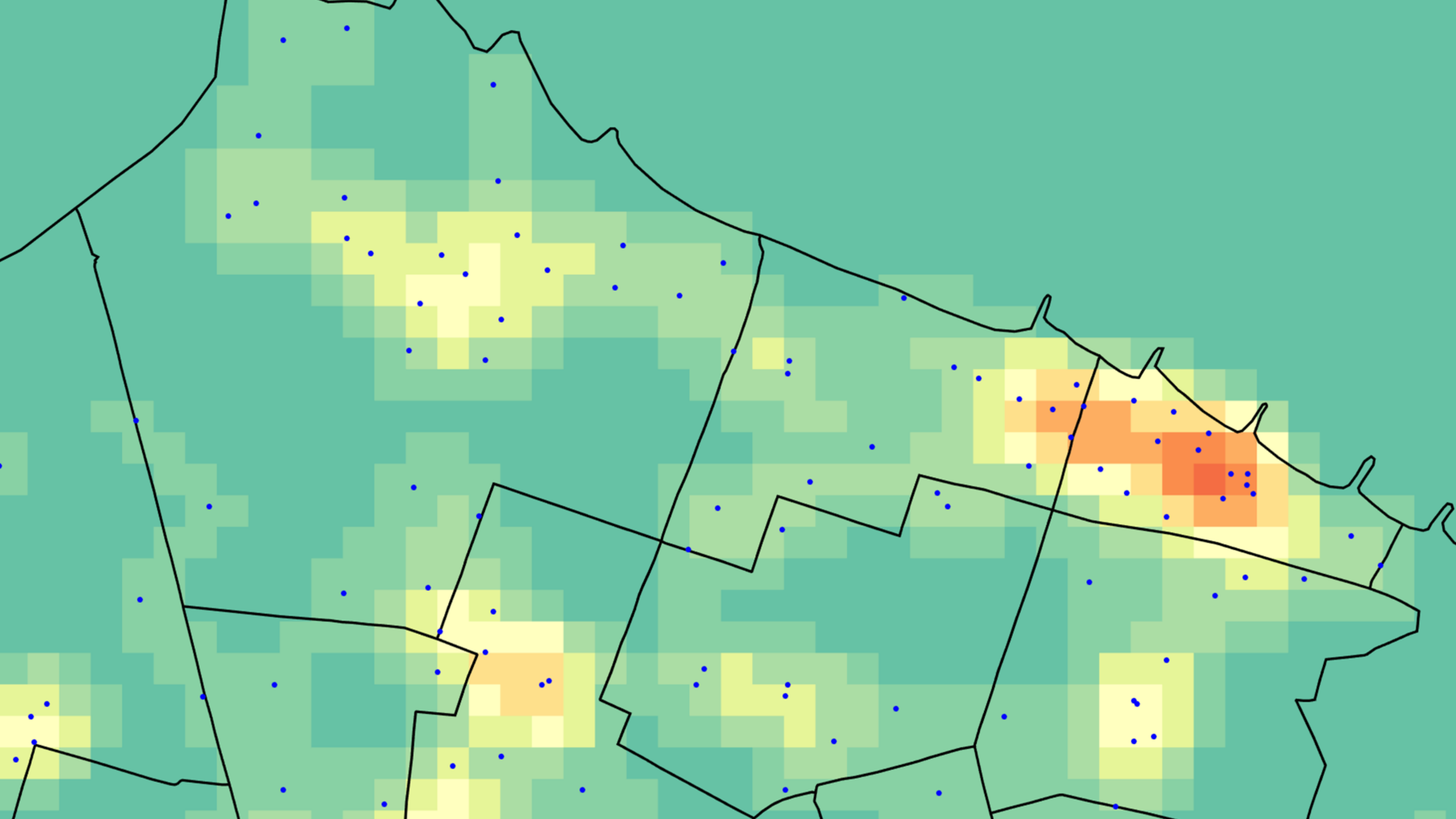
Normal kernel



Quartic kernel



Triangular kernel



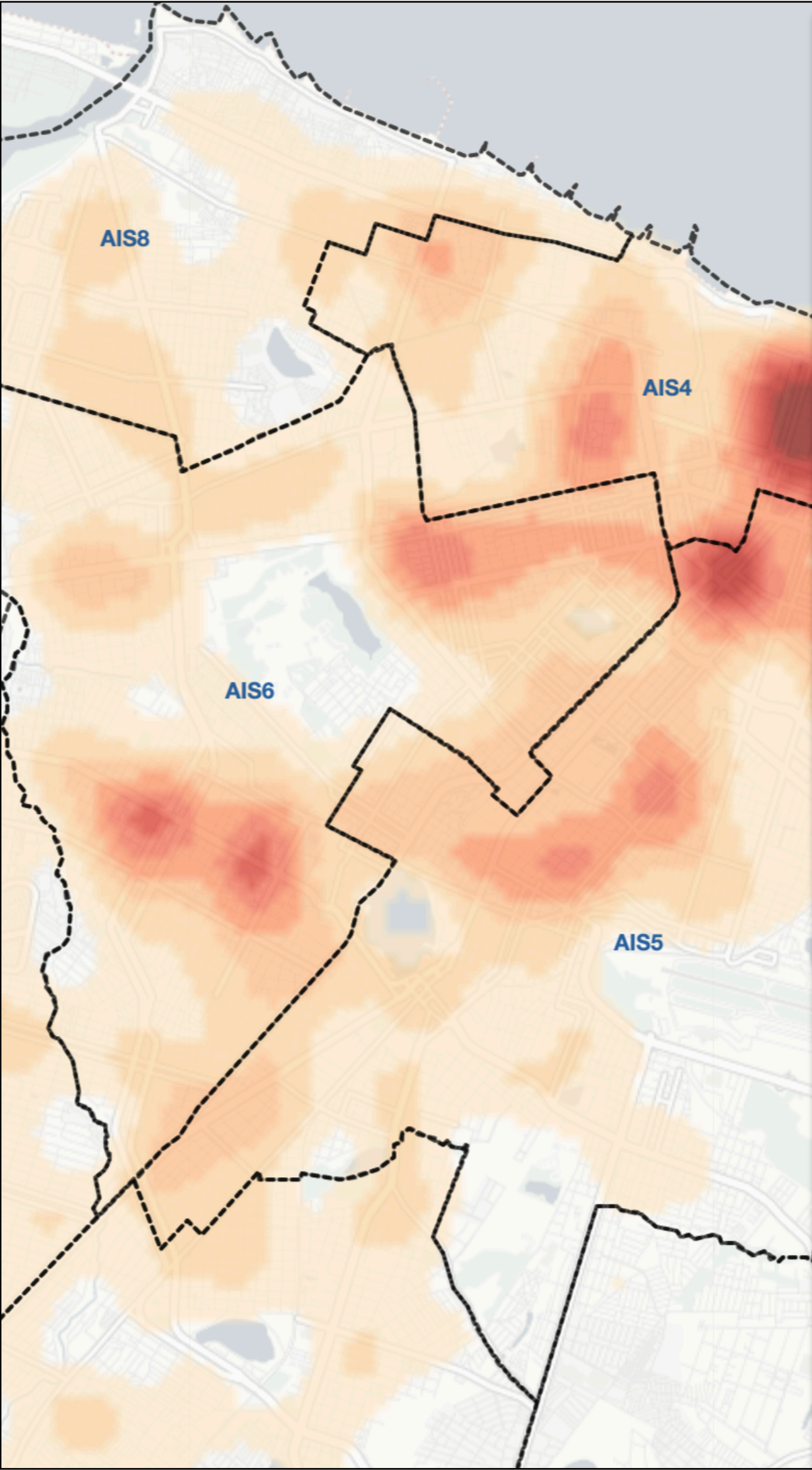
The problems with KDE

It is very difficult to see more than one KDE map in the same region, without losing information or having too much clutter.

