

Técnicas de *clustering* para detectar patrones espaciales de criminalidad en jóvenes y adultos en Medellín. Octubre del 2013 a noviembre del 2014

Clustering techniques used to detect spatial patterns of criminality among young people and adults in Medellin. October 2013 through November 2014

Técnicas de clustering para detectar padrões espaciais de criminalidade em jovens e adultos em Medellín. Outubro de 2013 a novembro de 2014

FECHA DE RECEPCIÓN: 2015/09/25 FECHA CONCEPTO EVALUACIÓN: 2015/12/16 FECHA DE APROBACIÓN: 2016/01/28

María Alejandra Arango González
Ingeniera Matemática.
Investigadora, Fundación Casa de las Estrategias,
Medellín, Colombia.
investigacion@casadelasestrategias.com

Juan Diego Jaramillo Morales
M. A. Estudios culturales.
Subdirector de Investigación en la Fundación Casa de las Estrategias,
Medellín, Colombia.
estrategias@casadelasestrategias.com

Lucas Jaramillo Escobar
Político.
Director de la Fundación Casa de las Estrategias,
Medellín, Colombia.
direccion@casadelasestrategias.com

Para citar este artículo / To reference this article / Para citar este artigo: Arango, M. A., Jaramillo, J. D. & Jaramillo, L. (2016). Técnicas de clustering para detectar patrones espaciales de criminalidad en jóvenes y adultos en Medellín. Octubre del 2013 a noviembre del 2014. *Revista Criminalidad*, 58 (1): 25-45.

RESUMEN

Se busca encontrar patrones espaciales de agrupación de la criminalidad entre jóvenes y población total en Medellín, Colombia, durante el período comprendido entre octubre del 2013 y noviembre del 2014. Para esto se creó una malla hexagonal de ciudad y buscamos clústeres de agrupamiento entre trece variables delictuales. Para encontrar los clústeres usamos las técnicas de *subtractive clustering* y *fuzzy c-means clustering*. Al correr los clústeres encontramos microcorredores territoriales donde la

criminalidad alta se consolida durante varios períodos de tiempo y patrones temporales que muestran cómo algunas zonas de alta criminalidad se van formando de modo gradual. Adicionalmente, se buscaron patrones espaciales de criminalidad entre jóvenes, y se encontró que este grupo etario suele presentar mayor variabilidad en la dinámica criminal y territorios de injerencia más pequeños que el resto de la población.

PALABRAS CLAVE

Comportamiento delictivo, medición de la criminalidad, homicidio, estadísticas policiales, delincuente juvenil (fuente: Tesoro de política criminal latinoamericana - ILANUD).

ABSTRACT

Our aim is to find clusters of spatial patterns of criminality among young people and the total population in Medellín, Colombia, within the period between October 2013 and November 2014. For this purpose, a hexagonal city network was created and we looked for groupings into clusters among thirteen tort/delict variables. In order to find the clusters, we used the *subtractive clustering* and *fuzzy c-means clustering*. When running them, we found

territorial microcorridors where high criminality is consolidated during several periods of time and temporal patterns showing how some high criminality zones are being gradually shaped. Additionally, spatial patterns of criminality were sought among youths, and it was found that, usually, this age group tends to exhibit higher variability in criminal dynamics and meddling territories smaller than the rest of the population.

KEY WORDS

Criminal behavior, criminality measuring, homicide, police statistics, juvenile offender (Source: Tesoro de política criminal latinoamericana - ILANUD).

RESUMO

Procura-se encontrar padrões especiais de agrupação da criminalidade entre jovens e a população total em Medellín, Colômbia, durante o período entre outubro de 2013 e novembro de 2014. Para isso, uma malha hexagonal da cidade foi criada e nós procuramos clusteres entre treze variáveis criminosas. A fim de encontrar clusteres nós usamos as técnicas de *clustering sustractivo* e *fuzzy c-means clustering*. Quando executar os clusteres nós encontramos microbrokers territoriais onde o criminalidade alta é con-

solidada durante diversos períodos de tempo e padrões temporários que mostram como algumas zonas de alta criminalidade formam-se gradualmente. Adicionalmente, padrões espaciais de criminalidade entre jovens foram procurados, e achou-se que este grupo etário costuma apresentar maior variabilidade na dinâmica criminosa e territórios da ingerência menores do que o resto da população.

PALAVRAS - CHAVE

Comportamento criminoso, medição da criminalidade, homicídio, estatísticas policiais, delinquente juvenil (fonte: Tesoro de política criminal latinoamericana - ILANUD).

Introducción

El homicidio es una de las variables que más han influenciado sobre las decisiones de política de seguridad en Medellín. Su variación, sus diferentes etapas y niveles han marcado coyunturas políticas importantes, y en tiempos electorales este es uno de los fenómenos que más se discuten. Incluso, al revisar la bibliografía sobre violencia en Medellín, en los últimos 30 años, el homicidio sigue siendo la variable principal en los análisis sobre violencia (Cf. Blair, Grisales & Muñoz, 2009).

El problema que aquí se presenta es que la conflictividad en los barrios y comunas de Medellín no es solo una suma y variabilidad de una cantidad de homicidios dada, estos conflictos suelen presentar otros indicadores que muestran diferentes tipos de conflictividad. E. g., no es lo mismo una zona donde hay homicidios por violencia intrafamiliar o por conflictos de convivencia, que una en la cual los homicidios son por sicariato, muy tecnificados y estratégicamente distribuidos. Además, hay otros delitos diferentes al homicidio que van marcando las características de un conflicto en una zona específica. Desde las capturas por estupefacientes, las hechas por diferentes tipos de hurtos, las incautaciones de drogas y armas, van marcando la especificidad o caracterización de uno u otro lugar. De este modo, zonas con un mismo nivel de homicidios, pero con diferencias en otros delitos, van mostrando lugares donde la criminalidad tiene distintos tipos de articulaciones con la población y diferente incidencia.

En este sentido, un estudio que busque detectar fenómenos más allá del homicidio podría empezar por buscar delitos relacionados y agrupaciones de estos. De esta forma, ya no se trabaja con números solamente, sino con niveles de criminalidad por zonas, por agrupaciones espaciales de fenómenos. En suma, se trata de encontrar patrones espaciales que diferencien los distintos tipos de criminalidad que pueden incidir sobre el número de homicidios de una u otra forma y del mismo modo, ver las poblaciones que están involucradas en estas zonificaciones de criminalidad.

Sobre este tema de agrupación espacial y violencias se encuentran estudios principalmente alrededor del homicidio. En particular, existen levantamientos de índices espaciales para determinar causalidades del homicidio en un sector específico (Cf. Loaiza, 2012). Otros, también sobre homicidios, agrupan en forma espacial las distintas formas que puede tomar un homicidio en un territorio dado (Cf. Perversi, Valenga, Fernández, Britos, & García, 2007). Ya en otros casos se encuentran estudios que buscan patrones

espaciales en la agresividad juvenil (Cf. Devadoss & Felix, 2013) o la violencia contra la mujer (Cf. Fioredistella & Mastrangelo, 2015).

Varios de estos estudios nos interesan por sus metodologías y agrupaciones de violencias distintas (Cf. Ingram & Kurtis, 2014, Di Martino & Sessa, 2009). No obstante, para el caso de Medellín no se encontró mucha bibliografía, y en particular ninguna agrupación que cree nuevas espacialidades alrededor de fenómenos delincuenciales, sino que muchos de los estudios solo usan la división político-administrativa de la ciudad. Asimismo, no se hallaron agrupaciones de variables distintas al homicidio que puedan dar una explicación más profunda alrededor de las violencias y, en específico, del homicidio.

Por todo lo anterior, este artículo en particular se interesa en encontrar agrupaciones de delitos en determinados espacios en Medellín, que expliquen distintas formas de criminalidad, tanto para jóvenes¹ como para el resto de la población. En específico, buscamos a través de distintas técnicas de “clustering” encontrar agrupaciones de delitos que den explicaciones de fenómenos delictivos en el espacio y el tiempo².

Se trabaja la variable jóvenes porque se considera que muchos de los estudios que se han hecho sobre ellos en la ciudad tienden a criminalizar un rango etario y no dan cuenta de las distintas formas de conexión que tiene esta población con el conflicto. De modo específico, creemos que debe medirse su incidencia real en el conflicto y los diferentes tipos delitos de los que son víctimas y victimarios.

Metodología

Cuando hacemos referencia a diferentes tipos de criminalidad, buscamos tomar distintos indicadores relacionados con esta, y agruparlos para detectar las distintas zonas en las que estos aparecen. Por esto, para este análisis usamos las variables que aparecen en la tabla 1.

Así las cosas, usamos los diez delitos que mayor cantidad de capturas presentan en el período dado, junto con las incautaciones de armas, las de drogas y los homicidios de la ciudad³. Tomamos estas variables porque dan cuenta de varios fenómenos a la vez. Primero, queremos ver los delitos que más

1 Entendemos por jóvenes el rango etario que va desde los 10 años hasta los 28, según la definición de la OMS (Organización Mundial de la Salud) y la Corte Constitucional de Colombia (Ley 26 de 2013).

2 Este artículo hace parte de una investigación mayor (con fuentes cualitativas y cuantitativas), que la Fundación Casa de las Estrategias ha venido adelantando con *Open Society Foundations* (OSF), titulada “Descontando a los jóvenes del homicidio en Medellín”. No obstante, los resultados aquí publicados son exclusivos y no se usan en dicha investigación.

3 Todas estas variables tienen como fuente la Policía Nacional de Colombia.

Tabla 1
Codificación de variables

Variable	Codificación	Descripción
Incautación de armas	INCARMA	Incautaciones de armamento realizadas (o registradas) por la Policía Nacional
Artículo 376. Tráfico, fabricación o porte de estupefacientes	ART. 376	Capturas por este delito
Artículo 239. Hurto a personas	ART. 239	Capturas por este delito
Artículo 229. Violencia intrafamiliar	ART. 229	Capturas por este delito
Artículo 239. Hurto a entidades comerciales	ART. 239B	Capturas por este delito
Artículo 111. Lesiones personales	ART. 111	Capturas por este delito
Artículo 291. Uso de documento falso	ART. 291	Capturas por este delito
Artículo 365. Fabricación, tráfico, porte o tenencia de armas de fuego, accesorios, partes o municiones	ART. 365	Capturas por este delito
Artículo 429. Violencia contra servidor público	ART. 429	Capturas por este delito
Artículo 447. Receptación	ART. 447	Capturas por este delito
Artículo 103. Homicidios	ART. 103	Capturas por este delito
Incautación de drogas	INCADRO	Incautaciones de sustancias realizadas (o registradas) por la Policía Nacional
Homicidios	HOMI	Homicidios registrados por la Policía Nacional

ocupan la operatividad policial en lo cotidiano; segundo, queremos detectar delitos que no tienen que ver con patrones de criminalidad estructurada, como las capturas por violencia intrafamiliar, y tercero, tenemos los delitos que implican comportamientos entre estructuras criminales, como son los homicidios, las incautaciones de drogas y de armas, y las capturas por estos dos delitos.

