SPEC REU R Resources: Data Management 3 – Group Work

Alix Ziff, Miriam Barnum, Claudia Salas Gimenez and Ben Graham

February 2025

For this assignment, we will continue working with the IDC Powersharing dataset to analyze how different types of political powersharing—constraining, dispersive, and inclusive—vary across regions. This exercise will get us practicing using the countrycode package to map countries to region and reinforce the importance of data aggregation and how it helps reveal meaningful patterns in large datasets.

By the end of this groupwork, you will further develop your ability to approach research projects, where you have a specific end goal and must manipulate the available data to achieve it. You will also continue practicing data aggregation to analyze regional trends over time and create visualizations that effectively communicate insights from the dataset.

Initial Setup

To begin, set the appropriate working directory, load the necessary libraries, and import the `IDC_analysis_master_MB_20210414.rds` dataset.

For reference, the IDC Powersharing dataset was introduced in the research paper "*De Jure Powersharing 1975–2019: Updating the Inclusion, Dispersion, and Constraints Dataset"* by Ziff, Barnum, Abadeer, Chu, Jao, Zaragoza, and Graham (2024).

You can find the research paper and the online appendix here.

```r
# Set working directory
#setwd("YourFolderPath")

# Load required libraries
library(dplyr)
library(tidyr)
library(ggplot2)
library(countrycode)

# Load the data
idc_controls <- readRDS("IDC_analysis_master_MB_20210414.rds")
```

How powersharing mechanisms differ across regions?

How do different forms of powersharing—constraining, dispersive, and inclusive—vary across regions, and how have these patterns changed over time? Rather than focusing on a single indicator, in this groupwork assignment, we will examine these three dimensions of political powersharing to explore how political systems allocate power and ensure representation across the world.

Let's try to break down the analysis process. Since our goal is to compare these powersharing mechanisms across regions and over time, we need to:

1. We need to aggregate the data to compute regional averages for each of the three powersharing indicators. Since we currently don't have a `region` variable, we must first assign countries to regions to ensure meaningful comparisons.
2. To visualize and compare patterns, we can use a grouped bar plot with facets. Bar plots help compare different values across categories, such as regions, and faceting the plot by powersharing type allow us to compare constraining, dispersive, and inclusive powersharing side by side.
3. Instead of looking at how these patterns have changed over time (e.g., by decade), we will concentrate on the most recent year available in the dataset to get a snapshot of current regional differences.

By the end of this assignment, you will have a clear visualization of how different powersharing mechanisms vary regionally and how they have evolved over time.

Exercise 1: Mapping Countries to Regions

First, we need to create a `region` variable, and we will use the `countrycode` package to assign each country to a region using country.

So far, whenever we have aggregated data or selected specific countries, we have relied on country names. While this approach works within a single dataset, it becomes problematic when integrating data from multiple sources. Some common issues include:

- Some datasets use uppercase ("United States"), while others use lowercase ("united states").
- Differences in spelling, such as "Côte d'Ivoire" vs. "Ivory Coast".
- Some countries have changed names over time (e.g., "Burma" vs. "Myanmar").
- Some datasets use abbreviations like "USA", while others spell out "United States of America".

To avoid these issues, we'll use Gleditsch and Ward numeric country codes (`gwno`) to map counties to regions, which uniquely identify each country and remain consistent across different datasets.

In order to construct the `region` variable, we will use the `countrycode()` function to map `gwno` codes to region based on the World Development Indicators (WDI) classification. Let's try it out!

Note: There are seven small states or territories are missing their corresponding `gwno` codes (e.g. Grenada or Saint Lucia). For the purpose of this exercise, we are going to drop those variables, which might affect the results slightly.

```r
# Map gwno codes to WDI regions
idc_controls <- idc_controls %>%
  mutate(region = countrycode(gwno, "gwn", "region"))
```

This code assigns a region to each country in the dataset based on its `gwno` code.