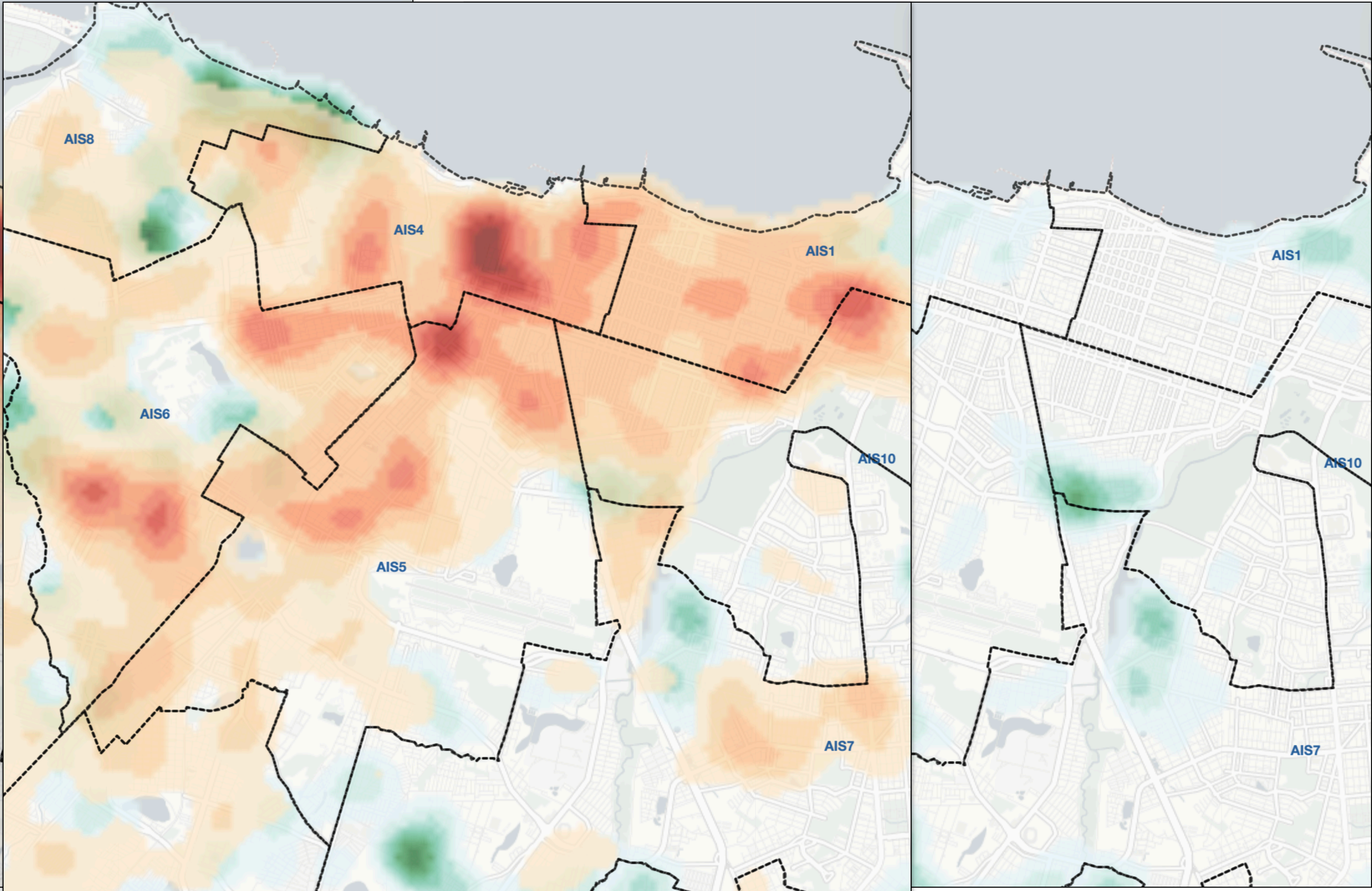
Producing a hotspot map using KDE at interactive speed may require not so small cell sizes:

Result is a pixelated and not accurate hotspot map

CAP in Fortaleza in 2015



CAL in Fortaleza in 2015



Marching Squares KDE - MSKDE

MSKDE

[Queiroz Neto et al., 2016].

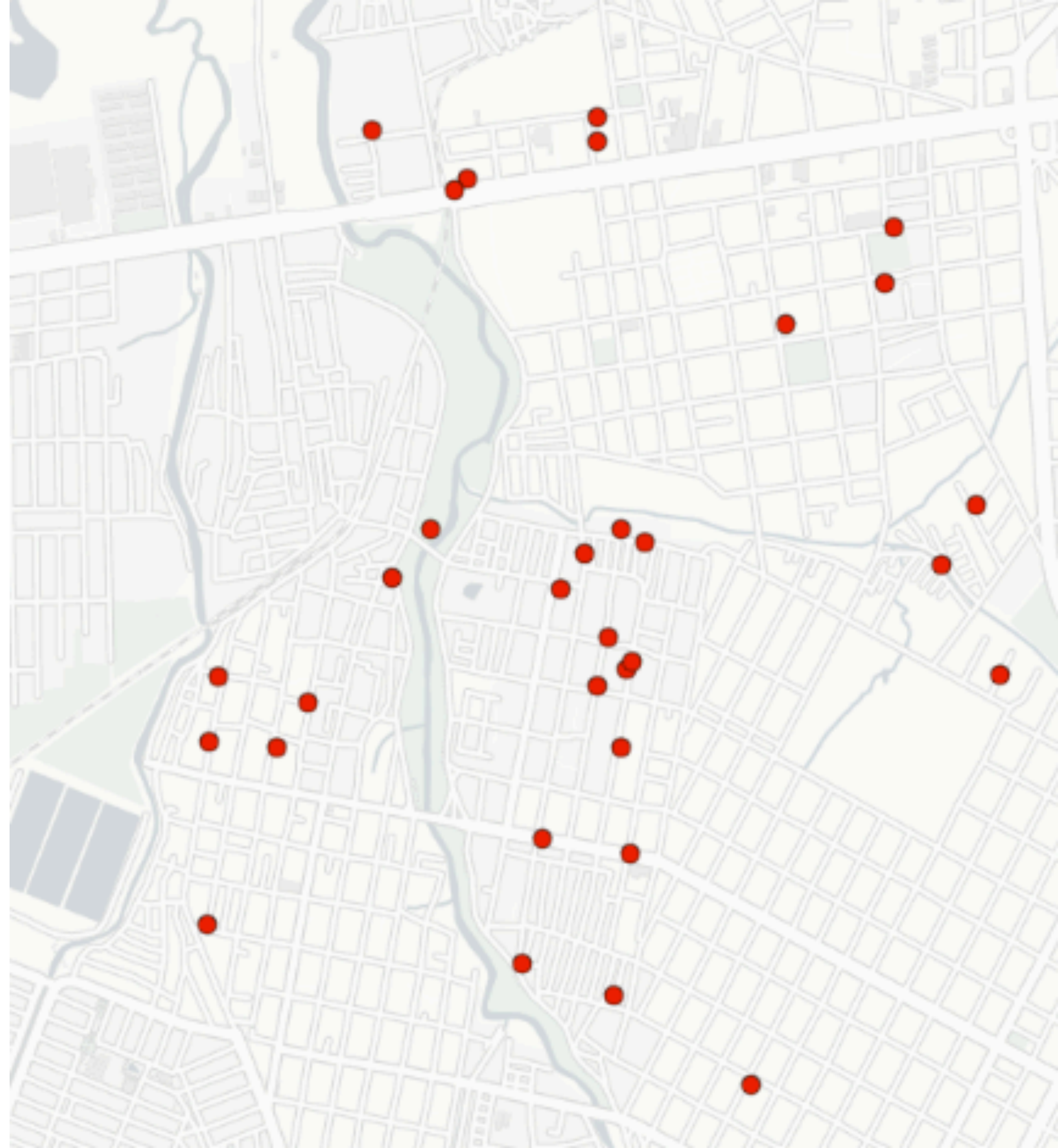
It is a fast technique to generate a hotspot map based on contour lines extracted from a KDE map.

It uses Marching Squares, a very efficient algorithm for generating contour lines in a regular grid representing a two-dimensional scalar field.

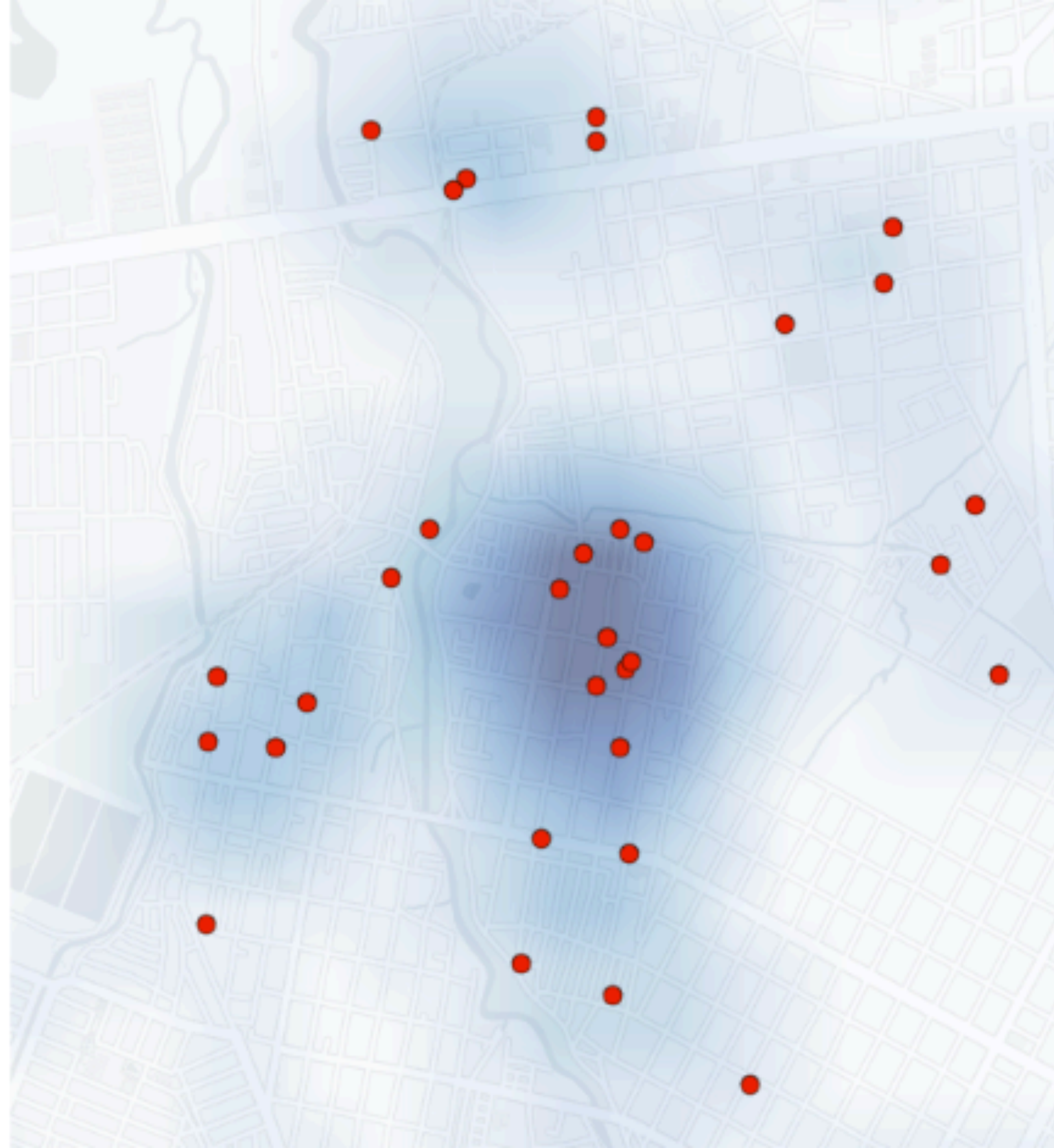
The result is a hotspot map with rounded and smooth boundaries, more accurate than the KDE hotspot map of the same cell size.

