Project proposal

Team 8

Contents

0.1 1. Introduction 1
0.2 2. Data 2
0.3 3. Preliminary results 2
0.4 4. Data analysis plan 3
0.5 Appendix 4

0.1 1. Introduction

Our research question is: How did COVID 19 mortality, unemployment, and lockdown severity affect domestic violence trends during the first wave of the pandemic, across 25 different countries?

Understanding these dynamics requires looking beyond surface indicators to the social effects of crisis in usually hidden spaces. While COVID 19's impact on the economy, healthcare, and institutions is well documented, less attention has been paid to private realms like the home, intimate relationships, and domestic vulnerability.

This analysis explores how large scale crises that confine people at home, such as the pandemic, affect patterns of domestic violence focusing on stressors like mortality, unemployment, and enforced proximity.

This question is difficult to answer, both because such violence is often underreported, and because the available data is fragmented across sources, uneven in quality, and covers different countries and time frames.

We are using four datasets that include monthly domestic violence indicators, mortality rates from COVID 19, unemployment data, and lockdown policy information. Not all datasets contain all countries, but we believe the overlap is sufficient to draw meaningful insights. We have structured our data month by month for the first half of 2020 and aligned it across countries using consistent identifiers and variables. Our approach is to visualize and analyze trends, and then apply statistical models to test relationships between key variables. Specifically, we aim to explore how reported domestic violence changed during the pandemic and whether it is linked to key pandemic related stressors. We will use regression models with non linear and interaction terms, and test variables such as cumulative lockdown duration and lagged unemployment to assess their predictive power. This approach is unique in combining diverse international datasets and applying data science to a sensitive, underexplored issue. We hypothesize that domestic violence increased as lockdowns extended in duration, and that there is a measurable association between violence trends and key pandemic related stressors such as unemployment and mortality.

The potential contribution of this research is in clarifying how public health policy may affect private and often hidden aspects of individual life, beyond commonly studied economic or institutional factors, thereby supporting the development of more intelligent and sensitive responses to future crises.

0.2 2. Data

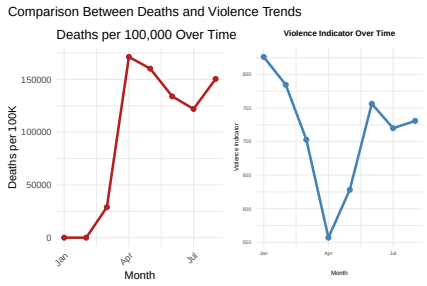
We are using four datasets to explore how the COVID-19 pandemic affected domestic violence trends across 25 countries. Our main dataset contains monthly reports of domestic violence. The other three datasets include lockdown policies, COVID-related deaths, and unemployment rates. Some of our datasets are of different measurements; we are looking to see trends in every country as its own, so the measurement does not matter nor need to be normalized. Also, some of the datasets do not contain all the countries from our main dataset, but because we are working on 25 different countries and looking for trends generally and not specifically by country, it is acceptable to research some factors on a partial dataset of countries as long as we have more than 10-15 countries and their other factors exist and have a difference between them.

A full description of each dataset and the features we are using is provided in the README.md file in the `/data` folder and appears in the appendix of this document.

0.3 3. Preliminary results

In this analysis, we examine the relationship between trends in deaths per 100,000 people and reported violence incidents during the first months of the COVID-19 pandemic.

We plotted two time series graphs side-by-side: One shows the monthly deaths per 100,000 people. The other shows the average violence indicator over the same period.

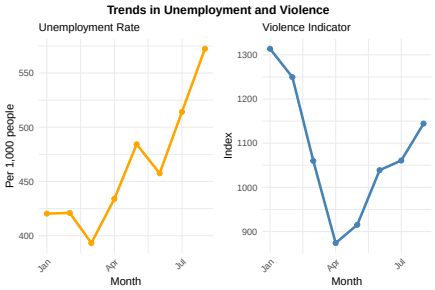


The graphs reveal that during the COVID-19 outbreak peak around April 2020, the number of deaths per 100,000 increased sharply. In contrast, the violence indicator shows a sharp decrease at the exact same time. After April, we observe an inverse pattern: while deaths begin to decline, violence starts to rise again.

These opposite trends suggest a potential negative association between mortality levels and violence during the pandemic — possibly reflecting factors such as lockdown severity, reduced social interaction, or underreporting of violence.

