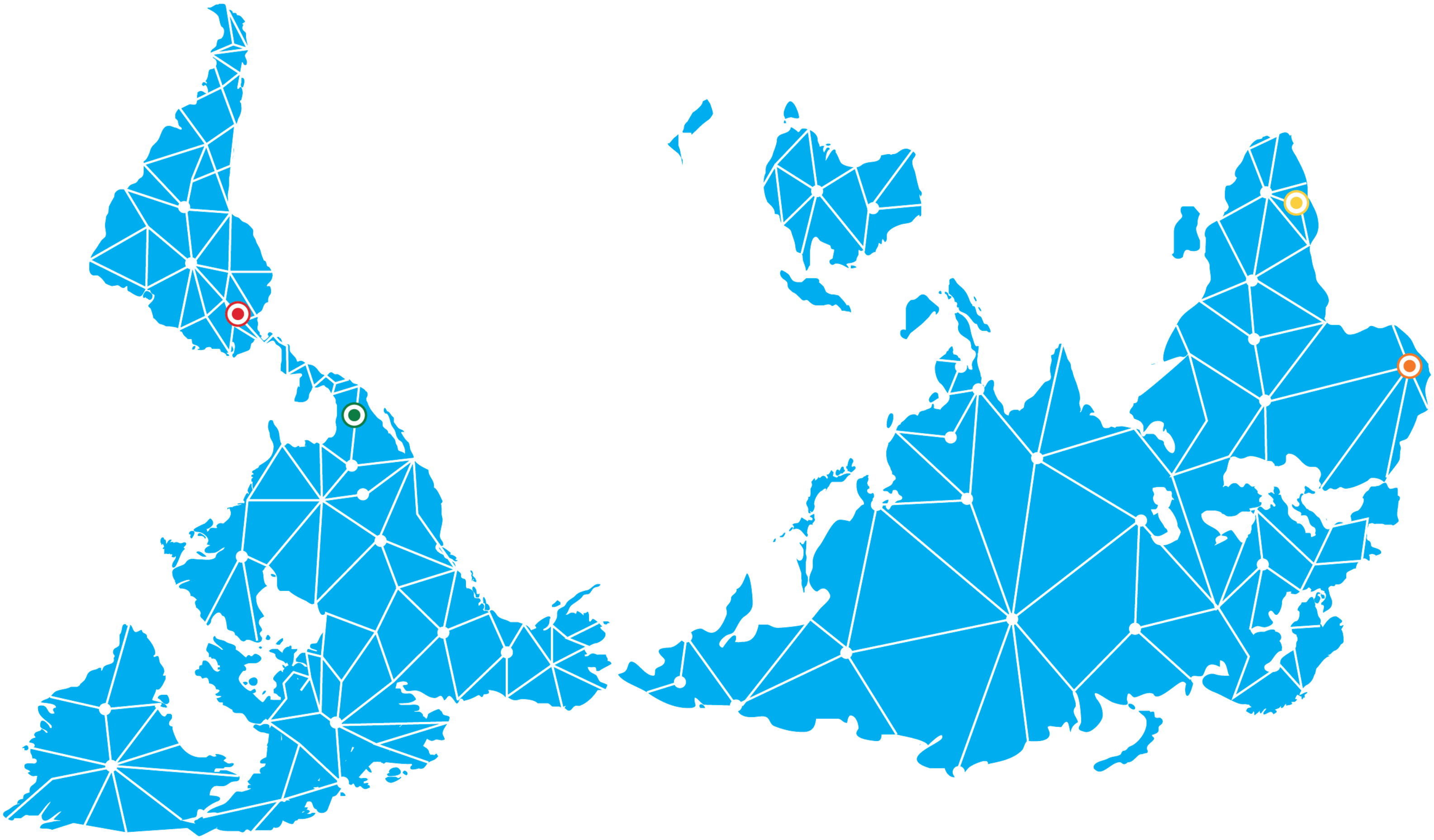
BIG DATA

TO ADDRESS GLOBAL DEVELOPMENT CHALLENGES

2018

DATA-POP ALLIANCE



1. Socio-physical Vulnerability to Flooding in Senegal.
2. Characterizing and analyzing urban dynamics in Bogota.
3. Understanding the Relationship between Short and Long Term Mobility.
4. Large-Scale Mapping of Citizens' Behavioral Disruption in the Face of Urban Crime Shocks.

DATA-POP ALLIANCE

Partners:

Cloud to Street

FLOWMINDER.ORG

mit media lab

HARVARD HUMANITARIAN INITIATIVE

Funded by:

AFD AGENCE FRANÇAISE DE DÉVELOPPEMENT

BIG DATA
TO ADDRESS GLOBAL DEVELOPMENT CHALLENGES
2018

DATA-POP ALLIANCE

THE IMPACT OF CRIME SHOCKS ACROSS GENDER AND SOCIOECONOMIC GROUPS: A LARGE-SCALE MAPPING OF BEHAVIORAL DISRUPTION

Rodrigo Lara Molina* - Data-Pop Alliance, Media Lab-HHI-ODI.
Alejandro Noriega* - MIT Media Lab.
Eaman Jahani* - Institute for Data, Systems, and Society; MIT.
Julie Ricard - Data-Pop Alliance, Media Lab-HHI-ODI.
Alex Pentland - MIT Media Lab, Data-Pop Alliance, Media Lab-HHI-ODI, Institute for Data, Systems, and Society; MIT.
*Authors contributed equally to this work.

DATA-POP ALLIANCE

mit media lab

Funded by:
AFD
AGENCE FRANÇAISE DE DÉVELOPPEMENT

THE IMPACT OF CRIME SHOCKS ACROSS GENDER AND SOCIOECONOMIC GROUPS: A LARGE-SCALE MAPPING OF BEHAVIORAL DISRUPTION

Rodrigo Lara Molina*
Data-Pop Alliance, Media Lab-HHI-ODI.

Alejandro Noriega*
MIT Media Lab.

Eaman Jahani*
Institute for Data, Systems, and Society; MIT.

Julie Ricard
Data-Pop Alliance, Media Lab-HHI-ODI.

Alex Pentland
MIT Media Lab, Data-Pop Alliance, Media Lab-HHI-ODI, Institute for Data, Systems, and Society; MIT.

*Authors contributed equally to this work.

In recent decades the world has seen a simultaneous trend towards becoming more peaceful overall, but also towards higher homicide rates surging in focal regions in the developing world. Although abundant research exists on the nature and sociology of crime, few studies look into the damaging impact of crime and violence on the daily lives of affected communities. The present study proposes the use of societal-scale behavioral data—card transactions' metadata—to elicit such impact. On the crime side, we use detailed homicide records for an entire middle-income country to identify salient crime shocks at the local level. On the behavioral side, we use debit card transaction volumes throughout the country to extract behavioral indices. We show that crime shocks have a substantial effect on communities' consumption patterns. Moreover, we show that the effects of crime shocks distribute differently across population subgroups defined by gender and socioeconomic status— e.g., with reductions of up to 7% in females' average volume of transactions—potentially exacerbating social inequalities. We conclude this work with policy recommendations on the use of 'big data' sources to monitor and help.

