Modern Global Scenario

Mini-Project
Multidimensional Data Analysis
DMKM

Carlos López Roa
me@mr3m.me

February 4, 2016

Abstract

As a foreign student in the DMKM master, the author has posed several times the question: *What makes the wealth of a country?*, Of course there exists several types of wealth, namely: Demographic, Economic, Natural, Cultural, Energetic, or even Military; Still a clear answer to the above question is to be searched and, if possible, found in this study. For this purpose a dataset of several indicators of the countries integrating the United Nations was retrieved, preprocessed and studied using the factorial methods of **PCA**[1], **CA**[2] and **FDA**[3]. At the end some conclusions are drawn about the nature of dominance of the countries.

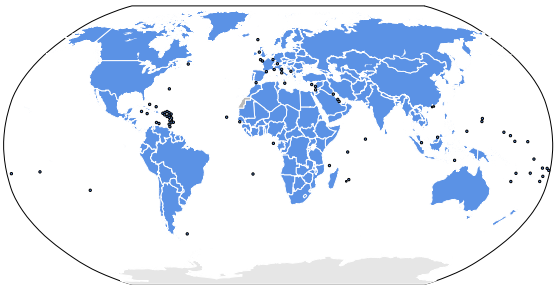


Figure 1-1: Map showing the member states of the United Nations.

Contents

1 Data preparation

Information about 90 selected indicators[4] of the 193 countries (figure **??**) integrating the United Nations was retrieved from the `Wolfram|Alpha Knowledgebase` [**?**] using a licensed copy of the `Wolfram Mathematica` software version `10.0`. A full overview of the variables used is available in the appendix **??**. A dictionary of the countries name and code can be found in appendix **??**

1.1 Dataset Preprocessing

As noted by [**?**] PCA might have troubles with variables which include big outliers, because in the normalization process, the mean of the sample resides far from both the majority of observations and the outliers. To bolster this issue, several strategies have been proposed, in particular, to make new synthetic variables taking the log of the original variables reducing the skewness. This was made for all non-negative variables spanning several orders of magnitude which were heavily left skewed, as shown in figure **??**. This can also help to densify the center of mass of the observations. The previous claim was proved experimentally as shown in the figure **??**.

Code for the querying and preprocessing transformations can be found in the appendix **??**.

2 Data processing

The resultant file `s1.xls` was imported to a licensed copy of `COHERIS SPAD` version `8.2.18`. The general schema of the process diagram is showed on figure **??**.

First some descriptive statistics, and several[5] normed PCA analysis on the different groups of variables was made as exploratory analysis. Then a normed PCA of the selected continuous variables was carried out to reduce the dimensionality of the dataset. Also a CA was made for two categorical variables to explore the relation between the continent and the economic sector, and finally a FDA was made to predict the economic sector based on the continuous variables.

2.1 Exploratory PCA

Since the data is composed of 84 continuous variables, a more extensive approach was taken with the PCA

[1] Principal Component Analysis
[2] Correspondance Analysis
[3] Factorial Discriminant Analysis
[4] Demographic, Economic, Energetic, Communication, Geography

[5] A thing to note here is that no significance anlysis, or quality of representation analysis of the factors was made because this was only exploratory analysis

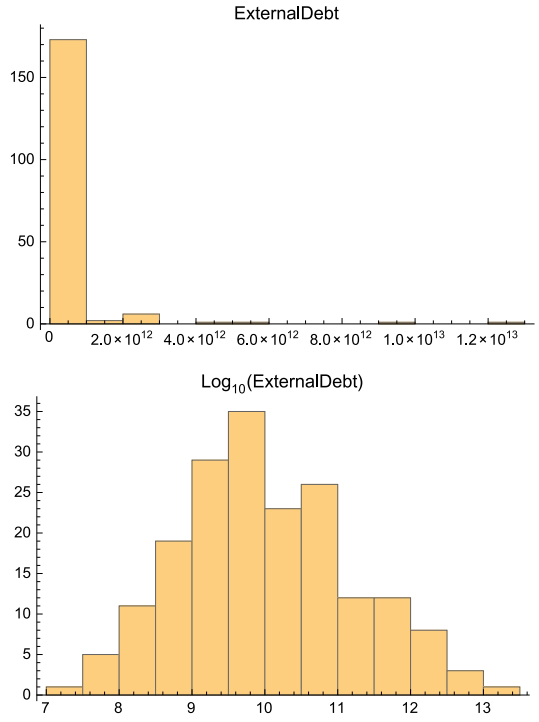


Figure 1-2: Histogram of the heavily skewed variables that span several orders of magnitude where treated taking the $\log_{10}$ to reduce the skewness as a measure of robustness. As seen in the figure, the new synthetic variable has a distribution closer to the Normal. The skewness goes from 6.73 to 0.31 respectively

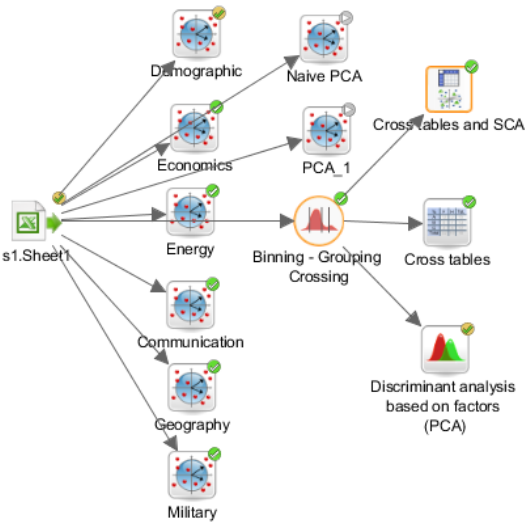


Figure 2-3: General process diagram of the analysis. In the first stage six different PCA were carried out as exploratory analysis. Then a PCA analysis was done in a subset of the continuous variables, a CA was made after binding modalities on two variables and a FDA on factors to explain a categorical variable.

method. Namely, exploring each group of variables separately as exploratory analysis to give some insight in the nature of the variables of each group.

The categorical variable `Continent` has frequencies as showed in figure **??** .

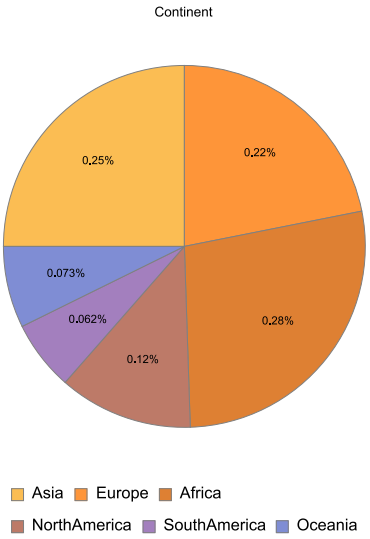


Figure 2-4: The total of 193 countries being studied are distributed amongst their respective continents as showed in the chart.

2.1.1 Demographics

A normed PCA of the `Demographic` group of variables was carried out leading to the results shown in figure **??**. Around 73% of the total inertia can be explained in the first two factors. There exists a high correlation between variables such as: `Poverty Fraction`, `Birth Rate Fraction`, `Total Fertitility Rate`, `Population Growth` and `Infant Mortality Fraction`. Group which is in contraposition with the highly correlated variables: `Literacy Fraction`, `Life Expectancy` and `Median Age`. A group of correlated variables stays in perpendicular relation with these contraposition and is composed of the variables `Annual Births`, `Child Population`, `Population`, `Annual Deaths`, `Adult Population` and `Elderly Population`. The variables `Death Rate Fraction` and `Migration Rate Fraction` are misrepresented since they contain several missing values.

When we plot the observations in this factors we can see that countries such as `NG: Federal Republic of Nigeria`, NE: Republic of Niger, TD: Republic of Chad lead the first quadrant[6], that is the one associated with high `Poverty Fraction`, high `Birth Rate Fraction`, etc. In the second quadrant countries such as NR: Republic of Nauru, TV: Tuvalu and PW: Republic of Palau all three in Oceania lead, that is they can be characterized by low `Annual Births` and `Population`. In the third quadrant we find countries

[6] in the geometric sense

such as AT: Republic of Austria, DE: Federal Republic of Germany and GB: United Kingdom of Great Britain and Northern Ireland that are characterized by high `Literacy Fraction`, high `Life Expectancy`, low `Poverty Fraction` and `Population Growth`. In the fourth quadrant we find countries such as IN: Republic of India, CN: People's Republic of China and ID: Republic of Indonesia which are characterized by high `Population`, high `Elderly Population` , `Annual Births` and `Annual Deaths`.

