

The international political economy data resource

Benjamin A. T. Graham¹ · Jacob R. Tucker¹

Published online: 30 June 2017
© Springer Science+Business Media, LLC 2017

Abstract Quantitative scholars in international relations often draw repeatedly on the same sources of country-year data across a diverse range of projects. The International Political Economy Data Resource seeks to provide a public good to the field by standardizing and merging together 951 variables from 78 core International Political Economy data sources into a single dataset, increasing efficiency and reducing the risk of data management errors. Easier access to data encourages researchers to perform more robustness checks in their own work and replicate others' published results more often. It also and makes it easier for teachers of quantitative research methods to assign realistic exercises to their students. This resource will be updated and expanded annually. The full resource is available via the Harvard Dataverse Network, with versions also available via the Niehaus Center for Globalization and Governance at Princeton University and NewGene.

Keywords International political economy · Political economy · Time-series-cross-sectional · Country-year · Data · Dataset

1 Introduction

Cross-national quantitative research in international relations is a data-intensive undertaking. Many researchers create novel measures of the constructs most critical to the question at hand – i.e., their variable(s) of interest. However, between control variables in the core model and a range of alternative measures employed as robustness checks, it is not uncommon for a single paper to utilize thirty or forty variables drawn from a dozen distinct datasets. In papers that use latent variable models to estimate

✉ Benjamin A. T. Graham
benjamin.a.graham@usc.edu

Jacob R. Tucker
jacobtuc@usc.edu

¹ University of Southern California, Los Angeles, CA, USA

unobservable variables of interest, the number of indicator variables used in a single paper can rise into the hundreds.¹ Thus, gathering together and compiling a diverse array of variables has become a substantial part of the labor of quantitative scholarship in international relations.

Country-year data in international relations in general and international political economy (IPE) in particular are fragmented in their sources, drawn from economics, political science, sociology, international business, and other fields. Given the low readership of most journals by scholars from outside the discipline, even identifying the most appropriate existing data sources is challenging. Once identified, merging multiple datasets together requires much labor and great care, as errors at this stage can undermine all the work that follows. Scholars of international security have long utilized EUGene for a compilation of important variables.² The IPE Data Resource brings together a much broader range of variables, focusing on those of interest to political economy scholars broadly construed, including political scientists, development- and macro-economists, and scholars of international business.

Despite the diversity of topics studied, many political economy projects utilize the same standard set of core country-year data sets to measure economic, political and social conditions across countries and over time. This area of overlap has only grown as economic factors have become more central to the study of political and military outcomes and as political factors have been more fully considered in analyses of economic outcomes and business strategy. This creates an opportunity for large efficiency gains through coordination. Currently, the process of merging these core datasets together is repeated scholar by scholar, project by project. This resource aims to improve and streamline this data management process.

Increased efficiency in data management not only reduces errors and saves time, it encourages researchers to perform their work more thoroughly. The cost of identifying, obtaining, and preparing alternative measures deters scholars from running additional robustness tests that would increase the quality of their work. All the measures we work with as quantitative social scientists are flawed in some way; use of more alternative measures reduces the odds that the results we publish are driven by these flaws. A centralized source for data is particularly valuable in an interdisciplinary topic area such as IPE where information about new data resources can be slow to diffuse across disciplinary boundaries.

In addition to increased robustness checks, aggregation resources can be effective limiters of p-hacking by increasing the ease of replication and introducing greater data transparency. Altering model specifications until the desired level of statistical significance is reached is a known issue in social science research (e.g., Monogan III 2013; Humphreys et al. 2013). Over the last several decades, the amount of data made publically available has increased substantially (e.g., Coppedge et al. 2016; Cruz, Keefer, and Scartascini 2016; Marshall and Cole, 2014; Fouré, Bénassy-Quéré, and

¹ See, for example, Fariss (2014) assessing government respect for human rights, Hollyer et al. (2014) assessing government transparency, or Fariss, Graham and Kenwick (2016) measuring the security of property rights.

² EUGene refers to the Expected Utility Generation and Data Management Program (Bennett and Stam 2000). NewGene, a successor to EUGene, will be released soon (Bennett, Poast and Stam 2015).

