Preparation of Papers for IEEE Sponsored Conferences & Symposia*

Ningyuan Xiong[1] and Xinyu Cai[2]

***Abstract*— In this assignment, we were asked to analyze a given data set from Correlates of War Project with respect to the history of wars and conflict among states. Through analyzing the giving data set, we have some interesting findings on the history of wars among states.**

I. INTRODUCTION

COW is a project about collecting, disseminating, and using quantitative data in the international relations field, and it was founded in 1963. It collects data about the history of wars and conflicts among states. The goal of the project has been the accumulation of scientific knowledge about war and it is an academic resource. The current director of COW is Zeev Maoz from the University of California, Davis. D. Scott Bennett is currently associate director from Pennsylvania State University. The project also has an Advisory Board with 9 voting members in total. An institution and an individual will maintain a data set and the related documentation for a period of time.

There are 15 data sets developed by this project:

- COW Country Codes: the list of states with COW abbreviations and ID numbers.
- State System Membership (v2016): fluctuating composition of the state system since 1816 and countries corresponding to the standard Correlates of War country codes.
- COW War Data, 1816 - 2007 (v4.0): the list of wars that are stored in the database which includes Non-State War data set (v4.0), Intra-State War data set (v5.1), Inter-State War data set (v4.0), and Extra-State War data set (v4.0).
- Militarized Interstate Disputes (v5.0): all instances of when one state threatened, displayed, or used force against another.
- National Material Capabilities (v6.0): National Material Capabilities such as military expenditure, military personnel, energy consumption, iron and steel production, urban population, and total population are included in this data set.
- Militarized Interstate Dispute Locations (v2.1): geographic locations of MIDs in latitude/longitude coordinates, per dispute and per incident.
- World Religion Data (v1.1): detailed information about religious adherence worldwide since 1945.
- Formal Alliances (v4.1): This data set records all formal alliances among states between 1816 and 2012, including mutual defense pacts, non-aggression treaties, and ententes.
- Direct Contiguity (v3.2): The Direct Contiguity data set registers the land and sea borders of all states since the Congress of Vienna, and covers 1816-2016.
- Territorial Change (v6): This data set records all peaceful and violent changes of territory from 1816-2018.
- Colonial/Dependency Contiguity (v3.1): The Colonial/Dependency Contiguity data set registers contiguity relationships between the colonies/dependencies of states (by land and by sea up to 400 miles) from 1816-2016.
- Intergovernmental Organizations (v3): Although the number of intergovernmental organizations (IGOs) grew dramatically during the late 20th century, they have been part of the world scene for much longer. This data set tracks the status and membership of such organizations from 1815-2014.
- Defense Cooperation Agreement Dataset: This dataset covers bilateral defense treaties from 1980-2010.
- Diplomatic Exchange (v2006.1): The Diplomatic Exchange data set tracks diplomatic representation at the level of chargé d'affaires, minister, and ambassador between states from 1817-2005.
- Trade (v4.0): This data set tracks total national trade and bilateral trade flows between states from 1870-2014.

Looking through the data set, we found the capability of each country varies from year to year, and each variation has some clue we could dig into. For example, the United States has had a very high military capability over the years, but there are some sharp increases and sharp decreases in the 1900s. After World War two, America became the world's richest country, hence the increase. Not long after that, the Great Depression might be the biggest causation that leads to the sharp decrease. After gaining back power, several wars causes a lot to the military capability, hence declined again as expected. Other countries also have increases and decreases due to their reasons. Therefore, we found this data set very comprehensive and meets our expectations.

II. DATA

The data set in the COW project is not the latest version. The newest version is 6.0. For the data set, version 5.0 expands the data to 2012 and adds additional documentation of data sources. For the documentation, an additional subsection labeled "2017 update" at the conclusion of each NMC component section is added.

In the data set of the Correlates of War Project, there are 15171 data of 243 countries in total, and eleven variables containing details of the changes in their military capabilities of each states.These variables are country code, years, military expenditure, military personnel, iron and steel

production, primary energy consumption, total population, the urban population, the Composite Index of National Capability (CINC), and version. In the data set, military expenditure and military personnel are each state's total military budget and the size of a state's military personnel in each year. Total population and urban population is the size of a state's total civilian population and urban civilian population in each year. Iron and Steel production reflects a state's production of pig iron (1816-1899) and steel (1900-2012), and the Primary Energy Consumption is a state's consumption of energy (metric ton coal equivalent) in each year. The CINC reflects an average of a state's share of the system total of each element of capabilities in each year, weighting each component equally.

