Check for updates

RESEARCH ARTICLE

Multivariate analysis and characterization of low impact crime in Mexico City [version 1; peer review: 1 approved, 1 approved with reservations, 1 not approved]

Johny Pambabay-Calero ¹, Sergio Bauz-Olvera¹, Rubén Flores-González², Carlos Piña-García³

¹Faculty of Natural Sciences and Mathematics, ESPOL, Polytechnic University, Guayaquil, Guayas, 09-01-5863, Ecuador

²Centro de Estudios de Opinión y Análisis, Universidad Veracruzana, Xalapa, Veracruz, 91030, Mexico

³LARSI, Universidad Veracruzana, Xalapa, Veracruz, 91030, Mexico

V1 First published: 21 Dec 2021, 10:1299
<https://doi.org/10.12688/f1000research.72990.1>

Latest published: 21 Dec 2021, 10:1299
<https://doi.org/10.12688/f1000research.72990.1>

Abstract

Background: Multivariate analysis is fast becoming a key instrument that can be used to address crimes or incidents. It may be helpful to assess government policies on crime prevention.

Methods: To distinguish between the 25 official crime categories reported in Mexico City (Jan 2019 - Jun 2019), principal coordinates analysis was used to determine the quality of a characteristic in this context. This study used cluster analysis via K-means and Biplot based on time and location in terms of crime occurrence.

Results: The results obtained from this preliminary analysis indicates that around 70% of crime occurrence is shown in the following boroughs: Cuauhtémoc, Iztapalapa, Gustavo A. Madero, Benito Juárez, Álvaro Obregón, Coyoacán, and Miguel Hidalgo.

Conclusion: There are two factors that contribute to the difficulties in crime analysis in Mexico City, namely, the lack of people's trust in authorities and the insufficiency of tools for data analysis. The latter is an integral part in achieving justice for the victims of crimes because it impedes the process of observing patterns and predicting the perpetrators' next actions, which may help in solving a number of types of crimes. It is then imperative for law enforcement to utilize data analysis tools that aid in identifying crime patterns and trends, such that the occurrences of crime show a downward trend and consequently increase the people's trust in law enforcement agencies.

Keywords

biplot, crime analysis, k-means clustering, multivariate analysis, principal coordinates analysis

Open Peer Review

Approval Status

	1	2	3
version 1
21 Dec 2021	view	view	view
<hr/>			
1. Nathalia Tejedor Flores , Universidad Tecnológica de Panamá, Panama City, Panama			
2. N. Sofía Huerta-Pacheco , Universidad Nacional Autónoma de México, Mexico City, Mexico			
3. José Luis Vicente Villardón , University of Salamanca, Salamanca, Spain			

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Johnny Pambabay-Calero (jpambaba@espol.edu.ec)

Author roles: **Pambabay-Calero J:** Conceptualization, Methodology, Software, Writing – Original Draft Preparation; **Bauz-Olvera S:** Conceptualization, Methodology, Software; **Flores-González R:** Methodology, Resources, Writing – Review & Editing; **Piña-García C:** Methodology, Resources, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2021 Pambabay-Calero J *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Pambabay-Calero J, Bauz-Olvera S, Flores-González R and Piña-García C. **Multivariate analysis and characterization of low impact crime in Mexico City [version 1; peer review: 1 approved, 1 approved with reservations, 1 not approved]** F1000Research 2021, 10:1299 <https://doi.org/10.12688/f1000research.72990.1>

First published: 21 Dec 2021, 10:1299 <https://doi.org/10.12688/f1000research.72990.1>

Introduction

Lack of meaningful crime data has existed as a social problem for many years in Mexico City. However, there is a growing interest in extracting useful knowledge from data related to crime reports. Crime analysis is a major problem due to the lack of trust in authorities and insufficiency of data. Therefore, statistical methods and applications of pattern discovery are necessary to study the impact of crime on communities. In this regard, it would be necessary and advisable to provide important insights into data crime analysis (Chadee *et al.*, 2019). Hence, it could conceivably be hypothesized that these results could be used to prevent crime and to develop tailored strategies to tackle crime and gradually build community trust.

It should be noted that this type of research involves a multidisciplinary approach i.e., different academic disciplines including social sciences, statistics, economics, and data science are combined to provide the necessary knowledge, skills, and confidence to analyze and compute data related to crime. Thus, statistics becomes a fundamental discipline to gain insights in terms of displaying, analyzing, interpreting, and presenting data.

This manuscript reviews the evidence for official crime data in Mexico City with the aim to determine the extent of crime occurrences and identify statistical features that permit us to establish variables that can provide the basis for quantitative characterization of crime reports (D'ancona and Ángeles, 1999). It is hoped that this research will contribute to a deeper understanding of crime analysis through the data availability provided by the government of Mexico City (Procuraduría, 2018).

Recently, researchers have shown an increased interest in data collection methods to predict and analyze criminal behavior (Aghababaei and Makrehchi, 2016; Malleson and Andresen, 2015; Wang and Gerber, 2015). However, most of these studies have been carried out in the U.S.A. (Chen, Cho, and Jang, 2015; Flores, 2017; Yadav *et al.*, 2019; Yang *et al.*, 2018). Another interesting study was based on methamphetamine-related crime using Google trends (Gamma *et al.*, 2016). Similarly, there is a growing body of literature that recognizes the importance of using social networks analysis as data sources (D'Avanzo *et al.*, 2017). Recently researchers have examined the effects of crime patterns and antisocial behavior in Mexico City through Big Data-driven models for predicting crime rates (Piña-García and Ramírez-Ramírez, 2019).

In Mexico City, there is an urgent need to address the safety problems caused by antisocial behavior. In this context, the main disadvantage of crime analysis in Mexico City is the lack of trust in authorities and insufficiency due to a lack of reliable data. Another common obstacle faced by many researchers is the inaccurate way in how crime reports in Mexico City have been recorded. Most of these records show a low-quality dataset that involves significant issues such as: missing data, human errors, and inconsistencies in crime information. Despite the limitations mentioned above, it is still possible to obtain and identify the variability of crime trends (Arango, 2003).