Todas estas variables se tomaron en un corte transversal comprendido entre octubre del 2013 y noviembre del 2014. Lo importante de esta muestra de variables es que presentan baja correlación, es decir, son variables que capturan distintos compor-

tamientos criminales en una misma zona, y de este modo se pueden detectar las distintas formas que puede tomar espacialmente la criminalidad. La tabla 2 presenta la correlación entre variables. Hay unas correlaciones altas esperadas, como la incautación de drogas con las capturas por estupefacientes, pero el resto de variables mantienen correlaciones por debajo de 0,5 y 0,6, lo que permite que no haya problemas de colinealidad en los clústeres; lo anterior quiere decir que estos últimos están conformados por variables que describen diferentes fenómenos y no el mismo fenómeno explicado por varias variables.

Tabla 2
Correlación entre variables

Variable	INCARMA	ART. 376	ART. 239	ART. 229	ART. 239B	ART. 111	ART. 291	ART. 365	ART. 429	ART. 447	ART. 103	INCADRO	HOMI
INCARMA	1												
ART. 376	0,38	1											
ART. 239	0,40	0,52	1										
ART. 229	0,27	0,26	0,25	1									
ART. 239B	0,25	0,15	0,41	0,11	1								
ART. 111	0,42	0,65	0,86	0,33	0,32	1							
ART. 291	0,37	0,29	0,42	0,19	0,25	0,38	1						
ART. 365	0,71	0,49	0,46	0,34	0,26	0,50	0,40	1					
ART. 429	0,40	0,60	0,69	0,33	0,33	0,74	0,40	0,53	1				
ART. 447	0,38	0,73	0,59	0,30	0,16	0,69	0,31	0,47	0,54	1			
ART. 103	0,41	0,49	0,50	0,30	0,18	0,57	0,27	0,48	0,49	0,50	1		
INCADRO	0,36	0,95	0,52	0,27	0,13	0,66	0,25	0,46	0,56	0,78	0,52	1	
HOMI	0,47	0,50	0,37	0,40	0,18	0,51	0,27	0,55	0,46	0,49	0,56	0,51	1

Análisis espacial y georreferenciación

Uno de los problemas que presenta la georreferenciación de variables de criminalidad en Medellín es el tamaño de los polígonos (de las zonas de medición). Por lo general, la información se divide en comunas y corregimientos, y algunos pocos análisis la presentan en el ámbito barrial. Esta información, a pesar de ser importante para tomadores de decisiones a nivel global y para ordenamiento general de los dispositivos de seguridad y convivencia en la ciudad, no sirve para optimizar las intervenciones. Por lo anterior, en aras de entender la complejidad de los fenómenos criminales en la ciudad, se necesitan análisis espaciales que apunten al nivel microterritorial, con posibilidades de agrupamiento de sectores y análisis globales, sin las limitaciones espaciales que impone la división político-administrativa de la ciudad. En suma, se trata de no usar las divisiones estándar (ciudad, comuna, barrio, etc.), sino de crear un nuevo mapa de fenómenos que establezca otras espacialidades, según los diferentes delitos.

Así las cosas, en este análisis se usó un método de clusterización para agrupar y georreferenciar los distintos delitos. En este sentido, se buscó una medida de hexágono estándar, con la cual dividir la ciudad y poder encontrar agrupaciones y patrones espaciales de comportamiento criminal en esta parrilla de hexágonos. Esta medida de hexágono usó

como referente la tasa de homicidios de Medellín. De este modo, cada hexágono de esta medición tiene proporcionalmente la misma tasa de homicidios que el total de ciudad⁴.

Una vez definido el tamaño de los hexágonos, se pasó a buscar patrones y agrupaciones espaciales de los delitos explicados con anterioridad. Para esto se usaron los siguientes pasos metodológicos:

Normalización

La normalización de las variables es necesaria para eliminar la dependencia de ellas con respecto a las unidades de medida empleadas. Esto hace los resultados comparables con otros que no lo eran en forma directa.

El método de normalización implementado transforma linealmente el conjunto de datos usando la ecuación 1.

$$\text{Valor}_{\text{normalizado}} = \frac{\text{valor} - \text{mínimo}}{\text{máximo} - \text{mínimo}}$$

Ecuación 1: normalización

⁴ La tasa de homicidios de Medellín para el período analizado fue de 34 homicidios por cada cien mil habitantes. Es por ello que se crea un “grid” hexagonal (panal o parrilla de hexágonos), donde cada hexágono tiene un promedio de 2.500 habitantes (8 cuadras en zonas urbanas) y cuya tasa de homicidios es 0,86 por cada cien mil habitantes (esta tasa es proporcional de ciudad, dada la cantidad de hexágonos y de personas por hexágonos). Se usa la tasa de homicidios dada, que es la variable con mejor información y menor error en su conformación.

Esta transformación logra que variables con unidades muy diferentes (tipos de delitos) se conserven en un mismo rango comparable: [0,1].

Clusterización

La clusterización⁵ es uno de los muchos métodos o algoritmos de clasificación y modelado de sistemas. El propósito de realizar clústeres es identificar agrupamientos “naturales” en un conjunto muy grande de datos y generar una representación concisa del comportamiento del sistema. En este caso es la agrupación de hexágonos, dadas las características que cada uno de estos tiene según los delitos que incluye.

Las técnicas de clusterización se usan comúnmente en conjunto. En este caso se eligieron los métodos: “*Subtractive clustering*” y “*Fuzzy C-means clustering*” (Hammouda & Karray, 2000).

Subtractive clustering

Cuando no se tiene una idea clara de cuántos clústeres deberían agrupar un conjunto de datos (de hexágonos), la técnica “*Subtractive clustering*” permite estimarlos. En vez de crear una cuadrícula de la dimensión del problema como “*Mountain clustering*”, elige entre el conjunto de datos los posibles clústeres, lo que permite reducir la computación de la dimensión del problema a su tamaño. Muchas veces los clústeres no están ubicados en uno de los puntos de los datos, pero en la mayoría de los casos son buenas aproximaciones.

Dado que cada punto es un centro en potencia, la medida de densidad (cuántos hexágonos son cercanos a un centro) en el punto x_i se define en la ecuación 2.

$$D_i = \sum_{j=1}^n \exp\left(\frac{-\|x_i - x_j\|^2}{\left(\frac{r_a}{2}\right)^2}\right)$$

Ecuación 2: medida de densidad en un punto

⁵ Existen diferentes técnicas de clusterización. Las principales son: **K-means Clustering**: Encuentra los centros de los clústeres tratando de minimizar una función de costo: la distancia. **Fuzzy C-means Clustering**: Es una técnica mejorada de “K-means Clustering”, en la que cada punto pertenece a un clúster en un grado específico. También trata de minimizar una función de costo. **Mountain Clustering**: Esta técnica construye una función de densidad (montaña) para cada posible posición en el conjunto de datos, y escoge la de mayor densidad como su centro. Luego hace nuevamente este proceso, hasta encontrar la cantidad de clústeres deseados. **Subtractive Clustering**: Es similar a “Mountain Clustering”, excepto que no calcula la función de densidad en todas las posibles posiciones, sino que usa las posiciones de los puntos para calcular la densidad, lo que reduce el número de cálculos de manera significativa.

Dónde r_a es una constante positiva que representa el radio del vecindario (cuánto se está dispuesto a alejarse de un centro). Si un punto tiene varios puntos rodeándolo, tendrá una mayor densidad.

El primer centro x_{c1} será el punto que mayor densidad D_{c1} tenga o que más esté rodeado de otros puntos. Luego la medida de densidad para cada conjunto de puntos x_i estará dada por la ecuación 3.

$$D_i = D_i - D_{c1} \exp\left(\frac{-\|x_i - x_{c1}\|^2}{\left(\frac{r_b}{2}\right)^2}\right)$$

Ecuación 3: medida de densidad para un conjunto de puntos

Dónde r_b es una constante positiva que define un vecindario con reducciones en la medida de densidad. Así, el conjunto de puntos cercanos al primer centro tendrá una medida de densidad reducida.

El siguiente clúster será el que mayor función de densidad tenga. El proceso continúa hasta que se cree una cantidad suficiente de clústeres (con el radio dado).

Fuzzy C-means Clustering

El “*Fuzzy C-means clustering*” emplea una partición “fuzzy”, en la que cada punto o conjunto de datos pertenece a varios grupos en un grado específico, dado por una función de pertenencia que está entre 0 y 1. También utiliza una función de costo que trata de minimizar para encontrar las particiones.

La matriz de pertenencia U tiene elementos cuyos valores están entre 0 y 1. La suma de los valores de pertenencia para cada punto es igual a 1, tal como se muestra en la ecuación 4.

$$\sum_{i=1}^c u_{ij} = 1, \forall j = 1, \dots, n$$

Ecuación 4: suma de los valores de pertenencia

La función de costo (la penalización o cobro por alejarse de un centro) para el “*Fuzzy C-means Clustering*” es una generalización de la distancia euclidiana, que se presenta en la ecuación 5.

$$J(U, c_1, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2$$

Ecuación 5: función de costo Fuzzy C-means Clustering

Donde,

- u_{ij} está entre 0 y 1.
- c_i es el centro del clúster del grupo i .
- $d_{ij} = |c_i - x_j|$ es la distancia euclidiana entre el centro del clúster i y el conjunto de puntos o el punto j -ésimo.
- m es un exponente ponderador.

El objetivo es que este costo sea mínimo, es decir, que los hexágonos encuentren su centro más cercano. Las condiciones necesarias para encontrar el mínimo con esta función de costo están dadas por las ecuaciones 6 y 7.

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}$$

Ecuación 6: condición para encontrar el mínimo en función de costo

$$u_{ij} = \frac{y}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}$$

Ecuación 7: condiciones para encontrar el mínimo en función de costo

El algoritmo trabaja iterativamente hasta que no encuentre otro mínimo bajo los siguientes pasos:

- Paso 1: inicializar la matriz U con valores aleatorios entre 0 y 1, cumpliendo con la restricción dada en la ecuación 4.
- Paso 2: calcular los centros de los clústeres usando la ecuación 6.
- Paso 3: computar la función de costo de la ecuación 5 con estos valores. Parar si no hay cambios significativos en el costo.
- Paso 4: calcular la matriz U usando la segunda condición o ecuación 7.

Distancia euclidiana

Después de encontrar los clústeres generales, se necesita hallar a qué clúster pertenece cada zona filtrada por determinadas características. Dado que los clústeres ya están definidos, se toma la distancia euclidiana no generalizada como función de costo para encontrar su pertenencia. La distancia euclidiana se presenta en la ecuación 8.

$$J = \sum_{i=1}^c J_i = \sum_{i=1}^c \left(\sum_{k, x_k \in G_i} \|x_k - c_i\|^2 \right)$$

Ecuación 8: distancia euclidiana

Dónde G_i es cada grupo de datos y c_i los centros definidos.

En este caso, al finalizar se elige el clúster que tenga mayor grado de pertenencia. Usando el “Subtractive clustering” se obtienen las cantidades de clústeres que se exponen en la tabla 3 para cada valor en los radios.