The `continent` has been plotted as a supplementary variable, showing that Africa is located in the first quadrant, Oceania and North America in the second, Europe and South America in the third and Asia in the fourth. A thing to note is the position of the member of the G8[7], as a representative of the world dominance, we observe that they are grouped around the second factor negative direction, that is, demographically speaking, countries with high `Population`, high `Life Expectancy` and `Literacy Fraction` but still low `Poverty Fraction`, `Total Fertitility Rate` and `Population Growth`.

2.1.2 Economics

A normed PCA of the `Economic` group of variables was carried out leading to the results shown in the figure **??**. Around 58% of the total inertia can be explained in the first two factors. There exists a high correlation between the variables: `GDP`, `GDP At Parity`, `Government Receipts`, `Government Expenditures`, `Government Debt`, `Foreign Exchange Reserves`, `External Debt`, this correlation is also very high with the first factor in the positive direction. Also the variables `GDP Real Growth`, `Industrial Production Growth` and `Exchange Rate` are highly correlated between each other and also to the second factor in the positive direction. The variable`Labor Force` is in the first quadrant, as a combination of the first and second factor in the positive direction, also the variable `GDP Per Capita` is in the fourth quadrant as a combination but in the negative direction of the second factor, and the variable `Unemployment Fraction` is in contrapositive with `Unemployment Fraction`. The variables `Inflation Rate` and `Price Index` are misrepresented because they contain several missing values.

When we plot the observations in the factor plane we can see that, in the first factor, the countries with highest values are: US: United States of America, JP: Japan, DE: Federal Republic of Germany, FR: French Republic, that is, countries with high `GDP`, `External Debt` and both `Government Receipts`, `Government Expenditures`, `Government Debt`, `Government Expenditures`. Contrary to this, countries with the most negative value in the first factor NR: Republic of Nauru , KI: Republic of Kiribati, and TO: Kingdom of Tonga in Oceania. In the second factor, the countries with the most positive value are: ET: Federal Democratic Republic of Ethiopia, AO: Republic of Angola, TZ: United Republic of Tanzania in Africa. And in the opposite direction we have LU: Grand Duchy of Luxembourg, EE: Republic of Estonia, IS: Iceland.

The `Continent` and the `Sector Labor Fractions` has been plotted as supplementary variables, showing that Europe has the highest value in the first factor, still keeping negative value in the second, whereas Asia and South America have both high value in both factors, Africa is in the second quarter and North America and Oceania have both negative values in the first and second factor. The `Sector Labor Fractions` Services is in the fourth quadrant, close to Industry, Agriculture is in the second quarter, and Industry And Services is in the third quarter. A thing to note is the locus of the countries members of G8, seven out of 8 remain really close, apart from anyone else, in the fourth quadrant, that is with high `External Debt`, `GDP`, and the highest `GDP Per Capite` but low `Exchange Rate`, `GDP Real Growth`, `Industrial Production Growth` and low `Unemployment Fraction`.

2.1.3 Energy

A normed PCA of the `Energy` group of variables was carried out leading to the results shown in the figure **??**.

Around 56% of the total inertia can be explained in the first two factors. There exists a high correlation between the variables: `Oil Imports`, `Oil Consumption`, `Electricity Production`, `Electricity Consumption`, and `Electricity Imports` in the first quadrant. Also the variables `Natural Gas Reserves`, `Oir Reserves`, `Natural Gas Production` and `Oil Production` are highly correlated between each other in the fourth quadrant. The variables `Oil Exports` and `Natural Gas Consumption` are also situated in the fourth quadrant.

When we plot the observations in this factorial plane we can see that US: United States of America is far from any other observation with the highest value in the first factor, that is a combination of `Oil Imports`, `Oil Consumption`, `Electricity Production`, `Electricity Consumption`, `Electricity Imports`, `Natural Gas Reserves`, `Oil Reserves`, `Natural Gas Production` and `Oil Production` . Countries in the first quadrant, that is associated with `Oil Consumption`, `Electricity Imports`, `Electricity`

[7] Group of the eight most industrialized countries: FR, DE, JP, GB, US, CA, IT and RU

Consumption and Electricity Production: ES: Kingdom of Spain, FR: French Republic, IT: Italian Republic, JP: Japan. In the other hand, countries situated in the fourth quadrant, that is associated with Natural Gas Reserves, Oir Reserves, Natural Gas Production and Oil Production: RU: Russian Federation, IR: Islamic Republic of Iran and CA: Canada. Countries close to the First factor, that is in a positive combination of the previous two groups are as previously mentioned, US, GB and IN. We must not forget that with synthetic logarithmic variables, now every distance in the play represents orders of magnitude, thus, the outlier US has significantly more energy consumption, production and trading that any other country. On the contrary, the country closer to the first factor axis but with negative coordinates is SL: Republic of Sierra Leone that is, with the less energy consumption, production and trading.

If we pay attention to the members of the G8 we see US, CA, RU in the fourth quadrant, that is of Oil producers, and the remaining GB, DE, JP, IT, FR of Oil importers and Electricity Producers. But still both are in the positive to far positive side of the First factor.

2.1.4 Communication

A normed PCA of the Communication group of variables was carried out leading to the results shown in the figure **??**.

Around 65% of the total inertia can be explained in the first two factors. There exists a high correlation between the variables: Airports, Television Stations, Internet Users, AM/FM Radio Stations, and Road Length with each other and with the first factor. Also the variables Merchant Ships, Merchant Ships Dead Weight and Merchant Ships Gross are highly correlated between each other in the fourth quadrant.

When we plot the observations in this factorial plane we can see that US: United States of America is far from any other observation with the highest value in the first factor, followed by BR: Federative Republic of Brazil, RU, CN, MX, that is with high volume of land and radio communications. Countries with high values in the sector factor: PA: Republic of Panama, LR: Republic of Liberia that is with high values in Merchant Ships.

If we pay attention to the members of the G8 we see US in the far positive factor, that is with high values in communication infrastructure, the remaining members remain close to each other with high values in the first factor.

2.1.5 Geography

A normed PCA of the Geography group of variables was carried out leading to the results shown in the figure **??**.

Around 57% of the total inertia can be explained in the first two factors. There exists a high correlation between the variables: Area, Water Area, Boundary Length, Coastline Length, and Arable Land Area with each other and with the first factor. Also the variables Irrigated Land Fraction and Arable Land Fraction are highly correlated between each other in the negative direction of the second factor. There exist opposition by the variable Lowest Elevation in this second factor. And the variables Crops Land Area and Irrigated Land Area are correlated and in the fourth quadrant.

When we plot the observations in this factorial plane we can see that US, CA, RU, CN, IN have the highest values in the first factor, that is related with the size of the country. In the negative direction of the second factor we find MD: Republic of Moldova with high values in the Arable Land Fraction and Irrigated Land Fraction. On the other hand, countries with low values in the first factor are MC Principality of Monaco, VC Saint Vincent and the Grenadines, BB: Barbados that is, small countries.

If we pay attention to the members of the G8 we see US, CA, RU in the far positive factor, that is with high values in size the remaining members remain close to each other with not that high values in the first factor.

2.1.6 Military

A normed PCA of the Military group of variables was carried out leading to the results shown in the figure **??**.

Around 79% of the total inertia can be explained in the first two factors. There exists a high correlation between the variables: Military Fit Population, Military Age Rate and Military Age Males with each other and with the first factor. Also the variable Military Expenditure Fraction is highly correlated with the second factor. And the variable Military Expenditure is in the first quadrant.

When we plot the observations in this factorial plane we can see that CN, IN, US have the highest values in the first factor, that is related with the Military Fit Population of the country. In the negative direction of the second factor we find ST: Democratic Republic of Sao Tome and Principe. On the other hand, countries with high values in the second factor are OM: Sultanate of Oman and QA State of Qatar, that is, with high values in Military Expenditure Fraction.

If we pay attention to the members of the G8 we see US, CA, RU in the far positive factor, that is

with high values in `Military Fit Population` the remaining members remain close to each other with not that high values in the first factor.

2.2 PCA

Having explored these groups a variables, a normed PCA was made selecting `Life Expectancy`, `Population`, `Population Growth`, `Total Fertility Rate` in the DEMOGRAPHICS; `External Debt`, `Foreign Exchange Reserves`, `GDP`, `GDP Per Capita`, as active variables.

