Case Study 1: Descriptive and Time Series Analysis of Homicide in East Asia

David Buil-Gil and Laura Bui

2023-11-14

Add some context in text.

Data from The World Bank Data repository (<https://data.worldbank.org/indicator/VC.IHR.PSRC.P5?most_recent_year_desc=true&locations=>).

We want to compare rates of homicide between East Asia and the rest of the world, and visualise patterns over time.

We begin by loading the required packages in R:

library(here)  
library(dplyr)  
library(tidyr)  
library(stringr)  
library(ggplot2)

We open the dataset of interest, and select East Asian countries (China, Hong Kong, Macau, Japan, Mongolia, North Korea, South Korea, Taiwan):

#Read csv file with data  
data\_homicides <- read.csv(here("data/API\_VC.IHR.PSRC.P5\_DS2\_en\_csv\_v2\_5996865.csv"),  
 skip = 3) #Skip first three rows (no data)  
  
#Make list of countries of interest  
countries\_interest <- c("China", "Hong Kong SAR, China", "Japan",  
 "Korea, Rep.", "Macao SAR, China", "Mongolia",  
 "Korea, Dem. People's Rep.", "World")  
  
#Select countries in the list  
data\_homicides <- data\_homicides %>%  
 filter(Country.Name %in% countries\_interest)

We are only interested in those countries that have complete data between 2000 and 2020. We select columns of interest, tidy up the column names, filter out countries without complete data, and gather columns (years) to rows.

data\_homicides <- data\_homicides %>%  
 #Select columns of interest  
 select(Country.Name, X2000:X2020) %>%  
 #Remove X from column names  
 rename\_with(~str\_replace(., "^X", ""), starts\_with("X")) %>%   
 #Filter out incomplete cases  
 filter(complete.cases(.)) %>%  
 #Gather columns to a rows  
 gather(Year, "Homicide rate", -Country.Name)

We loose Mongolia, China and North Korea.

We may be interested in knowing the average homicide rate across countries, and its variance over the years.

data\_homicides %>%  
 group\_by(Country.Name) %>%  
 summarize(mean = mean(`Homicide rate`),  
 variance = var(`Homicide rate`))

## # A tibble: 5 × 3  
## Country.Name mean variance  
## <chr> <dbl> <dbl>  
## 1 Hong Kong SAR, China 0.524 0.0534  
## 2 Japan 0.394 0.0135  
## 3 Korea, Rep. 0.793 0.0187  
## 4 Macao SAR, China 0.782 0.436   
## 5 World 6.17 0.160

Finally, we may want to display it in a temporal line graph:

ggplot(data\_homicides, aes(x = Year, y = `Homicide rate`,   
 group = Country.Name)) +  
 #Visualise each country with a different color  
 geom\_line(aes(color = Country.Name)) +  
 #Add title  
 ggtitle("Intentional homicide rate (per 100,000 people)") +  
 #Classic graph theme  
 theme\_classic()

