Codebook for Data Collection

Adee Rose Weller

To measure violence against politicians, I collected an original dataset on fatal attacks against local politicians, building from newspaper reports. Collected during 2021 and 2022, this dataset contains the record of fatal attacks against mayoral candidates, pre-candidates, mayors, and former mayors from January 1, 2013, to June 30, 2021. In total, there were 414 recorded attacks. This data was collected by scraping local, state, and national newspaper articles containing keywords relating to assassinations, attacks, mayors, candidates, and former mayors.[1] The articles were identified using keywords such as "*alcalde*" and "*asesinado*," as well as cross-referenced with other articles, reports and more limited datasets.

In this dataset, the date of the attack, state and municipality of the attack and of the victim's home, victim's profession and any time spent in office, party affiliation (if applicable), and gender are all recorded. I also recorded the source(s) of the report and a few sentences summary of the details of the attack, including what happened, when, how and to whom, as well as any important comments. These include references to potential perpetrators if the article discussed it, any allegations of corruption or ties to criminal groups, or past experiences of the victim with threats or other attacks. While smaller acts of violence, like threats or arson, might not be covered consistently, fatal violence is likely to be reported, even if only by smaller, more regionally specific papers.

To identify attacks by organized criminal groups, I examine three key indicators. First, if the victim was shot, we can likely associate the attack with organized crime. Guns are significantly more difficult to attain in Mexico than in the United States,[2] and the use of them is closely associated with the actions of organized criminal groups (Dube, Dube, and García-Ponce 2013). The second indication that the crime was committed by an organized group is that multiple people were involved (Pérez Esparza and De Paz Mancera 2018). Finally, when an individual is kidnapped and taken to a second location before they are killed indicates an organized act, rather than a crime of passion.

[1] See Appendix B for more information about how this dataset was collected and the coding practices that were used.

[2] There is only one legal gun shop in the country, located in Mexico City, and access to it is heavily restricted by the Ministry of National Defense (Dube, Dube, and García-Ponce 2013).

Since this dataset was built from newspaper reports, we also must consider how organized crime might influence that reporting. Mexico has one of the highest rates of assassinations of journalists in the world, acts committed both by criminal groups and by state actors (Ahmed 2017). In some places where criminal groups can both engage in politics and control the reporting of journalists, this could present a challenge to this analysis. In such cases, at the highest degrees of criminal engagement, the measure presented here may indicate little engagement, despite killings occurring. Alternatively, criminal groups may actually *encourage* this kind of reporting to maximize its signaling and chilling effects to nearby areas, strategically using the violence for its visible purpose (Brancati and Penn 2022).

To combat this, I used as many outside sources for cross-referencing and verification as I could. After building this dataset, the information was cross-referenced from the *Votar Entre Balas* dataset from CIDE, which has attacks from 2018 to the present, reports from *Alcaldes de México*, the National Mayors Association of Mexico, and the ACLED dataset for Mexico. These sources are somewhat limited in timeframes, but when combined, can increase the confidence in this data. However, it is important to note that this is not, nor does it claim to be, a perfect list of attacks during the period.

1. Sources

    - For information on the outcome, data can be gathered from newspaper articles, mainly by hand and a small amount using Python to collect articles.
    - For biographical information, can use campaign websites or newspaper articles about candidacy, or Facebook posts
    - Key words for finding articles are *"alcalde"* or *"alcaldesa"* or *"edil"* or *"corregidor"* or *"presidente municipal"* (mayor), *"alcaldía"* (mayoral seat), *"candidato"* or *"candidata"* (candidate), *"ex candidato"* or *"ex candidata"* (former candidate), *"asesinado"* or *"asesinan"* or *"muerto"* or *"mató"* or *"hallan"* (killed), *"fallecimiento"* (death), *"atacó"* (attacked), *"tiró"* (shot), *"secuestrado"* (kidnapped), *"México"* (Mexico), as well as state-specific and/or municipality-specific names.
    - Top websites for confirmation: *Mileno, Reforma, El Tiempo, Infobae, alcaldesdemexico, INEGI, Reuters, WRadio.com.mx, Telemundo, BBC*
    - *Votar entre Balas* data for cross-referencing: using reports of attacks from 2018 to 2021. Matched identifying characteristics and attack information.
    - ACLED data for cross-referencing: Filtered to Mexico, read through each observation and removed any that did not fit the coding criteria outlined in the paper.

Then, cross-referenced newspaper articles with the data to increase internal validity.

- When there was a disconnect on the dates of an attack, but the other details remained the same, I recorded the earlier date for consistency. Usually, these dates were a day apart for an attack that occurred overnight, so the true date is unclear.

2. Process

- Every event was given a unique ID code.
- The day, month, and year of the attack were recorded.
- The type of attack, the perpetrator(s) and victim, their gender, profession, as well as victim's party were recorded.
- The attack was geo-coded.
- The sources and their scope were recorded.
- A paragraph summary of the attack was written.
- The number of total fatalities was recorded.