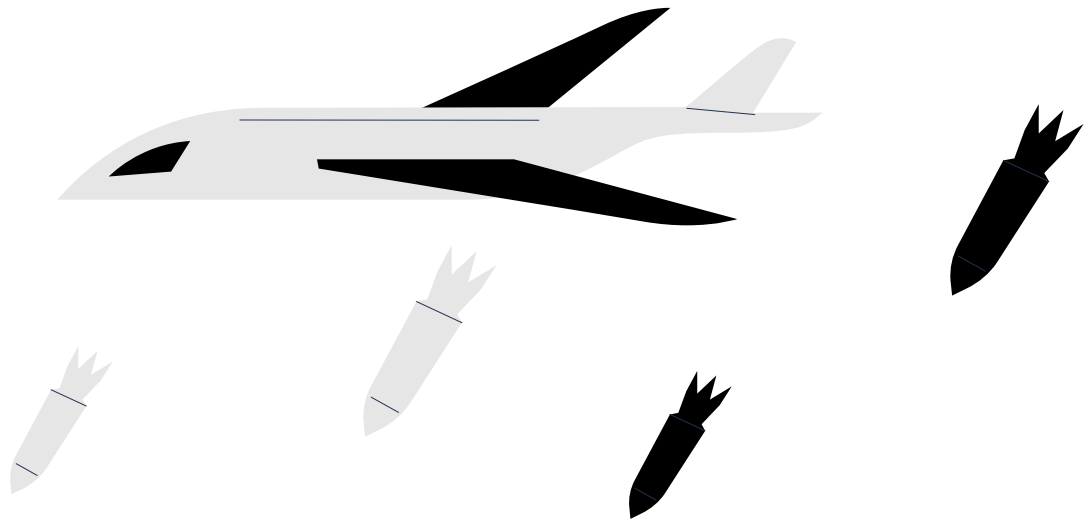
UNDERSTANDING INTERNATIONAL CONFLICT

Using Statistical Methods

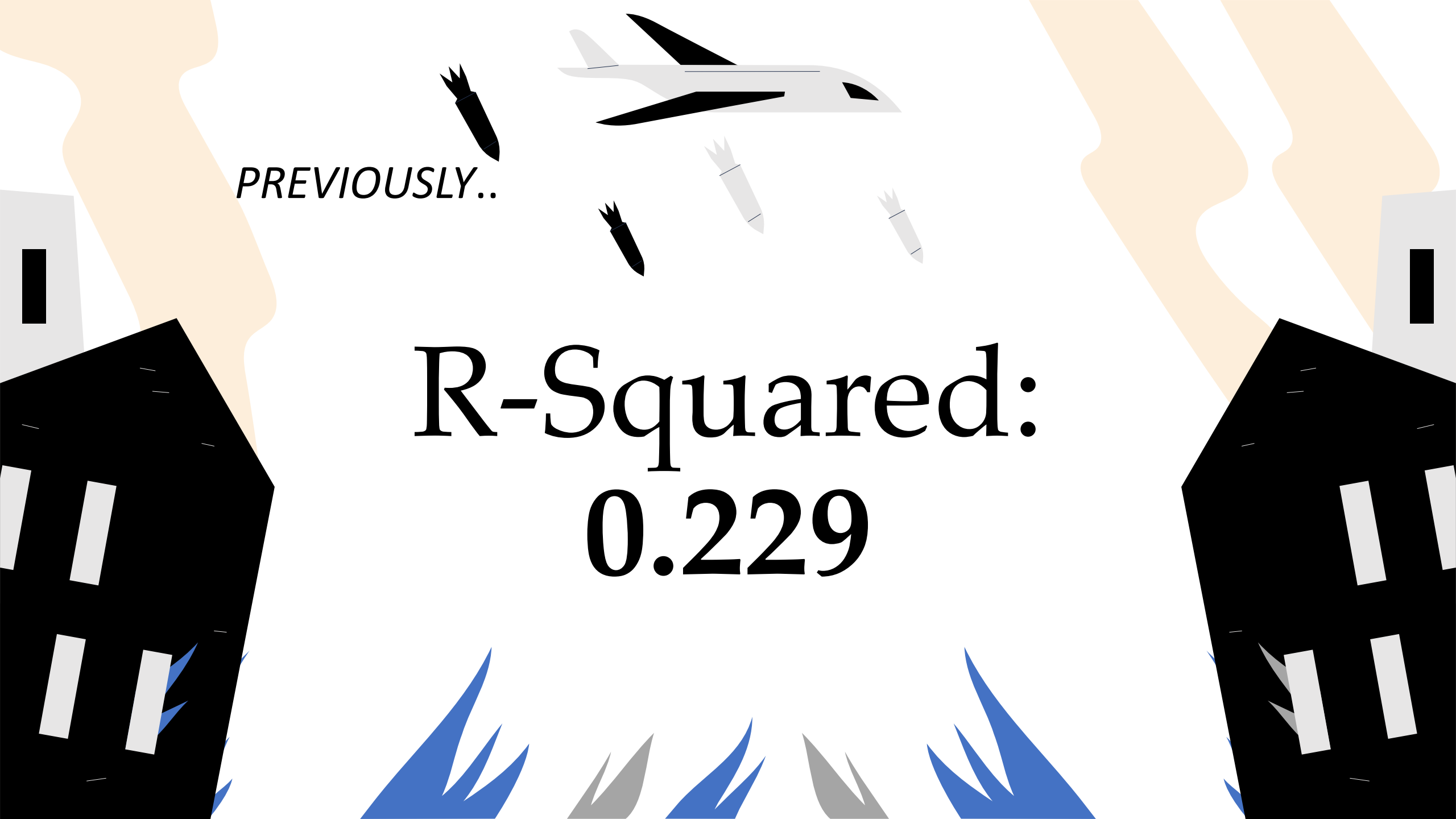


“War does not determine who’s right- only who’s left”

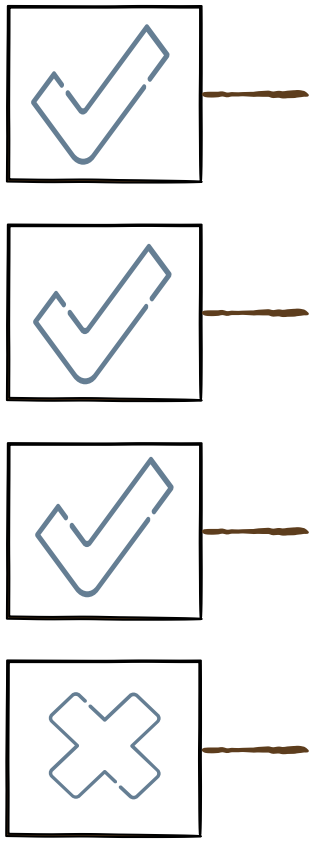
01

HYPOTHESIS

Can statistical methods be used to understand and predict war?


