**Dataset Information and Variable Definitions:**

**country\_pair\_IPDs.csv**

This dataset will likely be the one of most interest to you. The dataset is essentially panelized at the country-pair-year level. It runs from 2001 to 2023. The way we identify a unique country pair is through the variable “sorted”. This variable just puts the ISO3 code that is alphabetically first in the pair as the first in the sorted string (We needed this when working with flow variables).

You will see 3 variables, ipd\_all\_full, ipd\_econ\_full, and Ipd\_all\_short. These are all ideal point distances between countries but use different data-generating processes. The first uses the full sample of UN votes for all vote categories, the second uses the subset of only economic votes (by construction these votes start in the 1970s), and the third uses the sample of all UN votes but beginning after 1991.

For the next set of variables that all start with net\_allign, we take a countries IPD with the US and with China to construct a US-China specific ideal point distance normalized on [-1,1], ie how have countries shifted in their geopolitical proximity between the two countries. Intuitively, US is 1 and China is -1. The suffixes af,ef,as correspond to all votes full sample, economic votes full sample, and all votes short sample respectively (referring to the IPDs they were generated from). The addition of ‘b’ at the end of the variable name just means it is the net\_allignment of the Code\_2 country. Finally, EU is a dummy for if *either* of the countries in the country-pair are EU member states.

**country\_pair\_bloc\_classifications.csv**

This dataset is structured slight differently. For our analysis we use the IPDs and net alignment variables to generate distributions of aligned, nonaligned, and unaligned or US aligned, nonaligned, and China aligned to understand their impacts on a myriad of variables.

When you see a variable that starts with UC that implies the US-China distribution was used (for that given vote subset) to generate the distributions. If you see bloc, that implies that we used the raw IPDs to generate the distribution. The af,ef,as definitions follow from the previous dataset. The number in the variable name (21 or 23) signals if we used 2021 or 2023 to generate the bloc distributions. This is quintessential to our analysis since the distribution is fixed for all years. For example, if the US and Canada are considered aligned using the full sample of votes in 2021, they will be considered aligned for all years. As such, we have 1 observation for each country-pair since it would be identical for every year.

You can pick and choose what variables and distributions you use for your analysis but hopefully this will be sufficient for your analysis! Let me know if you have any questions!