Case Study 3: Mapping Kidnappings in Mexico

David Buil-Gil, Reka Solymosi and Eon Kim

2024-01-25

In this spatial analysis activity, we explore the spatial distribution of kidnapping crime in Mexico. Kidnapping, defined as the unlawful abduction or captivate of individuals against their will, carries profound security implications. Within the context of Mexico, kidnapping is closely linked to organised crime, particularly drug cartels, and poses a challenge to national security (Jones, 2013; Massa and Fondevila, 2021; Ochoa, 2012, 2019). Kidnapping in Mexico is largely associated to the demand for illicit substances in the United States, leading to substantial financial costs for Mexican criminal organisations. These groups also exploit individuals, including children, for various criminal purposes, including kidnappings. Understanding the geographic distribution of these crimes may help developing better-informed crime prevention policies targeting areas that concentrate most offenses.

In this exemplar study we utilize police-recorded crime data released by the National Public Security System (SNSP) through the official portal of the Mexican Government <https://www.gob.mx/sesnsp/acciones-y-programas/datos-abiertos-de-incidencia-delictiva?state=published>.

We begin by loading the required packages in R.

library(here) # to identify the path to the data  
library(readr) # to read in CSV data  
library(dplyr) # for data wrangling  
library(ggplot2) # for data visualisation  
library(sf) # for spatial data manipulation and visualisation  
library(viridis) # for colour schemes

The data are saved in a CSV (comma-separated data) and we employ the read\_csv() function from the readr package, specifying the Latin1 encoding to correctly import the crime data into the data\_Mexico object.

#Read csv file with crime data  
data\_Mexico <- read\_csv(here("data/IDM\_nov2023.csv"), locale = locale(encoding = "Latin1"))

We examine the dataset and identify the three variables of interest. - TIPO (in English ‘TYPE’: type of crime) - AÑO (in English ‘YEAR’: year of crime recorded) - ENTIDAD (in English ‘ENTITY’: state of Mexico)

The data contains 18 different crime types across 32 federal entities in Mexico from 2011 to 2017. To prepare the data for analysis, we first reshape the dataset by filtering the records specific to the year 2017 and the crime type SECUESTRO (i.e., kidnapping).

data\_Mexico <- data\_Mexico %>%  
 filter(TIPO == "SECUESTRO") %>% # filter SECUESTRO  
 filter(AÑO == 2017) # filter 2017

We create a table which contains two columns: the state name (ENTIDAD) and the total number of kidnappings for each state. The data records kidnapping crimes monthly, with separate columns for each month of the year. To calculate the annual total for each state, we use the rowSums() function, and specify the columns to be included using the accross() function. It is important to set the na.rm parameter to TRUE to ensure that missing data (NAs) are just treated as 0 (i.e., no kidnapping crimes occurred in those months).

data\_Mexico <- data\_Mexico %>%  
 mutate(sum\_secuestro = rowSums(across(8:19), # create new variable of sums  
 na.rm = TRUE)) # treat NA as 0 here

We then compute the total number kidnappings in each state (ENTIDAD) in 2017. By using the group\_by() and summarise() functions, we create a frequency table, which is saved as a new object called data\_Mexico\_states. In this table, each row corresponds to each state, providing a overview of kidnappings in 2017.

#Calculate number of crimes in each state  
data\_Mexico\_states <- data\_Mexico %>%  
 group\_by(ENTIDAD) %>% #group by state  
 summarise(secuestro = sum(sum\_secuestro)) # sum all kidnappings in the state

To explore the data before mapping the frequencies, we look at the top 3 states with the highest number of kidnappings. The state Mexico has the highest numbers of kidnappings, totaling 173 crime records, followed by Veracruz with 172 incidents.

top\_n(data\_Mexico\_states, 3, secuestro)

## # A tibble: 3 × 2  
## ENTIDAD secuestro  
## <chr> <dbl>  
## 1 MEXICO 173  
## 2 TAMAULIPAS 140  
## 3 VERACRUZ 172

We observe the top three states (MEXICO, TAMAULIPAS, and VERACRUZ) stand out, with remarkably larger number of kidnappings that the rest of the country. To assess if these states are more at risk to kidnappings, we considered the kidnapping rate per population. It is possible that the three states have higher kidnapping counts because they have large populations, where more people can lead more crime incident, including kidnappings. To account this, we calculate kidnapping rates per 100,000 residents. We obtained census data from the National Institute of Statistics and Geography (INEGI): <https://www.inegi.org.mx/app/tabulados/default.html?nc=mdemo02>.

#Read csv file with population data  
population <- read\_csv(here("data/Population2010.csv"))

## Rows: 32 Columns: 2  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): STATE  
## dbl (1): Population2010  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

To link the population data with the kidnapping data, we combined the two tables using a common identifier, which remains as ENTIDAD in the crime data and STATE in INEGI’s population statistics. We employ the left\_join() function.

