Immigration and Crime: An International Perspective

Data reproduction of the paper Immigration and Crime: An International Perspective by Tseng and Nhundu

Introduction

Choosing a paper

1. Inside the Box: Safety, Health, and Isolation in Prison (Western, 2021)
2. Immigration and Crime: An International Perspective (Marie & Pinotti, 2024)

The first one was written by an author who used R to do most of the work, so the professor suggested that we choose another one.

This presentation is an reproducibility project of looking at two graphs from the paper Immigration and Crime: An International Perspective by the authors Olivier Marie and Paolo Pinotti (Marie & Pinotti, 2024).

Marie, O., & Pinotti, P. (2024). Immigration and Crime: An International Persp
Journal of Economic Perspectives, *38*(1), 181–200.
https://doi.org/10.1257/jep.38.1.181

References

Marie, O., & Pinotti, P. (2024). Immigration and Crime: An International Perspective. *Journal of Economic Perspectives*, *38*(1), 181–200. https://doi.org/10.1257/jep.38.1.181

Western, B. (2021). Inside the Box: Safety, Health, and Isolation in Prison. *Journal of Economic Perspectives*, *35*(4), 97–122. https://doi.org/10.1257/jep.35.4.97

Figure 2a Data

Figure 2

Figure 2 Immigration and Homicides in 55 Countries, 1990–2019

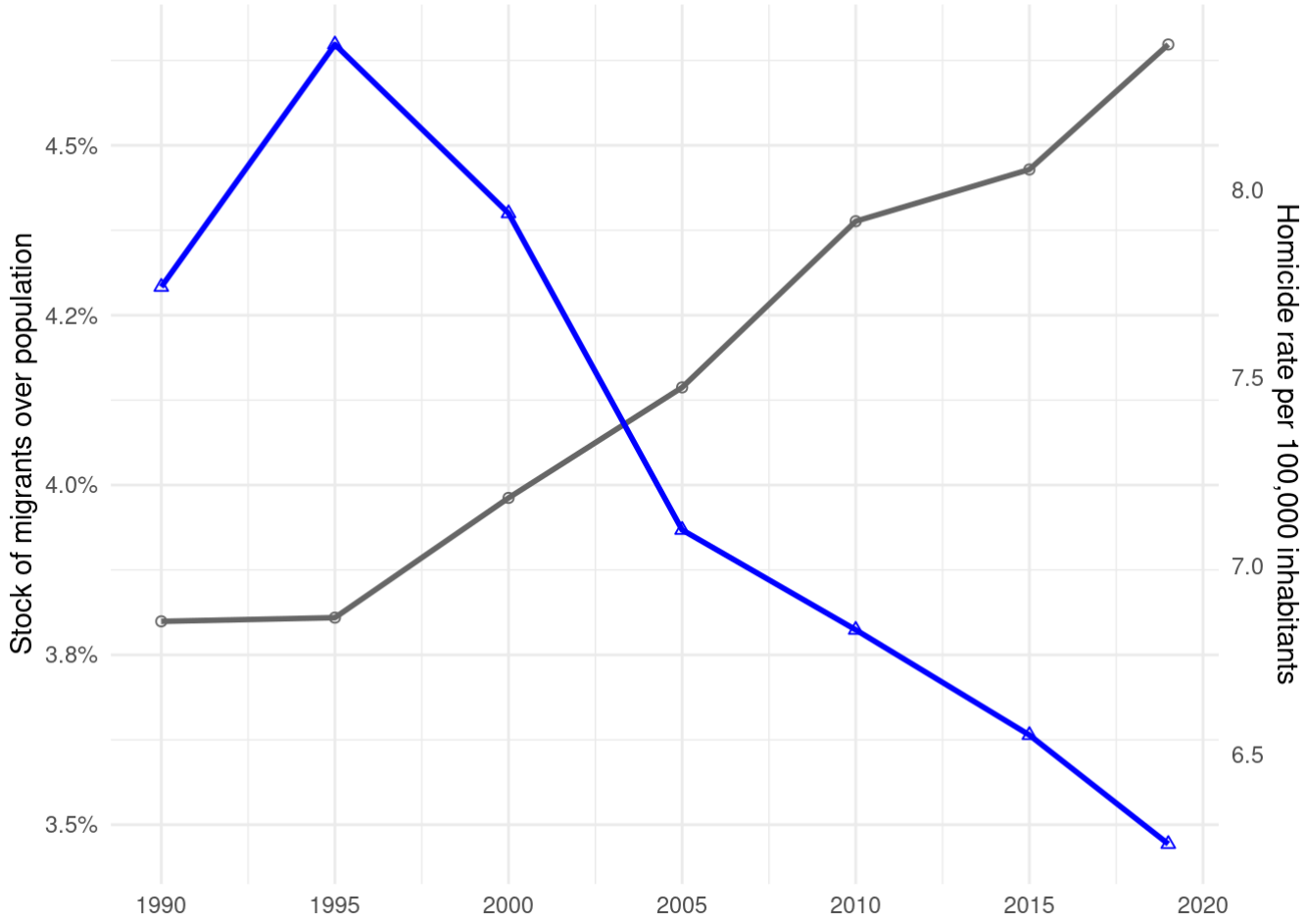
```
iso3c_vec <- c(
  "ARM","AUS","AUT","AZE","BGR","BIH","BLR","BRA","CAN","CHE","COL","CRI",
  "DEU","DNK","ECU","ESP","EST","FIN","FRA","GBR","GEO","GRC","HKG","HND",
  "HRV","IND","IRL","ITA","JAM","JPN","KGZ","KOR","LKA","LTU","MAR","MDA",
