Documentation for COW Direct Contiguity Data - Version 3.20
Paul R. Hensel
13 February 2017

Overview

Version 3.2 of the Correlates of War Direct Contiguity data identifies all direct contiguity relationships between the homeland territory of states in the international system from 1816 through 2016. Contiguity relationships involving colonial or other non-homeland territories are included in the COW Colonial Contiguity data set, which is available separately.

Citation

In any papers or publications that utilize this data set, users are asked to give the version number and cite the data set, as follows:

> Correlates of War Project. *Direct Contiguity Data, 1816-2016*. Version 3.2.

Users are asked to cite the current article of record for the data set, as follows:

> Douglas M. Stinnett, Jaroslav Tir, Philip Schafer, Paul F. Diehl, and Charles Gochman (2002). "The Correlates of War Project Direct Contiguity Data, Version 3." *Conflict Management and Peace Science* 19 (2):58-66.

Additional details of the basic coding process were elaborated in

> Charles S. Gochman (1991). "Interstate Metrics: Conceptualizing, Operationalizing, and Measuring the Geographic Proximity of States since the Congress of Vienna," *International Interactions* 17 (1): 93-112.

Description of Data

The classification system for contiguity is comprised of five categories, one for land contiguity and four for water contiguity. Land contiguity is defined as the intersection of the homeland territory of the two states in the dyad, either through a land boundary or a river (such as the Rio Grande along the US-Mexico border). Water contiguity is based on whether a straight line of no more than a certain distance can be drawn between a point on the border of one state, across open water (uninterrupted by the territory of a third state), to the closest point on the homeland territory of another state. Four different levels of water contiguity are recorded, based on the distance between the two states' territories: up to 12 miles (reflecting the widely recognized 12-mile limit for territorial waters), 24 miles (reflecting the maximum distance at which two states' 12-mile territorial limits can intersect), 150 miles (from the original 1816-1965 version of the data set, reflecting what was considered the average distance that a sailing ship could travel in one day), and 400 miles (the maximum distance at which two 200-mile exclusive economic zones can intersect).

The coding for each dyad is always the closest form of contiguity for that year. Thus, if a dyad shares a land border and is also separated by a stretch of water less than 400 miles, that dyad will be coded based on the closest form of contiguity (the land border). If the type of contiguity for a dyad changes during a given year, only the closest classification is recorded for that year, in order to prevent multiple records for one dyad-year from appearing in the data set.

A variety of sources have been used in the process of collecting these data. The sources used for earlier versions of the data are identified in Gochman (1991); more recent updates have relied on various editions of the *Times Atlas of the World*, *National Geographic Atlas of the World*, *Hammond World Atlas*, and *Oxford Atlas of the World*.

The downloadable data archive includes a text file with this codebook, as well as three data files that are provided in .csv comma-delimited format, which can easily be read into any spreadsheet or statistical software:

- **contdir.csv**: The master data file (with one entry per contiguity relationship)
- **contdird.csv**: A directed dyad-year-level version of the master data
- **contdirs.csv**: A state-year-level data set calculated from the master data

Variable List

Direct Contiguity Master File: ContDir

This file contains the master records for the Direct Contiguity data. It is the one from which all other files are ultimately derived and to which all future changes will be made. It has a non-directed format in order to prevent asymmetries from accidentally being introduced into the data set. The variable names are given in the first record in the file. The format of the remaining records is as follows:

• Dyad: A variable that combines the two states' COW country codes in a way that facilitates merging with other dyadic data. This takes the format AAABBB, where AAA is the lower value of the two country codes and BBB is the higher value.
• StateLNo: COW state number ("country code") of lower numbered member of this dyad
• StateLAb: COW state abbreviation of lower numbered member of this dyad
• StateHNo: COW state number ("country code") of higher numbered member of this dyad
• StateHAb: COW state abbreviation of higher numbered member of this dyad
• ContType: Type of contiguity relationship
    1: Separated by a land or river border
    2: Separated by 12 miles of water or less
    3: Separated by 24 miles of water or less (but more than 12 miles)
    4: Separated by 150 miles of water or less (but more than 24 miles)
    5: Separated by 400 miles of water or less (but more than 150 miles)
• Begin: Year and month when this contiguity relationship began (YYYYMM)
• End: Year and month when this contiguity relationship ended (YYYYMM)
• Notes: Additional information about the contiguity relationship such as the body of water that is crossed (for sea contiguity), the portion of the state's territory from which the distance was calculated, and the reasons for the beginning and (where relevant) ending of the relationship.
• Version: Data set version number

Dyadic Direct Contiguity: ContDirD

This file is derived from the master file and includes annual records for each (directional) pair of states that is directly contiguous. Dyad-years in which direct contiguity is absent are not included in the file to reduce greatly the size of the file. Users may assume that if no record is present for a

given dyad-year, no direct contiguity exists between the state pair in that year. The variable names are given in the first record in the file. The format of the remaining records is as follows:

• Dyad: A variable that combines the two states' COW country codes in a way that facilitates merging with other dyadic data. This takes the format AAABBB, where AAA is the lower value of the two country codes and BBB is the higher value.
• State1No: COW state number ("country code") of first member of this dyad
• State1Ab: COW state abbreviation of first member of this dyad
• State2No: COW state number ("country code") of second member of this dyad
• State2Ab: COW state abbreviation of second member of this dyad
• Year: Year of this observation
• ContType: Type of contiguity relationship
    1: Separated by a land or river border
    2: Separated by 12 miles of water or less
    3: Separated by 24 miles of water or less (but more than 12 miles)
    4: Separated by 150 miles of water or less (but more than 24 miles)
    5: Separated by 400 miles of water or less (but more than 150 miles)
• Version: Data set version number

State-Level Direct Contiguity: ContDirS

This file is derived from the master file and includes annual records for all states in the COW interstate system. The variable names are given in the first record in the file. The format of the remaining records is as follows:

• StateNo: COW state number ("country code") of state
• StateAb: COW state abbreviation of state
• Year: Year of this observation
• Total: Total number of direct contiguities for this state this year (land or sea)
• Land: Total number of direct contiguities by land for this state this year (i.e., ContType = 1)
• Sea: Total number of direct contiguities by sea for this state this year (i.e., ContType > 1)
• Version: Data set version number

Version History

Version 1

The first version of the COW contiguity dataset was collected at the University of Michigan in the early 1970s.

Version 2

The first major revision was compiled by Philip Schafer of the Correlates of War project at the University of Michigan. This version updated the earlier dataset through 1993, corrected some anomalies in the previous data, and added the 400 mile contiguity relationship.

Version 3.0

For its next major revision, the dataset was split into two files: Direct Contiguity (containing data on contiguous relationships between the metropole borders of two states) and Colonial

Contiguity (containing data on contiguous relationships that included at least one colony or other dependency). Paul Diehl and several graduate students at the University of Illinois updated the Direct Contiguity dataset through 2000; several years later, Stuart Bremer and several graduate students at Penn State University updated the Colonial Contiguity dataset through 2002. At this time, the dataset was also converted to a non-directed dyad format.

Version 3.10

This was the first version of the dataset to be updated and maintained by Paul Hensel under the COW Project's data hosting program. The following changes were made in this update:

• Temporal coverage: Data extended through the end of 2006.

• Added months & corrected years: A number of entries were ended a year too early or begun a year too late so that the data didn't code the same border twice during the same year; this has been corrected by giving the month for an observation beginning or ending, which allows both phases of that relationship to be coded during the same year. This affected 62 beginning or ending dates in the Direct Contiguity data.

• Added notes: Only a few cases had notes in version 3.0; I have now added notes to every observation to explain why it began/ended when it did, and to list the body of water that is crossed (where relevant)

• Updated state-year-level data: The previous version of this file (contdirs) only included observations that had at least one form of contiguity. This has been updated with the COW interstate system membership list to include all years for all states. including observations with no land or sea borders. This is meant to prevent users from unwittingly omitting these border-less cases from any analyses that might be run; users who wish to focus only on states with at least one interstate border are free to delete all border-less observations (those where total = 0).

• St. Pierre & Miquelon (entity #1029): this was coded in version 3.x as becoming part of France (rather than a colony) from 1949-2000. That has now been reversed; other similar entities (e.g. French Guiana and Martinique) were still coded as French colonies during this time.

• New borders: Since the end of data coverage in version 3.0 (2000), East Timor and Montenegro have qualified for membership in the COW interstate system, leading to the coding of new borders for the new states and (in some cases) corresponding changes to Indonesia's and Yugoslavia's borders. Also, with the 2000 opening of the Oresund Bridge, the previous sea border between Sweden and Denmark becomes a land border (like the Saudi-Bahrain border with the 1986 opening of the King Fahd Causeway).

Version 3.20

This update fixed a few typos, added a few clarifications in the Notes field, and made the following more substantial changes:

• Temporal coverage: Data extended from 2006 through 2016.

• New borders: Since the end of data coverage in version 3.10 (2006), Kosovo and South Sudan have qualified for membership in the COW interstate system, leading to the coding of new borders for the new states and (in some cases) corresponding changes to Serbia/Yugoslavia's and Sudan's borders. Restored Romanian and Bulgarian borders with Russia across Black Sea upon Russian annexation of Crimean Peninsula.

• Updated several observations for consistency with changed COW system membership dates.

• Changed or updated COW country codes and abbreviations for Montenegro (#341/MNG);

USSR/Russia (RUS); and Romania (ROM).

Errors and Anomalies

Every effort has been made to identify and correct errors in this data set, but some anomalies may have escaped our attention. If you detect something suspicious or questionable in this data set that the Notes field in the master data file does not clear up, please contact the data host, Paul Hensel (phensel@unt.edu). A quick response to your message cannot be guaranteed, but your input will not be ignored and will help us with the future development of this data set.