Correlates of War

Composite Index of National Capability

A Study of States' Power and Influence using SQL

Report Authored By:

Nicholas Birosik

Table Of Contents

Project Abstract

This project finds its roots in A.F.K Organski's 1965 theory, Power Transition Theory. Organski's theory establishes a relationship between all of the state actors within the international system, categorizing them as either satisfied or unsatisfied with the international *status quo*. Briefly, the theory posits that if a dissatisfied state reaches Power Parity with the leading satisfied state, or Hegemon, of the international system and they are dissatisfied with the *status quo* imparted by the hegemon, they will go to war with said state. Our objective is to use this dataset do help further refine Power Transition Theory which is a key part of national security politics. We accomplish this by exploring the distribution of CINC scores during different periods of time, to determine the period of power transition between the Hegemon and an approaching power a period known as the most militaristically volatile; through this study we will be able to determine conclusively the significance of the CINC score, and explore metrics used to predict the power of a state..

Our research question that has guided this project is: Which method of measuring state (terminology for countries in the study of international relations) power best models their actual power score? This information will prove valuable to the entire International Relations field of study. It will find implications in the accurate application of key theories in that domain. One such example is predicting outcomes posited by Power Transition Theory, in cases of proposed conflict between states such as China and The United States.

Project Methodology

Data Sources

The first data source that we have included in this study is the fifth version 5.0 of the National Material Capabilities (NMC) data set (2012). NMC is a core component of the *The Correlates of War Project* (COW) started by the late professor J. David Singer at the University of Michigan and is revered as the gold standard for measuring a states' power by IR scholars. Contained within this dataset are 15,171 entries, each of which contains an abbreviation of the country's name (stateabb), the COW unique three-letter country code, the year, the military spending of that country (milex), the number of military personnel in millions (milper), the energy consumption of the country (pec), the iron and steel production (irst), the urban population (upop), the total population of the country in millions (tpop), and finally the CINC (Composite Index of National Capability) index value for the country for a particular year of observation. Some limitations of this dataset include the limited span of observation. This dataset observes each state between 1816-2012, so older data on conflicts such as Napoleon's invasion of Russian, and more recent data is not incorporated in this dataset. Notwithstanding this minor hindrance in the dataset, NMC provides a relative global power share rating for each state in a measure that sums and finds the average of each of the above six numbers. This number is known as the Composite Index of National Capability (CINC). This numeric value will prove to

be very useful when we need to compare nations quantitatively; it will lend insight into how the individual factors of the CINC score help shape the overall measure of a country's power.

The second dataset that we will be incorporating is the World Bank's *GDP per Capita in Current US Dollars.* This dataset is based on data sourced from the Organization of Economic and Economic Development (OECD) and World Bank National Accounts (WBNC) compendiums. This dataset contains gross domestic product (GDP) per country per year. A shortcoming that we have discovered upon initial analysis is that this dataset only provides GDP information over the years from 1960 to 2019. Our narrow period of analysis, as already limited by the NMC data, may well be further hindered because of the limited scope of data that this source affords us; but because we are looking at more recent studies to measure states' power in the China vs. United States power transition this dataset, though limited in scope, may provide just the right berth for our intended analysis.

Cleaning the Two Datasets Using R

The primary object of bringing in the GDP dataset was to join it with our NMC dataset. The only problem was that the dataset had the year of observation as the column names, and the three letter state code as the row names. We utilized the statistical computing language R to mutate our data frames so that they would be arranged in the same format. After we made use of the gather statement to arrange all of the variables properly, we proceeded to rename the WBD dataset's "Country Code" column to "stateabb", so that our two data frames would be consistent. Next we saved the WBD year column as a numeric value so that our year data types would be equivalent. Next we made use of R's left_join function to join the NMC and WBD data. We made an executive decision to merge the datasets in R because then we would not need to waste valuable time attempting to perform a haphazard non-native join in MonogoDB[1]. With this newly created Tidy Data, we were ready to proceed with importation into our selected SQL and NoSQL data tools: MySQL and MongoDB respectfully.

Data Management Tool Selection

For this project we chose to make use of two tools: MySQL and MongoDB. We chose MySQL because it is the second most utilized relational DB on the market, making it a competitive product. (solid IT) It made sense for the project to make use of this simple database system because it could accommodate the row-column form of our .CSV files that we had acquired from the aforementioned data sources. Simply, we have chosen to use MySQL as our data tool, as the schema will be more conducive to MySQL queries, and MySQL will be more effective in structuring our data based on our schema

For the NoSQL portion of this project we have chosen to use MongoDB. This is boils down to two main components: the easy to read with syntax, and the ability to store a large

[1] For your reference, the R queries that were executed for this cleaning and joining process are available in the Appendix, under CC.1.

number of different documents of varying data types. Both of these features will lend themselves towards our ability to effectively query the database and manipulate data. There is also an added benefit to using MongoDB, as it offers us a GUI called "Compass" that allows for us to move away from Mongo's primitive shell-based environment. Compass also allows use to assess specific metrics that we are interested in, as well as lending us insight into query execution time and performance speed. We also chose to use MongoDB, because of its dynamic data-reduction power to select multiple attributes yet only project (or return to us) the variables that we are interested in. This means we can have operations running in the background that reduce confusion with information unnecessary to our questions and overall research question. In short, the simple syntax, ability to display specific data, and the ability to gauge our chosen metrics that resulted in our choice of MongoDB as our NOSQL tool.

Loading the Data into the Data Tools

To load the data into the MySQL database, we made use of phpMyAdmin on our GoDaddy hosted server (See CC5). phpMyAdmin provided a graphical user interface where we chose the amount of rows we wanted. All we had to do was define the datatype and max length. Once the datatable was created within the database, we uploaded the .CSV file to the server. From that document, the server automatically ran an insertion query for over 15,000 observations in the dataset.

To load the dataset into MongoDB, we made use of the Mongo shell on our personal computer environments laptop to create the database server. Loading the dataset required an import statement similar to the one seen in MC7, which was the loading of a joined dataset of the NMC data and GDP data from the World Bank. In this statement we had to write in the Path-to-File, and merely input the name of the .csv file. Once the line of code ran, all 15,173 documents were added in the server and were able to be accessed using the methods and commands that Mongo affords.

Results

In this section, you will find a description of what queries we ran on the dataset, and their implications on navigating our way towards our research goals. You may note how the results of one query often lead the way for further analysis to be performed in each successive query.

SQL Queries

We ran three SQL queries to determine which states held over twenty percent of the world powershare (CC2), which countries in 2010 held over five percent world powershare (CC3), and finally an aggregate query which determined Pearson's product-moment coefficient to determine the substantive significance of iron and steel production on Gross Domestic Product (CC4). We follow with analysis of each query and how it aids towards our research question.

The first query (CC2) returned us a list of states who have held the majority of the world powershare over the course of the years available in the dataset. This query helps us explore trends of ebbing power over the course of time. What we see is a veritable three-person boxing match whereby the United Soviet Socialist Republic, the United States, and China are fighting for global dominance, starting after the end of the Second World War. Before that period of time we see that the United Kingdom reigned supreme up until the First World war. This can be attributed to the mercantilist and colonial boon in India, et cet. With this data we can more concretely identify periods of power transition – that is a period of time where the difference in the power metric of two countries of opposing regime types is so small that if war were to break out a victor could not reasonably be forecasted. Now with a group of countries listed by date order and a power ranking, we can identify states that are in this range to know where we could potentially apply Power Transition Theory; first, though, we would need to analyze the states on this list to ensure that their regime type was in fact in contest and that the lesser state was indeed dissatisfied with the Hegemon's implementation of the *status quo*.

