

# SPEC REU R Resources: Data Management 3 – Homework

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

February 2025

Welcome to the final homework of Data Management III—and the final assignment in our Data Management sequence! Throughout this course, you’ve learned essential skills in cleaning, merging, transforming, and summarizing data. However, data management is not an end in itself; rather, it serves as the foundation for rigorous analysis, meaningful insights, and compelling visualizations that help answer critical research questions.

The purpose of research extends beyond running functions or generating statistics—it is about uncovering patterns, testing hypotheses, and effectively communicating findings. In this assignment, you will continue to practice these principles by exploring important questions in International Political Economy (IPE) using the Master IPE Data Resource, introduced in:

Graham, Benjamin A.T. and Jacob R. Tucker. 2017. “The International Political Economy Data Resource.” *Review of International Organizations*. Online First.

Save your responses in your personal subfolder in the 412\_413 shared Google Drive folder. The R script should be titled HW\_DM3\_[YOUR INITIALS]. You can also save a copy of your R script to your own computer for future reference.

## Initial Setup

To begin, set the appropriate working directory, load the necessary libraries, and import the `Graham_Tucker_IPE_v5.RDS` dataset.

For reference, the Master IPE Data Resource was introduced in:

Graham, Benjamin A.T. and Jacob R. Tucker. 2017. “The International Political Economy Data Resource.” *Review of International Organizations*. Online First.

This data resource merges 97 of the most commonly used datasets related to the field of IPE, with data going back to 1500. The unit of analysis in this dataset is the country-year, with unique observations identified by the Gleditsch-Ward number (gwno) and year.

You can find the research paper [here](#), and download the codebook and replication materials from the [Harvard Dataverse website](#) to learn more about the included variables and their data sources.

For this assignment, we will analyze the impact of foreign direct investment (FDI) on growth. Specifically, we’ll look at whether FDI promotes economic growth in recipient countries. The hypothesis is that countries with higher FDI inflows contribute to GDP growth by increasing capital investment, technology transfer, and job creation.

However, if you’re interested in analyzing the relationship between different variables within the dataset, you are encouraged to complete this assignment by exploring the research question of your choice. This will help you think more deeply about how the research process works.

## Exercise 1: How Have FDI Inflows and GDP Growth Changed Over Time?

Before analyzing whether FDI affects economic growth, we first need to understand their long-term trends. If FDI inflows have been consistently high or low, fluctuations in economic growth might reflect the impact of other factors that we're not accounting for, rather than FDI inflows alone.

To identify trends over time, calculate the global averages for both FDI inflows and GDP growth across the years.

**Helpful Hint:** To measure economic growth, we'll use GDP growth (annual %) from the World Development Indicators (1960–2018), and to measure FDI inflows, we'll use the FDI Flows (Inward) in USD millions from UNCTAD (1970–2020). We will limit our analysis to the period between 1960 and 2021 because of data availability.

## Exercise 2: Visualizing Over-Time Trends

For an additional challenge, visualize how FDI inflows and GDP growth have evolved over time. What type of plot is most suitable for displaying trends over time?

### Bonus Exercise 1

As additional challenge, let's take the log of FDI inflows, and re-run Exercises 1 and 2: calculate the global averages for the log of FDI inflows, plot the over-time trend again, and compare it to the original figure in Exercise 2.

Why might a log transformation be useful? In many datasets, raw values can vary significantly in magnitude, often following a highly skewed distribution. This means a small number of observations have extremely large values compared to the rest. This is especially true for data on income, investment flows, or corporate revenues, where a few countries or firms dominate while many others have much smaller values.

Taking the log of a variable helps by compressing large values and spreads out smaller ones, making the data more evenly distributed; reduces the impact of extreme values, helping highlight underlying trends; and makes it easier to compare patterns across different scales.

Are there any differences between the original FDI inflows plot and the one with the log of FDI inflows?

## Exercise 3: Regional Patterns – Where Is FDI and Growth Concentrated?

However, looking at global trends alone might not provide enough insight. While global averages show fluctuation in FDI inflows and GDP growth over time, they do not tell us where FDI inflows is concentrated or which countries are driving the trend.

Thus, let's move beyond broad global trends and calculate the annual regional averages to uncover regional dynamics that shape the FDI-GDP growth relationship.

**Helpful Hint:** Remember to first create a `region` variable using the `countrycode` package to assign each country to a region using `country`.

### Bonus Exercise 2

For an bonus question, create a plot to illustrate differences in FDI inflows across regions. Make sure you challenge yourself by customizing the visualization to achieve publication-quality results. Are there any outlier in out data?

As mentioned in the previous walkthrough assignment, plotted all regions in the same figure can make the plot crowded and difficult to interpret. Therefore, it can be helpful to narrow the analysis to the regions of greatest interest. Choose a subset of three regions and generate an additional plot showing differences in FDI inflows across those regions.