17.835 Project Proposal
Luann Jung, John Posada, Noah Dasanaike

We will be producing a model that uses various input variables to forecast the chance of instability at a national level, allowing for qualitative analysis and case studies on how different input variables produce different outcomes. Potentially, depending on feasibility and run-time, we may also produce a forecast n-days into the future by running the model on a daily-basis n-days into the past, then simply running a regression analysis.

Data will be collected by an input-variable basis, such that predetermined indicators of violence such as demographics, historical tendency toward violence, economic factors, and state characteristics (i.e. authoritarian, democratic, etc.) will likely come from varying sources. Historical data of conflict occurrence will provide a baseline upon which our indicators will build; databases such as GDELT allow for computation of n-events of x-type over y-period of time. We will likely be using R or Python depending on the availability of certain APIs, with R being the most likely for merging purposes and Python for model-building. Organizations such as the OECD and the World Bank provide key economic indicators while the United Nations Statistical Division maintains social and demographic data for member states.

Using the merged data, we aim to develop an ensemble model consisting of constituent models that each perform prediction using a subset of the variables. Essentially, we will be considering different factors to inform multiple predictions of national violence and instability, then combine these predictions into a resultant probability of conflict. Examples of constituent models that we may explore include regression on conflict history, modeling of natural resources or disease, or trends of civil unrest. Particularly, we will be building out machine learning models in Python to generate predictions from the input data. After attaining constituent model results, we aim to use some form of Bayesian Model Averaging to aggregate them. In addition, we plan to visualize our results by generating graphs that help to understand the interplay of important indicator variables and violence. For example, we could create heat maps comparing the densities of the most relevant predictor variables from our models to the density of historical violence within each country. Ultimately, the result of this project will be a series of models which, when combined, output a singular likelihood of national violence/instability for each nation considered.