TABLE OF CONTENTS

1. Introduction

1.1 Crime and Violence

In recent decades the world has seen a trend towards becoming more peaceful overall (*1*). Simultaneously, higher civilian murder rates are still surging in focal regions in the developing world (while they are decreasing in rich countries): in 2016, 68 percent of violent deaths worldwide were murders, which is more than 3 times the number of deaths caused by war (*2*). Homicide rates are often used as a proxy for the overall regional levels of violence, among other reasons because they are the most systematically recorded crime.

Increasing violence in developing countries can be attributed to a multiplicity of factors. In some cases, it is strongly associated to factors such as organized crime, trafficking, and gang activity (*3*). Moreover, crime is explained to varying degrees by a broad range of structural and societal circumstances, such as demographic and socioeconomic factors (*4*, *5*), access to education (*6*) and weakness of institutions (*5*, *7*). In addition, more recent place-centric approaches have emphasized the impact of physical characteristics of cities on crime, such as the proportion of uninhabited homes (*8*).

Excluding war zones, Latin America and the Caribbean (LAC), is considered the most violent region in the world, where homicide rates registered in 2015 were four times higher than the world average (*3*). In fact, levels of violence in LAC are so disproportionately high, that they are considered at epidemic levels by the World Health Organization. Mexico ranks among the most violent countries in the region, where homicide rates have surged to unprecedented levels since 2007 (*9*), when president Felipe Calderon began a frontal assault on drug-trafficking organizations (*10*). Recently, in 2017, violence levels broke historic records, registering more than 29,000 homicides (INEGI, 2017).

The present work focuses on the effects that prevalent violence may have on the daily lives of local communities. Although the majority of violence is credited to conflict among criminal organizations, or among them and government forces, a recent study published by Open Society Foundations reports numerous atrocity crimes perpetrated against the civilian population since 2006, including killings, torture, and disappearances (*11*). Consequently, violence and insecurity have become the number one concern of Mexicans, above unemployment, corruption, and poverty—with 70% of Mexicans declaring it their top concern, according to the 2015 National Survey on the Quality and Impact of Government (ENCIG 2015).

1.2 Impact of Crime on Citizens' Lives

Whether originated by war or waves of crime, violence has pervasive impacts on communities and the development of cities and countries. Recent studies in Mexico have shown the impact of violence at a societal level, such as deteriorating effect on democratic institutions (*12*), economic costs (*9*, *13–15*) and urban transformation (*16*).

Moreover, violence creates fear and uncertainty, affecting not only the primary victim but also his/her family and community, also known as secondary victims (*17*). Only a few studies worldwide have explored the effects of crime and violence on daily activities of both primary and secondary victims, even the less focusing on LAC countries. A few have revealed the impact of victimization or fear of victimization on risky routine activities (*18*), on nighttime activities (*19*) and on the emotional and psychological health of individuals (*20*).

In Mexico, the National Victimization and Perception of Public Security Survey (ENVIPE), attempts to measure behavioral changes in correlation with perception of security. According to ENVIPE, in 2017 the three daily activities most affected by fear of crime are 'Letting children go out alone', 'Use jewelry' and 'Go out at night'. Further studies based on ENVIPE results, underlined gender differences in both perceptions of security and changes in behaviors in a high-crime environment (*21*, *22*).

However, to our knowledge, the disruption of daily activities such as mobility or consumption, and its differentiated effects on population subgroups, as defined by gender and socioeconomic status, has never been quantified or systematically evaluated.

1.3 Private Sector's Big Data for Public Good

Big data are defined as the digital traces (or digital breadcrumbs) passively emitted through the use of digital devices, such as call records, credit card transactions, GPS locations, etc. (*23*). Given its ubiquitous nature, crumbs can yield accurate information about human behavior (e.g., purchasing behavior or physical mobility and communications) at unprecedented levels of spatiotemporal granularity. Derived insights have been used to diagnose a diverse range of situations, opening to applications such as poverty mapping (*24*), infrastructure (*25*) and transportation planning (*26*). In the field of citizen security and crime prevention, researchers were able to predict crime ”hotspots” in cities such as London, Boston, and Bogota using aggregated human behavioral data captured from the mobile network infrastructure, together with basic demographic information (*27*). To date, these approaches have not looked into the impact of crime on the daily behavior of individuals.

This paper. In this paper, we map individuals' behavioral disruption in the face of urban crime shocks, by using societal-scale debit card transaction records. Section 2 describes the crime and behavioral datasets and how indices were computed. Section 3 summarizes the main results and Section 4 discusses results and future work.

2. Data and Methods

The data and methods used in this paper consist of two types: on the one hand crime data and the elicitation of crime shocks, and on the other hand consumption data and the construction of behavioral indices.

2.1 Crime Shocks

Crime data. We collected detailed crime records from the National Public Safety System of Mexico. More specifically, we computed monthly homicide rates—homicides for each 10,000 individuals—for each of the +2.4 thousand municipalities in the country. The rationale for focusing on homicides is threefold: their high prevalence in the country of focus, their high impact on citizen's perception of safety, and the fact that homicide statistics are among the most reliable among several crime types (e.g., as opposed to rape, which is often under-reported).

