War: What Is It Good For?

by Jeffrey Duong, Nancy Zha, Natasha Leodjaja, and Ryan Scibetta

Author contributions

- Jeffrey Duong contributed Principal Component computations, with a focus on dataset creation.
- Nancy Zha contributed Principal Component visualizations, with a focus on dataset manipulation.
- Natasha Leodjaja contributed regression models, with a focus on graphs and visualizations.
- Ryan Scibetta contributed the correlation matrix, with a focus on report writing and analysis.
- The work was evenly split, and everyone helped each other out.

Abstract

Prepare an abstract *after* you've written the entire report. The abstract should be 4-6 sentences summarizing the report contents. Typically:

How much money should a nation's government invest in its military? This report seeks to analyze how the military spending of a country correlates with its citizens' quality of life. We found that while larger military spending is negatively correlated with most quality-of-life indicators, the connection was far too noisy to conclude with any degree of certainty. In the process of our report, we also discovered that life expectancy, access to electricity, and government health care expenditure are strongly directly correlated.

Introduction

Background and Aims

This analysis concerns itself with goverment spending. How does a country's military spending correlate with quality-of-life for its citizens? Do nations with high military budgets tend to create better lives for their citizens? This report helps bring statistics to an interdisciplinary field with work by political scientists, economists, lawyers, and public administrators, and politicians. Our analysis can help inform politicians and public administrators on how to help their constituents, and if they don't listen, it can help advise the voters to elect someone who will.

Our data is collected by a variety of organizations, assembled by the World Bank. These sum to a collection of indicator variables that we can use as a rough estimate of a citizen's quality of life. We group these indicators into principal components to obtain a more broad understanding of quality-of-life than any one indicator can provide on its own.

We approached our project by curating the dataset/s to use for our visualizations. The majority of the work we did for this project was finding, curating, adjusting, fine-tuning, adding, dropping, pivoting, and sharing observations among countries in our dataset. With help from office hours, we came to a solution where we would group our observations by country, average the observations from 2000-2016, and drop any countries with fewer than four full years of data. We then explored and analyze the quality of life indicators against military expenditure via a correlation matrix, a principle component analysis, and a regression model. This allowed us to see which, and how much, specific indicators are influenced by how much a country spends on their military. We found that there is very little correlation between military expenditure and quality-of-life. Our analysis returned ambiguous results, and thus failed to reject the null hypothesis.

Materials and methods

Datasets

The data consists of quality of life indicators for the citizens living in specific countries, split by country and year. These nine indicators were chosen from our initial group of our initial 16 indicators because they have enough data to overcome missingness issues. Our 'MEG' indicator represents military expenditure, so we treat it seperately from the other eight indicators.

Indicators	Source	Collection Methods
MEG	Stockholm International Peace Research Institute (SIPRI)	Derived from the NATO definition
AER	World Bank	Nationally representative surverys (including censuses)
AL	Food and Agriculture Organization	Questionnaire supplemented with information from official secondary data sources
DGH	World Health Organization Global Health Expenditure database	Health accounts studies and government expenditure records
ICP	International Monetary Fund	Laspeyres formula from International Financial Statistics and data files
LEF	United Nations Population Division	Official statistics report websites, census reports and other statistical publications
LEM	United Nations Population Division	Official statistics report websites, census reports and other statistical publications
REP	Inter-Parliamentary Union (IPU)	Census of lower branch of national parliament
UEMP	International Labor Organization, ILOSTAT database.	Data was collected based on countries' responses

Our dataframe is organized by country and year. Our identifying variables are 'Country Name' / 'Country Code' and 'Year'. ('Country Code' is just an abbreviated version of 'Country Name'.) The population for this dataset is every country in the world, and the sampling frame is all of the

countries that reported some data requested by different institutions. Our sample consists of the countries that reported enough data to make statistical conclusions; therefore, our sample is a conveinience sample, with no scope of inference.