It is important to note that traditional surveys are still an acceptable and reliable proxy to explore opinion about crime occurrence among citizens. These studies are mainly carried out by the National Survey of Victimization and Perception on Public Safety¹ [ENVIPE], the Victimization and Institutional Effectiveness Survey² [ENVEI], the National Institute of Statistics and Geography³ [INEGI] and the Center for Educational Research and Development⁴ [CIDE] (Bergman and Sarsfield, 2009; Jasso López, 2013; Vilalta-Perdomo, 2010). In this regard, there is still uncertainty in these surveys-based studies as they may be biased due to inaccurate information e.g., false rumors, fake news, and hoaxes (Cisneros and Cunjama-López, 2011).

It was after the Mexico City elections in 2018 and the open data initiative proposed recently by the Mexico City government that it was possible to gain access to crime reports committed from Jan 2019 - Jun 2019 (Piña-García and Ramírez-Ramírez, 2019). This dataset shows a significant enhancement in terms of data structure, updated information, geolocated info, and computing performance. By providing a new approach mainly based on an open policy, the Mexico City government have released a significant amount of data related to crime reports that contain information such as date, location, and category of crime, among others.

Much of the previous research on crime analysis has been exploratory in nature. A significant analysis and discussion paper on the subject was presented by Fuentes-Flores and Sánchez-Salinas (2015). This research examined the use of

¹<https://www.inegi.org.mx/programas/envipe/2019/>

²<http://hdl.handle.net/10089/16309>

³<https://www.inegi.org.mx>

⁴<https://www.cide.edu>

institutional data to provide information on crime activity. Likewise, one longitudinal study found that homicide prevalence is associated to those places located around areas that are not occupied or lived in by people (Fuentes-Flores, 2015). Auto parts theft has been linked to social aspects such as: lack of scholars degree (lack of further education) and density of population (Sánchez-Salinas and Fuentes-Flores, 2016). In addition, Vilalta (2011) examined the trend in auto parts theft and described its regular fluctuations along short periods of time. In a previous study (Vilalta-Perdomo, 2009), it was observed that firearm carrying, the hour of the day, and offender prevalence were linked to the degree of violence displayed in criminal aggressions.

It is important to note that previous studies on crime exploration have highlighted several social factors such as: friendship networks on a local level, social relegation, ethnic and racial characteristics, and breakdown of parental relationships (Kubrin and Weitzer, 2003; Sampson and Groves, 1989; Felson and Clarke, 1998).

To better understand the mechanisms of crime patterns and its effects, Clarke (2016) analyzed data from the routine activity theory perspective to determine how crime occurs when three elements converge: (1) a motivated offender, (2) a suitable target, and (3) the absence of a capable guardian.

This study systematically reviews the Mexico City crime data with the aim to evaluate the impact of crime occurrence in an in-depth statistical analysis to identify spatio-temporal patterns of crime. In addition, this manuscript attempts to show that there are crime categories with a high rate of concentration. Thus, this research should make an important contribution to develop and encourage tailored strategies to tackle crime, and to evaluate government programs or policies related to crime prevention (Bandrés and Díez-Ticio, 2001).

The rest of the paper is structured as follows. The methods section is concerned with the techniques that are used in the statistical analysis for this study. Preliminary results will be given in a systematic and detailed way in the results section. Finally, a summary explaining the significance of these findings will be highlighted in the conclusions section.

Methods

One of the most well-known tools for identifying historic data, patterns, and trends is descriptive analytics. These analytic techniques are intended to identify measure of central tendency and statistical distribution. Moreover, there are methods such as principal co-ordinates analysis (PCoA) and K-means clustering that may be useful when *a-priori* knowledge is vague and cannot be fully described. These two techniques mentioned above will be our proposed framework for this study.

The biplot graph in the context of principal coordinates analysis is one of the more practical ways of displaying observations and variables of multivariate data in the same plot (Gabriel and Odoroff, 1990; Gower, 1966). In this regard, samples are displayed as points while variables are displayed either as vectors, linear axes, or nonlinear trajectories. In the case of categorical variables, category level points may be used to represent the levels of a categorical variable (Pambabay-Calero *et al.*, 2018).

In order to identify antisocial behavior and crime patterns, the Mexico City Police Department recorded 115,209 offenses from January 2019 to June 2019 through the Portal de Delitos del Departamento de Policía de la Ciudad de México web page.⁵ This information related to crime reports was made accessible to the public and shared via (Procuraduría, 2018). The procedure for generating the data was as follows:

- Access the Mexico City Police Department Crime Portal web page.⁵
- Access to Mexico City FGJ investigation folders.⁶
- Select the time period to the 1st January to 30th June 2019, due to the reclassification of crimes in July 2019.

Figure 1 provides a flowchart about our three stages approach: 1. Understanding, data gathering, and preparation from the Portal de Datos de la Ciudad de México web site (in order to get the data in the best format for multivariate statistical analysis). The data were organized in such a way that the variables were the columns and the observations (crime report) as rows. The first row represented the variable titles. Also, the data set has an identifier (research folder code);

⁵<https://datos.cdmx.gob.mx>

⁶https://archivo.datos.cdmx.gob.mx/fiscalia-general-de-justicia/carpetas-de-investigacion-fgj-de-la-ciudad-de-mexico/carpetas_ss_junio2020.csv

Figure 1. Methodology flowchart.

2. Discovery, data cleansing, filtering, and transformation (dealing with missing values, dealing with outliers, correcting typos, grouping sparse classes, dropping duplicates). In this step, the identified data errors are corrected to create complete and accurate data sets that are ready to be processed and analyzed. For example, faulty data was removed or manually fixed, missing values are filled in, and inconsistent entries are harmonized. 3. Application, analysis using PCoA and K-means, discussions, and conclusions (visually analyze the results of the clustering using the rstudio software ([R Core Team, 2018](#)) and drawing a rough conclusion on how k-means clustering and PCoA performs on dataset).

We analyze the impact of low impact crime in Mexico City based on 25 official categories identified from the Mexico City FGJ investigation reports: abuse of power (AOPow), abuse of trust (AOTRU), assault and battery (AABAT), burglary of home (BOHOM), business burglary (BOBUS), dispossession (DISPO), domestic violence (DOVIO), extortion (EXTOR), forgery of public documents (FOPDO), fraud (FRAUD), identity theft (IDTHE), injuries caused by a crash due to negligent or reckless driving (ICDRV), low impact crime (LOW), malicious property damage (MPDAM), minor drug trafficking (MDTRA), sexual assault (not rape) (SEXAT), negligent or reckless damage of a vehicle while driving (DADRV), robbery of minor belongings (PTHEF), petty theft of property inside a vehicle (THOVE), robbery to passenger of public transport with violence (RPVIO), robbery to passenger or driver of a private vehicle with violence (RVVIO), theft of auto-parts (THAPA), theft of bicycle or similar (THBIC), threats (THREA), and other offenses (OTHER). 25 categories were selected as five categories had very low percentages and were therefore categorized together into other offences (OTHER). We compare and analyze how people report a crime through the traditional system of the Mexico City Police department K-means Clustering via Principal Component analysis and using multivariate statistical analysis (see *extended data*, Table 1 for abbreviations) ([Pambabay-Calero, 2021](#)).

