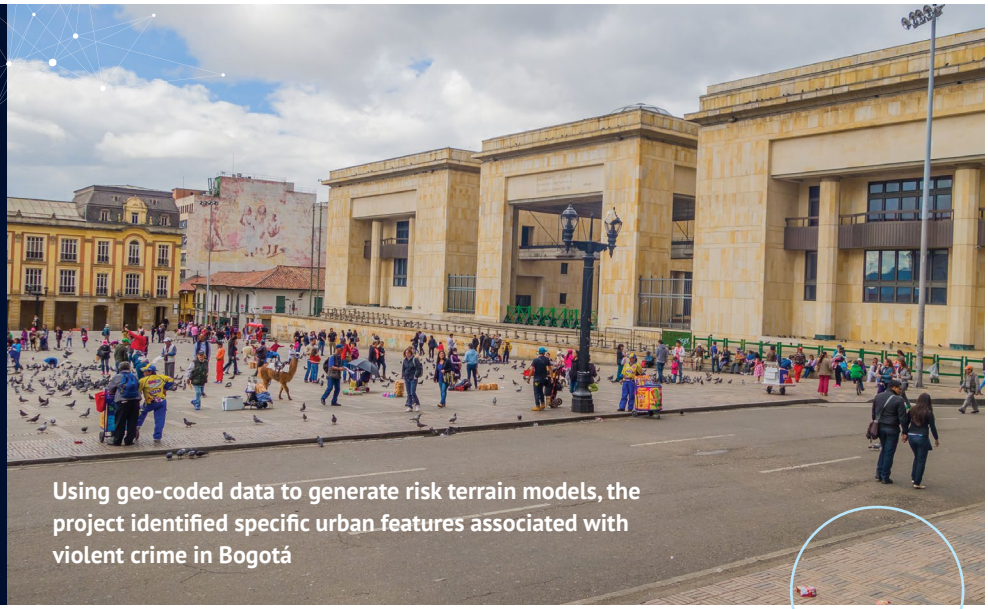
BIG DATA
CASE Study

WORLD BANK GROUP



Using geo-coded data to generate risk terrain models, the project identified specific urban features associated with violent crime in Bogotá

Understanding How Infrastructure Affects Crime

Camila Rodríguez and **Andrés Villaveces**

SUMMARY

Latin America is highly urbanized, with above-average crime rates. Its cities are typically unplanned, with high socioeconomic inequality, yet the association between crime and infrastructure has not been clearly defined or quantified. Colombia's capital, Bogotá, collects considerable geo-coded data on urban infrastructure and has reliable geo-coded information on population and crime. The recent development of the world's largest Bus Rapid Transit system has led to the modification of infrastructure in several parts of Bogotá. These changes present an opportunity for studying the association between crime and infrastructure. Drawing on rich data, this project quantified the occurrence of crimes in relation to specific characteristics in the built environment. Through risk terrain modeling (RTM), the team identified locations near public hospitals, schools, drugstores and bus stations as being associated with assault and homicide. The modeling also revealed peak times of day for crime, and predicted areas of the city more likely to experience future crime. Combined with local stakeholder perspectives, RTM analyses can reliably suggest action to reduce crime associated with particular environmental factors. The methods are widely applicable in other locations and for other crimes.

CHALLENGE

Latin America is one of the world's most urbanized regions, with crime levels higher than the global average. Urban violence and crime disproportionately affect young, economically active populations, which are the region's largest segment. Yet the effects of urban characteristics on crime in Latin American cities are little studied. Understanding of this relationship could inform urban planning that helps deter crime.