Generating an MSKDE map

Start with a set of crime events.



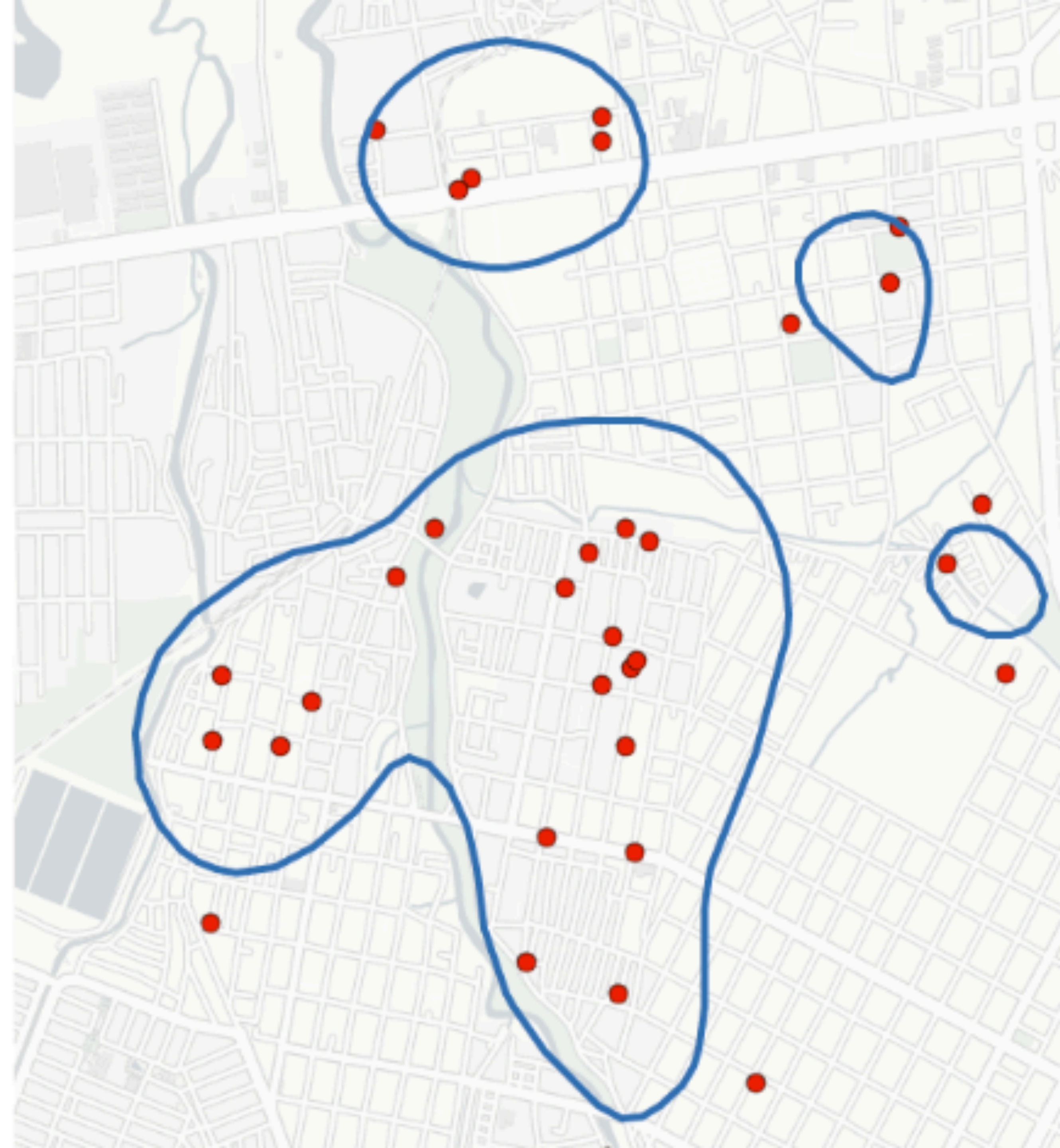
Generate a KDE map for
input set.



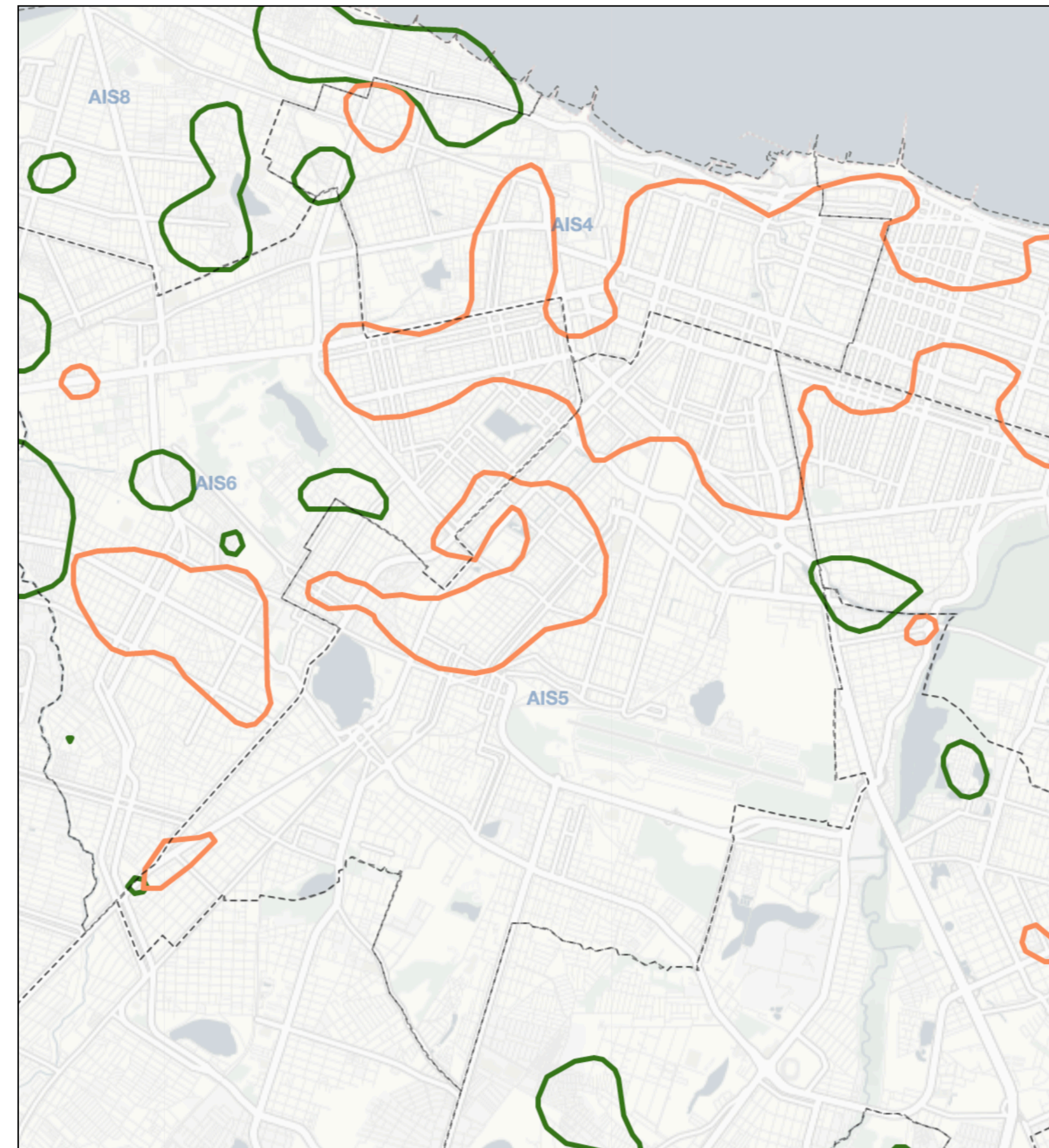
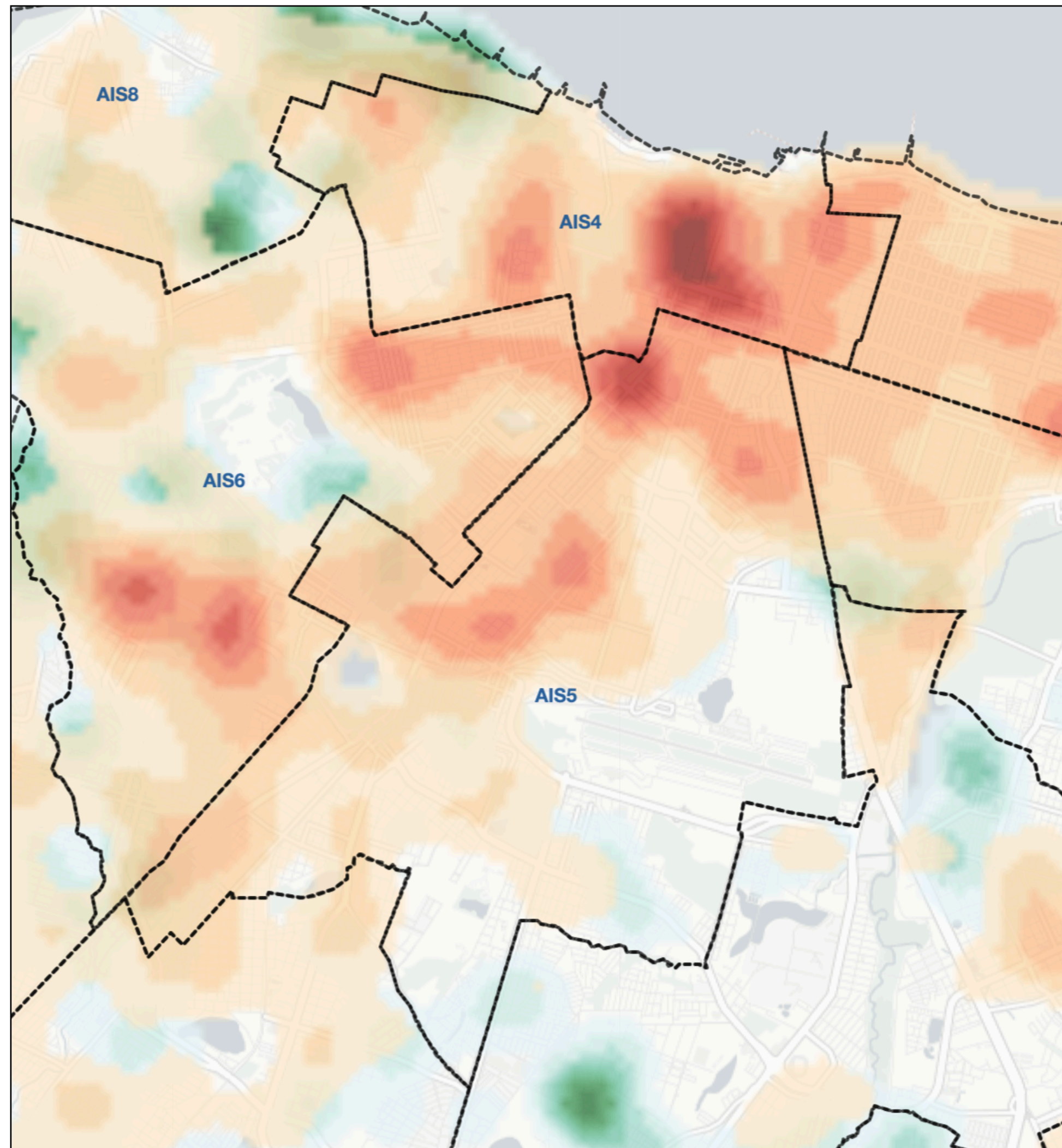
Run marching squares algorithm over the KDE scalar field.