Due to a large number of crime categories, in this study, we will use treemap graphs, which are perfect for displaying large amounts of hierarchical structure data (tree structure). This diagram is ideal for displaying hierarchies since it presents information in a compact and space-saving way.

Analysis using PCoA and K-means

A data set with 115,209 records was collected from the Mexico City police department reports were used for multivariate statistical analysis ([Pambabay-Calero, 2021](#)). The data set consisted of the following variables: year of the crime, month of the crime, day of the crime, crime, crime category, boroughs, prosecutor's office, time of the crime, sex of the victim, and age of the victim. The following variables were excluded from the analysis: agency, investigation unit, the neighborhood of the crime, street of the crime, longitude and latitude, as they did not provide relevant information for the present study.

PCoA provides a means of exploring and visualizing similarities or dissimilarities of data. One advantage of PCoA is that it starts with a similarity matrix or dissimilarity matrix and assigns for each item a location in a low-dimensional space ([Kruskal, 1964a, 1964b](#)) by means of a biplot. For this study, axis 1 (PCoA1) characterizes the municipalities where the highest number of crimes occur, and axis 2 (PCoA2) shows the municipalities where the occurrence of crimes is very low.

The term biplot refers to the simultaneous consideration of both the column space and the row space of the data matrix X . In general terms, the PCoA algorithm is mainly based on a classical approach related to multidimensional scaling (MDS) (Wickelmaier, 2003), where the main purposes are the following.

- To represents the similarities of data as distances in space (coordinates) to make them accessible to visual inspection and exploration.
- To verify if the differences, which distinguish some objects from others, are reflected in the representation obtained.
- To discover the relevant dimensions, present in the similarities.
- To explain the criteria of the similarities in terms of a rule that "emulates" a particular type of distance.

In this regard, we follow the traditional algorithm where matrix X (data matrix) can be derived into $B = XX' = L\Lambda L'$ through singular value decomposition, where Λ represents the diagonal matrix of the eigenvalues of B , and L the normalized eigenvector matrix. The procedure can be summarized in the following steps

- Compute the distance matrix D and D^2 (or also the dissimilarity matrix Δ and Δ^2).
- Apply the double centering on the matrix to obtain the matrix B , i.e.,

$$B = -\frac{1}{2}JD^2J, \quad J = I - n^{-1}(11')$$

- Obtain the eigenvalues of B , according to the spectral decomposition, i.e., $B = XX' = L\Lambda L'$.
- Let r be the dimensionality of the solution. The matrix Δ_1 denotes the diagonal matrix with the first r positive eigenvalues and L_1 the first r columns of L . Then the coordinate matrix, from the scaling, is given by $X = L_1\Lambda_1^{1/2}$

Traditionally, crime data has been assessed by measuring clustering analysis. Thus, we carried out a k-means clustering method with the aim to find clusters of comparable spatial extent. In this context, this manuscript followed the algorithm proposed in Joshi, Sabitha, and Choudhury (2017). The procedure consists of finding the partition with the smallest error, moving individuals from one cluster to another until the reduction of the partition error stabilizes. In short, it is a matter of relocating the individuals so as to obtain clusters with the lowest possible variability. The following is the strategy followed by k-means.

- Choose the first k objects in the sample as the initial k sets of mean vectors.
- Choose the most distant objects.
- Start with a value of k as large as necessary and proceed to form centroids of the groups spaced at a multiplot of standard deviation on each variable.
- Label the objects from 1 to n and choose the ones marked with the following numbers $\frac{n}{k}, \frac{2n}{k}, \dots, \frac{(k-1)n}{k}$ and n .
- Choose k and the initial group configuration based on prior knowledge of the problem.

The following variables were used for this algorithm: days of the week, sex of the victim, category of the crime, and age group. For the selection of the number of clusters, we analyzed the silhouette coefficient (Yuan and Yang, 2019), which shows that, with six clusters, the data are correctly adjusted.

Results

In order to understand how crime patterns occur in Mexico City, a series of measurements were performed. First, a percentage analysis on the principal official crime categories was depicted in Figure 2, where low impact crime (LOW) showed the highest percentage (see *extended data*, Table 1 for abbreviations) (Pambabay-Calero, 2021).

Figure 2. Percentage of crime categories.

Given that the LOW category is the greatest with 76%, it becomes necessary to study this category with a different level of granularity. **Figure 3** shows these sub-categories for LOW as determined by the laws of the country of Mexico, and their percentages (**Table 1**). It should be noted that OTHER - other offenses (unclassified information) has the highest proportion with nearly 20%, followed by DOVIO - domestic violence at 13.26%, BOBUS - business burglary at 8.97%, and THREA - threats at 7.32%.

Figure 3. Sub-categories related to low impact crime category.

Table 1. Description and abbreviation of low impact crimes.

