```
 # libraries
import numpy as np
import pandas as pd
import altair as alt
```

Group information

Group members:

1. Jeffrey Duong
2. Ryan Scibetta
3. Nancy Zha
4. Natasha Leodjaja

Contributions:

1. Member 1 (Jeffrey): Variables: AER, DRR, GEE, PPC. Additional Tasks: Plot1
2. Member 2 (Ryan): Variables: R&D, REP, TOUR, UEMP. Additional Tasks: copy editing, alphabetization of tables, background paragraphs
3. Member 3 (Nancy): Variables: NEI, AL, ICP, LEF, LEM. Additional Tasks: Plot2 and Plot3
4. Member 4 (Natasha): Variables: DGH, ME, & ITE. Additional Tasks: Uploaded the tidied file, tabled source/collection methods and variable summaries.

Background

Broadly speaking, we are interested in governance. How do different nations choose to spend the money they allocate? And how do those choices affect a country's citizens? This report helps bring statistics to an interdisciplinary field with work by political scientists, economists, lawyers, and public administrators, and, of course, politicians.

Our data is collected by a variety of organizations, each specializing in its field (i.e. Unemployment % sourced by the International Labour Organization). These sum to a collection of indicator variables that we can use as a rough estimate of a citizen's quality of life. We also have indicator variables We can compare our military expenditure indicator/s with our quality of life indicator/s and look for trends and patterns in our data.

There are many different philosophies about military spending. The United States alone spent an estimated $788 billion dollars on defense in 2020, according to the Stockholm International Peace Research Institute. Our project report may be able to shed some light on whether that money was well-spent. Do nations with high military budgets tend to create better lives for their citizens? We can help advise politicians and public administrators on how to help their constituents, and if they don't listen, we can help advise the voters to elect someone who will.

General Description

The data consists of quality of life indicators for the citizens living in specific countries, split by country and year.

Source/Collection Methods

Indicators	Source	Link	Collection
AER	World Bank*	https://data.worldbank.org/indicator/EG.ELC.ACCS.RU.ZS?view=chart	Nationally representative surverys
PPC	UNICEF	https://data.worldbank.org/indicator/SH.STA.ANVC.ZS?view=chart	Women in household surveys
DRR	UNISDR	https://data.worldbank.org/indicator/EN.CLC.DRSK.XQ?view=chart	Countries using average of self re
GEE	UNESCO Institute for Statistics	https://data.worldbank.org/indicator/SE.XPD.TOTL.GB.ZS?view=chart	Official responses to UNESCO's a
DGH	World Health Organization Global Health Expenditure database	https://data.worldbank.org/indicator/SH.XPD.GHED.CH.ZS?view=chart	Health accounts studies and gove
ME	Stockholm International Peace Research Institute (SIPRI)*	https://data.worldbank.org/indicator/MS.MIL.XPND.GD.ZS?view=chart	Derived from the NATO definition
MEG	Stockholm International Peace Research Institute (SIPRI)	https://data.worldbank.org/indicator/MS.MIL.XPND.ZS?view=chart	Derived from the NATO definition

Indicators	Source	Link	Collection
ITE	World Tourism Organization	https://data.worldbank.org/indicator/ST.INT.RCPT.XP.ZS?view=chart	Data was collected based on what travelers
UEMP	International Labor Organization, ILOSTAT database.	https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS?view=chart	Data was collected based on countries' resp
R&D	UNESCO Institute for Statistics	https://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS?view=chart	Self-reported metasurvey from countries' ow
REP	Inter-Parliamentary Union (IPU)	https://data.worldbank.org/indicator/SG.GEN.PARL.ZS?view=chart	Census of lower branch of national parliame
NEI	IEA Statistics © OECD/IEA 2014	https://data.worldbank.org/indicator/EG.IMP.CONS.ZS?view=chart	National energy data adjusted to conform to
AL	Food and Agriculture Organization	https://data.worldbank.org/indicator/AG.LND.ARBL.ZS?view=chart	Questionnaire supplemented with informatio
ICP	International Monetary Fund	https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG	Laspeyres formula from International Financ
LEM	United Nations Population Division	https://data.worldbank.org/indicator/SP.DYN.LE00.MA.IN?view=chart	Official statistics report websites, census rep
LEF	United Nations Population Division	https://data.worldbank.org/indicator/SP.DYN.LE00.FE.IN?view=chart	Official statistics report websites, census rep