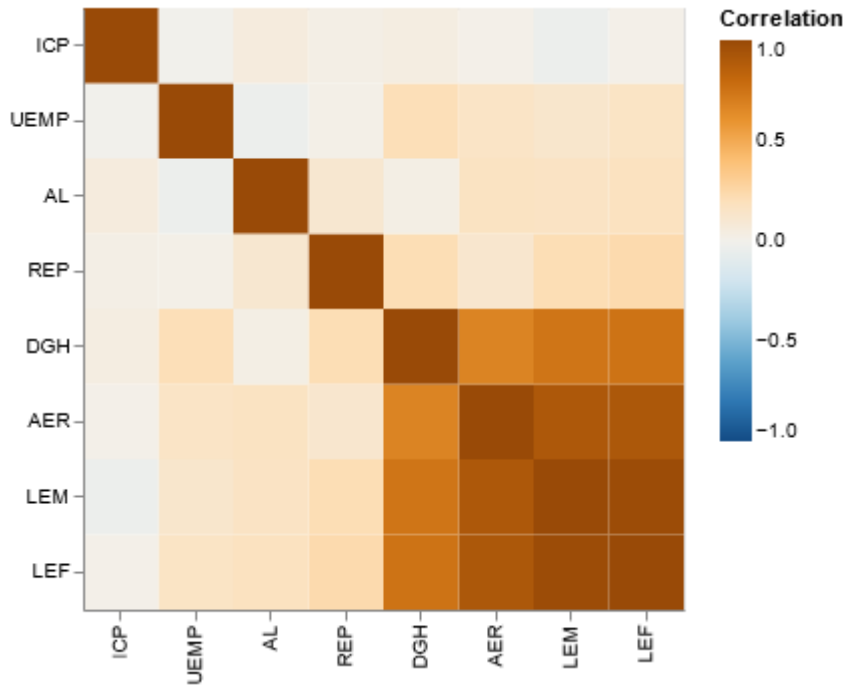
A convienience sample is not ideal, but in this case, it is our best option. We can maintain the integrity of our results by limiting our scope of inference to only the countries sampled. However, our sample still contains 194 observations. By slicing countries with low or no data points, we can guarentee that each indicator consists of an mean of at least four years of data. In addition, by including regional indicators like 'World' and 'Sub-Saharan Africa (excluding high income)', we're able to include data from the entire world, including countries that are not able to provide complete information on their own.

Methods

Exploratory analysis attempted to see if there were any clear correlations between military spending and quality of life indicators using simple scatter plots that mapped military spending and one indicator. Visual inspection did not find any strong links although small possible links can be seen in the graphs. Principal Component Analysis was then done on normalized data to find the indicators that drove the variation the most. The loadings that were each individually responsible for 6% of the variation were used to analyze. Giving a much clearer picture, a few specific indicators can be picked out that most heavily affect the variance. From here, regression models were created to see the general trend for each of the indicators in response to military spending.

Results

Figure 1 (Correlation Matrix)



This graph shows the correlation matrix of our quality-of-life indicator variables. We normalized our observations using Pearson's correlation coefficient, sorted them by ascending sum correlation, and displayed them using a linear blue-orange gradiant. We found that there is a strong positive correlation between health care expenditure, access to electricity, and life expectancy

Figures 2A and 2B (Principle Component Tables)