The second query (CC3) implements the same investigative style as the first; however, it narrows its focus on the present day to find states that are relevant to Power Transition Theory today. We isolated the year of observation to be 2010, the most recent data the dataset has available on complete CINC metrics for the world. What we found was that the three big players have changed yet again. They are the People's Republic of China, the United States of America, and India[2]. If we examine this data closely we see that China currently has a global powershare of over 20 percent and the United States has just over 14.8 percent. This is interesting for we see that the two states clearly crossed in power share in the 1990s. By Power Transition Theory, these two states are no longer in a period of power transition. Now we could reasonably forecast, noting diagram (Other References 2) that the relationship of these two states has matured into challenger preponderance, especially with a power differential of 5.2 percent, or nearly the entire state of India as of 2007. This leads to an interesting question, why wasn't Power Transition theory activated, and more puzzling still, why are the United States and China not at war? Is it that China, a socialist state which has taken to interning millions of Muslims (Buckly and Qin), restricting citizen liberty, and emplacing a nationwide facial-recognition based social credit surveillance network satisfied with the United State's projection of civil liberty, democratic ideals, and *laissez-faire* trade policy on the international forum? Or is it the more likely alternative that the metric of power, used to measure state power, is flawed? This leads to the next query which examines GDP as a metric for measuring state power.

Our final SQL Query (CC4) was used to determine the Person's Product-Moment Coefficient of correlation between the Iron and Steel Production and a measure of GDP for a given state. The resulting measure Pearson's product-moment coefficient, a measure of both statistical and substantive significance, of 0.5434473. This number can be interpreted as a

[2] It comes as no surprise that India has accumulated 7.9% of the global power share with a population of over 1 billion.

moderately strong positive correlation, meaning that as there is a rise in the dependent variable of IRST, the GDP also rises consistently. This query works by taking Person's coefficient formula (noted below in other sources one) and expressing it programmatically in SQL. This was achieved by using an aggregate function over and using aliased variables that are required by Pearson's function. Summed values of the independent variable, dependent variable, squares of each, and total sums. After we aggregated, we grouped by our aliased variables and executed the query to find the substantively significant results. This query is important because it helps answer why China may have the upper hand in CINC score but not with GDP, strictly based on a variable used in the calculation of both CINC and GDP.

Mongo Queries

We ran 3 different Mongo queries to determine different relations between not only the different states, but also between different variables. We analyzed which states had more than 15% of global power (MC1), what the average iron and steel production by country (MC3), and the average iron and steel production and average GDP of all the countries in the database (MC5). We also attempted to try to run a comparative analysis with MySQL in writing a query based on the Pearson's Product-Moment coefficient to determine the substantive significance of iron and steel production on Gross Domestic Product. However, MongoDB is not able to reasonably handle the modeling of this formula.

The first query (MC1) was among our first queries. We were initially interested in which states had the highest CINC score in the 21st century. As one of the focus of this project was to prove or disprove the Power Transition Theory, it was crucial that we look at which states were the most "powerful" in recent years. In looking into this data, we were able to see that only two countries in recent years have exceeded a rating of 0.15 (15% of the total global power), US and China. With China's recent revanchist and expansionist policies this puts the retrieved data into context with current world affairs. There is also the fact that the CINC scores of the two states have been increasing, however, China's score has slowed matched pace with the US, lending proof that the Power Transition Theory is correct, insofar as material capabilities. However this analysis only left us with further questions, leading to our development of MC3.

In MC3, we were attempting to find a more effective means of measuring power. Since iron and steel production (IRST) are key aspects of developed nations, we developed MC3 to analyze and find the average production levels for the countries in the dataset. In doing so we would be able to gleam whether or not this one factor could be a more accurate and efficient measurement of states' power. As it turns out, which MC3 was helpful insofar as we were able to easily pick out the wealthiest states, however for many countries this value was null or zero. While helpful in understanding the relations between power and IRST, it still did not answer the question as to how to develop a better measurement of power. This led to the development of MC5.

MC5 represents the culmination of our work with MongoDB. Before developing this query we previously discussed how GDP is increasingly becoming a major measure of global power and market share. For this reason we chose to join our dataset with data from the World Bank in RStudio, so as to add the GDP variable to our server under the new collection "Joined Frame." MC5 was developed to compare the averages of GDP and IRST among that states in the database. In finding these averages, we once again encountered the problems associated with null values. However, after exporting the data and eliminating null entries, we were able to develop OF4 and OF5 in excel which shows the different GDP and IRST values. In so doing it became apparent that even though states may have a high IRST value, their GDP is not dependent on this variable, and therefore GDP appears to be a more effective means of measuring power, as GDP encompasses states entire economies and not simply one sector. If GDP were included in the National Material Capabilities dataset, entirely different trends would develop and new implication could be gleaned. In short, our work with Mongo showed that, while a great measurement of military and energy capabilities, CINC score does not compensate for the fact that GDP can also have a major effect on government military spending and funding. It is therefore preferable to utilize the components that make up a CINC score as detailed above, in conjunction with the states GDP, giving us a better sense of a state's capabilities.

Discussion

This section is dedicated to discussing the speed with which the queries were executed, pros and cons of each data tool, and which tool we would proceed choose to continue the project using based on scalability needs.

Pros and Cons of Each Data Tool

In terms of the two data tools, each has their own benefits and shortcomings. In the case of MySQL, the primary benefit is its simplicity of syntax and versatility of querying – especially when performing aggregate functions. While MySQL is more user-friendly and intuitive when compared to MongoDB, it is far slower. These time differences are shown in the performance times of queries CC2 and MC1. Another tradeoff between the tool was the ease of uploading the data. As described in the "Loading the Data into the Data Tools" section, there is a few more steps with MySQL than are required by the MongoDB shell. While mostly negligible, this is still a slight difference in ease of use.

As for MongoDB, there are similar tradeoffs. While it is faster as shown and described above, Mongo's syntax is slightly harder to write. Indeed there were many times while working on this project that we had to take time to explore Mongo's documentation to find the right commands and syntax. Therefore in terms of querying, Mongo is moderately complex. However, Mongo rivals SQL in two categories, they are speed of querying and ease of data-loading. As described above Mongo makes it easier to load data meaning it is more user-friendly and it's queries execute faster thus leading to faster results.

Final Data Management Tool Selection

While though we are looking at a difference of 70 ms in query time difference between MySQL and its NoSQL counterpart, it is important to think about this from the perspective of negotiating big-data. Though in this instance our single query is trivial, when working with larger datasets, perhaps OLTPs, this 70ms multiple by a factor of millions of operations, for example, roughly equates to hours of CPU processing time being saved. It is important, therefore to explicitly mention this processing time difference between NoSQL tool MongoDB, and its lessor counterpart MySQL. Though this may be the case with Big Data, our data is small and cannot be typified by any of the V's of big data. We have come to realize that our future needs for expanding this project are quite narrow. With every annual release of the COW dataset, we are only looking at an additional 197 observations. Because there is such an insignificant increase in observations per year, it seems that the versatility offered by MonogDB, while though quantitatively better, does not meet our data computing needs. We run queries for this project using an *ad hoc* style. With that said, we are not constantly hitting this DB Engine with constant aggregate functions; our use is as needed to further our research goals. We will run perhaps one or two complex queries for each additional dataset. This alone does not justify using MongoDB which was built for working with big data. We select the MySQL tool as our data management tool because it augurs the best with our future needs: infrequent complex queries that otherwise are not yet supported by MongoDB. MongoDB simply offers to many complex features that we would not make use of going forward. An example is the clustering and sharding of data. This is a magnificent feature for scalability in large data warehouses, but is not particularly advantageous with a dataset comprised of a meger 15,000 observations. Finally with merging of large datasets it is not practical to follow Mongo's advanced join operation. It is more accessible to perform a left join in SQL, as opposed to jumping through hoops to perform a single join operation; especially when considering we have many datasets to join to the COW dataset. With these final datasets we do not anticipate altering the structure of the data beyond simply adding another attribute to the state entities. Because of this linear attribute adding protocol, we would not make use of Mongo's ability to effectively handle different structures of data; our current and planned schemas make the most sense when considered under the relational model afforded by MySQL.