PREVIOUSLY..
R-Squared:
0.229

PREVIOUS METHODOLOGY



Data Set	Selecting Variables and Structuring Data
Linear Model with outliers	Standard linear regression fit: 0.229 R-squared
Kernel Regression Fit	Depicts skewed data due to outliers
Understanding the Variables	Variables in question were chosen at random and were not informative

CAN NEW TECHNIQUES PROVIDE NEW INFORMATION?

Issues in the past:

- The dataset has highly correlated variables
- War Data is not necessarily linear- namely due to its randomness- and it is difficult to select features without bias
- We require a way to understand the variables of our data further, with or without outliers, that can give us more information for better models

Sounds like a job for.. Principal Component Analysis!

02

EDA:

The data and its context to the domain

DATASET ONE: UCPD WAR

TIME

The data is depicted from 1948-2020

LOCATION

The data is international- sorted by each country

SIDE A

Group One that is on the offense during war

SIDE B

Group Two that is on the defense during war

INCOMPATIBILITY

1= Territory
2 = Government
3 = Both

INTENSITY

1 = Battle
2 = War

REGION

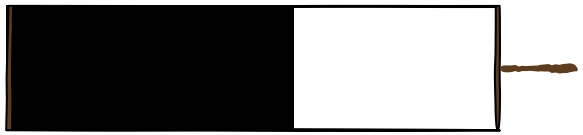
1 = Europe
2 = Middle East
3 = Asia
4 = Africa
5= America

CONFLICT TYPE

1= Extrasystemic
2= Interstate
3= Intrastate
4 = Internationalized Intrastate

DATASET ONE: UCPD WAR

location	side_a	side_b	side_b_id	side_b_2nd	i
Afghanistan	Government of Afghanistan	PDPA	291		2
Afghanistan	Government of Afghanistan	Jam'iyyat-i Islami-yi Afghanistan	292		2
Afghanistan	Government of Afghanistan	Harakat-i Inqilab-i Islami-yi Afghanistan , Hizb-i Islami-yi Afgh	293, 299, 294, 295, 292, 296		2
Afghanistan	Government of Afghanistan	Harakat-i Inqilab-i Islami-yi Afghanistan , Hizb-i Islami-yi Afgh	293, 299, 294, 297, 295, 292, 296		2
Afghanistan	Government of Afghanistan	Harakat-i Inqilab-i Islami-yi Afghanistan , Hizb-i Islami-yi Afgh	293, 299, 294, 297, 295, 292, 296		2
Afghanistan	Government of Afghanistan	Harakat-i Inqilab-i Islami-yi Afghanistan , Hizb-i Islami-yi Afgh	293, 299, 294, 297, 295, 292, 296		2
Afghanistan	Government of Afghanistan	Harakat-i Inqilab-i Islami-yi Afghanistan , Harakat-i Islami-yi A	293, 298, 299, 294, 297, 295, 292, 296		2
Afghanistan	Government of Afghanistan	Hizb-i Islami-yi Afghanistan, Hizb-i Wahdat, Junbish-i Milli-yi I	299, 300, 302		2
Afghanistan	Government of Afghanistan	Hizb-i Islami-yi Afghanistan, Hizb-i Wahdat, Junbish-i Milli-yi I	299, 300, 302		2
Afghanistan	Government of Afghanistan	Harakat-i Inqilab-i Islami-yi Afghanistan , Hizb-i Islami-yi Afgh	293, 299, 294, 297, 295, 292, 296		2
Afghanistan	Government of Afghanistan	Harakat-i Inqilab-i Islami-yi Afghanistan , Hizb-i Islami-yi Afgh	293, 299, 294, 297, 295, 292, 296		2
Afghanistan	Government of Afghanistan	Harakat-i Inqilab-i Islami-yi Afghanistan , Hizb-i Islami-yi Afgh	293, 299, 294, 297, 295, 292, 296		2
Afghanistan	Government of Afghanistan	Harakat-i Inqilab-i Islami-yi Afghanistan , Hizb-i Islami-yi Afgh	293, 299, 294, 297, 295, 292, 296		2
Afghanistan	Government of Afghanistan	Hizb-i Islami-yi Afghanistan, Hizb-i Islami-yi Afghanistan - Kha	299, 294, 300, 292, 296		2
Afghanistan	Government of Afghanistan	Hizb-i Islami-yi Afghanistan, Hizb-i Islami-yi Afghanistan - Kha	299, 294, 300, 292, 301		2
Afghanistan	Government of Afghanistan	Hizb-i Islami-yi Afghanistan, Hizb-i Islami-yi Afghanistan - Kha	299, 294, 300, 292		2
Afghanistan	Government of Afghanistan	Hizb-i Islami-yi Afghanistan, Hizb-i Wahdat, Jam'iyyat-i Islami	299, 300, 292		2
Afghanistan	Government of Afghanistan	Hizb-i Islami-yi Afghanistan, Hizb-i Wahdat, Junbish-i Milli-yi I	299, 300, 302, 303		2
Afghanistan	Government of Afghanistan	Jam'iyyat-i Islami-yi Afghanistan, Taleban, UIFSA	292, 303, 304		2
Afghanistan	Government of Afghanistan	UIFSA	304		2
Afghanistan	Government of Afghanistan	UIFSA	304		2
Afghanistan	Government of Afghanistan	UIFSA	304		2
Afghanistan	Government of Afghanistan	UIFSA	304		2
Afghanistan	Government of Afghanistan	UIFSA	304	Government of Iran	2
Afghanistan	Government of Afghanistan	Hizb-i Islami-yi Afghanistan	299		2
Afghanistan	Government of Afghanistan	Hizb-i Islami-yi Afghanistan, Taleban	299, 303		2
Afghanistan	Government of Afghanistan	Taleban	303		2
Afghanistan	Government of Afghanistan	Taleban	303		2
Afghanistan	Government of Afghanistan	Hizb-i Islami-yi Afghanistan, Taleban	299, 303		2
Afghanistan	Government of Afghanistan	Taleban	303		2
Afghanistan	Government of Afghanistan	Taleban	303		2
Afghanistan	Government of Afghanistan	Taleban	303		2
Afghanistan	Government of Afghanistan	Taleban	303		2
Afghanistan	Government of Afghanistan	Taleban	303		2