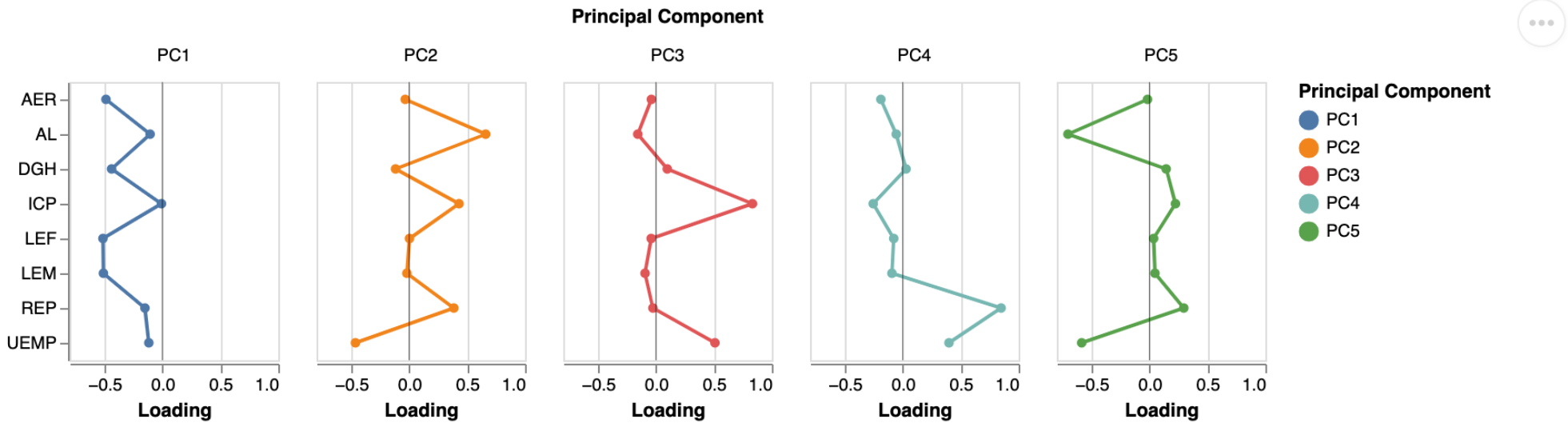
	Proportion of variance explained	Component	Cumulative Varaiance Explained
0	0.453370	1	0.453370
1	0.138709	2	0.592079
2	0.124011	3	0.716090
3	0.116125	4	0.832215
4	0.110554	5	0.942768
5	0.043013	6	0.985782
6	0.012599	7	0.998381
7	0.001619	8	1.000000

	PC1	PC2	PC3	PC4	PC5	Variable
0	-0.487014	-0.033975	-0.039490	-0.188852	-0.017335	AER
1	-0.438611	-0.117772	0.097613	0.028860	0.143839	DGH
2	-0.150635	0.385335	-0.026277	0.849221	0.295661	REP
3	-0.117126	-0.463697	0.509261	0.399379	-0.585470	UEMP
4	-0.105911	0.660262	-0.157048	-0.056202	-0.703668	AL
5	-0.007896	0.430260	0.832836	-0.254775	0.224049	ICP
6	-0.509147	-0.020014	-0.094277	-0.093340	0.047128	LEM
7	-0.513340	0.002594	-0.041324	-0.077721	0.035360	LEF

Our first table shows the proportion of variance each individual component explains as well as the cumulative proportion for multiple components. The first 5 components individually explain at least 11% of the total variation while cumulatively explaining 94% of the total variation. In particular, our first component explains 45% of the variation all on its own!

Our second table shows the weights of the variables that are combined to make up the specific principal component. Each row is a weight for one indicator in the dataset, and each column is a unique set of weights for the component.

Figure 3 (Principle Component Analysis)



Principal Component Analysis found that 5 principal components were responsible for over 94% of the total variation within the data. These components were found with normalized data and are compared to see which specific indicators drive each of the components.

The first component is driven heavily by low Access to Electricity (AER), low Domestic Government Health Expenditure (DGH), low Life Expectancy for Males (LEM), and low Life Expectancy for Females (LEF). These are the four strongly correlated variables in our correlation matrix, and having low values in all four may indicate an undeveolped country.

The second component is driven by low Unemployment Percentage (UEMP) and high Arable Land (AL). This may indicate a country consisting of family farmers; they'd have the farmable land, and the people would need to farm that land. Interestingly, the fifth component is also driven by low Unemployment Percentage (UEMP), but it has low Arable Land (AL) as well. This may reflect a well-off country that imports its crops.

The third component is explained by Inflation of Consumer Prices (ICP), and the fourth component is explained primarily by Government Representation (REP).