  "MEX","MUS","NLD","NOR","PAK","PAN","PHL","POL","PRI","PRT","ROU","RUS",
  "SGP","SVK","SVN","SWE","URY","USA","VEN"
)

# 1) Filter population
wb_pop_ts <- WDI(
    country   = "all",
    indicator = "SP.POP.TOTL",
    start     = 1990,
    end       = 2019,
    extra     = TRUE
  ) %>%
  filter(region != "Aggregates") %>%
  # generate iso3c
  mutate(code = countrycode(iso2c, "iso2c", "iso3c")) %>%
  # Keep the select countries
  filter(code %in% iso3c_vec) %>%
  select(code, year, pop_total = SP.POP.TOTL)

# 2) Combine and calculate weighted indicators
ts_df <- df %>%
  left_join(wb_pop_ts, by = c("code","year")) %>%
  group_by(year) %>%
  summarise(
    migr_w = sum(migr_pop * pop_total, na.rm = TRUE) / sum(pop_total, na.rm = TRUE),
    hom_w  = sum(homicide_rate * pop_total, na.rm = TRUE) / sum(pop_total, na.rm = TRUE)
  )

# 3) Calculate proportions and draw
sf <- max(ts_df$hom_w) / max(ts_df$migr_w)


ggplot(ts_df, aes(x = year)) +
  geom_line(aes(y = migr_w), color="grey40", linewidth=1) +
  geom_point(aes(y = migr_w), color="grey40", shape=21) +
  geom_line(aes(y = hom_w / sf), color="blue", linewidth=1) +
  geom_point(aes(y = hom_w / sf), color="blue", shape=24) +
  scale_x_continuous(breaks = seq(1990, 2020, 5)) +
  scale_y_continuous(
    name      = "Stock of migrants over population",
    labels    = percent_format(accuracy = 0.1),
    sec.axis  = sec_axis(~ . * sf, name = "Homicide rate per 100,000 inhabitants")
  ) +
```

```
  theme_minimal() +
  theme(axis.title.x = element_blank())
```