#Merge with crime data and calculate crime rates  
data\_Mexico\_states <- data\_Mexico\_states %>%  
 left\_join(population, by = c("ENTIDAD" = "STATE")) %>%  
 mutate(secuestro\_rate = secuestro / Population2010 \* 100000)

We examine the top 5 states with the highest kidnapping rate per 100,000.

top\_n(data\_Mexico\_states, 3, secuestro\_rate)

## # A tibble: 3 × 4  
## ENTIDAD secuestro Population2010 secuestro\_rate  
## <chr> <dbl> <dbl> <dbl>  
## 1 TABASCO 77 2238603 3.44  
## 2 TAMAULIPAS 140 3268554 4.28  
## 3 ZACATECAS 67 1490668 4.49

We observe that the states with the highest kidnapping rates are Tabasco and Zacatecas. Tamaulipas consistently stood out in both a high count and a high rate of kidnappings. This highlights the significant of exploring potential problem-solving interventions to mitigate kidnapping incident in this region.

To further our understanding, we examine the spatial distribution of kidnappings and inspect any potential spatial relationships among these states through a thematic map. The spatial data in JSON format used in this exercise was obtained from the following source: <https://github.com/strotgen/mexico-leaflet/>.

To link geographical information with kidnapping rates, we begin by importing geospatial data, which represents the Mexican state’s boundaries. We then standardise state names and merge the geospatial data with the kidnapping data to create a new data for spatial analysis.

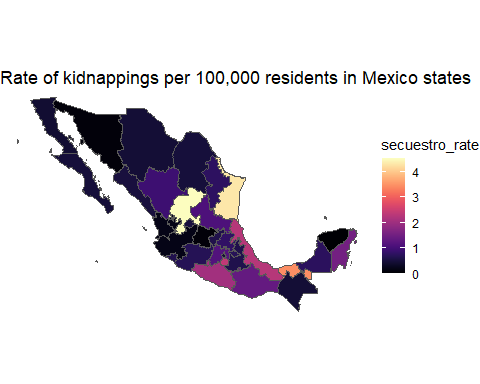
#Read geojson of Mexico states  
#states\_geojson <- st\_read("https://github.com/strotgen/mexico-leaflet/blob/master/states.geojson")  
states\_geojson <- st\_read(here("data/states.geojson"))

## Reading layer `states' from data source   
## `\\nask.man.ac.uk\home$\Documents\GitHub\crim-data-south2\data\states.geojson'   
## using driver `GeoJSON'  
## Simple feature collection with 32 features and 3 fields  
## Geometry type: MULTIPOLYGON  
## Dimension: XY  
## Bounding box: xmin: -118.4 ymin: 14.5321 xmax: -86.72404 ymax: 32.71865  
## Geodetic CRS: WGS 84

#Merge crime rates with geojson file  
states\_geojson <- states\_geojson %>%  
 mutate(state\_name = toupper(state\_name), #capital letters for consistency  
 state\_name = recode(state\_name, #rename some states for consistency  
 'DISTRITO FEDERAL' = 'CIUDAD DE MEXICO',  
 'MÉXICO' = 'MEXICO',  
 'MICHOACÁN DE OCAMPO' = 'MICHOACAN',  
 'QUERÉTARO' = 'QUERETARO',  
 'SAN LUIS POTOSÍ' = 'SAN LUIS POTOSI',  
 'VERACRUZ DE IGNACIO DE LA LLAVE' = 'VERACRUZ',  
 'NUEVO LEÓN' = 'NUEVO LEON',  
 'COAHUILA DE ZARAGOZA' = 'COAHUILA',  
 'YUCATÁN' = 'YUCATAN')) %>%   
 left\_join(data\_Mexico\_states, by = c("state\_name" = "ENTIDAD"))

