Codebook for Analysis Datasets, "The Effective Power of Military Coalitions"

This codebook covers the two analysis datasets distributed with the replication archive for "The Effective Power of Military Coalitions" by Brenton Kenkel and Kristopher W. Ramsay. For any questions about the data or replication materials, contact Brenton Kenkel at brenton.kenkel@gmail.com.

`kr_analysis_dispute.rda`

This RData file contains one object, `data_dispute`, a list of 10 data frames. Each data frame contains 2,101 dispute-level observations and contains the same columns. The only differences between them are the imputed values of missing variables.

In the following variable descriptions, an "opposing dyad" is a dyad comprised of countries on opposite sides of the dispute. When either or both sides consist of more than one country, we often take averages across all of the opposing dyads in the dispute.

The variables are:

- `id`: Unique identifier for each dispute. Used to link to participants in `kr_analysis_participant.rda`.
- `war`: Binary indicator for whether the dispute resulted in war, which we code as having 25+ fatalities.
- `win_a`: Binary indicator for whether the initiating side won, conditional on war occurring.
- `win_b`: Binary indicator for whether the defending side won, conditional on war occurring.
- `win_b_alt`: Binary indicator for whether the initiating side did not win, conditional on war occurring.
- `s_nowt`: Average unweighted S-score for opposing dyads.
- `s_cinc`: Average CINC-weighted S-score for opposing dyads.
- `contig`: Proportion of opposing dyads that share a land border or are separated by no more than 150 miles of water.
- `py`: Average years since the last MID between opposing dyads.
- `py_alt`: Modified version of `py`, where peace years for a dyad are reset to 0 whenever one or both states temporarily leaves the international system.
- `mp_any`: Binary indicator for whether there is a major power on at least one side of the dispute.

- `mp_all`: Binary indicator for whether there are major powers on both sides of the dispute.
- `polity2_min`, `polity2_max`: For each side of the dispute, we extract the minimum (least democratic) Polity IV regime type score among countries on the given side. The lower of these two values is `polity_min`; the greater is `polity2_max`.
- `n_states`: Total number of states involved in the dispute.
- `polity_a`, `polity_b`: Average Polity IV regime type score for the initiating and defending sides, respectively.
- `majpow_a`, `majpow_b`: Indicators for whether there is a major power on the initiating and defending sides, respectively.
- `n_states_a`, `n_states_b`: Number of states on the initiating and defending sides, respectively.

`kr_analysis_participant.rda`

This RData file contains one object, `data_participant`, a list of 10 data frames. Each data frame contains 4,962 dispute participant-level observations and contains the same columns. The only differences between them are the imputed values of missing variables.

Unless otherwise specified, all variables are coded as of the participant's first year of involvement in the dispute.

The variables are:

- `id`: Unique identifier for each dispute. Used to link to disputes in `kr_analysis_dispute.rda`.
- `ccode`: The participant's Correlates of War country code.
- `stateabb`: The participant's Correlates of War country abbreviation.
- `statenme`: The participant's Correlates of War country name.
- `fe_1`: Fixed effect term, grouping together countries that always appear together on the same side of all disputes.
- `fe_2`: Alternative fixed effect term, allowing for time-varying effects for major powers.
- `sidea`: Binary indicator for whether the participant is on the initiating side of the dispute.
- `year`: First year the participant was involved in the dispute.
- `distance`: Distance in miles from the state's capital to the MID location (set to 0 if the dispute occurs at or within the state's borders).

- `milex`: Participant's military expenditures, via the National Material Capabilities data.
- `milper`: Participant's military personnel, via the National Material Capabilities data.
- `irst`: Participant's iron and steel production, via the National Material Capabilities data.
- `pec`: Participant's primary energy consumption, via the National Material Capabilities data.
- `tpop`: Participant's total population, via the National Material Capabilities data.
- `upop`: Participant's urban population, via the National Material Capabilities data.
- `cinc`: Participant's Composite Index of National Capabilities, via the National Material Capabilities data.
- `polity2`: Polity IV project measure of participant's regime type.
- `cgdpp_madd`: Maddison project "CGDPpc" measure of participant's real GDP per capita.
- `rdgppc_madd`: Maddison project "RGDPNApc" alternative measure of participant's real GDP per capita.
- `pop_madd`: Maddison project estimate of participant's population.
- `gdp_pwt`: Penn World Tables "RGDPo" measure of participant's output-side real GDP.
- `rgdpe_pwt`: Penn World Tables "RGDPe" measure of participant's expenditure-side real GDP.
- `cgdpe_pwt`: Penn World Tables "CGDPe" measure of participant's expenditure-side real GDP.
- `cgdpo_pwt`: Penn World Tables "CGDPo" measure of participant's output-side real GDP.
- `pop_pwt`: Penn World Tables estimate of participant's population.
- `pop_wdi`: World Development Indicators estimate of participant's population.
- `gdp_wdi`: World Development Indicators measure of participant's GDP.
- `pct_imports`: Ratio of participant's total imports to its GDP, per World Development Indicators data.
- `majpow`: Binary indicator for whether the participant is a major power, as coded by Correlates of War.

- `latitude`: Latitude of participant's capital.
- `longitude`: Longitude of participant's capital.
- `quality`: Ratio of military expenditures to military personnel (0 if military personnel is 0).
- `milex_lag_pure` through `quality_lag_pure`: Lagged values of the corresponding National Material Capabilities values.
- `milex_lag` through `quality_lag`: Lagged values of the corresponding National Material Capabilities values, using the current-year value when the lagged value is unavailable due to the participant not being in the state system in the prior year.