Figure 2a

Figure 2

Figure 2 Immigration and Homicides in 55 Countries, 1990–2019



Immigration and Crime: An International Perspective

Figure 2b Data

Figure 2

Figure 2 Immigration and Homicides in 55 Countries, 1990–2019

```
iso3c_vec <- c(
  "ARM","AUS","AUT","AZE","BGR","BIH","BLR","BRA","CAN","CHE","COL","CRI",
  "DEU","DNK","ECU","ESP","EST","FIN","FRA","GBR","GEO","GRC","HKG","HND",
  "HRV","IND","IRL","ITA","JAM","JPN","KGZ","KOR","LKA","LTU","MAR","MDA",
  "MEX","MUS","NLD","NOR","PAK","PAN","PHL","POL","PRI","PRT","ROU","RUS",
  "SGP","SVK","SVN","SWE","URY","USA","VEN"
)

# 1) Filter population
wb_pop_ts <- WDI(
    country   = "all",
    indicator = "SP.POP.TOTL",
    start     = 1990,
    end       = 2019,
    extra     = TRUE
  ) %>%
  filter(region != "Aggregates") %>%
  # generate iso3c
  mutate(code = countrycode(iso2c, "iso2c", "iso3c")) %>%
  # Keep the select countries
  filter(code %in% iso3c_vec) %>%
  select(code, year, pop_total = SP.POP.TOTL)

# 2) Combine and calculate weighted indicators
ts_df <- df %>%
  left_join(wb_pop_ts, by = c("code","year")) %>%
  group_by(year) %>%
  summarise(
    migr_w = sum(migr_pop * pop_total, na.rm = TRUE) / sum(pop_total, na.rm = TRUE),
    hom_w  = sum(homicide_rate * pop_total, na.rm = TRUE) / sum(pop_total, na.rm = TRUE)
  )

# 3) Calculate proportions and draw
sf <- max(ts_df$hom_w) / max(ts_df$migr_w)
```

```
df %>%
  filter(code %in% iso3c_vec) %>%
  group_by(year) %>%
  summarise(n_country = n_distinct(code))
```

```
# A tibble: 7 × 2
   year n_country
  <dbl>     <int>
1  1990        55
2  1995        55
3  2000        55
4  2005        55
```

```
5  2010        55
6  2015        55
7  2019        55
```

```
wb_pop_ts %>%
  filter(code=="USA", year %in% c(1990, 2000, 2010, 2019))
```

```
  code year pop_total
1  USA 2019 330226227
2  USA 2010 309378227
3  USA 2000 282162411
4  USA 1990 249623000
```

```
# These pop_total should be officially
#   1990 ~ 248 million
#   2000 ~ 282 million
#   2010 ~ 309 million
#   2019 ~ 328 million
```

