Modeling Geopolitical Multipolarity

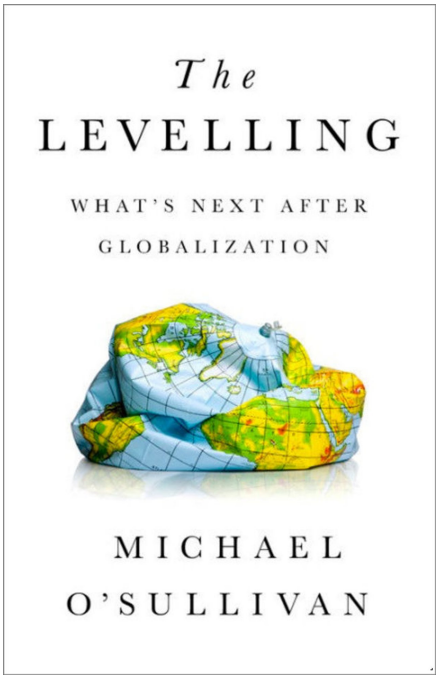
Christine Dorward, Andrew MacFadden, Nicholas Cappon,
Brandon Case, Nhan Do, Tanay Bhardwaj

MGIS650:02 - Modeling Geopolitical Multipolarity

Christine Dorward, Andrew MacFadden, Nicholas Cappon, Brandon Case, Nhan Do, Tanay Bhardwaj

Table of Contents

Executive Summary:



Polarity in geopolitical terms is defined as the way which power is distributed in the international system. There are three current models regarding polarity including unipolarity, bipolarity, and multipolarity. This type of system is not only related in the macro global context but the regional context as well.

These ideas are discussed by author Michael O'Sullivan in his book "The Levelling: What's Next After Globalization", where he discusses the impacts of a changing trends on the world in the future. His main theory is that the world is shifting towards a multipolar world instead of a globalized one where countries exert various degrees of regional as well as global dominance.

The main factors described in his model included Country GDP, Population Size, Regional Economic Role (Regional GDP), UN Human Development Index Relative to Region, Existence of an Imperial Legacy, Military Size & Sophistication (Absolute Military Spending, Number of fighter jets, Number of ships), and Regional Grouping Participation (NATO or European Union).

In order to examine the theory behind this book we created a unique dataset based on the variables described in his model. We then used this dataset to see if his theory was indeed accurate and if these variables could be used to predict geopolitical strength. We also included some additional variables not mentioned in his model that we thought would provide a more accurate representation as well.

After dividing the sovereign nations of the world into different geographical regions we calculated various regional ratios of variables instead of using the absolute global values to get a better picture of the regional effect of each state as well. We then created a regression model using the Comprehensive Index of National Capability (CINC) as a dependent variable to run our various chosen model variables against to see if they correlated.

We also collected around 100,000 tweets regarding the NATO summit in early December 2019 and analyzed them for sentiment to see how people were feeling about this important geopolitical organization.

In conclusion we determined that this model is a fairly accurate for determining a nations place on the CINC ranking. While this is not exactly what the author was theorized in his book it was as close as we could come given the data available and the time frame we had available to complete this project. In fact there are many limitations to this project mostly based around the availability and reliability of the data itself and the inability to quantify intangible variables for modeling geopolitical strength.

In the future we might want to get past historical data for many of these important variables and see how they have changed over time. New research also suggests it is far too simple to simply correlate GDP with the CINC as there are many past historical examples of this not accurately interpreting the reality of the growth of certain nations.

Introduction:

Globalization and multipolarity are geopolitical concepts that have transformed the way the world operates. In his book “The Levelling: What’s Next After Globalization,” Michael O’Sullivan claims that the world is heading towards multipolarity over globalization, but believes that the world is unprepared for this change due to the current worldwide trends of low growth and high debt.

Globalization is the integration of goods, technology, jobs and information on a global scale through free trade. During the peak globalization period, nations became more interdependent in relation to trade. Multipolarity is a system where power is distributed within the international realm, usually among at least three significant “poles.” Poles are typically regional powers with a large global outreach. These poles usually concentrate on wealth and military capabilities and have the capability of disrupting any arrangements that could disrupt their major interests.

Under O’Sullivans model he believes that the European Union, United States, China and India will all be considered poles but other countries like Japan and Russia will not. He believes all of these poles will take different approaches to economic policies, political freedoms, technology, war and culture. He suggests that some of these smaller counties will work to find their place in this new world and may even join in leagues of nation states like the past.

Multipolarity is largely dependent on the current distribution of power and influence of regional states. O’Sullivan stated that multipolarity could possibly be defined by factors other than those that are economic. Military, political and cyber freedoms, technological sophistication, financial sector growth and a sense of cultural confidence could also lead to strong multipolarity in a region.

According to him this could be interpreted down to specific variables for example he says:

“To try to synthesize what a pole entails, we can point toward several initial factors: size of a country’s GDP, size of its population, the existence of an imperial legacy, the extent of its regional economic role, its military size and sophistication (e.g., absolute spending, number of fighter jets and ships), its place on the UN Human Development Index relative to its region, and its participation (or not) in a regional grouping (such as NATO or the European Union).” (Economist)

We decided to focus on whether data related to the other factors stated above could help determine whether O’Sullivan’s theory held true.

Research & Methods:

In order to successfully model this concept we created our own data source from publicly available material provided by many different sources including the World Bank and the United Nations. Since some of these values for some countries were incomplete so we filled in the remaining blank spaces using information available through other reputable international sources. [See Limitations]

Based on the above mentioned variables we attempted to break them apart and see what specific data we could get for each one. The table below shows each one of his propose variables as well as the final variable we used in our calculations.

Author's Variable	**Our Variable**
Size of a country's GDP	Country GDP
Size of its population	Population Size
Existence of an imperial legacy	Imperial Legacy
Extent of its regional economic role	Country GDP Relative to Region
Military size and Sophistication	Global Firepower Index
Absolute Military Spending	Military Budget ($US BN)
Number of Fighter Jets	Total Air Assets
Number of Ships	Total Naval Assets
Participation (or not) in a regional grouping (such as NATO or the European Union).	Regional Grouping

In this report, we collected data from multiple sources to assess every country's geopolitical strength majorly based on economics and military aspects. In order to assess each country's military powers, we obtained data about military expenditures as well as other specific numbers about their armory from SPRI Fact sheet which contains the latest data from 2019 about the world militaries in 2018.

We also selected other factors such as population density, oil reserves, the Human Development Index (HDI) and also the corresponding regional HDI for each country. We also added in land area (km^2) and several inclusive indexes such as the Index of Economic Freedom (IEF) that takes many other national aspects into consideration.

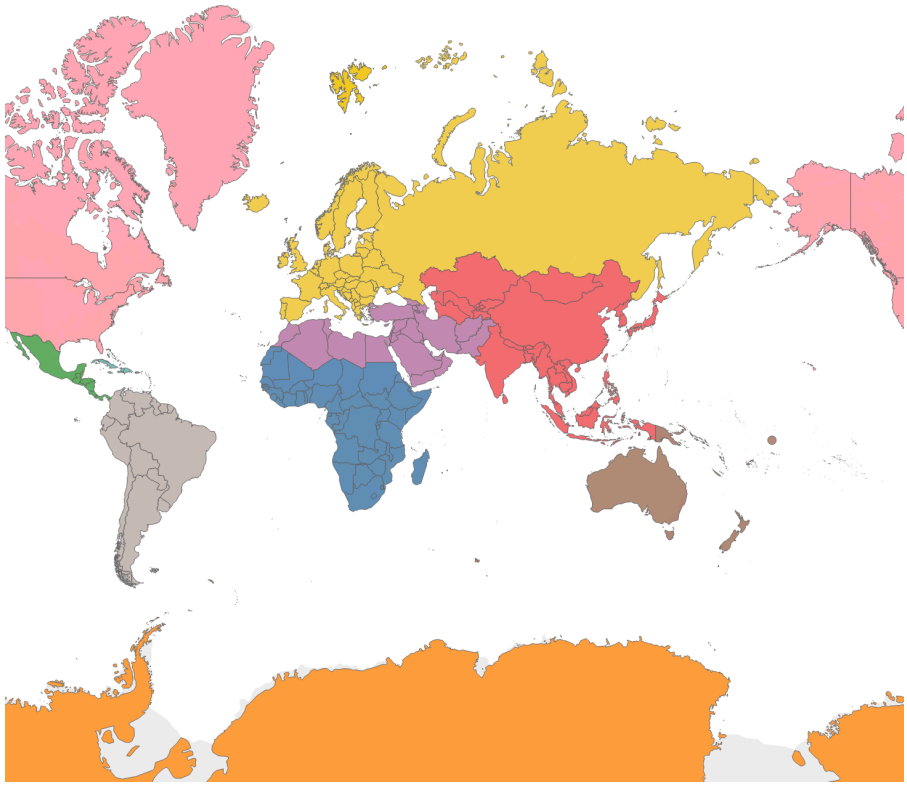
Country & Regional Grouping Scheme

In determining which countries to include in our dataset we also chose to eliminate countries that were not sovereign. This includes Hong Kong, Macau, and almost all of the British Islands as well as many United States territories.

