International Relations Database

Alden Felix

EPPS 6354

May 4, 2023

Contents

Abstract 3

Introduction 4

Data and Methods 5

Conclusion and Future Improvements 9

References 10

Abstract

Studies in international relations often involve numerous variables that cannot be found within a single dataset. Therefore, this project seeks to provide a high quality yet diverse database that will aid relevant studies in their data collection process. Data is collected from multiple credible sources, wrangled, and stored in a PostgreSQL server using R, and then deployed to the cloud in the form of an R Shiny application as proof of concept. The dataset currently includes coded event data, countries and their socioeconomic and demographic data, and dyadic variables such as trade between a pair of states, all for the period from 2015 to 2019.

Introduction

The purpose of this database project is to help fill in a gap in the existing international relations literature surrounding affinity communities. While conflict has been studied extensively, and predicting conflict through machine learning and big data is attracting attention, the same is not true for the study of network effects on global relationships and investigating the origin of these networks. This database will directly aid a machine learning project I am involved in that seeks to predict interstate affinity using socioeconomic and demographic variables. The data gathering process for the project revealed that the data we are seeking is currently scattered throughout multiple unrelated databases and contains many variables that are not of interest. With this database, the data collection process will be made much easier for our study and similar studies in international relations.

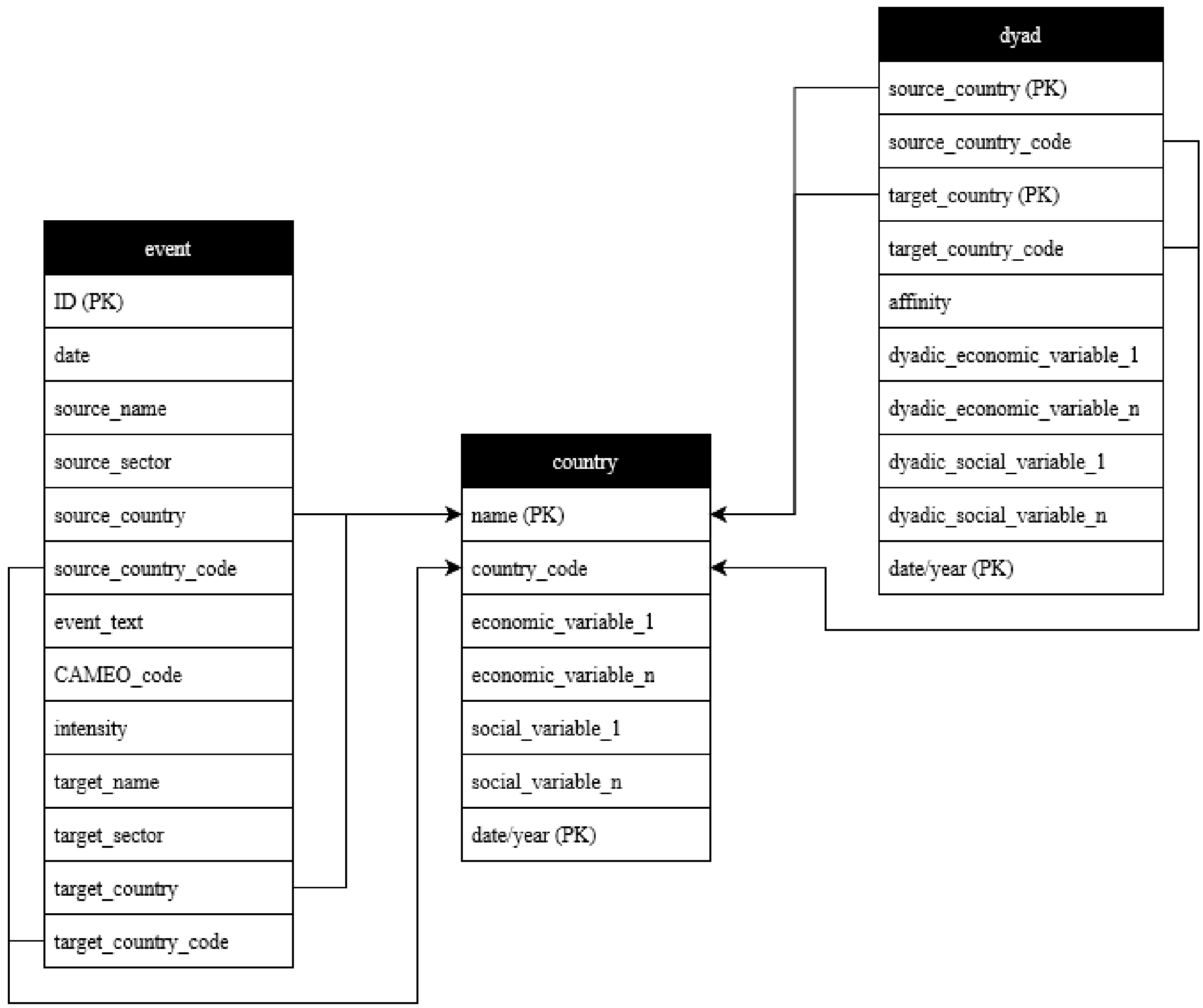
Studies on why relationships between states are formed and how these relationships are affected are significant to many actors, including states, international organizations, businesses, and individuals, and for different reasons. This might explain why predicting conflict, one extreme of a relationship, has become such a large field. Where this database fits in is the selection of model variables. Existing literature highlights the importance of variable selection in building accurate and well performing models (Perry 2013), but the variables used in literature vary greatly. In addition, the quality and form of variables found in publicly available sources is not homogenous, leading to the data collecting process becoming a hurdle for studies. These are all issues that this project aims to solve.