Tabla 3
Radio de clústeres

Radio	Cantidad de clústeres
$\geq 0,3$	1
0,2	2
0,1	3
0,04	4
0,03	5
$\leq 0,02$	6

Al estabilizarse en un r_a de 0,02, se eligen 6 como la cantidad de clústeres adecuados. Luego, implementando el “Fuzzy C-Means” se obtienen los diferentes clústeres en que se categorizarán los hexágonos, los cuales quedan expresados en la tabla 4.

De esta forma, cada delito está representado en cada clúster en cierto nivel (expresado porcentualmente según fue normalizado). Cada porcentaje expone en qué cantidad está expresado ese delito con respecto al total de delitos ocurridos en el tiempo analizado; cada uno de ellos varía en la formación del clúster. En total, la clusterización arrojó seis clústeres diferentes. Seis maneras distintas de agrupar las zonas con delitos en la ciudad. De estos seis clústeres encontramos las siguientes:

1. Este clúster (blanco) tiende a cero en todas las variables, son lugares sin ningún fenómeno de criminalidad significativo. A pesar de que son niveles muy bajos, se destaca que comparativamente los niveles más altos en este clúster son homicidios y capturas por violencia intrafamiliar.
2. Este clúster (verde) se presenta principalmente en zonas residenciales. Alrededor de familias, donde los delitos socioeconómicos (como los

Tabla 4
Clústeres

Clúster	1	2	3	4	5	6
INCARMA	0,0 %	29,8 %	10,7 %	48,8 %	69,8 %	100,0 %
ART. 376	0,0 %	52,4 %	21,1 %	56,5 %	76,6 %	100,0 %
ART. 239	0,0 %	17,4 %	12,7 %	53,9 %	73,4 %	100,0 %
ART. 229	0,0 %	100,0 %	20,8 %	22,2 %	24,7 %	25,8 %
ART. 239B	0,0 %	4,4 %	8,6 %	47,8 %	70,0 %	100,0 %
ART. 111	0,0 %	42,3 %	20,8 %	54,0 %	74,8 %	100,0 %
ART.291	0,0 %	26,3 %	15,7 %	52,8 %	73,9 %	100,0 %
ART.365	0,0 %	33,3 %	9,2 %	45,3 %	67,7 %	100,0 %
ART.429	0,0 %	52,0 %	20,5 %	56,7 %	75,1 %	100,0 %
ART.447	0,0 %	57,8 %	15,3 %	55,5 %	74,9 %	100,0 %
ART.103	0,0 %	48,9 %	19,4 %	56,7 %	75,0 %	100,0 %
INCADRO	0,0 %	60,2 %	21,7 %	56,9 %	77,0 %	100,0 %
HOMI	0,0 %	60,5 %	11,4 %	53,1 %	74,9 %	100,0 %

hurto) no tienen mucho impacto, pero otros, como la violencia intrafamiliar, el homicidio y las capturas por tráfico y porte de estupefacientes, tienen mayor predominancia.

- Para este clúster (amarillo) casi todos los delitos están en un nivel bajo. No obstante, los delitos más altos son la incautación de drogas, la violencia intrafamiliar, las capturas por estupefacientes y las capturas por lesiones personales.
- En este clúster (naranja) todos los delitos se encuentran en niveles medios, menos las capturas por violencia intrafamiliar, que están en un nivel bajo.
- Este clúster (rojo) es similar al anterior, pero con un mayor nivel en los indicadores diferentes a las capturas por violencia intrafamiliar. Llama la atención que es un clúster donde las incautaciones de drogas y las capturas por lo mismo son las más altas comparativamente con los indicadores. Podría asociarse a un clúster de marcada especialidad en el microtráfico.
- En este clúster (rojo oscuro) los indicadores alcanzan el máximo nivel para el período de tiempo analizado, exceptuando las capturas por violencia intrafamiliar.

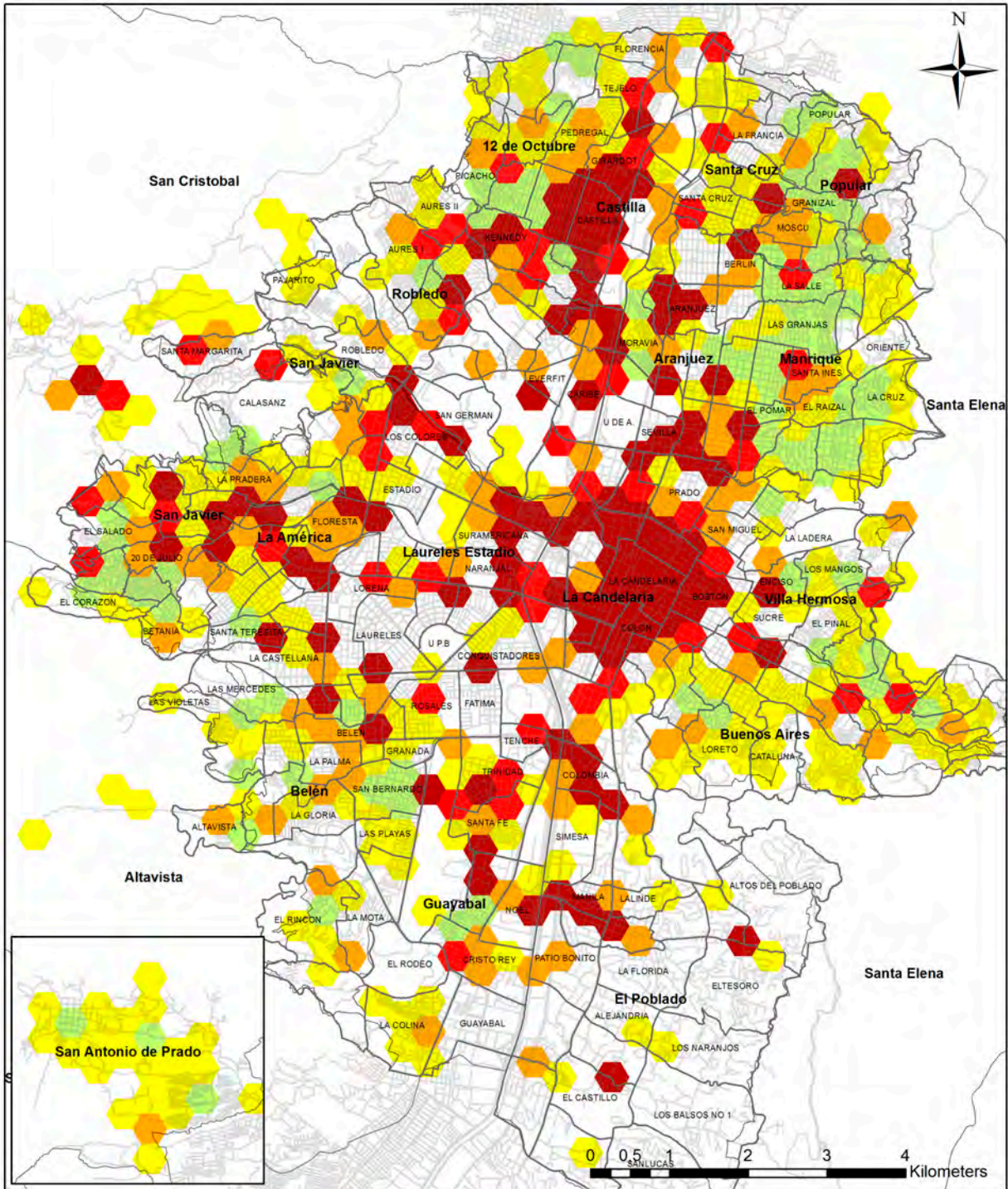
Resultados

Una vez obtenidos los tipos de clústeres, se procede al análisis general y particular, en aras de encontrar algunos patrones.

De este modo, cada clúster tiene una especialidad, un perfil de agrupación de delitos, que muestra cómo varios de estos se agrupan en una zona según la presencia de distintos actores. En general, se destaca que mientras los niveles de delitos criminales (homicidios, capturas por homicidios o por estupefacientes, incautación de drogas y de armas, entre otros) van aumentando, los delitos como capturas por violencia intrafamiliar van disminuyendo. Más allá de querer entender las relaciones numéricas entre clústeres, nos interesa acá comprender sus comportamientos espacio-temporales.

En particular, queremos ver los clústeres formados por jóvenes. De este modo, el análisis se divide en dos frentes fundamentales: movimiento temporal de los clústeres en el período comprendido entre octubre del 2013 y noviembre del 2014 y clústeres totales para estos períodos. Estas dos etapas están discriminadas por jóvenes, población total, y población total sin jóvenes.

Clústeres totales entre octubre del 2013 y noviembre del 2014



Clústeres

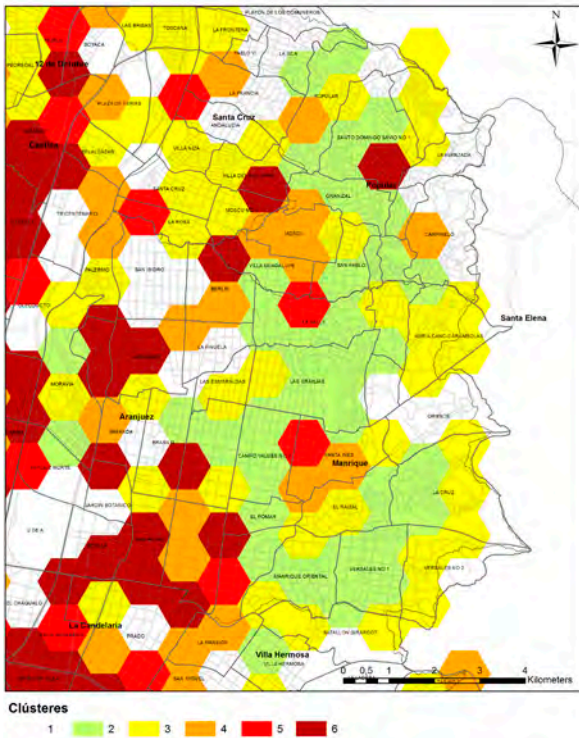


Mapa 1. Clústeres totales (oct./13-nov./14)

Fuente: Datos Policía Nacional. Elaboración propia.