Figures 4A-4H (Regressions)

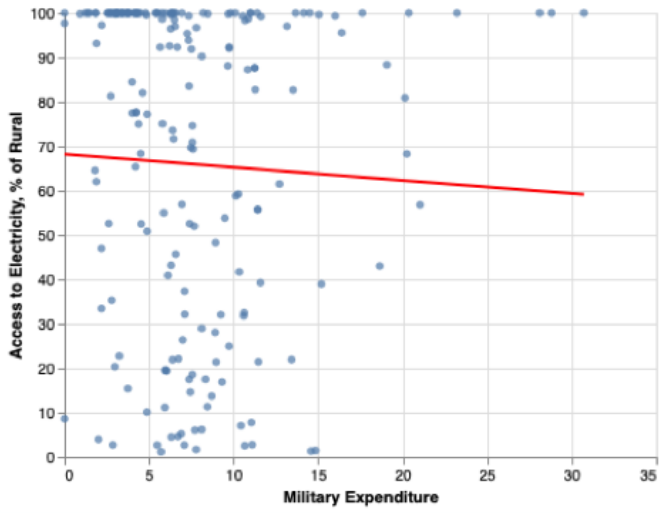


Figure 4A:
Military Expenditure vs Access to Electricity

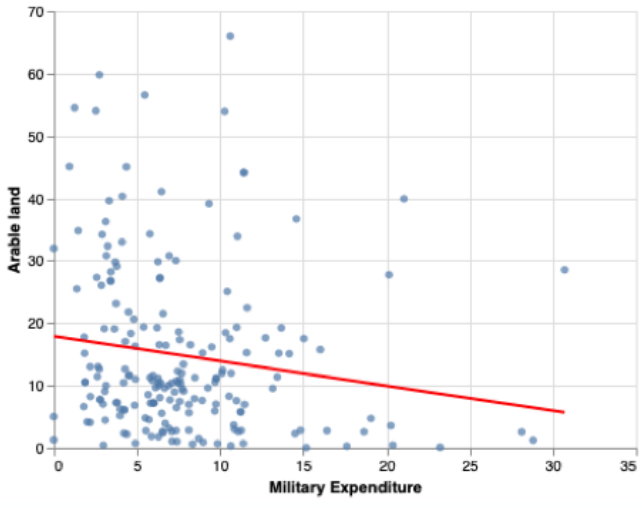


Figure 4B:
Military Expenditure vs Arable LandMilitary

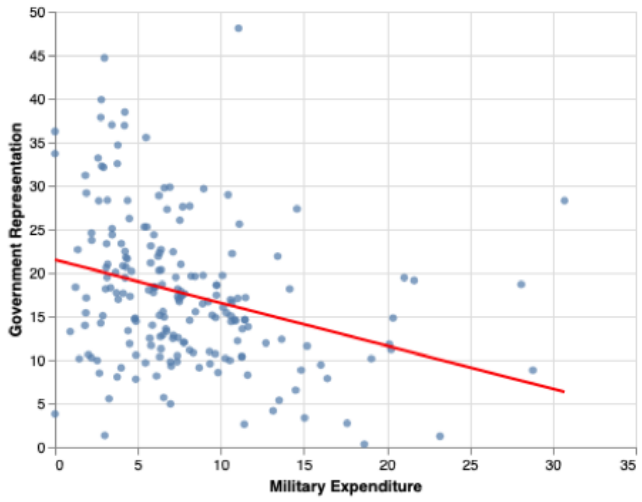


Figure 4C:
Expenditure vs Government Representation

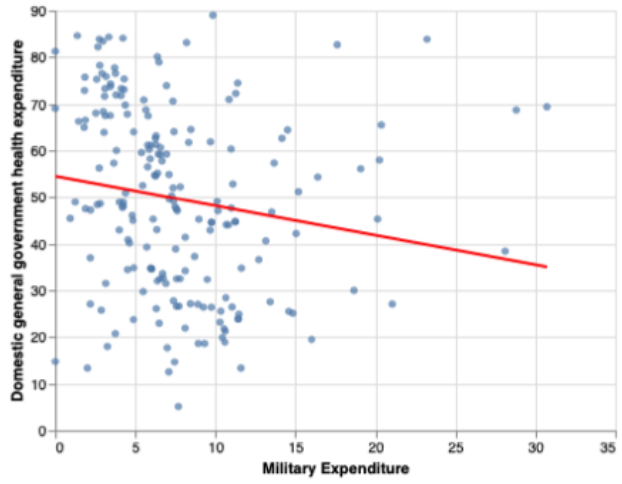


Figure 4D:
Military Expenditure vs Health Expenditure

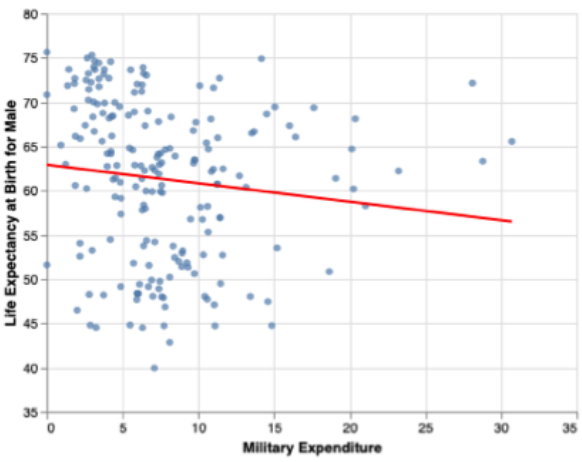


Figure 4E:
Military Expenditure vs Life Expectancy at Birth, Male

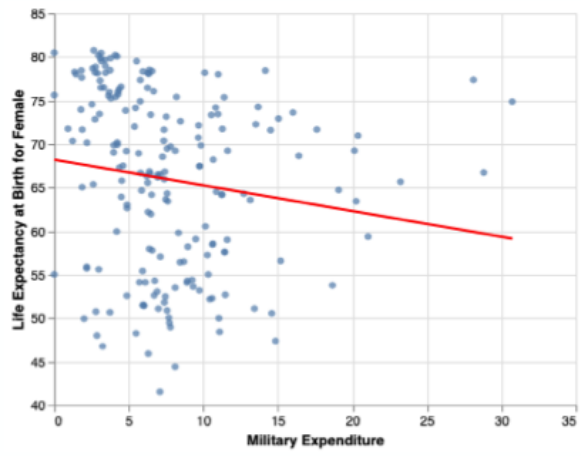


Figure 4F:
Military Expenditure vs Life Expectancy at Birth, Female

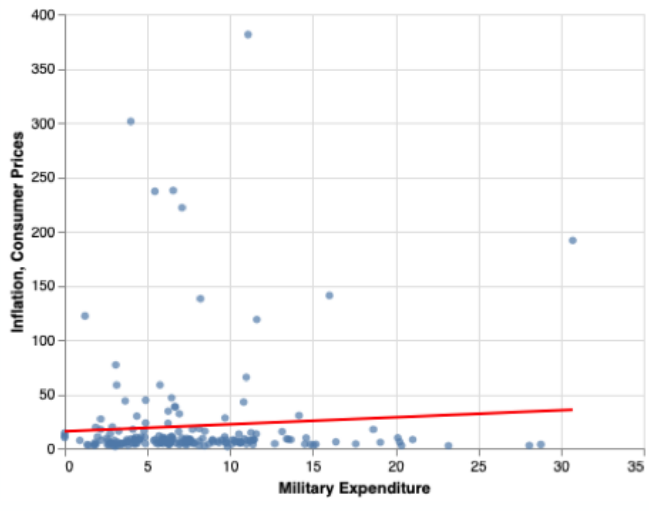


Figure 4G:
Military Expenditure vs Inflation, Consumer Prices

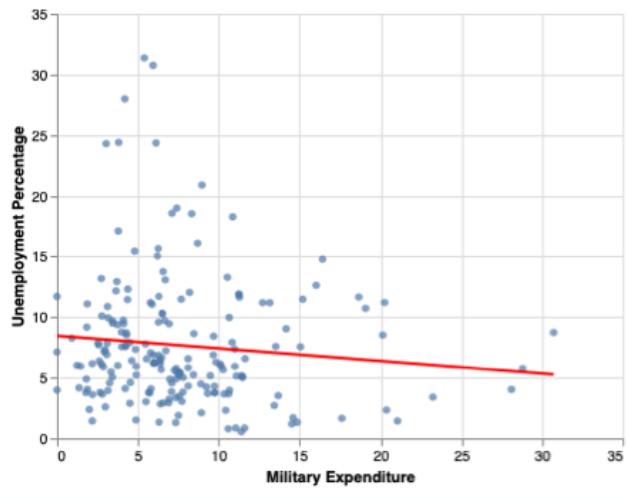


Figure 4H:
Military Expenditure vs Unemployment

For the regression, we decided to use simple linear regression and plot eight graphs for each of our eight quality-of-life indicators as well as a MLR summary. Based on the MLR R^2 value, we can see that 19.79% of variation in MEG is explained by our other variables.

Analyzing our regression plots, we see that military expenditure is slightly inversely correlated with inflation has a slight upward and stable slope as compared to the rest of the indicators such as government rep, life expectancies, etc with a down slope. Our lines of best fit claim that among the countries reporting data, the mean military expenditure increases with inflation at a rate