Among initiatives to address Bogotá's transport problems has been the creation of the world's largest Bus Rapid Transit (BRT) system. The development of the system's trunk lines includes several modifications to the cityscape. This project sought to evaluate the association between crime and infrastructure in Bogotá using the BRT trunk routes as an intervention and comparing them with other areas. Carried out in collaboration with Rutgers University, the study aimed to generate reliable estimates of risk for different crimes around BRT stations and throughout Bogotá. If it could quantify crime variations and their relationship with population flows through the different bus stations, this would help city authorities understand the impact of urban infrastructure modifications in relation to crime. Such information could inform future urban planning to help create environments that are non-conducive or even inhibitive to crime. The study also sought to highlight and estimate risks of crime in 18 different urban features, including schools, libraries, tourist attractions, bridges and clinics. These findings would have implications for cities elsewhere in the world, especially those with similar BRT systems.

Clusters of crime can be partially explained by landscape features that attract criminal behavior at certain times

INNOVATION

To capture Bogotá's overall characteristics, the team drew on rich infrastructure data, including land-use information, service network data (gas, water, sewage) and city block and street audit information captured via previously validated tools. Geo-coded crime data from 2012 then allowed them to estimate the correlation of homicide and assaults with overall city infrastructure features, land-use patterns, socio- economic level, cadastral information, socio- economic survey data and BRT or non-BRT sections of the city. The BRT analyses included data on population flows through stations to understand crime variations related to population density at specific times of the day.

The crime data were analyzed in different ways. The team first conducted a Nearest Neighbor analysis to assess clustering within the distribution of crimes. The results suggested that the distribution of crimes in Bogotá is significantly clustered. Kernel density mapping was then used to identify where, at more localized places within the study area, the highest concentrations of incidents of crime occur. Using hotspot analysis, the team found many micro-level locations around which crimes cluster. This patterning was statistically significant.

Drawing on the findings from these exercises and combining them with the infrastructural data for Bogotá, the project generated 16 different risk terrain models (RTMs) for assault incidents across the city. This modeling process uses a specific algorithm that identifies relationships between different layers of data and correlates them with crime using count regression models which are then linked to places on a digitized map. The approach represents spatial influences of crime risk factors as common geographic units, then combines separate layers of map (one per risk) to produce maps showing the intensity of

all risk factors at every location throughout a landscape – the 'risk terrain'. Risk terrain maps show where conditions are most conducive to crime. They help diagnose why crimes have clustered at certain places, and can help forecast where they are likely to occur in the future.

The 16 risk terrain models covered:

- ***The entire city for one calendar year***
  City-wide risk terrain models suggest that different economic strata (Low, Medium and High) of Bogotá correlate with crime incident locations, but each stratum influences assault and homicide differently. Low stratum blocks have an increased likelihood of both homicide and assault. Proximity to drug stores and medical clinics presented consistently high risk, with at least 50 percent greater likelihood ofeither assault or homicide. Places close to hospitals, public schools or BRT stations also presented significantly higher risks of assault or homicide compared to places lacking the defining landscape features of these public facilities (such as large pedestrian access routes).
- ***Different economic strata for one calendar year***
  To understand temporal and spatial incentives to criminal behavior within each stratum, the team performed hourly-based RTM analyses. A heat map summarizing the incidence of crimes by hour and day of the week indicated that violent crime clusters at certain times, peaking from Saturday evening until Sunday dawn.
- ***Highest concentration of assault incidents***
  Several RTM analyses explore these peak times in greater detail. The results suggest that proximity to drugstores and medical clinics increases the risk of assault and homicide significantly in low stratum sections of Bogotá. In medium stratum sections, proximity to public schools increases the likelihood of assault. Overall, crimes tend to occur mostly in the evenings.
- ***Peak and off-peak hours***
  Hourly-based risk terrain models were carried out by stratum, for peak and off-peak hours on the BRT network. These suggest that the risk of assault more than doubles near medical clinics during peak hours in lower stratum sections. In medium stratum sections, the likelihood of assault increases near private schools. The likelihood of homicide more than quadruples near BRT stations during peak hours in low stratum sections. The results suggest that crimes occur in different places at different times of the day.
- ***Gender-based maps for one calendar year***
  Gender-based risk terrain models were generated for each economic stratum to analyze the locations of assaults on female and male victims. The results suggest that proximity to medical clinics and drugstores in low stratum sections almost doubles the likelihood of assault for both male and female victims. In medium stratum sections, being near tourist attractions increases the risk of assault on men, while proximity to private schools increases the likelihood of women being assaulted more than fivefold.