Del mapa 1 se destacan varios corredores para tener en cuenta:

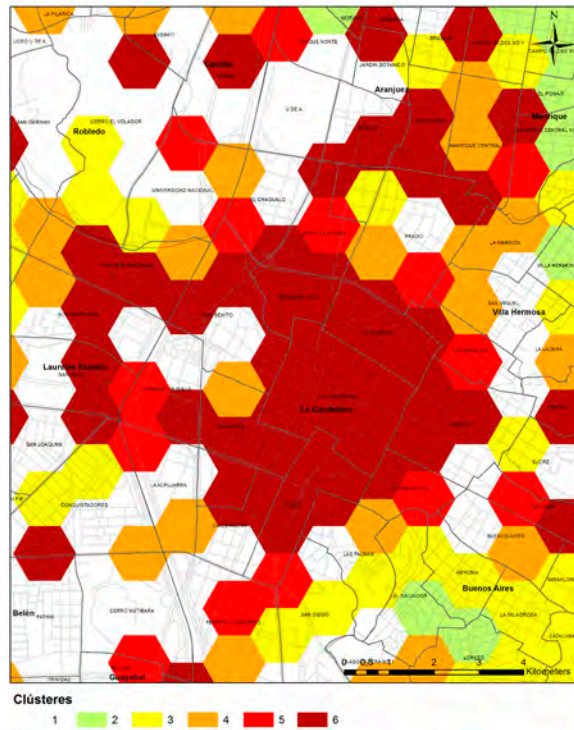
1) El primero es uno que habla sobre todo de capturas por estupefacientes, violencia intrafamiliar y homicidios. Este corredor va por toda la centralidad de la Comuna Nororiental, desde la frontera entre la Comuna 3 (Manrique) y la Comuna 8 (Villa hermosa), hasta las comunas 1 y 2 (Popular y Santa Cruz, respectivamente). Este corredor puede verse con mayor claridad en el mapa 2.



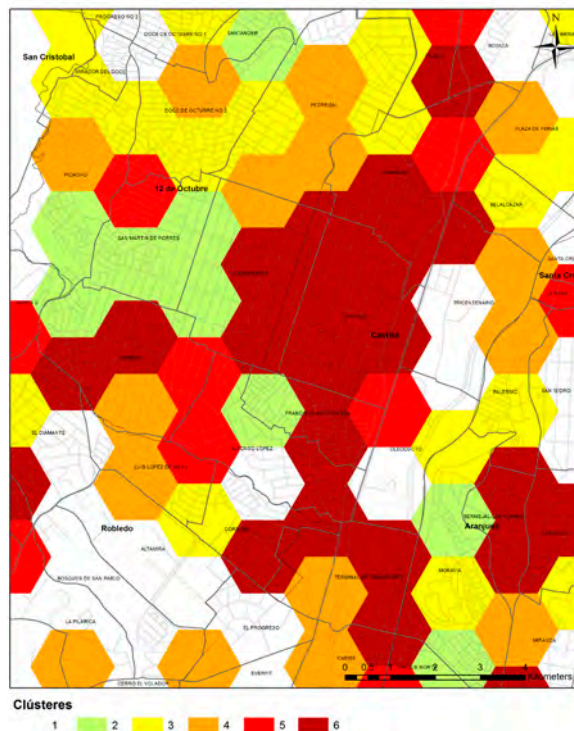
Mapa 2. Corredor 1 (oct./13-nov./14)
Fuente: Datos Policía Nacional. Elaboración propia.

2) El siguiente corredor (que aparece en el mapa 3) se encuentra en todo el centro de la ciudad (Comuna 10 - La Candelaria). Tiene todos los delitos agrupados en su máximo nivel posible para el período seleccionado, excepto las capturas por violencia intrafamiliar, y es una zona con pocos hogares residenciales o de familias.

3) El tercer corredor, expuesto en el mapa 4, es el comprendido entre las comunas 5 y 6 (Castilla y Doce de Octubre, respectivamente). También en este los delitos alcanzan el máximo nivel, en especial en los barrios Castilla y la Esperanza, que son fronterizos entre estas dos comunas.

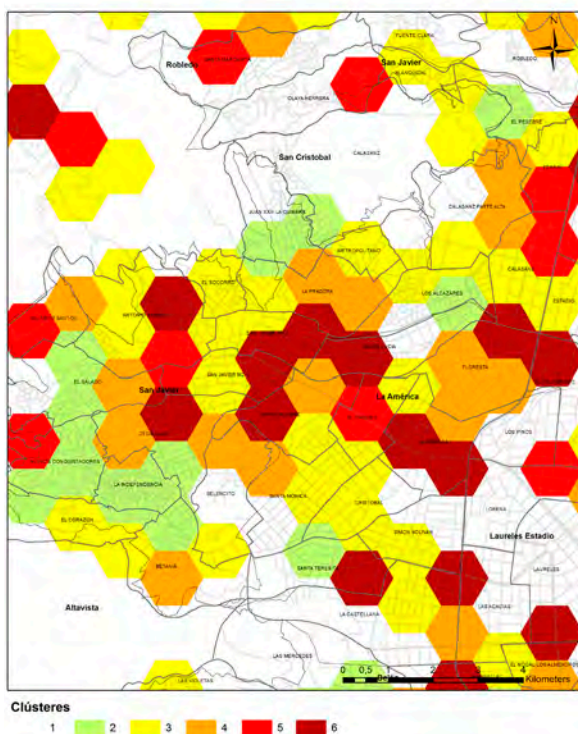


Mapa 3. Corredor 2 (oct./13-nov./14)
Fuente: Datos Policía Nacional. Elaboración propia.



Mapa 4. Corredor 3 (oct./13-nov./14)
Fuente: Datos Policía Nacional. Elaboración propia.

4) Otro corredor importante, también en máximo nivel de delitos (aunque en número de hexágonos es más pequeño que el resto), es el ubicado entre las comunas 12 y 13, por el corredor del Metro, que viene desde la Floresta, en la Comuna 12, y llega hasta San Javier, El Salado y Nuevos Conquistadores, en la Comuna 13. Este se presenta en el mapa 5.



Mapa 5. Corredor 4 (oct./13-nov./14)

Fuente: Datos Policía Nacional. Elaboración propia.

Estos son los principales clústeres de la ciudad que presentan patrones de distribución espacial duradera en el tiempo seleccionado. En estos también vale la pena ver que muchos de los colores de comportamiento siguen una distribución escalonada. Específicamente, esto se destaca en dos zonas:

1. En la zona nororiental los hexágonos verdes suelen estar rodeados de hexágonos amarillos. Esto explica que hay un corredor con profundización en delitos, como el de violencia intrafamiliar, homicidios y estupefacientes (verde), rodeado de unas franjas amarillas, que presentan todos los delitos en rangos bajos (incluyendo el homicidio).

2. En la zonas centro, noroccidente y centro-occidente ocurre algo similar, pero con los hexágonos rojos oscuros, rojos y naranjas (clústeres 6, 5 y 4, respectivamente). En ellas todos los delitos están en nivel alto, por lo común rodeadas de zonas donde todos los delitos tienden a ser altos (excepto los de violencia intrafamiliar).

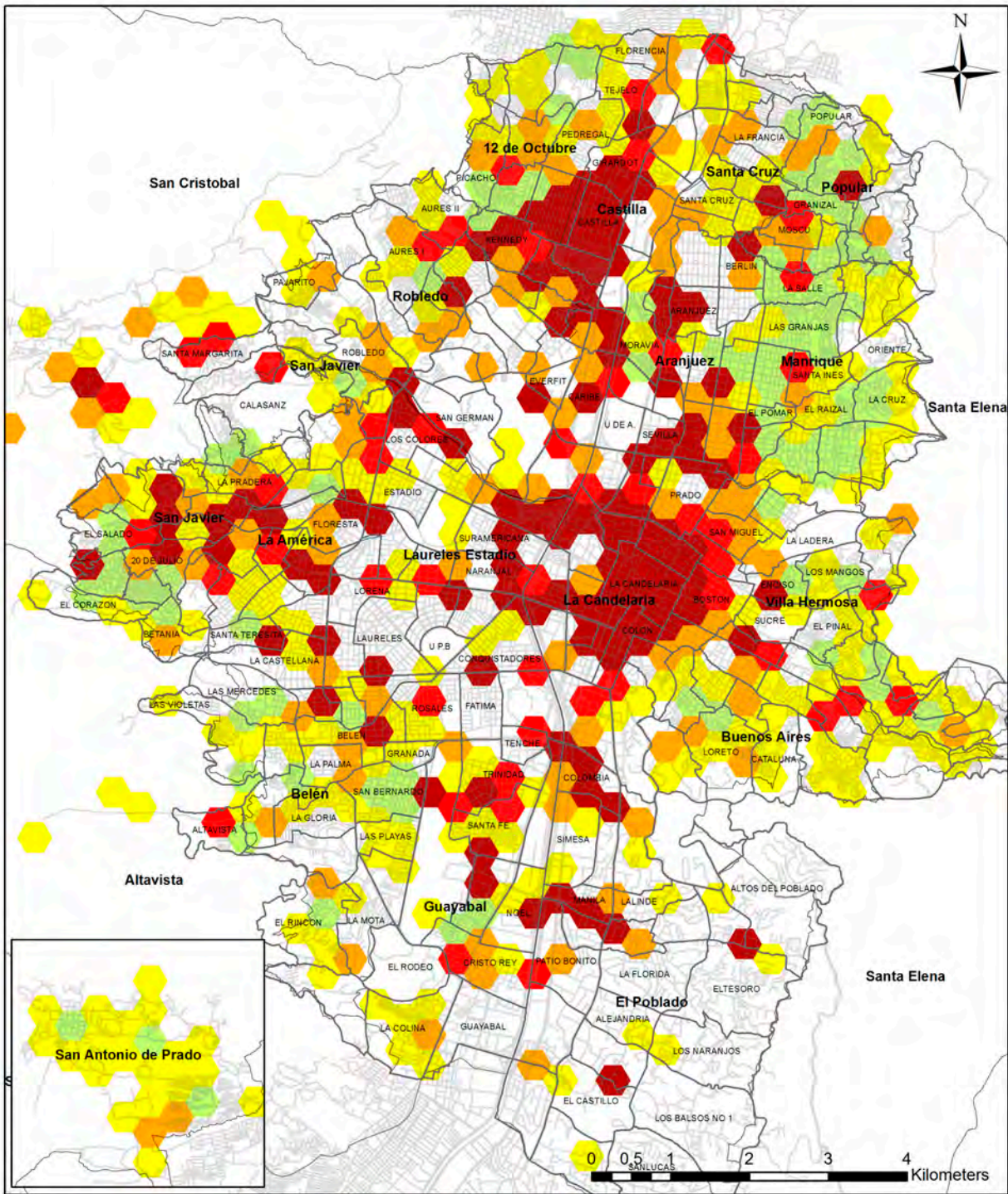
Ahora analizaremos el mismo mapa de clústeres de ciudad, pero solo para la población joven (10 a 28 años), con los siguientes resultados:

Los clústeres de jóvenes, visible en el mapa 6, presenta unos corredores similares a los del total de la población, teniendo en cuenta que este grupo etario es el 30,65 %⁶ de la población total de Medellín. No obstante, llama la atención que muchas de las zonas que en el mapa de clústeres totales aparecen como medianas (amarillas o naranjas), se vuelven rojas o rojas oscuras para el mapa de jóvenes. Es decir, aumentan delitos criminales y disminuyen otros de convivencia, como la violencia intrafamiliar. Este fenómeno se presenta principalmente en las zonas noroccidentales (comunidades 5 y 6, Castilla y Doce de Octubre), centro y centro-occidente (Comuna 13 - San Javier).

Del mismo modo, es importante ver el mapa sin jóvenes, es decir, el de clústeres donde está toda la población, exceptuando el rango etario entre 10 y 28 años.