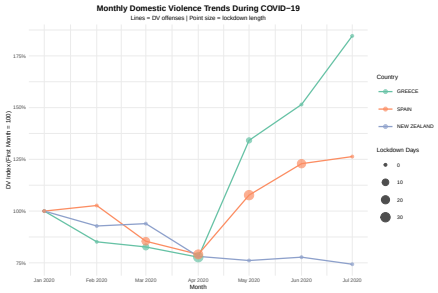
This analysis examines the global trends in unemployment per 1000 people and violence incidents during the early phase of the COVID-19 pandemic.

We present two side-by-side time series plots. The left plot shows the average global unemployment rate per 1000 people. The right plot shows the average global violence indicator over time.



Although the relationship between the two variables is less clearly defined compared to previous analyses, we still observe a notable shift around April 2020, when the COVID-19 pandemic began to peak globally. During this time the unemployment rate begins a steady increase that continues in the following months. In contrast, the violence indicator reaches its lowest point in April, followed by a gradual upward trend.

This simultaneous change suggests that the early stages of the pandemic may have impacted both economic and social stability. While the trends are not perfectly inversely correlated, the timing of their shifts hints at a possible indirect connection between rising unemployment and increases in violence.



In this graph we compare three countries that handled COVID-19 in a different way through national lockdowns: Spain which had serious and long lockdowns; Greece which had partial and short lockdowns; and New Zealand, which had no lockdowns during this time. Although it is only three countries, we can see the domestic violence rate trends are very different, and we would like to continue our research along with the other factors to fully understand what might cause the rise or fall of domestic violence during the beginning of the pandemic.

0.4 4. Data analysis plan

Our data analysis plan implements a multi method approach to examine how pandemic factors affected domestic violence rates across diverse global contexts. We will analyze temporal patterns in violence reporting alongside COVID-19 mortality, lockdown policies, and economic indicators to identify causal relationships and moderating factors. Response (Y) Variable: Violence indicator: Our primary outcome measure representing reported domestic violence incidents Explanatory (X) Variables:

•COVID mortality rates: Monthly deaths per 100,000 people
•Lockdown severity: Categorized levels of government restrictions
•Unemployment rates: Monthly unemployment figures per 1,000 people

Comparison Groups We will group countries across several dimensions: Geographic:

•Continental regions: Countries grouped by continental location
(Europe, Americas, Asia, Africa, Oceania)

Policy Response: • Lockdown severity (none/partial/full)

Temporal: • Pandemic phases (pre-pandemic, intermediate, second wave) • Seasonal periods to control for seasonal variations

Impact-Based: • Mortality impact: Countries with high/medium/low COVID-19 death rates Methods Data Visualization Using time series plots to visualize trends in violence reporting during different pandemic phases. This will help us identify patterns and test our hypothesis visually before conducting statistical analysis.

-Needed results: Time series plots showing violence increases as lockdowns extended in duration across different countries and regions Multiple Linear Regression Analyzing how lockdown duration, unemployment,

and mortality rates simultaneously affected violence indicators while controlling for multiple factors. This directly tests our hypothesis about associations between pandemic stressors and violence.

•Needed results: Significant positive coefficients for lockdown duration and unemployment variables, confirming the hypothesized associations between these factors and violence increases

Feature Engineering Creating derived variables including cumulative lockdown duration and percent changes from pre-pandemic baselines to better capture the effects of extended lockdowns on violence patterns.

•Needed results: Improved model performance when using cumulative lockdown measures, supporting the hypothesis that violence effects accumulate over lockdown duration

Model Evaluation

Using adjusted R-squared and residual analysis to assess model fit and check assumptions about our linear relationships. • Needed results: Adjusted R^2 values above 0.3 and normally distributed residuals confirming model validity and explanatory power