of 0.01. Whereas the mean military expenditure decreases with other indicators such as life expectancy at birth for women at a rate of -0.59. However, based on the R^2 value of each individual indicators, we can see that there is slight to no correlation between MEG with the indicators.

Indicators	R^2 Value
REP	0.0967445355953036
DGH	0.02715509138080518
AL	0.023323568441348395
AER	0.022436696237149476
LEF	0.022436696237149476
LEM	0.014330799563240907
UEMP	0.010106757974956734
ICP	0.004684361900999185

Discussion

This project looked at a handful of quality of life indicators against military expenditure for countries from 2000-2015 that reported at least four full years' data to the respective organizations responsible for each indicator. The analysis focused first on finding indicators that were correlated to each other, using a correlation table (Figure 1). This table showed a strong positive correlation between health care expenditure, access to electricity, and life expectancy (for both males and females). Our analysis then moved to finding which specific indicators drove the most variation in the data. Simple indicator vs military spending plots (Figures 5A-5H) visually showed great variation and using PCA (Figure 2A), 5 main components which drive 94% of the variation were found. The specific indicators which weight most heavily on each component are shown (Figure 2B). Finally, the analysis shifted towards an attempt to fit and predict each indicator as a response to military spending with multiple regression models (Figure 3).

The overall goal of this project was to find correlations between military spending and quality of life for citizens, and the analysis of the project suggests that there may be a slightly negative correlation between the two. Every regression model aside from the Inflation of Consumer Prices model (Figure 4G) has a fitted line with a slightly negative slope, implying a negative correlation with indicators like life expectancy and gender representation in government. However, the error variance for the models is extremely high, and our R^2 values are all extremely low. We cannot confidently conclude that there is a negative correlation between quality of life and military spending.

A way to gain more confidence in the conclusion is to group the countries into regions as richer countries could already have infrastructure in place to ensure quality of life for citizens while still spending large amounts into their military. This could be seen in contrast to countries that are extremely poor and cannot afford putting money into either the military or quality of life. Both of these types of data could be responsible for the extremely high amounts of variation in the data and splitting into regions could help create a clearer picture overall. Our dataset did not include direct indicators of wealth, but given more time, we would be interested in further deveolping our analysis in this vein.

Appendix

Figures 5A-5H (Scatter Plots)