En el mapa 7 es más fácil ver que muchos de los corredores y clústeres de agrupamiento de hexágonos se rompen, se desintegran un poco, lo que muestra que en cierto modo, y sin el ánimo de adelantar conclusiones, el rango etario de jóvenes da consistencia a los agrupamientos de clústeres rojos y rojos oscuros.

⁶ Según el último censo del DANE, del año 2005, con proyecciones hasta el año 2010.

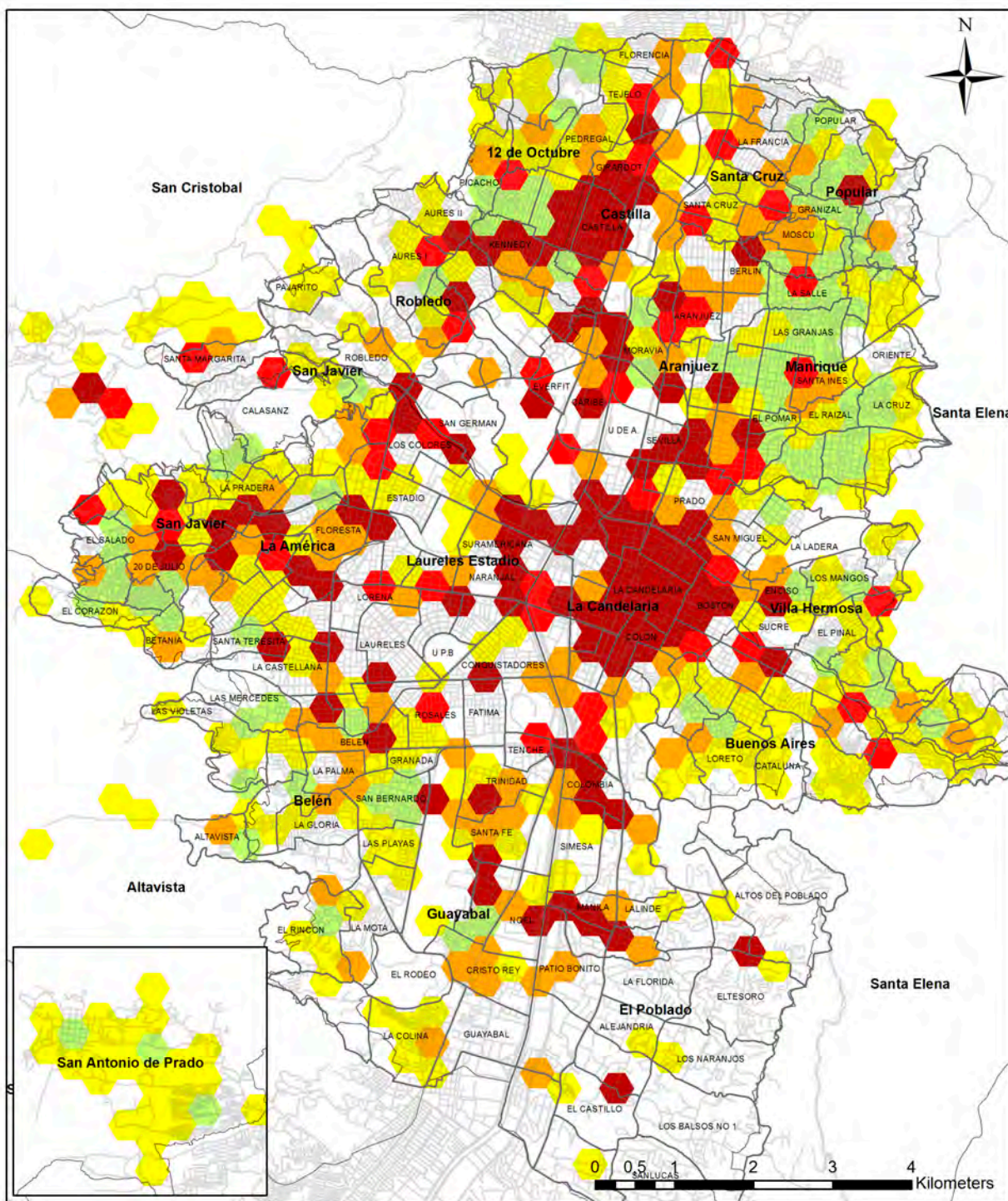


Clústeres



Mapa 6. Clústeres de jóvenes (oct./13-nov./14)

Fuente: Datos Policía Nacional. Elaboración propia.



Clústeres



Mapa 7. Clústeres sin jóvenes (oct./13-nov./14)

Fuente: Datos Policía Nacional. Elaboración propia.

Consistencia en hexágonos

Siguiendo en esta línea de análisis, es importante ver qué tan consistentes son los hexágonos, es decir, qué tanto cambian de color en el tiempo analizado. En la matriz de la tabla 5 se puede ver la probabilidad de que en el tiempo analizado los hexágonos cambien de color (es decir, cambie el fenómeno delictivo en esa zona). La diagonal central muestra la probabilidad de que un hexágono se conserve en su color; los clústeres clase uno (color blanco) tienen una alta probabilidad de conservarse blancos, ya que muchos de estos están en zonas poco habitadas de la ciudad, como los corregimientos (zona rural).

Tabla 5
Consistencia en hexágonos

Clúster	1	2	3	4	5	6
1	96,4%	1,2%	0,3%	0,8%	0,2%	1,0%
2	71,9%	12,6%	2,1%	4,3%	0,9%	8,1%
3	76,5%	14,5%	1,2%	3,6%	1,8%	2,4%
4	69,4%	6,2%	2,3%	8,2%	4,1%	9,9%
5	58,6%	6,9%	2,3%	9,8%	5,2%	17,2%
6	59,2%	8,3%	2,1%	7,9%	2,9%	19,5%

Lo primero que sale de esta matriz es que es poco probable que un sector con baja criminalidad (clústeres 1 o 2) se convierta en sectores de alta criminalidad en un período de tiempo dado; en cambio, sectores de alta criminalidad (clústeres 5 o 6) tienen mayor probabilidad (no mucha) de volverse de baja criminalidad. La movilidad de fenómenos, entonces, como se analizaba en los mapas anteriores, se da más entre sectores de criminalidad similar. En suma, las graduaciones se dan entre sectores con dinámicas criminales similares.

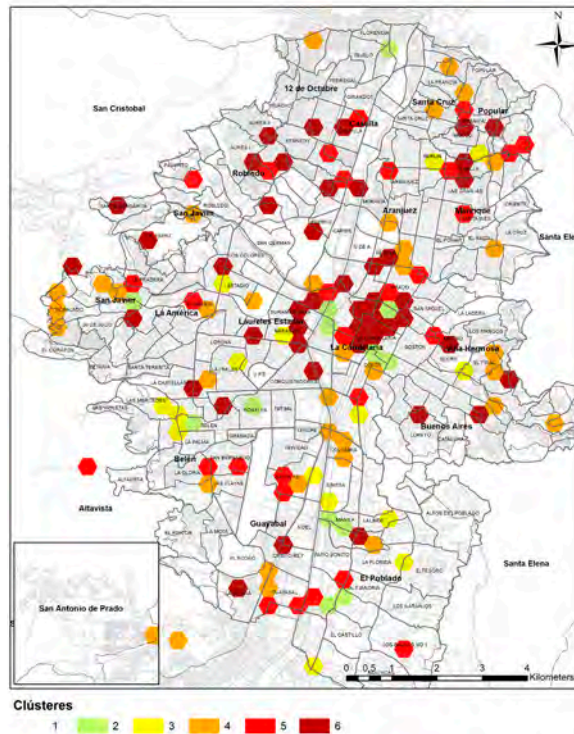
Al revisar la misma matriz, pero solo para la población joven (expuesta en la tabla 6), aparece que sigue una dinámica similar a la matriz de la población total, pero con la diferencia de que esta presenta probabilidades mayores de movilidad. Es decir, en general siguen el mismo patrón, pero los comportamientos criminales tienen mayor probabilidad de variar en un período determinado de tiempo cuando se trata de jóvenes.

Tabla 6
Consistencia de hexágonos para jóvenes

Clúster	1	2	3	4	5	6
1	96,8%	1,2%	0,2%	0,7%	0,2%	0,9%
2	69,3%	13,2%	1,6%	6,3%	1,8%	7,9%
3	80,2%	14,3%	1,6%	1,6%	1,6%	0,8%
4	47,4%	8,4%	0,8%	28,1%	6,2%	9,1%
5	38,6%	8,1%	0,9%	13,5%	15,7%	23,3%
6	49,3%	9,4%	1,1%	13,6%	5,3%	21,3%