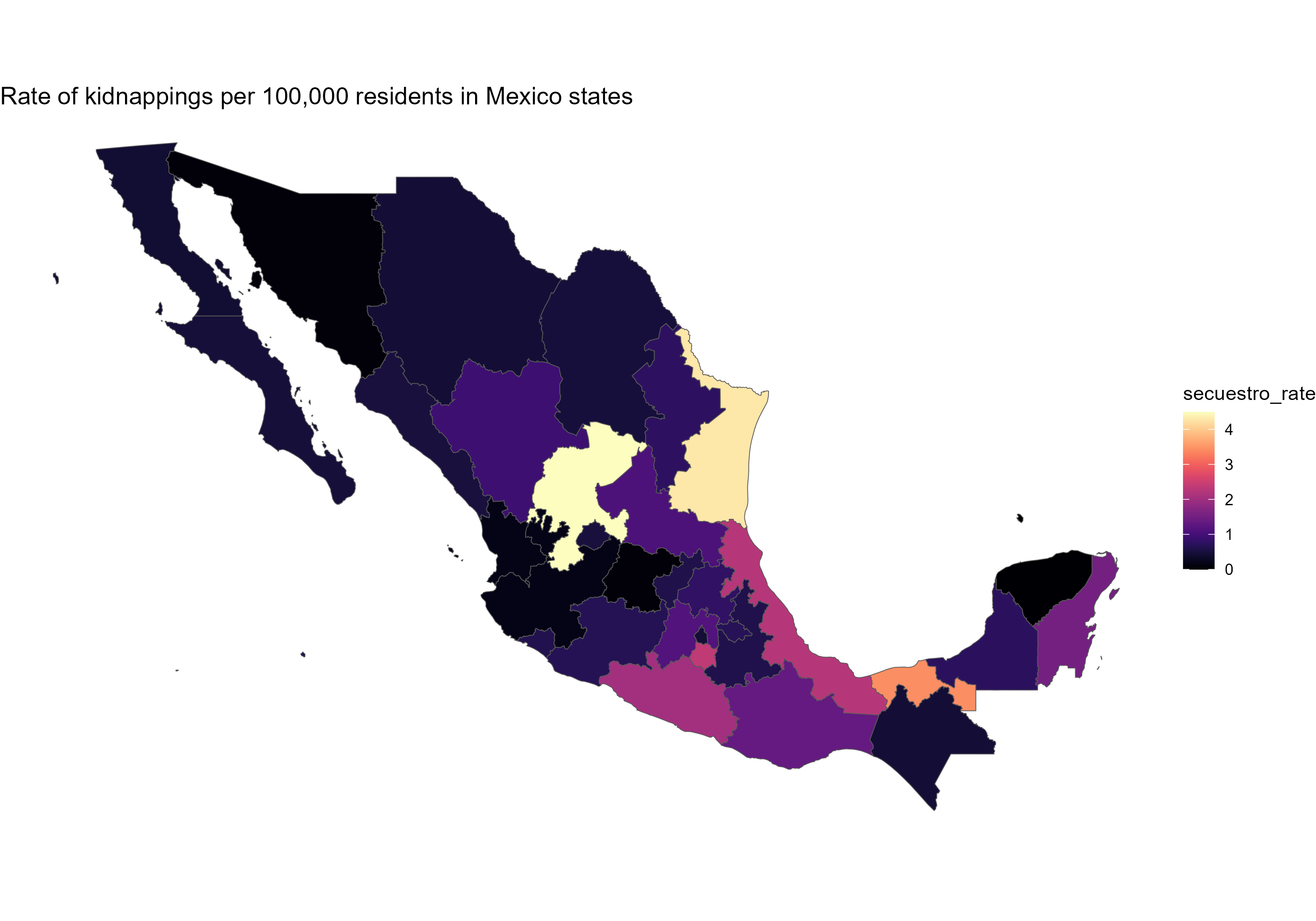
We generate a heatmap of kidnapping rates across Mexican states utilising the ggplot2 package and customise the fill colors. The output has been saved as a PNG file named ‘Mexico\_map.png’ in the specified directory.

ggplot(data = states\_geojson) +  
 ggtitle("Rate of kidnappings per 100,000 residents in Mexico states") +  
 geom\_sf(aes(fill = secuestro\_rate)) +  
 scale\_fill\_viridis(option = "magma") +  
 theme\_void()



ggsave(here("exemplar-activities/Mexico\_map.png"),   
 width = 10, height = 7)

The heatmap illustrates the rate of kidnapping per 100,000 population across 32 states in Mexico. Yellow/orange shades on the map indicate higher kidnapping rates, while darker shades represent lower rates. These variations provide valuable insights for targeted interventions, highlighting areas where increased law enforcement resources may be necessary to effectively address and mitigate kidnapping incidents. This spatial analysis underscores the influence of strategic choices in crime prevention.



**References**

Jones, N. (2013). The unintended consequences of kingpin strategies: kidnap rates and the Arellano-Félix Organization. Trends in Organized Crime, 16, 156–176. <https://doi.org/10.1007/s12117-012-9185-x>

Massa, R., and Fondevila, G. (2021). Criminal displacement in Mexico city’s metropolitan area: The case of kidnapping. International Journal of Law, Crime and Justice, 67, 100479. <https://doi.org/10.1016/j.ijlcj.2021.100479>

Ochoa. R. (2012). Not just the rich: New tendencies in kidnapping in Mexico City. Global Crime, 13(1), 1-21. <https://doi.org/10.1080/17440572.2011.632499>

Ochoa, R. (2019). Intimate crimes: Kidnapping, gangs and trust in Mexico. Oxford: Oxford University Press.