The countries in our dataset can be broken down into ten major regions based on cultural, geographical, political and other factors that allow for more statistical accuracy. Most of the regional groups correspond to the continent on which they are located but that is not always the case in regards towards areas like the Near East cultural area which encompasses North Africa as well as the Eastern Mediterranean and parts of Asia.

There is also the separation of Central America and the Caribbean based on various cultural and geographical factors which is likewise separated from North America and South America. We also chose to divide North Africa from Sub-Saharan Africa due to cultural reasons as well as the fact there are clear measures of statistical difference between these two areas that would skew our results if we included them together.

Legend	
Pink	North America
Green	Central America
Blue	Caribbean
Gray	South America
Yellow	Europe
Purple	Near East
Blue	Africa
Red	Asia
Brown	Oceania
Orange	Antarctica



The massive ice covered continent of Antarctica hosts no permanent human inhabitants save for the occasional research scientists. This gives it a value of zero for population, GDP, and all associated variables. This allows us to use Antarctica has a control group across the variables as Antarctica does not have a military either.

Variables:

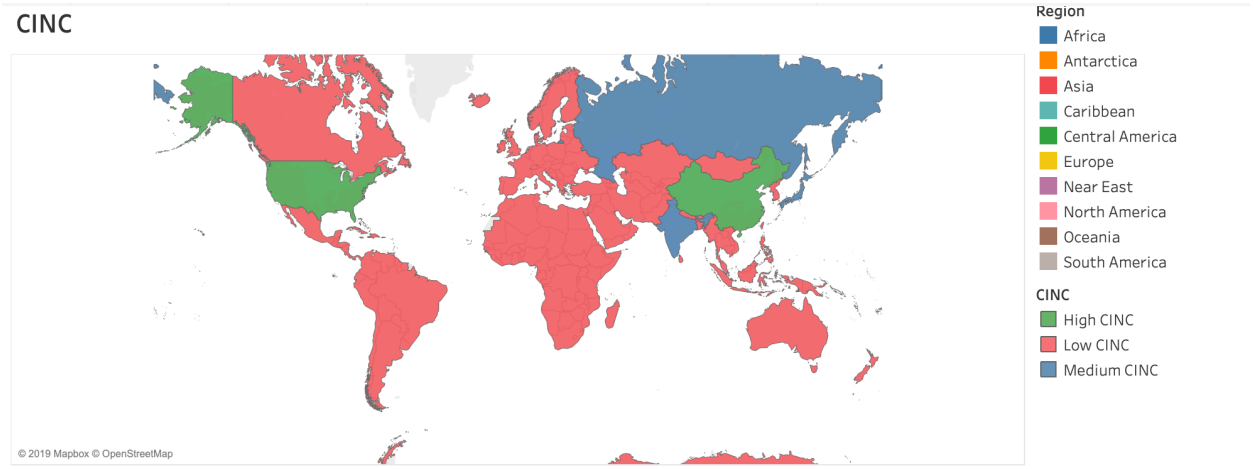
Dependent Variable:

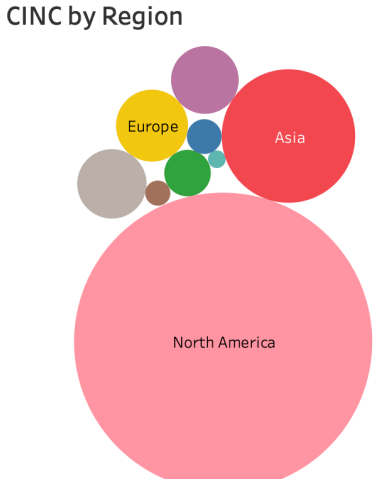
CINC (Comprehensive Index of National Capability):

This index is a measure of hard power rather than a comprehensive indicator of overall national capability. The measure is obtained by taking a simple average of six ratios. These ratios measure population, urban population, iron and steel production, primary energy consumption, military expenditure and military personnel respectively.

The Correlates of War webpage shows the variation of this index for the period 1816-2007. Even though it is primarily a 'hard power' index, it omits the significance of nuclear capabilities. Moreover, the role of modern technology aided weapons like drones has not been considered. Given that the world is more urban now than it is rural, giving equal weights to both population and urban population leads to "double-counting" effect of the population variable.

The graph above shows the CINC geographically and by region. Almost ⅔ of the world has a low CINC. Whereas North America has a very high CINC because United States has a very high CINC compared to any other country and there are only two countries included in this region, the USA and Canada which also has a higher than average CINC.



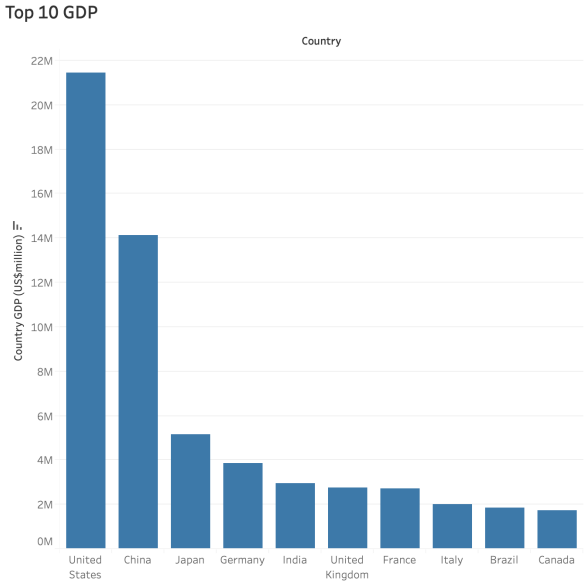


Independent Variables:

Population Size

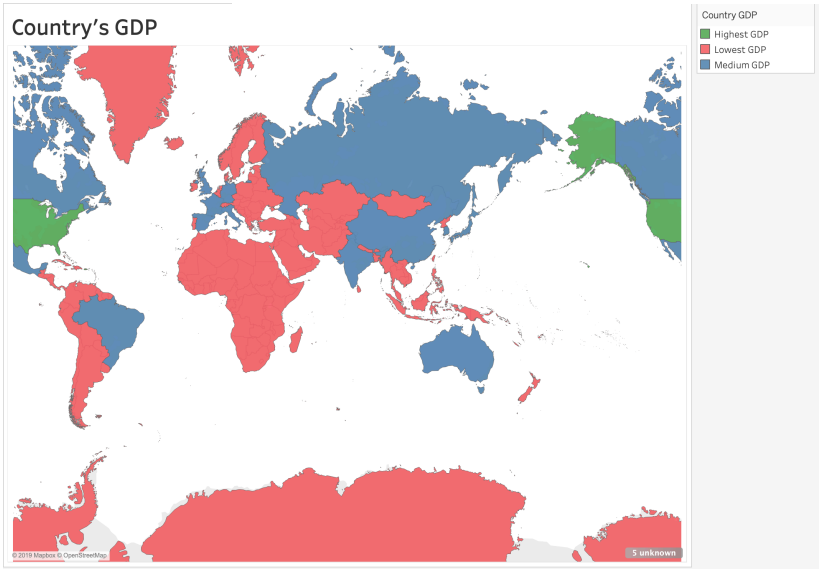
This data was calculated in absolute value based on the available data taken from the United Nations. This variable was later excluded from our analysis because it is included as part of the CINC composite variable.

Gross Domestic Product (GDP)



This number indicates a country's total domestic production value of finished goods and services in a specific time period. In this case, it shows the annual amount of each country. As this measures overall domestic production, it is often used to reflect a country's economic health. In other words, GDP provides an economic preview of a country, and thus, it is often used to estimate the size of an economy and its growth rate.

In general, the higher the GDP is, the better a country's economy is performing. There are several ways to calculate GDP by either expenditures, production, or incomes. Moreover, if GDP calculation also gets adjusted for inflation and population, GDP can provide deeper insights of country's economy and overall development. For those reasons, a country GDP is a key tool to guide policymakers, businesses and analysts in decision-making process as well as countries' economic development assessment.

