Capstone Project 2 – Milestone Report 1

My project attempts to predict coup attempts utilizing data that tracks a wide variety of variables, including leader characteristics, government types, political events in each country, as well as weather related data such as precipitation rates. This model will also determine help to determine which features of the model are the most important in determining whether a coup attempt will take place.

My project also attempts to analyze a specific subset of the data in order to predict coup attempt success rates. The data contains information regarding whether or not a coup was successful or not, and so I intend to utilize the model developed for predicting coup attempts to predict whether a particular coup attempt will be successful or not, and to determine whether the same features that are important in predicting coup attempts are the same characteristics that predict whether a coup attempt will be successful.

These are important questions to answer, as coup attempts are a particularly difficult phenomenon to predict since there is a large imbalance between positive and negative observations. Coup attempts are extremely rare, so predicting the likelihood can be extremely difficult, much like it is difficult to predict the likelihood of rare disease in a particular patient.

A model that enables us to predict the relative likelihood of coups would be of particular interest to organizations – both private and public – that analyze global risks, such as intelligence organizations like the CIA or insurance companies that insure assets held by clients in politically unstable countries.

For example, a mining company may make significant investments in a politically unstable country like the Democratic Republic of the Congo, where perhaps the greatest risk to those investments may be political in nature. A coup attempt may translate into a regime that is predisposed to renegotiating contracts or even seizing assets. Knowing the likelihood of such an event would be helpful for a company to more accurately quantify risks and would therefore be quite valuable.

The dataset itself comes from REIGN, the Rulers, Elections, and Irregular Governance dataset describes political conditions in every country each and every month. These conditions include the tenures and personal characteristics of world leaders, the types of political institutions and political regimes in effect, election outcomes and election announcements, and irregular events like coups, coup attempts and other violent conflicts. The latest dataset covers 200 countries, from 1950 to the present, and includes dozens of variables for each monthly snapshot. The dataset was created by Dr. Curtis Bell, and as of April 2019, it contains over 130,000 unique leader-month observations.

The data is directly available from the REIGN archive and downloads as a CSV file. Therefore, I easily pulled the data into a Jupyter notebook using Panda's "read_csv" importation method.