(Marie & Pinotti, 2024)

```
# Install the package
library(dplyr)
library(tidyr)
library(ggplot2)
library(ggrepel)
library(WDI)
library(countrycode)

# 1) select the 55 countries
iso3_codes <- c(
  "ARM","AUS","AUT","AZE","BGR","BIH","BLR","BRA","CAN","CHE","COL","CRI",
  "DEU","DNK","ECU","ESP","EST","FIN","FRA","GBR","GEO","GRC","HKG","HND",
  "HRV","IND","IRL","ITA","JAM","JPN","KGZ","KOR","LKA","LTU","MAR","MDA",
  "MEX","MUS","NLD","NOR","PAK","PAN","PHL","POL","PRI","PRT","ROU","RUS",
  "SGP","SVK","SVN","SWE","URY","USA","VEN"
)

# 2) Filter out 1990 & 2019 from df (the complete data frame read_dta() previously)
df2 <- df %>%
  filter(code %in% iso3_codes,             # filter the 55 countries
         year %in% c(1990, 2019)) %>%        # only 1990/2019
  mutate(
    ln_migr  = log(migr_pop),
    ln_homic = log(homicide_rate)
  )
cat("df2 rows:", nrow(df2), "\n")
```

```
df2 rows: 110
```

```
print(head(df2))
```

```
# A tibble: 6 × 9
  country code   year population homicide_rate pop1990 migr_pop ln_migr ln_homic
  <chr>   <chr> <dbl>      <dbl>         <dbl>   <dbl>    <dbl>   <dbl>    <dbl>
1 Armenia ARM    1990      3538.          5.03   3538.   0.186    -1.68   1.62
2 Armenia ARM    2019      2958.          1.69   3538.   0.0643   -2.74   0.525
3 Austra… AUS    2019     25203.          0.89  16961.   0.300    -1.21  -0.117
4 Austra… AUS    1990     16961.          2.21  16961.   0.233    -1.46   0.793
5 Austria AUT    2019      8955.          0.97   7724.   0.199    -1.62  -0.0305
6 Austria AUT    1990      7724.          1.15   7724.   0.103    -2.28   0.140
```

```
# 3) Get the total population from the World Bank
# (first grab all years, then join and then pivot)
wb_pop <- WDI(
  country   = "all",
  indicator = "SP.POP.TOTL",
  start     = 1990,
  end       = 2019,
  extra     = TRUE
) %>%
  filter(region != "Aggregates") %>%
  mutate(code = countrycode(iso2c, "iso2c", "iso3c")) %>%
  select(code, year, pop_total = SP.POP.TOTL)

# 4) Merge, widen, and calculate log-change
df_sc <- df2 %>%
  left_join(wb_pop, by = c("code","year")) %>%
  select(code, year, ln_migr, ln_homic, pop_total) %>%  # ← remove population、homicide_rate、r
  pivot_wider(
    names_from  = year,
    values_from = c(ln_migr, ln_homic, pop_total),
    names_sep   = "_"
  ) %>%
  mutate(
    dln_migr  = ln_migr_2019  - ln_migr_1990,
    dln_homic = ln_homic_2019 - ln_homic_1990
  ) %>%
  filter(!is.na(pop_total_1990))
```

```
# 5) Check the amout of rows of data
cat("Rows to plot:", nrow(df_sc), "\n")
```

```
Rows to plot: 54
```

```
df_tmp <- df2 %>%
  left_join(wb_pop, by = c("code","year")) %>%
  pivot_wider(
    names_from  = year,
    values_from = c(ln_migr, ln_homic, pop_total),
    names_sep   = "_"
  )
print(head(df_tmp))# check the rows after pivot
```

```
# A tibble: 6 × 12
  country   code  population homicide_rate pop1990 migr_pop ln_migr_1990
  <chr>     <chr>      <dbl>         <dbl>   <dbl>    <dbl>        <dbl>
1 Armenia   ARM        3538.          5.03   3538.   0.186         -1.68
2 Armenia   ARM        2958.          1.69   3538.   0.0643        NA
3 Australia AUS       25203.          0.89  16961.   0.300         NA
4 Australia AUS       16961.          2.21  16961.   0.233         -1.46
5 Austria   AUT        8955.          0.97   7724.   0.199         NA
6 Austria   AUT        7724.          1.15   7724.   0.103         -2.28
#   5 more variables: ln_migr_2019 <dbl>, ln_homic_1990 <dbl>,
#   ln_homic_2019 <dbl>, pop_total_1990 <dbl>, pop_total_2019 <dbl>
```

```
cat("Rows after pivot:", nrow(df_tmp), "\n")
```

```
Rows after pivot: 110
```

```
with(df_sc, summary(dln_migr))
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-2.6146 -0.2811  0.2761  0.2960  0.8171  3.1155
```

```
with(df_sc, summary(dln_homic))
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-2.27727 -0.88754 -0.59014 -0.45285 -0.04651  1.37231
```

```
library(ggplot2)
library(ggrepel)

ggplot(df_sc, aes(x = dln_migr, y = dln_homic)) +
  geom_point(shape = 21, fill = NA, color = "black") +
  geom_text_repel(aes(label = code), size = 3) +
  geom_smooth(method = "lm",
              se     = FALSE,
              aes(weight = pop_total_1990)) +
  theme_minimal(base_size = 12) +
  labs(
    title = "Immigration and homicides (pop. weighted)",
    x     = "Log change migration, 1990–2019",
    y     = "Log change homicides, 1990–2019"
  )
```