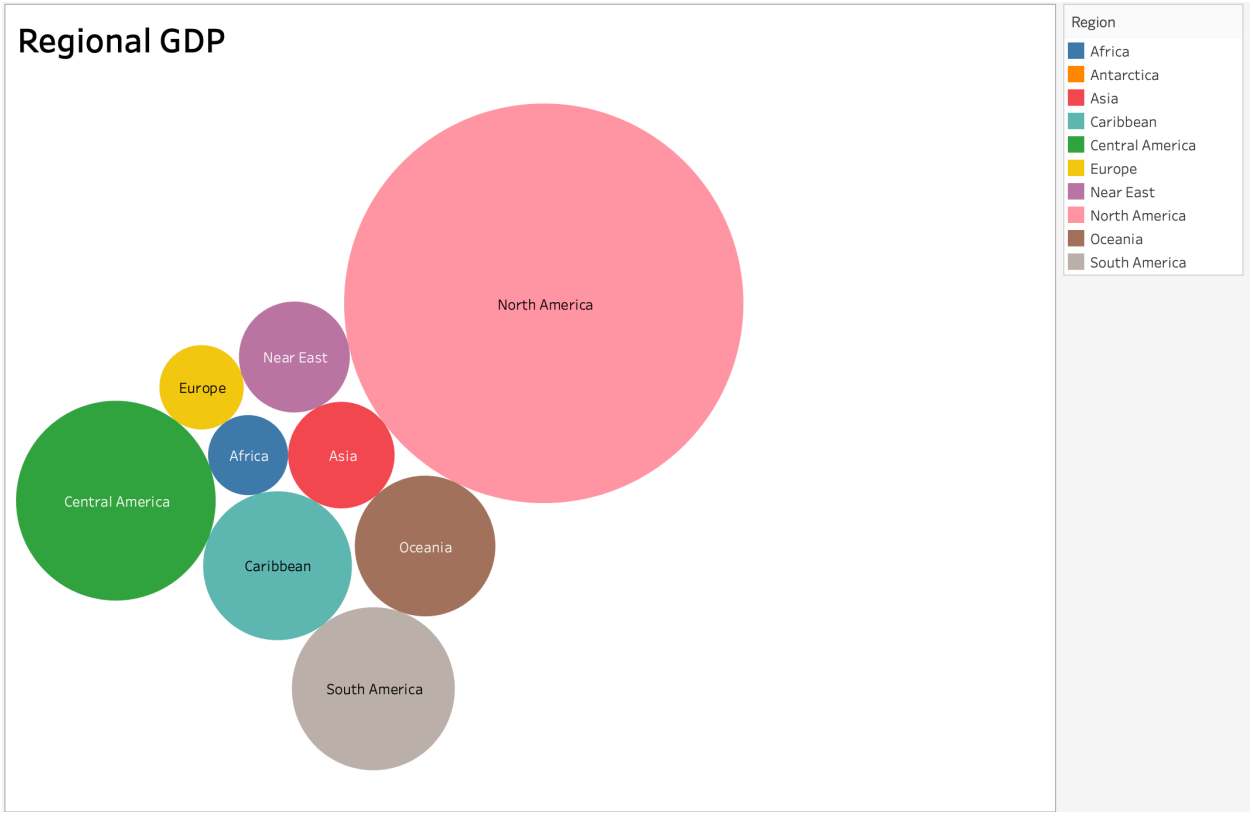


Throughout our analysis, GDP has been proved to be an essential factor since it has high correlation with almost all variables in our list which demonstrates how influential this indicator is in every sector of a country in general as well as our chief dependent variable in particular.

Almost ⅔ of the world has a low GDP, whereas countries like India and China has a medium GDP, on course of becoming having a major GDP in the future. The graph above shows the top 10 GDP worldwide and which region has the highest GDP.

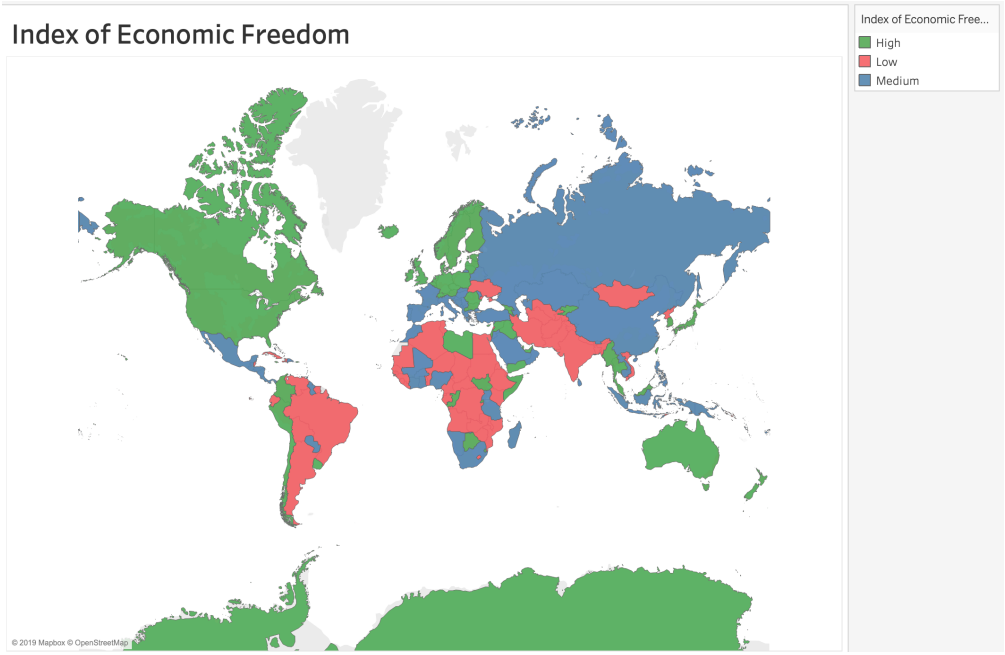
Regional Economic Role (Regional GDP)

The regional GDP was calculated by taking the sum of the GDP for each country in a particular region and then dividing each country by this total regional GDP sum. This was then converted into a percentage to give a representation of that country's GDP contribution to the region. The higher a percentage the more a contribution and thus the stronger a country is theorized to be.



Index of Economic Freedom

Economic freedom is the right of individuals to control their own living. Countries with the highest economic freedom ratings are those with the most personal freedom, voluntary exchange, security, and ability to compete in open markets. The Index of Economic Freedom helps viewers understand what helps an economy grow by evaluating aspects such as Rule of Law, Government Size, Regulatory Efficiency and Open Markets. The sub-variables within these categories are weighted, scaled and then ranked on a scale from 0 to 100.
The graphs shows the Index of Economic Freedom geographically and by region. The African region has the lowest IEF, whereas North America has the highest.



We initially thought this index could be used as the dependent variable due to the inclusiveness of economic variables playing into our analysis. We discovered through our initial statistical analysis that the Index of Economic Freedom did not correlate with the independent variables we are using in determining multipolarity. In addition, the VIF for several of the variables was significantly high, which could lead to an unreliable model, as shown below.

Regression Equation

Index of Economic Freedom Ranki = 57.41 + 0.000001 Country GDP (US$million)
- 1.661 Global Firepower Index - 0.000000 Populatio
+ 0.4522 High-technology exports (% of m
+ 1.659 UN HD Index

Coefficients

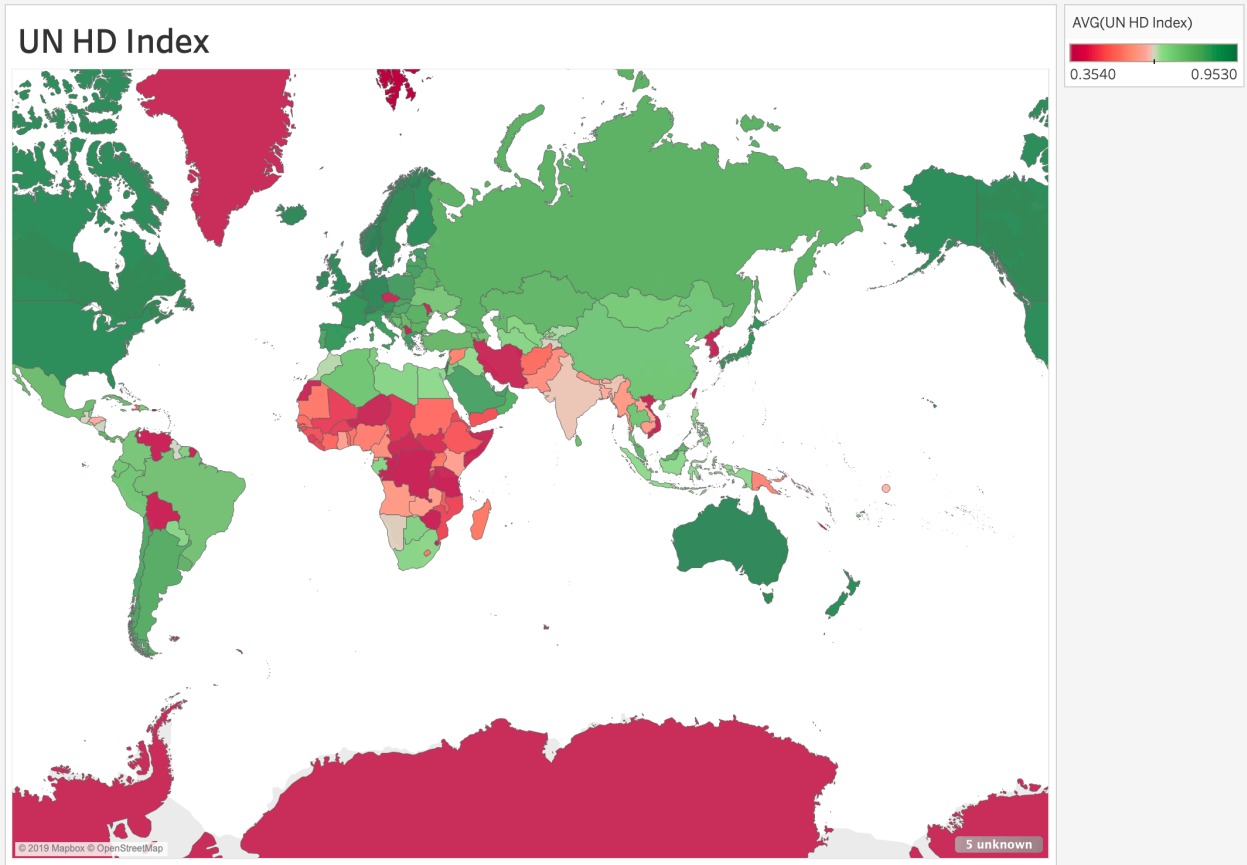
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	57.41	1.19	48.24	0.000	
Country GDP (US$million)	0.000001	0.000001	1.76	0.079	1.56
Global Firepower Index	-1.661	0.872	-1.91	0.058	331.71
Population	-0.000000	0.000000	-2.34	0.020	1.50
High-technology exports (% of m	0.4522	0.0931	4.86	0.000	4.66
UN HD Index	1.659	0.885	1.87	0.063	341.64