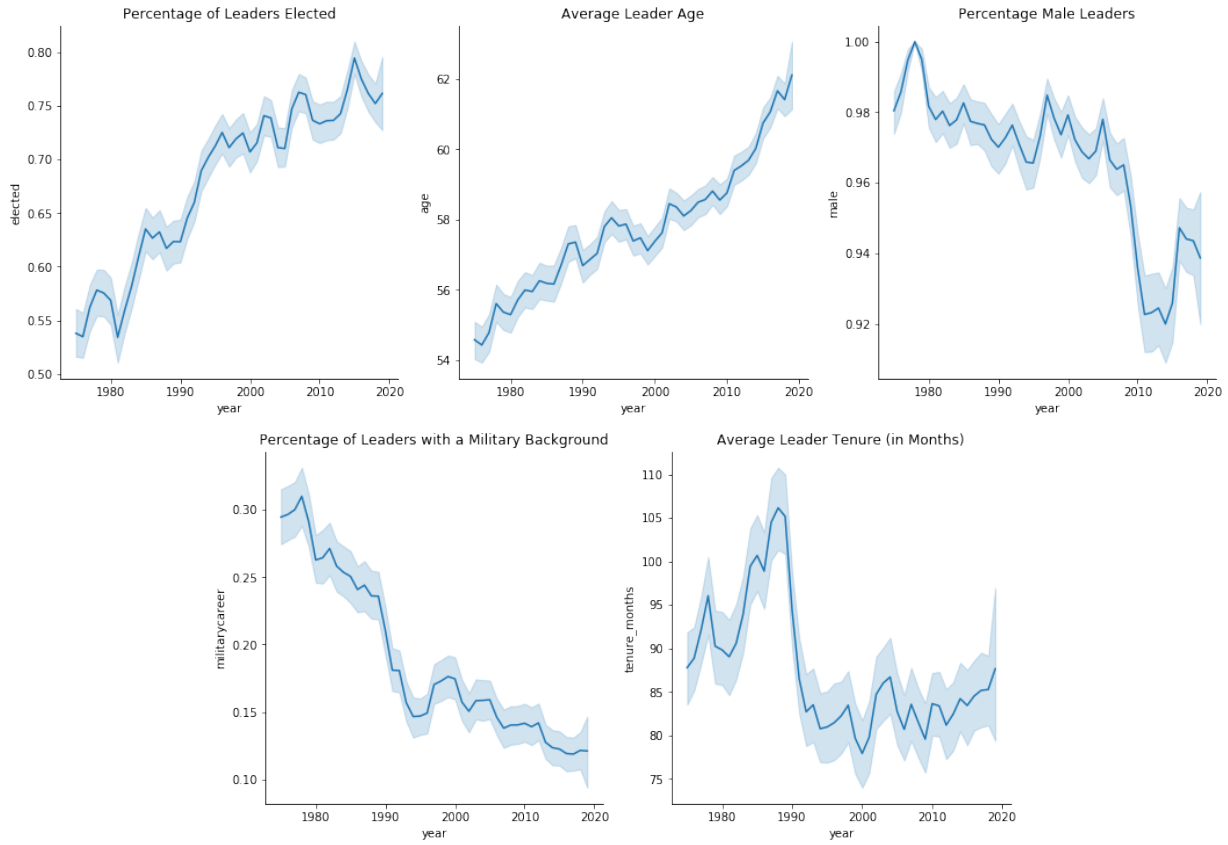
The first step in cleaning and wrangling the data was first, removing any missing values, and second, converting any character-based data into digestible, numerical values to be used by the model.

To accomplish the first step, I used the built-in "dropna" function, which removes any and all rows that even have a single missing value. This way, we would only be using complete rows of data in the model, increasing the accuracy and applicability of the model.

To accomplish the second step, I needed to convert the "government" feature to binary using one-hot encoding. I did this by utilizing the built-in "get_dummies" function, which served to split the single "government" feature into 16 separate features for each of the 16 distinct forms of government in the dataset.

After completing these two wrangling steps we are left with a dataset that has 51 separate and distinct features, containing 98,010 separate and complete observations (i.e. has values for each of the 51 features).

The next step in our analysis was to perform Exploratory Data Analysis (EDA). The first area of interest was to graphically represent some of the most important features in our dataset. In all we created 23 time series plots.



These 23 plots fall into 3 main categories: leader characteristics (of which there are 5), government types (of which there are 16), and coups (of which there are 2).

These graphs suggest several trends. First, there has been a distinct upward trend in the percentage of leaders elected. This seems to suggest that the world is becoming more democratic (even if the percentage of government types listed as democracies were not rising – which it is, as we will see).

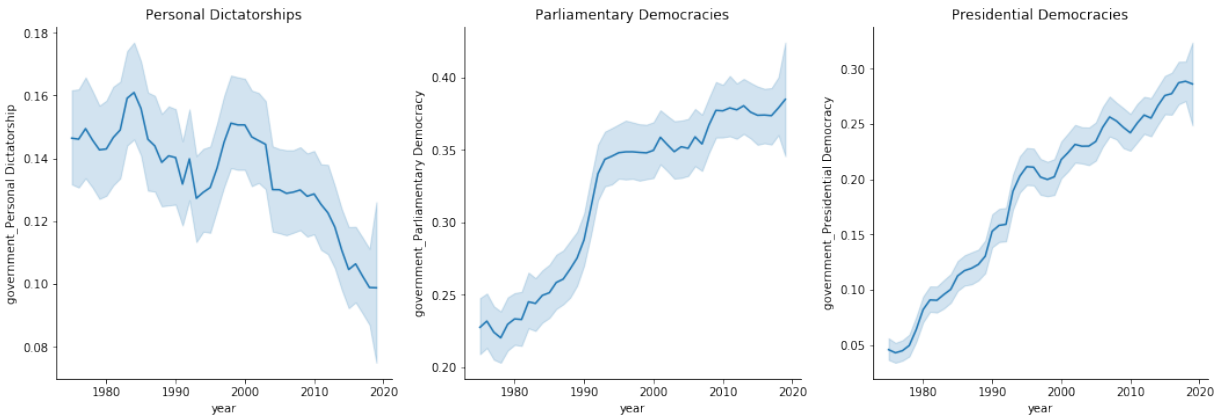
Second, the average age of a given leader has increased quite quickly over the past 70 years, from ~52 in 1950 to ~62 in 2019. This suggests either that the global median age has risen quite quickly, or that there has been a secular rise in the average age of leaders – perhaps as a result of the trends towards a larger percentage of elected leaders over that same time period.

