

La metodología de #HomicidiosCDMX

Los datos

La base de datos con la que se realizó el presente estudio proviene de un convenio firmado entre la Secretaría de Seguridad Pública de la Ciudad de México y México Evalúa. El convenio de colaboración contempla el intercambio de información para fortalecer el trabajo de la SSP-CDMX, diseñar mejores políticas públicas e incrementar la comprensión que se tiene del fenómeno criminal.

La base de datos contiene información generada por el primer respondiente y consiste en lo reportado por los policías que atienden avisos de algún evento delictivo. En este caso, la información contiene exclusivamente datos generados por los reportes de homicidios intencionales en la Ciudad de México para el periodo 2009 a 2016.

La base original contenía 6,418 expedientes de reportes de homicidios intencionales generados por el Primer Respondiente. Los reportes eran a partir de eventos y no de víctimas por lo que un evento podía tener más de un cuerpo registrado. La base de datos incluía los campos de fecha, calles de ocurrencia, colonia, delegación, sector policial, zona, hora de reporte, región policial y un campo con una descripción básica de los hechos.

En su versión original encontramos varias inconsistencias en la información. En específico, centramos nuestra atención en verificar la calidad de la georreferenciación. El primer paso fue contrastar los campos de calles, colonias y descripción para corroborar que los puntos hubieran sido reportados en la dirección correcta. Gracias a esta revisión se encontró que para 2009, 2010 y 2011, entre el 70% y el 80% de las observaciones estaban mal georreferenciadas. Por alguna razón, la ubicación correcta estaba a unas cuadras hacia el noreste de la coordenada que se había incluido en la tabla. Por el contrario, en los años restantes solo entre el 7% y el 13% de las observaciones tenían problemas en la georreferenciación. Adicionalmente, se utilizó la información contenida en el campo sobre la descripción del evento para verificar si la ubicación era la correcta. El proceso de corrección de coordenadas se realizó en *Google Maps* con el formato UTM, sistema de coordenadas WGS 84, zona 14N.

Durante la revisión se encontraron algunos casos en los que no fue posible señalar la localización exacta del evento. Tres fueron las principales razones: se desconocía la ubicación del hecho por lo que no estaba georreferenciado; el evento sí estaba ubicado pero la descripción correspondiente señalaba que el homicidio no sucedió en el lugar en donde se encontró el cuerpo; y la ubicación reportada era el hospital donde había ocurrido el fallecimiento.

Otro criterio para descartar observaciones fue cuando la información describía eventos distintos al homicidio intencional, como suicidio, homicidio culposo u otro delito. De forma similar se descartaron algunas observaciones en las que la descripción no aportaba los suficientes elementos para asegurar que se había cometido un homicidio intencional.

Gracias al trabajo de limpieza de datos se descartó el 27% de las observaciones. Las observaciones que finalmente se mantuvieron en la versión final están correctamente georreferenciadas y corresponden al fenómeno criminal que nos interesa estudiar.

Para #HomicidiosCDMX generamos dos tipos de mapas: mapas de calor y mapas de conglomerados espaciales (*clusters*).

Mapa de calor

Este proceso estima la variación de la densidad de eventos a lo largo de un área a partir de los puntos donde se ubican los eventos de interés. Con esta información se genera un mapa donde los valores de densidad reflejan la concentración de los puntos a partir de un área determinada, otorgándole a la mayor acumulación de puntos en un área valores más altos que a los puntos que se encuentran más dispersos. Con esta técnica, se pueden generar visualizaciones más precisas de los datos que con el simple mapeo de puntos o de tasas crudas, porque a diferencia de las tasas o los conteos simples, con este proceso el grado de dispersión en un área importa. Sin embargo, los conglomerados identificados con esta técnica no son conglomerados espaciales, ya que no han sido sometidos a una prueba estadística que verifique su covarianza con respecto a las áreas vecinas. Es decir, no se ha rechazado la hipótesis nula de Aleatoriedad Espacial.

Mapas de conglomerados espaciales

Estos mapas son parte del Análisis Espacial Exploratorio o *Exploratory Spatial Data Analysis* (ESDA). Esta herramienta consta de un conjunto de técnicas para la descripción de los datos espaciales y para preparar el camino para la implementación de los modelos espaciales. El análisis exploratorio nos permite describir y visualizar la distribución espacial de los datos para identificar patrones de asociación espacial o conglomerados (*clusters*) y ubicaciones atípicas.

Entre los elementos más usados en el análisis exploratorio están los índices de autocorrelación espacial a nivel local. La autocorrelación espacial es un procedimiento estadístico que detecta la presencia de covarianza entre los atributos de un área y los de las áreas circundantes. Para examinar la autocorrelación local se empleó la I de Moran (LISA). Con los índices locales se pueden identificar observaciones agrupadas en zonas y generar categorías sobre el tipo de asociación a la que pertenecen. Esta técnica puede identificar al menos cuatro tipos de categorías: los conglomerados de Alto-Alto (AA), los de Bajo-Bajo (BB) y dos tipos de valores atípicos espaciales (AB y BA). Las primeras dos categorías son conocidas en el ámbito no-técnico como *hot spots* y *cold spots*.

Un grupo de localidades recibe la categoría de *hot spot* o *cold spot* porque tiene valores por arriba o por abajo de la media de la variable de interés. El área identificada no representa un conglomerado, en realidad, es el centro de un conglomerado porque la categoría de *hot spot* o *cold spot* la adquiere cuando los valores de la variable covarían en la misma dirección y magnitud que los valores de las áreas que lo rodean. Para el caso de los valores atípicos espaciales, el área de interés es solo la localidad identificada.

Estos mapas se interpretan de la siguiente manera. Los *hot spots* y *cold spots* nos indican que el evento de interés (homicidios) a nivel Área Geoestadística Básica (AGEB) se concentran en una zona porque existe una correlación entre el evento y el espacio en donde ocurren (autocorrelación espacial). Se clasifican como *hot spots* porque la magnitud del evento supera la media del evento en el resto de unidades. Son identificados como *cold spots* cuando los valores están por debajo de la media. Los valores atípicos indican una localidad con alta magnitud rodeada de localidades con valores por debajo de la media (AB) o una localidad con baja magnitud, pero rodeada de localidades con valores por arriba de la media (BA).