

Project: Capstone Project 1: Project Ideas  
Daniel Loew

As I've discussed with my mentor, I'm primarily interested in using data science for social science projects that are in the service of the public good. With this in mind, I discovered three datasets that could be good sources for the first capstone project. They are detailed below.

1. The NYPD Complaint Data Historic dataset contains data on all valid misdemeanor, violation, and felony crimes reported by the New York Police Department between 2006 and 2018. It is openly supplied to the public through the NYC Open Data project at <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/ggea-i56i>. While reviewing the data dictionary, I saw that there were a few indicators of crime type, some of them more broad and some of them more specific. In the data dictionary, I saw that there was a fine crime code for hate crimes. I'd like to look at predictors of hate crimes and how they have changed in New York City since the 2016 election. I'm also interested in looking at predictors of cannabis crimes in New York City, as the NYPD's policing of cannabis crime and their related "stop and frisk" program has been very controversial. This policing tactic involved stopping thousands of young African-American and Latino males in New York City over many years. This program was not successful in stopping crime, and was judged in 2013 as being unconstitutional. It has been reported elsewhere that there is a large racial disparity in who gets arrested for cannabis and it is an ongoing discussion in civil rights, but I'd be interested in bringing the power of machine learning classification and regression to explore a larger picture for what predicts cannabis arrests.
2. I discovered a dataset of peace agreements throughout history (<https://www.kaggle.com/university-of-edinburgh/peace-agreements-dataset>), as well as a dataset on wars (<https://www.kaggle.com/umichigan/interstate-wars>). I would be interested in comparing these datasets to see if there is specific language or other factors that would predict whether certain wars end in peace agreements.
3. I also discovered a dataset on Chicago lobbying, that details what companies gave money to which politicians. I would be interested in getting a dataset of Chicago politicians' voting record to determine to try and run regression analyses to see if lobbying predicts voting records. The Chicago lobbying dataset can be found here: <https://www.kaggle.com/gfennema/chicago-lobbyist-data>