

Diplomado en Análisis de Información Geoespacial

Patrones aglomerados locales

Autora:
Claudia Coronel Enríquez

I El caso del análisis del crimen

La gente puede desconocer el riesgo que representan algunos lugares pero por sentido común sabe que la posibilidad de ser víctimas del crimen no es geográficamente constante. El análisis del crimen se dedica principalmente a localizar concentraciones del crimen o hot spots y gran parte del mapeo del crimen se dedica a su detección.

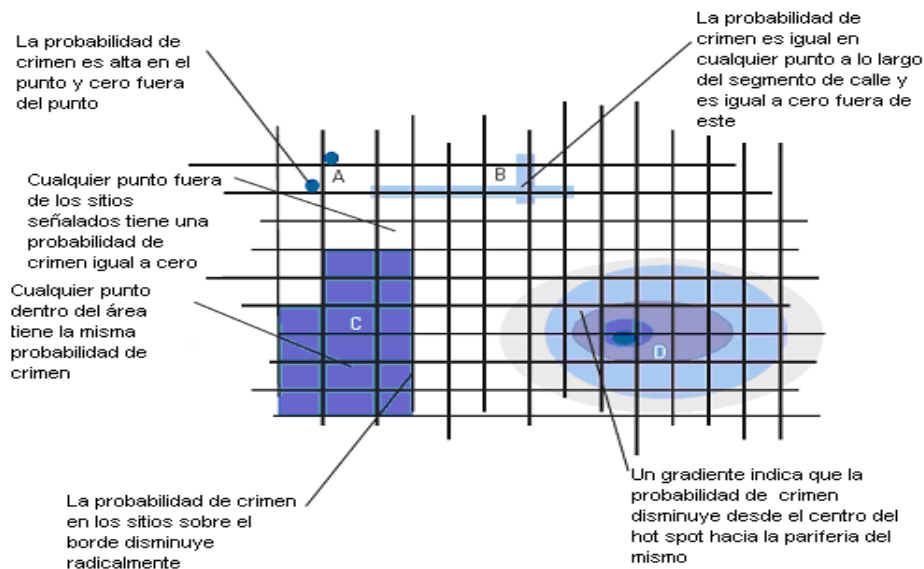
Un hot spot es un área que tiene un número superior a la media de crímenes o acontecimientos del desorden, o un área donde la gente tiene un riesgo más alto que el promedio de victimización.

Más importante, los factores que dan lugar a hot spot puntuales son diferentes de los factores que dan lugar a hot spots en calles, vecindarios conflictivos o ciudades con alta concentración de eventos criminales.

Identificar el nivel apropiado de análisis es crítico para entender el problema y determinar qué acción tomar. Las teorías del crimen son útiles para el mapeo del crimen solamente si se selecciona una teoría apropiada para el nivel de análisis y de acción.

Estas estrategias metodológicas son aplicables a cualquier proceso que se comporte con un patrón conglomerado local o hot spot. A continuación se muestran algunos ejemplos de la interpretación aplicable a diferentes representaciones de hot spots.

Mensajes en los mapas del crimen



Modificado de Eck, et al., (2005) Hot Spot.

II Herramientas de visualización y mapeo para identificar hot spots

Una preocupación central en el análisis de patrones de conglomerados locales es evaluar el grado de aleatoriedad espacial observado en los datos.

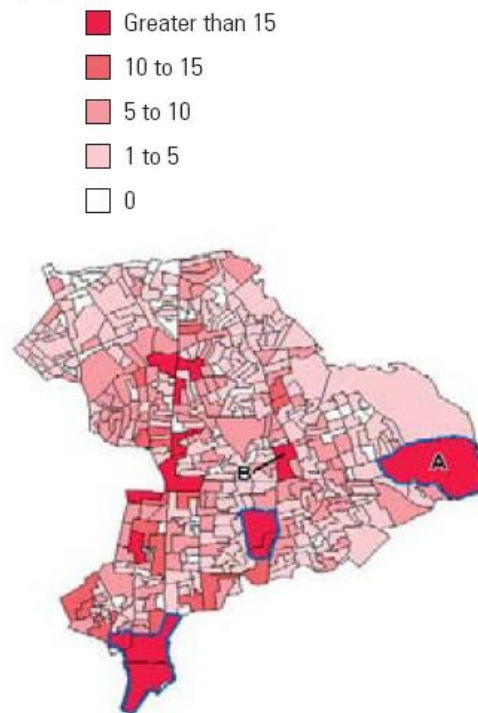
Los datos en formato vector representan características geográficas como puntos, líneas, o áreas, que se definen en términos de coordenadas X, Y. Los datos raster representan características geográficas como una rejilla de celdas en una superficie continua.