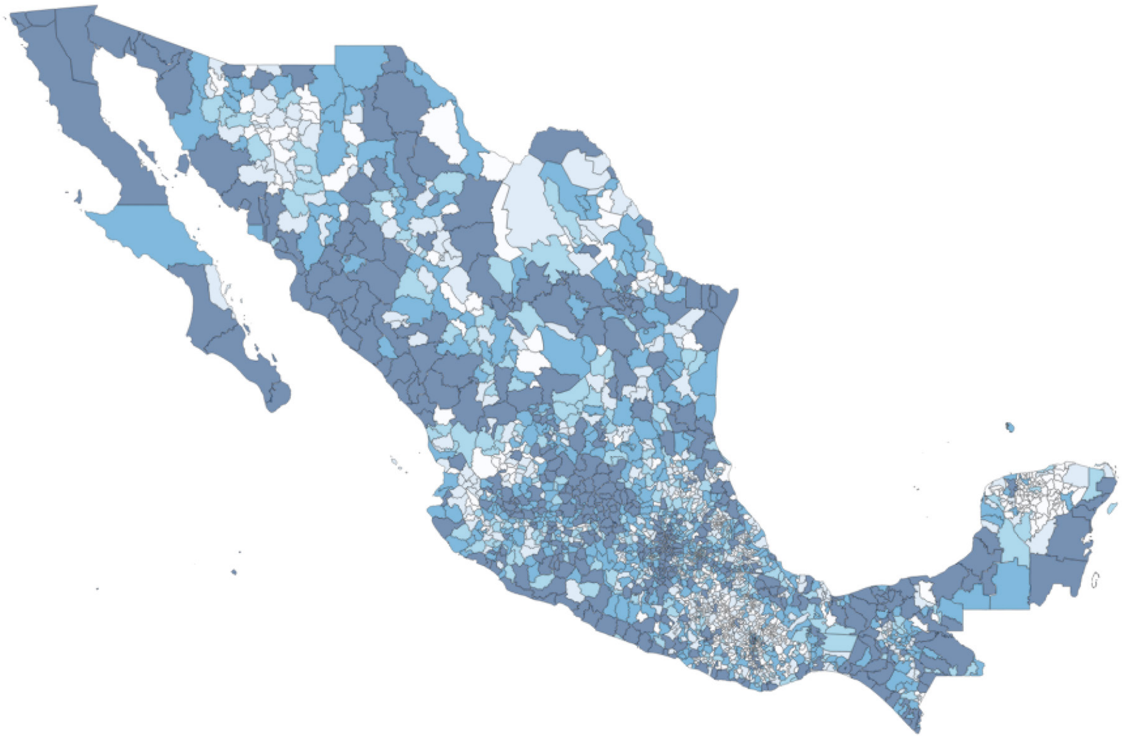


Figure 1: **Spatial partitions and homicide rates**. The map shows the 2.4k municipalities, color-coded by their respective murder rate during the entire period studied. Shades of blue correspond to quartiles of homicide rates, where darker blue indicates higher rates.

Crime shocks. We define a crime shock as a period of sustained and relatively low homicide rates, followed by a period of sustained and relatively high crime rates. In particular—and given the available temporal longitudinality at the intersection of the crime and behavioral datasets (one year from October 2014 to September 2015)—we use six-month windows, such that a shock consists of six months of relative peace, followed by six months of high crime rates. Figure 2 shows such shocks elicited for the municipalities of Mugica, Jalacingo, and Venustiano Carranza. More generally, Figure 3 shows the distribution of percentage change in homicide rate for all municipalities. In what follows, we consider that a municipality endures a crime shock when its six-month homicide rate increases by 75% or more, which corresponds to the .85 percentile (as shown in Figure 3).

As mentioned above, the current ubiquitous use of digital services and infrastructure generates "digital footprints" that can provide statistical information about societal dynamics. For example, mobile phone metadata can provide an approximation of individuals' mobility traces, and credit and debit card transaction records can provide an approximation of consumption and the activities that citizens engage with throughout the day.

2.2 Behavioral Indices.

Consumption data[1]. Here we focus on anonymized and aggregated card transaction records, from one of the leading banks operating in the Mexican market. From these we compute municipality- and monthly-level indices across the period of study (from October 2014 to September 2015) for all urban areas in the country. In particular, we measure the empirical distribution of average expenditure per person and compare them before and after the crime shocks. We

1 We refer to consumption and expenditure interchangeably.

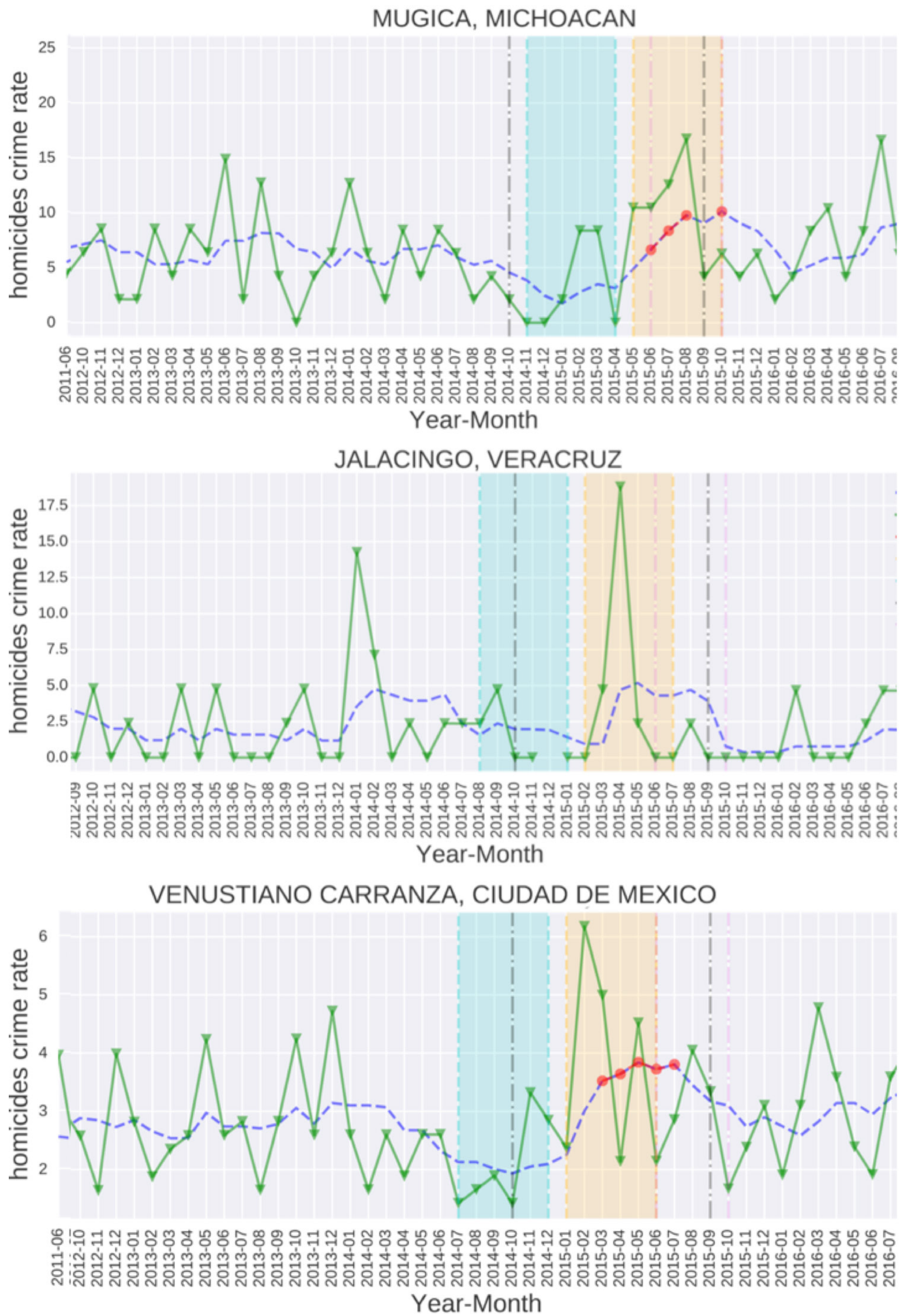


Figure 2: Crime shock examples. Examples of crime shocks in three different municipalities. The six-month windows of the relative peace are shown in blue, followed by six-month windows of high violence, shown in orange. Green lines denote monthly crime rates and dashed blue lines denote six-month moving averages. Red dots indicate six-month percentage changes which are 1.95 standard deviations greater than the average percentage change within the municipality, considering all available homicide data.