55%

TERRITORY

More than half the battles in the dataset have been over territory

75%

BATTLES

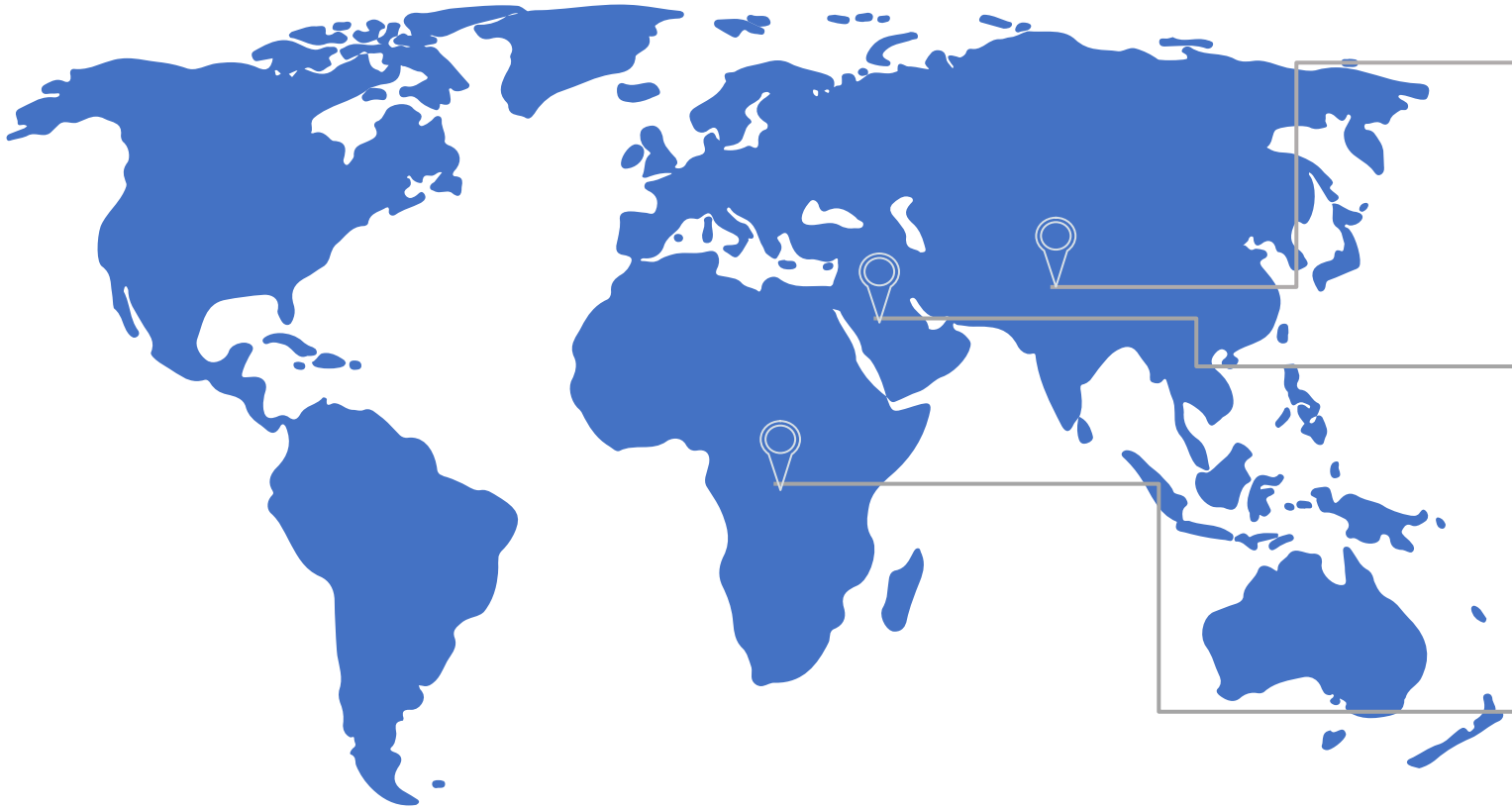
A vast majority of conflict is battle, <999 deaths by UCPD standard

74%

INTRASTATE

Conflict was a government vs. one or more rebel groups

WAR BY REGION



40% **ASIA**

14% **MIDDLE EAST**

32% **AFRICA**

DATASET TWO: WAR

Need: A numerical dataset

→

UCPD War data is a count of every instance of war with obsolete variables

→

Combine this count and attach it to numerical variables

DATASET TWO: WAR

	Country	Population	Land_Area	Exports_USD	Labor_Force	Military_Spend	GDP	Wars
1	Afghanistan	37171922	652860	8.500900e+09	10022144	1.980747e+08	1.805323e+10	94
2	Albania	2866376	27400	4.785178e+09	1399364	1.758867e+08	1.515643e+10	1
3	Algeria	42228415	2381741	4.523397e+10	12330675	9.583724e+09	1.749109e+11	52
4	Angola	30809787	1246700	4.138890e+10	12659938	1.983614e+09	1.013532e+11	81
5	Argentina	44494502	2736690	7.576658e+10	19952439	3.842852e+09	5.248197e+11	7
6	Armenia	2963243	28470	1.440800e+08	1153003	6.200000e+07	1.386000e+07	8
7	Azerbaijan	9939771	82654	2.548447e+10	5091128	1.672176e+09	4.711294e+10	20
8	Bangladesh	161376713	130170	4.056039e+10	68073235	3.649624e+09	2.740390e+11	21
9	Bhutan	771608	38117	8.082228e+08	360953	0.000000e+00	2.539553e+09	1
10	Bolivia	11353140	1083300	1.047027e+10	5618578	6.188421e+08	4.028765e+10	4
11	Bosnia and Herzegovina	3323929	51200	8.579141e+09	1187357	1.720733e+08	2.017741e+10	14
12	Brunei	428960	5270	7.045834e+09	215909	3.577088e+08	1.356735e+10	1
13	Burundi	11175379	25680	2.508663e+08	4808244	6.543660e+07	2.668496e+09	24