RESULTS

The results indicated that incidents of both homicide and assault cluster at certain places. This can be partially explained by certain features of the landscape that attract criminal behavior at certain times, linked to changing density of people. Such crime patterns occur beyond random chance and are statistically significant.

Findings were presented to key local and national stakeholders in Bogotá, who fed back useful

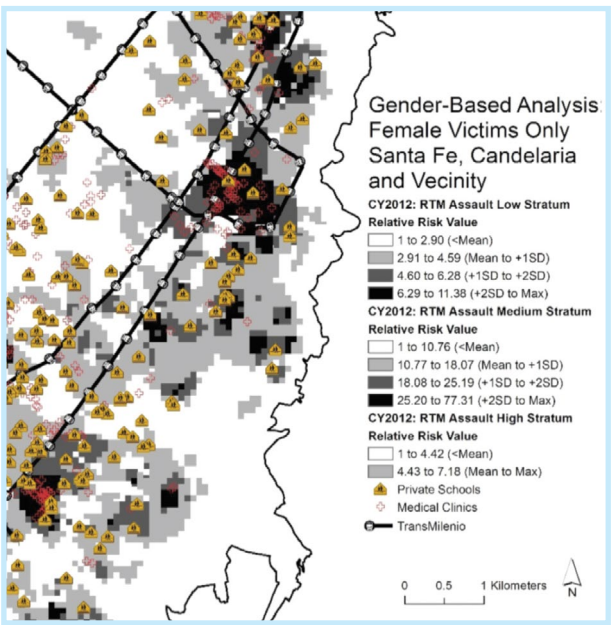
information for contextualizing the results. They suggested the clustering of homicide incidents during peak hours near BRT stations in low stratum sections was related to known criminal activities organized by certain 'street vendors' near a few BRT stations. They also explained that many people go to drugstores to buy prescription drugs unobtainable at hospitals. Offenders near these locations may see these people with cash as potential targets. Drug micro-trafficking may also be a factor, particularly in lower stratum sections.

RTM offers insights into the spatial dynamics of crime and how to mitigate key factors leading to criminal behavior, for example, by environmental modifications that improve passive surveillance and law enforcement activities. The approach can be applied to other crimes, such as residential or commercial robbery and vehicle thefts. The team has encouraged various entities to perform their own RTM analyses. Local stakeholders have already used the methodology to create maps of drug micro- trafficking in five Colombian cities for the Ministry of Justice. Their insider perspective increases the reliability and practical value of the RTM analyses.

LESSONS LEARNED

The project clearly showed how big data analytics reveals correlations which inform strategies to reduce crime.

- ***Draw on big data to inform decision-making***
  RTM is useful as a reliable diagnostic tool that can evaluate patterns of crime and orient prevention and enforcement activities more efficiently – including towards specific areas and at specific times. Being cross- sectional, the methodology established certain correlations, rather than causal associations, but in doing so it highlighted certain features of the environment that otherwise would not easily be detected by simpler statistical modeling.
- ***Use big data analytics to assess disparate data sources***
  Big data approaches are useful for integrating disparate but complementary information to provide a very rich environment that can be analyzed to respond to key questions – in this case, about crime. Analyzing big data can also establish emergent patterns of correlation or association which can be related to specific outcomes of interest.
- ***Involve local stakeholders to help explain findings***
  Insider perspectives and content analysis with local stakeholders help identify potentially plausible – though not causal – explanations for crime, given their knowledge of the context. Working with local stakeholders is important for generating explanatory hypotheses.



Risk terrain models help diagnose why crimes have clustered at certain places and forecast where they are likely to occur