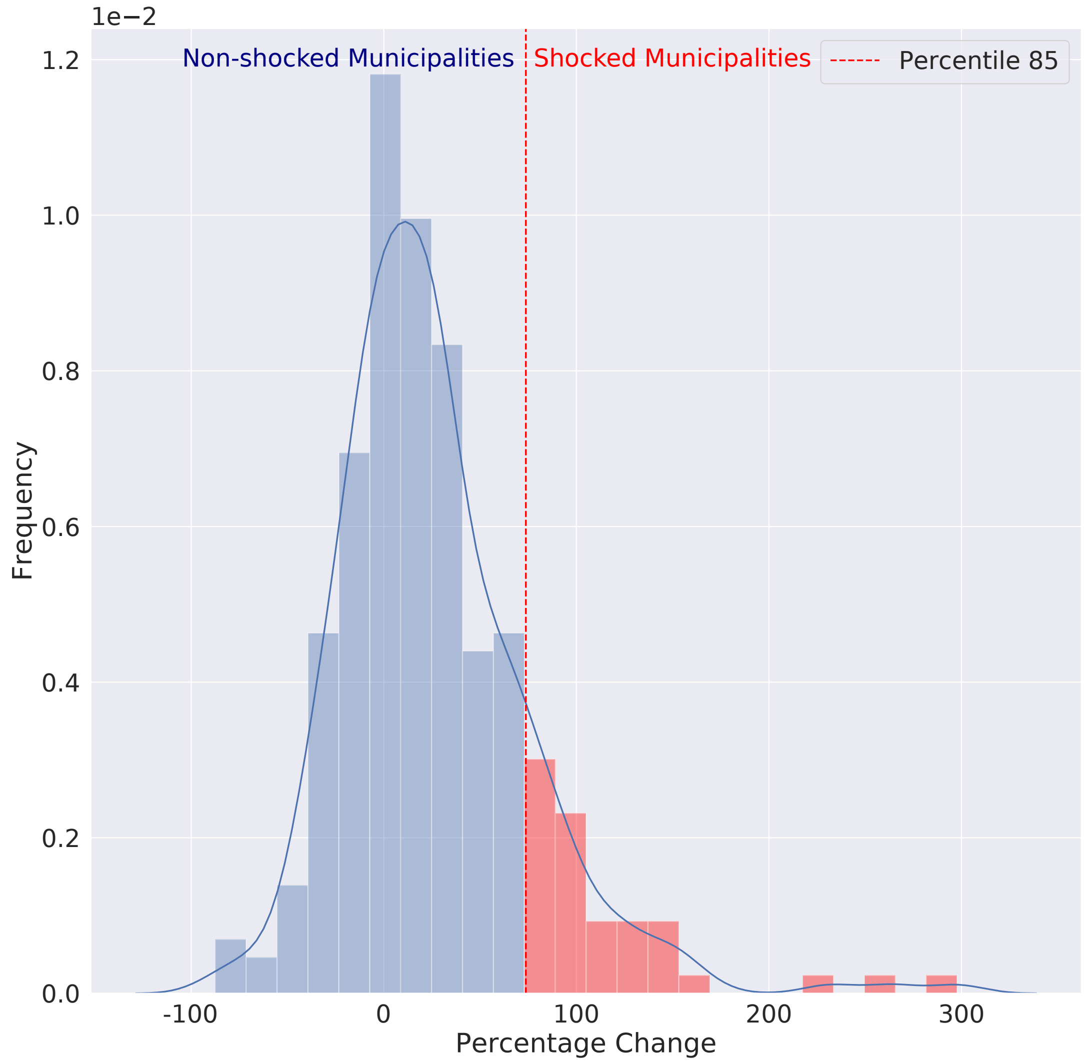


Figure 3: Distribution of changes in homicide rates of municipalities. Observations beyond the 75% change threshold are considered crime shocks. X axis is the percentage change of crime rate in the second six month period from the first six month period in October 2014 to September 2015.

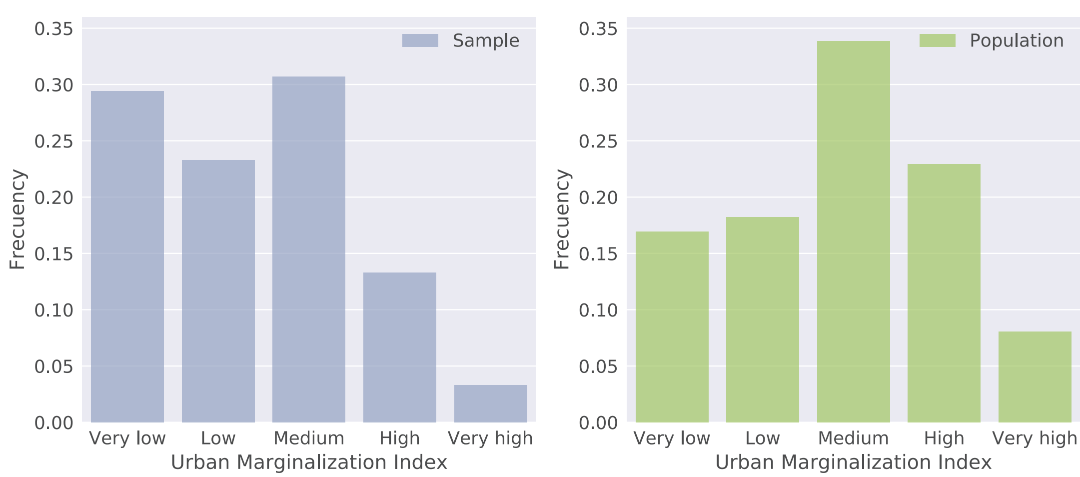


Figure 4: Socioeconomic distribution of consumption data sample vs. the country's population. The *Urban Marginalization* Index is the national statistic used to measure multidimensional deprivation and poverty, in particular associated with schooling, housing, income and residence in urban localities.

further subdivide the distribution per gender (male and female) and socioeconomic groups (defined by very low, low, medium, high and very high marginalization index) to investigate the differential effects across groups.

Socioeconomic distribution and marginalization index. It is relevant to note that our data sample is likely to differ from that of the overall population of the country, as we are restricted to individuals with access to basic financial services. To assess this socioeconomic representativity, we study the distribution of marginalization according to the urban marginalization index as defined by the national statistical office (INEGI)—a multidimensional measure of deprivation and poverty, in particular associated with schooling, housing, income and residence in urban localities. Figure 4 compares the sample and population distributions in terms of urban marginalization index. We note that, as expected, the sample is biased towards individuals in low marginalization segments. However, the sample does include a substantial amount of individuals across very-low, low, medium, and high neighborhood marginalization segments.

Section 3.2.2 studies heterogeneous effects of crime across these segments.