Para ello se han diseñado diferentes herramientas de visualización y mapeo aplicables a cualquier fenómeno estructurado espacialmente. En el caso que nos ocupa, estas herramientas permiten determinar comportamientos distinguibles por la concentración o casi ausencia de los procesos a estudiar.

a) Clasificación visual

Una herramienta muy empleada para destacar el comportamiento de los datos son los mapas de cloropletas, particularmente útiles para obtener un panorama general de la distribución espacial del proceso de interés. La distribución espacial de eventos del crimen, por ejemplo, se puede mapear utilizando unidades administrativas o estándares censales o bien mediante superficies.

Exhibit 9. Vehicle crimes mapped by census tract



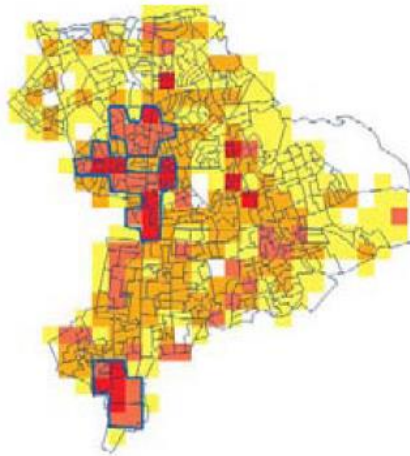
Eck, et al., (2005). Clasificación visual.

Clasificación visual por intervalos iguales de la variable, robos de vehículos por unidad de área.

Exhibit 10. 250-m quadrat thematic map



Vehicle crimes by 250-m quadrats



Eck, et al., (2005). Robo de vehículos.

Clasificación visual por intervalos iguales, empleando datos de densidad por unidad de área (celdas de 250m x 250m). Robos de vehículos por celda.

¿Qué diferencias observas en términos de la identificación de conglomerados locales de robos entre las dos representaciones?

- Mapas de cloropletas

Se clasifican visualmente los índices para cada unidad areal o las superficies, representando intensidades por unidad de área. Los mapas de percentiles y de caja nos dan una idea general del comportamiento de las variables pero también nos permiten identificar la presencia de hot spots en polígonos.

b) Análisis exploratorio de datos espaciales de eventos puntuales

La visualización y el mapeo son las herramientas más empleadas por los analistas de patrones hot spots, sin embargo se han desarrollado técnicas estadísticas robustas para facilitar el descubrimiento de patrones espaciales mediante el análisis exploratorio de datos espaciales (ESDA por sus siglas en inglés).

Estos métodos determinan si los eventos analizados se distribuyen aleatoriamente en el espacio o muestran evidencias de concentración o aglomeración.

1. Pruebas estadísticas globales o de primer orden.

Describen los patrones espaciales generales de los datos puntuales

- Media central.

La media de las distancias entre pares de puntos se calcula y se sitúa en el centroide del área cubierta por las observaciones.

- Desviación estándar de la distancia

Se calcula la desviación o dispersión en distancia de los puntos respecto de la media central.

- Elipse de la desviación estándar

Igual que el anterior privilegia la dirección en la cual se da la mayor dispersión de los puntos.

2. Pruebas preliminares de aglomeración en los datos

Permiten evaluar el grado en que la distribución espacial observada se asemeja a una distribución puramente aleatoria. Estas pruebas son de segundo orden porque evalúan los patrones locales o de vecindarios de datos puntuales.

- Índice del vecino más cercano (NNI siglas en inglés)

Es el cociente entre la distancia promedio observada de los vecinos más cercanos y la distancia promedio de una distribución aleatoria de los vecinos más cercanos.

Si es menor a 1 indica aglomeración, pero si es mayor a 1 indica un patrón uniforme en los datos. En el último caso, podría ser indicio de que el proceso es más regional.

- Prueba de autocorrelación espacial

Se estima mediante el índice de Moran o de Geary, indica dependencia entre las observaciones o autocorrelación espacial de la variable.

Entre más cercano a 1 o -1 indica fuerte autocorrelación y es indicador de heterogeneidad o aglomeración en los datos. Datos raster.

- Índice Ripley's K

Es una función basada en el índice del vecino más cercano, la cual compara el número de puntos observados en intervalos de distancia definidos respecto del número de puntos en una distribución aleatoria.

- Correlograma espacial

Se tratad de un gráfico que representa la variación del índice de Morán en intervalos de distancia. Con este correlograma se puede identificar el dominio de escala espacial de una variable. (Aplicable también a datos areales).

3. Técnicas para generar aglomerados en datos puntuales

Tradicionalmente la diferenciación espacial de comportamientos se ha determinado por métodos de extracción de regiones, es decir, por regionalización mediante métodos de clasificación.

En específico se han generado algoritmos que proceden a la clasificación o discriminación de clases, basados en estos métodos tradicionales estadísticos pero restringidos por criterios espaciales.