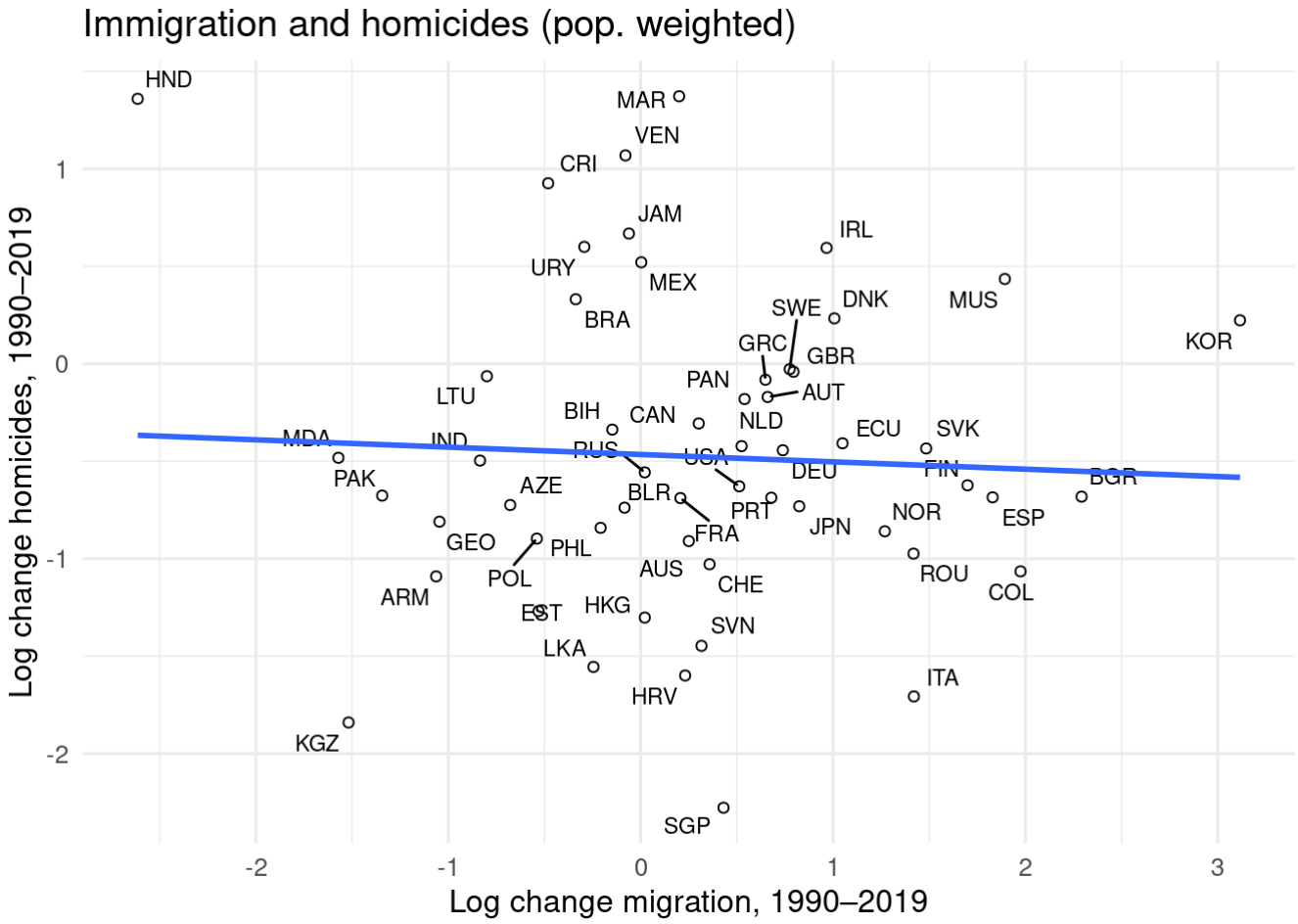
References

Marie, O., & Pinotti, P. (2024). Immigration and Crime: An International Perspective. *Journal of Economic Perspectives*, *38*(1), 181–200. https://doi.org/10.1257/jep.38.1.181

Figure 2b

Figure 2

Figure 2 Immigration and Homicides in 55 Countries, 1990–2019



Immigration and Crime: An International Perspective

Problem and Solution

Problem 1

Figure 2. Problem 1

The gray “immigrant proportion” line is always close to 0. This is because the dual Y axes of ggplot2 are not truly “independently scaled”, but share a set of numerical spaces

Solution:

1. Define a scaling parameter, say sf, to scale the murder rate to the range of 0–0.055: Use max(ts $hom_w)/max(ts$migr_w) to automatically calculate sf sf <- max(ts$hom_w)/max(ts$migr_w)

2. Multiply hom_w by sf in the graph to make it fall in the range of 0–0.055:

p1 <- ggplot(ts, aes(x = year)) + geom_line(aes(y = migr_w), color = “grey40”, linewidth = 1) + geom_point(aes(y = migr_w), color = “grey40”, shape = 21) + geom_line(aes(y = hom_w / sf), color = “blue”, linewidth = 1) + geom_point(aes(y = hom_w / sf),color = “blue”, shape = 24) + scale_y_continuous(name = “Stock of migrants over population”, # Use ^ on the right axis to “reverse” the true murder rate sec.axis = sec_axis(~ . * sf, name = “Homicide rate per 100,000 inhabitants”)) + scale_x_continuous(breaks = seq(1990, 2020, by = 5)) + theme_minimal() + theme(axis.title.x = element_blank()) print(p1) Here we use y = hom_w / sf when drawing the blue line (equivalent to “compressing” the murder rate to the small interval of the proportion of immigrants), Then use sec_axis(~ . * sf) to “expand” the label on the right back to the true murder rate value.