0.5 Appendix

0.5.1 Data README

```
---
output:
  pdf_document: default
  html_document: default
---
# Data Dictionary

## 1. covid_monthly_domesticviolence.xlsx
**Description**: this is a monthly count of domestic violence cases across 34 countries
during the beginning of COVID-19

**Columns:**
- ‘Region‘: self explenatory
- ‘Sub region‘: not-used
- ‘Country‘: country indicator
- ‘Indicator‘: the domestic violence indicator - sexual/physical, female/male victim
- ‘Oct_2019‘: monthly indicator (has numeric count per country per indicator)
- ‘Nov_2019‘: monthly indicator (has numeric count per country per indicator)
- ‘Dec_2019‘: monthly indicator (has numeric count per country per indicator)
- ‘Jan_2020‘: monthly indicator (has numeric count per country per indicator)
- ‘Feb_2020‘: monthly indicator (has numeric count per country per indicator)
- ‘Mar_2020‘: monthly indicator (has numeric count per country per indicator)
- ‘Apr_2020‘: monthly indicator (has numeric count per country per indicator)
- ‘May_2020‘: monthly indicator (has numeric count per country per indicator)
- ‘Jun_2020‘: monthly indicator (has numeric count per country per indicator)
- ‘Jul_2020‘: monthly indicator (has numeric count per country per indicator)
- ‘Aug_2020‘: monthly indicator (has numeric count per country per indicator)

**Structure (‘glimpse‘)**:
Rows: 129
```

```
Columns: 15
$ Region       <chr> "Europe", "Americas", "Europe", "Europe", "Europe", "Europe", "Americas", "Amer...
$ `Sub region` <chr> "Southern Europe", "Latin America and the Caribbean", "Southern Europe", "South...
$ Country      <chr> "ALBANIA", "ANTIGUA AND BARBUDA", "BOSNIA AND HERZEGOVINA", "BOSNIA AND HERZEGO...
$ Indicator    <chr> "Sexual violence or physical assault by IPFM*  (domestic violence): Total numbe...
$ Oct_2019     <dbl> 112, 5, 109, 30, 3, 27, 9903, 1224, 3888, 0, 1, 0, 1, NA, NA, NA, NA, NA, 176, ...
$ Nov_2019     <dbl> 81, 5, 104, 32, 2, 30, 9886, 1193, 3983, 0, 1, 0, 0, NA, NA, NA, NA, NA, 155, 2...
$ Dec_2019     <dbl> 113, 1, 131, 34, 3, 31, 10883, 1304, 4224, 0, 1, 0, 0, NA, NA, NA, NA, NA, 184,...
$ Jan_2020     <dbl> 111, 6, 73, 38, 5, 33, 11178, 1222, 4294, 0, 2, 0, 0, NA, NA, NA, NA, NA, 190, ...
$ Feb_2020     <dbl> 90, 6, 95, 57, 2, 55, 9899, 1124, 3680, 0, 0, 0, 2, NA, NA, NA, NA, NA, 222, 5,...
$ Mar_2020     <dbl> 90, 7, 88, 49, 5, 44, 9578, 1035, 3626, 0, 1, 0, 2, NA, NA, NA, NA, NA, 209, 1,...
$ Apr_2020     <dbl> 96, NA, 86, 40, 5, 35, 7776, 878, 2889, 0, 1, 0, 0, 7617, 0, 9, 1716, 538, 190,...
$ May_2020     <dbl> 127, NA, NA, NA, NA, NA, NA, NA, NA, 0, 0, 0, 0, 9674, 0, 3, 2216, 575, 222, 0,...
$ Jun_2020     <dbl> 159, NA, NA, NA, NA, NA, NA, NA, NA, 0, 3, 0, 1, 7690, 0, 5, 1703, 443, 292, 0,...
$ Jul_2020     <dbl> 158, NA, NA, NA, NA, NA, NA, NA, NA, 0, 2, 0, 0, 8167, 1, 1, 1985, 512, 250, 0,...
$ Aug_2020     <dbl> 174, NA, NA, NA, NA, NA, NA, NA, NA, 0, 5, 1, 2, 7289, 0, 2, 1684, 296, 295, 0,...
```