UN Human Development Index (HDI)

HD Index or Human Development Index (HDI) is an index measured by UN to be a criteria for assessing the development of a country, based on a country's human factors and its potential capabilities instead of relying on economic growth. By aggregating key dimensions of human developments such as lifespan, education and living standards, HDI is a simple index that provides a picture of a country's human development. Unlike the HDRO index, HDI doesn't cover human factors like inequalities, poverty, human security, or empowerment but it majorly focus on human development factors.

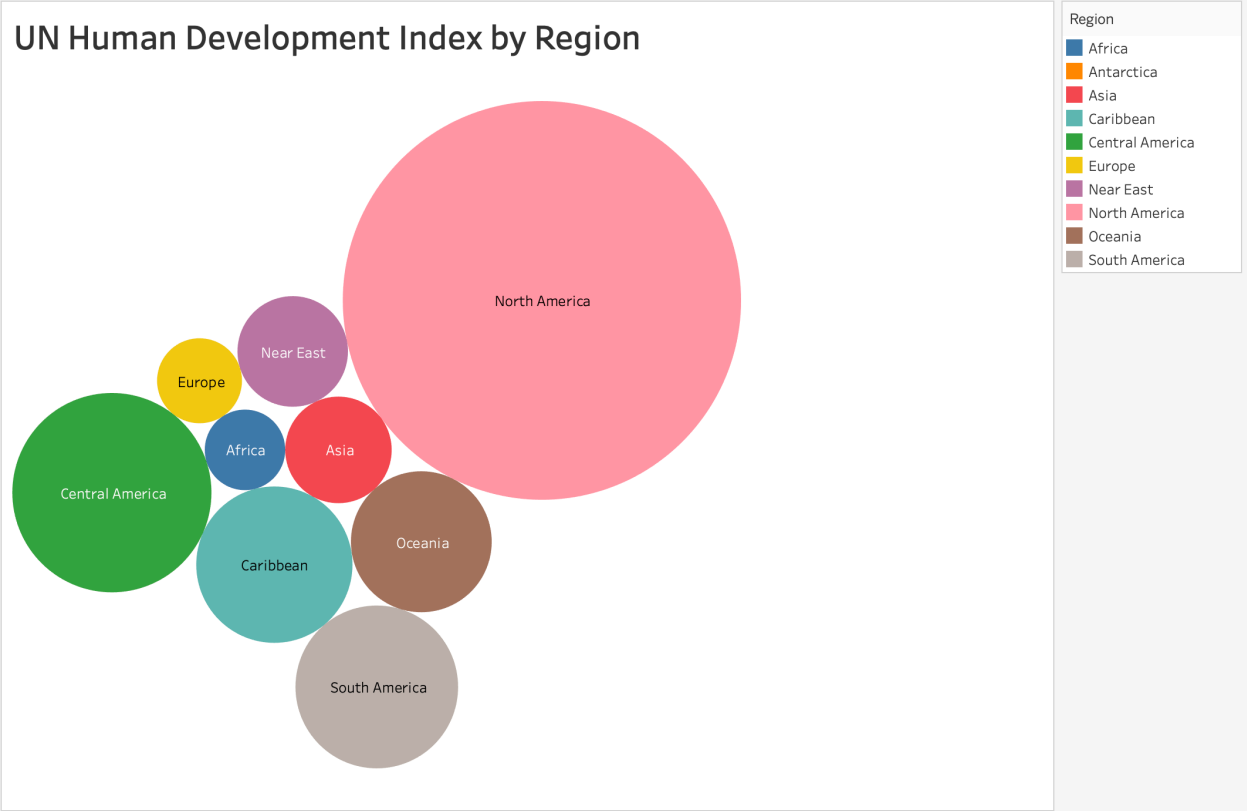
That's why in HDI the health dimension is assessed by life expectancy at birth, the education dimension is measured by average years of schooling for adults and the standard of living dimension is measured by gross national income per capita. In sum, although a fuller picture of a country's level of human development requires more analysis of other indicators, the inclusion of HDI in our model provides more human development insights to our assessment of countries.

In traditional approaches, country valuation is often based on its survival and thus, there has been a centrality of power in geopolitics. According to the paper by the RAND Corporation that we are referencing, the focus of country valuation has shifted to a few indices which possess incremental improvements over the traditional single and multivariable approaches to assess national power. For those reasons, we decided to include GDP, Global Firepower Index and UN HD Index in our country's analysis.



UN Human Development Index Relative to Region

This variable was calculated by taking the sum of the UN HDI for each country in each region and then dividing each country by this total regional UN HDI sum. This value gives the relative contribution of each country's HDI to the total region. Much like regional GDP the higher the percentage the stronger a nation is theorized to be in geopolitical terms.



Imperial Legacy

Imperial legacy is defined as a nation that has held colonial territories outside of its borders and also known as an “empire”. To set parameters for this variable the scope of countries included in this category are known to have had colonial territories within the last 200 years. The proper name of the empire as well as the known start and ending dates are given in the table below.

Nations with an Imperial Legacy (Former Colonial Powers)				
Flag	Country	Empire	Start	End
	Austria	Austro-Hungarian Empire	1867	1918
	Belgium	Belgian Empire	1885	1962
	Brazil	Empire of Brazil	1822	1889
	China	Qing Dynasty	1644	1911
	Denmark	Danish Empire	1536	1953
	France	French Empire	1534	1980/Present
	Germany	German Empire	1884	1918
	Hungary	Austro-Hungarian Empire	1867	1918
	India	Mughal Empire	1526	1857
	Italy	Italian Empire	1885	1960
	Japan	Japanese Empire	1868	1945
	Netherlands	Dutch Empire	1602	1975
	Pakistan	Mughal Empire	1526	1857
	Portugal	Portuguese Empire	1415	1999
	Russia	Russian Empire	1721	1917
	Spain	Spanish Empire	1492	1898
	Sweden	Swedish Empire	1638	1878
	Thailand	Kingdom of Siam	1782	1932
	Turkey	Ottoman Empire	1453	1922
	United Kingdom	British Empire	1707	1997/Present
	United States	United States Territories	1817	Present

Military Size & Sophistication

Military are highly armed forces that are maintained by the sovereign state. Military can be further divided into various branches like Navy, Air-Force and Army, which in turn have various subdivisions. Data for this was gathered from multiple sources including the SIPRI Military Expenditure Database as well as the International Institute for Strategic Studies in their publication the “The Military Balance” for 2019.

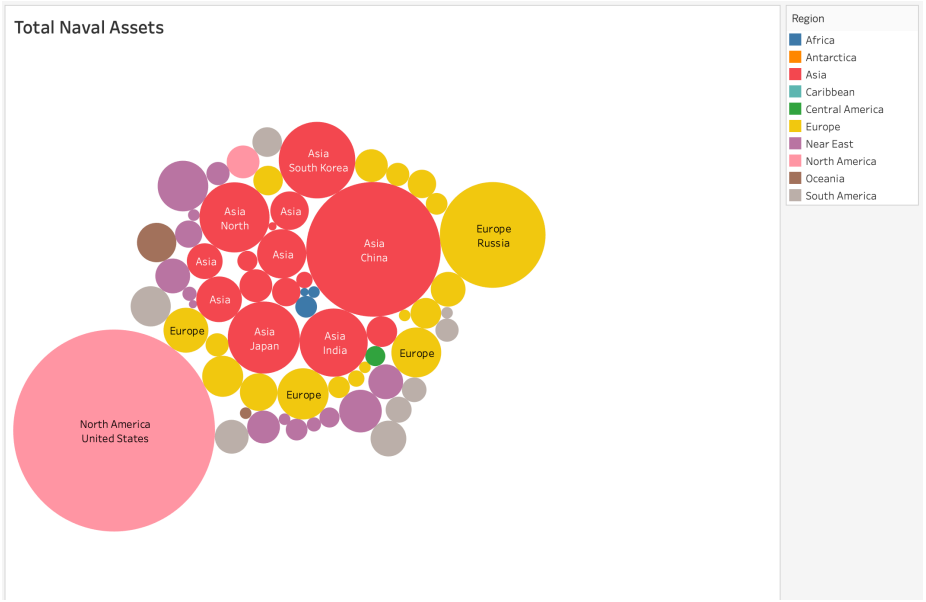
We ran simple regression for each of these variables against the CINC to determine which components of the military correlated best. Based on the table below the results were ranked in order of their R-squared value in descending order.