Fontagné 2012), heightening the possibility of replication studies, which are critical to resolving this issue (Laitin 2013). A key contribution of this resource is to provide the data necessary for scholars to move beyond simply reproducing published results using author's published replication data, and allowing them to more fully probe the robustness of published results, testing whether those results are sensitive to the use of alternative measures for key variables. By bringing the most commonly used variables in IPE literature into a single resource that is regression-ready and easily accessible, the cost of such robustness testing is reduced. Having such a large number of variables available does, of course, make it easier for a scholar to simply download the data and run regressions until one with significant results appears. However, it also makes it easier for replication studies to expose and overturn such tenuous results.

Efficient data management is valuable in teaching as well as research. Professors teaching quantitative methods courses sometimes shy away from assigning "real world" data exploration exercises due to the difficulty of pulling together the necessary datasets for their students. Similarly, one of the motivations for this resource comes from a common assignment in graduate statistics courses – to reproduce and probe the robustness of a core finding in the literature. Having a dataset like this on hand from which students can draw additional/alternative regressors makes it easier for students to complete such a project in the span of a semester.

This project standardizes and merges together 78 commonly used datasets drawn from a range of disciplines into a single, analysis-ready country-year dataset.³ Political institutions, economic and social conditions, and political violence are all tightly intertwined, and thus, while the content area of these data is focused on IPE, we expect this resource to be of interest to scholars doing cross-national work across the social sciences. The diversity of disciplines we hope will be served by the resource is reflecting the diversity of the fields represented in the data: in addition to political science and economics, the resource contains component datasets related to international business (e.g., Kingsley and Graham 2017); law (e.g., Linzer and Staton 2015); education (e.g., Barro and Lee 2013); religion (e.g., Brown and James 2015); and geography (e.g., Mayer and Zignago 2011).

The resource will be updated and expanded annually, incorporating new data sources as they become available. The boundaries of the field of IPE are inherently arbitrary; these annual updates allow additional data to be added to the resource as requested by its users.⁴

2 Method

The IPE data resource is organized into four general topic areas: politics (37 datasets); economics (31 datasets); social conditions (8 datasets); and geography (2 datasets). See Table 1 in the Appendix for a full list. From some of the largest datasets, which contain hundreds of variables, we retain those variables we expect to be of most use to scholars working in areas related to IPE.

³ This count, and all details in this article, refer to version 1.5 of the data. Additional datasets will be added with each update.

⁴ Please send suggestions for additions or revisions to benjamin.a.graham@usc.edu.

While we expect this data resource to be quite useful to security and civil conflict scholars, this resource includes only a limited number of measures of interstate and intrastate war.⁵ We largely leave it to scholars in these areas to merge in their dependent variable(s) of interest, which may be quite varied in form.

The most critical decision regarding how to compile country-year data is the choice of unique country ID numbers. Over time, countries are born, conquered, merged and divided. Following every merger or division, a time-series-cross-sectional dataset must either treat the resulting units as continuations of previous countries, in which case they retain the same ID number, or as new entities with a new number. In this project, we use the ID numbers associated with the Gleditsch-Ward system (Gleditsch and Ward 1999). However, the resource also includes Correlates of War (COW) codes, International Financial Statistics (IFS) codes, and the country codes from the World Development Indicators.⁶

The Gleditsch-Ward system asserts continuation of political units through most mergers and divisions. Thus, modern Yemen, which was formed through the merger of People's Democratic Republic of Yemen and the Arab Republic of Yemen in 1990, is treated as a continuation of the Arab Republic of Yemen. Imperial Russia, the Soviet Union, and modern Russia all retain the same number. When Estonia, Latvia, and Lithuania regain independence in 1991, they regain the same ID numbers they had prior to their loss of independence in 1940.

This presumption of continuity is appropriate to most of the core questions in IPE, as well as the study of intrastate violence and some (though not all) comparative politics questions.⁷ In most cases, the same firms that are operating the day before a merger or division of countries continue to operate after it; debts that were owed continue to be owed; contracts remain in place; social identities and divisions persist; and in many cases even the identity of political elites and the structure of many political institutions remains the same (e.g., Acemoglu and Robinson 2006; Alston 1996). Thus the Gleditsch-Ward system treats the economic and social unit (the country) as persisting as well, albeit in altered form. However, to facilitate use of this resource by security scholars, we make available a version of the dataset that is time-series-set for COW codes, which are the dominant set of country identifiers in the study of interstate war.⁸ It

⁵ The one exception is the Major Episodes of Political Violence Data (Marshall 1999), which includes economic damage within its index of conflict severity. We view this as a conflict measure particularly appropriate for use as an independent variable in analyses of economic outcomes.