Summary Statistics (`summary`):

```
    Region           Sub region          Country           Indicator            Oct_2019      
 Length:129         Length:129         Length:129         Length:129         Min.   :   0.0  
 Class :character   Class :character   Class :character   Class :character   1st Qu.:   3.0  
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :  31.0  
                                                                             Mean   : 554.7  
                                                                             3rd Qu.: 251.0  
                                                                             Max.   :9903.0  
                                                                             NA's   :24      
    Nov_2019          Dec_2019          Jan_2020          Feb_2020          Mar_2020      
 Min.   :    0.0   Min.   :    0.0   Min.   :    0.00   Min.   :    0.0   Min.   :   0.0  
 1st Qu.:    3.0   1st Qu.:    3.0   1st Qu.:    2.75   1st Qu.:    2.0   1st Qu.:   3.0  
 Median :   32.0   Median :   49.5   Median :   34.00   Median :   35.0   Median :  33.0  
 Mean   :  515.5   Mean   :  585.0   Mean   :  568.65   Mean   :  540.0   Mean   : 467.5  
 3rd Qu.:  269.0   3rd Qu.:  372.2   3rd Qu.:  307.75   3rd Qu.:  288.2   3rd Qu.: 218.0  
 Max.   :10508.0   Max.   :11642.0   Max.   :13018.00   Max.   :12409.0   Max.   :9750.0  
 NA's   :16        NA's   :15        NA's   :17         NA's   :17        NA's   :12      
    Apr_2020          May_2020          Jun_2020          Jul_2020          Aug_2020      
 Min.   :   0.0   Min.   :   0.00   Min.   :   0.0   Min.   :   0.0   Min.   :   0.00  
 1st Qu.:   2.0   1st Qu.:   3.25   1st Qu.:   5.0   1st Qu.:   2.0   1st Qu.:   2.75  
 Median :  35.0   Median :  51.00   Median :  54.0   Median :  49.0   Median :  29.00  
 Mean   : 444.7   Mean   : 572.08   Mean   : 610.2   Mean   : 611.7   Mean   : 604.20  
 3rd Qu.: 216.5   3rd Qu.: 360.75   3rd Qu.: 331.0   3rd Qu.: 342.2   3rd Qu.: 295.25  
 Max.   :7776.0   Max.   :9674.00   Max.   :9122.0   Max.   :9377.0   Max.   :9056.00  
 NA's   :11       NA's   :43        NA's   :40       NA's   :43       NA's   :53       
```

---

2. death_rates.csv

Description: a 3 column dataset that will tell us the daily and by that monthly deaths from COVID-19 for 2020 and on if we need it

Columns:

- `Entity`: the country indicator
- `Day`: daily date from the start of 2020
- `Daily new confirmed deaths due to COVID-19 per million people(rolling 7-day average, right-aligned)`

a 7 day rolling avarage of new deaths from COVID-19

Structure (‘glimpse‘):
Rows: 480,085
Columns: 3
$ Entity
$ Day
$ ‘Daily new confirmed deaths due to COVID-19 per million people (rolling 7-day average, right-aligned)

Summary Statistics (‘summary‘):
```
    Entity              Day
 Length:480085      Length:480085
 Class :character   Class :character
 Mode  :character   Mode  :character



 Daily new confirmed deaths due to COVID-19 per million people (rolling 7-day average, right-aligned)
 Min.   :  0.0000
 1st Qu.:  0.0000
 Median :  0.0000
 Mean   :  0.6629
 3rd Qu.:  0.2464
 Max.   :129.2137
```

---

3. coronanet_release_Croatia.csv
Description: an example dataset of the policies in this country,
we will take the lockdowns and interpert from that,
we have a dataset like this for each of our countries (we are explaining only some columns)

Columns:
- ‘description‘: the full description of the policy/lockdown for our use and filtering
- ‘date_announced‘: when was it announced
- ‘date_start‘: when did it start
- ‘date_end‘: when did it end
- ‘date_end_spec‘: not used or important
- ‘country‘: country indicator
- ‘init_country_level‘: is it national, regional or provencial
- ‘domestic_policy‘: is it domestic or worldwide
- ‘type‘: specified if its lockdown or something else
- ‘target_who_what‘: is it for all residents or just visitors or based on age
- ‘compliance‘: mandatory/voluntary

Structure (‘glimpse‘):
Rows: 1,076
Columns: 63
```
$ description            <chr> "September 14, 2020: With Bulletin N. 262, the Italian Ministry of He...
$ date_announced         <date> 2020-09-14, 2020-04-07, 2020-02-25, 2020-03-11, 2020-03-11, 2020-03-...
$ date_start             <date> 2020-09-14, 2020-04-07, 2020-02-25, 2020-03-13, 2020-03-13, 2020-03-...
$ date_end               <date> 2020-09-15, NA, 2020-02-25, 2020-03-30, 2020-03-30, 2020-03-30, 2020...
$ date_end_spec          <chr> "The policy has a clear end date", "The policy's end date is unknown ...
```

```
$ country                <chr> "Albania,Andorra,Armenia,Austria,Azerbaijan,Belarus,Belgium,Bosnia an...
$ init_country_level     <chr> "National", "National", "National", "National", "National", "National...
$ domestic_policy        <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
$ type                   <chr> "New Task Force, Bureau or Administrative Configuration", "New Task F...
$ target_who_what        <chr> "All Residents (Citizen Residents + Foreign Residents)", "All Residen...
$ compliance             <chr> "Mandatory (Unspecified/Implied)", "Mandatory (Unspecified/Implied)",...
```

