# Data Challenge 6

**Due Thursday, April 4**

This week we're talking about the role of interdependence as a pathway to peace, but as summarized in the lecture notes, when we take some of these ideas to the level of international relations, the data may not align with theory. In particular, we found that economic interdependence as captured through volume of bilateral trade corresponded with a higher probability that one country initiated a militarized dispute with another. While it also was true that high trade dependence by one side on the other corresponded with a decline in the probability that that side started a fight, trade dominance made a side more likely to do so.

Taken together, this analysis showed that economic ties might be a double-edged sword at the level of international politics. However, there are other ways that countries can be connected that might be a better pathway to peace. We can access a few measures of these alternative ways countries can be connected from the `{cepiigeodist}` R package, which you can install using the code below:

```
## install {cepiigeodist}
devtools::install_github("pachamaltese/cepiigeodist")
```

This package provides access to a series of indicators of "distance" between countries, where distance can mean the literal geographic distance between countries, but also cultural and social distance. Measures of the latter include whether one country was a former colonizer of the other, whether two countries share the same colonizer, whether they share a common language, and whether once upon a time the two countries were once part of the same country. Once you install the `{cepiigeodist}` package, you can use the code on the next page to create a dataset for analysis that will be very similar to the one we used in the lecture notes. However, it now will include these additional indicators of "closeness" between countries.

For this data challenge your objective is to extend the analysis in the lecture notes by adding the following variables to your statistical model:

- `comlang`: An indicator that takes the value 1 if a pair of countries either have the same official language or at least 9% of their population speaks the same language.
- `colony`: An indicator that takes the value 1 if one of the countries in the pair was ever a colonizer of the other.
- `comcol`: An indicator that takes the value 1 if both countries in the pair share the same colonizer.
- `smctry`: An indicator that takes the value 1 if both countries in the pair were ever the same country.

After adding these variables to your model (along with those already included in the model as per the lecture notes), write a short discussion of the results that addresses the following questions:

1. What kind of interdependence (if any) might these additional factors reflect between countries (economic, social, cultural, something else)?
2. Does accounting for these factors change the impact of trade and trade dependence on MID initiation?
3. Do any of these additional factors impact the likelihood of MID initiation in ways that are consistent with or surprising in light of the arguments that Blattman makes about interdependence?

When your done, render your work and submit to Canvas.

```r
## open packages
library(tidyverse)
library(peacesciencer)
library(cepiigeodist)

## make dataset
create_dyadyears(
  subset_years = 1870:2010
) |>
  add_gml_mids() |>
  add_spells() |>
  add_sdp_gdp() |>
  filter_prd() -> Data

## read in trade dataset
readRDS(
  url("http://svmiller.com/R/peacesciencer/cow_trade_ddy.rds")
) -> trade_dt

## merge the trade data with the main data
Data |>
  left_join(
    trade_dt
  ) -> Data

## merge in CEPII distances dataset
# ?dist_cepii # to learn more
dist_cepii |>
  mutate(
    ccode1 = countrycode::countrycode(
      iso_o, "iso3c", "cown"
    ),
    ccode2 = countrycode::countrycode(
      iso_d, "iso3c", "cown"
    )
  ) -> dist_cepii_to_merge

Data |>
  left_join(
    dist_cepii_to_merge
  ) -> Data
```

```r
## make trade vars for analysis
# rescaling function
rescale <- function(x) {
  x <- rank(x)
  x <- x - min(x, na.rm = T)
  x <- x / (max(x, na.rm = T))
  x <- log(x * 9 + 1)
  x
}
# make changes
Data |>
  mutate(
    # treat NA = 0 for trade
    across(flow1:flow2, ~ replace_na(.x, 0)),
    # make dyad indicator
    dyad = paste(ccode1, ccode2, sep = "-"),
    # measure of total trade
    total_trade = flow1 + flow2,
    # the log-rank of total trade
    lrank_trade = rescale(total_trade),
    # trade dependence of side 1 and 2
    trade_dep1 = (1000000 * total_trade) / exp(wbgdp2011est1),
    trade_dep2 = (1000000 * total_trade) / exp(wbgdp2011est2),
    # log ratio of side 1 trade dependence
    lrat_trade_dep1 = rescale(trade_dep1) - rescale(trade_dep2),
    # log-rank of dyadic distance
    lrank_dist = rescale(dist),
    # common language
    comlang = pmax(comlang_off, comlang_ethno)
  ) -> Data
```