2.3 CA

2.4 FDA

3 Data postprocessing

4 Conclusions and interpretations

A Appendices

A.1 Dataset Description

Index	Property	Unit	Group
1	CountryCode	None	Identification
2	FullName	None	Identification
3	Continent	None	Identification
4	IndependenceYear	None	Demographic
5	AdultPopulation	People	Demographic
6	AnnualBirths	PeoplePerYear	Demographic
7	AnnualDeaths	PeoplePerYear	Demographic
8	BirthRateFraction	PeoplePerPersonPerYear	Demographic
9	ChildPopulation	People	Demographic
10	DeathRateFraction	PeoplePerPersonPerYear	Demographic
11	ElderlyPopulation	People	Demographic
12	InfantMortalityFraction	PeoplePerPerson	Demographic
13	LifeExpectancy	Years	Demographic
14	LiteracyFraction	PeoplePerPerson	Demographic
15	MedianAge	Years	Demographic
16	MigrationRateFraction	PeoplePerPersonPerYear	Demographic
17	Population	People	Demographic
18	PopulationGrowth	PeoplePerPersonPerYear	Demographic
19	PovertyFraction	None	Demographic
20	TotalFertilityRate	PeoplePerPerson	Demographic
21	CurrencyCode	None	Economic
22	ExchangeRate	PerUSDollar	Economic
23	ExternalDebt	USDollars	Economic
24	ForeignExchangeReserves	USDollars	Economic
25	GDP	USDollarsPerYear	Economic
26	GDPAtParity	USDollarsPerYear	Economic
27	GDPPerCapita	USDollarsPerYearPerPerson	Economic
28	GDPRealGrowth	USDollarsPerYearPerYear	Economic
29	GovernmentDebt	USDollars	Economic
30	GovernmentExpenditures	USDollarsPerYear	Economic
31	GovernmentReceipts	USDollarsPerYear	Economic
32	IndustrialProductionGrowth	PerYear	Economic
33	InflationRate	PerYear	Economic
34	LaborForce	People	Economic
35	PriceIndex	None	Economic
36	UnemploymentFraction	None	Economic
37	SectorLaborFractions	None	Economic
38	ExportPartnersFractions	None	Economic
39	ImportPartnersFractions	None	Economic
40	ElectricityConsumption	KilowattHoursPerYear	Energy
41	ElectricityExports	KilowattHoursPerYear	Energy
42	ElectricityImports	KilowattHoursPerYear	Energy
43	ElectricityProduction	KilowattHoursPerYear	Energy
44	NaturalGasConsumption	CubicMetersPerYear	Energy
45	NaturalGasExports	CubicMetersPerYear	Energy
46	NaturalGasImports	CubicMetersPerYear	Energy
47	NaturalGasProduction	CubicMetersPerYear	Energy
48	NaturalGasReserves	CubicMeters	Energy
49	OilConsumption	BarrelsPerDay	Energy
50	OilExports	BarrelsPerDay	Energy
51	OilImports	BarrelsPerDay	Energy
52	OilProduction	BarrelsPerDay	Energy
53	OilReserves	Barrels	Energy

54	Airports	None	Communication
55	AMRadioStations	None	Communication
56	CellularPhones	None	Communication
57	FMRadioStations	None	Communication
58	InternetHosts	None	Communication
59	InternetUsers	People	Communication
60	MerchantShips	None	Communication
61	MerchantShipsDeadWeight	MetricTons	Communication
62	MerchantShipsGross	RegisterTons	Communication
63	PavedAirports	None	Communication
64	PhoneLines	None	Communication
65	RadioStations	None	Communication
66	RailwayLength	Kilometers	Communication
67	RoadLength	Kilometers	Communication
68	ShortWaveRadioStations	None	Communication
69	TelevisionStations	None	Communication
70	UnpavedAirports	None	Communication
71	ArableLandArea	SquareKilometers	Geography
72	ArableLandFraction	None	Geography
73	Area	SquareKilometers	Geography
74	BoundaryLength	Kilometers	Geography
75	CoastlineLength	Kilometers	Geography
76	CropsLandArea	SquareKilometers	Geography
77	CropsLandFraction	None	Geography
78	HighestElevation	Meters	Geography
79	IrrigatedLandArea	SquareKilometers	Geography
80	IrrigatedLandFraction	None	Geography
81	LandArea	SquareKilometers	Geography
82	LowestElevation	Meters	Geography
83	WaterArea	SquareKilometers	Geography
84	MilitaryAgeFemales	People	Military
85	MilitaryAgeMales	People	Military
86	MilitaryAgePopulation	People	Military
87	MilitaryAgeRate	PeoplePerYear	Military
88	MilitaryExpenditureFraction	None	Military
89	MilitaryExpenditures	USDollarsPerYear	Military
90	MilitaryFitPopulation	People	Military

A.2 Countries Dictionary

AF	Islamic Republic of Afghanistan	Asia
AL	Republic of Albania	Europe
DZ	People's Democratic Republic of Algeria	Africa
AD	Principality of Andorra	Europe
AO	Republic of Angola	Africa
AG	Antigua and Barbuda	NorthAmerica
AR	Argentine Republic	SouthAmerica
AM	Republic of Armenia	Asia
AU	Commonwealth of Australia	Oceania
AT	Republic of Austria	Europe
AZ	Republic of Azerbaijan	Asia
BS	Commonwealth of The Bahamas	NorthAmerica
BH	Kingdom of Bahrain	Asia
BD	People's Republic of Bangladesh	Asia
BB	Barbados	NorthAmerica
BY	Republic of Belarus	Europe
BE	Kingdom of Belgium	Europe
BZ	Belize	NorthAmerica
BJ	Republic of Benin	Africa
BT	Kingdom of Bhutan	Asia
BO	Plurinational State of Bolivia	SouthAmerica
BA	Bosnia and Herzegovina	Europe
BW	Republic of Botswana	Africa
BR	Federative Republic of Brazil	SouthAmerica
BN	Brunei Darussalam	Asia
BG	Republic of Bulgaria	Europe
BF	Burkina Faso	Africa
BI	Republic of Burundi	Africa
KH	Kingdom of Cambodia	Asia
CM	Republic of Cameroon	Africa
CA	Canada	NorthAmerica
CV	Republic of Cape Verde	Africa
CF	Central African Republic	Africa
TD	Republic of Chad	Africa
CL	Republic of Chile	SouthAmerica
CN	People's Republic of China	Asia
CO	Republic of Colombia	SouthAmerica
KM	Union of the Comoros	Africa
CR	Republic of Costa Rica	NorthAmerica
HR	Republic of Croatia	Europe
CU	Republic of Cuba	NorthAmerica
CY	Republic of Cyprus	Asia
CZ	Czech Republic	Europe
CD	Democratic Republic of the Congo	Africa
DK	Kingdom of Denmark	Europe
DJ	Republic of Djibouti	Africa
DM	Commonwealth of Dominica	NorthAmerica
DO	Dominican Republic	NorthAmerica
TL	Democratic Republic of Timor-Leste	Asia
EC	Republic of Ecuador	SouthAmerica
EG	Arab Republic of Egypt	Africa
SV	Republic of El Salvador	NorthAmerica
GQ	Republic of Equatorial Guinea	Africa
ER	State of Eritrea	Africa
EE	Republic of Estonia	Europe
ET	Federal Democratic Republic of Ethiopia	Africa

FJ	Republic of the Fiji Islands	Oceania
FI	Republic of Finland	Europe
FR	French Republic	Europe
GA	Gabonese Republic	Africa
GM	Republic of The Gambia	Africa
GE	Georgia	Asia
DE	Federal Republic of Germany	Europe
GH	Republic of Ghana	Africa
GR	Hellenic Republic	Europe
GD	Grenada	NorthAmerica
GT	Republic of Guatemala	NorthAmerica
GN	Republic of Guinea	Africa
GW	Republic of Guinea-Bissau	Africa
GY	Cooperative Republic of Guyana	SouthAmerica
HT	Republic of Haiti	NorthAmerica
HN	Republic of Honduras	NorthAmerica
HU	Hungary	Europe
IS	Iceland	Europe
IN	Republic of India	Asia
ID	Republic of Indonesia	Asia
IR	Islamic Republic of Iran	Asia
IQ	Republic of Iraq	Asia
IE	Ireland	Europe
IL	State of Israel	Asia
IT	Italian Republic	Europe
CI	Republic of Cote d'Ivoire	Africa
JM	Jamaica	NorthAmerica
JP	Japan	Asia
JO	Hashemite Kingdom of Jordan	Asia
KZ	Republic of Kazakhstan	Asia
KE	Republic of Kenya	Africa
KI	Republic of Kiribati	Oceania
KW	State of Kuwait	Asia
KG	Kyrgyz Republic	Asia
LA	Lao People's Democratic Republic	Asia
LV	Republic of Latvia	Europe
LB	Lebanese Republic	Asia
LS	Kingdom of Lesotho	Africa
LR	Republic of Liberia	Africa
LY	Great Socialist People's Libyan Arab Jamahiriya	Africa
LI	Principality of Liechtenstein	Europe
LT	Republic of Lithuania	Europe
LU	Grand Duchy of Luxembourg	Europe
MK	Republic of Macedonia (FYROM)	Europe
MG	Republic of Madagascar	Africa
MW	Republic of Malawi	Africa
MY	Malaysia	Asia
MV	Republic of Maldives	Asia
ML	Republic of Mali	Africa
MT	Republic of Malta	Europe
MH	Republic of the Marshall Islands	Oceania
MR	Islamic Republic of Mauritania	Africa
MU	Republic of Mauritius	Africa
MX	United Mexican States	NorthAmerica
FM	Federated States of Micronesia	Oceania
MD	Republic of Moldova	Europe
MC	Principality of Monaco	Europe
MN	Mongolia	Asia