3. Results

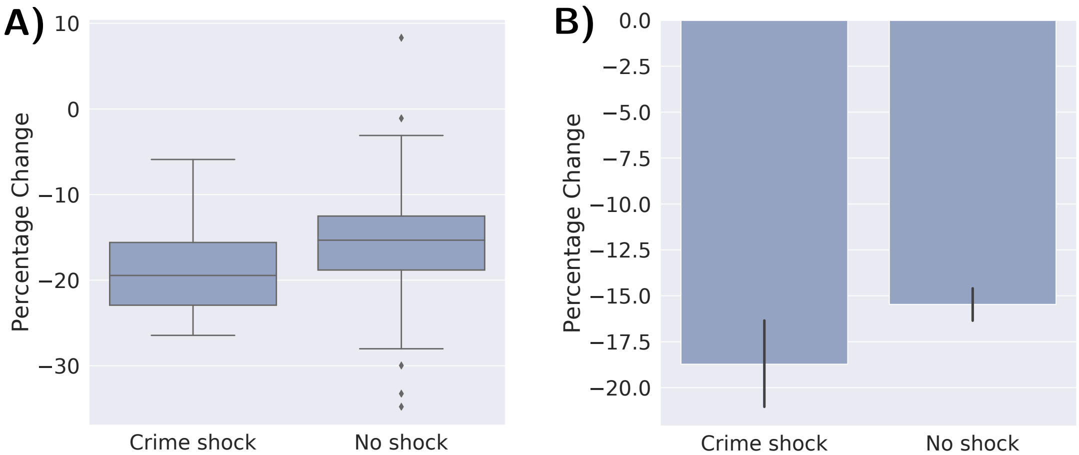


Figure 5: Effect of crime shocks on citizen's consumption patterns. A) Box-plot of average change of individual expenditure in each municipality. B) Bar chart representation with the 95% confidence interval. The shock condition shows that expenditure decreases significantly more in municipalities under crime stress. Expenditure is measured as the average expenditure per client within each municipality. The no shock condition corresponds to municipalities that did not experience a crime shock, and acts as a baseline comparison. Bars correspond to 95% confidence interval around the mean percentage change in expenditure.

3.1 Behavioral Disruptions in the Face of Crime Shocks

Figure 5 summarizes findings about the economic impact of crime shocks on the communities' consumption patterns and the overall economy. The y-axis for the crime shock condition is negative as it corresponds to the percentage decrease in average expenditure activity per person. The expenditure change of no-crime shock condition is also negative, mainly due to seasonality effects of the December period, and it serves as a baseline comparison against the shock condition. As expected, it is observed that the expenditure drop in the shock condition is larger than in the no shock condition, since a persistent wave of criminal activity dissuades households to consume within their municipality and encourages preservation in the face of uncertainty. The difference between the two conditions is above 3% (*p-value* $< .05$).

Results in Figure 5 are generated from average expenditure of residents from municipalities where the debit card transaction data covers more than 2% of the overall population, according to population extrapolations based on the last available census. Figure 8 in Supplementary Materials shows the distribution of sample coverage across municipalities, as well as the 2% threshold applied. In addition, we restrict the analysis to urban populations with more than 75k residents (in accordance with official population-based urban/rural classification). Finally, the no shock condition serves as counterfactual baseline for comparison against expenditure in municipalities that undergo a crime shock, hence controlling for seasonality effects, such as increased expenditure in December.

3.2 Differential Impact to Population Subgroups

As noted in (*28*), different subgroups in the population are affected differently by waves of violence. In this section, we explore these differential impacts across two subgroup dimensions, gender and socioeconomic status. If the effect of violence across gender groups (male and female) are different, one would expect it can exacerbate already existing inequalities especially if violence interferes more with female lives than males'. The same mechanism that lead to differential impacts across gender can lead to deepening inequalities across other dimensions of social or socio-economic status, as suggested by (*28*). Hence, we investigate whether waves of violence disrupt economic activity at different rates across neighborhoods with low and high urban marginalization index (as described in Section 2.2).

3.2.1 Across Gender

Women's concerns and behaviors in the public space differ from men's in ways that can be broadly associated with gender roles and a set of beliefs about safety (*29*). As evidenced by researchers, the experience of violence, including victimization and perception of security, is different among men and women (*18–20*), leading to differential behavioral changes. These changes in behavior can further exacerbate existing inequalities, such as unequal endowment with social capital (*30*) and access to new economic opportunities, ultimately slowing prospects of personal growth. Given these facts, any difference in how crime dissuades economic activity and social mobility across gender groups would mean that a crime shock would potentially exacerbate the existing level of inequality between men and women. In particular, women could be more constrained for a long period after a crime shock, hence limiting their access to opportunities in comparison to men.

This study quantifies the differential effect of crime across gender groups. Figure 6 shows our main results comparing the effect of crime shocks between men and women, based on changes in their respective expenditures aggregated at the municipality level. As hypothesized, the behavioral disruption effected by crime shocks is stronger for females

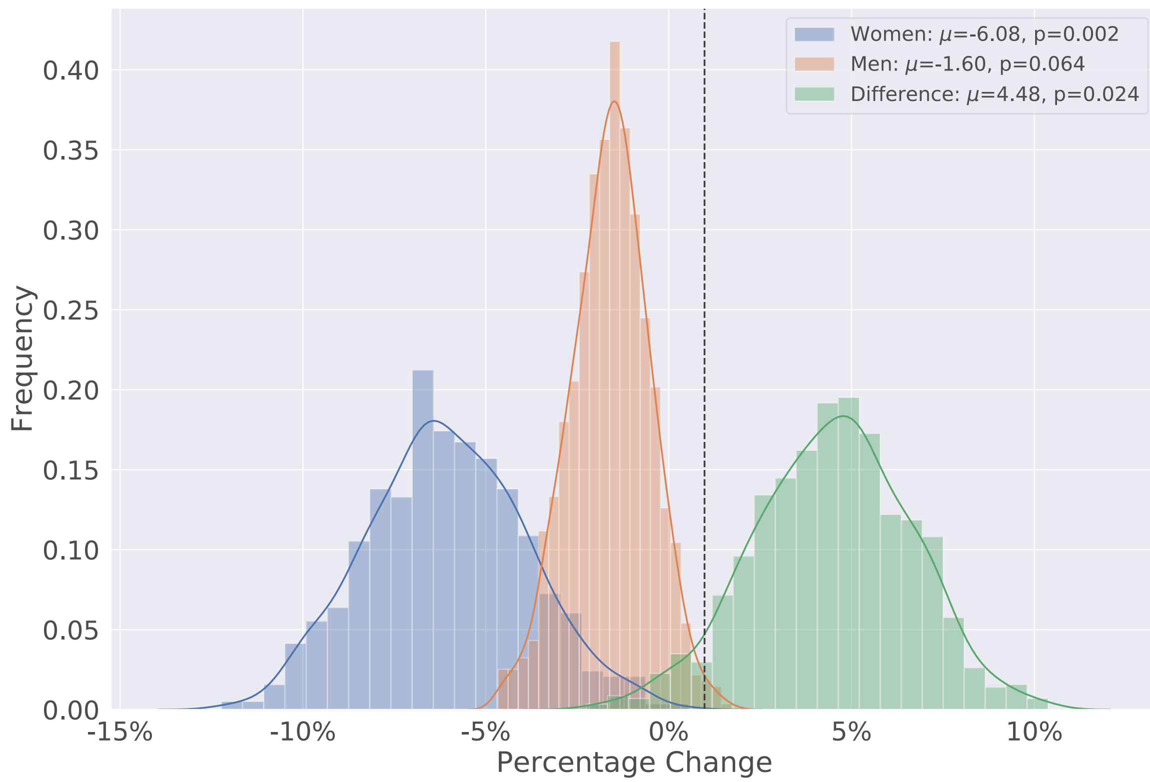


Figure 6: Effects of crime shocks on expenditure across gender. X-axis is the percentage change of expenditure after the crime shock when the trend from no crime shock condition is subtracted. The mean observed effect for males is −1.6%, and −6.1% for females, yielding a gender disparity of 4.5%. The p-values for male and female refer to the hypothesis that the activity change after the crime shock is zero. The p-value for the difference refers to the hypothesis that the difference between the two groups is zero. The distribution of difference is obtained through bootstrapping of difference between males and females.