There are 23 variables in the supplementary dataset. The "statenme" is the state's name. "milexsource" and "milexnote" represent the source and note of the military expenditures, and "milpersource", and "milpernote" military expenditures, represent the source and note of the military personnel. "irstsource" and "irstnote" represent the source and notes of Iron and Steel Production, and "irstqualitycode", and "irstanomalycode" represent the quality code and anomaly code of Iron and Steel Production. Similarly, there are "pecsource", "pecnote", "pecqualitycode", and "pecanomalycode" of energy consumption, "tpopsource", "tpopnote", "tpopqualitycode", and "tpopanomalycode" of the total population, "upopsource", "upopnote", "upopqualitycode", and "upopanomalycode" of the urban population. "upopgrowth" and "upopgrowthsource" represent the growth rate and growth rate source of urban population.

These data might be collected from each country, and thus generate some problems. In the military personnel, it is easy to see that during the course of each state's foreign policy, state often has an incentive to exaggerate their troop strengths. In the military expenditure, it was often difficult to identify and exclude civil expenditures from reported budgets of less developed nations. In the total population data, there are two difficulties with territorial boundary changes and assuming a constant growth rate in regression.

In question 5 (Fig. 1), variable correlations between Irst and Pec, Pec and Upop, Upop and Tpop, and Cinc score with all other variables are very high. Variables like Upop and Tpop might be correlated by nature, for the urban population is part of the total population. Pec and Upop are also correlated by nature, for the number of primary energy consumption are determined by the population in the urban areas.

If we include positively correlated and negatively correlated variables as explanatory variables in the same predictive model, they will both cause multicollinearity[1], which causes the following problems:

1) The coefficient estimates can swing wildly based on which other independent variables are in the model. The coefficients become very sensitive to small changes in the model.
2) Multicollinearity reduces the precision of the estimated coefficients, which weakens the statistical power of your regression model. You might not be able to trust the p-values to identify independent variables that are statistically significant.

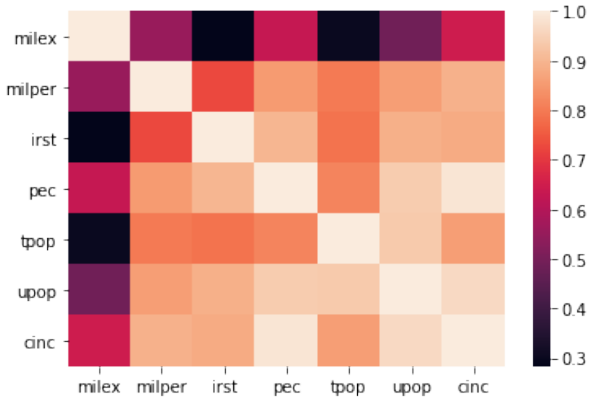

Fig. 1. HeatMap for Correlation Matrix

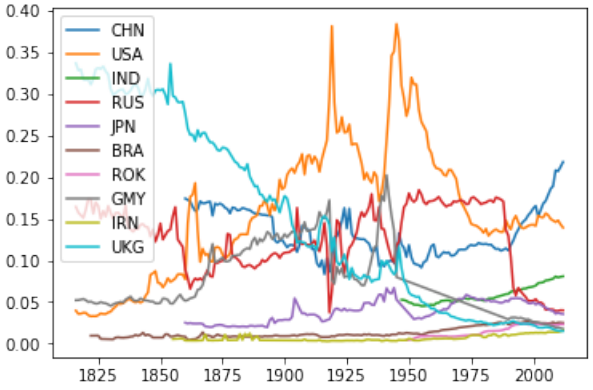

Fig. 2. Line Graph of Top 10 Countries

III. ANALYSIS & RESULT

In this assignment, we used the CINC score to analyze the capability of 243 different countries. In question 3 (Fig.2), we found the top 10 most capable countries in year 2012 are China, United States, India, Russia, Japan, Brazil, South Korea, German, Iran, and United Kingdom, and we used a line graph to show their changes. By observing the changes in CINC for these 10 countries over time, it is easy to find that these changes are associated with historical events. For example, the U.S has two peaks in 1925 and 1950, and there is a large decrease between and after these two periods. The former decrease is probably due to the Great Depression, while the later one could be attributed to the beginning of the Korean War and Vietnam War. The increase of the cinc at the end of 80s of China is due to the Opening-up policies instituted by the Chinese government. Besides, there is a great decrease for Russia at the end of 80s which is the direct result of the dissolution of the USSR.