Next Steps

Merging Other Datasets

We have a need to merge our data with additional datasets that contain vectors of state power measurement, such that we can run the appropriate regressions and find the best way to measure a states' power. One such data set, beyond the World Bank's GDP measure is the North Atlantic Treaty Organization's (NATO) state rankings, where by the list the relative scores of member and non-member states by reporting on the efficacy of each entity. We could also merge

our dataset with the World Happiness Index to determine if that has implications on a states' global standing. We are attempting to determine the proper combination of Betas (variables) in our linear model will best predict past occurrences of Power Transition theory using new power estimation models, this way we can rule out any omitted variable bias or any latent measures of variables that are measuring for the same thing (additive variable bias). The more datasets that we join, as we hope to do in the future, the more likely we are to find the most refined metrics of both statistical and substantive significance for determining a states' global power score.

What Questions Are Answered by this Research

This research has great implications in the field of International Relations scholarship. For finding the true measure of a state's power in the international system means understanding previously posited theories and contextualize them in meaningful ways. Take for example Power Transition Theory. If we could deduce the true metric for determining a state's standing, we could predict interstate conflict between nations such as China and the United States; this would enable governmental organizations, such as intelligence services, to be wary of looming conflict between such states; this might have the ability to deter war before it even starts. Determining this metric does not stop at Power Transition Theory, it has direct implications in many other theories that account for state's power in the international system. Enhancing the neoliberal model against competing realist theories is but one such example. In this way our study is valuable for it directly adds new knowledge to the domain of International Relations.

Classification of States and Joining Addition Data

Going forward we will follow the data pipeline process represented by the EER diagram listed under 'Other Figures 3'. Turn your attention to that diagram now.

This EER diagram represents the current implementation of the attributes that make up a state; note the two primary keys underlined are CountryID and Year of Observation which follow Chen Notation. Each additional joined data set will become yet another attribute of this State entity. The database will hold many observations of such states.

You may also take note of the listed specialization. The listed total disjoint specialization is a product of our ambitions for future implementation and the advancement of our project as briefly described in the "Merging Other Datasets" section header. This is a very important classification in the case of our work with Power Transition Theory. PTT posits a conditional anarchical international system, whereby there are some states that are satisfied with the *status quo* (how the world is currently running under standards emplaced by the hegemon) and those that are not. A state's regime type usually determines this. We want to augment the datatable so that we can clearly see where the states rank within each realm, satisfied or dissatisfied. This will help elucidate for us those states that may go to war with each other in periods of power transition; for it will identify which states are considered to be dissatisfied. We could re-run CC1

and CC2 with this augmented data frame to identify potential competition for the title of Hegemon of the International System.

Conclusion

An initial challenge we faced was the join statement in Mongo. We had remedied this problem by joining using R, and concluded that going forward since joining in mongo was such a convoluted process that, perhaps, Mongo might not be the right data tool for our project. We have looked at our current and future schema and determined that the rigidity of our data fits perfectly in line with what MySQL offers, and that we would not be making use of Mongo's additional features, such as semi-structure, ability to adapt to millions of documents, etc. While it is true that Mongo ran queries 70ms faster than SQL, Monogo proved incapable of running a complex Pearson's Product-Moment regression which was something that we needed for our analysis upon juxtaposition of the two. The GUIs were found to be comparable, however when running anything beyond a mere filter and project statement in Compass, this Mongo GUI proved to be limiting in comparison to the open query shell afforded by phpMyAdmin. If we were to scale up with these products we would only be adding 192 observations per year, which is quite miniscule in comparison to what MongoDB was meant to handle; as such, future growth with the Mongo brand does not seem feasible, for it continues not to augur well with our research goals, ambitions, and needs of this research project. Additionally, since we are not worried about speed, but rather infrequent complex queries that we can use to answer any questions as they arise, we have made the choice to move forward with the MySQL data tool.

References

Data Sets

National Material Capabilities (v5.0 1816-2012) adapted and expanded from:

Singer, J. David, Stuart Bremer, and John Stuckey. (1972). "Capability Distribution, Uncertainty, and Major Power War, 1820-1965." in Bruce Russett (ed) Peace, War, and Numbers, Beverly Hills: Sage, 19-48. https://correlatesofwar.org/data-sets/national-material-capabilities

GDP per Capita (Current US$):

The World Bank. "GDP per Capita (Current US$)." *The World Bank Databank*, World Bank National Accounts Data, and OECD National Accounts Data Files., 7 Mar. 2019, data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2015&start=1960.

Works Cited

Buckley, Chris, and Amy Qin. “Muslim Detention Camps Are Like 'Boarding Schools,' Chinese Official Says.” The New York Times, The New York Times, 12 Mar. 2019, www.nytimes.com/2019/03/12/world/asia/china-xinjiang.html.

solid IT. “Engines Ranking.” Knowledge Base of Relational and NoSQL Database Management Systems, 1 Dec. 2019, db-engines.com/en/ranking.

Constantinov, Calin, Mihai L. Mocanu, and Cosmin M. Poteras. "Running complex queries on a graph database: A performance evaluation of neo4j." Annals of the University of Craiova 12.1 (2015): 38-44.

Factor, Phil. “Statistics in SQL: Pearson's Correlation.” *Simple Talk*, Simple Talk, 3 Aug. 2017, www.red-gate.com/simple-talk/blogs/statistics-sql-pearsons-correlation/.

Lemke, Douglas and Werner, Suzanne. 1996. “Power Parity, Commitment to Change, and War”. International Studies Quarterly, 40: 235–260.

Li, Jiexing, et al. "Robust estimation of resource consumption for sql queries using statistical techniques." Proceedings of the VLDB Endowment 5.11 (2012): 1555-1566.

Organski, A. F. K. 1958. World Politics, New York: Alfred A. Knopf.

Prasad, Bakshi Rohit, and Sonali Agarwal. "Comparative Study of Big Data Computing and Storage Tools." International Journal of Database Theory and Application 9.1 (2016): 45-66.

Ray, James Lee and Russett, Bruce. 1996. “The Future as Arbiter of Theoretical Controversies: Predictions, Explanations, and the End of the Cold War”. British Journal of Political Science

Wemer, Suzanne and Kugler, Jacek. 1996. “Power Transitions and Military Buildups”. In Parity and War: Evaluations and Extensions of “The War Ledger”, Ann Arbor, MI: The University of Michigan Press.