ME	Republic of Montenegro	Europe
MA	Kingdom of Morocco	Africa
MZ	Republic of Mozambique	Africa
MM	Union of Myanmar	Asia
NA	Republic of Namibia	Africa
NR	Republic of Nauru	Oceania
NP	Federal Democratic Republic of Nepal	Asia
NL	Kingdom of the Netherlands	Europe
NZ	New Zealand	Oceania
NI	Republic of Nicaragua	NorthAmerica
NE	Republic of Niger	Africa
NG	Federal Republic of Nigeria	Africa
KP	Democratic People's Republic of Korea	Asia
NO	Kingdom of Norway	Europe
OM	Sultanate of Oman	Asia
PK	Islamic Republic of Pakistan	Asia
PW	Republic of Palau	Oceania
PA	Republic of Panama	NorthAmerica
PG	Independent State of Papua New Guinea	Oceania
PY	Republic of Paraguay	SouthAmerica
PE	Republic of Peru	SouthAmerica
PH	Republic of the Philippines	Asia
PL	Republic of Poland	Europe
PT	Portuguese Republic	Europe
QA	State of Qatar	Asia
CG	Republic of the Congo	Africa
RO	Romania	Europe
RU	Russian Federation	Asia
RW	Republic of Rwanda	Africa
KN	Federation of Saint Kitts and Nevis	NorthAmerica
LC	Saint Lucia	NorthAmerica
VC	Saint Vincent and the Grenadines	NorthAmerica
WS	Independent State of Samoa	Oceania
SM	Republic of San Marino	Europe
ST	Democratic Republic of Sao Tome and Principe	Africa
SA	Kingdom of Saudi Arabia	Asia
SN	Republic of Senegal	Africa
RS	Republic of Serbia	Europe
SC	Republic of Seychelles	Africa
SL	Republic of Sierra Leone	Africa
SG	Republic of Singapore	Asia
SK	Slovak Republic	Europe
SI	Republic of Slovenia	Europe
SB	Solomon Islands	Oceania
SO	Somalia	Africa
ZA	Republic of South Africa	Africa
KR	Republic of Korea	Asia
ES	Kingdom of Spain	Europe
LK	Democratic Socialist Republic of Sri Lanka	Asia
SD	Republic of the Sudan	Africa
SR	Republic of Suriname	SouthAmerica
SZ	Kingdom of Swaziland	Africa
SE	Kingdom of Sweden	Europe
CH	Swiss Confederation	Europe
SY	Syrian Arab Republic	Asia
TJ	Republic of Tajikistan	Asia
TZ	United Republic of Tanzania	Africa
TH	Kingdom of Thailand	Asia

TG	Togolese Republic	Africa
TO	Kingdom of Tonga	Oceania
TT	Republic of Trinidad and Tobago	NorthAmerica
TN	Tunisian Republic	Africa
TR	Republic of Turkey	Asia
TM	Turkmenistan	Asia
TV	Tuvalu	Oceania
UG	Republic of Uganda	Africa
UA	Ukraine	Europe
AE	United Arab Emirates	Asia
GB	United Kingdom of Great Britain and Northern Ireland	Europe
US	United States of America	NorthAmerica
UY	Oriental Republic of Uruguay	SouthAmerica
UZ	Republic of Uzbekistan	Asia
VU	Republic of Vanuatu	Oceania
VE	Bolivarian Republic of Venezuela	SouthAmerica
VN	Socialist Republic of Vietnam	Asia
YE	Republic of Yemen	Asia
ZM	Republic of Zambia	Africa
ZW	Republic of Zimbabwe	Africa

A.3 Preprocessing Code

```
SetDirectory[NotebookDirectory[]]
(*Index of the Selected Variables *)
vars1 = {33, 84, 30, 111, 1, 8, 9, 14, 25, 41, 45, 114, 132, 133, 148,
    154, 187, 188, 189, 212, 36, 60, 66, 80, 87, 88, 89, 90, 94, 95,
   96, 112, 116, 126, 190, 216, 202, 64, 108, 52, 53, 54, 55, 169,
   170, 171, 172, 173, 176, 177, 178, 179, 180, 4, 7, 22, 79, 120,
   121, 150, 151, 152, 182, 184, 191, 194, 199, 204, 208, 219, 11, 12,
    13, 17, 28, 34, 35, 100, 123, 124, 127, 134, 222, 155,
   156, 157, 158, 159, 160, 163};
(*Prints the variable Map*)
Prepend[{Range[Length[vars1]], CountryData["Properties"][[vars1]],
     CountryData["US", #, "Units"] & /@
      CountryData["Properties"][[vars1]]}\[Transpose], {"Index",
    "Property", "Unit"}] // TableForm;
(*Retrieves the Selected variables of the countries of the United \
Nations from the Wolfram|Alpha Knowledge Base*)
s1 = Transpose[
   ParallelTable[
    CountryData[CountryData["UN"][[j]],
     CountryData["Properties"][[i]]], {i, vars1}, {j,
     Length[CountryData["UN"]]}]];
(*Converts Quantity objects to plain plain text*)
q1 = Flatten[
   Position[
    Table[AnyTrue[QuantityQ /@ (s1\[Transpose][[j]]), TrueQ], {j,
      Length[s1\[Transpose]]}], True]];
For[ii = 1, ii <= Length[s1], ii++,
 s1[[ii, q1]] = QuantityMagnitude[s1[[ii, q1]]]
 ]
(*Takes the log base 10 of a subset of the selected variables*)
log = Complement[
   Range[Length[s1\[Transpose]]], {1, 2, 3, 21, 37, 38, 39}, {8, 10,
    12, 13, 14, 15, 16, 18, 19, 20, 28, 32, 33, 35, 36, 72, 78, 80,
    88}];
For[ii = 1, ii <= Length[log], ii++,
 s1[[All, notlog[[ii]]]] = Log[10, s1[[All, notlog[[ii]]]]]
 ]
(*Converts Entity Object to plain text*)
s1[[All, 3]] = CanonicalName[s1\[Transpose][[3]]];
s1[[All, 4]] = Map[Part[#, 1] &, Normal /@ s1[[All, 4]]];
s1[[All, 37]] =
  Part[#, 1, 1] & /@ (Sort[#, #1[[2]] > #2[[2]] &] & /@ s1[[All, 37]]);
s1[[All, 38]] =
  CanonicalName[
   Part[#, 1, 1] & /@ (Sort[#, #1[[2]] > #2[[2]] &] & /@
      s1[[All, 38]])];
s1[[All, 39]] =
  CanonicalName[
   Part[#, 1, 1] & /@ (Sort[#, #1[[2]] > #2[[2]] &] & /@
      s1[[All, 39]])];
(*Signals correctly the missing Data for output*)
s1 = Replace[s1,
   Missing["NotAvailable"][[1, 1]] -> Missing["NotAvailable"], 2];
s1 = Replace[s1,
   CanonicalName[Missing["NotAvailable"][[1, 1]]] ->
    Missing["NotAvailable"], 2];
```

```
s1 = Replace[s1,
   QuantityMagnitude[Missing["NotAvailable"]] ->
    Missing["NotAvailable"], 2];
s1 = Replace[s1,
   QuantityMagnitude[Missing["NotApplicable"]] ->
    Missing["NotAvailable"], 2];
s1 = Replace[s1, "NotApplicable" -> Missing["NotAvailable"], 2];
(*Removes undesired countries*)
s1 = Select[
   s1, ! IntersectingQ[{#[[1]]}, {"CX", "CC", "FK",
       Missing["NotApplicable"], "NU", "NF", "PN", "SJ", "TK", "VA",
       "WF", "SS"}] &];
(*Save binaries of the computation*)
s1 >> "s1.mx"
(*Retrieve the binaries*)
<< s1.mx;
(*Export to excel*)
Export["s1.xls",
 Insert[s1, CountryData["Properties"][[vars1]], 1]]
```