In question 4 (Fig. 3), we used a rigdeline plot to compare the distributions of the cinc variable for these 10 countries. By observing the plot, Iran, ROK and BRA have an approximate bimodal distribution. The distribution of IND is

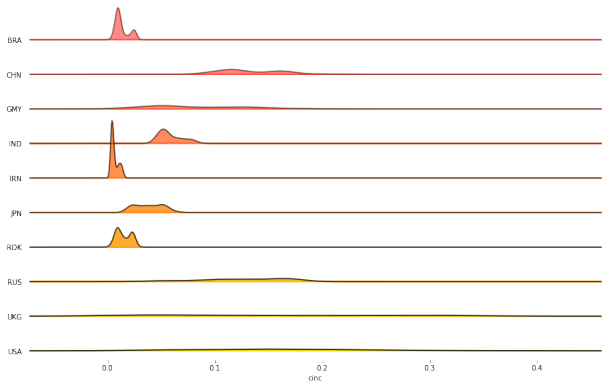

Fig. 3. Ridgeline plot of Top 10 Countries

right-skewed. Countries with high peaks indicates that the cinc value barely changed over the years, while for those relatively-flat distribution, the cinc value changed a lot during these years like USA, UKG etc. Reflecting the ridgeline plot with the line graph in Q3, the relatively-flat distribution corresponds with the line graph that shows a clear trend.

In question 6, we calculate the pairwise Manhattan and Euclidean distances between countries. From the result, we find the top 10 smallest Manhattan and Euclidean distance pairs are the same except the order of the pairs is different. The ten smallest pairs are Tuvalu and Nauru, St. Kitts & Nevis and Marshall Islands, Kiribati and Tonga, Monaco and Liechtenstein, Grenada and St. Vincent and the Grenadines, Grenada and Tonga, St. Vincent and the Grenadines and Tonga, Monaco and San Marino, Liechtenstein and San Marino, St. Lucia and Samoa. An interesting finding is that the top 3 smallest country pairs are all island countries. The smallest pair is Tuvalu and Naura. They are both in the Pacific Ocean, and the distance between them is only 1395 kilometers. Thus, it's reasonable that these two countries have similar National Material Capabilities. The fourth smallest distance pair is Monaco and Liechtenstein. They are both in Europe and are included in the top 6 smallest countries in the world. One more interesting finding is that these 10 countries are all in the rank of 50 smallest country in the world[3]. The explanation could be that the population and resources are similar due to the size of the country, and small countries typically have fewer people and expenditure on military.

In question 7, we used the 2012 cinc value to create a world map that colors the countries according to the amount of national capability. We found a good world map method from a website[2], and add our own data into it. By observing the graph, it is easy to find that Asia seem to have power imbalance. China, Japan, South Korea and India has relatively high cinc value, while other countries in Asia has relatively lower cinc values such as Thailand, Myanmar and Pakistan. Besides, Europe is also power imbalance. France and Germany have cinc scores higher than other European countries. However, Africa and South America have more power balance. Except for Brazil, all other countries in these two regions have cinc scores ranging from 0.0001 to 0.009.

Countries with approximately similar military capabilities are easier to have conflicts. For example, conflicts happened a lot in Middle East Asia and Africa among low military capabilities countries. Countries with higher cinc value tend to have other types of conflicts such as economic or international diplomatic issues.

REFERENCES

[1] Frost, J., Nousheen, R., Taiwo, Omari, J., Bayarbaatar, Amanda, Abid, M. A. R., Wolde, B., David, Yj, Louise, Hornos, F., Data, D., Heather, Prince, Grenci, J., Hans, Garg, P., Pavithra, ... Chakraborty, S. (2021, August 25). Multicollinearity in regression analysis: Problems, detection, and solutions. Statistics By Jim. Retrieved September 20, 2021, from https://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/.

[2] Han, D. C. (2020, July 6). How to make a coronavirus world map in python. Medium. Retrieved September 20, 2021, from https://towardsdatascience.com/how-to-make-a-coronavirus-world-map-in-python-734c9fd87195.

[3] The 100 smallest countries in the world. TitleMax. (2021, April 13). Retrieved September 20, 2021, from https://www.titlemax.com/discovery-center/lifestyle/the-100-smallest-countries-in-the-world/.