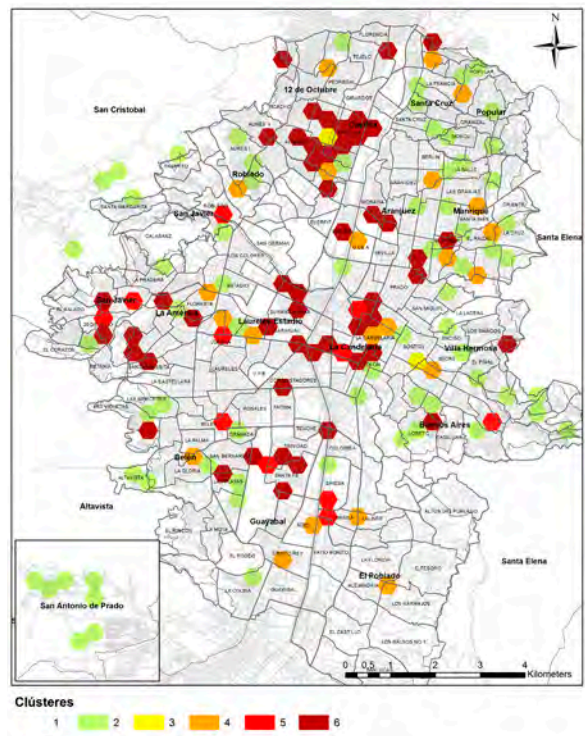
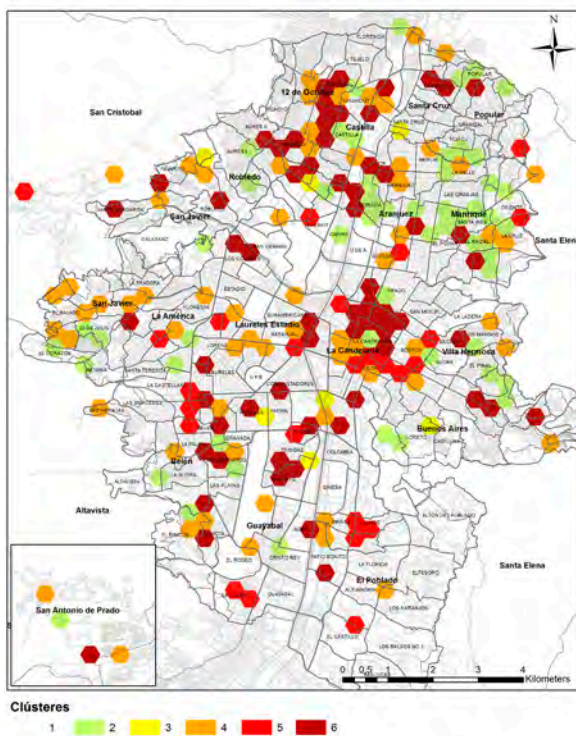
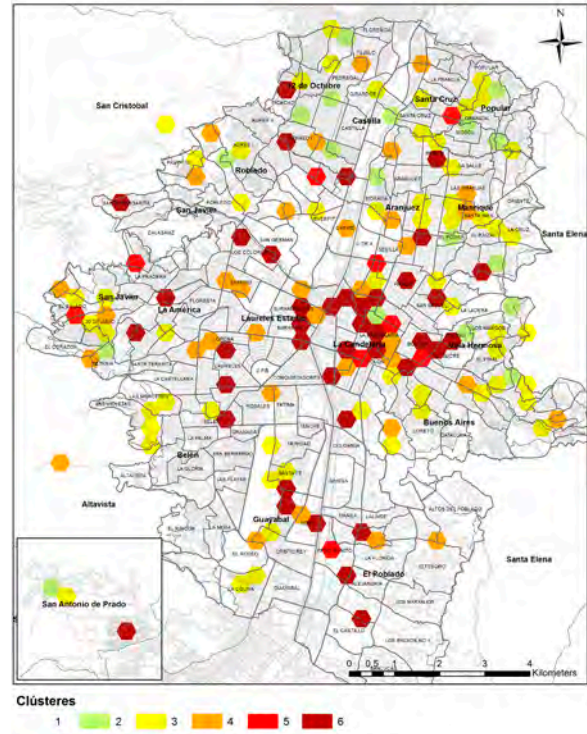
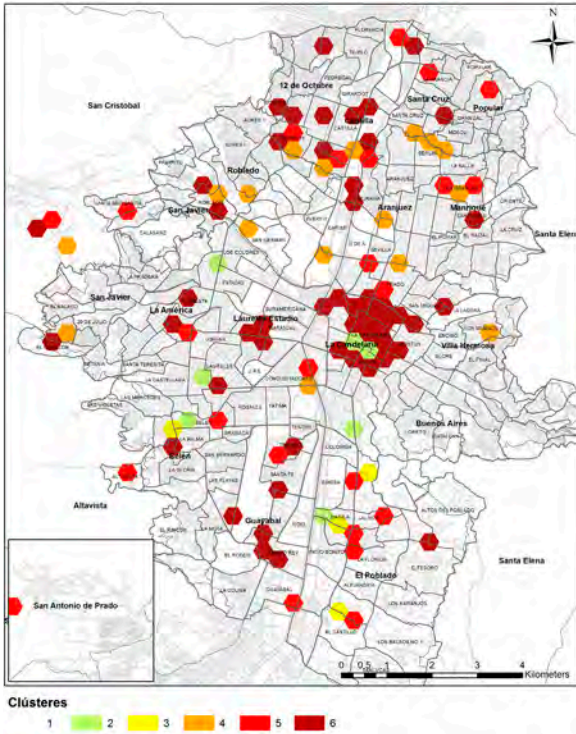
Análisis de clustering temporal octubre 2013 a noviembre de 2014

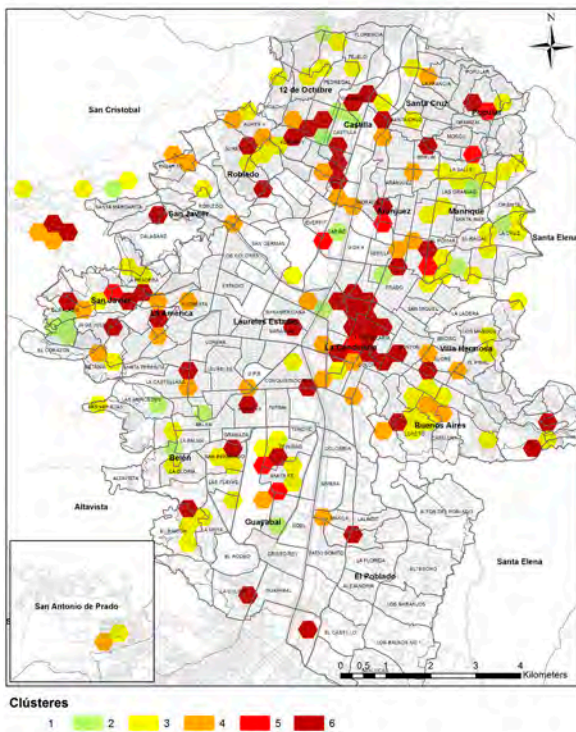
Para ver las movilizaciones específicas de la variación entre clústeres es importante analizar el fenómeno a lo largo de los 14 meses analizados; a continuación se presentan los meses más importantes:



Mapa 8. Clústeres, octubre 2013

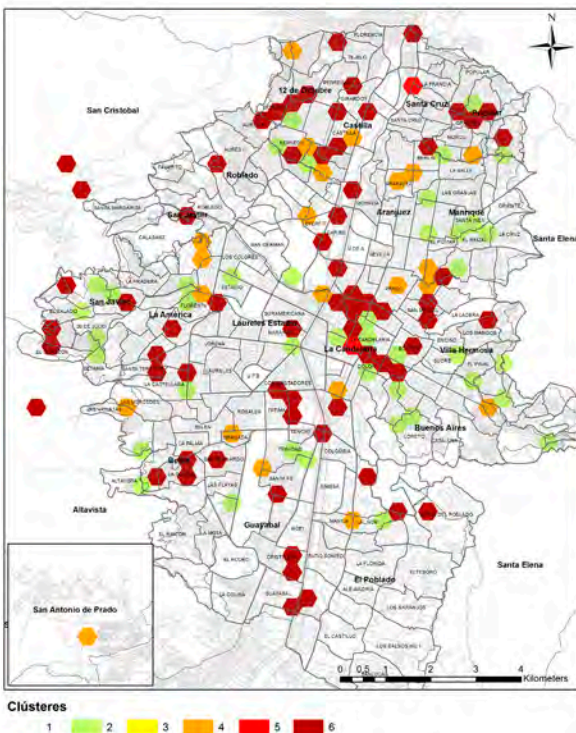
Fuente: Datos Policía Nacional. Elaboración propia.





Mapa 13. Clústeres, septiembre 2014

Fuente: Datos Policía Nacional. Elaboración propia.



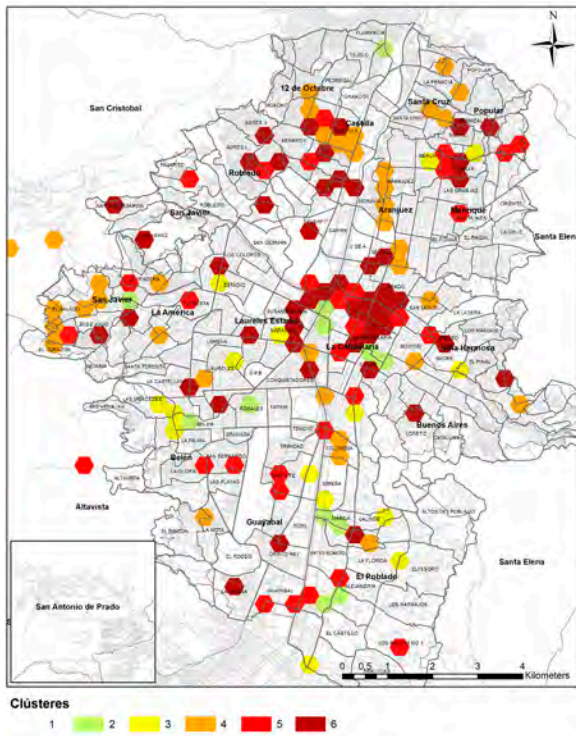
Mapa 14. Clústeres, noviembre 2014

Fuente: Datos Policía Nacional. Elaboración propia.

Al ver el comportamiento de estos meses, se destaca que en general los sectores van consolidándose de menor a mayor; es decir, los que están en nivel 2 o 3 tienden a convertirse durante el tiempo en sectores 5 o 6. No obstante, es más improbable que pase lo contrario: sectores que están en 5 o 6 pasen a convertirse en 1, 2 o 3. En algunos períodos de tiempo, como el comprendido entre mayo y julio del 2014, los sectores se mantienen entre 3 y 5, pero eventualmente aumenta su nivel de criminalidad hasta llegar a 6.

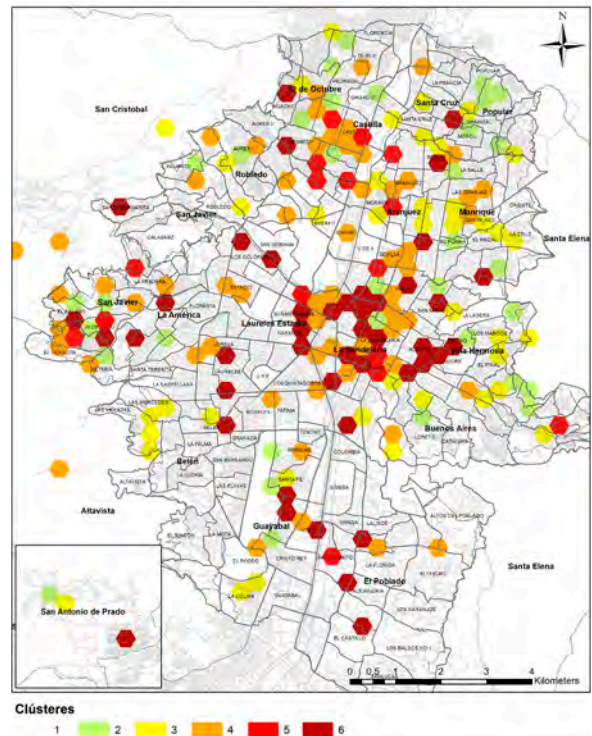
Al ver el detalle específico de este comportamiento y analizarlo en el ámbito espacial, se destaca que las zonas noroccidental y centro-occidental suelen variar a lo largo de los meses entre clústeres de baja y alta criminalidad, mientras que el centro de la ciudad conserva una criminalidad media y alta durante el período analizado. En estos 14 meses analizados también se ve un esparcimiento y reordenamiento de los niveles de las características de la criminalidad, al pasar de unos clústeres bastantes dispersos por toda la ciudad, a una consolidación de zonas, específicamente en los corredores mencionados antes. Una de estas zonas es la noroccidental, donde es importante mostrar que desde mediados del 2014 en adelante se consolida una zona con clústeres de alta criminalidad, que van marcando unos corredores que duran varios períodos continuos.