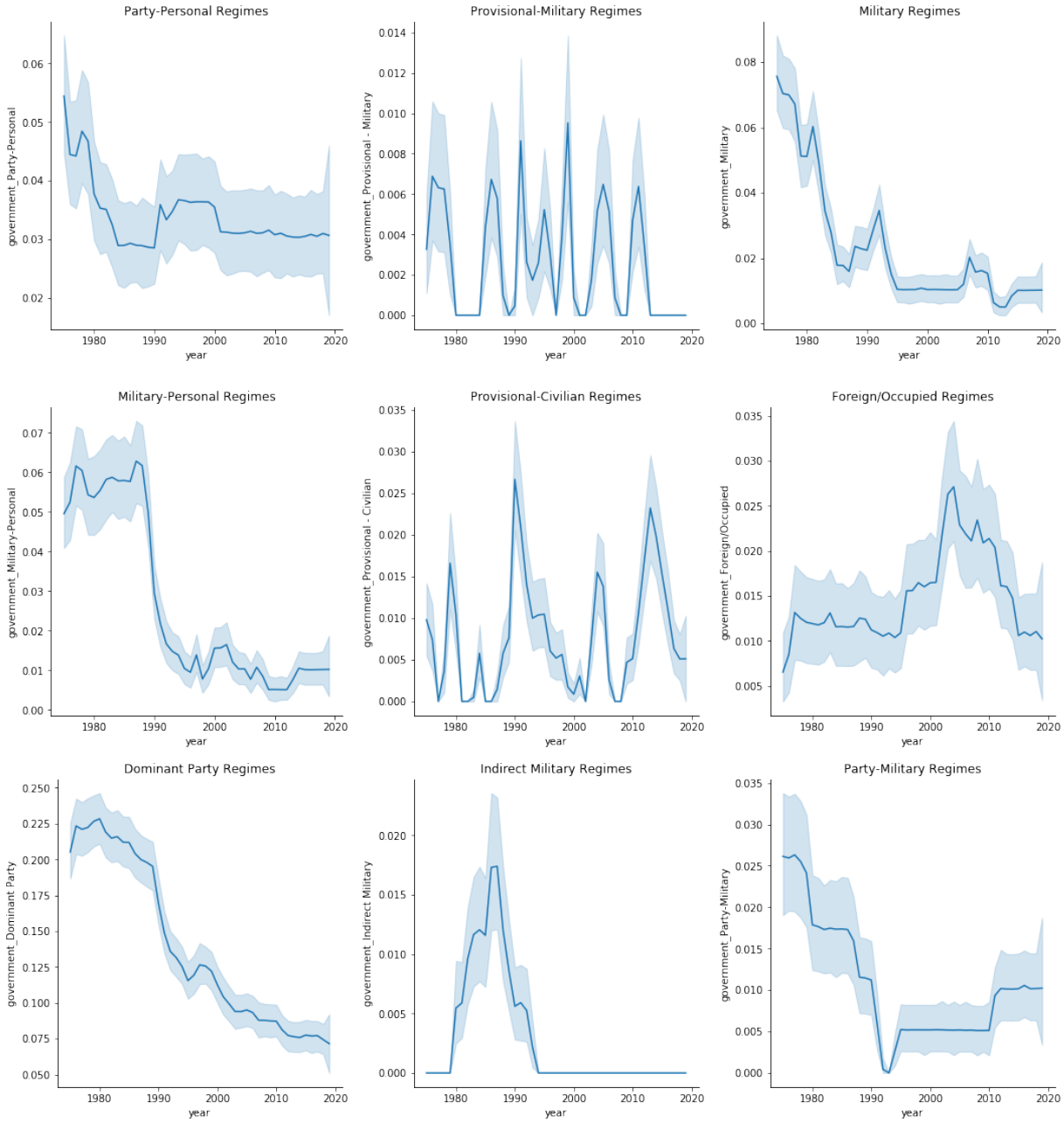
Third, the percentage of leaders that are male has declined over the past 70 years, at first slowly (from roughly 1950 to 2010), and then all of a sudden over the past decade (i.e. from 2010 to 2019). The trend has reversed a bit in recent years, but the overall trajectory is towards greater male/female parity. However, true parity (i.e. 50% male and 50% female) still seems a long way off – even the highest female leadership rate entailed a roughly 12 to 1 male leader to female leader ratio.