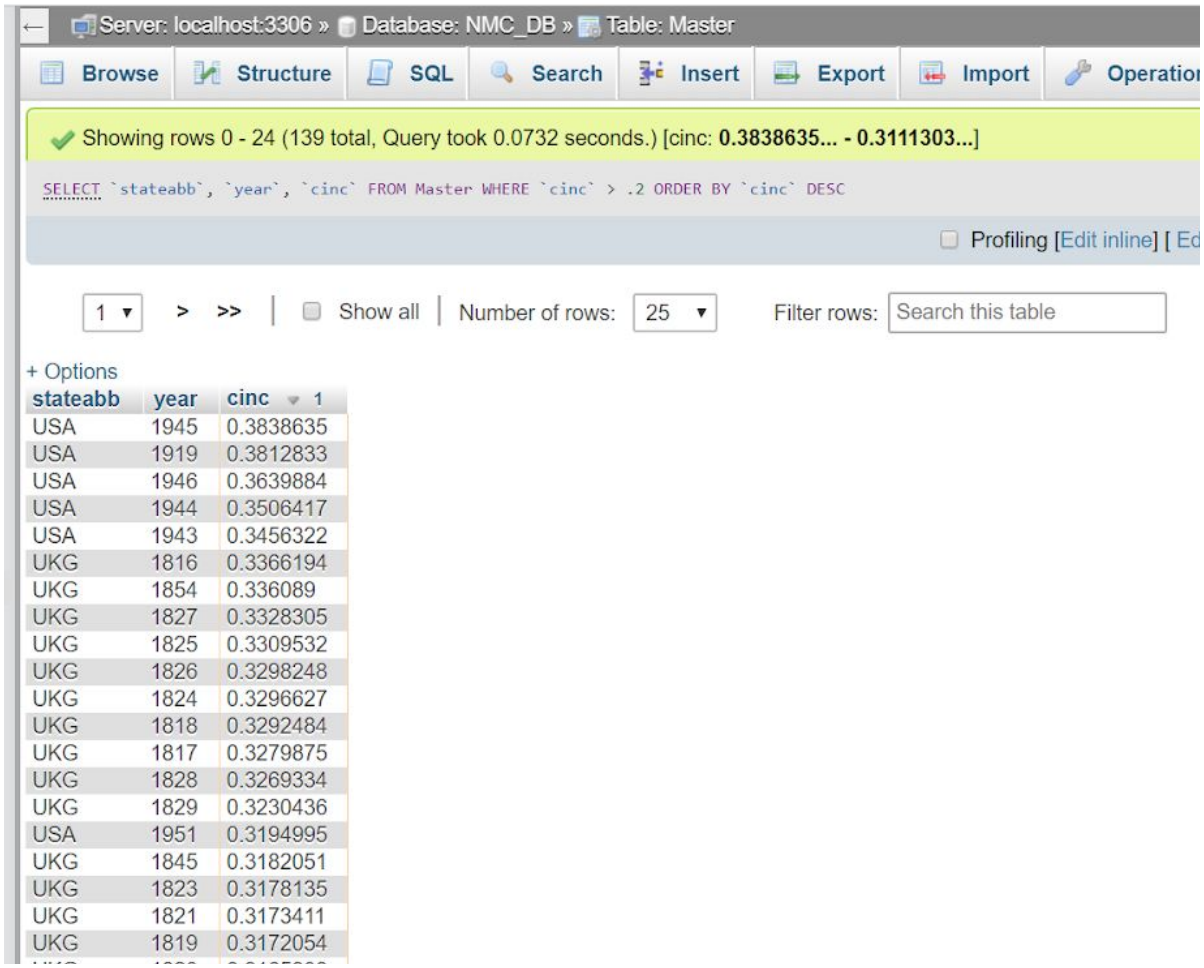
Appendix

SQL and R Code Chunks (CC)

1.

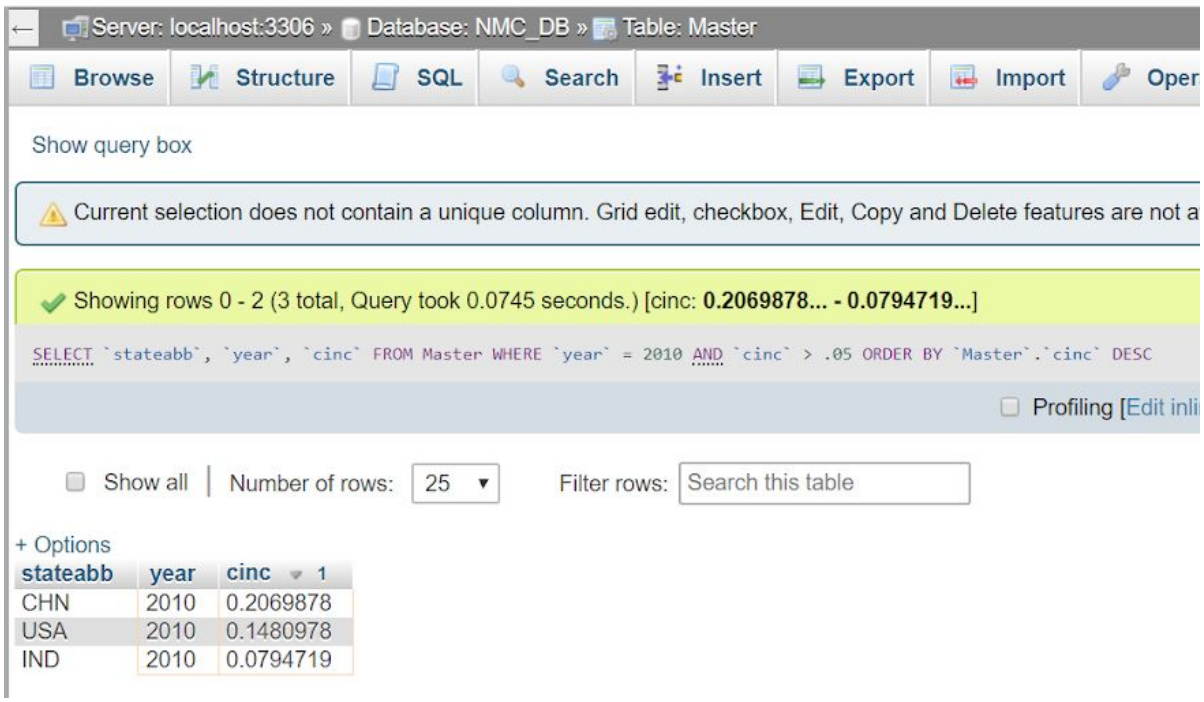
```r
# clean up workspace environment
rm(list = ls())

# all packages used for the assignment
library(tidyr)
library(plyr)
library(readr)
library(dplyr)
library(ggmap)
library(party)
library(rpart)
library(rpart.plot)
library(reticulate)
library(DataComputing)

# Read the NMC data in.
NMC<-readr::read_csv("NMC_5_0.csv")

# Read the World Bank Data
WBD <- readr::read_csv("wbd.csv")

# Inspect the data
head(NMC)
head(WBD)
nrow(NMC)
nrow(WBD)

# Gather function to get World Bank Data (WBD) to the same narrow format as the National Material Capabilites (NMC) data
# Gather all the years, GDP, by all the years less the county code
WBD <-
  WBD %>%
  gather(year, GDP, -'Country Code') # Perserves country code column

head(WBD)

# Now we are going to join WBD GDP data with NMC data by Country Code and Year -- note: there are going to be records only from 1960 and beyond.

# Rename country code to stateabb for ease of natural join (plyr package)
WBD <-
  WBD %>%
  plyr::rename(c("Country Code"="stateabb"))

# Make WBD year column data type numeric so it is consistent with NMC
WBD$year <- as.numeric(WBD$year)

# Left join by year and state abbreviation equivalency
combinedDataFrame <-
  NMC %>%
  left_join(WBD, by = c("year"="year","stateabb"="stateabb"))

```

2.



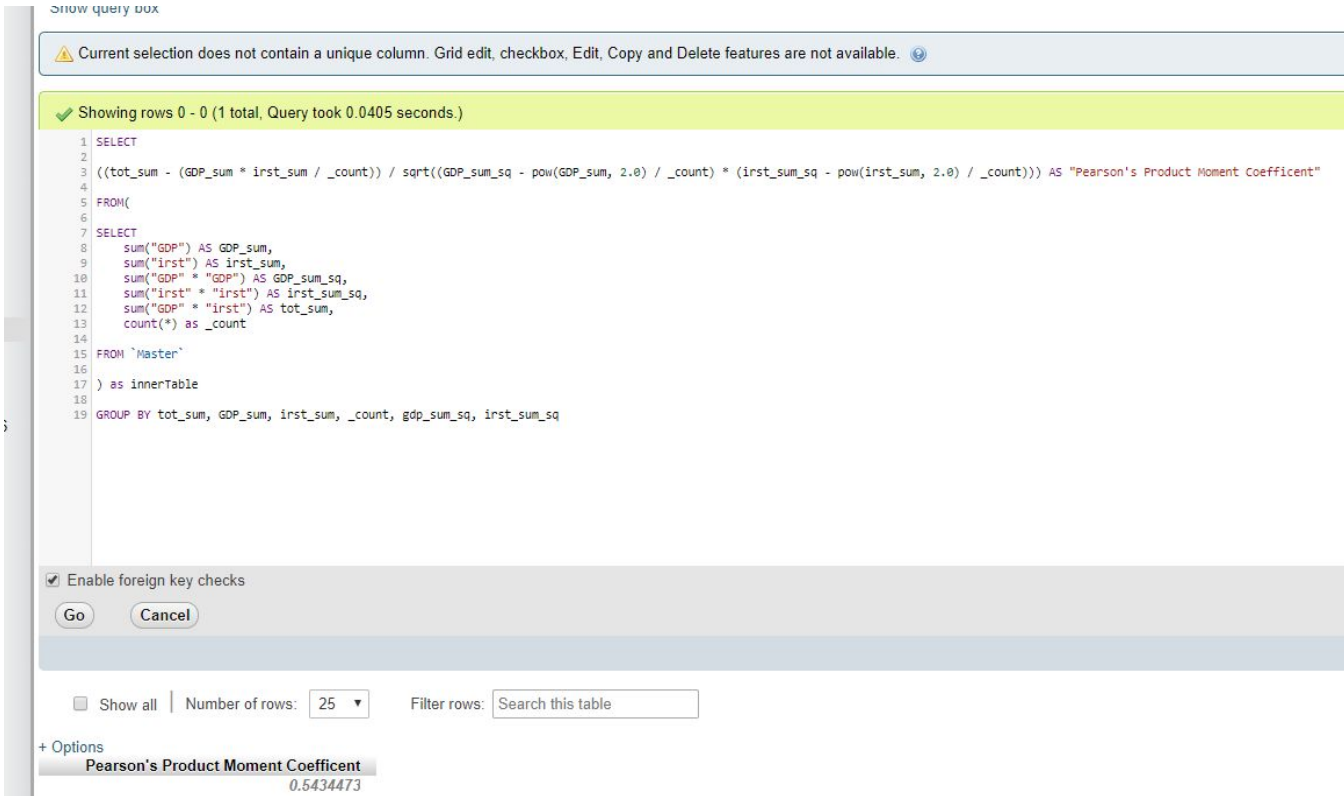
stateabb	year	cinc ▼ 1
USA	1945	0.3838635
USA	1919	0.3812833
USA	1946	0.3639884
USA	1944	0.3506417
USA	1943	0.3456322
UKG	1816	0.3366194
UKG	1854	0.336089
UKG	1827	0.3328305
UKG	1825	0.3309532
UKG	1826	0.3298248
UKG	1824	0.3296627
UKG	1818	0.3292484
UKG	1817	0.3279875
UKG	1828	0.3269334
UKG	1829	0.3230436
USA	1951	0.3194995
UKG	1845	0.3182051
UKG	1823	0.3178135
UKG	1821	0.3173411
UKG	1819	0.3172054