Additional sources:

- AL: SE4ALL Global Tracking Framework led jointly by the World Bank, International Energy Agency, and the Energy Sector Management Assistance Program.
- ME: Yearbook: Armaments, Disarmament and International Security.
- MEG: Yearbook: Armaments, Disarmament and International Security.
- ITE: Yearbook of Tourism Statistics, Compendium of Tourism Statistics and data files.
- LEM: National Statistical Offices, Eurostat, United Nations Statistical Division, U.S. Census Bureau, Secretariat of the Pacific Community.
- LEF: National Statistical Offices, Eurostat, United Nations Statistical Division, U.S. Census Bureau, Secretariat of the Pacific Community.

Sampling Design/Scope of Inference

The population for this dataset is every country in the world, and the frame is all of the countries that report the data requested by different institutions. Because the sample is equal to the frame, the sample mechanism is census.

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'.)

This is an administrative dataset so there is no scope of inference. We don't believe our lack of a scope of inference is a significant limitation for our particular topic, because our sample frame encompasses almost every country in the world. While we can't extrapolate to other countries, our data includes all member countries of the World Bank, of which there are 189. We counted 217 data points per year for one of our indicator variables, unemployment.

Name	Variable description	Type	Units of measurement
AER	Access to Electricity, % of Rural	Numeric	% of access
AL	Arable land	Numeric	% of land area
DGH	Domestic general government health expenditure	Numeric	% of current health expenditure
DRR	Disaster Risk Reduction progress	Numeric	Score from 1-5
GEE	Government Expenditure on Education	Numeric	% of total expenditure on education
ICP	inflation, consumer prices	Numeric	annual %
ITE	International tourism, expenditure for travel items	Numeric	Current US$
LEF	life expectancy at birth for female	Numeric	Calendar year
LEM	life expectancy at birth for male	Numeric	Calendar year
PPC	Pregnant women recieving Prenatal Care	Numeric	% of women who visisted health care professional during pregnancy
ME	Military expenditure	Numeric	% of GDP
MEG	Military Expenditure	Numeric	% of total expenditure on military
NEI	Net energy imports	Numeric	% of energy use
R&D	Research and Development	Numeric	% of total expenditure on research and development, as a % of GDP
REP	Government Representation	Numeric	% of women in national parliaments
TOUR	International Tourism	Numeric	ratio of international receipts to total exports
UEMP	unemployment percent	Numeric	Calendar year

```
# load tidied data and print rows
final = pd.read_csv('sample_data/project.csv')
final = final.drop(columns={'Unnamed: 0'})
final.head()
```

	Country Name	Country Code	Year	AER	DRR	GEE	PPC	DGH	ME	ITE	REP	R&D	TOUR	UEMP	MEG	AL	NEI	ICP	LEM
0	Aruba	ABW	1960	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	64.084	6
1	Afghanistan	AFG	1960	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	31.718	3
2	Angola	AGO	1960	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	36.305	3
3	Albania	ALB	1960	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	61.309	6
4	Andorra	AND	1960	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

2. Initial explorations

Basic properties of the dataset

Dimensions: (16104,20)

Missing values: Yes, there are missing values due to countries' relucatance to disclose information.

Variable summaries: Provide simple variable summaries for the most important variables in your dataset. Preferably, you'll do this for all variables, but if you have a large number, you might need to prioritize and focus on the ones most of interest. What exactly you do is a little case-specific, but think of things like means and variances, min/max, number of levels and observation counts for categorical variables, etc.