Offense	Description	Code
Abuse of power	A public servant exerts violence upon another person without a legitimate cause	AOPOW
Abuse of trust	A person uses for his/her own profit the possessions of another, who has conceded the tenure, but not the dominion	AOTRU
Assault and battery	A person punches or somehow beats another one	AABAT
Burglary of home	Home burglary without any form of violence against persons	BOHOM
Business burglary	Burglary performed on a shop, store, or some kind of business facility	BOBUS
Dispossession	A person occupies land, or property not owned by him/her, through the use of force, deceit, or stealth	DISPO
Domestic violence	Violence is exerted by someone against a relative. Relatives include adopted children, people under one's care, and other not blood relatives	DOVIO
Extortion	A person blackmails another one to do some- thing, to stay, or to tolerate something, and obtains a profit from this	EXTOR
Forgery of public documents	A person forges public documents such as credit cards or cash	FOPDO
Fraud	A person maliciously obtains profit of the errors of another one	FRAUD
Identity theft	A person uses the identity of another one or concedes the usurpation of one's identity in order to commit a criminal offense	IDTHE
Injuries caused by a crash due to negligent or reckless driving	Due to reckless driving, a person harms another one	ICDRV
Low impact crime	Low impact crime	LOW
Malicious property damage	A person intentionally destroys or damages another one's property	MPDAM
Minor drug trafficking	A person is found in possession of prohibited drugs in such an amount that is presumed that the drugs would be sold or distributed	MDTRA
xxx		
xxx		
Sexual assault (not rape	A person forces upon another an act of sexual nature, or forces another person to watch an act of sexual nature, without the intent of reaching coitus	SEXAT
Negligent or reckless damage of a vehicle while driving	While driving recklessly, a person crashes and destroys or deteriorates someone else's vehicle	DADRV
Petty theft	Robbery of minor belongings	PTHEF
Petty theft of property inside a vehicle	A person robs belongings stored within a vehicle	THOVE
Robbery to passenger of public transport with violence	A person uses violence to rob belongings from persons within a public vehicle	RPVIO
Robbery to passenger or driver of a private vehicle with violence	A person uses violence to rob belongings from persons within a private vehicle	RVVIO
Theft of auto-parts	Robbery of auto-parts	THAPA
Theft of bycycle or similar	Theft of bicycles or mobility devices that rely on human propulsion	THBIC
Threats	A person threatens another one with whom he/she shares a familiar or social bond	THREA
Other offenses	Other offenses	OTHER

Moreover, by using principal co-ordinates analysis (PCoA) in a low dimensional space, we carried out a PCoA on the following variables: crime, time, and location (see extended data, Table 2 for abbreviations) (Pambabay-Calero, 2021). This analysis permits to observe similarities between those variables. Thus, it can be seen in Figure 4 that DOVIO - domestic violence is well clustered with locations as Cuauhtémoc and Iztapalapa in the following time (in hours): 12pm - 15pm, 15pm - 18pm and 18pm - 21pm. The other cluster is formed by the rest of the 24 crime categories, i.e., AOPOW, AOTRU, AABAT, BOHOM, BOBUS, DISPO, EXTOR, FOPDO, FRAUD, IDTHE, ICDRV, LOW, MPDAM, MDTRA, SEXAT, DADRV, PTHEF, THOVE, RPVIO, RVVIO, THAPA, THBIC, THREA and OTHER.

With the aim to gain insights about between the aforementioned crime categories; we display a biplot graph in Figure 5; in this figure it can be noted that the crimes for domestic violence, negligent or reckless damage of a vehicle while

Figure 4. Principal Coordinates Analysis in two dimensions on the following variables: crime, time and location.

Figure 5. Biplot in two dimensions of the variables: crime, time, and location.

driving, and malicious property damage, present the highest level of correlation. Similarly, these categories are highly correlated with the time window of 21pm to 6am, mostly on Sundays. These results were obtained with the databases (see underlying data database_of_victim_complaints.xlsx, [Pambabay-Calero, 2021](#)).

It is important to note that locations with the highest rate of crime are depicted on axis 1 (PCoA1) at these locations; Cuauhtémoc, Iztapalapa, Gustavo Madero, Benito Juárez, Coyoacán, and Miguel Hidalgo.

In addition, [Figure 5](#) on axis 2 (PCoA2) reveals that domestic violence and assault and battery crimes increase on Sundays between 3am to 6am. In the same Figure, about axis 1 (PCoA1), it is noted that the time of highest crime incidence is between 18 and 21 hours in Gustavo Madero and Coyoacán.

In Cuauhtémoc the most frequent crime is petty theft of property inside a vehicle in the period 15pm - 18pm. Meanwhile, the same crime increases in Benito Juárez and Miguel Hidalgo from 6am - 9am. Finally, Iztapalapa shows that the highest occurrence of crime is cases of dispossession between 6am - 9am.

It should be highlighted that LOW behavior might be measured in terms of locations and time as well. The correlation between location and time is showed in a tree map with the aim to provide a comparative analysis based on proportions. [Figure 6](#) compares the results of the correlational analysis based on weekdays and locations. As can be seen from the figure, Cuauhtémoc, Iztapalapa, and Gustavo A Madero, have the highest number of low impact crimes. A closer inspection of the tree map shows that proportionally speaking these crimes take place from Wednesdays to Saturdays.

Similarly, [Figure 7](#) shows a tree map with the time interval for the top three locations: Cuauhtémoc, Iztapalapa, and Gustavo A. Madero. From this data, we can see that the number of low impact crimes committed has the highest proportion from 9am to 15pm.

Figure 6. A comparative analysis based on the proportions of low crime occurrence according to location.

Figure 7. A comparative analysis based on the proportions of low crime occurrence according to the time.

The results of this study indicate that it is possible to find correlations in terms of crime, time, and day across all locations included in this study. Through a k-means clustering analysis, six clusters were proposed to group LOW categories into coherent themes depicted in Figure 8. Themes were defined by the authors after the clustering analysis was performed and a discussion was held about the common characteristic of crime categories in each cluster. In theme one it is possible to find those categories related to 'mobility' from 12pm to 15pm across all regions: injuries caused by a crash due to negligent or reckless driving, robbery to passenger of public transport with violence, robbery to passenger or driver of a private vehicle with violence, theft of auto-parts, theft of bicycle or similar, extortion, and abuse of power. Though extortion and abuse of power are not mobility crimes per se, they could be related to misbehavior by police officers in charge of mobility surveillance. These crimes are mostly associated to men and women between 15 and 24 years old. There is no evidence of a particular day of the week where these crimes occur more frequently.

In theme two, most of the crimes are linked to the 'burglary of home' category. In this context this crime presents the highest number of occurrences around 9am-12pm mostly on Fridays, Saturdays, and Sundays. There is no association between the sex and age of the victim and the crime. Theme three 'theft from stores or parked vehicles' consisted of the crimes of business burglary, identity theft, petty theft of property inside a vehicle, and malicious property damage.

These crimes present the highest number of occurrences mostly on Tuesdays, Wednesdays, and Thursdays between the hours of 6am-9am. There is no association between the sex and age of the victim and the crime.

