Case Study 3: Geographic distribution of crime in Mexico

David Buil-Gil and Reka Solymosi

2024-01-25

CONTEXT

We analyse police-recorded crime data made available in the webside 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)  
library(dplyr)  
library(ggplot2)  
library(sf)  
library(viridis)

We open this dataset and select only those rows that refer to our crime type of interest: kidnappings (or ‘secuestro’ in Spanish). We are particularly interested in exploring the geographic distribution of crimes in 2017; the last year when data was made available.

#Read csv file with crime data  
data\_Mexico <- read.csv(here("data/IDM\_nov2023.csv"))  
  
#Select crime type of interest in dataset, and only records for 2017  
data\_Mexico <- data\_Mexico %>%  
 filter(MODALIDAD == "PRIV. DE LA LIBERTAD (SECUESTRO)") %>%  
 filter(ANO == 2017)

Calculate number of crimes each year in each state (‘ENTIDAD’ in the database).

#Calculate number of crimes across months  
data\_Mexico <- data\_Mexico %>%  
 mutate(freq = rowSums(select(., 8:19), na.rm = TRUE))  
  
#Select variables of interest only  
data\_Mexico <- data\_Mexico %>%  
 select(ANO, ENTIDAD, TIPO, freq)  
  
#Calculate number of crimes in each state  
data\_Mexico\_states <- data\_Mexico %>%  
 group\_by(ENTIDAD) %>%  
 summarize(freq = sum(freq))

We have now created a new database called ‘data\_Mexico\_states’ which includes the count of kidnappings for each state in 2017. We can execute ‘top\_n(data\_Mexico\_states, 3, freq)’ and observe that the State of Mexico concentrates the largest number of kidnappings, 173, followed by Veracruz, with 172. On the other end, Yucatan recorded 0 kidnappings in 2017. On average, 35.9 kidnappings were recorded across the 32 states of Mexico (‘mean(data\_Mexico\_states$freq)’).

This however may mask .BLABLABLA.. need for calculating rates.. BLABLA. We calculate rates of kidnappings per 100,000 residents.

We can download data about the population size for each state from the website of 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"))  
  
#Merge with crime data and calculate crime rates  
data\_Mexico\_states <- data\_Mexico\_states %>%  
 left\_join(population, by = c("ENTIDAD" = "STATE")) %>%  
 mutate(crime\_rate = freq / Population2010 \* 100000)

According to calculated crime rates, the states with the highest rates of kidnappings per capita are Zacatecas and Tamaulipas, both with over 4 kidnappings per 100,000 residents.

Finally, we want to display crime rates in maps.

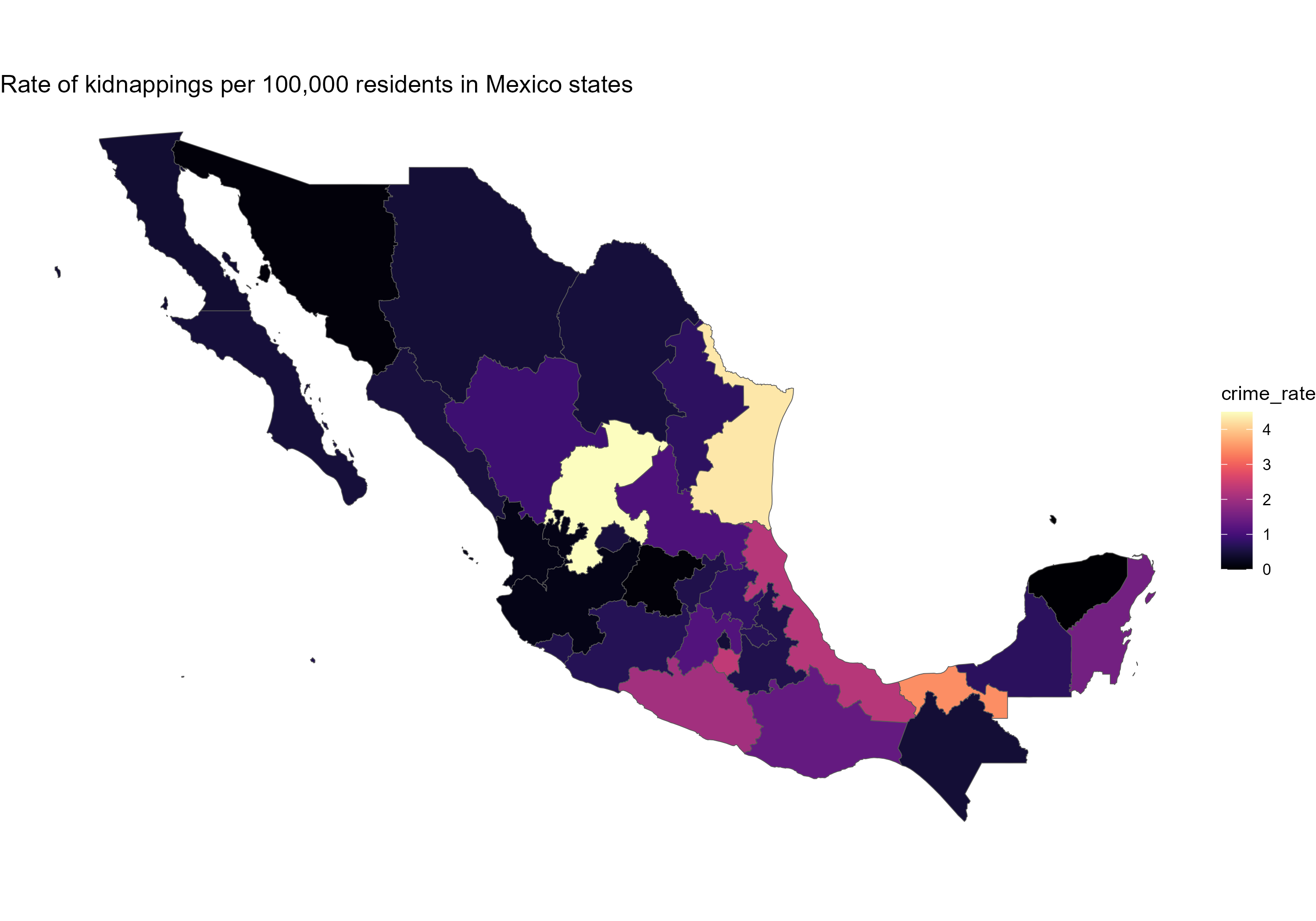
Source of shapefile: <https://github.com/strotgen/mexico-leaflet/>

#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"))

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



XXXX interpret

**References**

<https://link.springer.com/article/10.1007/s12117-012-9185-x> <https://doi.org/10.1080/17440572.2011.632499> <https://doi.org/10.1016/j.ijlcj.2021.100479>