Indicators	Mean	Min	Max	Variance	Standard Deviation	Median
AER	75.03936721076224	0.01	100	1148.508597	33.889653	97.3186011169581
AL	13.163365916236433	0.00117061750073164	73.3886456172697	166.62685107579102	12.908402344046726	9.578063879077892
DGH	49.71815334916497	0.89142376	99.46162415	433.5059741416053	20.820806279815518	50.2744190450432
DRR	3.2710843373493974	1.0	4.75	0.549093	0.741008	3.25
GEE	14.717284205155227	0.0	47.27874	21.751390	4.663839	14.281864643096899
ICP	20.88649330187525	-30.243160228172897	23773.131774101603	92305.28304603275	303.8178451737698	4.97148571692381
ITE	29740519025.90272	9999.999776482578	1399675840946.35	1.1112176368795985e+22	105414308178.7097	765000000.0
LEM	61.58691052397916	16.285999999999998	84.1	114.4125953835555	10.696382350288134	64.0

Indicators	Mean	Min	Max	Variance	Standard Deviation	Median
LEF	66.1953109585959	22.394000000000002	87.7	139.4307103169767	11.808078180507474	69.131
ME	2.8018281962788474	0.0	117.34982320869	8.674400879098885	2.9452335865086976	2.09752650927828
MEG	7.721218361764516	0.0	57.478097622027505	37.38692725795918	6.1144850362037175	6.140559979309545
NEI	-69.14188666548525	-17632.7661127925	100.0	237199.88766825027	487.0317111526212	13.023777247450749
PPC	83.46087306226664	15.4	100.0	322.432093	17.956394	90.25
R&D	1.0288121222396516	0.0054399999999999	4.9527800000000015	0.8259959190357258	0.9088431762607484	0.725771559459431
REP	17.189401243052945	0.0	63.75	111.3701195242153	10.553204230195458	16.241958392514
TOUR	15.545022186516292	0.0009561945847061	93.4860024389904	300.64324466555587	17.33906700677853	8.45631572178083
UEMP	7.863912813842509	0.10999999940395402	37.970001220703104	32.15378520674475	5.670430777881408	6.322333925970966