than males. Crime suppresses female activity (6.1% decline on average) at a rate more than three times larger than males (1.6% decline on average), leading to a gender gap of 4.5%. The diff-in-diff activity across genders is statistically significant at a 0.05 level (p=0.024) obtained after bootstrapping the distribution of mean difference.

3.2.2 Across Socio-economic Status

We repeat the same analysis for groups with different socioeconomic status. We use the marginalization index released as part of annual Mexico Census for different districts (there are multiple districts within one municipality). Districts with high marginalization tend to have lower access to education and other public services, and districts with low marginalization tend to be from high socio-economic status. We expect behavioral impacts of crime to hit areas with higher marginalization harder, due to higher mobility and economic constraints. For the purpose of comparison, we merged groups with very low and low marginalization into one group denoted as low and the rest into the second group denoted as high marginalization.

Figure 7 shows our main results on differential effects of crime shocks on districts with different marginalization status. Each district is assigned with a percentage change in average per-person expenditure after subtracting the baseline value from the no crime shock condition. The figure illustrates the distribution of these extra changes in expenditure after a crime shock across different districts with high and low marginalization. The results indeed confirm our expectation that within a municipality affected by a crime shock, districts of higher socioeconomic status (low marginalization) are less impacted by crime shocks, than districts with lower socioeconomic status (high marginalization), as compared to districts in a municipality not affected by a crime shock. Crime suppresses economic activity in high and low marginalized districts by −6.7% and −3.7%, on average respectively, leading to a 3.0% socioeconomic gap in the behavioral impact of crime. The diff-in-diff activity across low and high marginalization is statistically significant at a 0.1 level obtained after bootstrapping the distribution of differences across the groups. This mean difference is only marginally significant

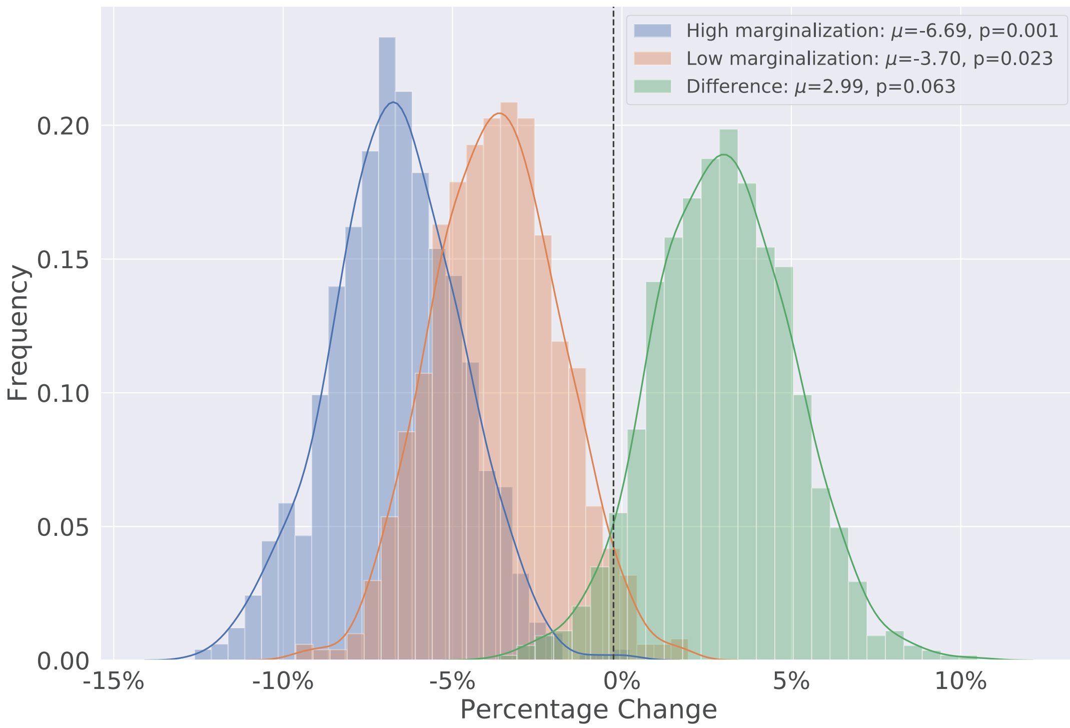


Figure 7: Effects of crime shocks on expenditure across different socioeconomic status. X-axis is the percentage change of expenditure after the crime shock when the trend from no crime shock condition is subtracted. The p-values for high and low marginalization refer to the hypothesis that the activity change after the crime shock is zero. The p-value for the difference refers to the hypothesis that the mean difference between the two groups is zero. The distribution of difference is obtained through bootstrapping of difference between the two groups.

(*p-value* = .063) mainly due to small sample size. As mentioned in Section 4, in near-future work we will increase the size of crime shock samples by accessing larger longitudinal periods of overlapping card and crime data, in order to increase the precision of differential effects' estimates, as well as move into a cross-sectional approach across gender and socio-economics.

4. Discussion and Future Work

In this work we have shown the effect that crime and violence have on citizens' daily lives. We undertook a novel approach that uses societal-scale information from anonymized card transaction metadata, to compute indices related to the consumption patterns of the population. We show that crime shocks have a negative effect on citizens' consumption patterns and that this effect can be unequally distributed across population subgroups.

A couple of data-related challenges constrained the extent of the analysis here presented. Most importantly, the present study relied on available data spanning one calendar year, for which both card transaction and crime data overlapped. This longitudinality constraint fixed the pre- and post-shock six-month window periods, hence reducing substantially the amount of crime shocks studied. Consequently, finer grain analysis, such as cross-sectional socioeconomic and gender analysis was not attainable due to power constraints; and similarly the study of heterogeneities across municipality types. In contrast, we have recently started a close collaboration with Banorte, the second largest bank in the Mexican market. Through a direct, on-site collaboration with their analytics team, we look forward to build and extend substantially on the results and framework here presented. In particular, by leveraging the project's relevant longitudinality from one to five years.

The results here presented are first steps towards mapping, understanding, and helping manage the effects of crime shocks on affected communities. Several avenues of work lie immediately ahead. First, regarding categories of consumption, it exists a rich categorization of merchants, which can be used to elicit category-specific effects. Future work will extract hierarchical meta-categories, to provide insight into how different domains of people's routine are differentially affected (e.g., entertainment and transportation).