⁶ IFS codes are used by the International Monetary Fund. World Development Indicator codes are included beginning with version 2.0.

⁷ It is notable that the Uppsala/PRIO civil conflict datasets are all based on Gleditsch-Ward numbers (Gleditsch et al. 2002).

⁸ Time series setting refers to the process of establishing unit and time variables within Stata to identify unique time-series-cross-sectional observations. The primary version of this resource contains some observations that are duplicates by COW code and year. In particular, some component datasets include data for both Serbia (in isolation) and Yugoslavia (in total) in the same years. We can retain this information in the Gleditsch-Ward-based version of the data, where modern Serbia is treated as a continuation of the Kingdom of Serbia and as distinct from Yugoslavia, but must drop this information from the COW version, where Yugoslavia and Serbia share the same country code.

is the COW version of the data that is integrated into the forthcoming NewGene software.

Many of the component datasets included in the IPE data resource do not include Gleditsch-Ward numbers in their native form, and no other numbering system is hegemonic. Some datasets use IFS codes, others use COW codes, etc. Thus, we append Gleditsch-Ward IDs on the basis of country names, carefully remove duplicate observations by ID and year,⁹ and then proceed to merge the datasets together.

Data cleaning involves difficult choices, and choices optimal for one research project may be inappropriate for another. Thus, it is important for scholars using this resource to be able to track and review every step that is taken from when the raw data are downloaded until the final dataset is compiled. To this end, extensive documentation is posted along with the final dataset. The codebook for this project provides the name of each dataset, a description of each variable, citations to the sources from which the data are drawn, notes regarding any data cleaning that was necessary to integrate each component dataset into the resource, and links to the relevant codebooks whenever possible. We also post the raw component data files themselves; the Stata .do files that clean these data files and append the Gleditsch-Ward Numbers (one per raw data file); and the Stata .do file that merges all of the prepped data files together. Thus, every step in the process is transparent, and scholars can review and revise any of the data cleaning choices made.

3 Accessing the resource

The full version of the resource, including all code and documentation is available via the Harvard Dataverse.¹⁰ The resource will be updated and expanded annually. Version 1.5 of the dataset was released on August 30, 2016, and version 2.0 is scheduled for release in July 2017. Newly released versions will be posted first to Dataverse; Dataverse automatically archives all past versions.¹¹ The archived versions are critical, as they allow scholars to replicate published studies using the same version of the data used in the original study.

In order to maximize accessibility and serve diverse user communities, we also make the resource available via two other mechanisms. First, an interactive interface hosted by the Niehaus Center for Globalization and Governance at Princeton University, which allows users to download subsets of variables, countries, and years.¹² Second, to aid security scholars in particular, version 1.5 of the IPE Data Resource is

⁹ In this process, we remove the observation with the least information. In most cases at least one of the duplicate observations contains exclusively missing values.

¹⁰ <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/X093TV>

¹¹ <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/X093TV> Versions 1.0–1.4 are not on Dataverse and are available from the authors upon request. Other sites that host interactive versions of the resource may experience a lag in posting updated versions of the resource after they become available via Dataverse.

¹² <http://ncgg.princeton.edu/irdataverse.php>

scheduled for inclusion in version 1.0 of NewGene, the successor to EUGene, a software program facilitating the creation of dyadic and directed-dyadic datasets for the study of interstate war (Bennett, Poast and Stam 2015).¹³ While both of these alternative sites provide the codebook to the version of the resource posted, the code used in the standardization and merging process and the archive of outdated versions are only provided via Dataverse.

4 Existing resources

A number of scholars and institutions have taken the step of providing catalogs of links to frequently used datasets. For example, Paul Hensel maintains the International Relations Data Site, which is essentially a catalog of links to individual datasets created by various scholars.¹⁴ The Quality of Governance Institute (QoG) provides the resource most similar to that which we offer here, but with more of a comparative politics focus than an IPE focuses (Teorell et al. 2016). Thus, the largest difference between these two resources is in the substantive topics they cover: 106 datasets are compiled by QoG and 78 datasets here, of which only 28 overlap.¹⁵ However, these resources also differ in their approach to country codes. The QoG (rightly) takes an approach to country codes that is tailored to the study of domestic institutions and processes. Every time countries merge or divide in the QoG data, the resulting units are given new country IDs. Thus, in the QoG data, a unified Yemen is treated as neither a continuation of South Yemen or of North Yemen, but rather as an entirely new entity. This is appropriate in many comparative politics contexts because the nature of domestic politics has been fundamentally altered, but as discussed in the method section, it is less ideal for the study of international relations. Thus, the IPE data resource uses Gleditsch-Ward and COW IDs instead.