Simple Regression against Composite Index of National Capability (CINC)				
Rank	**Variable**	**P-Value**	**R-Sq**	**R-Sq (adjusted)**
1	Amphibious War-Ships	<0.0001	73.12%	72.98%
2	Frigates	<0.0001	66.01%	65.82%
3	Military Satellites	<0.0001	65.61%	65.41%
4	Battle Tanks	<0.0001	59.08%	58.83%
5	Military Aircraft	<0.0001	51.00%	50.71%
6	Attack Helicopters	<0.0001	48.50%	48.19%
7	Destroyers	<0.0001	46.21%	45.89%
8	Corvettes	<0.0001	39.25%	38.89%
9	Nuclear Submarines	<0.0001	37.12%	36.74%
10	Aircraft Carriers	<0.0001	33.73%	33.33%
11	Nuclear Weapons	<0.0001	29.61%	29.23%
12	Non-Nuclear Submarines	<0.0001	28.75%	28.33%
13	Cruisers	<0.0001	28.10%	27.67%
14	Absolute Infantry	<0.0001	20.91%	20.44%
15	Per 1000 Capita Infantry (Total)	0.8024	0.04%	0.00%

The above table suggests that countries should focus their military efforts on these sectors however, this may not be applicable in all cases due to technology and geography constraints. Overall one of the most important on this list that does not get included in our regression model are military satellites which give an extra edge to any country in intelligence, surveillance and reconnaissance. These types of intangible military variables are not accounted for in our model and form a major limitation in our analysis.

Total Naval Assets

Total naval assets was calculated by adding up the aircraft carriers, amphibious war-ships, cruisers, destroyers, frigates, corvettes, nuclear submarines, and non-nuclear submarines for a nation.

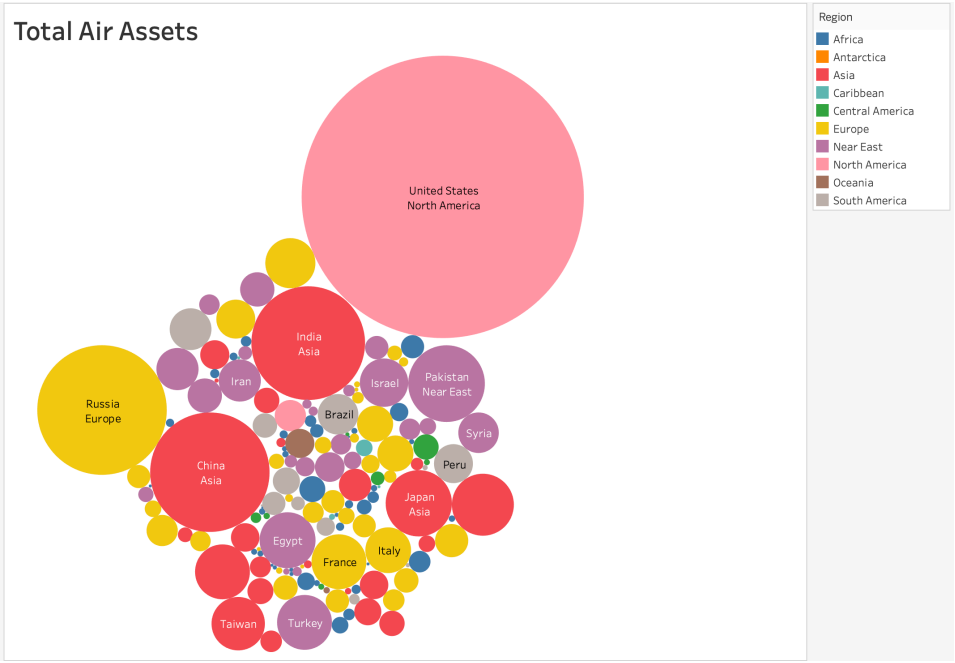


However, this variable leaves out the existence of small speedboats, drones, unconventional weapons or the use of privateers or state-sanctioned pirates.

The graph shows the military assets by region. North America, United States in particular has the highest number of military satellites, nuclear submarine and air assets, whereas battle tanks are present in Asia, Europe, North American region almost uniformly.

Total Air Assets

Total air assets was calculated by adding up military aircraft and attack helicopters. Excluded from this but also important in our analysis are military satellites and nuclear bombs which can be delivered through various air, land and naval means.

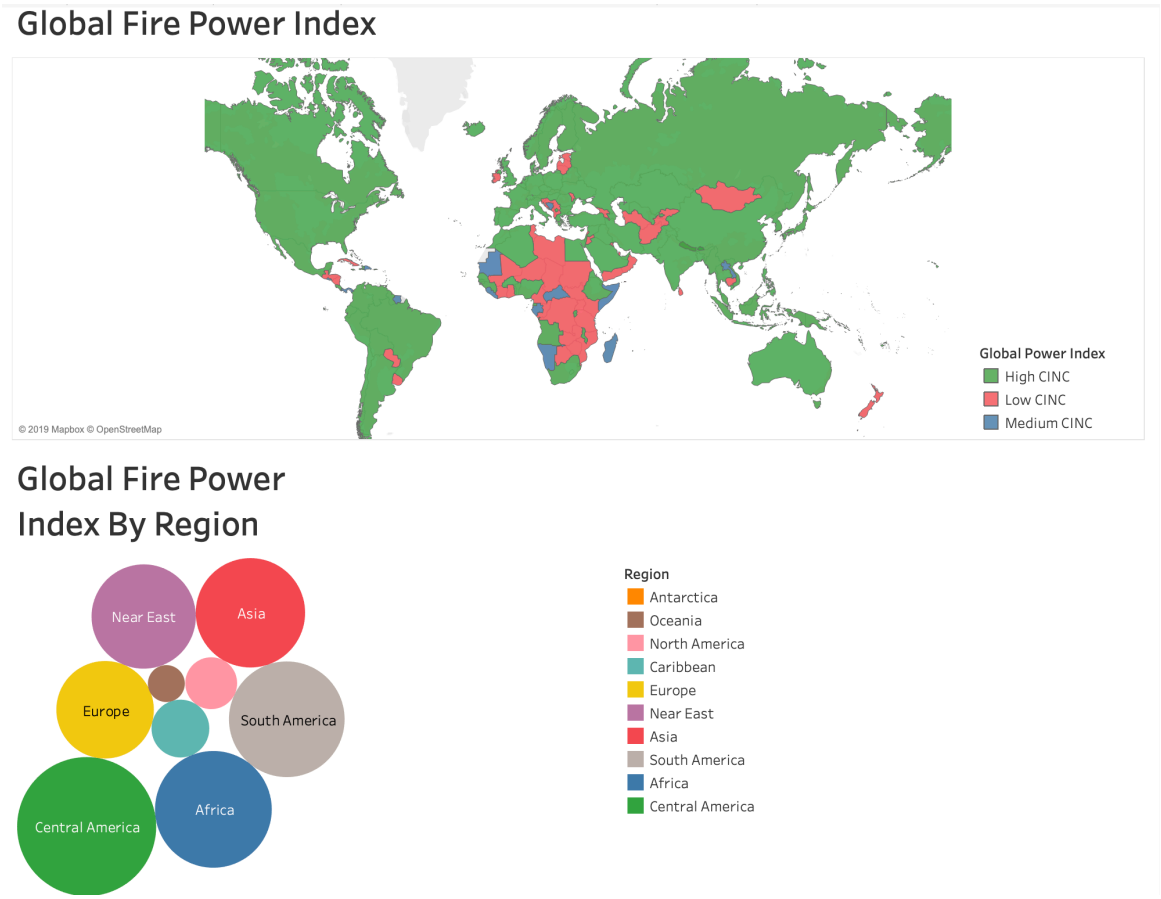


Global Firepower Index (GFP):

The Global Firepower ranking provides real time analytical data concerning the modern military power of a country in all three phases of combat (land, air, and sea). This index takes into account over 55 different factors, which includes the number of specific vehicles and weapons each country possesses, to determine the military capabilities of a country.

In simplistic terms, this index acts as a rankings of sorts, where it tries to determine the victor if two countries fought one another in a conventional war setting (one country versus one country). Thus, the highest ranked country would be favored in a military war no matter the opponent, while the lowest ranked would be expected to lose against everyone else.

More than ⅔ of the countries has a high Global Fire Power Index, this indicates that all the countries are determined to have a strong military because of the present geopolitical scenario. Also, there is no difference region wise. Majority of region has almost the same GPF.



Regional Grouping Participation (NATO or European Union)

This variable assigned each country either a score of 1 if in either the EU or NATO or 2 if not. We attempted to find other regional groupings with the military equivalent of NATO or the political of the EU but we could not. Given the seemingly incomplete nature of this variable and its corresponding low correlation with the CINC we decide to exclude it from our model.

