Global Energy Dynamics

a Data Analysis Project

Panagiotis Podiotis

University of Macedonia

Department of International and European Studies

Thessaloniki, Greece

The present paper can be considered a Data Analysis project. Energy-related data within the CIA World Factbook (hereinafter Factbook) (CIA, 2019, p. About) were explored. The Factbook[1] was chosen as the source of data for this project for the following reasons:

1. It provides a big volume of data across a variety of sectors[2].
2. High confidence can be placed on the data since the CIA is one of the most recognized and prestigious US governmental agencies specializing on intelligence.
3. Data are open-source and available for download.
4. The author has worked extensively with the Factbook by having converted it in dataset format (Podiotis, 2020) and by having conducted various ML sub-projects with it.
5. There is a general lack of extensive datasets which cover International Relation topics in such a wholistic manner.

Data Science provides with almost endless research approaches. In an effort to limit the scope of the paper and increase the quality of research, it was decided that few specific tasks will be performed and always in regard to energy-related data:

1. Comparative Data Visualizations between Factbook's Regions (Continents). Visualizations will cover cumulative and averaged resource reserves, production, imports, exports and electricity production sources.
2. Various implementations of clustering covering different subgroups of data, regions and ultimately countries.

By fulfilling the objectives described above, the author aims to:

1. Lay the ground for future research and hypothesis with the findings of this paper.
2. Provide with a better understanding of global dynamics.
3. Uncover the structure of the world (and countries) in terms of energy. Moreover, key regions and countries will be uncovered.

All of the above tasks were out with the help of Python 3.7 & 3.8 (Python Foundation, 2001) programming language, within the JetBrains PyCharm Community Edition 2019.2.2 x64 (JetBrains, 2000) Integrated Development Environment.

---

[1] The latest available version (2018 version, uploaded on Jan 04, 2019 01:51 PM – downloaded 15/05/2020) of the CIA World Factbook which was downloaded through the official web portal was used.
[2] The Factbook covers 12 thematic areas for each entry (country), namely: History, People and Society, Government, Economy, Energy, Geography, Communications, Transportation, Military, Terrorism, Transnational Issues. Each Thematic area (category) consists of numeric or textual data points.

The Data

All data used in this paper are drawn from the CIA World Factbook. The latest available version (2018 version, uploaded on Jan 04, 2019 01:51 PM – downloaded 15/05/2020) of the CIA World Factbook which was downloaded through the official web portal was used. The Factbook was converted into a dataset by downloading and processing data from its online version[3] in which certain missing values were imputed with a Mean Average Percentage Error smaller than 15%. The Dataset is in a two-dimensional table format with individual entities (countries/territories) being the rows and columns representing different data (ex. GDP, Population etc.). More than forty columns contain data related directly or indirectly to energy out of which fourteen were selected for the present paper. The 2018 version of the CIA World Factbook consists of data drawn during the period 2015-2018. The classification of countries into regions/continents was adopted directly from the Factbook. Specifically:

Table 1 - Factbook's Regions and respective countries.