The IPE Data Resource also has similarities to the data management components of the Expected Utility Generation and Data Management Program (EUGene) developed by Bennett and Stam (2000), the World Development Indicators (WDI), which are compiled by the World Bank, the *GROW^{HP}* project created by Girardin et al. (2015), and to the PRIO-GRID dataset (Tollefsen, Strand, and Buhaug 2012). EUGene and its forthcoming successor, NewGene, cover a range of variables often used in models of interstate conflict, while WDI covers measures of economic and human development, largely eschewing measures of political institutions. Thus, a key distinction is again the substantive area of focus of each data aggregation effort. PRIO-GRID brings together a range of economic, social, and political information on the basis of geographic space, rather than country boundaries. PRIO-GRID covers some of the same variables

¹³ Software available at <http://newgenesoftware.org>

¹⁴ <http://www.paulhensel.org/data.html> Accessed May 16, 2017.

¹⁵ This count was produced using the January 2016 version of the QoG Standard Data.

as this project, but it brings them together with regard to a different set of geographic units – quadratic grid cells rather than countries.

These data resources are complements rather than substitutes for one another. Just as NewGene incorporates a version of the IPE Data Resource into its interface, our dataset includes variables drawn from the WDI. Each resource serves a related, but distinct need of the quantitative social science community.

5 A note on citation

The component datasets compiled in this project were each constructed, through great cost and effort, by scholars and institutions that deserve credit for their work.¹⁶ The codebook to the IPE Data Resource provides the appropriate citation for each component dataset. Any publication that draws on data from this resource must cite to the original datasets from which the variables in question are drawn and this paper, with reference to the specific version of the IPE Data Resource utilized. This allows precise replication of any results produced using the resource and makes clear the source of any data-related errors.

6 Conclusion

This data resource serves the field by bringing together 78 datasets from political science, economics, international business, and sociology and carefully merging them together in a theoretically informed manner. The resource can be downloaded in its entirety, or subsets of variables, countries, and years can be selected. Datasets are merged together based on the Gleditsch-Ward system of country IDs, but an alternate version of the dataset employing the COW numbering system is also available. Because the project has the institutional support necessary to sustain annual updates, it will be able to remain relevant and up to date, updating existing data and incorporating new component datasets as they become available. A dyadic version of the resource represents the next logical step in this resource creation agenda. A dyad-year resource that includes information on bilateral trade and investment flows, joint membership in international organizations, and other dyadic information is under construction as a complement to the country-year resource described here.

The value of this project comes in the form of increased efficiency, increased research and teaching quality, and the diffusion of new data resources across disciplinary boundaries. As the costs of doing high quality work fall, we can all do better work and more of it.

¹⁶ Permission to repost data was obtained from the authors of each component dataset with the exception of datasets governed by creative commons licenses.

Appendix

Table 1 List of component datasets

Institution/ Authors	Dataset Name	# of countries and years	Topic	
Economic Datasets				
Gwartney, Lawson, and Hall	Economic Freedom of the World	158	1970–2013	Indicators of economic freedom
Freedom House	Freedom in the World	195	1972–2014	Measures of political and civil liberties
Heritage Foundation	Index of Economic Freedom	187	1995–2016	Index of economic freedom
Aizenman, Chinn, and Ito	Open Economy Trilemma	188	1960–2014	Monetary independence, exchange rate stability, and financial openness
Singer, Bremer, and Stuckey	COW Material Capabilities	212	1816–2001	CINC scores
Stockholm International Peace Research Institute	SIPRI Military Expenditures	172	1988–2015	Military expenditures
Claessens and van Horen	Bank Ownership	137	1995–2009	Annual values for number of foreign banks in each country
Bodea and Hicks	Central Bank Independence	82	1972–2010	Central bank independence score for each country
International Monetary Fund	Contract Intensive Money	184	1948–2012	Contract intensity of the money supply from the International Financial Statistics
Bauer, Cruz and Graham	IMF Agreement	188	1960–2008	Whether countries are under an IMF agreement in a given year
Beck, Demirgüç-Kunt, and Levine	Financial Development and Structure	175	1960–2011	Size, activity, and efficiency of financial intermediaries and markets
Levy-Yeyati and Sturzenegger	De facto exchange rate regime	183	1974–2004	De facto exchange rate regimes
Graham, Johnston, and Kingsley	LexisNexis-based information Measures	171	1992–2012	Number of business and investment relative articles written about a country in a given year (English)
World Bank	Newspaper Circulation Data	192	1961–2010	Data on newspaper circulation
Graham, Johnston, and Kingsley	I/B/E/S Analyst Coverage [IB]	94	1993–2015	Number of unique I/B/E/S analysts covering firms based in a country in a given year.
Forbes and Warnock		58	1973–2009	Surges, stops, flights and retrenchments in capital flows.