3.

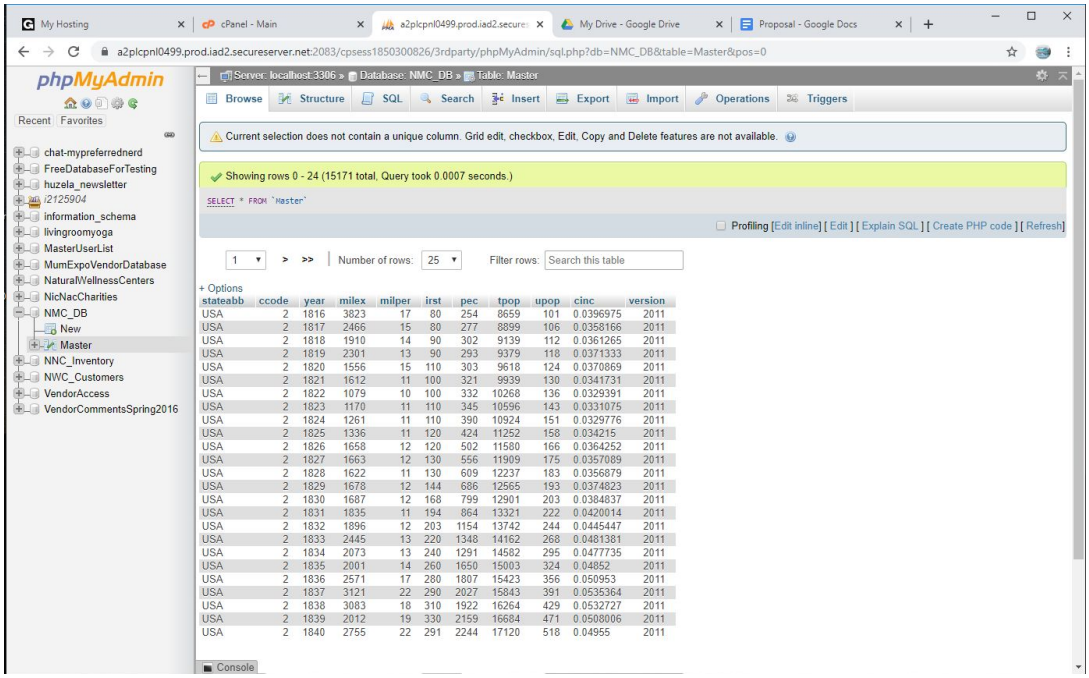


stateabb	year	cinc ▼ 1
CHN	2010	0.2069878
USA	2010	0.1480978
IND	2010	0.0794719

4.


Show query box
Current selection does not contain a unique column. Grid edit, checkbox, Edit, Copy and Delete features are not available.
Showing rows 0 - 0 (1 total, Query took 0.0405 seconds.)
SELECT
((tot_sum - (GDP_sum * irst_sum / _count)) / sqrt((GDP_sum_sq - pow(GDP_sum, 2.0) / _count) * (irst_sum_sq - pow(irst_sum, 2.0) / _count))) AS "Pearson's Product Moment Coefficent"
FROM(
SELECT
sum("GDP") AS GDP_sum,
sum("irst") AS irst_sum,
sum("GDP" * "GDP") AS GDP_sum_sq,
sum("irst" * "irst") AS irst_sum_sq,
sum("GDP" * "irst") AS tot_sum,
count(*) as _count
FROM `Master`
) as innerTable
GROUP BY tot_sum, GDP_sum, irst_sum, _count, gdp_sum_sq, irst_sum_sq
Enable foreign key checks
Go
Cancel
Show all
Number of rows: 25
Filter rows: Search this table
+ Options
Pearson's Product Moment Coefficent
0.5434473

5.