DATASET TWO: WAR

Country

Nation

Population

Average Value from 1948-2021

Land Size

Square Foot average of every country

Exports

Average Exports in USD

Labor Force

Average number of people in the working force

Military Spend

Average military spend in USD

GDP

Average population spend in USD

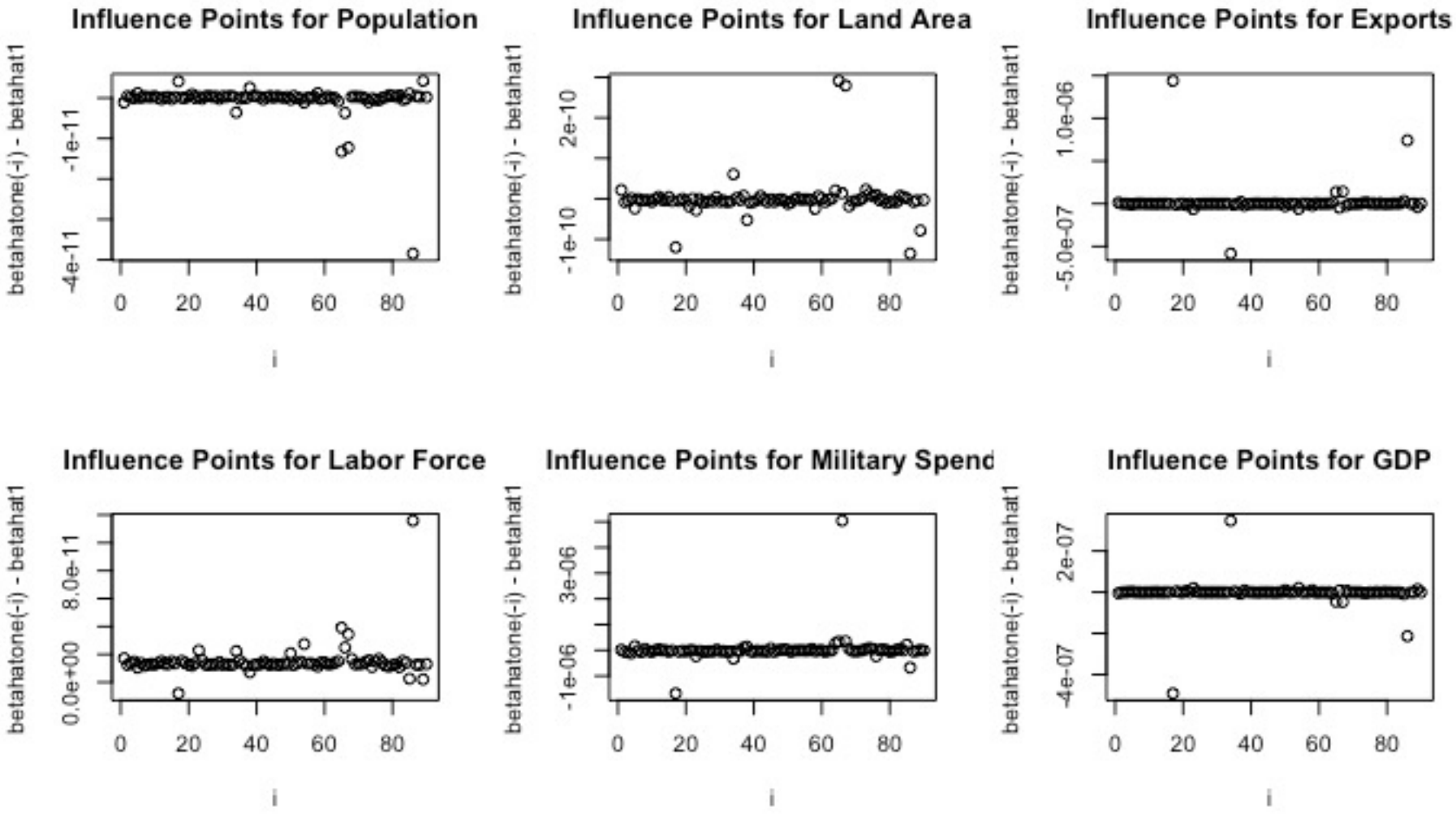
Wars

Total concatenated value (count of total wars in time frame)

INFLUENCE PLOTS

OUTLIERS

Every variable has its own outliers, which poses as a big issue since they are all highly influential. Note as well that the monetary values are very large.