Data and Methods

The event data included in the database comes from the Integrated Crisis Early Warning System's (ICEWS) publicly available coded event data. The large number of events makes working with the data computationally intensive, especially when studying international relations across extended time periods. Therefore, storing the events in an easily accessible database and querying only the needed data becomes essential. The other socioeconomic and demographic data comes from various sources, including the World Banks's World Development Indicators, Education Statistics, Gender Statistics, Health Nutrition and Population Statistics, and Doing Business data, and the International Monetary Fund, and Freedom House. The currently included variables were selected based on existing literature to be diverse and significant to international relations studies and can always be modified in the future.

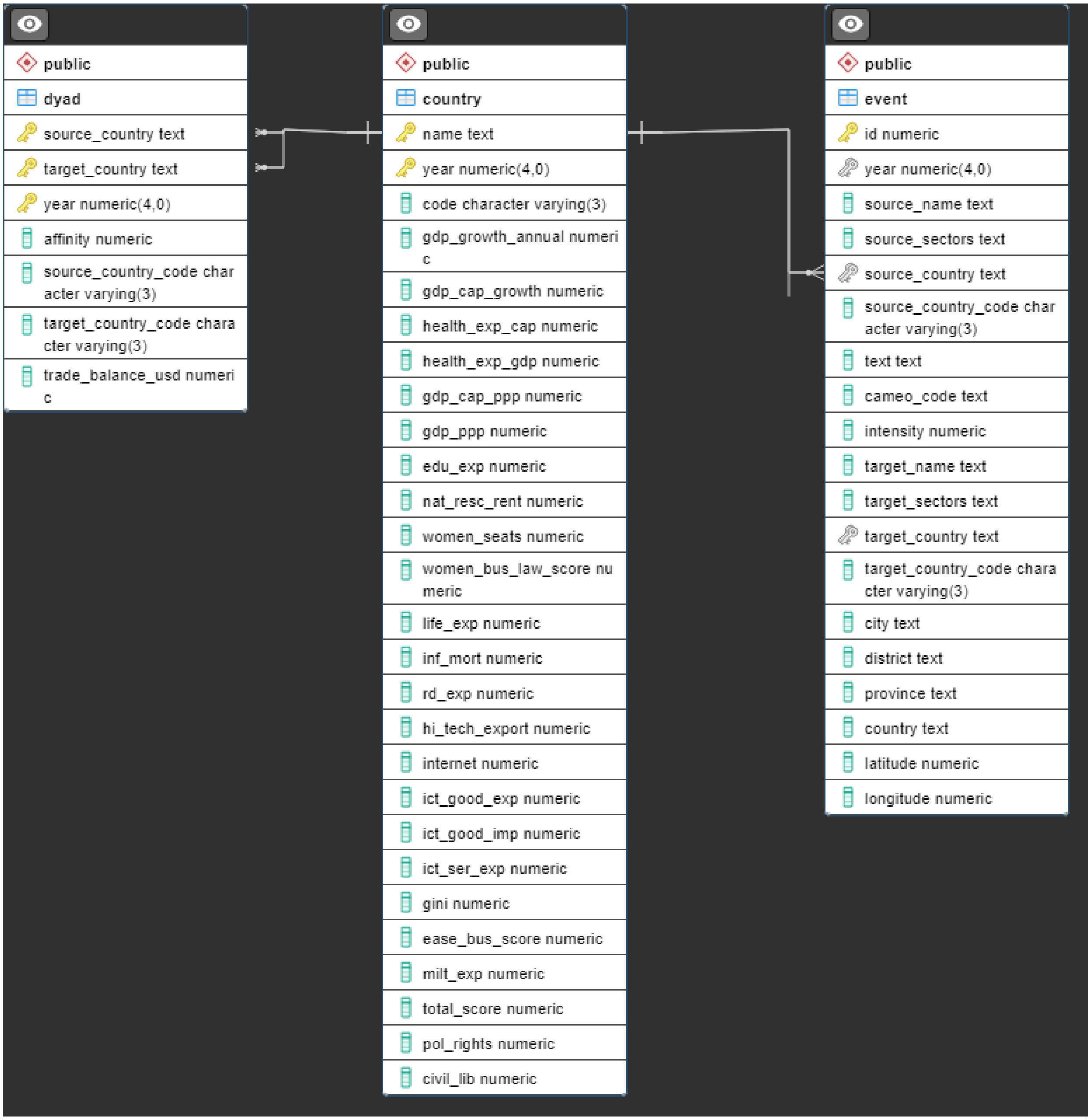
The data was manually imported into R studio and wrangled into appropriate data types and format before being exported into a local PostgreSQL server using the 'DBI' library. A relation schema was made beforehand and empty tables were created to store the data. The initial relation schema can be seen below.