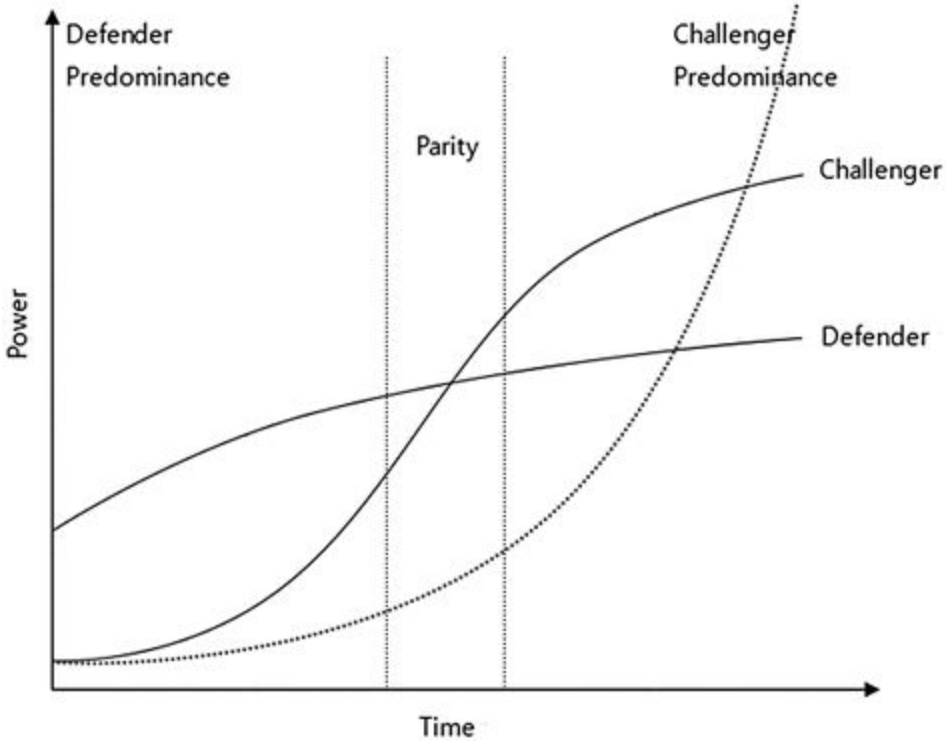

phpMyAdmin
Server: localhost:3306 » Database: NMC_DB » Table: Master
Browse
Structure
SQL
Search
Insert
Export
Import
Operations
Triggers
Current selection does not contain a unique column. Grid edit, checkbox, Edit, Copy and Delete features are not available.
Showing rows 0 - 24 (15171 total, Query took 0.0007 seconds.)
SELECT * FROM `Master`
Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]
Number of rows: 25
Filter rows: Search this table
+ Options
stateabb ccode year milex milper irst pec tpop upop cinc version
USA 2 1816 3823 17 80 254 8659 101 0.0396975 2011
USA 2 1817 2466 15 80 277 8899 106 0.0358166 2011
USA 2 1818 1910 14 90 302 9139 112 0.0361265 2011
USA 2 1819 2301 13 90 293 9379 118 0.0371333 2011
USA 2 1820 1556 15 110 303 9618 124 0.0370869 2011
USA 2 1821 1612 11 100 321 9939 130 0.0341731 2011
USA 2 1822 1079 10 100 332 10268 136 0.0329391 2011
USA 2 1823 1170 11 110 345 10596 143 0.0331075 2011
USA 2 1824 1261 11 110 390 10924 151 0.0329776 2011
USA 2 1825 1336 11 120 424 11252 158 0.034215 2011
USA 2 1826 1658 12 120 502 11580 166 0.0364252 2011
USA 2 1827 1663 12 130 556 11909 175 0.0357089 2011
USA 2 1828 1622 11 130 609 12237 183 0.0356879 2011
USA 2 1829 1678 12 144 686 12565 193 0.0374823 2011
USA 2 1830 1687 12 168 799 12901 203 0.0384837 2011
USA 2 1831 1835 11 194 864 13321 222 0.0420014 2011
USA 2 1832 1896 12 203 1154 13742 244 0.0445447 2011
USA 2 1833 2445 13 220 1348 14162 268 0.0481381 2011
USA 2 1834 2073 13 240 1291 14582 295 0.0477735 2011
USA 2 1835 2001 14 260 1650 15003 324 0.04852 2011
USA 2 1836 2571 17 280 1807 15423 356 0.050953 2011
USA 2 1837 3121 22 290 2027 15843 391 0.0535364 2011
USA 2 1838 3083 18 310 1922 16264 429 0.0532727 2011
USA 2 1839 2012 19 330 2159 16684 471 0.0508006 2011
USA 2 1840 2755 22 291 2244 17120 518 0.04955 2011
Console

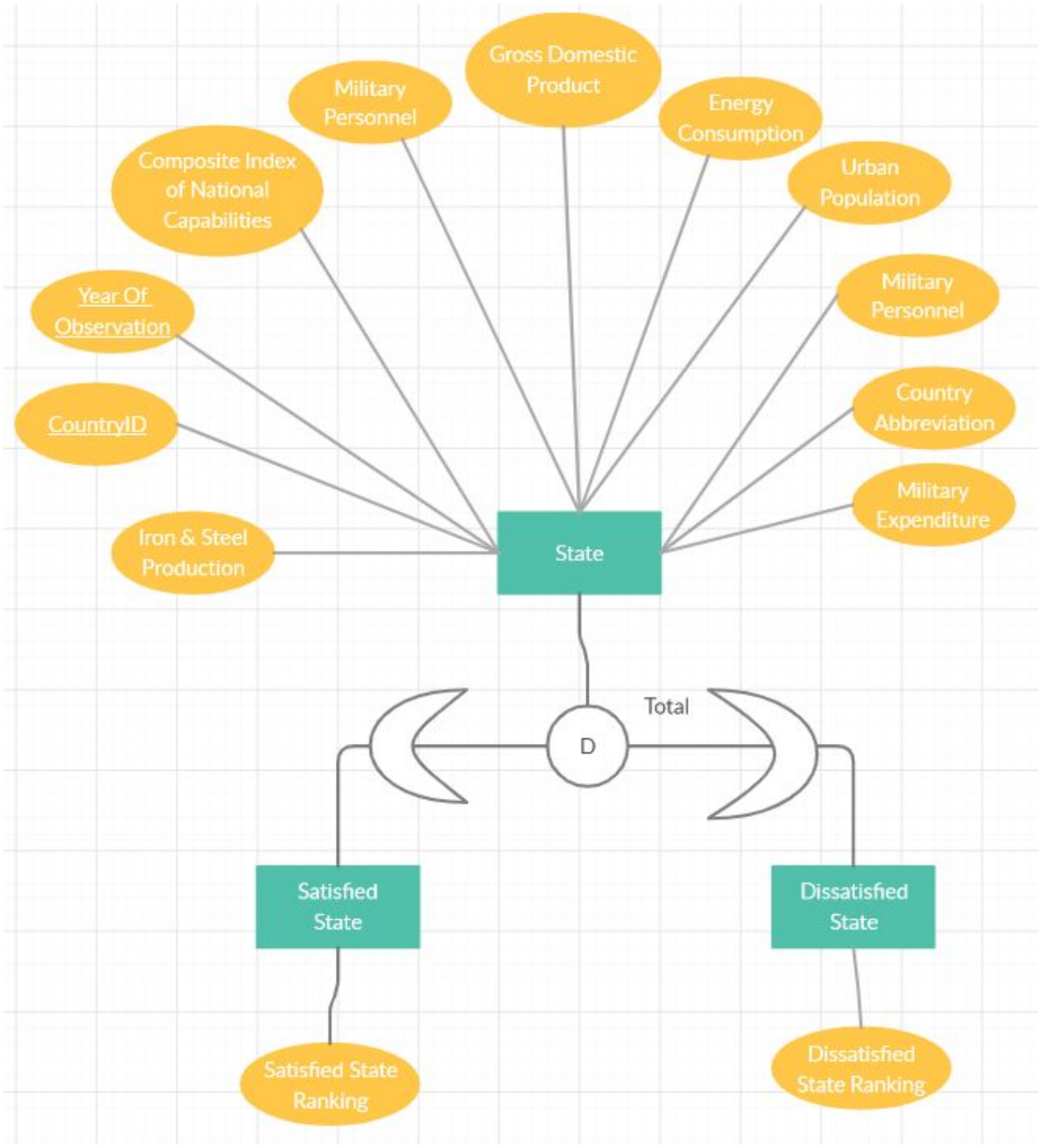
Other Figures (OF)

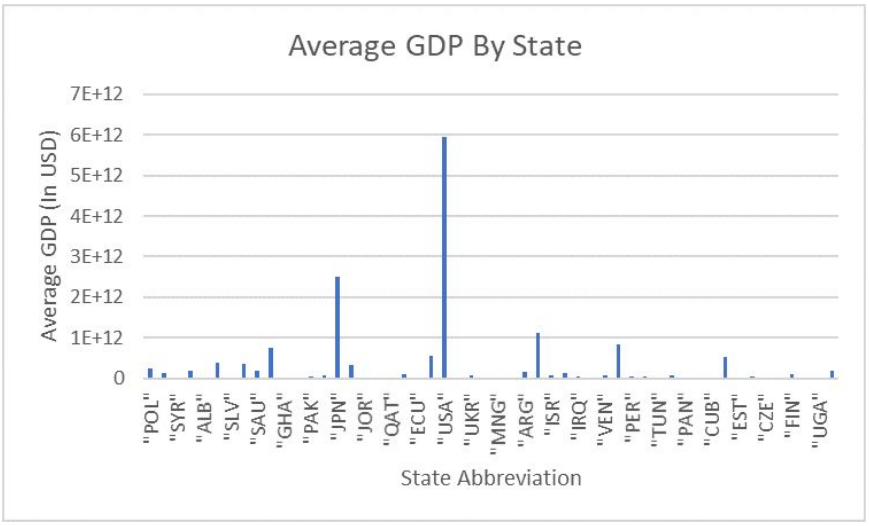
$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[\, n\Sigma x^2 - (\Sigma x)^2 \,][\, n\Sigma y^2 - (\Sigma y)^2 \,]}}$$