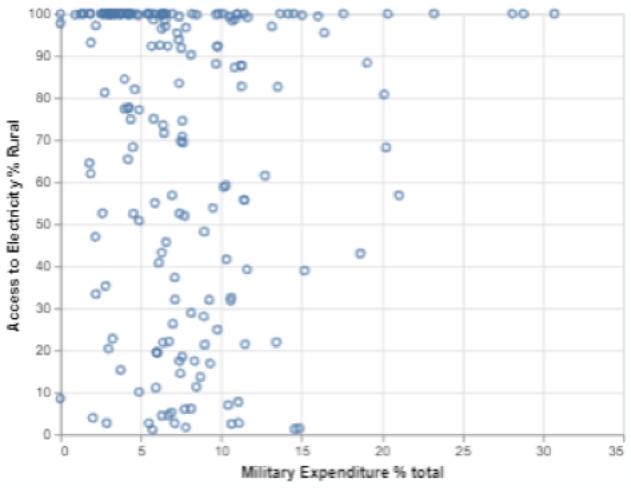


Figure 5A:
Military Expenditure vs Access to Electricity

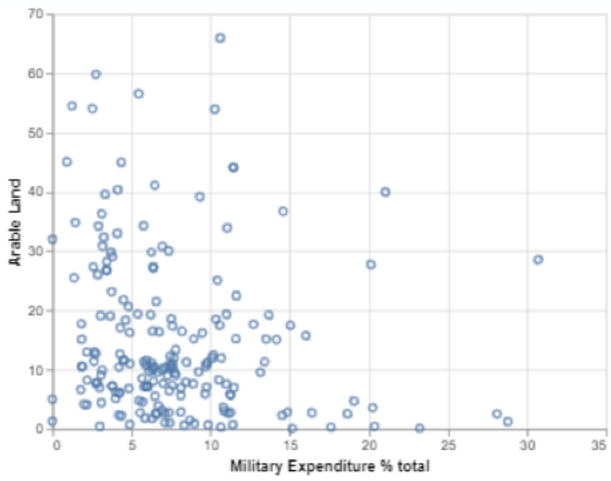


Figure 5B:
Military Expenditure vs Arable LandMilitary

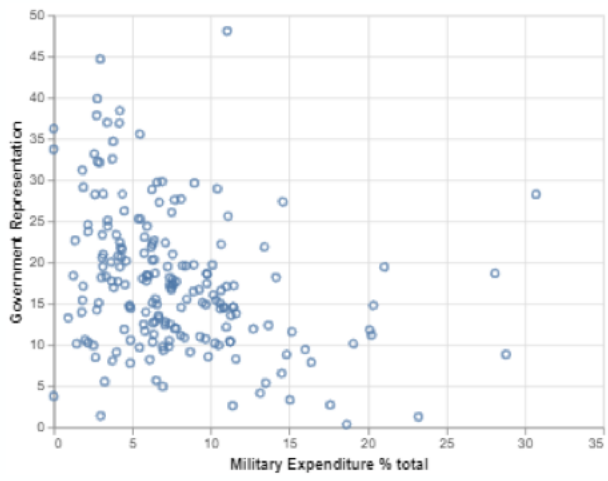


Figure 5C:
Expenditure vs Government Representation

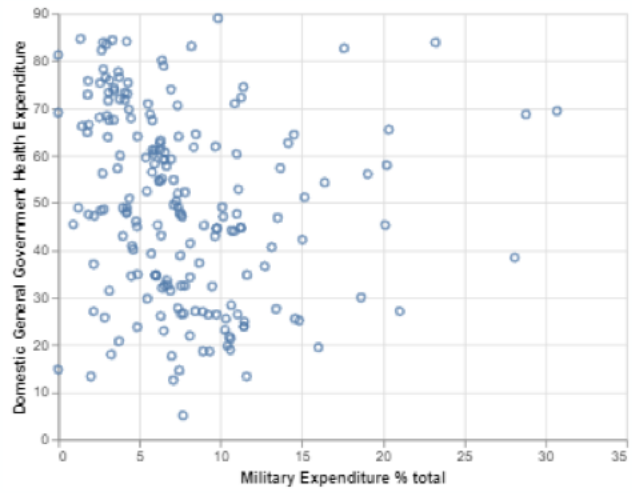


Figure 5D:
Military Expenditure vs Health Expenditure