Initial Relation Schema



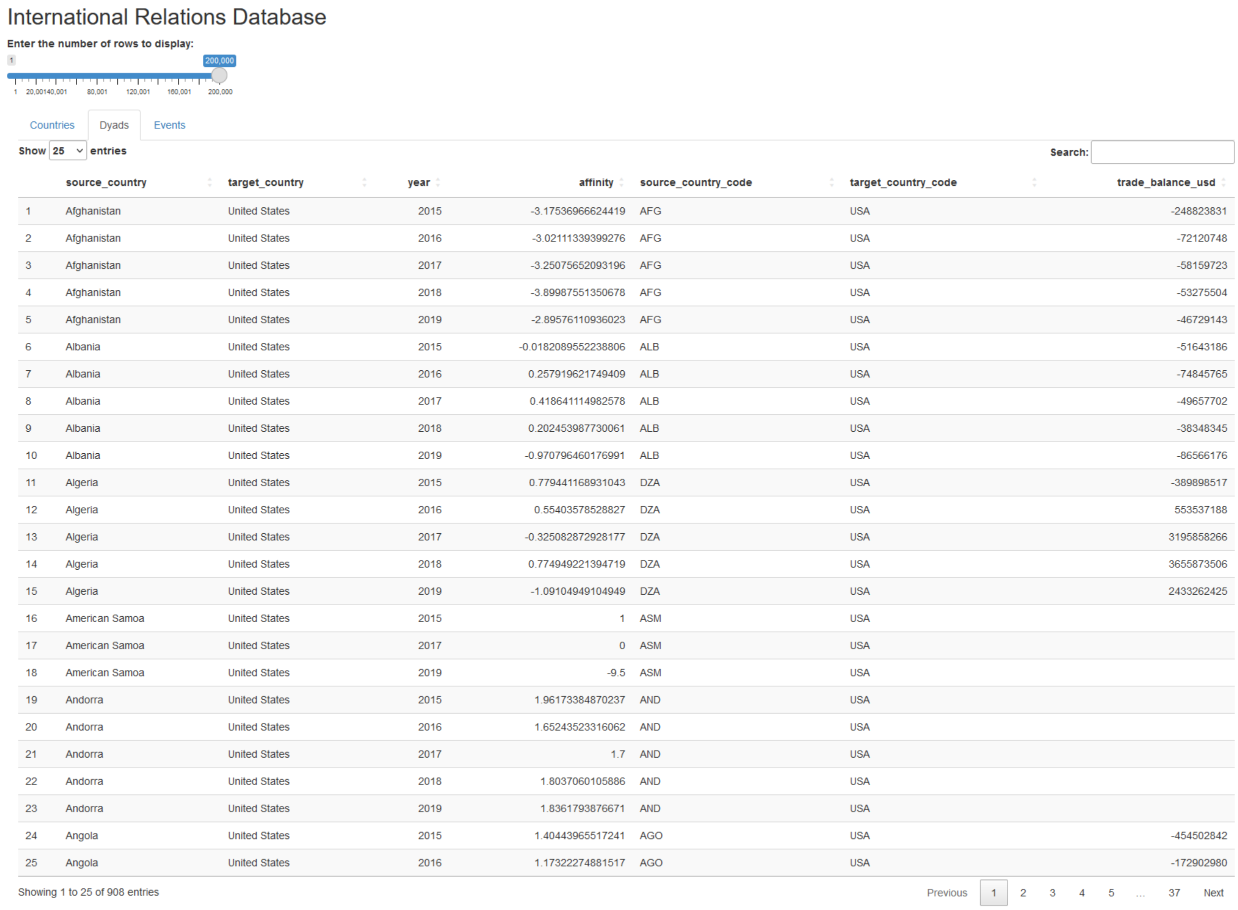
As seen in the schema, there are three tables in total. One for the event data, one to store the socioeconomic and demographic variables for each state for each year, and finally a table to store all state dyads and their relevant variables. The primary keys are represented by the '(PK)', and foreign keys are represented by variables that are connected to an attribute in another table by an arrow, so the country table does not have any foreign keys. The final relation schema can be seen below with the current list of attributes.

Final Relation Schema



The primary keys in the final relation schema are represented by a single yellow key while foreign keys are represented by multiple grey keys and a connection to an attribute in another table.

Once the data was stored correctly in the local PostgreSQL server, it was converted into SQLite format using DB Browser for SQLite. However, because of errors due to syntax variations between PostgreSQL and SQLite the primary keys and foreign keys were removed for the proof-of-concept cloud deployment. An app was built using R Shiny that allows basic searching and viewing of the database. An image of the application can be seen below. There is a slider to limit the total number of tuples that are pulled from the database, along with row limits and pagination.


International Relations Database

Enter the number of rows to display:

1 200,000

1 20,001 40,001 80,001 120,001 160,001 200,000

Countries Dyads Events

Show 25 entries Search:

	source_country	target_country	year	affinity	source_country_code	target_country_code	trade_balance_usd
1	Afghanistan	United States	2015	-3.17536966624419	AFG	USA	-248823831
2	Afghanistan	United States	2016	-3.02111339399276	AFG	USA	-72120748
3	Afghanistan	United States	2017	-3.25075652093196	AFG	USA	-58159723
4	Afghanistan	United States	2018	-3.89987551350678	AFG	USA	-53275504
5	Afghanistan	United States	2019	-2.89576110936023	AFG	USA	-46729143
6	Albania	United States	2015	-0.0182089552238806	ALB	USA	-51643186
7	Albania	United States	2016	0.257919621749409	ALB	USA	-74845765
8	Albania	United States	2017	0.418641114982578	ALB	USA	-49657702
9	Albania	United States	2018	0.202453987730061	ALB	USA	-38348345
10	Albania	United States	2019	-0.970796460176991	ALB	USA	-86566176
11	Algeria	United States	2015	0.779441168931043	DZA	USA	-389898517
12	Algeria	United States	2016	0.55403578528827	DZA	USA	553537188
13	Algeria	United States	2017	-0.325082872928177	DZA	USA	3195858266
14	Algeria	United States	2018	0.774949221394719	DZA	USA	3655873506
15	Algeria	United States	2019	-1.09104949104949	DZA	USA	2433262425
16	American Samoa	United States	2015	1	ASM	USA	
17	American Samoa	United States	2017	0	ASM	USA	
18	American Samoa	United States	2019	-9.5	ASM	USA	
19	Andorra	United States	2015	1.96173384870237	AND	USA	
20	Andorra	United States	2016	1.65243523316062	AND	USA	
21	Andorra	United States	2017	1.7	AND	USA	
22	Andorra	United States	2018	1.8037060105886	AND	USA	
23	Andorra	United States	2019	1.8361793876671	AND	USA	
24	Angola	United States	2015	1.40443965517241	AGO	USA	-454502842
25	Angola	United States	2016	1.17322274881517	AGO	USA	-172902980

Showing 1 to 25 of 908 entries

Previous 1 2 3 4 5 … 37 Next

The database was limited to only dyads and events with the US as the target country because of concerns about cloud computing resources, so there are only 1,000 tuples in the country table, 1,000 tuples in the dyads table, and 150,000 events in the events table.

Conclusion and Future Improvements

This database is an initial step in aiding research in international relations and highlights the importance of publicly available and easily accessible high-quality data. The local PostgreSQL server in its current state serves as an excellent example of this as it greatly aided my machine learning study that predicts interstate affinity. However, the currently deployed Shiny application has the potential for significant improvements. Foremost, the absence of primary and foreign keys due to inconsistent syntax between PostgreSQL and SQLite should be addressed to ensure normal functioning of the database before any other changes are made. After this is achieved, more attributes can be added to enhance the relevance of the database to international relations studies beyond those dealing primarily with interstate affinity.

Regarding the accessibility of the application as a method of initial data exploration, the filtering options can be expanded beyond what currently exists. For example, filtering by attribute and across tables is one much needed function. In addition, adding user defined plots such as histograms and scatterplots that are dynamically adjusted according to selected data would aid exploratory data analysis.

References

Kuhn, & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Halterman, A., Bagozzi, B. E., Beger, A., Schrodt, P., & Scraborough, G. (2023, April 3). PLOVER and POLECAT: A New Political Event Ontology and Dataset. https://doi.org/10.31235/osf.io/rm5dw

Helle, V., Negus, A.S., & Nyberg, J. (2018). Improving armed conflict prediction using machine learning : ViEWS+. Pauls, & Cranmer, S. J. (2017). Affinity communities in United Nations voting: Implications for democracy, cooperation, and conflict. *Physica A*, *484*, 428–439. https://doi.org/10.1016/j.physa.2017.04.177

Perry. (2013). Machine Learning and Conflict Prediction: A Use Case. *Stability (Norfolk, VA)*, *2*(3), 56–. https://doi.org/10.5334/sta.cr

Bazzi, Blair, R. A., Blattman, C., Dube, O., Gudgeon, M., & Peck, R. (2022). The Promise and Pitfalls of Conflict Prediction: Evidence from Colombia and Indonesia. *The Review of Economics and Statistics*, *104*(4), 764–779. https://doi.org/10.1162/rest_a_01016