Table 1 (continued)

Institution/ Authors	Dataset Name	# of countries and years	Topic
	Surges, Stops, Flight and Retrenchment Data		
UNCTAD (UN Conference on Trade and Development)	FDI flows & stocks	197	1970–2014 Inward FDI flows and stocks as reported to the UNCTAD
U.S. Bureau of Economic Analysis	Outward FDI Stocks, Historical Cost Basis	191	1982–2014 Outward FDI stocks from the U.S.
U.S. Bureau of Economic Analysis	Outward FDI Flows, by industry	59	1999–2014 Outward FDI flows from the U.S., separated by industry.
U.S. Bureau of Economic Analysis	Outward FDI flows for PPE	196	1997–2013 Outward FDI flows from the US for net property, plant and equipment
US Department of the Treasury	List of Tax Havens	211	1983–2015 List of know tax havens
Ashford	Oil Exports Data	188	1960–2001 Data on oil revenues generated from the WDI, BP Statistical Index, and the Energy Information Administration
Feenstra, Inklaar and Timmer	PENN World Tables 8.0	166	1950–2011 A range of economic indicators
World Bank	World Development Indicators	195	1960–2014 Information about national economic development such as food exports, interest payments on external debts, armed forces personnel.
IPE Data Resource	Augmented WDI Data	161	1950–2011 GDP Total, GDP/Capita, GDP Growth, Population, labor/capital ratio
Organisation for Economic Co-operation and Development	OECD Statistics	34	2010–2014 Long-term interest rates
International Labor Organization	ILOSTAT	201	2000–2013 Unit labor cost
Li	Tax incentives	202	Measures of certain types of tax incentives
World Bank	Doing Business Indicators	181	2004–2013 Variables that measure barriers to starting a business within each country
World Bank	Global Financial Development Database	197	1960–2014 Characteristics of the banking sector within each country.

Table 1 (continued)

Institution/ Authors	Dataset Name	# of countries and years	Topic
World Bank	MFN Tariff Data	183	MFN tariff rates
Political Datasets			
Alvarez, Cheibub, Limongi, and Przeworski	Regime Type	135	Regime variables that classify each country's government type
Marshall and Gurr	Polity IV	184	Information about political regimes
Coppedge et al.	Varieties of Democracy	170	Indicators of democracy
Beck, Clarke, Groff, Keefer, et al.	Database of Political Institutions	176	Information about the party system, political parties, and political orientations within each country.
Geddes, Wright, and Frantz	Autocratic Regimes	120	Information about political regimes and regime failures
Boix, Miller, and Rosato	Regime Type	211	Binary democracy measure
Boix, Miller, and Rosato	System Level Democracy	194	These system-level democracy measures are built from the Boix et al. data.
Strom, Gates, Graham and Strand	Powersharing	175	Underlying political structures and their ability to support credible policy commitments
Linzer and Staton	Measure of Judicial Independence	198	Latent measure of judicial independence.
La Porta et al.	Judicial Checks and Balances	71	Measures of judicial independence and judicial review
Kaufmann, Kraay, and Mastruzzi	Worldwide Governance Indicators	198	Measures of political stability, government accountability, and rule of law within a country
Marshall and Cole	State Fragility Index	166	Stability of each country based on social, economic, and political factors.
Bertelsmann Stiftung	Sustainable Governance Indicators	41	Measures of sustainable governance practices.
World Justice Project	Rule of Law Index	102	Measurements on how the rule of law is experienced by the general public.
Bertelsmann Stiftung	Transformation Index	130	Indicators of the quality of democracy, a market economy and political management in developing and transitioning countries.