A.4 Figures

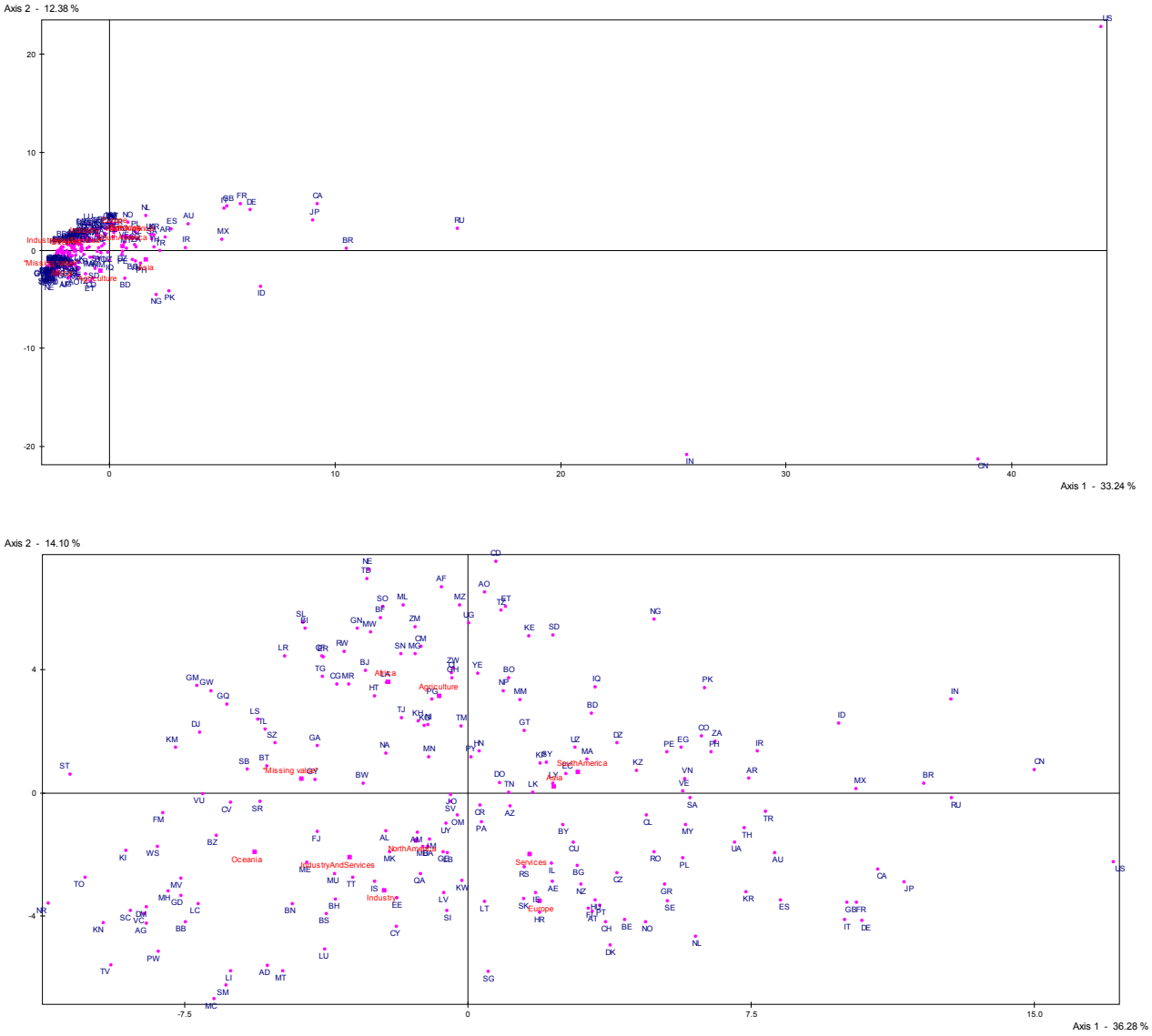
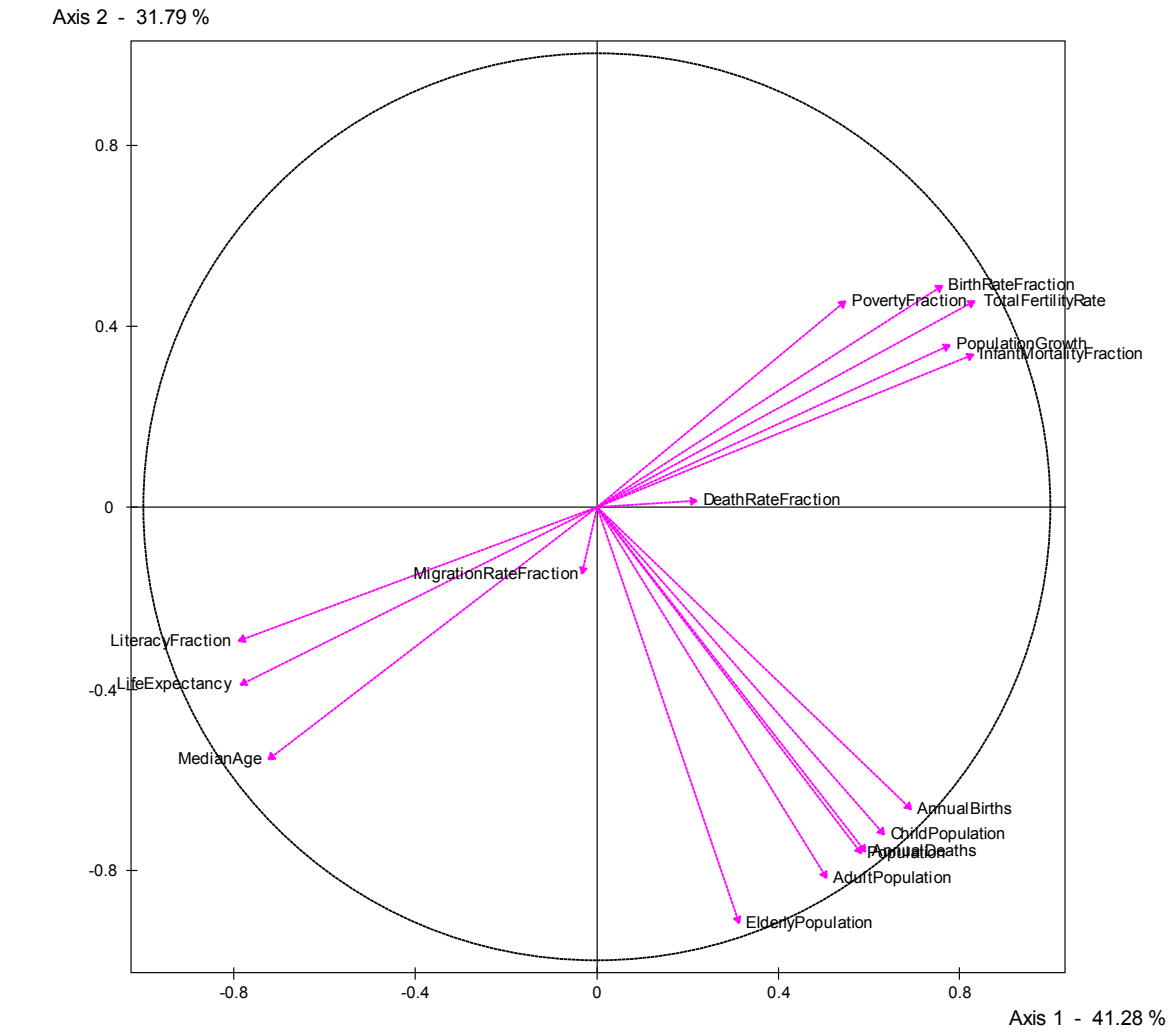


Figure A-5: Comparative results of the *naive* PCA using original and synthetic variables respectively. Naive PCA consists on a PCA with all continuous variables as active variables. In the PCA with the original variables, the distribution of the cloud of points is heavily left skewed, a few outliers can be seen to the right. In the PCA done with synthetic variables we see a more normal distribution of the cloud of points, thus densifiyng the center of mass still preserving roughly the order of previous observation points and the inertia of the factors.