En este sentido, es significativo ver el mismo comportamiento pero para los clústeres de jóvenes. A diferencia del análisis total, en el de jóvenes es más difícil ver un comportamiento estable, toda vez que tienden a bajar y a subir durante varios períodos de tiempo. Además, vemos que suelen mantenerse más tiempo en niveles más altos de criminalidad, como se ve en los períodos a inicios del 2014, cuando se mueven alrededor del 4 y el 5:



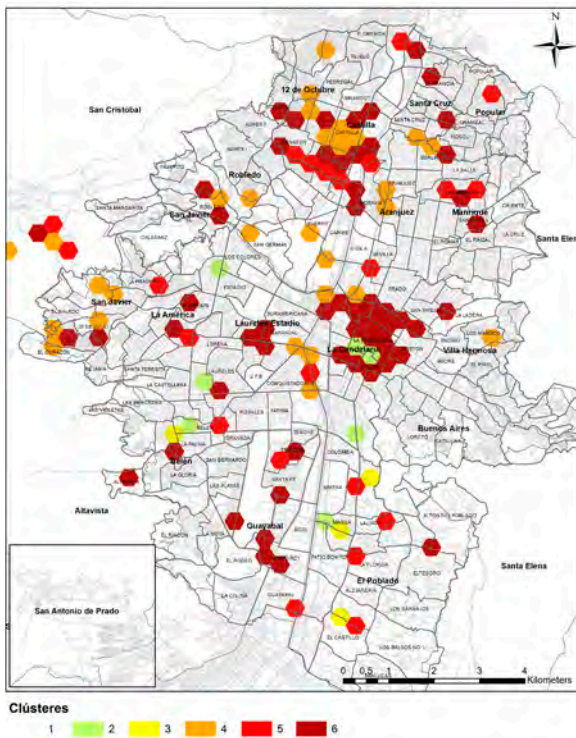
Mapa 15. Jóvenes clústeres, octubre 2013

Fuente: Datos Policía Nacional. Elaboración propia.



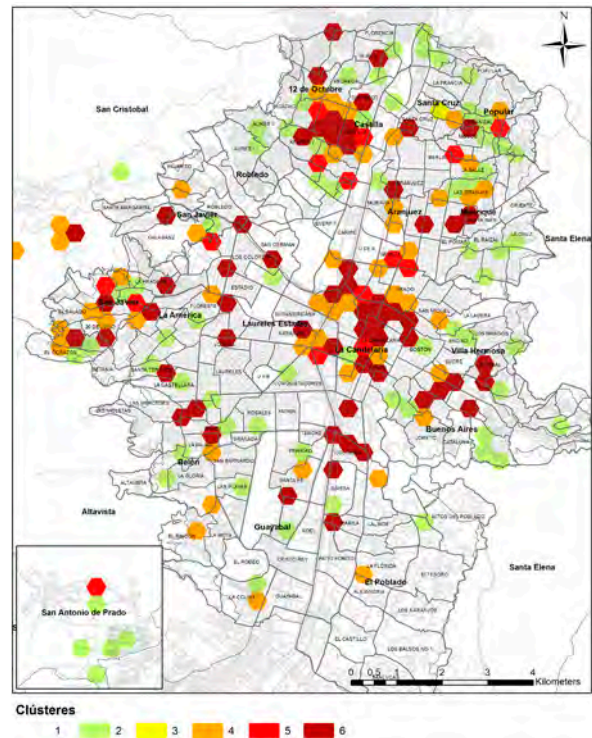
Mapa 17. Jóvenes clústeres, febrero 2014

Fuente: Datos Policía Nacional. Elaboración propia.



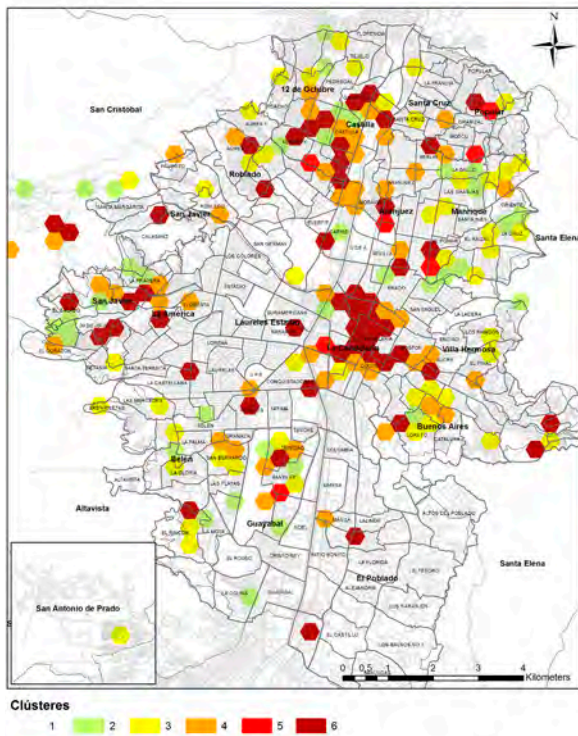
Mapa 16. Jóvenes clústeres, noviembre 2013

Fuente: Datos Policía Nacional. Elaboración propia.



Mapa 18. Clústeres jóvenes, agosto 2014

Fuente: Datos Policía Nacional. Elaboración propia.

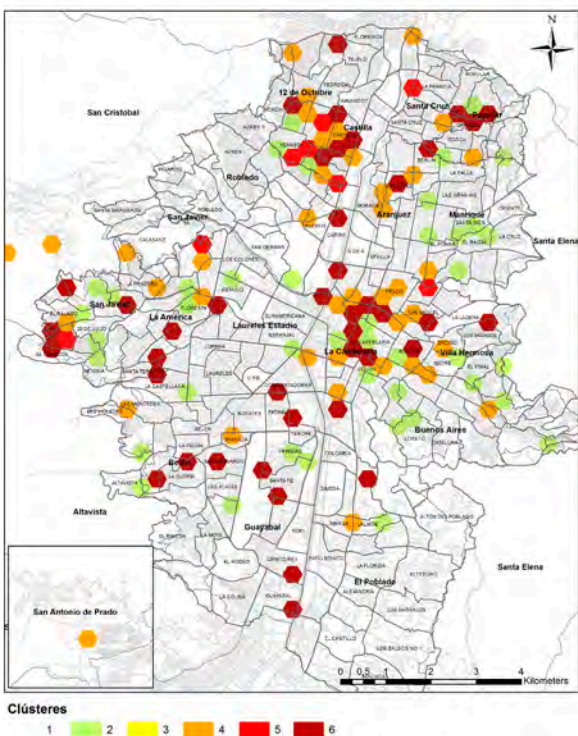


Mapa 19. Clústeres jóvenes, septiembre 2014

Fuente: Datos Policía Nacional. Elaboración propia.

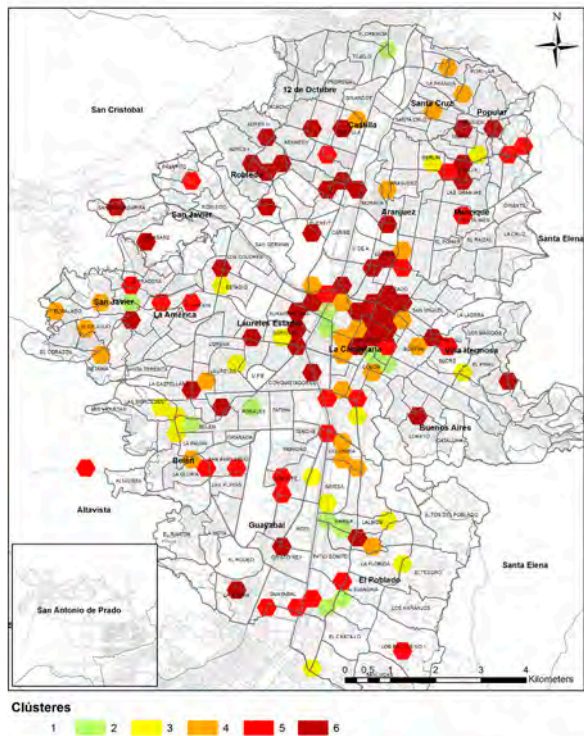
En resumen, los clústeres de jóvenes presentan una consolidación de sectores con alta criminalidad. A pesar de que constituyen los que tienen mayor variabilidad, en el período de tiempo analizado se van formando unos sectores estables en los corredores antes mencionados. Lo particular de estos sectores es que tienden a ser pequeños territorialmente, pero perduran en el tiempo.

Para contrastar lo anterior, a continuación se presentan los mapas mensuales más significativos para toda la población, exceptuando los jóvenes. Lo que sale de ello es que la variabilidad que aparece en el movimiento de fenómenos criminales suele darla la población joven, toda vez que sin este rango etario se ve poca variación entre meses, y se mantiene estable entre los niveles 2 y 5. Esta estabilidad se refleja en el ámbito espacial, toda vez que, a diferencia de los clústeres para el total de la población, en los de jóvenes no hay variabilidad en las zonas noroccidental y centro-occidental, como sí la hay en los mapas mensuales del total de la población.



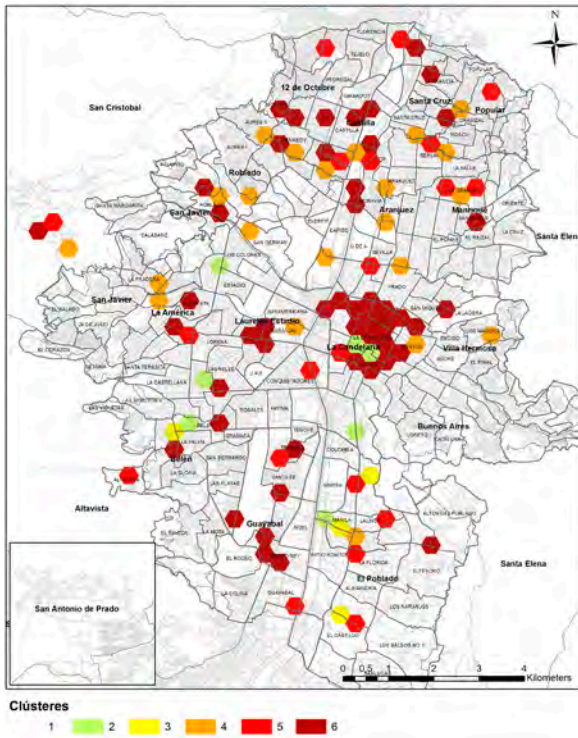
Mapa 20. Clústeres jóvenes, noviembre 2014

Fuente: Datos Policía Nacional. Elaboración propia.



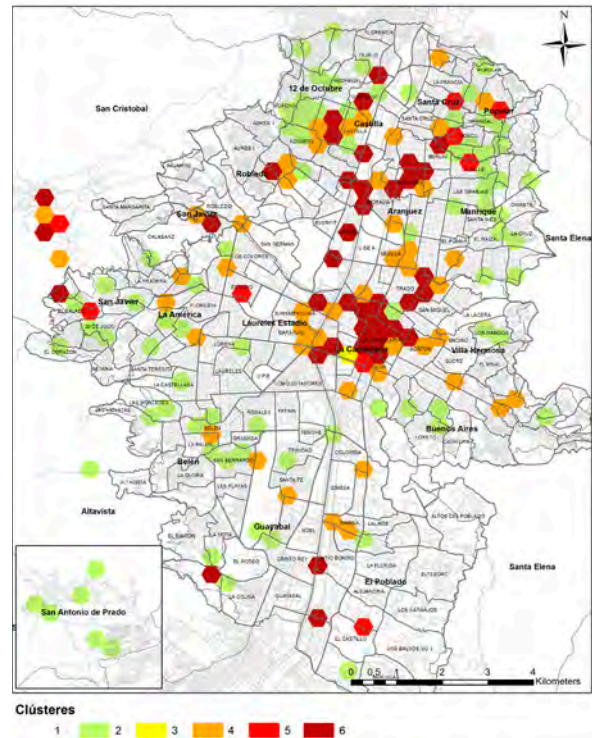
Mapa 21. Clústeres sin jóvenes, octubre 2013

Fuente: Datos Policía Nacional. Elaboración propia.



