

EPPS 6354 Information Management Progress Report

Alden Felix

April 13, 2023

Overall Database Purpose

The purpose of this database project has not changed from the initial proposal. The purpose is to help fill in a gap in the existing literature surrounding affinity communities and directly aid a machine learning project I am involved in that seeks to predict interstate affinity using multiple socioeconomic and demographic variables.

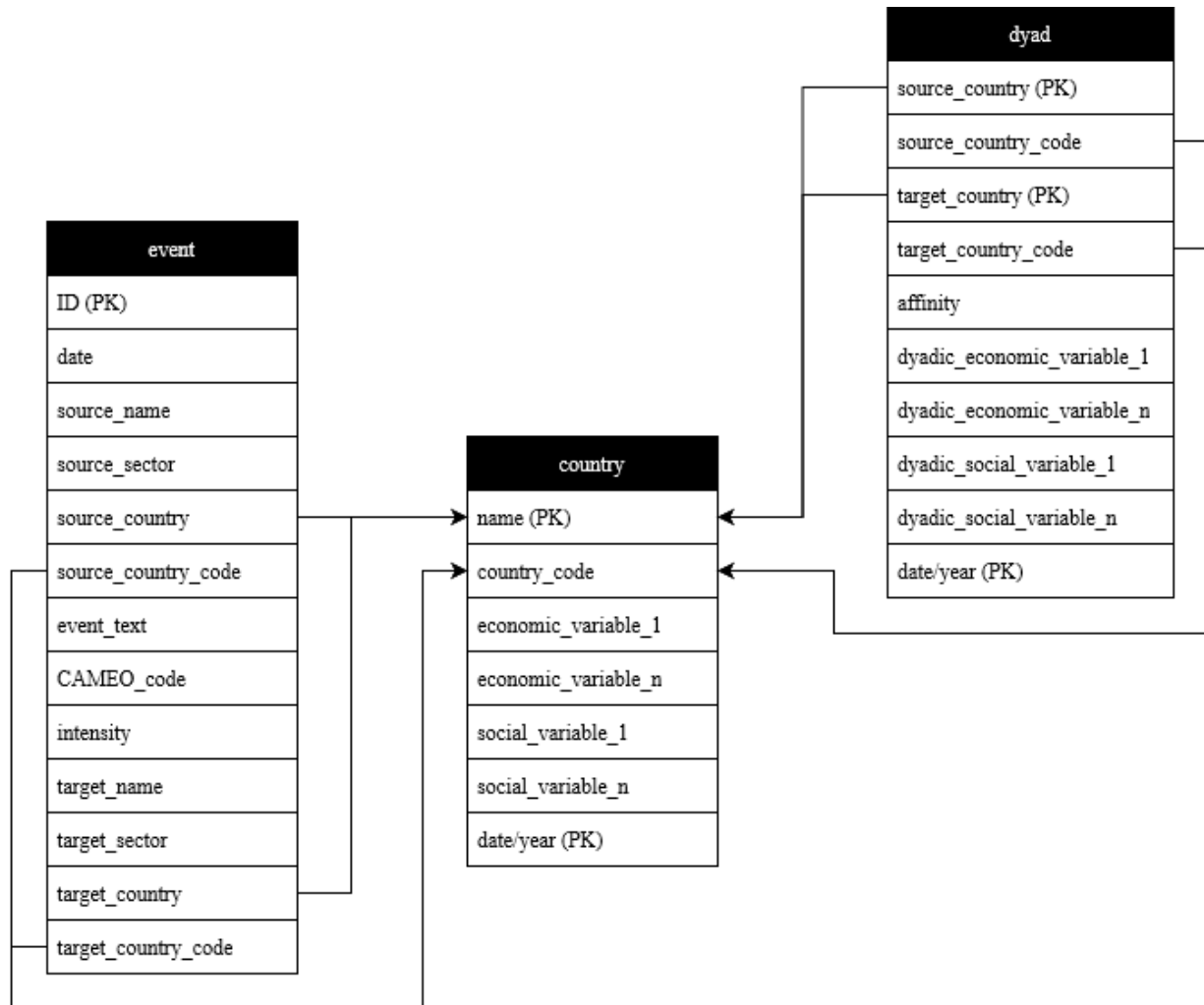
Schema

The machine learning project has progressed to a point where the relevant variables for the study have been decided on and the data has been gathered and wrangled into a form that can be used to train a model. Therefore, the relation schema of the database is also finalized for the most part. Variables can still be added or removed as needed but I do not see the primary key variables changing at this point. The initial relation schema and the finalized schema can be seen below.