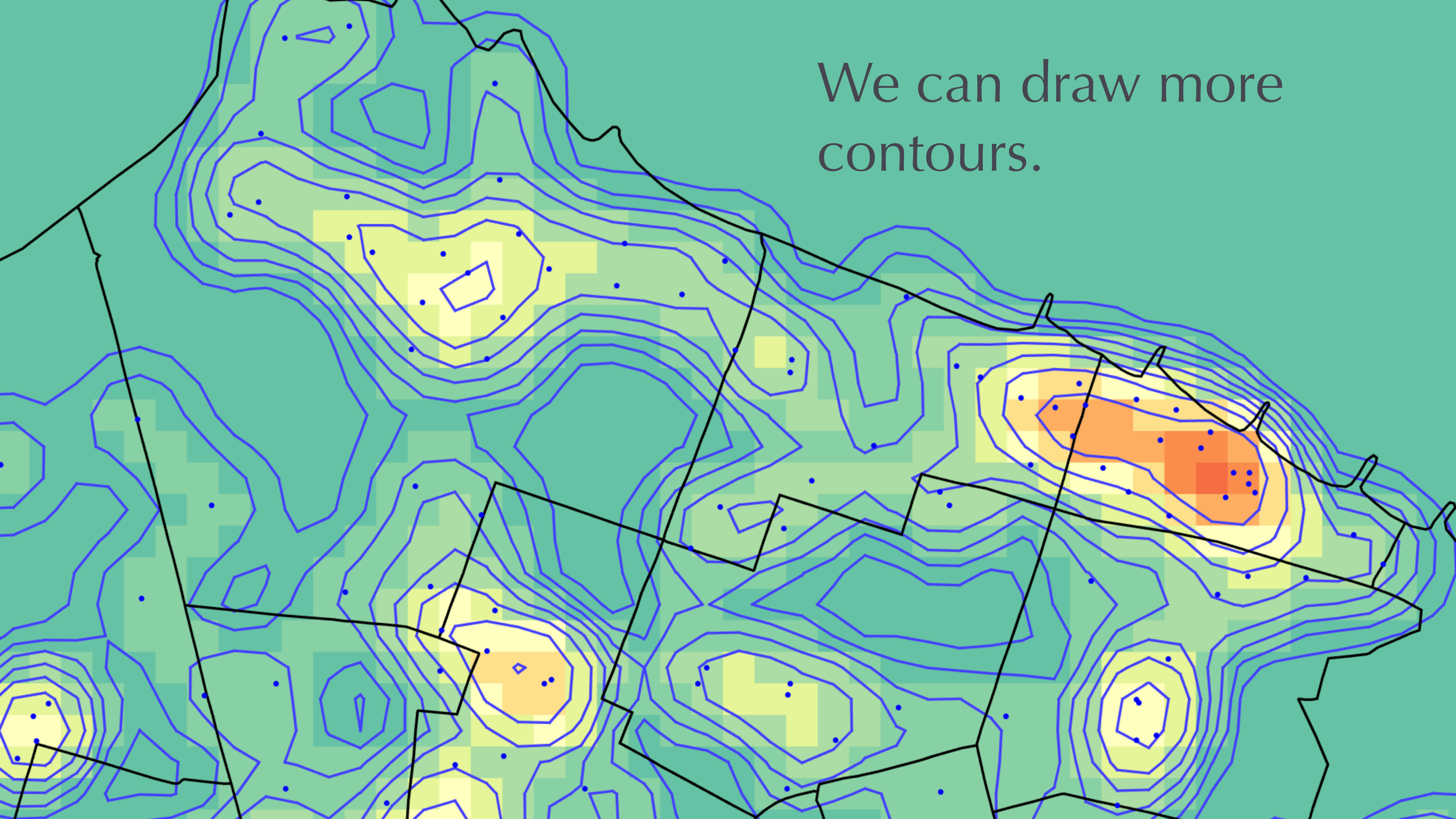
The final *MSKDE* map is obtained by removing the KDE scalar field.



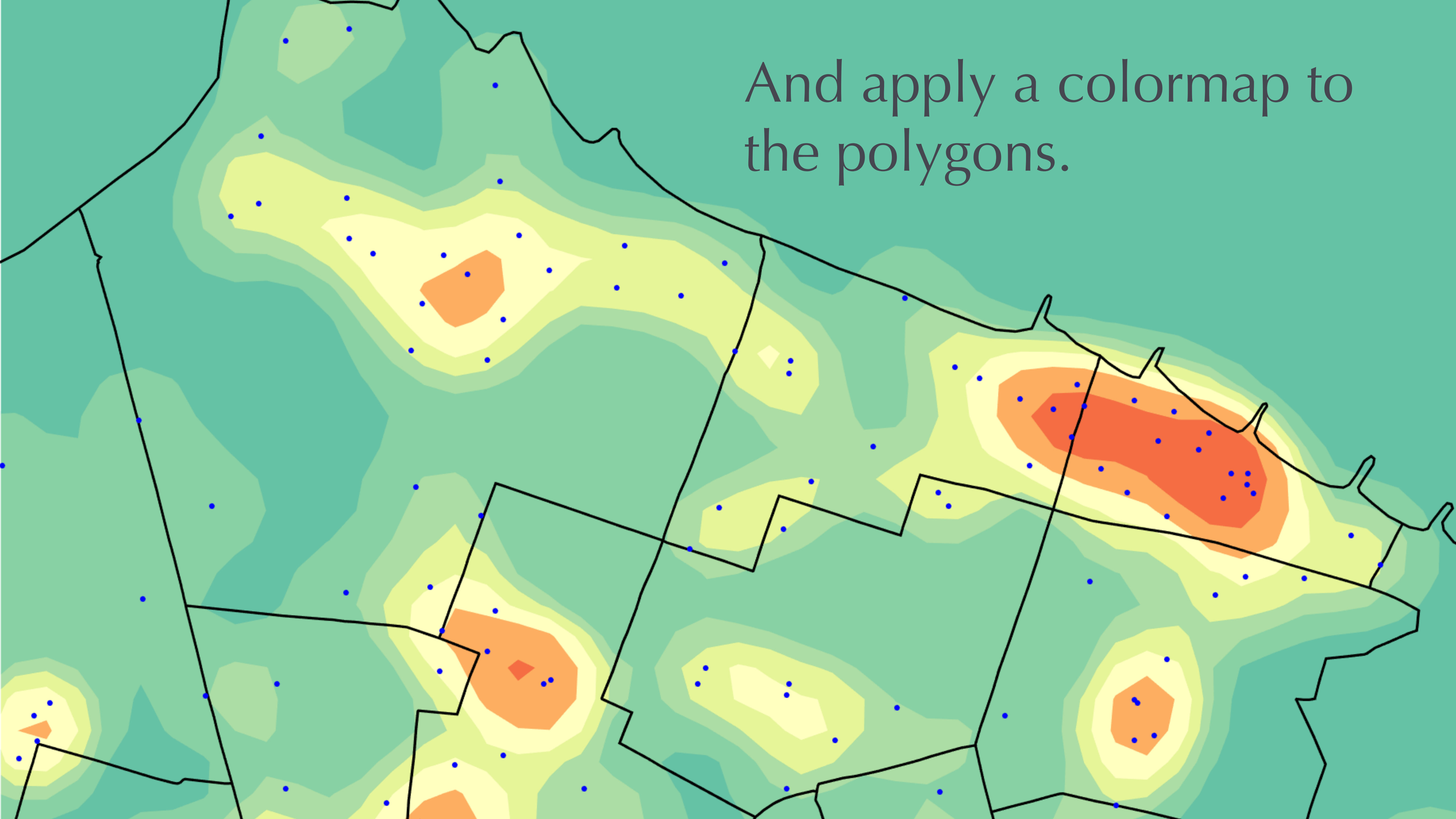
CAP and **CAL** in Fortaleza in 2015



We can draw more
contours.