Regarding to theme four, 'crimes against vulnerable people', the crimes of sexual assault (not rape), threats, fraud, abuse of trust, dispossession, and petty theft are the most frequent crimes. The most frequent victims are in between 0-14 years old and greater than 65 years old. There is no evidence of a particular day of the week where these crimes occur more frequently.

Figure 8. K-means cluster analysis.

With respect to theme five, ‘weekend crimes’, the most frequent crimes are as follows: negligent or reckless damage of a vehicle while driving, forgery of public documents, and assault and battery; these crimes have the greatest occurrence on Mondays from 3am-6am. There is no association between the sex and age of the victim and the crime

Finally, theme six provides those crimes related to ‘minor drug trafficking, and domestic violence’. In this case, most of the victims are in between 25 and 64 years old and the highest rates of these crimes happen between the hours of midnight to 3am. There is no association between the sex of the victim and the day of the week with the crime.

Discussion

The present results are another test of the routine activity theory of crime. In a city as large as Mexico City, stereotyped situations that facilitate minor crime are produced in a routine fashion related to well scheduled mass activities such as work or leisure. There is a clear difference in crime occurrence between the central city district Cuauhtémoc —which concentrates offices, malls, restaurants, and touristic locations (Gasca-Zamora, 2017; Salas-Benítez and López-López, 2019) — and a peripheral district, Iztapalapa, which houses about a fifth of the city dwellers, in conditions of social vulnerability (Rodríguez-Cortés, 2017). As the weekend sets in, a sort of wave of crime moves from the peripheral areas towards the central areas of the city.

These findings support the literature that relies on routine activities theory to explain why crime concentrates in particular spaces at particular times (e.g., Campedelli, Aziani, and Favarin (2021); Schnell, Grossman, and Braga, 2019; Vilalta *et al.*, 2018). Explanations are reduced to the interaction of three main factors, the presence of victims, the presence of offenders, and the absence of authority. Yet, the spatial location of victims and offenders is not static and so, the matter of mobility of victims and perpetrators is a field of study that has not been thoroughly explored.

In relation to mobility, the characterization developed in this study, may mean one of two things: either offenders follow the displacements of victims, or different offenders target victims as they develop different activities within the city. Also, the activities of victims follow a general work-leisure cycle. Song *et al.*, (2019) state that though the mobility of law-abiding citizens determines the characteristics of crime location, offenders seem to act not far from their neighborhoods, thus supporting the hypothesis of victims falling to offenders living in different parts of the city. However, further studies are needed to assess the matter.

In terms of public policy, our research envisions the possibility to use recent data to organize crime in situational categories which could help improve police efficiency.

The divulgation of these situational categories could also help non-governmental organizations (NGOs) that help victims of crime or that stand for a better access to justice, to push for changes in prevention policies and to inform citizens about situational risks and possible preventive measures.

A deep sociological analysis of the cost-benefit of placing a complaint is required to disentangle the true characteristics of crime from those characteristics related to the operation of prosecuting offices in Mexico; for example: schedules of prosecution offices, location of prosecution offices, documentation needed to file a complaint, availability of prosecution clerks, and other considerations that may tamper the quality of the data available for analysis. In this study, sociological considerations of routines related to crime reporting are not considered, and that is a serious limitation of this and other similar studies.

The National Victims Survey in Mexico (ENVIPE) states that an investigation folder (such as the ones used in this research to get data) filed in a prosecution office has only in 7.9% of the actual offenses committed (INEGI, 2020). Thus, data gathered from prosecution offices must be combined with data from surveys in order to obtain a more accurate panorama of crime, particularly in highly underreported offenses, such as domestic violence and extortion (INEGI, 2020; Saavedra-Hernández, 2009).

This situational categorization, however, is incomplete without data about police behavior or observations about security measures taken by the state in specific locations to prevent minor crime. Such observations would cover the “absence of a capable guardian” part stated by the routine activities theory. Some reflections on the matter have been made with regards to the effectiveness of the use of video cameras against non-violent crime (Vilalta *et al.*, 2018). Yet, considering the amount of crime observed and its regularity, the absence of capable guardians may be the rule and not the exception in the city.

Conclusions

The aim of the present research was to examine how crime reports can be used as an acceptable proxy measure of antisocial behavior, this data was captured in the database (database_of_victim_complaints.xlsx) and graphs 4 through 8 were generated from the file (victim_data_figures_4_8.xlsx) (Pambabay-Calero, 2021). It should be noted that by using this sort of information it is possible to find unsafe areas in Mexico City; and to estimate crime occurrence peaks as well.

This study has identified those city boroughs that show the highest number of reported crimes in terms of low impact crime. In this context, our preliminary results can be helpful to develop tailored strategies to tackle crime.

The findings of this study suggest that Cuauhtémoc and Iztapalapa show high rate of criminal acts due to their demographic factors. This observation may support the hypothesis as Iztapalapa has the highest population density and social relegation. On the other hand, despite the fact that Cuauhtémoc presents a similar population density, its levels in growth and change in terms of human development are more than acceptable in comparison with Iztapalapa (Ángeles-Castro, 2013; Caudillo and Flores, 2016).

Data availability

Underlying data

Open Science Framework (OSF). charactLowImpact. <https://doi.org/10.17605/OSF.IO/PT3GV>. (Pambabay-Calero, 2021).

This project contains the following underlying data:

- Data file 1. database_of_victim_complaints.xlsx, (data on complaints collected from source including dates, age, gender, type of crime, crime description and location)
- Data file 2. victim_data_figures_4_8.xlsx, (data to generate graphs in figures 4, 5, 6, 7, 8)

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

Extended data

Open Science Framework (OSF): charactLowImpact. <https://doi.org/10.17605/OSF.IO/PT3GV>. (Pambabay-Calero, 2021).

This project contains the following extended data:

- Table 1. (Description and abbreviation of low impact crimes).

- Table 2. (Abbreviations of municipalities).
- Figure 1. (Methodology flowchart).
- Figure 2. (Percentage of crime categories).
- Figure 3. (Sub-categories related to low impact crime category).
- Figure 4. (Principal Coordinates Analysis in two dimensions on the following variables: crime, time and location).
- Figure 5. (Biplot in two dimensions of the variables: crime, time, and location). • Figure 6. A comparative analysis based on the proportions of low crime occurrence according to location.
- Figure 7. A comparative analysis based on the proportions of low crime occurrence according to the time.
- Figure 8. K-means cluster analysis.