OUTLIERS BY VARIABLE

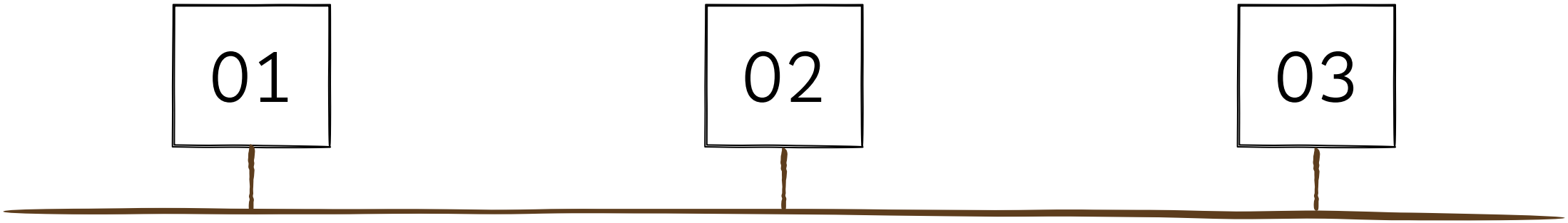
Population	10	Bangladesh, China, India, Indonesia, Japan, Mexico, Nigeria, Pakistan, Russia, U.S.
Land Size	11	Bangladesh, China, India, Indonesia, Japan, Mexico, Nigeria, Pakistan, Russia, U.S., Vietnam
Exports	8	Algeria, Argentina, China, Colombia, DR Congo, India, Russia, Saudia Arabia, U.S.
Labor Force	16	China, Denmark, France, India, Indonesia, Ireland, Japan, Mexico, Russia, Saudia Arabia, South Korea, Thailand, Turkey, UK, U.S., Vietnam
Military Spend	11	China, France, India, Israel, Japan, Romania, Rwanda, South Africa, Turkey, UK, U.S.

03

PCA

Principal Component Analysis

PCA OBJECTIVES



Data Reduction

"Though p components are required to reproduce the total system variability, much of this variability can be accounted for by a small number k of the principal components. : these k principal components can then replace the initial p variable."

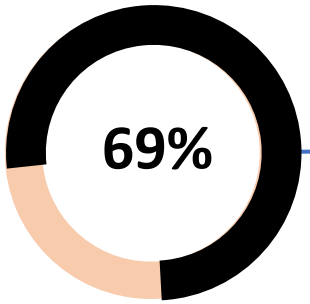
Interpretation

"PCA reveals relationships that were not previously suspected and allows interpretations." Additionally, these new relationships can bring research questions that surface from the extended knowledge on the domain.

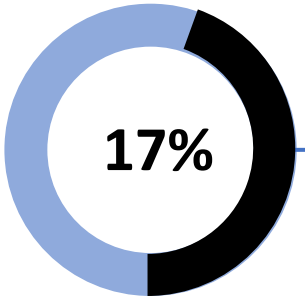
Intermediary

"Serve as intermediate steps in larger investigations" PCA is often used as a task in machine learning before equipping the model with its data.