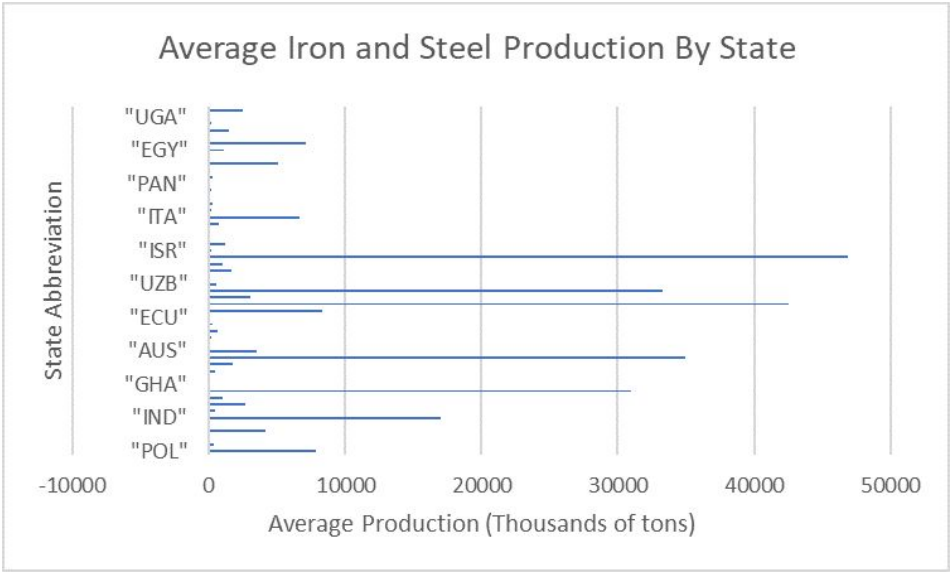
1.

2.



3. 
Gross Domestic Product
Military Personnel
Energy Consumption
Composite Index of National Capabilities
Urban Population
Year Of Observation
Military Personnel
CountryID
Country Abbreviation
Military Expenditure
Iron & Steel Production
State
Total
D
Satisfied State
Dissatisfied State
Satisfied State Ranking
Dissatisfied State Ranking

4. 

5. 

Mongo Code Chunks (MC)

1. db.NMC.find({"cinc":{$gt: 0.15}, "year":{$gt: 2000}}, {"_id":0, "stateabb":1, "cinc":1, "milex":1, "milper":1,"irst":1,"pec":1, "tpop":1, "upop":1})

```
db.NMC.find({"cinc":{$gt: 0.15}, "year":{$gt: 2000}}, {"_id":0, "stateabb":1, "cinc":1, "milex":1, "milper":1,"irst":1,"pec":1, "tpop":1, "upop":1})
{ "stateabb" : "USA", "milex" : 348555000, "milper" : 1414, "irst" : 91587, "pec" : 3227356, "tpop" : 290270, "upop" : 157321, "cinc" : 0.1537646 }1})
{ "stateabb" : "USA", "milex" : 404920000, "milper" : 1427, "irst" : 93677, "pec" : 3243291, "tpop" : 292883, "upop" : 159669, "cinc" : 0.1519185 }
{ "stateabb" : "USA", "milex" : 455908000, "milper" : 1450, "irst" : 99681, "pec" : 3267987, "tpop" : 295487, "upop" : 162073, "cinc" : 0.1515293 }
{ "stateabb" : "USA", "milex" : 495326000, "milper" : 1474, "irst" : 94897, "pec" : 3267376, "tpop" : 298166, "upop" : 164533, "cinc" : 0.1565189 }
{ "stateabb" : "USA", "milex" : 521840000, "milper" : 1546, "irst" : 98557, "pec" : 3245355, "tpop" : 300943, "upop" : 166465, "cinc" : 0.1544305 }
{ "stateabb" : "CHN", "milex" : 46049000, "milper" : 2310, "irst" : 150906, "pec" : 1914161, "tpop" : 1271850, "upop" : 612933, "cinc" : 0.1654946 }
{ "stateabb" : "CHN", "milex" : 68963000, "milper" : 2270, "irst" : 182249, "pec" : 2023332, "tpop" : 1295322, "upop" : 294634, "cinc" : 0.152607 }
{ "stateabb" : "CHN", "milex" : 75500000, "milper" : 2250, "irst" : 222336, "pec" : 2458728, "tpop" : 1302810, "upop" : 307049, "cinc" : 0.1601365 }
{ "stateabb" : "CHN", "milex" : 29873000, "milper" : 2255, "irst" : 355790, "pec" : 3137042, "tpop" : 1318177, "upop" : 334517, "cinc" : 0.1713495 }
{ "stateabb" : "CHN", "milex" : 35223000, "milper" : 2255, "irst" : 421024, "pec" : 3468714, "tpop" : 1326146, "upop" : 347428, "cinc" : 0.1787783 }
{ "stateabb" : "CHN", "milex" : 84303000, "milper" : 2253, "irst" : 272798, "pec" : 2861534, "tpop" : 1310414, "upop" : 319782, "cinc" : 0.1681193 }
{ "stateabb" : "CHN", "milex" : 46174000, "milper" : 2255, "irst" : 489712, "pec" : 3706324, "tpop" : 1334344, "upop" : 360866, "cinc" : 0.185799 }
{ "stateabb" : "CHN", "milex" : 60187000, "milper" : 2105, "irst" : 512339, "pec" : 3857902, "tpop" : 1342733, "upop" : 376397, "cinc" : 0.1893755 }
{ "stateabb" : "CHN", "milex" : 70381000, "milper" : 2185, "irst" : 577070, "pec" : 4197495, "tpop" : 1351248, "upop" : 391335, "cinc" : 0.2081897 }
{ "stateabb" : "CHN", "milex" : 76361000, "milper" : 2285, "irst" : 638743, "pec" : 4529730, "tpop" : 1359821, "upop" : 407821, "cinc" : 0.2069878 }
{ "stateabb" : "CHN", "milex" : 90221000, "milper" : 2285, "irst" : 683883, "pec" : 5043897, "tpop" : 1368440, "upop" : 423604, "cinc" : 0.2122435 }
{ "stateabb" : "CHN", "milex" : 102643000, "milper" : 2285, "irst" : 731040, "pec" : 5333707, "tpop" : 1377065, "upop" : 440254, "cinc" : 0.2181166 }
>
```

2. db.NMC.explain("executionStats").find({"cinc":{$gt: 0.15}, "year":{$gt: 2000}}, {"_id":0, "stateabb":1, "cinc":1, "milex":1, "milper":1,"irst":1,"pec":1, "tpop":1, "upop":1})

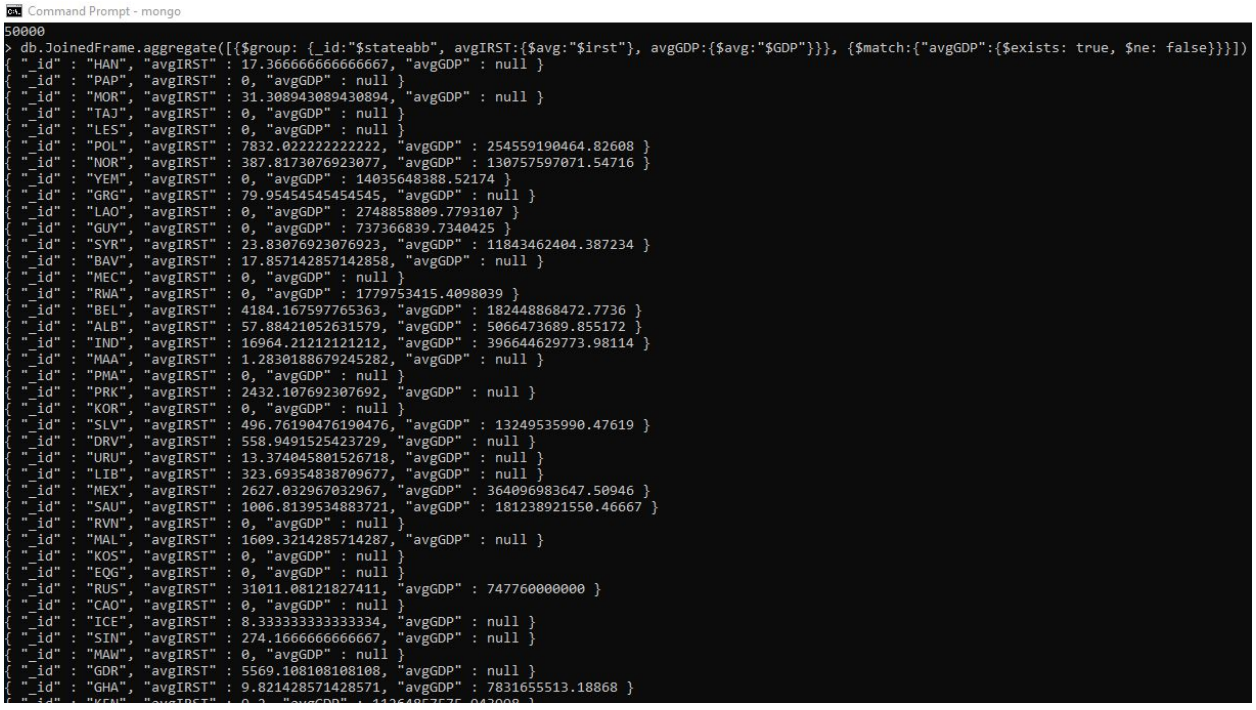
```
},
"executionStats" : {
        "executionSuccess" : true,
        "nReturned" : 17,
        "executionTimeMillis" : 11,
        "totalKeysExamined" : 0,
        "totalDocsExamined" : 15171,
        "executionStages" : {
                "stage" : "PROJECTION_SIMPLE",
```