- Aglomeración espacial jerárquica

Consiste en determinar posibles agrupaciones de puntos empleando el método del vecino más cercano. En este caso el analista debe especificar una distancia umbral para la búsqueda de pares de puntos que se encuentren dentro del radio definido cuya distribución no sea resultado del azar, también se debe especificar un número mínimo de puntos para cada agrupación.

- Aglomeración K-means

Se selecciona a priori un número de clases a discriminar y el algoritmo define semillas o descriptores de clase para comenzar el proceso de aglomeración. En este caso los criterios de designación son: una varianza mínima dentro del grupo y el aumento de la diferencia entre clases.

- Mapas de autocorrelación espacial local. LISA

Consiste en evaluar la autocorrelación espacial entre observaciones sobre una variable y utilizar este parámetro para determinar clases por el valor y sentido de la autocorrelación.

Para ello se define un intervalo de distancia en el que se sabe que la dependencia espacial es significativa (basado en un correlograma espacial). No se trabaja con eventos aislados sino con la intensidad de la variable de interés, registrada para localidades o pares de coordenadas X, Y.

Estos mapas están asociados con el diagrama de dispersión de Morán donde las clases corresponden a la ubicación de las observaciones en los cuadrantes del diagrama.

4. Técnicas para aglomerados en datos de unidades areales (polígonos)

- Mapas de autocorrelación espacial local (LISA)

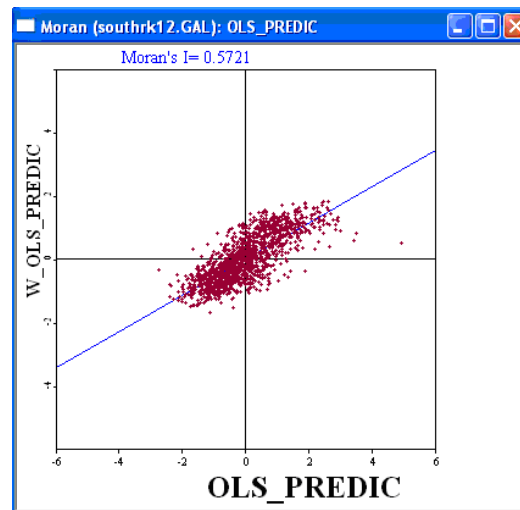
Al igual que los mapas de datos puntuales, consiste en evaluar la autocorrelación espacial entre observaciones sobre una variable.

En este caso se trabaja con intensidades o valores de una variable para un área definida, por ejemplo, número de homicidios por km^2 para un periodo de tiempo dado o índices para unidades administrativas como agebs o colonias. Pueden utilizarse para identificar el dominio de escala de la variable analizada.

- Diagrama de dispersión del índice de Morán

Se representa la variación de un proceso respecto de la misma variable, transformada en forma de promedios ponderados por vecindario. La pendiente del diagrama es el Índice de Morán para probar la presencia de autocorrelación espacial en los datos, su valor nos revela si los patrones observados en mapas de cloropletas se deben a la estructura espacial de la variable o al azar.

La disposición de las observaciones en los cuadrantes del diagrama de Morán indica el tipo de autocorrelación: positiva o directa o negativa o inversa. En el caso del esquema se observa que la nube de puntos se ubica principalmente en los cuadrantes con autocorrelación positiva.



Anselin, (2005).Diagrama de dispersión.

III Consideraciones generales sobre aglomerados locales

- Diferenciación social-Heterogeneidad ambiental

Un elemento clave en el análisis de hot spots es la comprensión de los procesos geográficos, que contribuye a la presencia o ausencia de un fenómeno de interés en el territorio. Por ejemplo, los lugares tienen características específicas que contribuyen a la diferenciación espacial y temporal del crimen.

Los procesos sociales son no estacionarios (la media no es constante en traslaciones). Pero aún si un mismo tipo de crimen es examinado en diferentes lugares, las relaciones entre variables serán diferentes debido a la heterogeneidad ambiental, en cierta medida. Esta heterogeneidad se expresa en forma de un gradiente, tablero de ajedrez, saltos, etc.

- Dependencia espacial

Para detectar la presencia de aglomerados locales, el grado de las relaciones espaciales entre observaciones debe identificarse.

Esta autocorrelación es conocida como dependencia espacial y está basada en la primera Ley de Geografía, la cual dicta que todo está relacionado con todo, pero las cosas más cercanas están más relacionadas.

La dependencia espacial debe medirse para establecer el rango de distancia en el cual un incidente está relacionado con incidentes vecinos.

- Dependencia temporal

Explorar primero, si es posible, la presencia de una autocorrelación temporal. Selecciona los periodos que sincronizan la dependencia temporal con la dependencia espacial en los análisis, por ejemplo, rango de días, estaciones, eventos, implementaciones de políticas públicas, etc.

Normalmente, una buena representación en el contexto temporal consistirá en series de mapas hot spot diferentes entre sí en cuanto a los patrones espaciales.