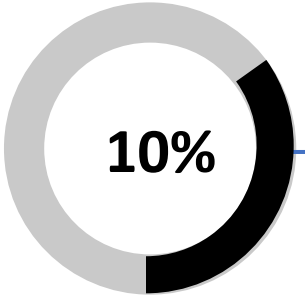
SUMMARY OF COMPONENTS



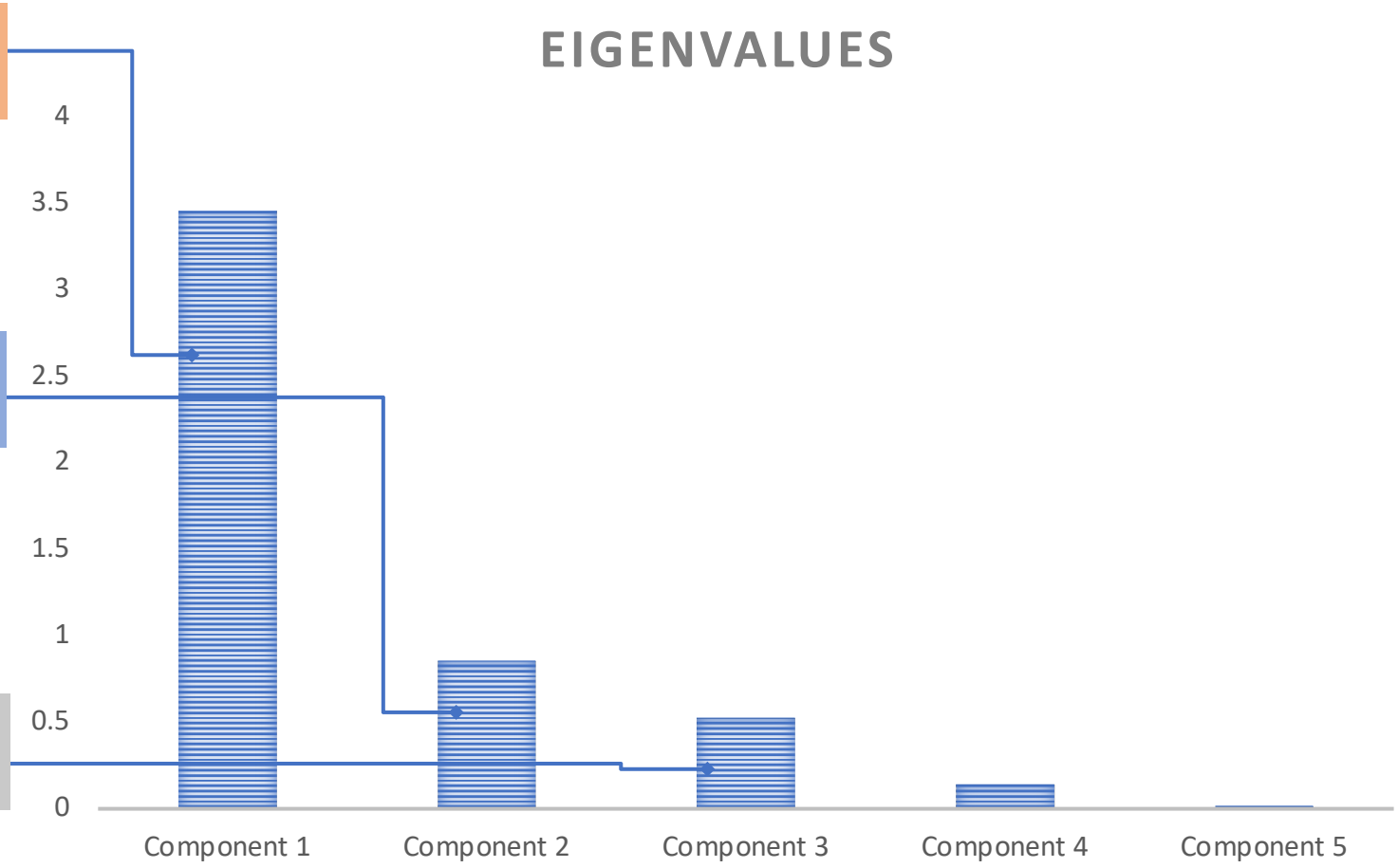
Variance explained by Component 1



Variance explained by Component 2



Variance explained by Component 3



EIGENVECTORS: CORRELATION MATRIX

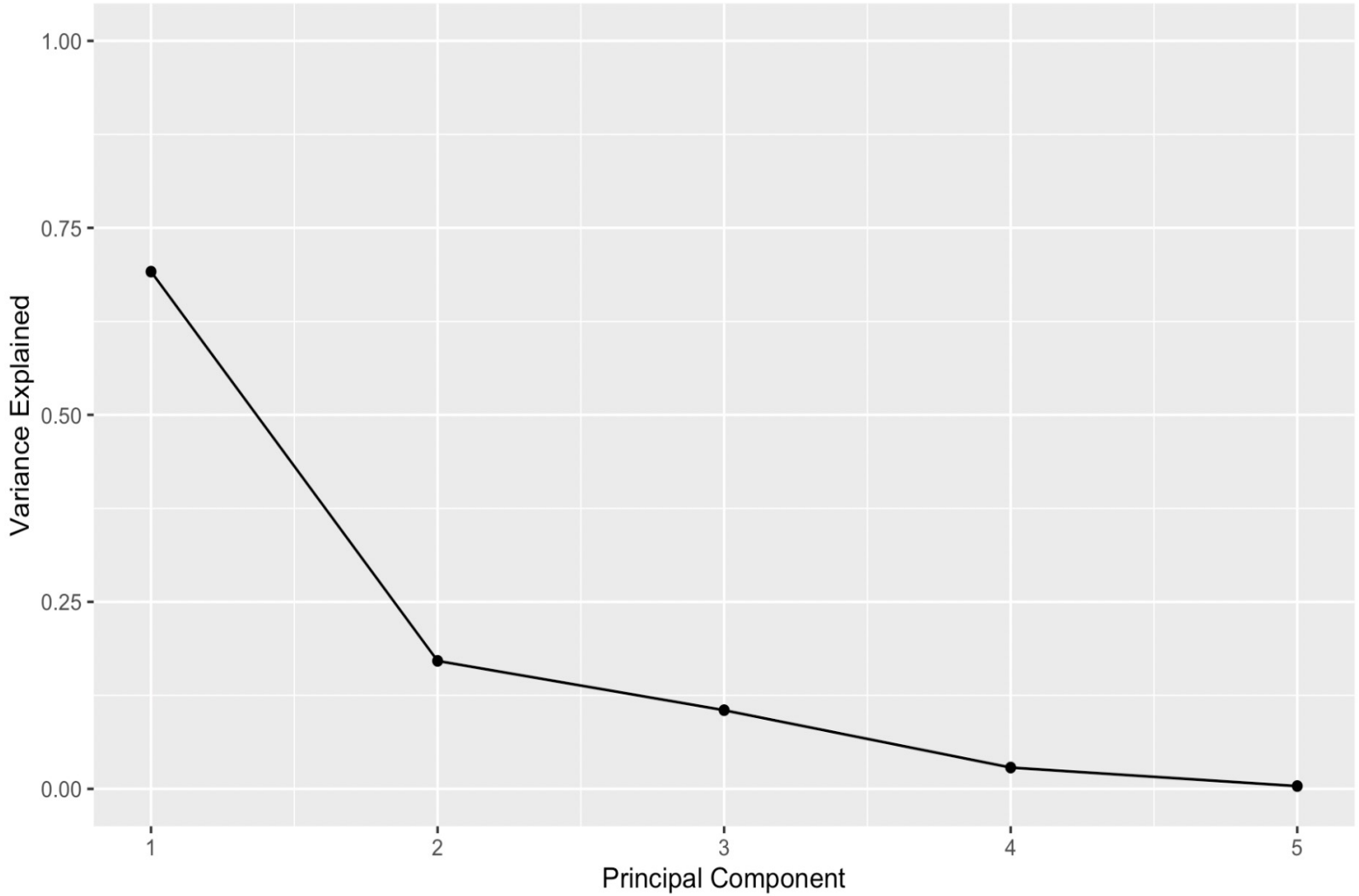
	People Component Variables that the "people" are in control of	**Economic Component** Variables that explain working class and spend	**Land Component** Land size has an exceptionally high eigenvalue	**Monetary Component** Exports and Military spend are monetary values	**Labor Component** Population and labor force explain variability
Population	0.459761	**0.5335745**	-0.09585589	-0.30563192	**0.63349703**
Land Size	0.3954616	-0.2263718	**0.88826757**	-0.05686546	0.01062996
Exports	0.4876755	-0.2712185	-0.23964172	**0.7667188**	0.20815183
Labor Force	0.4815583	**0.4639383**	-0.08531171	0.03149660	**-0.73796528**
Military Spend	0.4031176	**-0.6125813**	-0.37025145	**-0.56080799**	-0.10319193

COMPONENTS TO KEEP

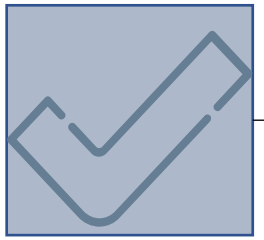
The First and Second Component have high correlation values

The Third Component explains 10% of variability but is almost exclusively Land Size

Best Bet: Including Three Components for modeling

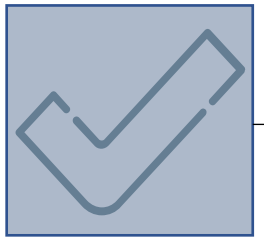


USE CASE: DATA REDUCTION



Using top components to reduce data from 5 to 3 dimensions. Though this data set is small, it could be useful for an expanded version