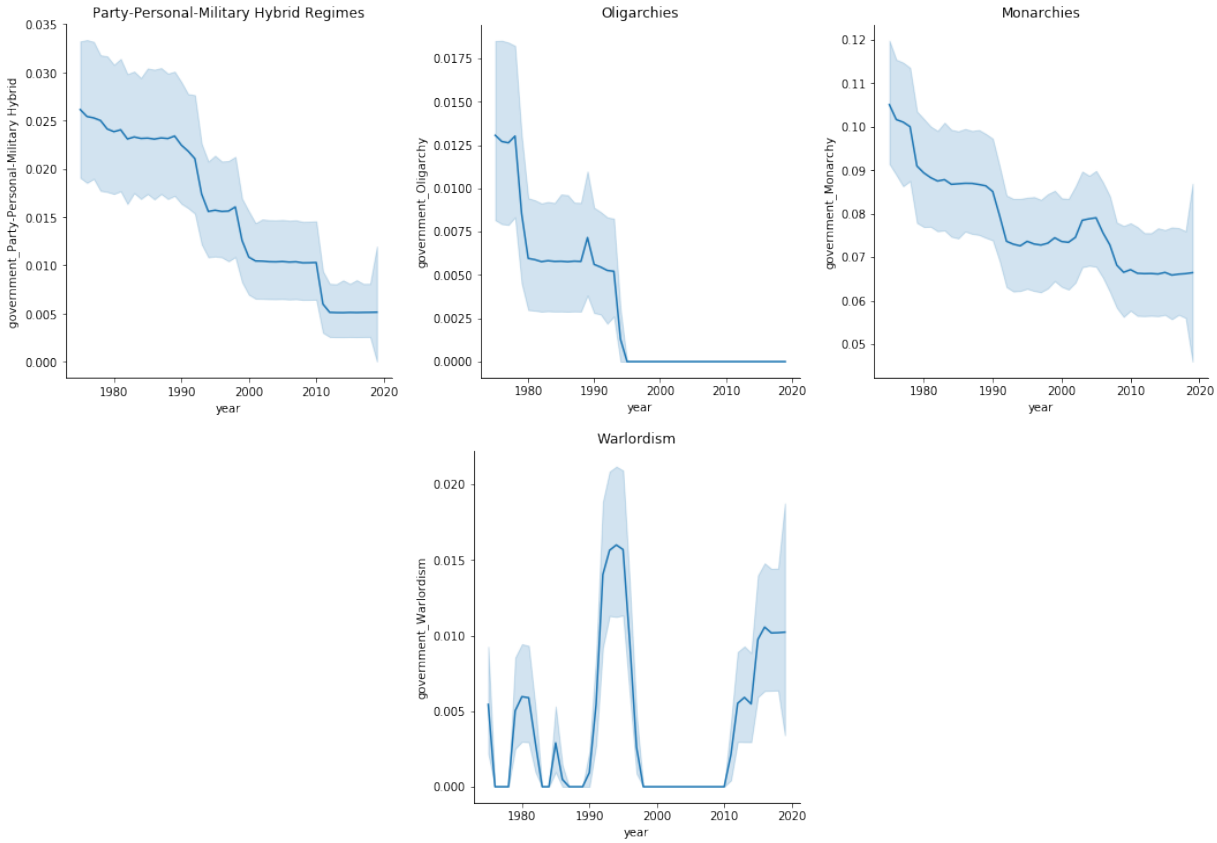
Fourth, the percentage of leaders with a military background has declined. This may go hand-in-hand with the increase in democratization we see. It is much more likely that as military regimes were replaced by civilian led governments, the leadership became less and less militarized. However, we do not expect that this trend will continue to zero, as many democratic societies (such as the United States) prize military service in a prospective leader.