Exercise 2: Aggregating Powersharing Data by Region and Year

Now that we have assigned countries to regions, let's aggregate the data to compare patterns across time. To do so, we'll group the data by region and year to calculate annual regional averages for each type of powersharing.

```r
# Compute regional averages for different types of powersharing
powersharing_regional <- idc_controls %>%
  group_by(region, year) %>%
  summarise(`Constraining` = mean(constraining_powersharing, na.rm = TRUE),
            `Dispersive` = mean(dispersive_powersharing, na.rm = TRUE),
            `Inclusive` = mean(inclusive_powersharing, na.rm = TRUE)) %>%
  filter(!is.na(region))
```

```r
# Preview results
head(powersharing_regional)
```

```
## # A tibble: 6 x 5
## # Groups:   region [1]
##   region                year Constraining Dispersive Inclusive
##   <chr>                <dbl>        <dbl>      <dbl>     <dbl>
## 1 East Asia & Pacific  1975       -0.220    -0.142     -0.117
## 2 East Asia & Pacific  1976       -0.335    -0.118     -0.123
## 3 East Asia & Pacific  1977       -0.333    -0.119     -0.124
## 4 East Asia & Pacific  1978       -0.335    -0.106     -0.122
## 5 East Asia & Pacific  1979       -0.196    -0.0468    -0.123
## 6 East Asia & Pacific  1980       -0.206    -0.0175    -0.122
```

Exercise 3: Create the Bar Plot

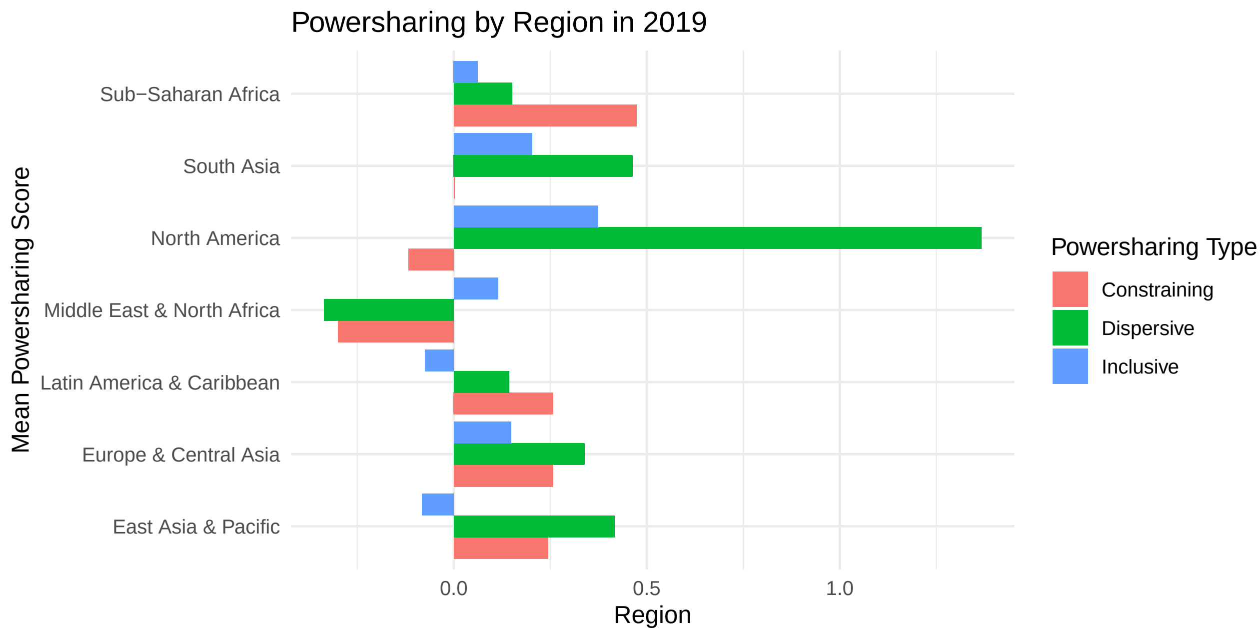
Lastly, create a bar plot to compare regional patterns in powersharing across the most recent year available in the dataset.

```r
# Identify the most recent year in the dataset
latest_year <- max(powersharing_regional$year, na.rm = TRUE)

# Filter the dataset so we only keep observations from that year
powersharing_regional_most_recent <- powersharing_regional %>%
  filter(year == latest_year)

# Convert data to long format and filter for the year 2019
powersharing_long <- powersharing_regional_most_recent %>%
  pivot_longer(
    cols = c(`Constraining`,
             `Dispersive`,
             `Inclusive`),
    names_to = "Powersharing Type",
    values_to = "Mean Value"
  )

# Create a horizontal grouped bar plot for 2019
ggplot(powersharing_long, aes(y = region,
                              x = `Mean Value`,
                              fill = `Powersharing Type`)) +
  geom_bar(stat = "identity", position = "dodge") +
  ## 'position = "dodge"' ensures bars for different types are side-by-side
  labs(title = paste("Powersharing by Region in", latest_year),
       x = "Region",
       y = "Mean Powersharing Score",
       fill = "Powersharing Type") +
  theme_minimal() +
  theme(legend.position = "right")
```