Exploratory analysis

Plot 1: Average % Military expenditure from total expenditure by Year(1988-2019)
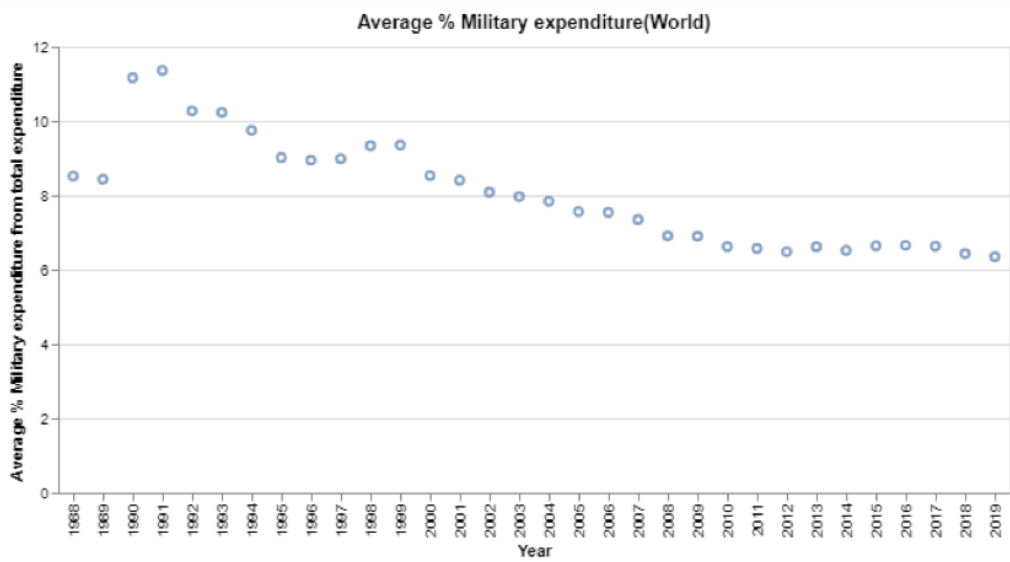
Plot 2: Military Expenditure VS. Military expenditure (% of general government expenditure)
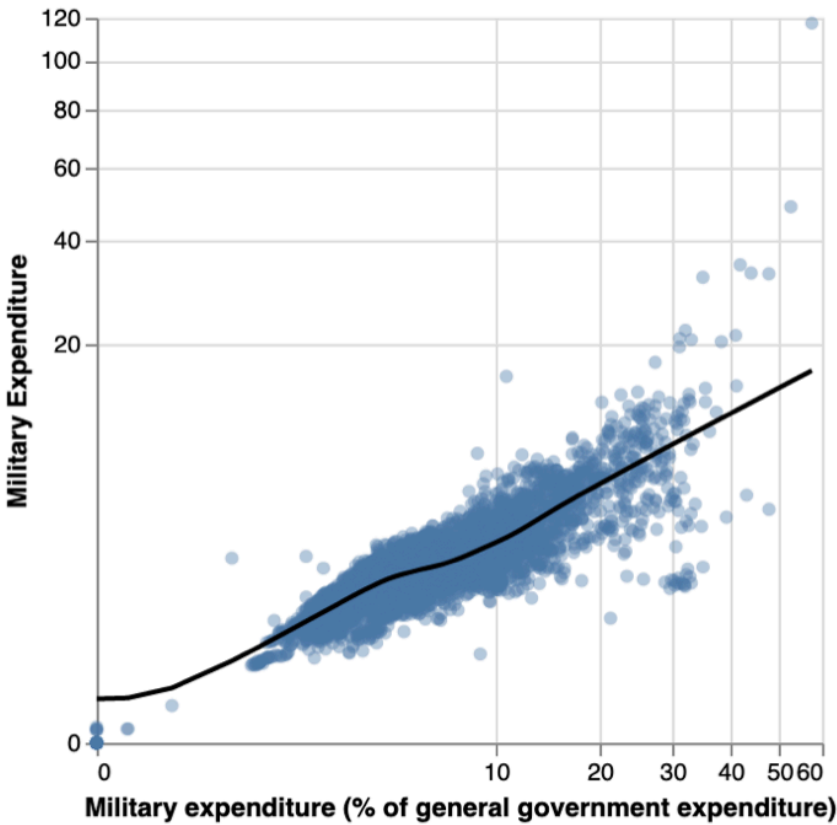
Plot 3: Average Military expenditure(% of general government expenditure) and Average Military Expenditure for Years from 1960 to 2020

Analysis:

- Plot 1 indicates that Average percent of Military expenditure from total expenditure is decreasing from 1988 to 2019.

- From the Military expenditure (% of general government expenditure) VS. Military expenditure plot (plot2), it's clear that Military expenditure (% of general government expenditure) and Military expenditure has a positive relationship. As one grows, the other also grows.

- It is really interesting that for the data available for both Average Military expenditure (% of general government expenditure) and Average Military expenditure (plot3), they have a very similar pattern from 1988 to 2020. As one goes down, the other also goes down.

Questions

1. How are citizens quality of life affected by their countries' military expenditure?
2. Which specific aspect of a citizen's life is most highly correlated with their country's military expenditure?

Proposed approaches

1. Using our tidied data, we will explore and analyze the quality of life indicators against military expenditure by plotting them and taking note of trends. Then, we will interept our indicators to see whether or not military spending does in fact heavily affect citizen's quality of life.
2. We will take a deeper, more mathematical approach than the first one, creating a heatmap of the correlation matrix, using PCA, and building regression models to see which specific indicators influence how much a country spends on their military.