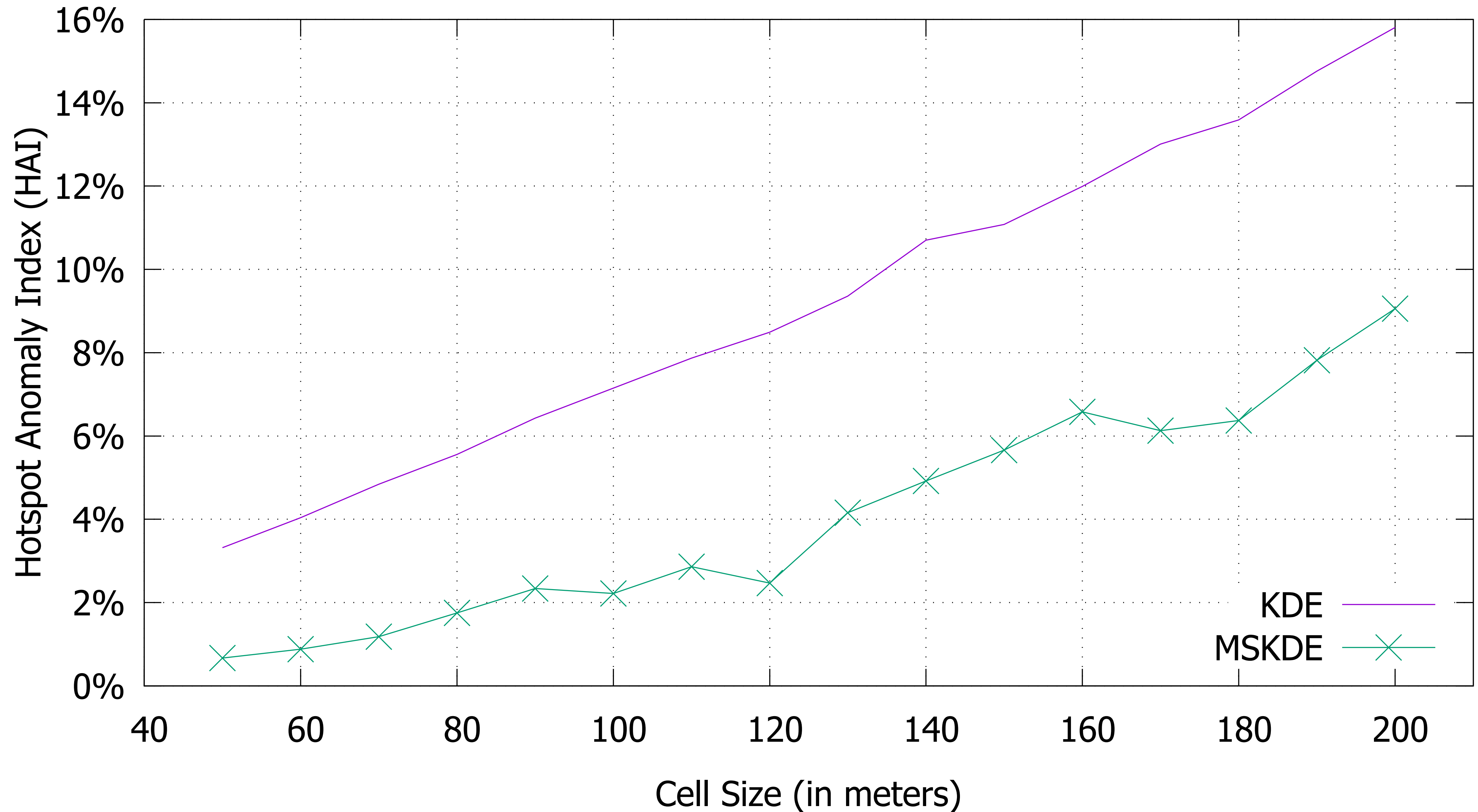


And apply a colormap to
the polygons.



MSKDE Accuracy

Anomaly x Cell Size for KDE and MSKDE hotspot maps



KDE			MSKDE		
Cell Size	Time	HAI	Cell Size	Time	HAI
50	31.68	3.32%	120	5.17	2.47%
60	18.37	4.04%	130	4.02	4.16%
70	14.05	4.84%	140	3.76	4.92%
80	10.47	5.56%	150	3.10	5.66%
90	9.01	6.43%	160	2.76	6.58%
100	6.94	7.15%	170	2.25	6.13%
110	5.74	7.87%	180	2.17	6.37%
120	5.04	8.49%	190	2.02	7.82%
130	3.93	9.36%	200	1.97	9.06%

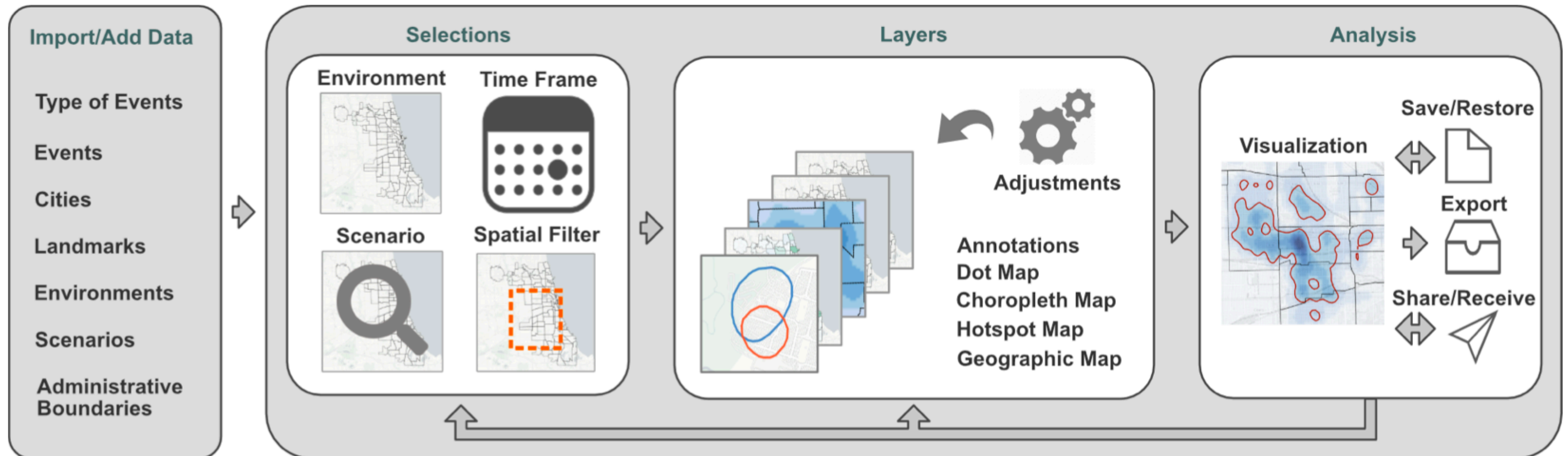
Goals

During the first two years of collaboration, we realized that crime hotspot analysis would **benefit from a tool that enabled spatial comparisons.**

We also recognized that **such tool should be as simple as possible**, in order for non-technical users to perform hotspot analysis tasks easily.

CrimeWatcher

CrimeWatcher's Workflow



- Events
- Landmarks
- Administrative Boundaries
- Kernel Density Estimation
- MSKDE

☒ Create MSKDE ☐ Use same threshold of

Cell Size:

Bandwidth:

Density Percentage:

☒ Integral ☐ Area

Opacity Border:

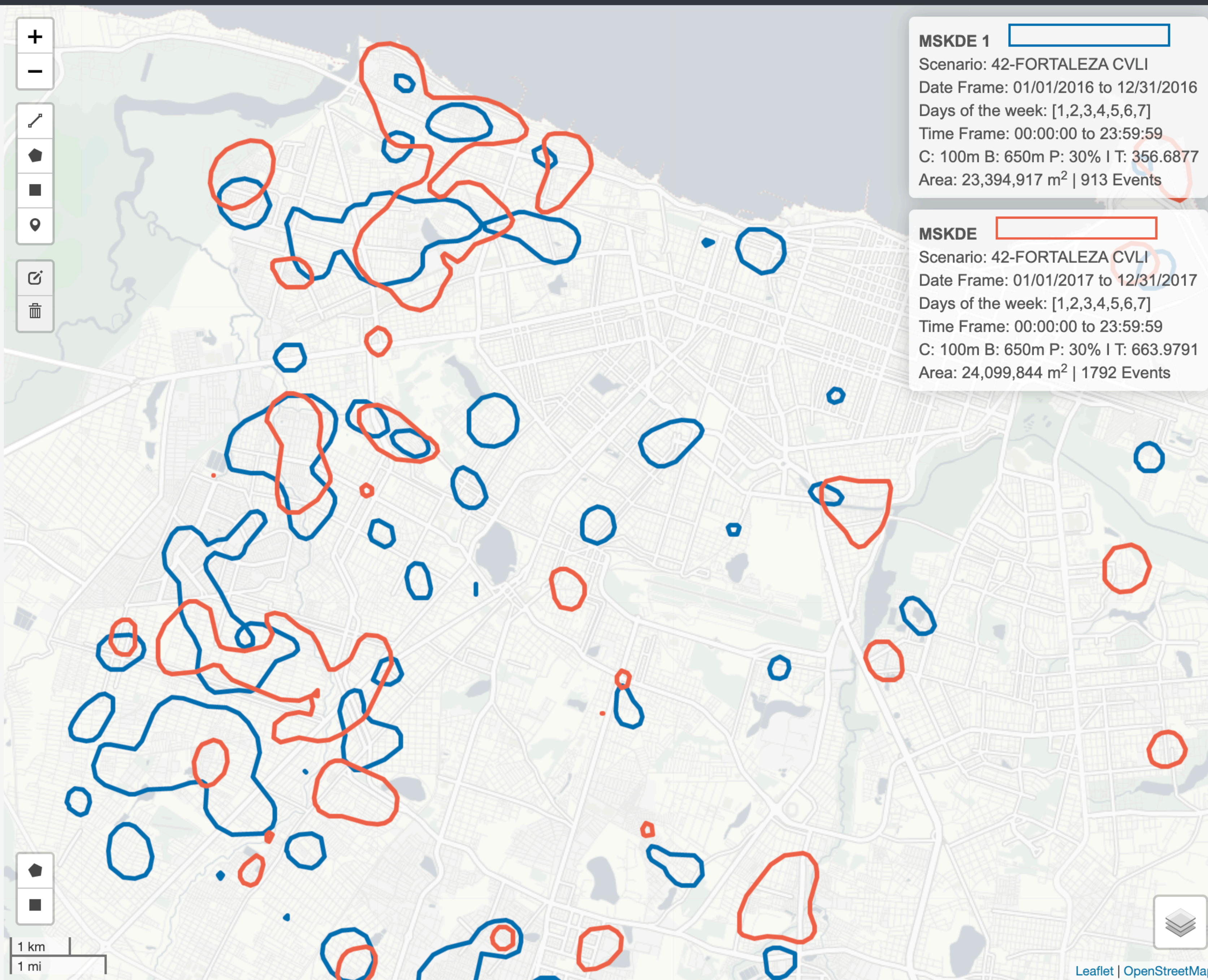
Opacity Fill:

Border: Weight: Fill:

☒ MSKDE Legend

Animation

Spatial Filters



Save Analysis

Scenario Selection

Time Selection

Events

Landmarks

Administrative Boundaries

Kernel Density Estimation

MSKDE

Spatial Filters

Annotations

Layers

[Refresh List](#)

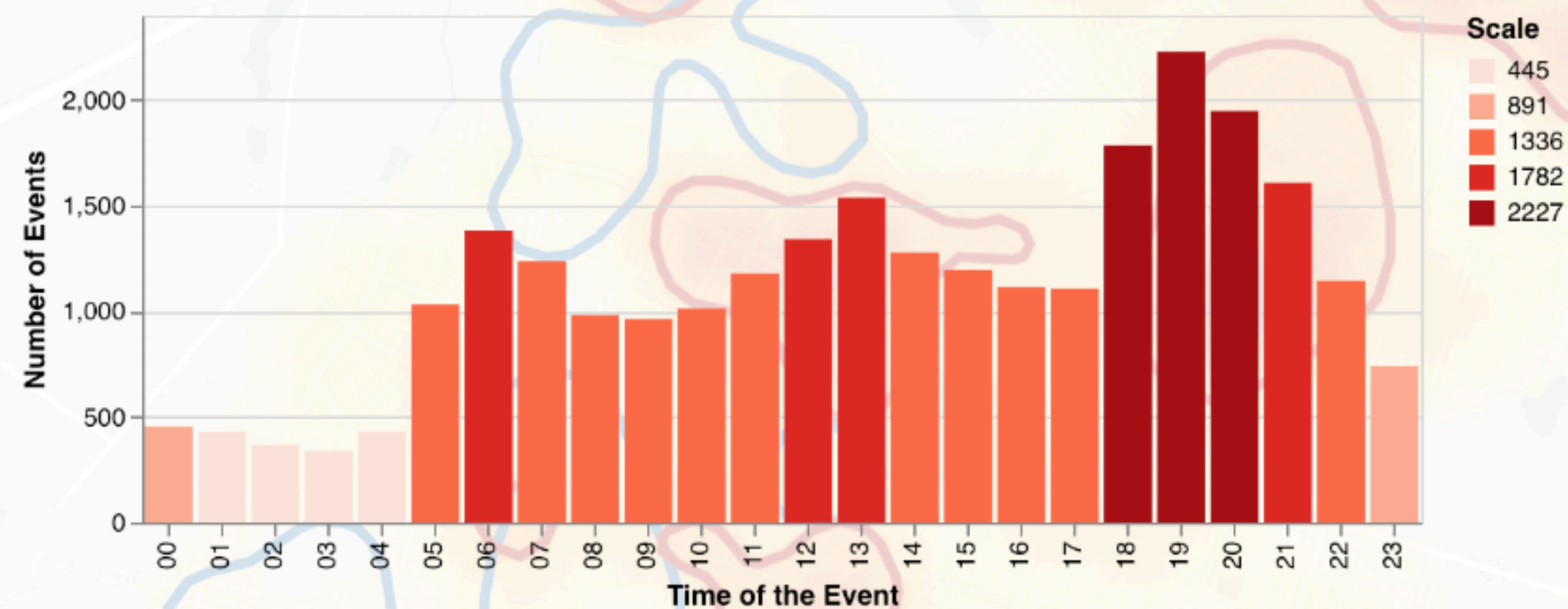
Remove Layer

1	Annotations
2	Events
3	Land Marks
4	CAL
5	CAP

Base



Distribution of Events per Time:



Scenario: 64-CAP
Date Frame: 01/01/2018 to 11/30/2018
Time Frame: 00:00:00 to 23:59:59
Events: 26827



1 km

1 mi

CAL

Scenario: 42-CAL

Date Frame: 01/01/2018 to 11/30/2018

Time Frame: 00:00:00 to 23:59:59

Percentage: 30% (Integral)

T: 722.5187 C: 100m B: 800m

Area: 26,863,615 m² | 1347 Events

CAP

Scenario: 64-CAP

Date Frame: 01/01/2018 to 11/30/2018

Time Frame: 00:00:00 to 23:59:59

Percentage: 30% (Integral)

