Technical Appendix

Integrating Political Strike Data with V-Dem and WhoGov

Author Contact

Rodrigo M. Medel
Universidad de Chile
rodrigo.medel.sierralta@gmail.com

1. Overview of the Datasets

This function was developed to facilitate the integration of **event-level data on political strikes** with **country-year-level variables** from the **Varieties of Democracy (V-Dem)** and **WhoGov** datasets.

- **Political Strikes Dataset**: A manually curated event-level dataset covering political strikes in 18 Latin American countries between 1990 and 2020, coded by type (local, sectoral, general).
  See: https://doi.org/10.7910/DVN/YZRT6V
- **V-Dem Dataset**: Provides detailed annual measures of democracy, institutions, and political variables for all countries.
  See: https://www.v-dem.net
- **WhoGov Dataset**: Contains data on the composition and duration of governments, including party affiliation and ideology of prime ministers.
  See: https://politicscentre.nuffield.ox.ac.uk/whogov-dataset/

The challenge addressed by this function is the **aggregation of event-level data** to match the **country-year unit of analysis** used in the V-Dem and WhoGov databases.

2. Objective of the Function

The purpose of the `link_strike_types_to_vdem_whogov()` function is to:

- Aggregate political strikes by **type** (local, sectoral, general) and **count them per country and year**
- Generate a **balanced country-year panel** (1990–2020) including zeros for years without strikes
- Merge this panel with corresponding V-Dem and WhoGov variables
- Produce a unified dataset for research on the relationship between contentious politics and political institutions

3. R Function Code

```
link_strike_types_to_vdem_whogov <- function(strikes_data, vdem_data, whogov_data) {

    # Step 1: Define country lookup table
    country_lookup <- c(`1` = "Argentina", `2` = "Bolivia", `3` = "Brazil", `4` = "Chile",
        `5` = "Colombia", `6` = "Costa Rica", `7` = "Ecuador", `8` = "El Salvador",
        `9` = "Guatemala", `10` = "Honduras", `11` = "Mexico", `12` = "Nicaragua",
        `13` = "Panama", `14` = "Paraguay", `15` = "Peru", `16` = "Dominican Republic",
        `17` = "Uruguay", `18` = "Venezuela")

    # Step 2: Prepare the strikes dataset
```

```
    strikes_data <- strikes_data |>
        dplyr::mutate(country_name = dplyr::recode(as.character(pais), !!!country_lookup),
            year = as.integer(yrinicio), type = as.character(type)) |>
        dplyr::filter(!is.na(type), year >= 1990, year <= 2020)

    # Step 3: Generate full country-year panel (1990-2020)
    countries <- unique(strikes_data$country_name)
    years <- 1990:2020
    full_panel <- expand.grid(country_name = countries, year = years) |>
        dplyr::arrange(country_name, year)

    # Step 4: Count strikes per type by country and year
    strike_counts <- strikes_data |>
        dplyr::group_by(country_name, year, type) |>
        dplyr::summarise(n = dplyr::n(), .groups = "drop") |>
        tidyr::pivot_wider(names_from = type, values_from = n, values_fill = list(n = 0)) |>
        dplyr::rename_with(.fn = ~paste0("n_", gsub(" ", "_", tolower(.x)), "_strikes"),
            .cols = -c(country_name, year))

    # Step 5: Join full panel with strike counts
    panel_with_strikes <- dplyr::left_join(full_panel, strike_counts, by = c("country_name",
        "year")) |>
        dplyr::mutate(dplyr::across(dplyr::starts_with("n_"), ~tidyr::replace_na(.x,
            0)))

    # Step 6: Ensure year variables in V-Dem and WhoGov are integers
    vdem_data$year <- as.integer(vdem_data$year)
    whogov_data$year <- as.integer(whogov_data$year)

    # Step 7: Merge with V-Dem and WhoGov datasets
    merged_vdem <- dplyr::left_join(panel_with_strikes, vdem_data, by = c("country_name",
        "year"))
    merged_all <- dplyr::left_join(merged_vdem, whogov_data, by = c("country_name",
        "year"))

    return(merged_all)
}
```

4. User Guide

Inputs

- `strikes_data`: The Political Strikes Dataset with at least the variables `pais`, `yrinicio`, and `type`
- `vdem_data`: The V-Dem Dataset with `country_name` and `year`
- `whogov_data`: The WhoGov Dataset with `country_name` and `year`

Output

A merged dataset with:

- Counts of **local**, **sectoral**, and **general** strikes per country-year
- V-Dem institutional indicators (e.g., `v2x_polyarchy`)
- WhoGov government indicators (e.g., `gov1_party`, `gov1_ideology`)

Usage Example

```
linked_data <- link_strike_types_to_vdem_whogov(strikes_data, vdem_data, whogov_data)
```

The function facilitates comparative analysis of how institutional and government-level characteristics interact with patterns of contentious politics.