Table 1 (continued)

Institution/ Authors	Dataset Name	# of countries and years	Topic
World Bank	CPIA Database	197	2005–2015 Measures of the extent to which private economic activity is facilitated by an effective legal system.
Graham	Bureaucratic Risk and Policy Risk	205	1988–2013 Political risk indices
Cingarella and Richards	CIRI Human Rights dataset	200	1981–2007 Government adherence to 13 internationally recognized human rights
Fariss	Latent Human Rights Protection	180	1949–2013 Information about several different measures of human rights protection
Mosley and Uno	Collective Labor Rights	90	1985–2002 Indicators of labor rights for developing nations
Reporters Without Borders	Freedom of the Press	177	2012–2014 Alternate freedom of the press measure
Freedom House	Freedom of the Press	198	1993–2012 Measures of freedom of the press
Hollyer, Rosendorff, and Vreeland	Transparency	125	1980–2010 HRV index value for each country in a given year
Transparency International	Corruption Perceptions	216	1946–2012 Transparency International corruption index
Monty G. Marshall	Major Episodes of Political Violence	176	1946–2010 Information on major wars and episodes of political violence in a country
Northaus, Oneal, and Russett	MID Propensity	165	1951–2001 Level of military spending in each country and its effect on democracy in that country.
Fortna	UN Peacekeeping	30	1958–2004 Information on the presence of UN Peacekeeping Forces
Cornett, Wood and Haschke	Political Terror Scale	192	1976–2014 Measures of political violence and terror
Cornett, Gibney and Haschke	Societal Violence Scale	194	2013 Measures of societal violence
Gleditsch, Wallensteen, Eriksson, Sollenberg, and Strand	UCDP/PRIO Armed Conflict	118	1946–2014 Measures of conflict intensity and involvement
Lacina and Gleditsch	PRIO Battle Deaths Dataset	117	1946–2008 Annual battle deaths
Uppsala Conflict Data Program	UCDP Battle Deaths Dataset	94	1989–2014 Annual battle deaths
Strand (PRIO)	Onset of Armed Conflict	176	1946–2010 List of armed conflict onsets

Table 1 (continued)

Institution/ Authors	Dataset Name	# of countries and years	Topic
Park	Patent protection Index	122	1960–2010 Index of strength of patent protection and patent rights in each country
Zhao	Strength of Property Protection	47	1993–2001 Cross-sectional measure of strength of property rights protection
Schwab (World Economic Forum)	Global Competitiveness Dataset	39	2004–2010 Government transparency
Multiple Sources	Bilateral Investment Treaties	190	1959–2008 Number of bilateral investment treaties signed by each country
IPE Data Resource	Membership in WTO, IMF, EU, and NATO	222	1946–2013 Dummy variable specifying which international organization each country belongs to
Social and Cultural Datasets Barro and Lee	Educational Attainment	146	1950–2010 Information about the years of schooling, primary, secondary, tertiary for population 25 and over
Organisation for Economic Co-operation and Development	Program for International Student Assessment Scores	70	2000–2012 Scores from an international assessment of 15-year-old students' literacy in reading, mathematics, and science.
World Bank	Education Statistics	197	1970–2015 General education statistics
Wimmer, Cederman, and Min	Ethnic Power Relations	165	1946–2013 Ethnic groups and their political sway in each country
UN Office on Drugs and Crime	Crime Data	196	2000–2014 Measures of the levels of crime.
NationsOnline	Languages	250	1800–2013 First two official languages in spoken in each country
Brown and James	Religious Characteristics of States	198	1815–2010 Estimates of religious demographics
United Nations	UN Emigrant Stock Data	188	1990, 2000, 2010, 2013 Global stock of emigrants from each country
Geographic Datasets	Archipelagos Data	58	N/A Number of archipelagos
Graham and Strom	Bilateral distance to U.S.	222	1800–2015 Geographic distance, linguistic ties and historical ties to U.S.
Mayer and Zignago (CEPII)			