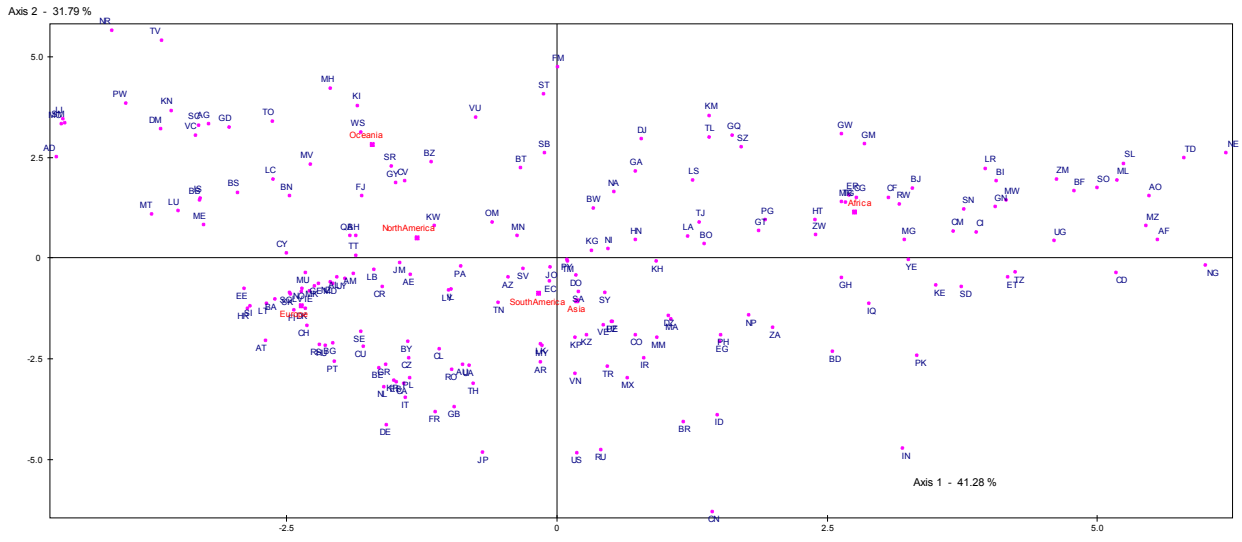


Figure A-6: Results of the normed PCA carried out on the Demographic group of variables in the set.

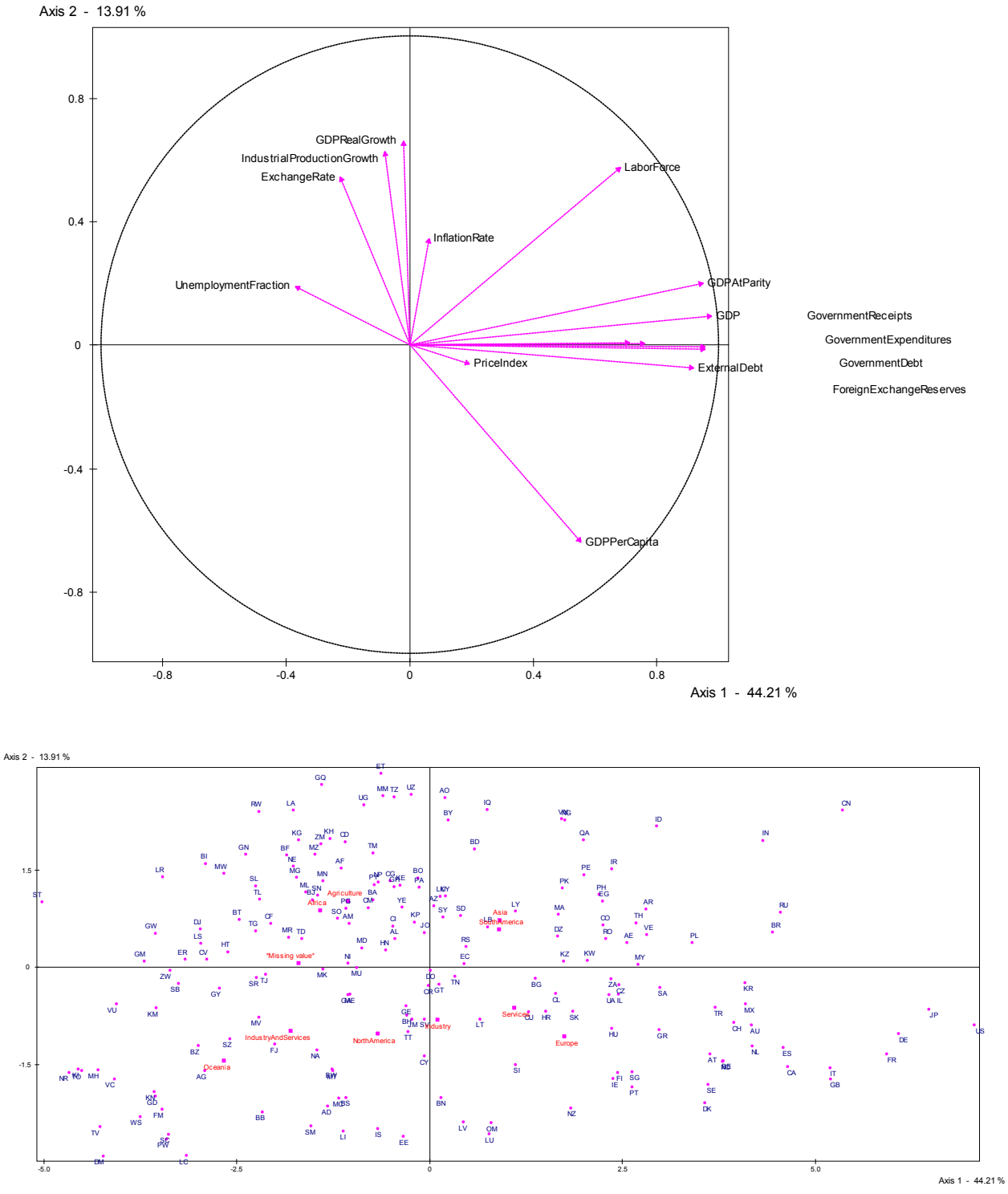


Figure A-7: Results of the normed PCA carried out on the Economic group of variables in the set.