---

4. UNE_TUNE_SEX_AGE_NB_M-filtered-2025-05-18.csv
Description: an unemployment monthly count by country and age ranges

Columns:
- ‘ref_area.label‘: country indicator
- ‘source.label‘: the source from which the data came from
- ‘indicator.label‘: the indicator which is usually unemployment but by what count
- ‘sex.label‘: sex
- ‘classif1.label‘: age range or other classifiers
- ‘time‘: month and year
- ‘obs_value‘: count value by the count on the indicator
- ‘obs_status.label‘: wether the count is reliable or not (mostly reliable)
- ‘note_classif.label‘: notes on the classifier
- ‘note_indicator.label‘: notes on the indicator
- ‘note_source.label‘: notes on the source

Structure (‘glimpse‘):
Rows: 7,970
Columns: 11

```
$ ref_area.label       <chr> "Australia", "Australia", "Australia", "Australia", "Australia", "Austr...
$ source.label         <chr> "LFS - Labour Force Survey", "LFS - Labour Force Survey", "LFS - Labour...
$ indicator.label      <chr> "Unemployment by sex and age (thousands)", "Unemployment by sex and age...
$ sex.label            <chr> "Total", "Total", "Total", "Total", "Total", "Total", "Total", "Total",...
$ classif1.label       <chr> "Age (Youth, adults): 15+", "Age (Youth, adults): 15-64", "Age (Youth, ...
$ time                 <chr> "2020M12", "2020M12", "2020M12", "2020M12", "2020M12", "2020M12", "2020...
$ obs_value            <dbl> 859.814, 843.707, 301.252, 558.562, 859.814, 301.252, 432.529, 109.927,...
$ obs_status.label     <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
$ note_classif.label   <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
$ note_indicator.label <chr> "Frequency: Monthly", "Frequency: Monthly", "Frequency: Monthly", "Freq...
$ note_source.label    <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,...
```