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

Acknowledgments

We are greatly indebted to anonymous referees for many helpful comments.

References

- Aghababaei S, Makrehchi M: **Mining social media content for crime prediction**. 2016 *IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. 2016; (pp. 526–531). [Publisher Full Text](#)
- Ángeles-Castro G: **Crecimiento económico y desarrollo humano en la ciudad de México con respecto a un entorno nacional: una perspectiva neoclásica y dualista**. *Economía Sociedad y Territorio*. 2013; **13**(42): 431–457. [Publisher Full Text](#)
- Arango A: **Indicadores de Seguridad Pública en México: La Construcción de un Sistema de Estadísticas Delictivas**. 2003. [Reference Source](#) | [Publisher Full Text](#)
- Bandrés E, Díez-Ticio A: **Delincuencia y acción policial. un enfoque económico**. *Revista de Economía Aplicada*. 2001; **9**(27): 5–33.
- Bergman M, Sarsfield R: **Encuesta de Victimización y Eficacia Institucional (ENVEI) - Informe Anual 2008**. Mexico City: Centro de Investigación y Docencia Económicas, Programa de Estudios para la Seguridad Pública y Estado de Derecho; 2009. [Reference Source](#)
- Caudillo C, Flores S: **Tendencias espacio-temporales en la segregación. Tendencias territoriales determinantes del futuro de la Ciudad de México**. 2016.
- Campedelli GM, Aziani A, Favarin S: **Exploring the Immediate Effects of COVID-19 Containment Policies on Crime: an Empirical Analysis of the Short-Term Aftermath in Los Angeles**. *American Journal of Criminal Justice*. 2021; **46**: 704–727. [PubMed Abstract](#) | [Publisher Full Text](#)
- Chadee D, Ng Ying NK, Chadee M, et al.: **Fear of crime: the influence of general fear, risk, and time perspective**. *Journal of Interpersonal Violence*. 2019; **34**(6): 1224–1246. [PubMed Abstract](#) | [Publisher Full Text](#)
- Chen X, Cho Y, Jang SY: **Crime prediction using Twitter sentiment and weather**. 2015 *systems and information engineering design symposium*. 2015; (pp. 63–68). [PubMed Abstract](#) | [Publisher Full Text](#)
- Cisneros J, Cunjama-López E: **El color del miedo bajo el desorden del paisaje urbano en la Ciudad de México y la zona metropolitana**. *Criminalidad*. 2011; **53**(1): 275–292. [Reference Source](#)
- Clarke RV: **Situational crime prevention**. *Environmental Criminology and Crime Analysis: Second Edition*. 2016, Jan; (Vol. **19**: pp. 286–303). [Publisher Full Text](#)
- D'ancona C, Ángeles M: **Metodología cuantitativa: estrategias y técnicas de investigación social**. *Síntesis*. 1999.
- D'Avanzo E, Pilato G, Lytras M: **Using Twitter sentiment and emotions analysis of google trends for decisions making**. *Program*. 2017; **51**(3): 322–350. [Publisher Full Text](#)
- Felson M, Clarke RV: **Opportunity makes the thief**. *Police Research Series, Paper 98*. 1998. 1–36.
- Flores RD: **Do anti-immigrant laws shape public sentiment? a study of Arizona's sb 1070 using twitter data**. *American Journal of Sociology*. 2017; **123**(2): 333–384. [Publisher Full Text](#)
- Fuentes-Flores C: **El impacto de las viviendas deshabitadas en el incremento de delitos (robo a casa habitación y homicidios) en Ciudad Juárez, Chihuahua, 2010**. *Frontera Norte*. 2015; **27**(54): 4–196. [Publisher Full Text](#)
- Fuentes-Flores C, Sánchez-Salinas O: **Contexto sociodemográfico de los homicidios en México DF: un análisis espacial**. *Revista Panamericana de Salud Pública*. 2015; **38**: 450–456. [PubMed Abstract](#) | [Reference Source](#)
- Gabriel KR, Odoroff CL: **Biplot in biomedical research**. *Statistics in Medicine*. 1990; **9**(5): 469–485. [PubMed Abstract](#) | [Publisher Full Text](#)
- Gamma A, Schleifer R, Weinmann W, et al.: **Could Google Trends be used to predict methamphetamine-related crime? an analysis of search volume data in Switzerland, Germany, and Austria**. *PLoS One*. 2016; **11**(11): e0166566. [PubMed Abstract](#) | [Publisher Full Text](#)
- Gasca-Zamora J: **Centros comerciales de la ciudad de México: el ascenso de los negocios inmobiliarios orientados al consumo**. *EURE*. 2017; **43**(130): 73–96. [Publisher Full Text](#)

Gower JC: **Some distance properties of latent root and vector methods used in multivariate analysis.** *Biometrika*. 1966; **53**(3-4): 325–338.

[Publisher Full Text](#)

Jasso López C: **Percepción de inseguridad en México.** *Revista Mexicana de Opinión Pública*. 2013; **15**: 12–29.

[Publisher Full Text](#)

INEGI: **Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE). Principales resultados.** 2020. August 16, 2021.

[Publisher Full Text](#) | [Reference Source](#)

Joshi A, Sabitha AS, Choudhury T: **Crime analysis using k-means clustering.** *2017 3rd International Conference on Computational Intelligence and Networks (CINE)*. 2017; **6**: (pp. 33–39).

[Publisher Full Text](#) | [PubMed Abstract](#)

Kruskal JB: **Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis.** *Psychometrika*. 1964a; **29**(1): 1–27.

[Publisher Full Text](#)

Kruskal JB: **Nonmetric multidimensional scaling: a numerical method.** *Psychometrika*. 1964b; **29**(2): 115–129.

[Publisher Full Text](#)

Kubrin CE, Weitzer R: **New directions in social disorganization theory.** 2003, Nov; (Vol. **40**)(No. 4):

[Publisher Full Text](#)

Malleson N, Andresen MA: **Spatio-temporal crime hotspots and the ambient population.** *Crime Science*. 2015; **4**(1): 10.

[Publisher Full Text](#)

Pambabay-Calero JJ, Bauz-Olvera SA, Nieto-Librero AB, et al.: **An alternative to the cochrane-q) statistic for analysis of heterogeneity in meta-analysis of diagnostic tests based on biplot.** *Investigación Operacional*. 2018; **39**(4): 536–545.