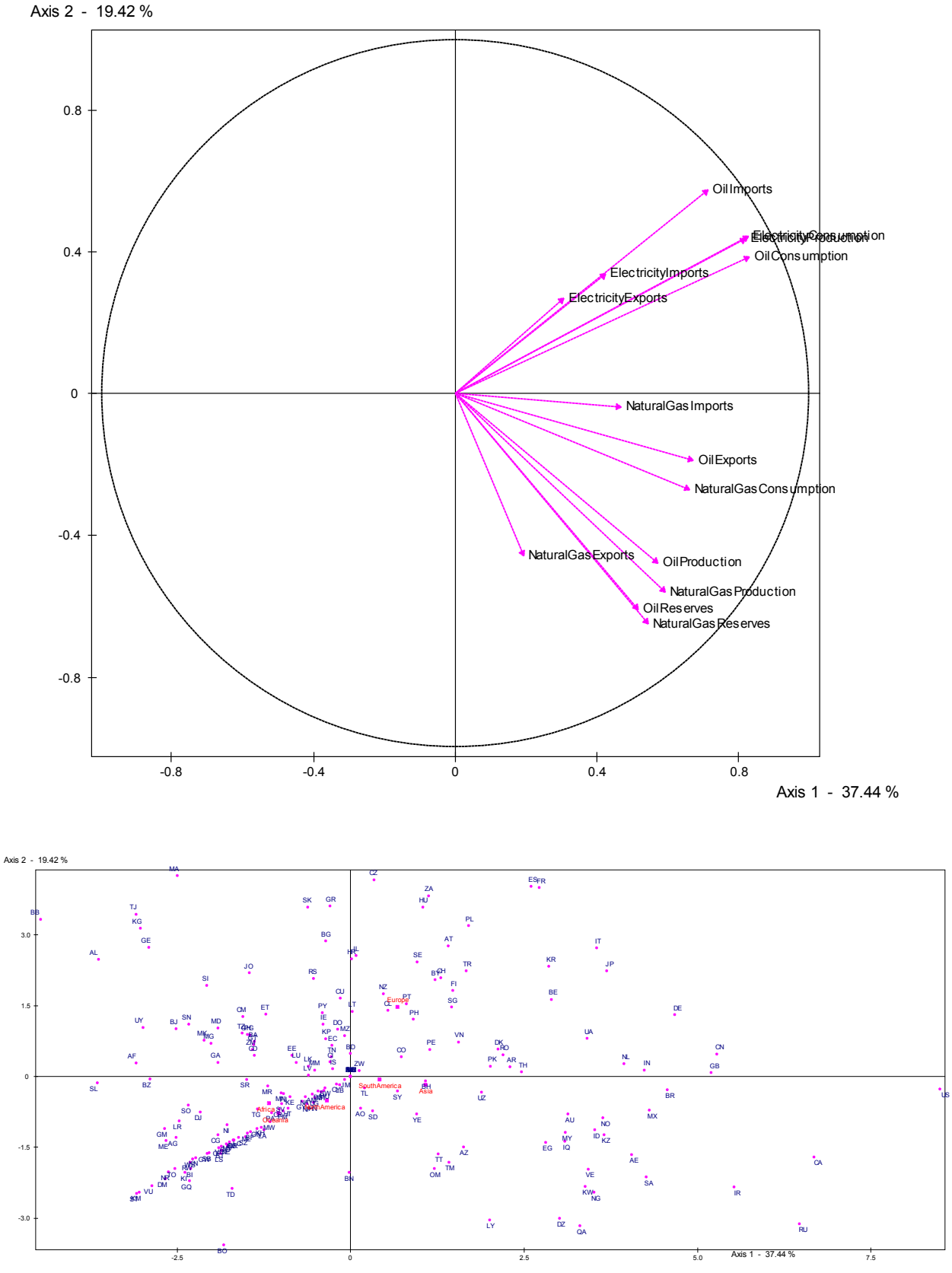


Figure A-8: Results of the normed PCA carried out on the Energy group of variables in the set.

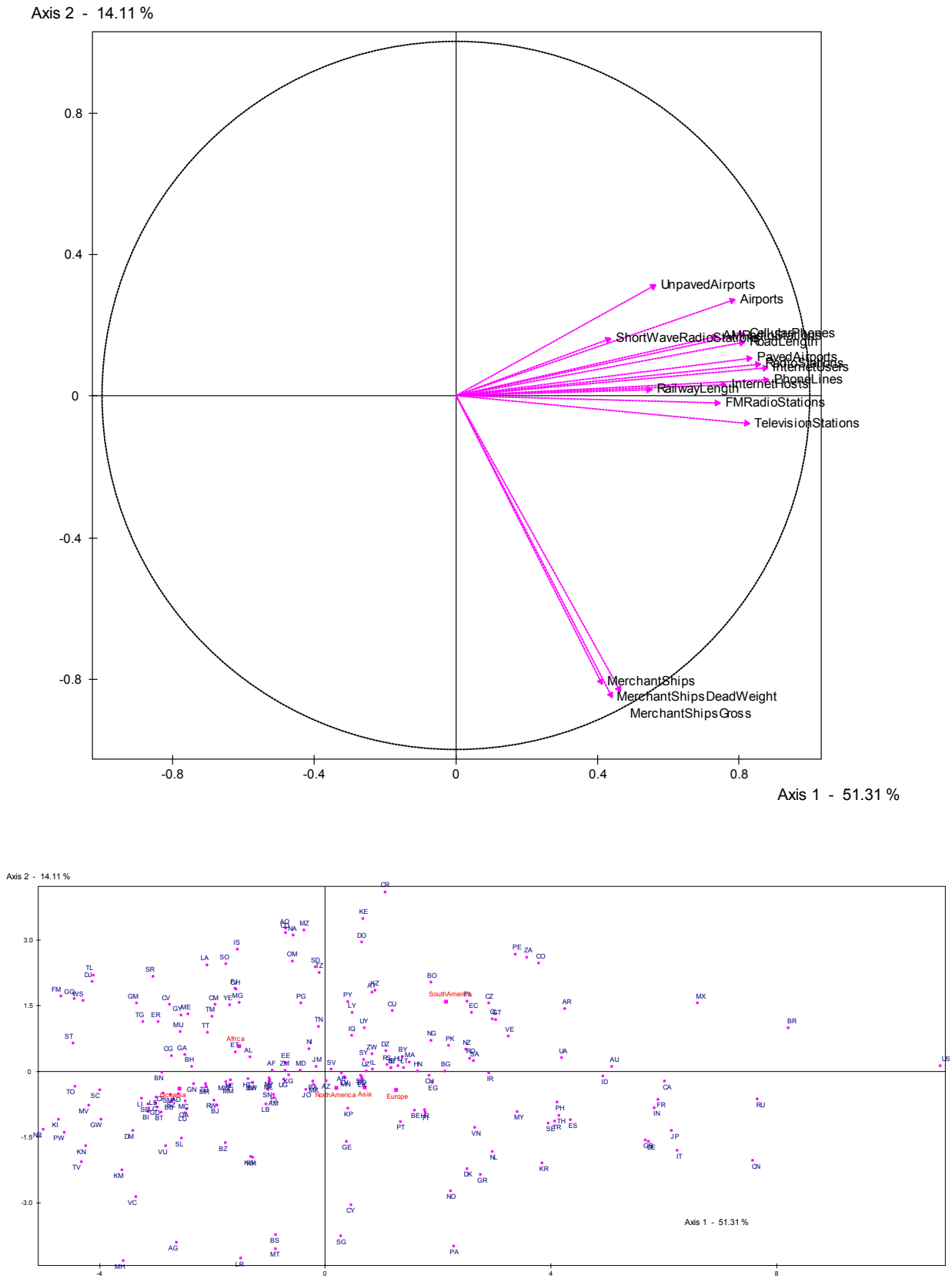
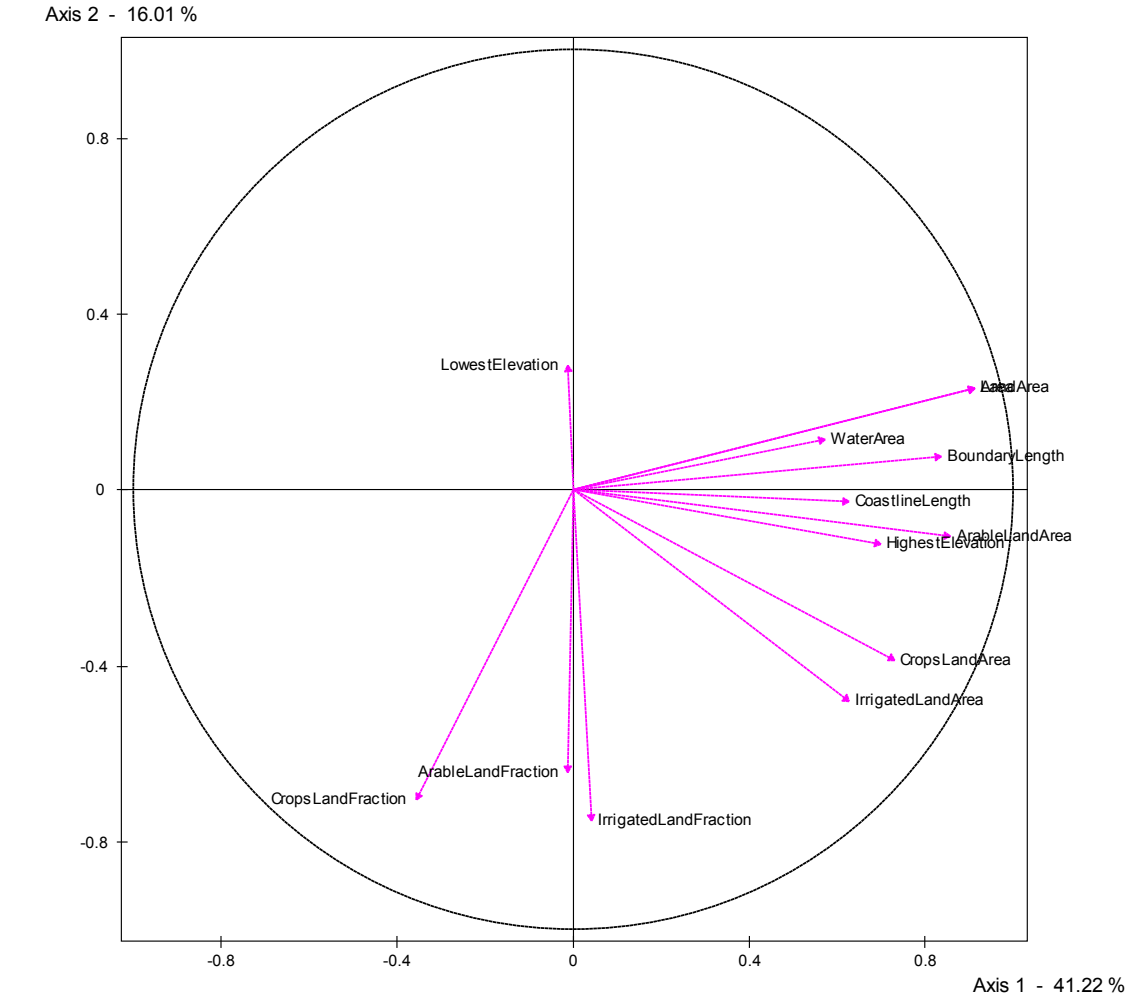


Figure A-9: Results of the normed PCA carried out on the Communication group of variables in the set.