Second, future work will explore at a finer grain how crime shocks unequally affect different community subgroups. In particular, by conducting a cross-sectional analysis at the intersection of gender and socioeconomic status, which is of great relevance towards translating observational insights into policy interventions. Third, future work will explore complementarities of other sources of data—mobile phone records in particular—and the additional insight they can provide in terms of effects on the mobility patterns of individuals. This additional data can ultimately be used to map more complex effects of high relevance, such as the relationship between crime shocks and community segregation. Finally, we expect future work to pursue causal insights from these wealth of observational data, by building detailed studies that leverage natural experiments in the data, or more sophisticated econometric estimators.

Overall, this paper proposes the use of digital footprints (metadata) to build indices of communities' behavioral patterns, at a large-scale and fine grain, in order to monitor the effects of crime and other types of shocks to local communities. The guiding vision of this work is a prospective ecosystem where the public sector, the private sector, and civil society participate in monitoring, communicating and managing the dynamics of disruption and recovery of communities in the face of crime, violence, and other adverse events. Today, several elements of such ecosystem already exist—the societal-scale behavioral data, technologies for safe data sharing (*31*), and policy-support platforms that aim at aiding local policy decisionmaking (*32*)—thus we expect these elements to connect and form such capabilities in the near future.

5. References and Notes

1. S. Pinker, *The better angels of our nature*: A *history of violence and humanity*. Penguin, 2011.
2. T. Economist, "How to cut the murder rate." https://www.economist.com/leaders/2018/04/05/how-to-cut-the-murder-rate, 2018 (accessed October 7, 2018).
3. C. J. Vilalta Perdomo, J. G. Castillo, and J. A. Torres, "Violent crime in latin american cities," tech. rep., Inter-American Development Bank, 2016.
4. C. Vilalta and R. Muggah, "Violent disorder in ciudad juarez: A spatial analysis of homicide," *Trends in organized crime*, vol. 17, no. 3, pp. 161–180, 2014.
5. C. Vilalta and R. Muggah, "What explains criminal violence in mexico city? a test of two theories of crime," *Stability: International Journal of Security and Development*, vol. 5, no. 1, 2016.
6. S. Machin, O. Marie, and S. Vujic, "The crime reducing effect of education," ´The *Economic Journal*, vol. 121, no. 552, pp. 463–484, 2011.
7. I. Chatterjee and R. Ray, "Crime, corruption and the role of institutions," *Indian Growth and Development Review*, vol. 7, no. 1, pp. 73–95, 2014.
8. C. M. Fuentes Flores, "El impacto de las viviendas deshabitadas en el incremento de delitos (robo a casa habitacion y homicidios) en ciudad juárez, chihuahua, 2010," *Frontera norte*, vol. 27, no. 54, pp. 171–196, 2015.
9. G. Robles, G. Calderon, and B. Magaloni, "Las consecuencias econ ´omicas de la violencia del narcotrafico en m éxico," tech. rep., IDB Working Paper Series, 2013.
10. J. Chabat, "Combatting drugs in mexico under calderon: the inevitable war," 2010. ´
11. O. S. J. Initiative et al., "Undeniable atrocities: Confronting crimes against humanity in mexico," *Open Society Foundations*, *New York*, 2016.
12. G. Trejo and S. Ley, "Municipios bajo fuego," *Nexos*. Febrero. En: http://www. nexos. com. mx, 2015.
13. L. Jaitman, D. Caprirolo, R. Granguillhome Ochoa, P. Keefer, T. Leggett, J. A. Lewis, J. A. Mejía-Guerra, M. Mello, H. Sutton, and I. Torre, "The costs of crime and violence: New evidence and insights in latin america and the caribbean," 2017.

14. N. Ajzenman, S. Galiani, and E. Seira, “On the distributive costs of drug-related homicides,” *The Journal of Law and Economics*, vol. 58, no. 4, pp. 779–803, 2015.

15. D. Rebolledo Sanchez, “La violencia como limitante para el desarrollo y el crecimiento económico en el estado de guerrero,” *Revista Mexicana de Ciencias Agr ícolas*, no. 12, 2015.

16. A. T. V. Montemayor, *Drug Violence, Fear of Crime and the Transformation of Everyday Life in the Mexican Metropolis*. University of California, Berkeley, 2016.

17. U. N. O. on *Drugs and Crime, Global study on homicide 2013: trends, contexts, data*. UNODC, 2013.

18. M. Averdijk, “Reciprocal effects of victimization and routine activities,” *Journal of Quantitative Criminology*, vol. 27, no. 2, pp. 125–149, 2011.

19. G. S. Mesch, “*Perceptions of risk, lifestyle activities, and fear of crime,” Deviant Behavior*, vol. 21, no. 1, pp. 47–62, 2000.

20. J.-A. Gale and T. Coupe, “The behavioural, emotional and psychological effects of street robbery on victims,” *International Review of Victimology*, vol. 12, no. 1, pp. 1–22, 2005.

21. R. B. Gutierrez and E. F. Molina, “El miedo al delito en las mujeres en méxico. expresión y determinantes,”

22. M. E. Avila, B. Martínez-Ferrer, A. Vera, A. Bahena, and G. Musitu, “Victimization, perception of insecurity, and changes in daily routines in mexico,” *Revista de saude publica*, vol. 50, p. 60, 2016.

23. A. . Pentland, “The data-driven society,” *Scientific American*, vol. 309, no. 4, pp. 78–83, 2013.

24. J. Blumenstock, G. Cadamuro, and R. On, “Predicting poverty and wealth from mobile phone meta-data,” *Science*, vol. 350, no. 6264, pp. 1073–1076, 2015.

25. E. A. Martinez-Cesena, P. Mancarella, M. Ndiaye, and M. Schlapfer, “Using mobile phone data forelectricity infrastructure planning,” *arXiv preprint arXiv*:1504.03899, 2015.

26. M. Berlingerio, F. Calabrese, G. Di Lorenzo, R. Nair, F. Pinelli, and M. L. Sbodio, “Allaboard: a system for exploring urban mobility and optimizing public transport using cellphone data,” in Joint *European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 663–666, Springer, 2013.

27. A. Bogomolov, B. Lepri, J. Staiano, E. Letouze, N. Oliver, F. Pianesi, and A. Pentland, “Moves on the street: Classifying crime hotspots using aggregated anonymized data on people dynamics,” *Big data*, vol. 3, no. 3, pp. 148–158, 2015.

28. I. Carvalho and D. A. Lewis, “Beyond community: Reactions to crime and disorder among inner-city residents,” *Criminology*, vol. 41, no. 3, pp. 779–812, 2003.

29. C. B. Gardner, “Safe conduct: Women, crime, and self in public places,” *Social problems*, vol. 37, no. 3, pp. 311–328, 1990.

30. N. Lin, “Inequality in social capital,” *Contemporary Sociology*, vol. 29, no. 6, pp. 785–795, 2000.

31. T. Hardjono, D. Shrier, and A. Pentland, “Opal/enigma,” in TRUST:: DATA: *A New Framework for Identity and Data Sharing*, ch. 3, pp. 79–99, ,: Visionary Future LLC, 2016.

