

# STAT 527 Project Update

Joseph David

May 27 2022

I have decided to use the Armed Conflict Location and Event data (ACLED) specifically for Latin America. The data includes almost 200,000 observations of 29 variables, which I think will be too much so I will likely take data only from recent years and include only the variables for Location, Time, Civilians Present, and perhaps a few other.

I have not performed any data analysis yet, but I have taken notes from the project proposal feedback to change the types of regression I perform. I plan to use many different types of models to regress the number of fatalities on certain variables, including tree-based models, smoothing splines, kernel estimators, and boosting. With these models, I hope to perform ensemble learning with all of them. For each model, I intend to cross-validate in order to tune hyperparameters, such as number of knots for smoothing splines, tree depth, and kernel bandwidth.

Following the feedback, I see that using a Wilcoxon Rank Sum and Kolmogorov-Smirnov tests is not useful so I will not be doing that anymore.