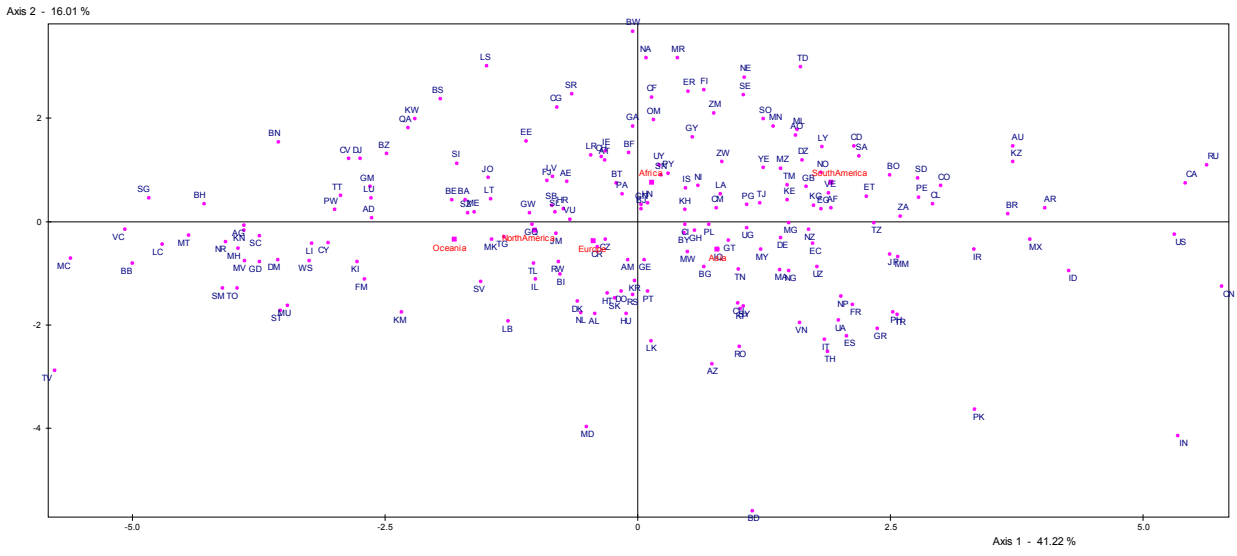
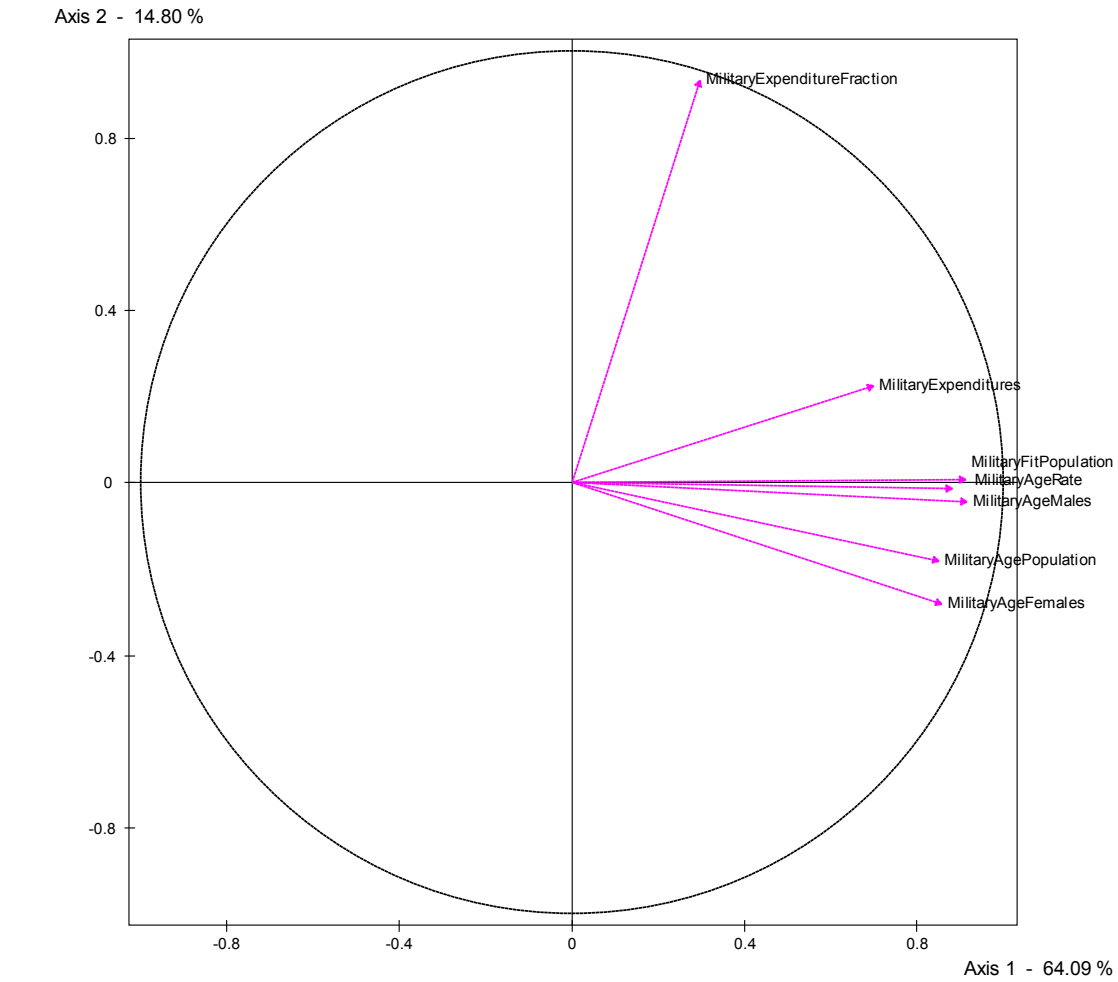


Figure A-10: Results of the normed PCA carried out on the Geography group of variables in the set.



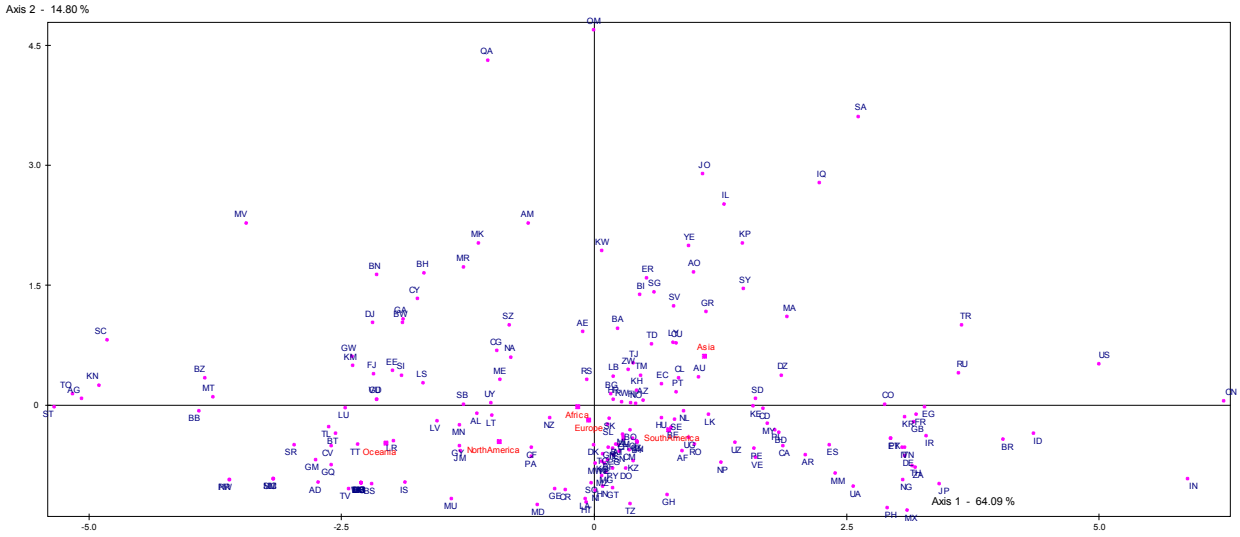


Figure A-11: Results of the normed PCA carried out on the Military group of variables in the set.