Bonus Exercise

Let's say that we want to take this analysis a step further and explore how these patterns have evolved over time. Understanding how political systems allocate power requires more than just a snapshot; trends over time can reveal regional differences in governance structures and allow us to examine how political institutions adapt over time.

Choose two region in our data to show the evolution of constraining, dispersive, and inclusive powersharing mechanisms over time and generate a plot for each region. What type of plot do you think is best suited for this exercise?

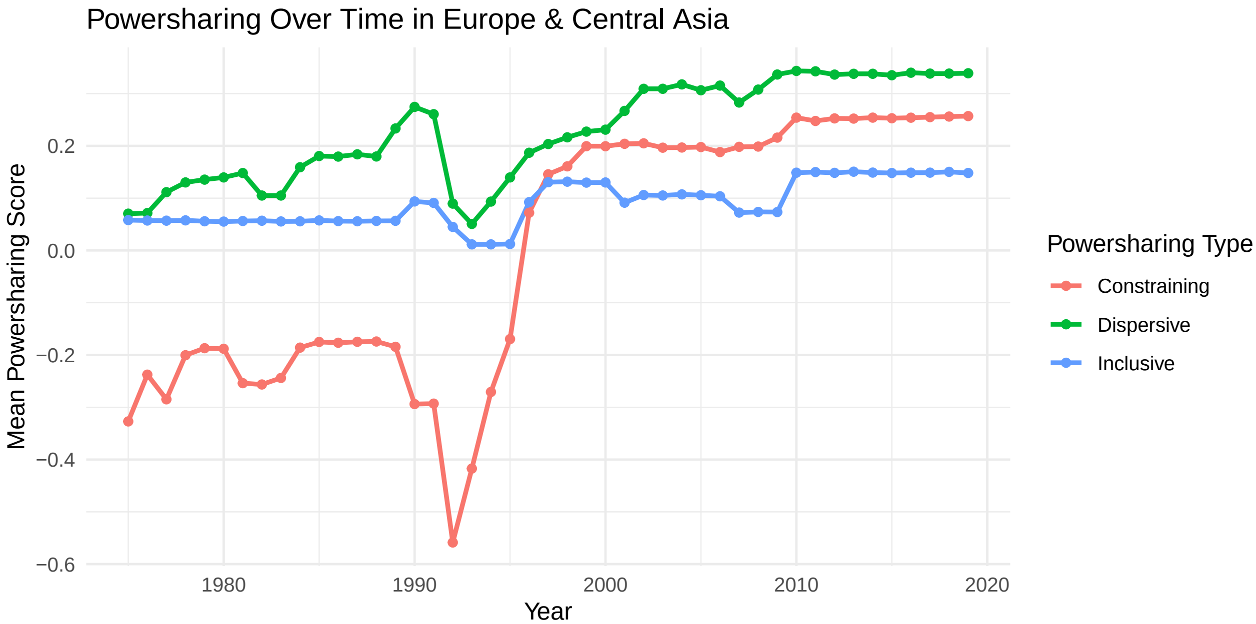
```r
# For this exercise, I chose Europe & Central Asia and Sub-Saharan Africa as the
# regions. Keep in mind that your plots may look different depending on the regions
# you choose.

# Plot for Europe & Central Asia
## Filter the dataset so we only keep observations for Europe & Central Asia
powersharing_eca <- powersharing_regional %>%
  filter(region == "Europe & Central Asia")

## Reshape to long format
powersharing_eca_long <- powersharing_eca %>%
  pivot_longer(
    cols = c(`Constraining`,
             `Dispersive`,
             `Inclusive`),
    names_to = "Powersharing Type",
    values_to = "Mean Value"
  )

## Plot the over-time trend
ggplot(powersharing_eca_long, aes(x = year,
                                  y = `Mean Value`,
                                  color = `Powersharing Type`)) +
```

```
  geom_line(size = 1) +
  geom_point() +
  labs(
    title = "Powersharing Over Time in Europe & Central Asia",
    x = "Year",
    y = "Mean Powersharing Score",
    color = "Powersharing Type"
  ) +
  theme_minimal() +
  theme(legend.position = "right")
```