T: 124.884 C: 100m B: 800m

Area: 33,171,296 m² | 26827 Events

☐ Waze

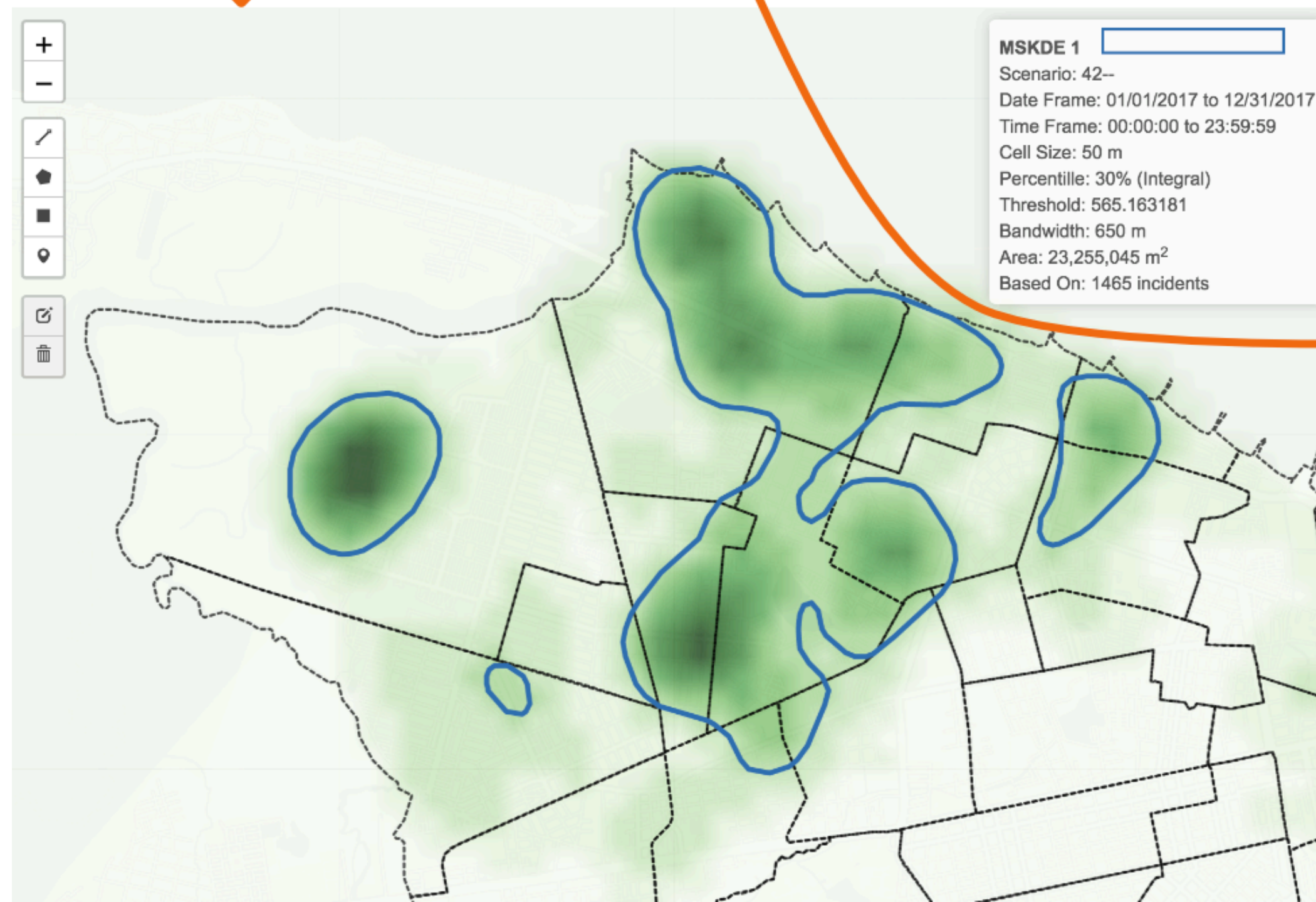
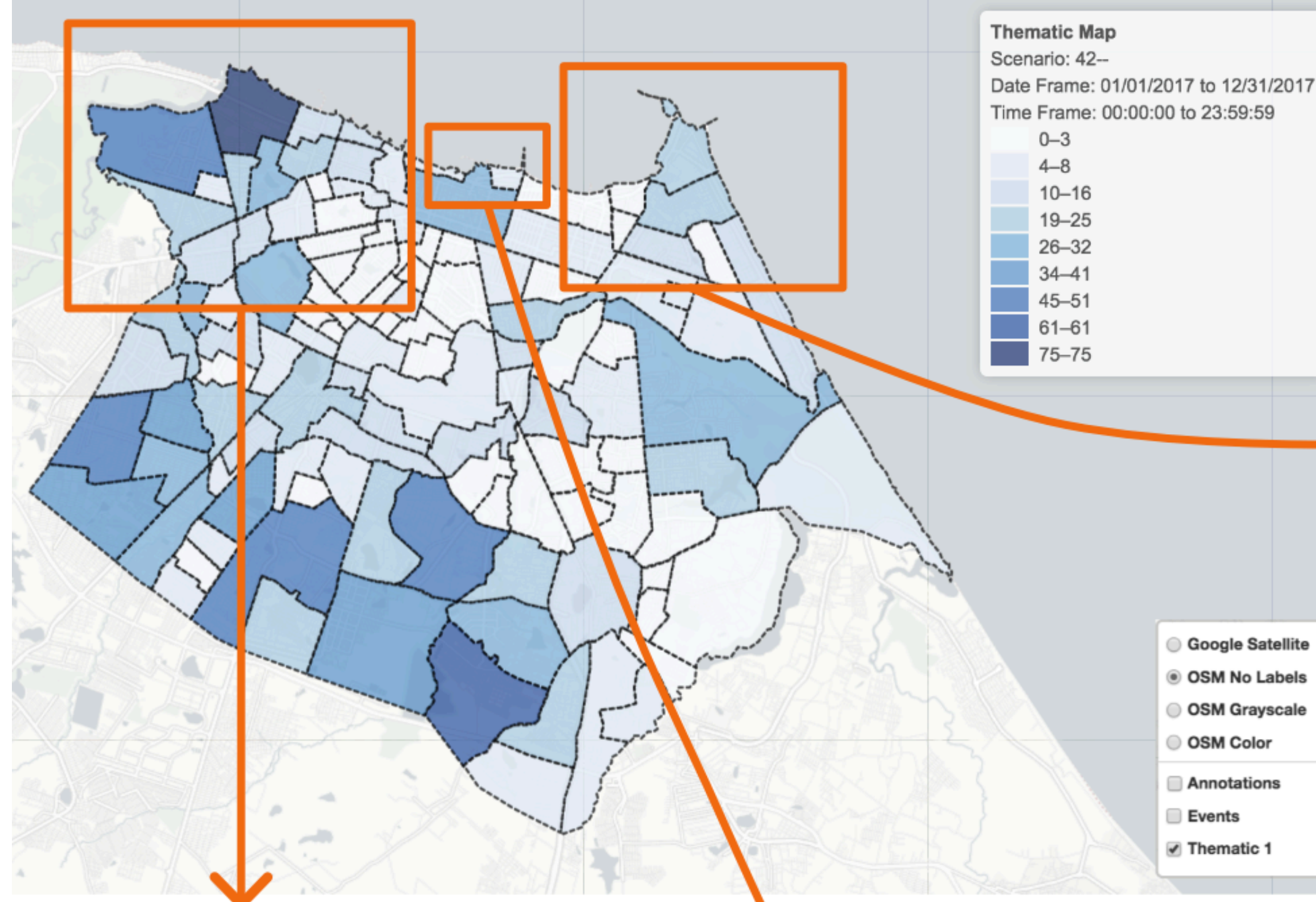
☐ Google Hybrid

☒ OSM No Labels

☐ OSM Grayscale☐ OSM Color

☒ Annotations

☐ Events☒ CAL☒ CAP



Case Study: Drug Abuse in Tippecanoe County, USA (2016-2017)

The collaboration

Lafayette Police Department (LPD) and the Purdue Police Department (PUPD).

8 meetings between November 16, 2017, and May 23, 2018.

The police knew that the recorded number of drug abuse incidents was increasing, and they wanted to know more about how and where that increase was taking place.

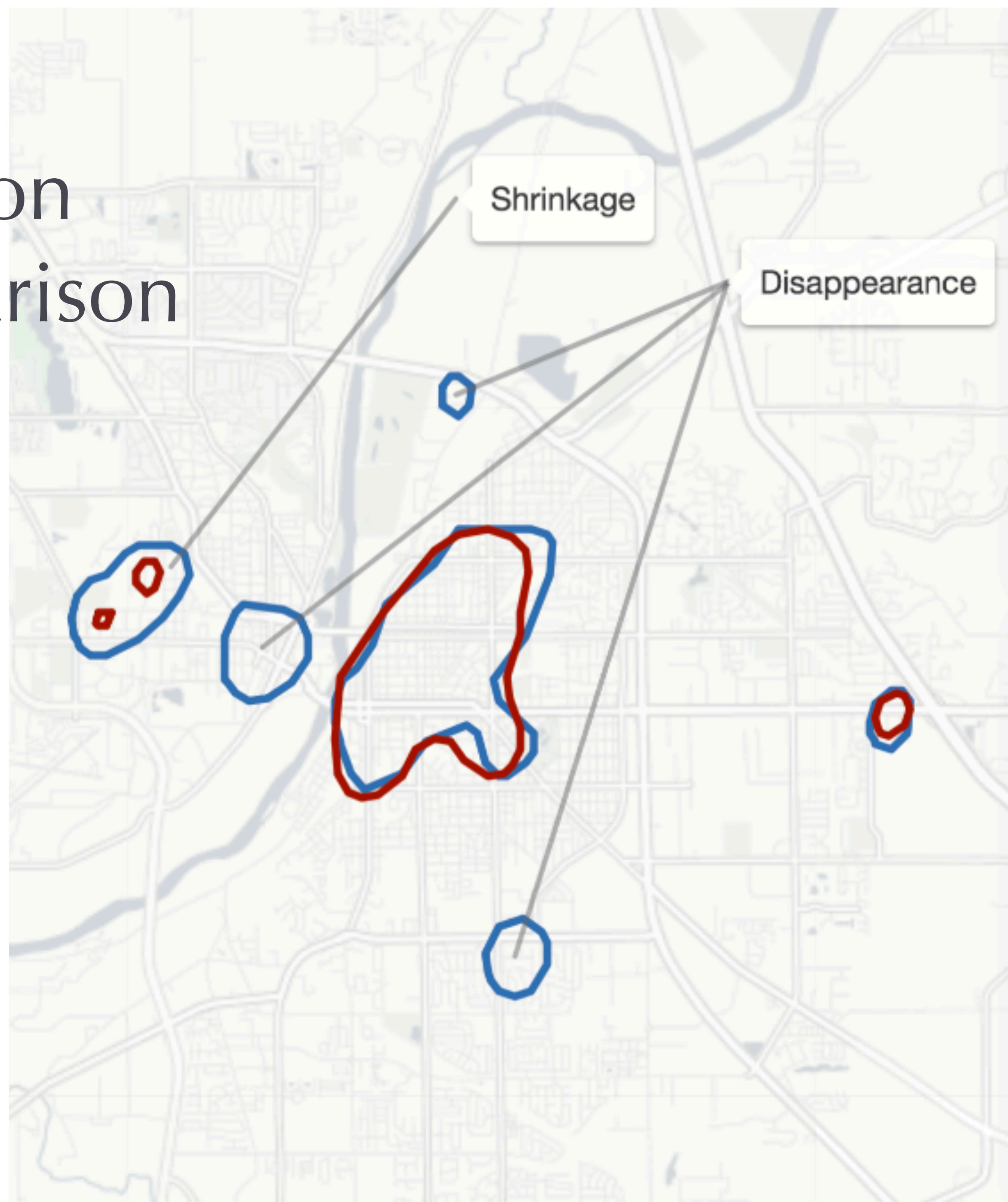
The crime dataset

Incidents recorded from January 1, 2016, to December 31, 2017 were separated into 2 partitions:

Partition 1: 1,288 incidents of 2016.

Partition 2: 2,129 incidents of 2017.