USE CASE: INTERPRETABILITY



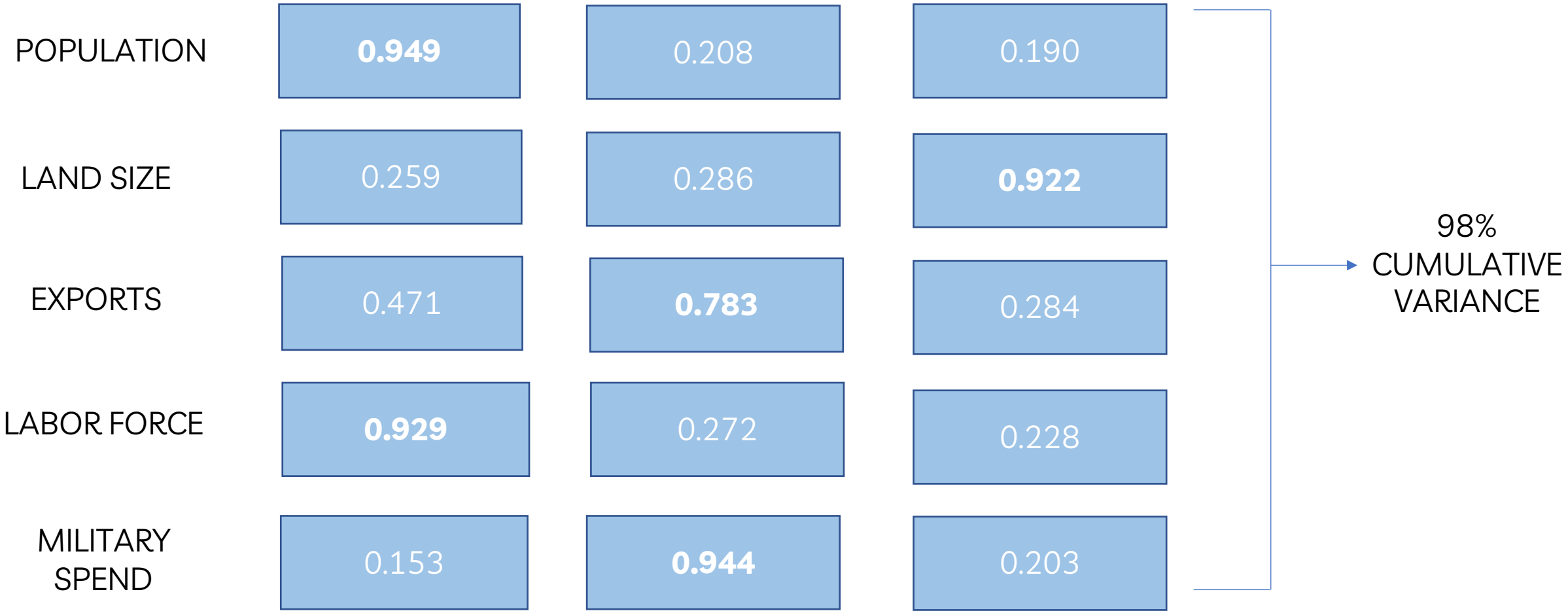
- Population and Labor Force having the highest variability: points to circulation of opinion, mass dissatisfaction of government and time on hands
- Land size having high influence in the third component- give room to civilians!
- People, Economy, and Land can be major components combined to explain war

04

Factor Analysis

Investigating the components further

PCA METHOD: VARIMAX ROTATION LOADINGS



FACTOR ANALYSIS OBJECTIVE

In general …

- "Describe the relationships among many variables (covariance) in terms of a few underlying, but unobservable, random quantities called factors
- "Model the interrelationships between items with the factors."

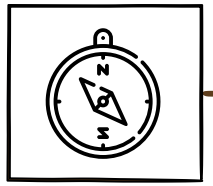
In this case …

- Understand the variables further
- Re-establish the relationship mapped out from PCA
- Uncover new interrelationships not foreseen prior

05

Further Discussion

KEY TAKEAWAYS

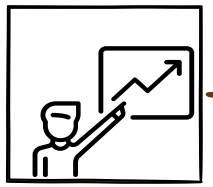


PRINCIPAL COMPONENT ANALYSIS

Was able to provide discernible information on the relationship between random variables that were mapped without cause.

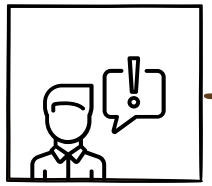
FACTOR ANALYSIS

Mapped out the intuitive relationship of the coupled variables (Population & Labor, Exports & Military Spend, etc.) that could be difficult to detect in a bigger data set



CURRENT MODELING

War is not linear! Though a tested R-squared of 0.389 was accomplished it was after removing outliers, leading to bias. Non-parametric methods could lead to a better result.



FUTURE MODELING

Understanding war is doable with good techniques and data. But even with advanced machine learning models, **predicting** war is much harder.

IRAN

51
INCIDENCES

Iran was responsible for 1.7% of wars in the dataset.

100%
INTRASTATE

Conflict where the government was on the offense and rebel groups were on the defense

78%
KDPI & MEK

KDPI - “Democratic Party of Iran and Kurdistan “ and MEK- ”People’s Mojahedin Organization of Iran” were in battle with the Iranian government for almost 60 years

IRAN

100%

INTRASTATE

Conflict where the government was on the offense and rebel groups were on the defense

78%

KDPI & MEK

KDPI - “Democratic Party of Iran and Kurdistan “ and MEK- ”People’s Mojahedin Organization of Iran” were in battle with the Iranian government for almost 60 years

2

INTERNATIONAL

Iran has had international involvement in territories not of its own in Iraq and Israel

78%

KDPI & MEK

KDPI - “Democratic Party of Iran and Kurdistan “ and MEK- ”People’s Mojahedin Organization of Iran” were in battle with the Iranian government for almost 60 years

2

INTERNATIONAL

Iran has had international involvement in territories not of its own in Iraq and Israel

2

INTERNATIONAL

Iran has had international involvement in territories not of its own in Iraq and Israel

9
INCIDENCES

Only 0.03% of the dataset involves South Korea.