[Reference Source](#)

Piña-García C, Ramírez-Ramírez L: **Exploring crime patterns in Mexico City.** *Journal of Big Data*. 2019; **6**(1): 65.

[Publisher Full Text](#)

Procuraduría G d J: **Carpetas de investigación PGJ de la Ciudad de México - Datos CDMX. 2019-08-16.** 2018.

[Reference Source](#)

R Core Team: **R: A language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing; 2018.

[Publisher Full Text](#) | [Reference Source](#)

Rodríguez-Cortés L: **Building citizenship. The struggle for housing in eastern Mexico City.** *Latin American Perspectives*. 2017; **44**(44): 176–190.

[Publisher Full Text](#)

Saavedra-Hernández LE: **Acceso a la justicia de las mujeres que viven violencia conyugal. El caso de Emilia.** [Conference presentation] XXVII Congreso de la Asociación Latinoamericana de Sociología. VIII Jornadas de Sociología de la Universidad de Buenos Aires. Asociación Latinoamericana de Sociología. Buenos Aires, Argentina. 2009.

[Reference Source](#)

Salas-Benítez CM, López-López A: **Efectos espaciales de la tematización cultural para la recreación y el turismo en los corredores culturales peatonales del Centro Histórico de la Ciudad de México.** *Investigaciones*

Geográficas. 2019; **98**: 1–17.

[Publisher Full Text](#)

Sampson RJ, Groves WB: **Community Structure and Crime: Testing Social-Disorganization Theory.** *American Journal of Sociology*. 1989, Jan; **94**(4): 774–802.

[Publisher Full Text](#)

Sánchez Salinas OA, Fuentes Flores CM: **El robo de vehículos y su relación espacial con el contexto sociodemográfico en tres delegaciones centrales de la Ciudad de México (2010).** *Investigaciones Geográficas*. 2016; **89**(89): 107–120.

[Publisher Full Text](#)

Schnell C, Grossman L, Braga A: **The routine activities of violent crime places: A retrospective case-control study of crime opportunities on street segments.** *Journal of Criminal Justice*. 2019; **60**: 140–153.

[Publisher Full Text](#)

Song G, Bernasco W, Liu L, et al.: **Crime Feeds on Legal Activities: Daily Mobility Flows Help to Explain Thieves' Target Location Choices.** *Journal of Quantitative Criminology*. 2019; **35**: 831–854.

[Publisher Full Text](#)

Vilalta-Perdomo C: **Los factores de la violencia durante el delito en México: armas de fuego, momento del día y experiencia criminal.** *Estudios Sociológicos*. 2009; **27**(79): 211–235.

Vilalta-Perdomo C: **El miedo al crimen en México: Estructura lógica, bases empíricas y recomendaciones iniciales de política pública.** *Gestión y Política Pública*. 2010; **19**(1): 3–36.

[Reference Source](#)

Vilalta C: **El robo de vehículos en la ciudad de México.** *Gestión y Política Pública*. 2011; **20**(1): 97–139.

[Reference Source](#)

Vilalta CJ, Sánchez TW, Fondevila G, Ramírez M: **A descriptive model of the relationship between police CCTV systems and crime. Evidence from Mexico City.** *Police Practice and Research*. 2018; **20**(2): 105–121.

[Publisher Full Text](#)

Wang M, Gerber MS: **Using twitter for next-place prediction, with an application to crime prediction.** *2015 IEEE Symposium Series on Computational Intelligence*. 2015; **107**: (pp. 941–958).

[Publisher Full Text](#)

Wickelmaier F: **An introduction to MDS.** *Sound Quality Research Unit*. Denmark: Aalborg University; 2003; **46** (5): 1–26.

Yadav N, Kumar A, Bhatnagar R, et al.: **City crime mapping using machine learning techniques.** *International Conference on Advanced Machine Learning Technologies and Applications*. 2019; (pp. 656–668).

Yang D, Heaney T, Tonon A, et al.: **Crimetelescope: crime hotspot prediction based on urban and social media data fusion.** *World Wide Web*. 2018; **21**(5): 1323–1347.

[Publisher Full Text](#)

Yuan C, Yang H: **Research on K-Value Selection Method of K-Means Clustering Algorithm.** *J*. 2019; **2**(2): 226–235.

[Publisher Full Text](#)

Pambabay-Calero J: **charactLowImpact.** *Open Science Framework*. 2021, September 14.

[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 22 February 2022

<https://doi.org/10.5256/f1000research.76607.r118439>

© 2022 Vicente Villardón J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

José Luis Vicente Villardón

Department of Statistics, University of Salamanca, Salamanca, Spain

The biplot references are not adequate. Gabriel and Odoroff (1990) present the biplot in medical studies, but does not use Principal Coordinates. Gower (1966) uses principal coordinates, but not biplots. Both concepts together are presented by Gower in different papers, but a summary of all can be found in Gower & Hand (1995). This would be a much better citation for the joint use of principal coordinates and biplots.

The so called "category level points" are not defined in Pambabay-Calero et al. (2018). They use a biplot based on principal components for frequency data. Category level points refer to a different concept that can be found also in Gower and Hand (1995), and is related to a modified version of Multiple Correspondence analysis.

The references of Kruskal(1964a, 1964b) do not describe PCoA (principal co-ordinates analysis) nor its relation to a biplot, rather describe Multidimensional Scaling that, although is a very close technique, is not the same as PCoA. Again, the right reference is Gower and Hand (1995).

The initial matrix for the analysis is not clearly described in the paper. What does the matrix have in its rows and columns? The locations? The times? The crimes? Is it a frequencies matrix? Is it an indicator matrix?

From the graphs it is not possible to infer the shape of the initial data introduced in the algorithm. The exact form of the input data matrix should be described clearly, (what are the rows and the columns?).

The description of the principal coordinates algorithm is confusing. First the inner product B is obtained from the data matrix X and later from the dissimilarity matrix D. In any case, the procedure to calculate dissimilarities from the raw data should be specified.

No clear method for the calculation of the biplot is described. The relation among principal coordinates and biplots is not clearly described, so, there is not enough information to replicate the calculations and the results. If finally, the used method is a classical biplot, the references to PCoA are not justified. Classical biplots are related to PCoA of the Euclidean distance matrix, but

that does not justify including that technique in the description.