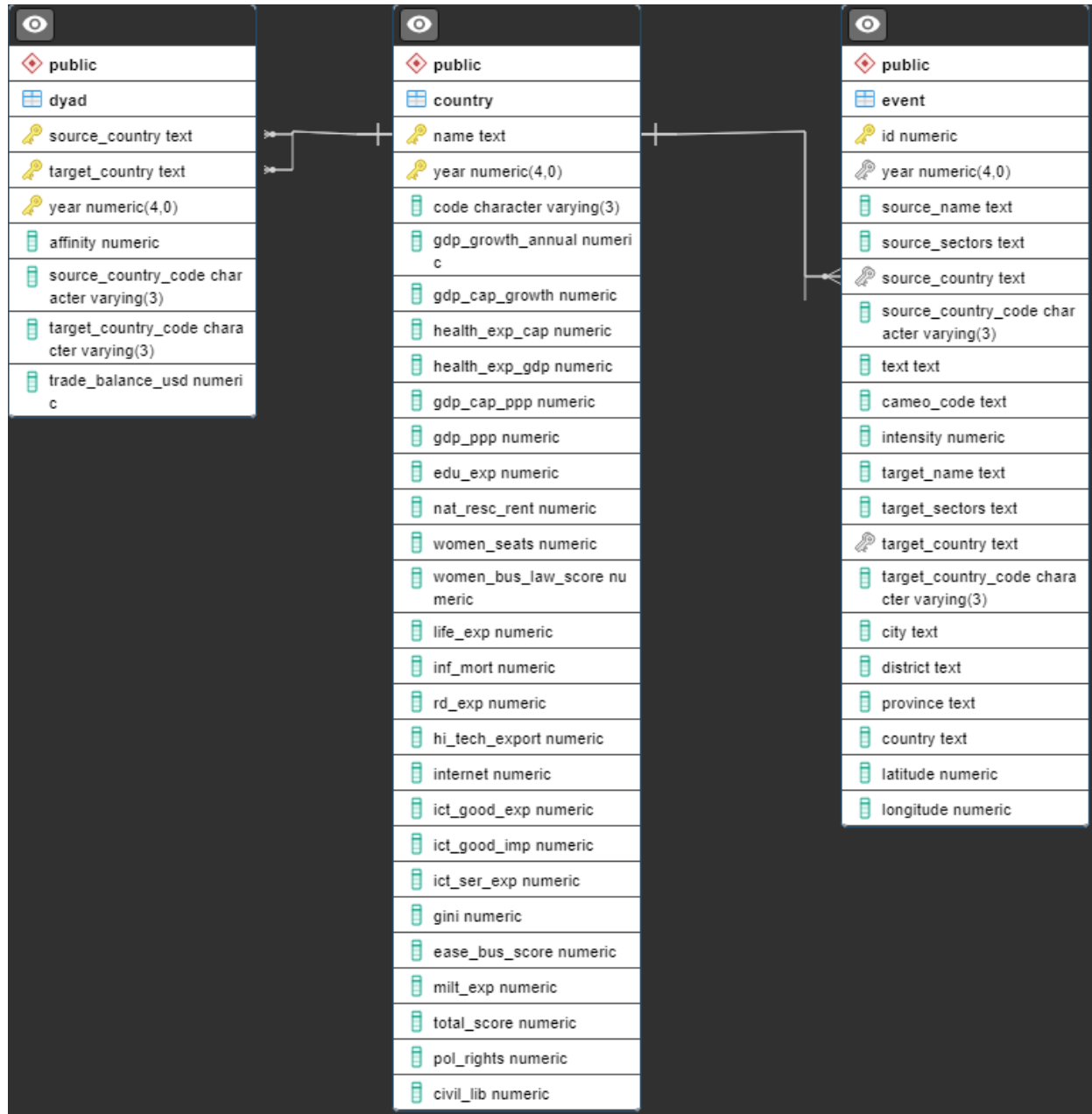
Database

The database has also been populated with all the variables, except for the event table which currently only contains events where the US is the target country. This is for the sake of keeping the database manageable in size, as there are over 3 million events across all time periods in the study.

Initial Database Relation Schema



Final Database Relation Schema



Work in Progress

With the finalized schema and all the data stored, the Shiny application which will serve as the face of the database can be worked on. The current idea for the application is a simple table viewer that will have multiple filters to make the specific data the user wants easy to view.