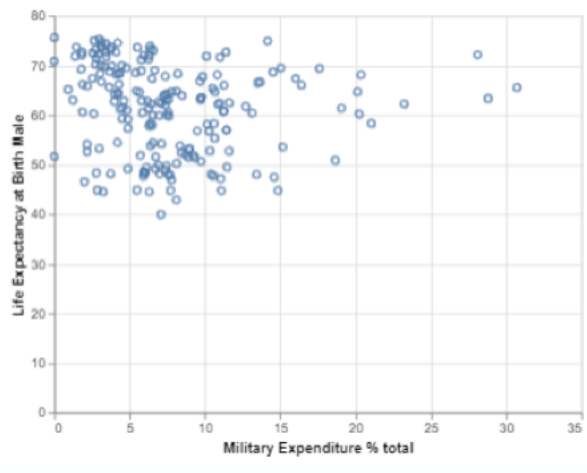


Figure 5E:
Military Expenditure vs Life Expectancy at Birth, Male

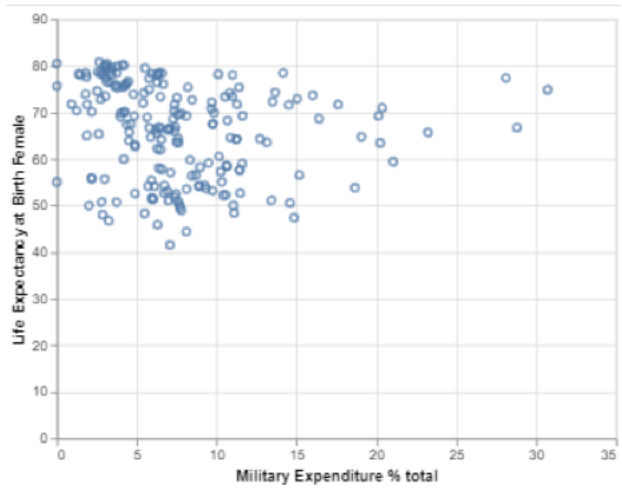


Figure 5F:
Military Expenditure vs Life Expectancy at Birth, Female

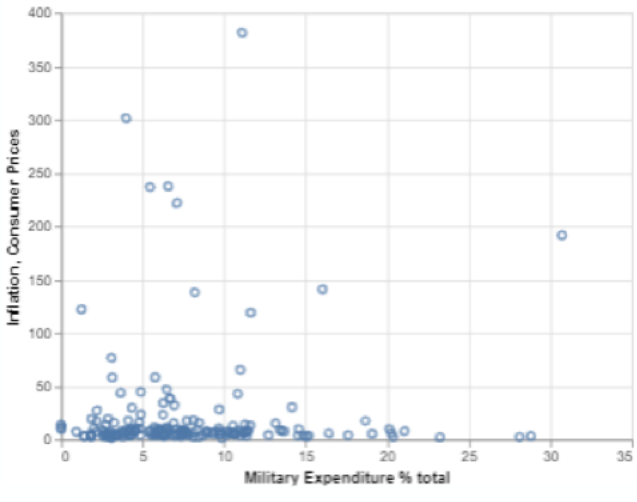


Figure 5G:
Military Expenditure vs Inflation, Consumer Prices

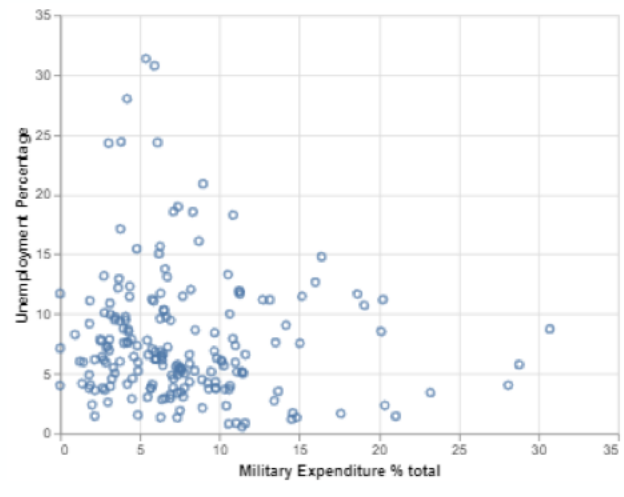


Figure 5H:
Military Expenditure vs Unemployment