The exact R packages used for the calculations should be reported. If the calculations were made using routines of the authors should also be reported. A package to construct biplots is MultBiplotR. With this package the authors can avoid all the problems described later and improve the representations.

Figures 4 and 5 are not adequate. PCoA analysis is interpreted using Euclidean (Pythagorean) distances. For that interpretation to be possible, one unit in the scale of the X axis must be exactly the same as one unit in the Y axis. That is not the case in both figures.

A biplot is a joint representation of the rows and columns of a data matrix, so figures 4 and 5 should be combined in a single plot or represented with the same scale. It is quite obvious that the scales of both plots are quite different, so the whole procedure should be redone.

The quality of the plot also needs to be improved, both figures are very low quality and figure 5 is even not complete.

With all the deficiencies in the representation of biplots, I cannot judge the accuracy of the results. If I understand correctly and the input matrix contains frequencies and a classical biplot has been used, the accuracy of the analysis is not clear.

As a summary, the description and application of the statistical methods must be significantly improved for the paper to be acceptable. Obviously, the results must be modified consequently if necessary.

References

1. Gower J.C, Hand D.J: Biplots. *Chapman and Hall/CRC*. 1995.
2. Vicente-Villardón J. L: MultBiplotR: Multivariate Analysis Using Biplots in R. R package version 2.1.3. 2022. [Reference Source](#)

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

No

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: My area of research is multivariate statistics, especially biplots.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 24 January 2022

<https://doi.org/10.5256/f1000research.76607.r118444>

© 2022 Huerta-Pacheco N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

N. Sofía Huerta-Pacheco

Cátedras CONACYT—Ciencia Forense, Facultad de Medicina, Universidad Nacional Autónoma de México, Mexico City, Mexico

This article describes the behavior and distribution of the crimes presented during a period in Mexico City, which showed findings that allow the identification of the areas with the highest incidence of crime, as well as the temporality in which they happen. Therefore, it presents and identifies multivariate patterns of crimes in Mexico as mentioned in the title.

- I suggest that your manuscript should undergo extensive English revisions.
- The article does not present a common thread that guides the reader.
- It is necessary to complement the article with updated information related to reports made, for example, the Observatory of Mexico City and Data Civica reports.
- Some figures and tables have errors.
- If the abbreviations were already explained in Table 1, it is not required describing them in the text.
- I recommend omitting the days-of-the-week variable in Figure 4 since it presents an independent behavior compared to the other variables and affects the patterns of association.
- I advise using the Moran index to see the homogeneity of the evaluated sites.
- In the conclusions, it is not necessary to emphasize where the information is.

Reviewer Questions

1. **Is the work clearly and accurately presented and does it cite the current literature?**
The work lacks up-to-date information reported by official sources.

2. **Is the study design appropriate and does the work have academic merit?**
The design is adequate; however, when presenting the descriptive results of the proportionality of the crimes, it loses the focus in trying to describe them all. Besides, when

the information is concluded, it becomes insufficient.

3. Are sufficient details of methods and analysis provided to allow replication by others?

The methods present sufficient description that can confuse the reader, hence I recommend making a description to make it more succinct.

4. If applicable, is the statistical analysis and its interpretation appropriate?

The proposed methods are correct; however, the level of variability of representation in two factors are not described as well as the level of contribution of the selected variables.

5. Are all the source data underlying the results available to ensure full reproducibility?

The original data is shared. However, the database is missing a description of the variables as well as the filtered base of the variables, since of the 115,209 data, at least 4,280 have no geographic information to classify them within some delegation, such as the description of the process in Figure 1.

6. Are the conclusions drawn adequately supported by the results?

I believe that there is information that should not be in the conclusions, and there is a lack of further explanation of the most outstanding findings, for example, the patterns of crime schedules.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Computational Statistics, Multivariate Statistics, Statistics Education, Forensic Science, and Affective Computing

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 18 January 2022

<https://doi.org/10.5256/f1000research.76607.r118447>

© 2022 Tejedor Flores N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nathalia Tejedor Flores

Centro de Investigaciones Hidráulicas e Hidrotécnica, Universidad Tecnológica de Panamá,
Panama City, Panama

The manuscript: "Multivariate analysis and characterization of low impact crime in Mexico City". The main objective of this article is to determine the extent of crime occurrences and identify statistical features that permit us to establish variables that can provide the basis for the quantitative characterization of crime reports.

The reasons why I gave this article the status of "approved", are as follows:

- The writing style is concise with no repetition or padding, and the manuscript kept the interest throughout the sections and is well structured.
- **Does the title accurately summarise the content?**
 - The title "Multivariate analysis and characterization of low impact crime in Mexico City", summarises the content accurately.
- **New, original ideas or research findings?**
 - This article presents two multivariate methods that uses crime in Mexico City data to know the basis for the quantitative characterization of crime reports, which is interesting to know.
- **Does it integrate tested ideas into a practical theory or model?**
 - Yes, the manuscript integrates tested ideas into a practical theory or model.
- **Are related or competing explanations acknowledged?**
 - Yes, the manuscript has related explanations acknowledged.
- **Is the argument readily apparent and clear to the reader?**
 - Yes, the argument is readily apparent and clear to the reader, the authors present a clear explanation of the methodology and results.
- **Are all relevant sources (stated or unstated) acknowledged?**
 - Yes, all relevant sources are well-acknowledged.

The only correction or clarification to be made by the authors is as follows:

- This article presents multivariate methods (principal co-ordinates analysis (PCoA) and Biplot), but on page 5 in the section of methods the author mention Principal components analysis (PCA), is that correct?

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Multivariate analysis

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 19 Jan 2022

Johny Pambabay, ESPOL, Polytechnic University, Guayaquil, Ecuador

Dear Reviewer,

We would like to thank the reviewer for careful and thorough reading of this manuscript and for the thoughtful comments and constructive suggestions, which help to improve the quality of this manuscript. Please see below, for a point-by-point response to the reviewers' comments and concerns.

Reviewers' Comments to the Authors:

The only correction or clarification to be made by the authors is as follows:

- This article presents multivariate methods (principal co-ordinates analysis (PCoA) and Biplot), but on page 5 in the section of methods the author mention Principal components analysis (PCA), is that correct?

Author response:

Thank you for pointing this out. The reviewer is correct. The correct term is principal co-ordinate analysis. This will be corrected in the article.

Thank you for your attention.

Best regards,

Johny Pambaby

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research