Problem 2

No dots on scatter plot

Figure 3. Problem 2 The problem is in the plot data frame, all rows corresponding to x = dln_migr or y = dln_homic are treated as NA, so ggplot automatically discards them.

Solution

Keep only the columns I need before pivoting

df_sc <- df2 %>% left_join(wb_pop, by = c(“code”,“year”)) %>% select(code, year, ln_migr, ln_homic, pop_total) %>% # ← Throw away population, homicide_rate, and migr_pop pivot_wider(names_from = year, values_from = c(ln_migr, ln_homic, pop_total), names_sep = “_”) %>% mutate(dln_migr = ln_migr_2019 - ln_migr_1990, dln_homic = ln_homic_2019 - ln_homic_1990) %>% filter(!is.na(pop_total_1990))

Problem 3

Unable to center image caption

Original Graph

Solution

Make a styles.css file and typed figure > figcaption {text-align: center;font-size: 0.8em;} and Reference it in YAML header

Figure 4 Data

Immigration and Crime: An International Perspective

Figure 4 Summary of Estimates of the Impact of Immigration on Crime

```
# Load necessary package
# Tidyverse Includes ggplot2, dplyr, etc.
if (!requireNamespace("pacman", quietly = TRUE)) install.packages("pacman")
pacman::p_load(haven, dplyr, tidyr, ggplot2, scales, ggrepel, patchwork, tidyverse, ggh4x)


# Create the data

data_for_plot <- tribble(
  ~Study, ~Type, ~Category, ~Estimate, ~LowerCI, ~UpperCI,
  "BBP, 2012\nItaly", "OLS", "Property", 0.084, 0.02912, 0.13888,
  "BBP, 2012\nItaly", "Shift share", "Property", 0.046, -0.248, 0.34,

  "BFM, 2013\nUnited Kingdom", "OLS", "Property", -0.061, -0.125, 0.003,
  "BFM, 2013\nUnited Kingdom", "Shift share", "Property", -0.386, -0.54476, -0.22724,

  "SPE, 2014\nUnited States", "OLS", "Property", 0.123, 0.00344, 0.24256,
  "SPE, 2014\nUnited States", "Shift share", "Property", 0.108, -0.13112, 0.34712,

  "ADU, 2021\nChile", "OLS", "Property", 0.01, -0.0096, 0.0296,
  "ADU, 2021\nChile", "Shift share", "Property", 0, -0.0784, 0.0784,

  "BBP, 2012\nItaly", "OLS", "Violent", 0.003, -0.16164, 0.16764,
  "BBP, 2012\nItaly", "Shift share", "Violent", -0.036, -0.312556, 0.240556,

  "BFM, 2013\nUnited Kingdom", "OLS", "Violent", -0.007, -0.0364, 0.0224,
  "BFM, 2013\nUnited Kingdom", "Shift share", "Violent", -0.074, -0.26216, 0.11416,

  "SPE, 2014\nUnited States", "OLS", "Violent", 0.065, -0.05848, 0.18848,
  "SPE, 2014\nUnited States", "Shift share", "Violent", 0.01, -0.29184, 0.31184,

  "ADU, 2021\nChile", "OLS", "Violent", 0, -0.0196, 0.0196,
  "ADU, 2021\nChile", "Shift share", "Violent", 0.02, -0.0192, 0.0592
)

# Ensure 'Study' is ordered correctly for plotting
data_for_plot$Study <- factor(data_for_plot$Study,
                              levels = c("ADU, 2021\nChile",
                                         "SPE, 2014\nUnited States",
                                         "BFM, 2013\nUnited Kingdom",
                                         "BBP, 2012\nItaly"))

# Ensure 'Type' is a factor with OLS ordered before Shift share for dodging control
data_for_plot$Type <- factor(data_for_plot$Type,
                             levels = c("OLS", "Shift share"))

# Define a common dodging position to slightly offset the lines vertically
```