100%
TERRITORY

All incidences involve one particular government (!) over territory

1
OPPONEN

North Korea is the o
major opponent on S
A (offensive) for all
incidences

100%
TERRITORY

All incidences involve one particular government (!) over territory

1
OPPONENT

North Korea is the only major opponent on Side A (offensive) for all 9 incidences

16
SUPPORTE

16 governments listed side_b_2nd, also know as allies in the confli

1

OPPONENT

North Korea is the only major opponent on Side A (offensive) for all 9 incidences

16

SUPPORTERS

16 governments listed in side_b_2nd, also known as allies in the conflict

16

SUPPORTERS

16 governments listed in side_b_2nd, also known as allies in the conflict

INDIA

216
INCIDENCES

Not only an outlier for every X variable, but also an outlier for Y- India has the 2nd highest number of wars in the data set after Myanmar

71%
INTRASTATE

The majority of conflict in India (other than civilian conflict) is Intrastate, and the remainder is Interstate (Pakistan)

40
YEA

The most preva
over territor
Kashmir war, w
been ongoing s
on the eclipse
year war involvi
partie

INDIA

71%

INTRASTATE

The majority of conflict in India (other than civilian conflict) is Intrastate, and the remainder is Interstate (Pakistan)

40

YEARS

The most prevalent war over territory is the Kashmir war, which has been ongoing since 1984 on the eclipse of a 40 year war involving several parties

1

NIS

At least one N
Indian ethn

INDIA

40

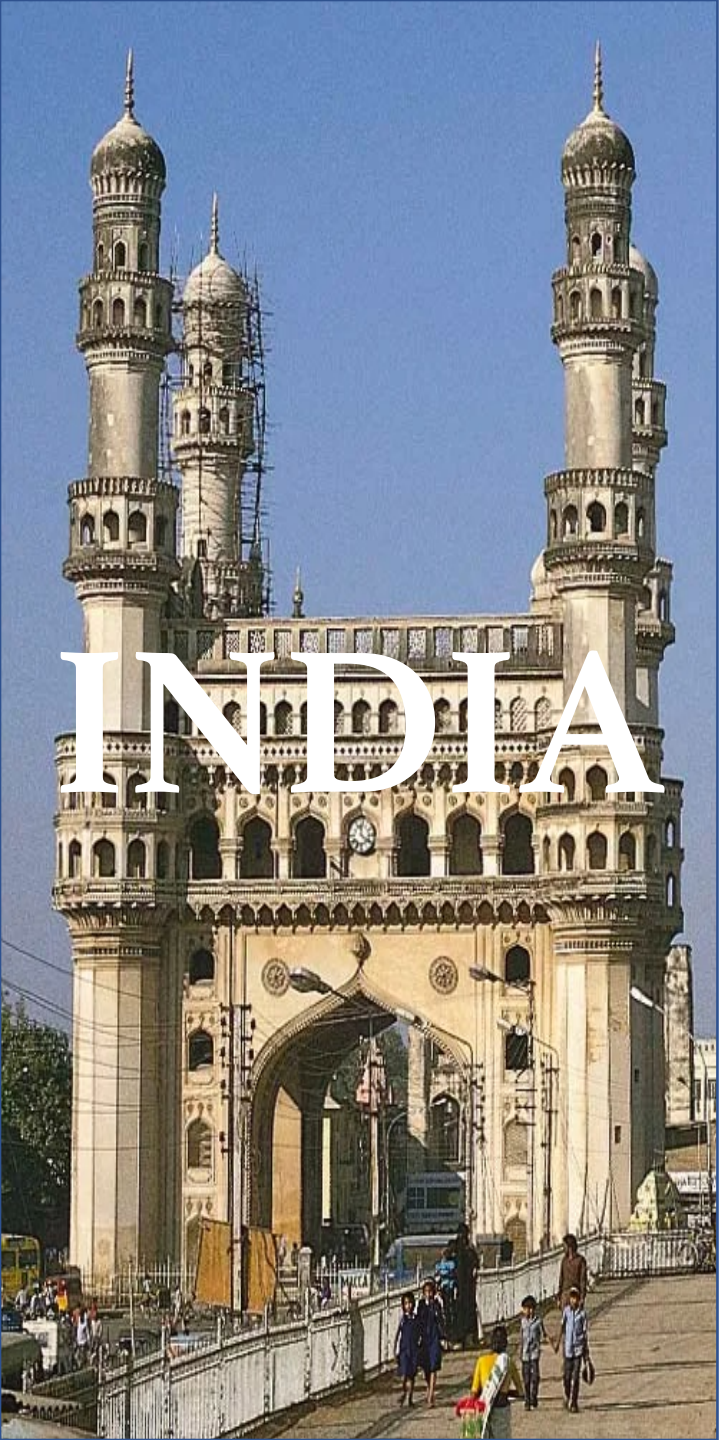
YEARS

The most prevalent war over territory is the Kashmir war, which has been ongoing since 1984 on the eclipse of a 40 year war involving several parties

1

NISH

At least one Nishath is of Indian ethnicity (me)



1

NISH

At least one Nishath is of Indian ethnicity (me)

RUSSIA

58
INCIDENCES

Though this count must have risen after 2021, the dataset features 58 wars on Russian territory

35%
INTENSITY

35% of the incidences were unfortunately at an intensity level of 2, meaning over 1,000 deaths per incident

100%
INCOMPATIBILITY

All involved parties and conflict was regarding dissatisfaction over government or an issue revolving the government.

RUSSIA

35%
INTENSITY

35% of the incidences were unfortunately at an intensity level of 2, meaning over 1,000 deaths per incident

100%
INCOMPATIBILITY

All involved parties and conflict was regarding dissatisfaction over government or an issue revolving the government.

4
TERRITORIES

Major regions involved in conflict with Russia: Caucus Emirate(group), Ukraine, Chechnya, Novorossiya

RUSSIA

100%

INCOMPATIBILITY

All involved parties and conflict was regarding dissatisfaction over government or an issue revolving the government.

4

TERRITORIES

Major regions involved in conflict with Russia: Caucus Emirate(group), Ukraine, Chechnya, Novorossiya

4

TERRITORIES

Major regions involved in conflict with Russia: Caucus Emirate(group), Ukraine, Chechnya, Novorossiya