Simple Regression for Various Variables

The table below gives the various outputs for running simple regression between the CINC and each individual variable. After doing this the data was sorted by R-squared values in descending order and the values with the highest R-squared and lowest p-value are shown at the top.

Simple Regression against Composite Index of National Capability (CINC)				
Rank	**Variable**	**P-Value**	**R-Sq**	**R-Sq (adjusted)**
1	Country GDP (US$million)	<0.0001	83.20%	83.11%
2	Total Naval Assets	<0.0001	71.31%	71.51%
3	Absolute Population	<0.0001	67.45%	67.28%
4	Total Air Assets	<0.0001	51.17%	50.91%
5	Military Budget (US$billion)	<0.0001	35.11%	34.71%
6	Regional GDP	<0.0001	33.18%	32.83%
7	Land Area (km2)	<0.0001	27.18%	26.80%
8	Imperial Legacy	<0.0001	27.14%	26.76%
9	Regional UN HDI	<0.0001	10.39%	9.91%
10	Global Firepower Index	0.0002	10.02%	9.36%
11	Regional Grouping	0.0001	7.44%	6.95%
12	UN HDI	0.0536	1.96%	1.44%
13	IEF	0.3432	0.54%	0.00%
14	Population Density	0.7789	0.04%	0.00%
15	Oil Reserves	0.5108	0.46%	0.00%

It can be interpreted based on this table that each of the variables selected by the author are an accurate predictor of the CINC. This corroborates his theory so we decided to select all the following variables and run a multiple regression model to see if they accurately determined the CINC as a collective.

Regression Model:

CINC = 0.000246 - 0.00322 Regional GDP + 0.000000 Country GDP (US$million)
- 0.00552 Imperial Legacy - 0.000106 Military budget (US$ BN)
+ 0.000041 Total Naval Assets + 0.000008 Total Air Assets

Coefficients:

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.000246	0.000355	0.69	0.490	
Regional GDP	-0.00322	0.00316	-1.02	0.311	1.80
Country GDP (US$million)	0.000000	0.000000	20.44	0.000	15.74
Imperial Legacy	-0.00552	0.00115	-4.81	0.000	1.38
Military budget (US$ BN)	-0.000106	0.000006	-17.04	0.000	17.41
Total Naval Assets	0.000041	0.000039	1.06	0.289	43.81
Total Air Assets	0.000008	0.000002	5.61	0.000	48.74

Model Summary:

S	R-sq	R-sq(adj)	R-sq(pred)
0.0040124	96.29%	96.15%	24.10%

Analysis of Variance (ANOVA):

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	6	0.065262	0.010877	675.63	0.000
Regional GDP	1	0.000017	0.000017	1.03	0.311
Country GDP (US$million)	1	0.006727	0.006727	417.87	0.000
Imperial Legacy	1	0.000372	0.000372	23.09	0.000
Military budget (US$ BN)	1	0.004677	0.004677	290.49	0.000
Total Naval Assets	1	0.000018	0.000018	1.13	0.289
Total Air Assets	1	0.000507	0.000507	31.48	0.000
Error	156	0.002511	0.000016		
Total	162	0.067774			

Analysis:

After analysis of the authors model we determined that many of the variables that he selected have very high VIF values associated with the CINC. This means each of these variables are correlated strongly with the CINC which we proved earlier as well We decided to take our analysis one step further and see if there were any other variables that correlated well with the CINC that could be used to predict this phenomena.

In order to do this we performed a stepwise regression on our entire dataset to pick the best subset of variables that can be used to predict the CINC. We wanted to see if the variables selected matched the ones the author had selected. As previously discussed this dataset included many variables outside of the scope of the author's theory including the UN HD Index and Global Firepower Index.

Among our total variables, we removed irrelevant variables and ones that already accounted for in other indices. For example, the calculation of CINC itself already has Population and Military number (active and reserve) so we have to eliminate them out of our equation to avoid double-counting and confounding variables. The Land Area, otherwise, is factor we identified with potential multicollinearity with unseen factors in CINC calculation so we decided to eliminate this factor as well. Thus, we are left with:

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.0170438255	0.00528	3.23	0.002	
Country GDP (US$million)	0.0000000088	0.000000	24.19	0.000	1.10
UN HD Index	-0.0174101718	0.00629	-2.77	0.007	1.23
Global Firepower Index	-0.0021903786	0.000938	-2.34	0.021	1.28

Country GDP, Population, Land Area (Km²), UN HD Index, and Global Firepower Index.

Statistical Analysis:

From this, we can see the t test is used to determine whether each variable's significance. Since all chosen variables have p-values smaller than 0.05, we can conclude that each of the individual independent variables is significant.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	0.055149	0.018383	219.44	0.000
Country GDP (US$million)	1	0.049009	0.049009	585.02	0.000
UN HD Index	1	0.000641	0.000641	7.65	0.007
Global Firepower Index	1	0.000457	0.000457	5.45	0.021
Error	117	0.009802	0.000084		
Total	120	0.064951			

From this F test (test for overall significance) results, as all chosen variables have p-values smaller than 0.05, we can conclude that the test gives us sufficient statistical evidence that overall relationship between CINC and the set of independent variables is significant.

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.0091528	84.91%	84.52%	53.61%

According to this, with this model, one can best explain and predict CINC index with 3 variables GDP, Global Firepower Index and UN HD Index with the multiple coefficient of determination at almost 85%. Hence, we can safely say the least squares line provided a good fit because 85% of the variability in our dependent variable CINC has been explained by the estimated regression equation.

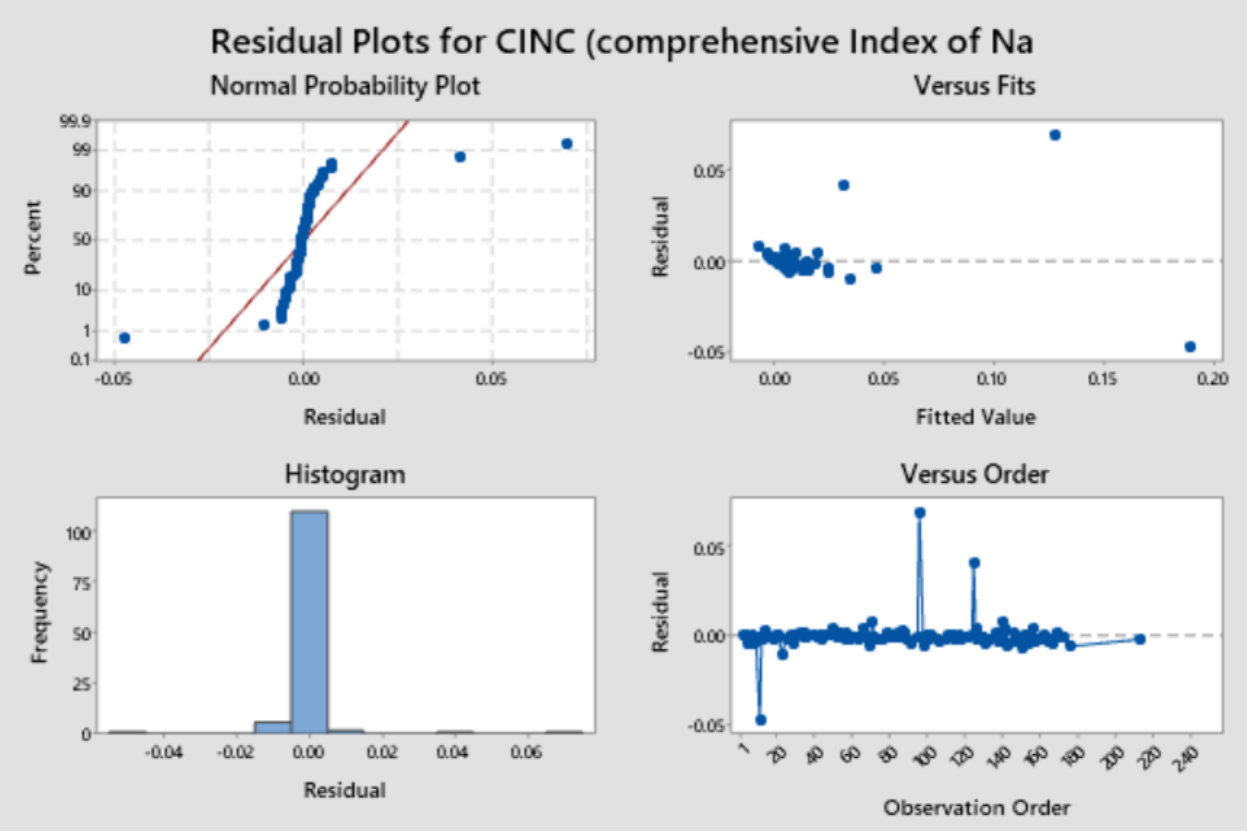
Regression Model Validation

We used residual plots to assess our model. First, to check for any skewness or outliers in the data, we check histogram of residuals which shows us little to no outliers in our data meaning there is very minimal skewness to threaten the validity of our model as the residuals are normally distributed.