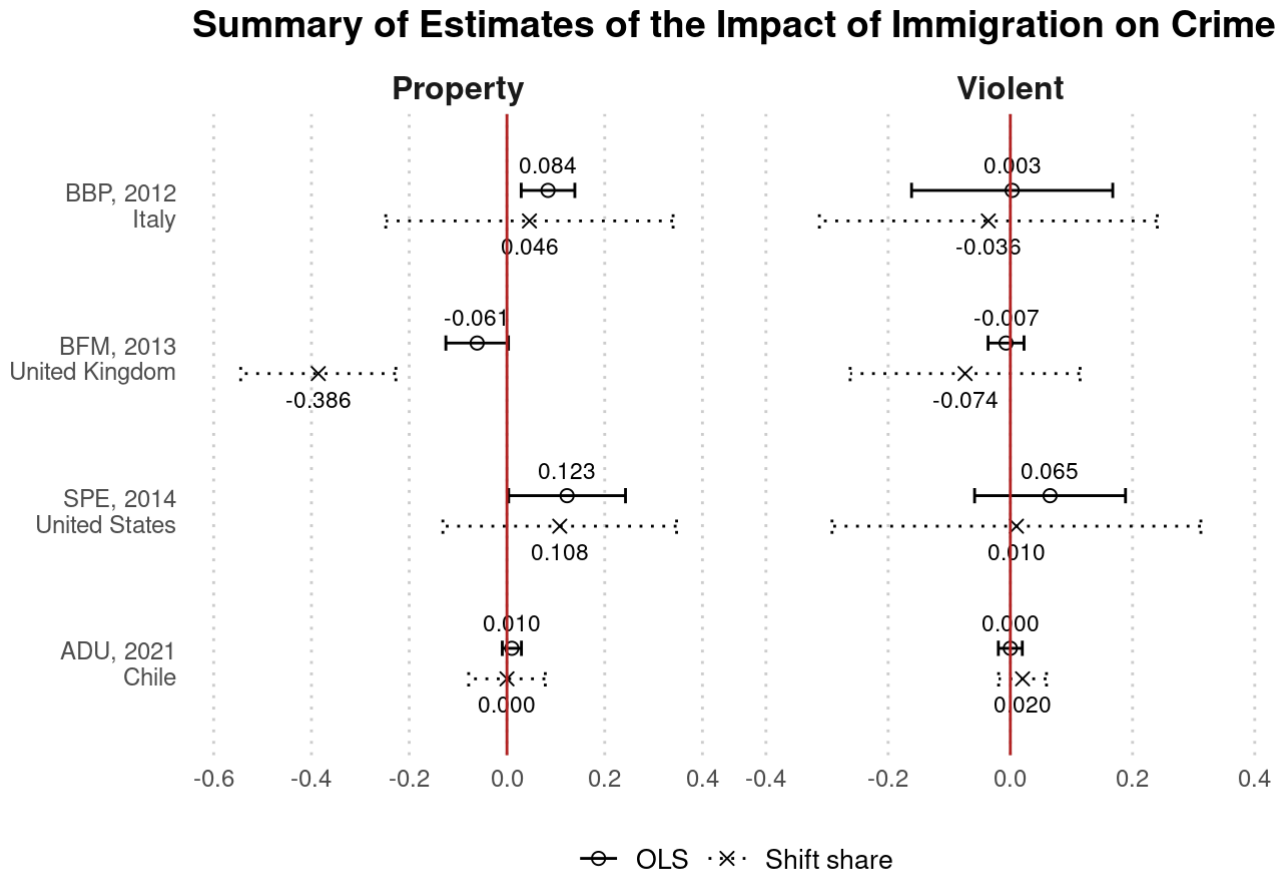
```
dodge_pos <- position_dodge(width = -0.4) # Negative width to put OLS above Shift Share

ggplot(data_for_plot, aes(y = Study, x = Estimate, color = Type)) +
  # Add horizontal error bars for confidence intervals
  geom_errorbarh(aes(xmin = LowerCI, xmax = UpperCI, linetype = Type),
                 height = 0.2, # Controls the vertical thickness of the bar
                 position = dodge_pos, # Use the defined dodging position
                 linewidth = 0.5) +
  # Add points for the estimates
  geom_point(aes(shape = Type),
             size = 2, # Size of the point
             position = dodge_pos) +
  # Add text labels for estimates
  geom_text(aes(label = scales::label_number(accuracy = 0.001, trim = TRUE)(Estimate), # <<--
                group = Type, # Group by Type for dodging
                # Adjust vertical position (vjust) based on Type
                vjust = ifelse(Type == "OLS", -1, 2) # OLS text above, Shift share text below
                ),
            position = dodge_pos, # Apply the same dodge as points/errorbars
            hjust = 0.5, # Center text horizontally on the point
            size = 3, # Font size for the labels
            color = "black") + # Ensure text is black
  # Add a vertical line at x = 0 (no effect)
  geom_vline(xintercept = 0, linetype = "solid", color = "firebrick", linewidth = 0.5) +
  # Facet by Category (Property vs. Violent) with independent x-axes
  facet_wrap2(~ Category, scales = "free_x", axes = "x") +

  # Use facetted_pos_scales from ggh4x to set different x-axis scales per facet
  facetted_pos_scales(
    x = list(
      Category == "Property" ~ scale_x_continuous(
        limits = c(-0.6, 0.4),
        breaks = seq(-0.6, 0.4, by = 0.2),
        labels = scales::number_format(accuracy = 0.1)
      ),
      Category == "Violent" ~ scale_x_continuous(
        limits = c(-0.4, 0.4),
        breaks = seq(-0.4, 0.4, by = 0.2),
        labels = scales::number_format(accuracy = 0.1)
      )
    )
  ) +

  # Customize scales, labels, and theme
  scale_shape_manual(values = c("OLS" = 1, "Shift share" = 4)) +
  scale_color_manual(values = c("OLS" = "black", "Shift share" = "black")) +
  scale_linetype_manual(values = c("OLS" = "solid", "Shift share" = "dotted")) +
  labs(
    title = "Summary of Estimates of the Impact of Immigration on Crime",
    y = NULL,
    x = NULL,
    color = "Type",
    shape = "Type"
  ) +
```

```
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
    axis.text.y = element_text(size = 9),
    axis.text.x = element_text(size = 9),
    legend.position = "bottom",
    legend.title = element_blank(),
    legend.text = element_text(size = 10),
    panel.grid.major.y = element_blank(),
    panel.grid.minor.y = element_blank(),
    panel.grid.major.x = element_line(linetype = "dotted", color = "gray80"),
    panel.grid.minor.x = element_blank(),
    strip.text = element_text(size = 12, face = "bold"),
    plot.margin = unit(c(0.5, 0.5, 0.5, 0.5), "cm")
  )
```