Hotspot identification and comparison



MSKDE 1

Scenario: 55-LPD DRUG ALL

Date Frame: 01/01/2016 to 12/31/2016

Time Frame: 00:00:00 to 23:59:59

Cell Size: 100 m

Percentage: 30% (Integral)

Threshold: 14.100953

Bandwidth: 650 m

Area: 6,593,499 m²

Based On: 1260 incidents

MSKDE 2

Scenario: 55-LPD DRUG ALL

Date Frame: 01/01/2017 to 12/31/2017

Time Frame: 00:00:00 to 23:59:59

Cell Size: 100 m

Percentage: 30% (Integral)

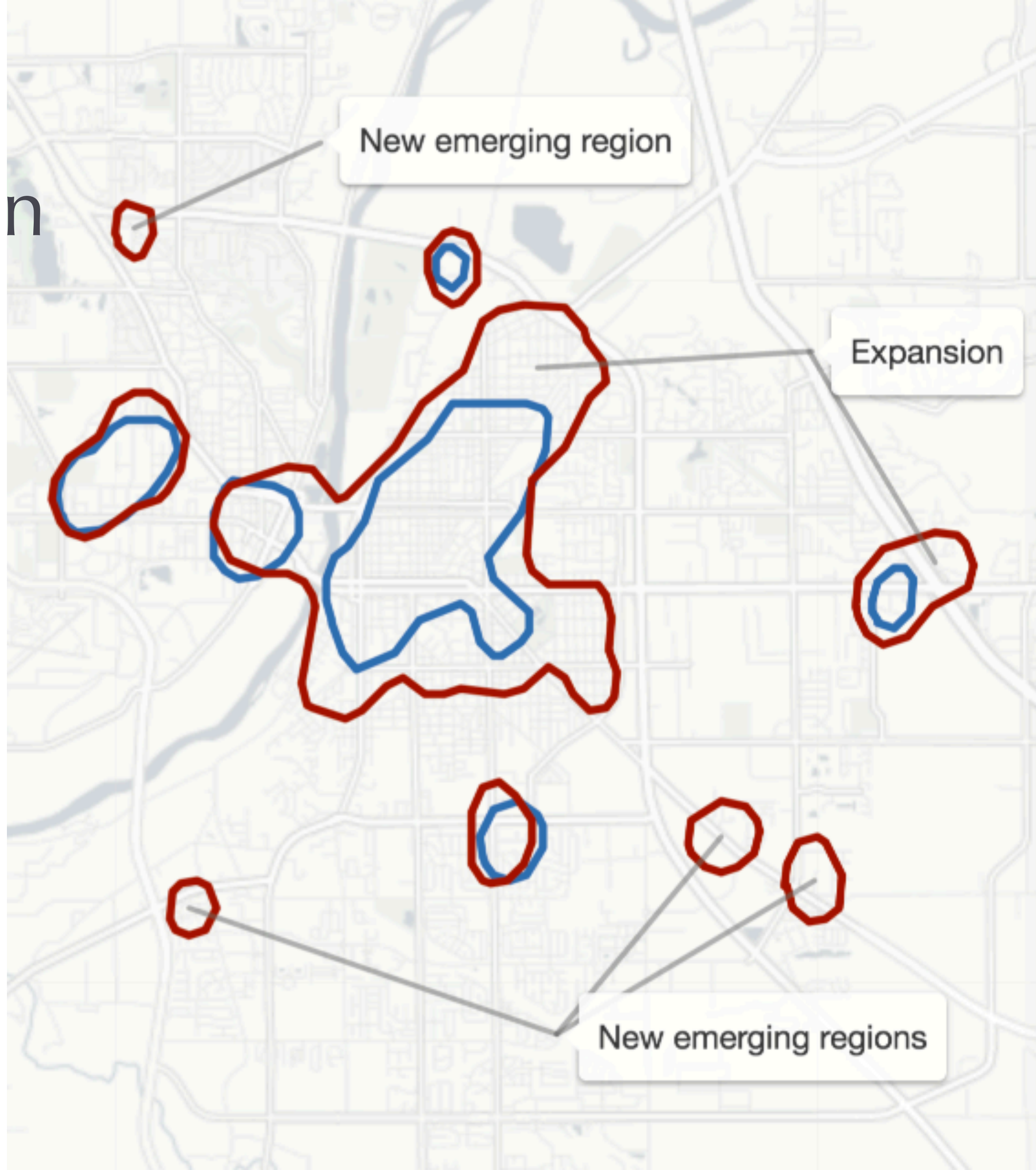
Threshold: 30.635459

Bandwidth: 650 m

Area: 4,350,281 m²

Based On: 2139 incidents

Hotspot evolution



MSKDE 1

Scenario: 55-LPD DRUG ALL

Date Frame: 01/01/2016 to 12/31/2016

Time Frame: 00:00:00 to 23:59:59

Cell Size: 100 m

Percentage: 30% (Integral)

Threshold: 14.100953

Bandwidth: 650 m

Area: 6,593,499 m²

Based On: 1260 incidents

MSKDE 2

Scenario: 55-LPD DRUG ALL

Date Frame: 01/01/2017 to 12/31/2017

Time Frame: 00:00:00 to 23:59:59

Cell Size: 100 m

Threshold: 14.100953

Bandwidth: 650 m

Area: 14,563,326 m²

Based On: 2139 incidents

LESSONS LEARNED

04

Lesson 01: Data quality varies considerably

Usually Crime Against Life (CAL) data is good (geocoded).

Most of the other types are not geocoded.

Data preparation workflow is manually done.

Lesson 02: Work directly with the Police officers

Scientists and Police officers usually do not speak the same language.

Listen and observe a lot.

Police officers can be resistant to change.

They have a lot of hidden knowledge.

Find someone to “take you under their wings”.

Lesson 03: Focus on simplicity, with sophistication

Police officers are usually overloaded and don't have technical expertise nor time to use complex general tools.

You can't hide the human interpretation from the police officer in the frontline.

If it takes more than 2 clicks they will not use.

Lesson 04: Research value is not enough to put a system in production

Prototypes built to publish papers don't work.

A mixed Research & Development team is more sustainable.

Lesson 05: Be prepared to make baby steps

You will spend a lot of time talking and convincing people.

Progress can be slow.

CrimeWatcher

Não seguro | 172.25.100.182:1096/select_analysis/29/

Português Brasileiro (pt-br) | Sair andrea.martins

Análise: AIS-6 CVLI

2018

Cenário: 42-FORTALEZA CVLI

Período: 01/01/2018 a 31/12/2018

Horário: 00:00:00 a 23:59:59

Tamanho da Célula: 100 m

Porcentagem: 28% (Integral)

Limite: 284.327923

Largura de Banda: 400 m

Área: 1.623.701 m²

Baseado em: 141 Eventos

Bairros Fortaleza

Seleção de Eventos

Cenário: 42-FORTALEZA CVLI

Período: 01/09/2018 a 30/09/2018

Horário: 00:00:00 a 23:59:59

Salvar Análise

Seleção de Cenário

Seleção de Janela de Tempo

Eventos

Locais

Divisões Administrativas

Estimativa de Densidade K

MSKDE

Filtro Espacial

Anotações

Camadas

Base

Ambiente: Fortaleza

Comunidade Góes

Av. Mister Hull

Av. Manoel

Subestação Coelco

11:19

Análise: AIS-6 CVLI

2018

Cenário: 42-FORTALEZA CVLI

Período: 01/01/2018 a 31/12/2018

Horário: 00:00:00 a 23:59:59

Tamanho da Célula: 100 m

Porcentagem: 28% (Integral)

Limite: 284.327923

Largura de Banda: 400 m

Área: 1.623.701 m²

Baseado em: 141 Eventos

Bairros Fortaleza

de Eventos

42-FORTALEZA CVLI

01/09/2018 a 30/09/2018

Horário: 00:00:00 a 23:59:59

CR / CE



CrimeWatcher

Inicio Cadastros Experimentos Sobre

Ambiente: Fortaleza

- Salvar Análise
- Seleção de Cenário
- Seleção de Janela de Tempo
- Eventos**
- Locais
- Divisões Administrativas
- Estimativa de Densidade K
- MDDE
- Filtro Espacial
- Anotações
- Cenários
- Base

Português Brasileiro (pt-br) Sair andrea.martins

Análise: AIS-9 CVP

2017

Cenário: 64-CVP Fortaleza Geral
Período: 01/01/2017 a 30/09/2017
Horário: 00:00:00 a 23:59:59
Tamanho da Célula: 100 m
Porcentagem: 30% (Integral)
Limite: 17.664415
Largura de Banda: 400 m
Área: 3,553,671 m²
Baseado em: 1683 Eventos

2018

Cenário: 64-CVP Fortaleza Geral
Período: 01/01/2018 a 31/12/2018
Horário: 00:00:00 a 23:59:59
Tamanho da Célula: 100 m
Porcentagem: 44% (Integral)
Limite: 17.737778
Largura de Banda: 400 m
Área: 5,468,461 m²
Baseado em: 2222 Eventos

Bairros Fortaleza

Seleção de Eventos

Cenário: 64-CVP Fortaleza Geral
Período: 01/09/2018 a 30/09/2018
Horário: 00:00:00 a 23:59:59
Eventos: 332

CrimeWatcher

Apps

CrimeWat

Selec

Divi

Estim

CONCLUSION

05

The development is far from being finished

We are working on making the interface more directed to the type of analyses the officers need to do without taking more than 2 clicks.

The system (GeoInteligência) is now being deployed in 15 states.

Many other lessons to be learned.

Lessons learned in developing a crime analytics solution



Dr. Emanuele Santos
emanuele@dc.ufc.br
Universidade Federal do Ceará

Summer School in Data Science
FGV EMAP
February 11th, 2020