Second, we use the normal probability plot to verify the assumption that the residuals are normally distributed according to our model. The normal probability plot of the residuals is approximately linear supporting the condition that the residuals are normally distributed.

Third, to verify the assumption that the residuals have a constant variance, we observe the Residuals versus fits graph: as residuals scattered around zero, the data has a mean of zero and the vertical width of the scatter doesn't appear to increase or decrease across the fitted values, so we can assume that the variance in the error terms/residuals is constant.

Finally, we viewed residuals versus order of data plot to check for autocorrelation and find that there is little autocorrelation between residuals to verify that the residuals are uncorrelated with each other.



Residual analysis conclusion: The above 4 graphs proves that our model is valid.

As we performed practical test on our model by calculating CINC using our regression model and compared it with actual CINC to see the differences, we have the following results:

As we can see, the majority of difference values is within 0 and 0.02 range while there are only 4 outliers with high difference values. For those reasons, we can say that our regression model is well validated to use as prediction model for future CINC.

Country	Country GDP	UN HD Index	Global Firepower Index	Current CINC	Calculated CINC	Difference
Norway	417627.00	0.95252202	0.6103	0.002	0.003	0.001
Switzerland	715360.00	0.94399757	0.5459	0.001	0.006	0.005
Australia	1376255.00	0.938631285	0.3277	0.008	0.012	0.004
Ireland	384940.00	0.938410059	2.1579	0.001	(0.001)	0.001
Germany	3863344.00	0.93604342	0.2097	0.024	0.034	0.010

Difference Range		Counts
0	0.01	172
0.01	0.02	76
0.02	0.03	1
0.03	0.04	0
0.04	0.05	2
0.05	0.06	0
0.06	0.07	0
0.07	0.08	1
Total		252

CINC difference	
Max	0.07
Min	0.00
Average	0.01
Median	0.00

Regression Model Application:

As we tried predicting next year CINC, we require next year's GDP, Global Firepower Index and UN HD Index. Since GDP is the essential factor in our model and future data prediction is unavailable for some variables, we assume the other two factors remain unchanged while calculating the new CINC. With the annual growth rate of GDP and current year GDP, we calculated the next year GDP for each country. As we applied our regression model, we have the following results for CINC next year and CINC ranking:

Obviously, for this prediction to work properly, we have to acquire better prediction methods for other factors as well as for GDP. Since the annual growth rate of GDP was used, the errors can occur if unusual GDP growth arises in this year or next year. In a better approach in the future, we can use past 10 year GDP values to calculate the average growth rate (this will minimize the effects of unusual GDP growth in certain years). The same can be said for other factors used in this regression model.Not to mention that one of our major interest observations (China) turns out to be one of the significant outliers, which means our prediction for it will not be as correct as other observation. Another limitation of our prediction is that our regression model is build upon current year dataset so possibly this model doesn't apply for next year's CINC.

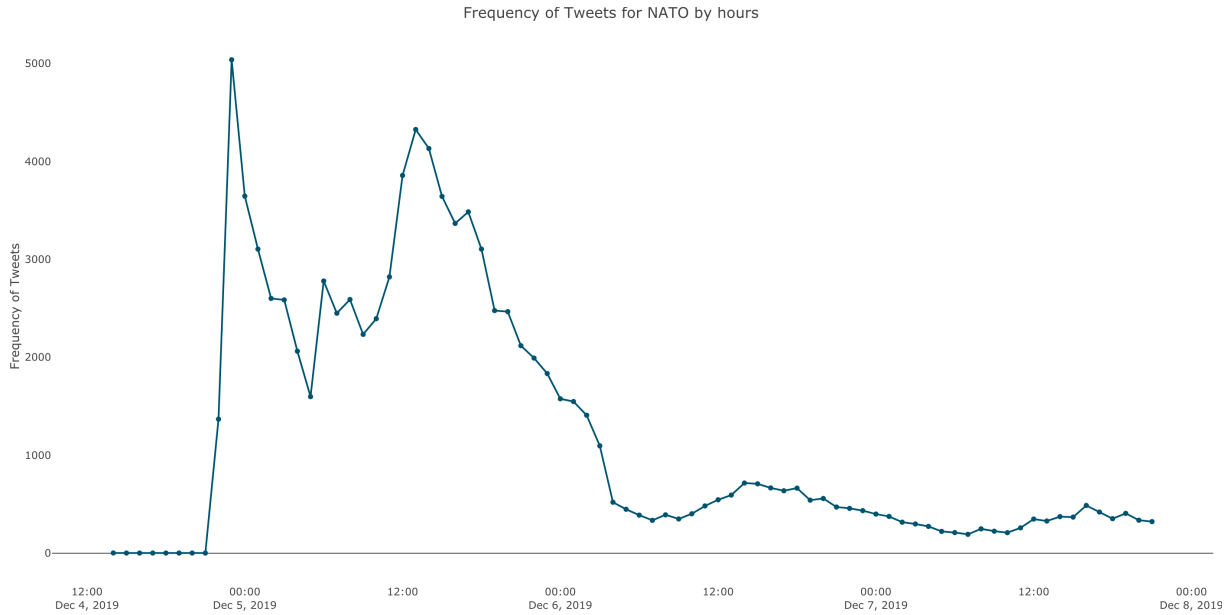
Country	Future GDP	UN HD Index	Global Firepower Index	CINC
United States	22,051,975.56	0.923913589	0.0615	0.20
China	15,073,413.76	0.751703947	0.0673	0.14
Japan	5,195,090.48	0.909152957	0.1707	0.05
Germany	3,918,408.38	0.93604342	0.2097	0.03
India	3,140,541.29	0.639832911	0.1065	0.03

Rank	Current CINC	Predicted CINC	GDP	UN HD Index	Global Firepower Index	Population
1	China	United States	United States	Norway	United States	China
2	United States	China	China	Switzerland	Russia	India
3	India	Japan	Japan	Australia	China	United States
4	Japan	Germany	Germany	Ireland	India	Indonesia
5	Russia	India	India	Germany	France	Pakistan
6	Brazil	Russia	United Kingdom	Iceland	Japan	Brazil
7	Germany	South Korea	France	Hong Kong	South Korea	Nigeria
8	South Korea	France	Italy	Sweden	United Kingdom	Bangladesh
9	United Kingdom	United Kingdom	Brazil	Singapore	Turkey	Russia
10	France	Taiwan	Canada	Netherlands	Germany	Mexico

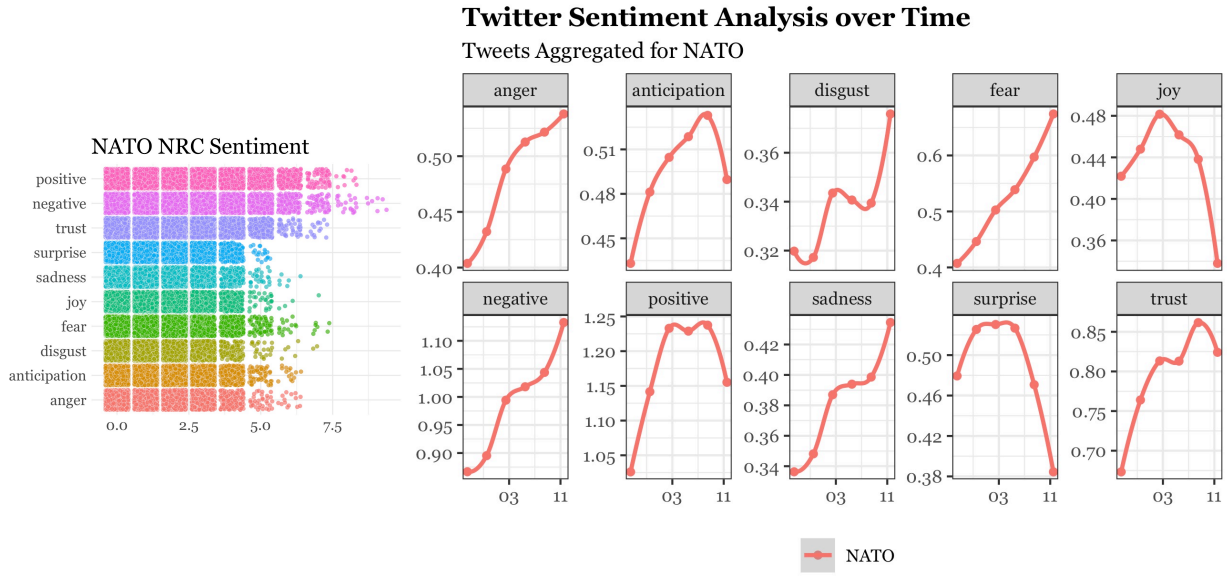
Twitter Sentiment Analysis:

Based on this project we decided to run a Twitter sentiment analysis on a London NATO summit that occurred on December 3rd, 2019 and ended on December 4th, 2019. We collected around 100,000 tweets between December 4th and December 8th to capture the incident and see how people felt in the aftermath.