References

- Acemoglu, D., & Robinson, J. A. (2006). De facto political power and institutional persistence. *The American Economic Review*, 96(2), 325–330.
- Alston, L. J. (1996). Empirical Work in Institutional Economics: An Overview. In L. J. Alston, T. Eggertsson, & D. C. North (Eds.), *Empirical Studies in Institutional Change* (Vol. 1996, pp. 25–30). Cambridge University Press.
- Barro, R. J., & Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104, 184–198.
- Bennett, D. S., & Stam, A. C. (2000). EUGene: A Conceptual Manual. *International Interactions*, 26, 179–204.
- Bennett, D. Scott and Poast, Paul and Stam, Allan C. 2015. “NewGene: A conceptual manual.” Available at SSRN: http://ssrn.com/abstract_id=2596297
- Brown, Davis, and Patrick James. 2015. “Religious Characteristics of State Dataset, Phase 1: Demographics.” <http://www.thearda.com/Archive/Files/Descriptions/BROWN.asp>. Accessed on September 15, 2015.
- Coppedge, M., Gerring, J., Lindberg, S. I., Skaaning, S-E., Teorell, J., Andersson, F., et al. (2016). Methodology V6. Available at SSRN: <https://ssrn.com/abstract=2951040>.
- Cruz, C., Keefer, P., & Scartascini, C. (2016). *The Database of political institutions 2015 (DPI2015)*. Inter-American Development Bank <https://publications.iadb.org/handle/11319/7408>.
- Fariss, C. J. (2014). Respect for human rights has improved over time: Modeling the changing standard of accountability. *American Political Science Review*, 108(02), 297–318.
- Fariss, C. J., Benjamin A.T. Graham, and Michael Kenwick. 2016 “Secure from what? Theorizing and measuring property rights security.” Working Paper. <https://ssrn.com/abstract=2989195>.
- Fouré, J., Bénassy-Quéré, A., & Fontagné, L. 2012. “The Great Shift: Macroeconomic Projections for the World Economy at the 2050 Horizon,” CEPII Working paper 2012–03. <http://www.cepii.fr/CEPII/en/publications/wp/abstract.asp?NoDoc=4179>
- Girardin, L., Hunziker, P., Cederman, L.-E., Bormann, N.-C., et al. (2015). *GROWUP - geographical research on war*. ETH Zurich: Unified Platform <http://growup.ethz.ch/>.
- Gleditsch, K. S., & Ward, M. D. (1999). Interstate system membership: A revised list of the independent states since 1816. *International Interactions*, 25(4), 393–413.
- Gleditsch, N. P., Wallensteen, P., Eriksson, M., Sollenberg, M., & Strand, H. (2002). Armed conflict 1946–2001: A new dataset. *Journal of Peace Research*, 39(5), 615–637.
- Hollyer, J. R., Peter Rosendorff, B., & Vreeland, J. R. (2014). Measuring transparency. *Political Analysis*, 22(4), 413–434.
- Humphreys, M., de la Sierra, R. S., & der Windt, P. V. (2013). Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. *Political Analysis*, 21(1), 1–20.
- Kingsley, Allison and Benjamin A.T. Graham. 2017. “The capital effects of information voids in emerging markets.” *Journal of International Business Studies* 48, 324.
- Laitin, D. D. (2013). Fisheries management. *Political Analysis*, 21(1), 42–47.
- Linzer, D. A., & Staton, J. K. (2015). A Global measure of judicial independence, 1948–2012. *Journal of Law and Courts*, 3(2), 223–256.
- Marshall, M. G. (1999). *Third world war*. Lanham, MD: Rowman & Littlefield Publishers.
- Marshall, Monty and Benjamin R. Cole. 2014. “Global report 2014: Conflict, governance, and State fragility.” Center for Systemic Peace. <http://www.systemicpeace.org/vlibrary/GlobalReport2014.pdf>
- Mayer, T., and S. Zignago. 2011. “Notes on CEPII’s distances measures: the GeoDist Database.” CEPII Working Paper 2011–25. <http://www.cepii.fr/anglaisgraph/workpap/pdf/2011/wp2011-25.pdf>.
- Monogan III, J. E. (2013). A case for registering studies of political outcomes: An application in the 2010 House elections. *Political Analysis*, 21(1), 21–37.
- Teorell, Jan, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Anna Khomenko, and Richard Svensson. 2016. The quality of government standard dataset, version Jan16. University of Gothenburg: The quality of government Institute, <http://www.Qog.Pol.Gu.Se> doi:10.18157/QoGStdJan16.
- Tollefsen, A. F., Strand, H., & Buhaug, H. (2012). PRIO-GRID: A unified spatial data structure. *Journal of Peace Research*, 49(2), 363–374.

Reproduced with permission of copyright owner.
Further reproduction prohibited without permission.