32. “Plataforma preventiva.” Secretaria de Desarrollo Social (SEDESOL), Mexico, http://plataformapreventiva.gob.mx/. Accessed: 2018-06-15.

6. Supplementary Material

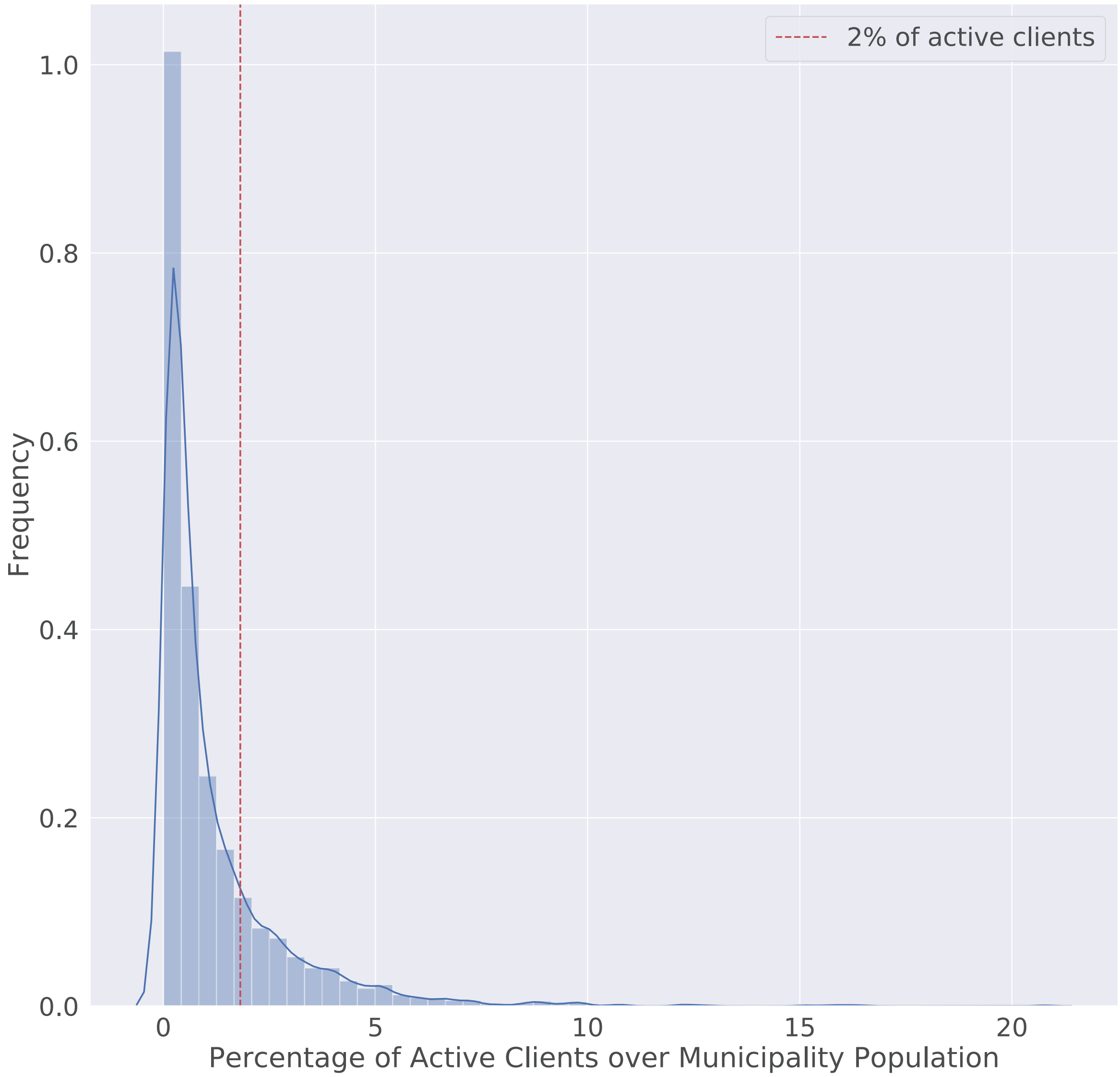


Figure 8: **The distribution of percentage of municipality population present in the debit transaction data. The municipalities below the 2% threshold in dashed red are discarded.** Only municipalities where active bank clients—clients which made at least one transaction during the period of analysis—represent more than 2% of the official estimated population were included in the study, in order to ensure a minimum degree of representativity in levels of expenditure in each municipality.

Papiers de Recherche de l'AFD

Les *Papiers de Recherche de l'AFD* ont pour but de diffuser rapidement les résultats de travaux en cours. Ilss'adressent principalement aux chercheurs, aux étudiants et au monde académique. Ils couvrent l'ensemble des sujets de travail de l'AFD : analyse économique, théorie économique, analyse des politiques publiques, sciences de l'ingénieur, sociologie, géographie et anthropologie. Une publication dans les Papiers de Recherche de l'AFD n'en exclut aucune autre.

L'Agence Française de Développement (AFD), institution financière publique qui met en oeuvre la politique définie par le gouvernement français, agit pour combattre la pauvreté et favoriser le développement durable. Présente sur quatre continents à travers un réseau de 72 bureaux, l'AFD finance et accompagne des projets qui améliorent les conditions de vie des populations, soutiennent la croissance économique et protègent la planète. En 2014, l'AFD a consacré 8,1 milliards d'euros au financement de projets dans les pays en développement et en faveur des Outre-mer.

Les opinions exprimées dans ce papier sont celles de son (ses) auteur(s) et ne reflètent pas nécessairement celles de l'AFD. Ce document est publié sous l'entière responsabilité de son (ses) auteur(s).

Les *Papiers de Recherche* sont téléchargeables sur : http://librairie.afd.fr/

AFD Research Papers

AFD Research Papers are intended to rapidly disseminate findings of ongoing work and mainly target researchers, students and the wider academic community. They cover the full range of AFD work, including: economic analysis, economic theory, policy analysis, engineering sciences, sociology, geography and anthropology. *AFD Research Papers* and other publications are not mutually exclusive.

Agence Française de Développement (AFD), a public financial institution that implements the policy

defined by the French Government, works to combat poverty and promote sustainable development. AFD operates on four continents via a network of 72 offices and finances and supports projects that improve living conditions for populations, boost economic growth and protect the planet. In 2014, AFD earmarked EUR 8.1bn to finance projects in developing countries and for overseas France.

The opinions expressed in this paper are those of the author(s) and do not necessarily reflect the position of AFD. It is therefore published under the sole responsibility of its author(s).

AFD Research Papers can be downloaded from: http://librairie.afd.fr/en/

AFD, 5 rue Roland Barthes
75598 Paris Cedex 12, France
ResearchPapers@afd.fr
ISSN 2492 - 2846

BIG DATA

TO ADDRESS GLOBAL DEVELOPMENT CHALLENGES

2018