Figure 4

Immigration and Crime: An International Perspective

Figure 4 Summary of Estimates of the Impact of Immigration on Crime



Immigration and Crime: An International Perspective

#Figure 2 Immigration and Homicides in 55 Countries, 1990–2019

$$\Delta Ym,2017-2008=\beta\Delta migrm,2017-2008+\gamma\epsilon mt$$

$$\Delta migrm,\ 2017-2008=\sum n\theta m,2008n\times\Delta lnMIGR2017-2008n$$

$$\theta m,2008n=\sum n'MIGRm,2008n'MIGRm,2008n$$

Some times more complicated graphs are not a good idea to replicate

Ensure that the authors even make it possible for most people to replicate

Immigration and Crime: An International Perspective

1. (Marie & Pinotti, 2024) (Western, 2021) (Bell et al., 2013) (Spenkuch, 2013) (Ajzenman et al., 2023)

References

Ajzenman, N., Dominguez, P., & Undurraga, R. (2023). Immigration, Crime, and Crime (Mis)Perceptions. *American Economic Journal: Applied Economics*, *15*(4), 142–176. https://doi.org/10.1257/app.20210156

Bell, B., Fasani, F., & Machin, S. (2013). Crime and Immigration: Evidence from Large Immigrant Waves. *The Review of Economics and Statistics*, *95*(4), 1278–1290. https://doi.org/10.1162/rest_a_00337

Marie, O., & Pinotti, P. (2024). Immigration and Crime: An International Perspective. *Journal of Economic Perspectives*, *38*(1), 181–200. https://doi.org/10.1257/jep.38.1.181

Spenkuch, J. L. (2013). Understanding the Impact of Immigration on Crime. *American Law and Economics Review*, *16*(1), 177–219. https://doi.org/10.1093/aler/aht017

Western, B. (2021). Inside the Box: Safety, Health, and Isolation in Prison. *Journal of Economic Perspectives*, *35*(4), 97–122. https://doi.org/10.1257/jep.35.4.97