The fifth and last trend we see is an initial drop and relatively stable tenure for leaders around the world. From 1950 until the mid 1980s we see a distinct increase in the average leader tenure, peaking at ~105 month (~9 years) only to see this trend reverse – precipitously – in the 1990s, whereby it has since stabilized around 82 months (~7 years).

The next group of plots includes government types, definitions of which were provided by the REIGN dataset administrators:



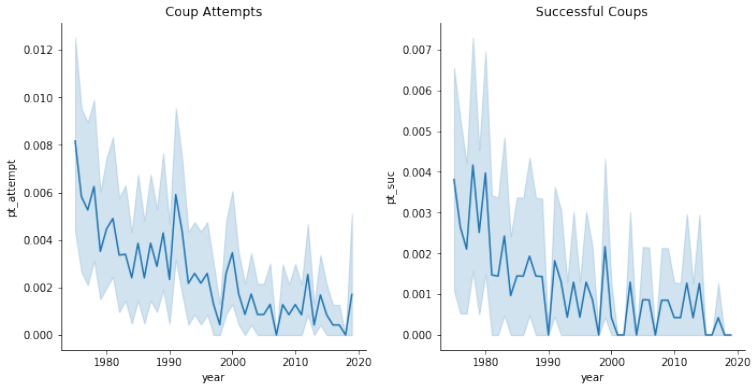

Party-Personal Regimes
government_Party-Personal
year
Provisional-Military Regimes
government_Provisional - Military
year
Military Regimes
government_Military
year
Military-Personal Regimes
government_Military-Personal
year
Provisional-Civilian Regimes
government_Provisional - Civilian
year
Foreign/Occupied Regimes
government_Foreign/Occupied
year
Dominant Party Regimes
government_Dominant Party
year
Indirect Military Regimes
government_Indirect Military
year
Party-Military Regimes
government_Party-Military
year



There are some clear trends that we can make out when we look at the plots of government types. Firstly, we see distinct in upward trend in Parliamentary and Presidential type democracies. Parliamentary democracies comprise less than a quarter of all government types in 1950 and Presidential democracies comprised less than one-twentieth. However, in the past 70 years both types of democracies have dramatically increased, to roughly 70% of all types of government in 2019.

The reverse has also been quite pronounced. The percentage of governments that comprise dictatorships, one-party, or military rule have declined to less than 30% of all government types.

The last and final group of graphs we plotted were coup attempts and successful coup attempts:



Here we see a clear trend. Both coup attempts and successful coups have declined, asymptotically, since 1950. This means that, over time, coups have become rarer events. This makes predicting them harder – as the imbalance between observed coups and non-coups widens. Indeed, of the ~98k observations, we witness only 240 coup attempts – roughly .02% of the total number of observations; and of these 240 coup attempts, only ~46% of have been successful. Any useful model must account for this imbalance.