The chart below shows the frequency of tweets between December 4th and December 8th. There is a clear spike between the 4th and the 6th that shows the incident along with the corresponding trail of decreased frequency in the following days after.

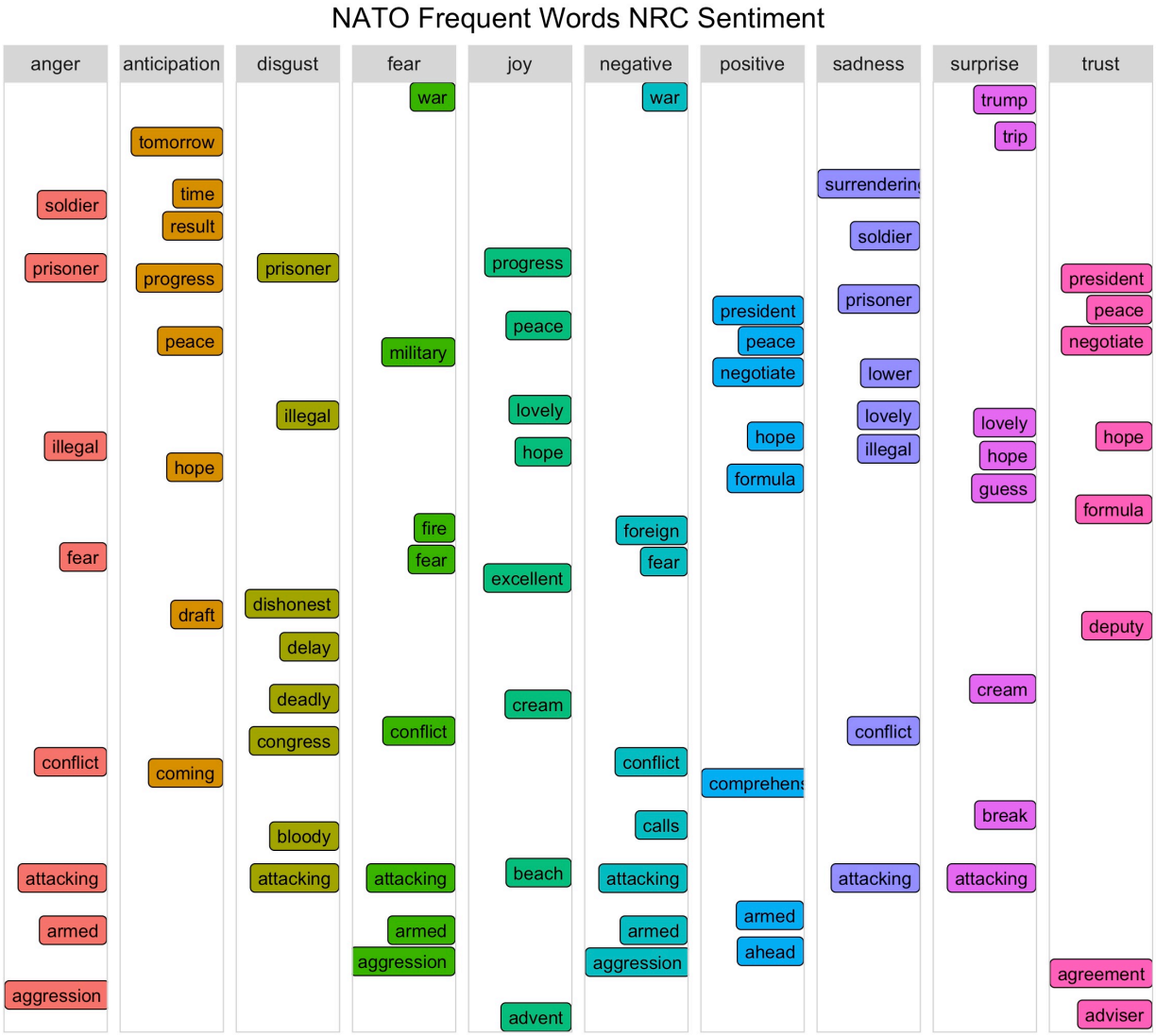


Based on this we ran generated a time series graph that plotted each individual sentiment over time which can be seen below.



Based on these sentiment charts we can determine that overall the sentiments related to anger, anticipation, fear, negative, positive, sadness and trust all increased while joy and surprise both decreased. Positive and trust also decreased after increasing which suggests that people were not pleased with the outcome of the summit.

Overall based on analysis of the sentiment over time regarding this NATO summit people were not happy about the current state of this organization and were not happy with the outcome of the summit. In order to find out possibly what happened we decided to sort the most frequent words for each sentiment into categories which were then plotted on the chart below. The most frequent words for each sentiment are plotted from top to bottom.



Based on this chart we can see that some of the most frequent words on the chart include war and there is a lot of anticipation towards progress. We would have liked to run this analysis for each specific day of tweets instead of on the larger dataset to get a better idea of what people were talking about day to day.

Limitations:

There were many limitations in this project that developed over the course of our analysis that if corrected, would lead to a smoother process and more reputable results. The first major limitation that we discovered while compiling our data was that there was significant missing information.

Some of these countries did not release significant data, which made it impossible to use them in any of our analysis. Countries that choose to not be associated with the UN / World Bank made it difficult to find data for these countries and in many cases we were not able to include them.

With that, each country also releases their data in different time increments, as each country undergoes a different reporting mechanism. As such, the different reporting standards and mechanisms each country uses may have made our analysis incorrect.

The values used in the finalized GFP ranking are constructed from information that is publicly available from various resources at the time of publishing - though it remains an impossibility to showcase presented values exactly due to limited reporting, a nation's lack of transparency, and other factors beyond our control. In some cases, estimates are used which may be based on previous years' data or, in its place, averages may be applied to complete the entry.

The model and steps we took is just one way someone could take to predict the next superpower. We cannot really predict with accuracy that a particular country or set of countries is going to be the next superpower because that would be naive. Theoretically, any major future current event (war, peace treaty, trade war, etc.) would impact some of our quantitative measures and change predictions.

In addition to this, most of the data that was derived is only quantitative in nature, as it is difficult to accurately gauge qualitative measures. Things like level of intelligence, quality of military and military strategy are very difficult to account for, and as such limits our analysis solely to quantity based measures.

There is significant debate among scholars about the actual relationship that CINC and GDP have together with some suggesting the strong correlation is actually misleading. There is evidence provided that contradicts this based on the United Kingdom and USA power transfer in the 19th century as well as a number of CINC-exclusive power transitions in the 20th century that GDP cannot account for. (Rauch 2017) This will require data from last 30-40 years at minimum and still the model may not be accurate. It will require resources and time to come up with a decent model and predictions.

Improvements & Future Work:

In the future, we hope to increase our scope and be able to realistically model future projections for many of these countries. While we did try to do some predictive analysis, we would feel much more confident if we had more historical data on which to base it off of. We would also like to look at more variables and see if there are other logical ones that can be used to help predict the CINC.

We would also like to look into finding a more accurate dependent variable to use as the CINC is the best available at the moment but is still not totally accurate. While it suggests China has a higher CINC than the United States, the USA is still the worlds strongest country and only global superpower. Therefore this variable itself is not an accurate predictor of geopolitical strength likely due to the presence of variables that either haven't been or cannot be quantified for regression analysis.

Conclusion:

Based on our analysis we can conclude that the author's theory is relatively correct regarding the correlation between various variables and the Composite Index of National Capability. While the real struggles exist in actually determine an accurate dependent variable and also with the data collection itself, this topic remains highly relevant in the modern world. Our sentiment analysis also reveals people feel pretty negatively about globalism right now if using the most recent NATO summit as a barometer. Overall, we would suggest a more in-depth and detailed analysis of this phenomena and we do not think that it can be so simply reduced to specific variables and formula equations.

Sources:

[1] The Economist. (06/28/2019) "Globalisation is dead and we need to invent a new world order" https://econ.st/34q6Tko

[2] "World Economic Outlook Database, October 2019". IMF.org. International Monetary Fund. 15 October 2019. https://bit.ly/2QNZ6sw

[3] Tian, Nan; Fleurant, Aude; Kuimova, Alexandra; Wezeman, Pieter D.; Wezeman, Siemon T. (28 April 2019). "Trends in World Military Expenditure, 2018" (PDF). Stockholm International Peace Research Institute. https://bit.ly/33m1dpY

[4] SIPRI (2019) "Data for all countries from 1988–2018 in constant (2017) USD". https://bit.ly/2XLdoeL

[5] A Survey of Indices Measuring Geopolitical Power. http://logos.nationalinterest.in/2014/03/a-survey-of-indices-measuring-geopolitical-power/

[6] Carsten Rauch (2017) "Challenging the Power Consensus: GDP, CINC, and Power Transition Theory". Security Studies, Volume 26, 2017, Issue 4, Pages 642-664. https://bit.ly/2qHrAtp

[7] HDRO (Human Development Report Office) (2019) "Human Development Report 2019 – "Human Development Indices and Indicators." United Nations Development Programme

[8] Global Firepower Index (2019) "2019 Military Strength Ranking" https://www.globalfirepower.com/countries-listing.asp