3. db.NMC.aggregate([{$group: {_id:"$stateabb", avgIRST:{$avg:"$irst"}}}])

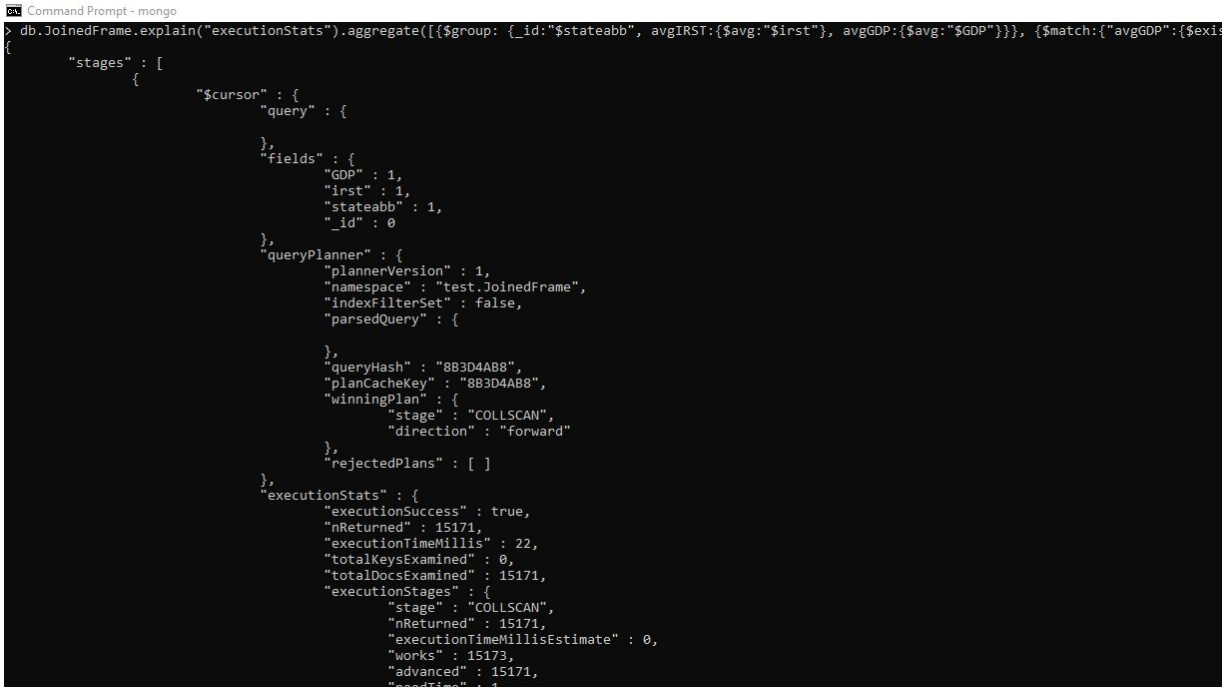
```
Command Prompt - mongo
> db.NMC.aggregate([{$group: {_id:"$stateabb", avgIRST:{$avg:"$irs
{ "_id" : "NIR", "avgIRST" : 1.0377358490566038 }
{ "_id" : "CHN", "avgIRST" : 46914.60130718954 }
{ "_id" : "FSM", "avgIRST" : 0 }
{ "_id" : "TUS", "avgIRST" : 0 }
{ "_id" : "FIN", "avgIRST" : 1450.96875 }
{ "_id" : "CAO", "avgIRST" : 0 }
{ "_id" : "MAC", "avgIRST" : 203.45 }
{ "_id" : "SWD", "avgIRST" : 1662.989847715736 }
{ "_id" : "GAB", "avgIRST" : 0 }
{ "_id" : "BAH", "avgIRST" : 0 }
{ "_id" : "COL", "avgIRST" : 164.2967032967033 }
{ "_id" : "JPN", "avgIRST" : 34912.6462585034 }
{ "_id" : "WSM", "avgIRST" : 0 }
{ "_id" : "TON", "avgIRST" : 0 }
{ "_id" : "BUI", "avgIRST" : 0 }
{ "_id" : "TAJ", "avgIRST" : 0 }
{ "_id" : "BOL", "avgIRST" : 0 }
{ "_id" : "BHM", "avgIRST" : 0 }
{ "_id" : "MLD", "avgIRST" : 690.2272727272727 }
{ "_id" : "GUI", "avgIRST" : 0 }
Type "it" for more
> DBQuery.shellBatchSize = 50000
50000
> db.NMC.aggregate([{$group: {_id:"$stateabb", avgIRST:{$avg:"$irs
{ "_id" : "SLV", "avgIRST" : 496.76190476190476 }
{ "_id" : "GAB", "avgIRST" : 0 }
```

4. db.NMC.explain("executionStats").aggregate([{$group: {_id:"$stateabb", avgIRST:{$avg:"$irst"}}}])

5. db.JoinedFrame.aggregate([{$group: {_id:"$stateabb", avgIRST:{$avg:"$irst"}, avgGDP:{$avg:"$GDP"}}}, {$match:{"avgGDP":{$exists: true, $ne: false}}}])



6. db.JoinedFrame.explain("executionStats").aggregate([{$group: {_id:"$stateabb", avgIRST:{$avg:"$irst"}, avgGDP:{$avg:"$GDP"}}}, {$match:{"avgGDP":{$exists: true, $ne: false}}}])



7. Import statement of joined data

```
Microsoft Windows [Version 10.0.17763.914]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\dchad>cd C:\Program Files\

C:\Program Files>cd MongoDB\Server\4.2\bin

C:\Program Files\MongoDB\Server\4.2\bin>mongoimport --db test --collection JoinedFrame --type csv --file :D\Join.csv --headerline
2019-12-12T20:31:55.401-0500    Failed: open :D\Join.csv: The filename, directory name, or volume label syntax is incorrect.
2019-12-12T20:31:55.403-0500    0 document(s) imported successfully. 0 document(s) failed to import.

C:\Program Files\MongoDB\Server\4.2\bin>mongoimport --db test --collection JoinedFrame --type csv --file :D\Join.csv --headerline
2019-12-12T20:32:16.108-0500    Failed: open :D\Join.csv: The filename, directory name, or volume label syntax is incorrect.
2019-12-12T20:32:16.109-0500    0 document(s) imported successfully. 0 document(s) failed to import.

C:\Program Files\MongoDB\Server\4.2\bin>mongoimport --db test --collection JoinedFrame --type csv --file D:\Join.csv --headerline
2019-12-12T20:33:17.124-0500    connected to: mongodb://localhost/
2019-12-12T20:33:17.503-0500    15171 document(s) imported successfully. 0 document(s) failed to import.

C:\Program Files\MongoDB\Server\4.2\bin>
```