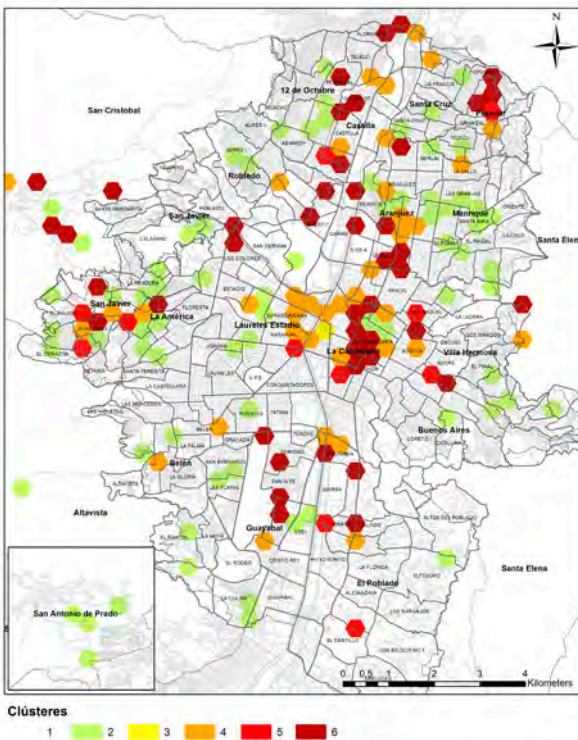
Mapa 22. Clústeres sin jóvenes, noviembre 2013

Fuente: Datos Policía Nacional. Elaboración propia.



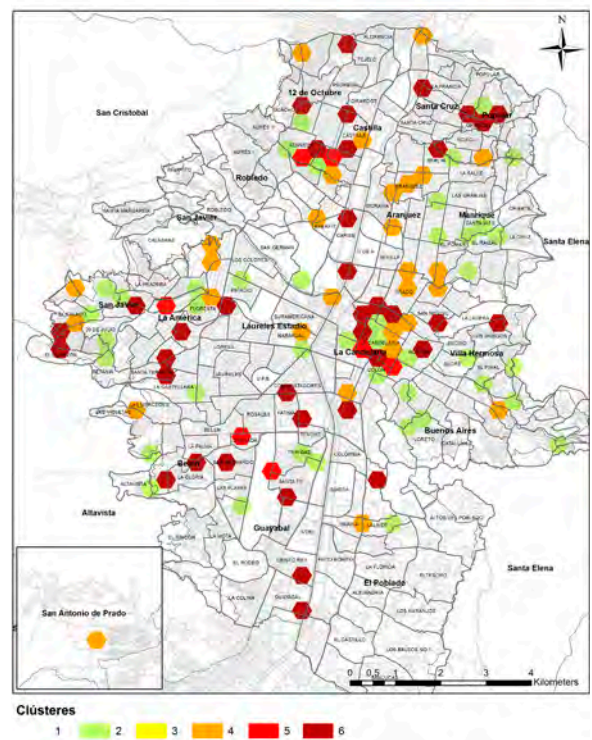
Mapa 24. Clústeres sin jóvenes, junio 2014

Fuente: Datos Policía Nacional. Elaboración propia.



Mapa 23. Clústeres sin jóvenes, mayo 2014

Fuente: Datos Policía Nacional. Elaboración propia.



Mapa 25. Clústeres sin jóvenes, noviembre 2014

Fuente: Datos Policía Nacional. Elaboración propia.

En estos mapas de clústeres sin jóvenes los corredores antes mencionados no presentan una durabilidad en el tiempo. No obstante, vale la pena mencionar que para los meses entre junio y septiembre del 2014, en casi todos los mapas hubo alta criminalidad en zonas específicas, sin importar la población que se analizara.

Discusión

Medellín ha tenido una reducción significativa en la tasa de homicidios en la última década; de esto se deducen muchas razones y se habla normalmente de una mejora progresiva. No obstante, cuando se trata de jóvenes, al mirar las tasas (tabla 7) para el período analizado vemos que son muy superiores a la tasa del total de la población.

Tabla 7
Tasas de homicidios por cada cien mil habitantes.
Total, jóvenes y menores de edad

Variable	Población total	Número de homicidios	Tasa de homicidios
Total	2.441.123	793	32,49
Jóvenes, 10-28 años	712.360	397	55,73
Menores de edad	555.878	61	10,97

Los jóvenes en Medellín, como víctimas o victimarios, siguen liderando los indicadores de criminalidad en la ciudad. Más allá de la consideración de que es un grupo etario que ocupa un tercio de la población, al hacer el análisis de clústeres, de compilación de delitos, el patrón no es diferente. En particular, vemos que la criminalidad de jóvenes consolida las dinámicas criminales en algunos sectores de la ciudad.

Sin embargo, vale la pena anotar que los sectores de más alta criminalidad, aquellos donde los jóvenes tienen elevada incidencia, son relativamente pequeños (de uno a tres barrios), mientras que los sectores con criminalidad baja, pero con niveles altos en otros delitos, como el de la violencia intrafamiliar, ocupan amplios corredores a lo largo de la ciudad.

Al mirar los clústeres exclusivamente de jóvenes, vemos que ocupan unos barrios muy específicos, una pequeña dispersión espacial, pero no

vemos que se amplíen en forma significativa a lo largo del tiempo, como sí lo hacen los clústeres de otras edades. Además, como se mencionaba antes, los clústeres de criminalidad baja y media (del 1 al 4) se esparcen y se mueven más fácil en el tiempo que aquellos de criminalidad alta (del 5 al 6).

En suma, el análisis de clústeres muestra que los de baja criminalidad, donde existen los delitos pero en nivel bajo, no tienen un patrón espacial estable y varían bastante con el paso del tiempo. Por otro lado, los clústeres de criminalidad alta tienden a formarse en pequeñas zonas, pero esto no aparece de repente, sino que son clústeres que van convirtiéndose, que se van formando y finalmente se consolidan: vienen desde el nivel 2 o 3 hasta llegar al 5 y 6, como se ve en el análisis mensual. Con esto es importante destacar que la consolidación de un clúster nivel 6 (rojo oscuro) está precedida por fenómenos alrededor de los estupefacientes (incautaciones o capturas), los cuales aparecen con mayor fuerza en el clúster tipo 4. Esto, según el período mostrado, puede presentar condiciones matemáticas para la predicción de fenómenos específicos en la ciudad.

Alrededor de esto es importante explicar que todos los clústeres suelen tener un nivel alto de homicidios, excepto el 1 y el 3 (blanco y amarillo, respectivamente), lo que habla de un fenómeno que está bastante esparcido a lo largo de la ciudad, pero que no siempre está ligado a las mismas características. Dicho lo anterior, es esencial mostrar que los clústeres rojos (5 y 6) suelen agrupar fenómenos delictivos donde están presentes muchas variables en nivel alto, incluyendo el homicidio, pero hay otros donde el homicidio es alto, pero está rodeado de otros delitos, como la violencia intrafamiliar y las dinámicas alrededor de los estupefacientes.

Para terminar, es importante mostrar que la clusterización permite ver que la criminalidad, o por lo menos la aquí agrupada, funciona bajo dinámicas de dispersión y consolidación. De modo específico, encontramos que durante 14 meses los fenómenos criminales suelen aparecer en torno a dinámicas de criminalidad baja, que primero emergen sin patrón alguno, luego se van agrupando, mientras va aumentando el nivel de criminalidad, hasta llegar a una consolidación de unas zonas específicas (y reducidas) en niveles altos (clústeres tipos 4, 5 y 6).

Esta clase de comportamientos, que destacamos para este tipo de agrupación de delitos, y en este período de tiempo, puede servir para optimizar zonas de intervención, y además para adelantarse a comportamientos de consolidación criminal en diferentes zonas. Lo esencial de este ejercicio es que puede ejecutarse en otros períodos

dos de tiempo, y se puede intentar con otras agrupaciones de delitos, según el perfil criminal de cada zona y las mediciones que se quiera obtener. Es importante destacar que esta técnica de agrupamiento puede ser ejecutada en cualquier período de tiempo, incluyendo diferentes variables, o incluso definiendo otras zonas, dado que es una de las técnicas más efectivas para mejorar la precisión. Este tipo de mejoras o cambios a la técnica se hace con el propósito de que esta no solo identifique o clasifique diferentes zonas, sino que permita efectuar evaluaciones en áreas definidas o en acciones efectuadas en esas áreas. Un cambio de clúster definitivo en el tiempo puede ser una buena medida de impacto o efectividad.

Fuentes consultadas

Bases de datos de la Policía Nacional de Colombia, 2013 y 2014.

Referencias

- Blair, E., Grisales, M. & Muñoz, A. M. (2009). Conflictividades urbanas vs. «guerra» urbana: otra «clave» para leer el conflicto en Medellín. *Universitas Humanística*, 67: 29-54.
- Devadoss, V. A. & Felix, A. (2013). Fuzzy clustering approach to study the degree of aggressiveness in youth violence. *International Journal of Computing Algorithm*, 2: 156-160.
- Di Martino, F. & Sessa, S. (2009). Implementation of the extended fuzzy c-means algorithm in geographic information systems. *Journal of Uncertain Systems*, 4 (3): 298-306.
- Fioredistella, D. & Mastrangelo, M. (2015). Men who kill women: Semantic maps for the identikit of the killer and murdered women. *Rivista Italiana di Economia Demografia e Statistica*, 69 (1): 191-198.
- Hammouda, K. & Karray, F. (2000). A comparative study of data clustering techniques [versión electrónica]. University of Waterloo. Recuperado de <http://www.pami.uwaterloo.ca/pub/hammouda/sde625-paper.pdf>.
- Ingram, M. & Curtis, K. (2014). Homicide in El Salvador's municipalities: Spatial clusters and the causal role of neighborhood effects, population pressures, poverty, and education [versión electrónica]. Wilson Center Latin American Program. Recuperado de https://www.wilsoncenter.org/sites/default/files/Homicides_El_Salvador.pdf.
- Loaiza, W. (2012). Distribución espacial del índice de propensión al homicidio (IPH) en las comunas de Cali, Colombia. *Perspectiva Geográfica*, 17: 169-192.
- Perversi, I., Valenga, F., Fernández, E., Britos, P. & García, R. (2007). Identificación y detección de patrones delictivos basada en minería de datos [versión electrónica]. Departamento de Ingeniería Industrial, ITBA. Recuperado de http://sedici.unlp.edu.ar/bitstream/handle/10915/20389/Documento_completo.pdf?sequence=1.