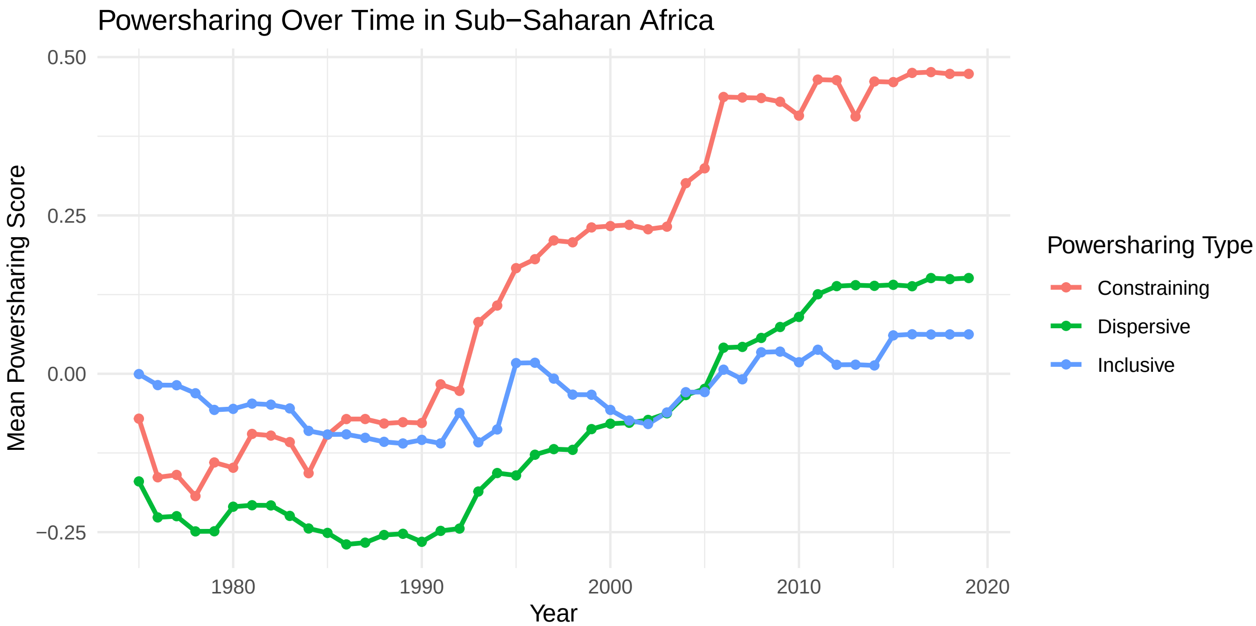



```
# Filter the dataset so we only keep observations for Sub-Saharan Africa
powersharing_ssa <- powersharing_regional %>%
  filter(region == "Sub-Saharan Africa")

## Reshape to long format
powersharing_ssa_long <- powersharing_ssa %>%
  pivot_longer(
    cols = c(`Constraining`,
             `Dispersive`,
             `Inclusive`),
    names_to = "Powersharing Type",
    values_to = "Mean Value"
  )

## Plot the over-time trend
ggplot(powersharing_ssa_long, aes(x = year,
                                  y = `Mean Value`,
                                  color = `Powersharing Type`)) +
  geom_line(size = 1) +
  geom_point() +
  labs(
    title = "Powersharing Over Time in Sub-Saharan Africa",
    x = "Year",
```

```
    y = "Mean Powersharing Score",
    color = "Powersharing Type"
  ) +
  theme_minimal() +
  theme(legend.position = "right")
```



Conclusion

In this groupwork, you practiced mapping countries to regions, aggregating data, and visualizing trends in political powersharing at the regional and decadal levels.

These skills are essential for transforming raw data into meaningful insights. Moving forward, you will apply these techniques in the homework assignment to solidify your understanding and eventually incorporate them into your own research projects.