0.5.2 Source code

```
knitr::opts_chunk$set(
  echo = FALSE,
  warning = FALSE,
  message = FALSE,
  fig.width = 6,
  fig.height = 4,
  out.width = "35%"
```

```
)
library(tidyverse)
library(broom)
library(htmltools)
library(readr)
library(dplyr)
library(ggplot2)
library(patchwork)


df <- read_csv("merged_deaths_violence_final.csv")


df_summary <- df %>%
  mutate(year_month = as.Date(paste0(year_month, "-01"))) %>%
  group_by(year_month) %>%
  summarise(
    avg_deaths = mean(monthly_avg_deaths, na.rm = TRUE),
    avg_violence = mean(violence_indicator, na.rm = TRUE)
  ) %>%
  mutate(deaths_per_100k = avg_deaths * 100000)


p1 <- ggplot(df_summary, aes(x = year_month, y = deaths_per_100k, group = 1)) +
  geom_line(color = "firebrick", size = 1.2) +
  geom_point(color = "firebrick", size = 2) +
  labs(title = "Deaths per 100,000 Over Time",
       x = "Month", y = "Deaths per 100K") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))


p2 <- ggplot(df_summary, aes(x = year_month, y = avg_violence, group = 1)) +
  geom_line(color = "steelblue", size = 1.2) +
  geom_point(color = "steelblue", size = 2) +
  labs(title = "Violence Indicator Over Time",
       x = "Month", y = "Violence Indicator") +
  theme_minimal() +
   theme(
  legend.title = element_text(size = 6),
  legend.text = element_text(size = 5),
  axis.text.x = element_text(size = 5),
  axis.text.y = element_text(size = 5),
  axis.title.x = element_text(size = 6),
  axis.title.y = element_text(size = 6),
  plot.title = element_text(size = 8, face = "bold", hjust = 0.5),
  plot.subtitle = element_text(size = 6, hjust = 0.5)
)

(p1 + p2) +
  plot_annotation(title = "Comparison Between Deaths and Violence Trends")
```

```r
library(dplyr)
library(readr)
library(stringr)


merge <- read_csv("merged_deaths_violence_final.csv", show_col_types = FALSE) %>%
  mutate(Entity = str_to_lower(Entity),
         year_month = as.character(year_month))

unemp <- read_csv("UNE_TUNE_SEX_AGE_NB_M-filtered-2025-05-18.csv", show_col_types = FALSE) %>%
  mutate(
    Entity = str_to_lower(`ref_area.label`),
    year_month = str_replace(time, "M", "-")
  )

unemp_summary <- unemp %>%
  group_by(Entity, year_month) %>%
  summarise(unemployment_rate = mean(obs_value, na.rm = TRUE), .groups = "drop")


merged_final <- merge %>%
  left_join(unemp_summary, by = c("Entity", "year_month"))

write_csv(merged_final, "merged_with_unemployment.csv")

library(dplyr)
library(readr)
library(ggplot2)
library(gridExtra)
library(grid)


df <- read_csv("merged_with_unemployment.csv")


df_filtered <- df %>% filter(!is.na(unemployment_rate))


monthly_avg <- df_filtered %>%
  group_by(year_month) %>%
  summarise(
    avg_unemployment = mean(unemployment_rate, na.rm = TRUE),
    avg_violence = mean(violence_indicator, na.rm = TRUE),
    .groups = "drop"
  )


monthly_avg$year_month <- as.Date(paste0(monthly_avg$year_month, "-01"))


p1 <- ggplot(monthly_avg, aes(x = year_month, y = avg_unemployment)) +
  geom_line(color = "orange", size = 1.2) +
  geom_point(color = "orange", size = 2) +
```

```
  labs(title = "Unemployment Rate",
       x = "Month", y = "Per 1,000 people") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(size = 11)
  )


p2 <- ggplot(monthly_avg, aes(x = year_month, y = avg_violence)) +
  geom_line(color = "steelblue", size = 1.2) +
  geom_point(color = "steelblue", size = 2) +
  labs(title = "Violence Indicator",
       x = "Month", y = "Index") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(size = 11)
  )


grid.arrange(
  p1, p2,
  ncol = 2,
  top = textGrob("Trends in Unemployment and Violence", gp = gpar(fontsize = 12, fontface = "bold"))
)



library(scales)
library(readr)
library(dplyr)
library(ggplot2)
library(lubridate)
library(forcats)

# === Load and prepare data ===
dv <- read_csv("merged_with_unemployment.csv") %>%
  rename(country = Entity) %>%
  mutate(month = as.Date(paste0(year_month, "-01"))) %>%
  filter(Indicator == "Sexual violence or physical assault by IPFM*  (domestic violence): Total number o
  mutate(country = toupper(country))

lockdowns <- read_csv("greece_spain_newzealand_lockdowns.csv") %>%
  mutate(country = toupper(country),
         month = as.Date(paste0(year_month, "-01")))

selected_countries <- c("GREECE", "SPAIN", "NEW ZEALAND")

# Merge & clean
merged <- dv %>%
  filter(country %in% selected_countries) %>%
  left_join(lockdowns, by = c("country", "month")) %>%
```

```r
  mutate(lockdown_length = replace_na(lockdown_length, 0)) %>%
  filter(format(month, "%Y-%m") != "2020-08")


# Clean and structure data
library(RColorBrewer)

# Assuming 'merged' is your combined dataset
plot_df <- merged %>%
  filter(format(month, "%Y-%m") != "2020-08") %>%
  mutate(
    country = factor(country, levels = c("GREECE", "SPAIN", "NEW ZEALAND")),
    month = as.Date(month)
  ) %>%
  group_by(country) %>%
  arrange(month) %>%
  mutate(
    dv_index = violence_indicator / first(violence_indicator) * 100
  ) %>%
  ungroup()

# Define a colorblind-friendly palette
cb_palette <- brewer.pal(n = 3, name = "Set2")

ggplot(plot_df, aes(x = month, y = dv_index, color = country)) +
  geom_line(size = 0.6) +
  geom_point(aes(size = lockdown_length), alpha = 0.7) +
  scale_color_manual(values = cb_palette) +
  scale_size_continuous(range = c(1, 4), name = "Lockdown Days") +
  scale_x_date(date_labels = "%b %Y", breaks = "1 month") +
  scale_y_continuous(labels = label_number(suffix = "%")) +
  labs(
    title = "Monthly Domestic Violence Trends During COVID-19",
    subtitle = "Lines = DV offenses | Point size = lockdown length",
    x = "Month",
    y = "DV Index (First Month = 100)",
    color = "Country"
  ) +
  theme_minimal(base_size = 6) +
  theme(
  legend.title = element_text(size = 6),
  legend.text = element_text(size = 5),
  axis.text.x = element_text(size = 5),
  axis.text.y = element_text(size = 5),
  axis.title.x = element_text(size = 6),
  axis.title.y = element_text(size = 6),
  plot.title = element_text(size = 8, face = "bold", hjust = 0.5),
  plot.subtitle = element_text(size = 6, hjust = 0.5)
)
```

```r
cat(readLines('../data/README.md'), sep = '\n')
```