Central America (32)	Middle East (19)	S. Asia (8)	Africa (55)	Europe (51)	Oceania (26)	S. America (13)	N. America (6)	E./S.E. Asia (21)	Central Asia (6)
Aruba, Antigua And Barbuda, Anguilla, Barbados, Bahamas, The, Belize, Cayman Islands, Costa Rica, Cuba, Dominica, Dominican Republic, El Salvador, Grenada, Guatemala, Haiti, Honduras, Jamaica, Montserrat, Sint Maarten, Nicaragua, Panama, Saint Martin, Puerto Rico, Saint Kitts And Nevis, Saint Lucia, Saint Barthelemy, Trinidad And Tobago, Turks And Caicos Islands, Curacao, Saint Vincent And The Grenadines, British Virgin Islands, Virgin Islands	United Arab Emirates, Azerbaijan, Armenia, Bahrain, Georgia, Gaza Strip, Iran, Israel, Iraq, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syria, Turkey, West Bank, Yemen	Afghanistan, Bangladesh, Bhutan, Sri Lanka, India, Maldives, Nepal, Pakistan	Algeria, Angola, Botswana, Benin, Burundi, Chad, Congo, DR Congo, Cameroon, Comoros, Central African Republic, Cabo Verde, Djibouti, Egypt, Equatorial Guinea, Eritrea, Ethiopia, Gambia, The, Gabon, Ghana, Guinea, Cote D'Ivoire, Kenya, Liberia, Lesotho, Libya, Madagascar, Malawi, Mali, Morocco, Mauritius, Mauritania, Mozambique, Niger, Nigeria, South Sudan, Guinea-Bissau, Rwanda, Seychelles, South Africa, Senegal, Saint Helena, Ascension, And Tristan Da Cunha, Sierra Leone, Somalia, Sudan, Togo, Sao Tome And Principe, Tunisia, Tanzania, Uganda, Burkina Faso, Namibia, Eswatini, Zambia, Zimbabwe	Albania, Andorra, Austria, Belgium, Bosnia And Herzegovina, Belarus, Bulgaria, Cyprus, Denmark, Ireland, Estonia, Czechia, Finland, Faroe Islands, France, Gibraltar, Guernsey, Germany, Greece, Croatia, Hungary, Iceland, Isle Of Man, Italy, Jersey, Kosovo, Latvia, Lithuania, Slovakia, Liechtenstein, Luxembourg, Moldova, Montenegro, Macedonia, Monaco, Malta, Netherlands, Norway, Poland, Portugal, Serbia, Romania, Slovenia, San Marino, Spain, Svalbard, Sweden, Switzerland, United Kingdom, Ukraine, Holy See (Vatican City)	American Samoa, Australia, Solomon Islands, Cocos (Keeling) Islands, Northern Mariana Islands, Cook Islands, Fiji, Micronesia, Federated States Of, French Polynesia, Guam, Kiribati, Christmas Island, New Caledonia, Niue, Norfolk Island, Vanuatu, Nauru, New Zealand, Pitcairn Islands, Palau, Marshall Islands, Tokelau, Tonga, Tuvalu, Wallis And Futuna, Samoa	Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Falkland Islands (Islas Malvinas), Guyana, Suriname, Paraguay, Peru, Uruguay, Venezuela	Bermuda, Canada, Greenland, Mexico, Saint Pierre and Miquelon, United States	Burma, Brunei, Cambodia, China, Hong Kong, Indonesia, Japan, Korea, North, Korea, South, Laos, Macau, Mongolia, Malaysia, Paracel Islands, Papua New Guinea, Philippines, Singapore, Thailand, Timor-Leste, Taiwan, Vietnam	Kyrgyzstan, Kazakhstan, Russia, Tajikistan, Turkmenistan, Uzbekistan

---

[3] For technical details see (Podiotis, 2020).

Out of all energy data columns, fifteen were selected for analysis in the scope of this paper. These columns cover five thematic groups described below.

Table 2 - Project's Data Categories and Columns.

Reserves (Proved)	Production	Exports	Imports	Electricity Production Sources	Ports & Terminals	Pipelines
Gas, Oil	Oil, Gas, Refined Petroleum Products	Oil, Gas, Refined Petroleum Products	Oil, Gas, Refined Petroleum Products	Fossil Fuels, Nuclear, Hydroelectric, Other/Renewable	LNG (import), LNG (export), Oil	Condensate (gas), Gas, LPG, Oil, Oil/Gas/Water

All energy quantities are measured in Giga Joules for uniformity and ease of comparison. The CIA measures oil production, exports and imports in barrels per day (bbl/day) and gas in year-cumulative cubic meters. All appropriate conversions have been made.

Clustering was done with Mean Shift[4] in order to remove all possible bias in cluster creation. Data were transformed to the same metric and then scaled before clustering. Descriptive statistics (mean, standard deviation, percentiles) were used to study each cluster. All of the above were implemented with libraries sklearn (scikit-learn developers, 2020) and pandas (Wes McKinney and the Pandas Development Team, 2020). Visualizations were done matplotlib (Hunter, 2007).

Region-cumulative data provide with a general description of the global dynamics. Caution should be given when generalizations between region and country-level assumptions are made

---

[4] Mean Shift is an unsupervised machine learning algorithm which performs cluster analysis. Its advantage is the fact that it automatically detects clusters based on density thus eliminating author's bias.

Proved Reserves

Proved Reserves

The Middle East has by far the biggest proven reserves of both natural gas and crude oil. It is important to note that the Middle East not only has the quantitative advantage in terms of resources but also the qualitative one. This is due to the fact that Middle Eastern oil is located near the ground and can be extracted easier with lower fixed costs. This geomorphological characteristic largely contributes to the ease with which Gulf states weaponize oil prices.

When the amount of resources is divided by the number of countries in the respective region, the picture changes. Central Asian countries are boosted to the first place in regards to natural gas while the difference between the Middle East and South, North America decreases. This means that resources in the Middle East are more unevenly distributed between countries. This is the case with Gulf Monarchies which posses the majority of crude oil reserves. Such is the case with Africa which is formed by many countries out of which few possess important reserves.

Yearly Resource Production

Country-Average Yearly Resource Production
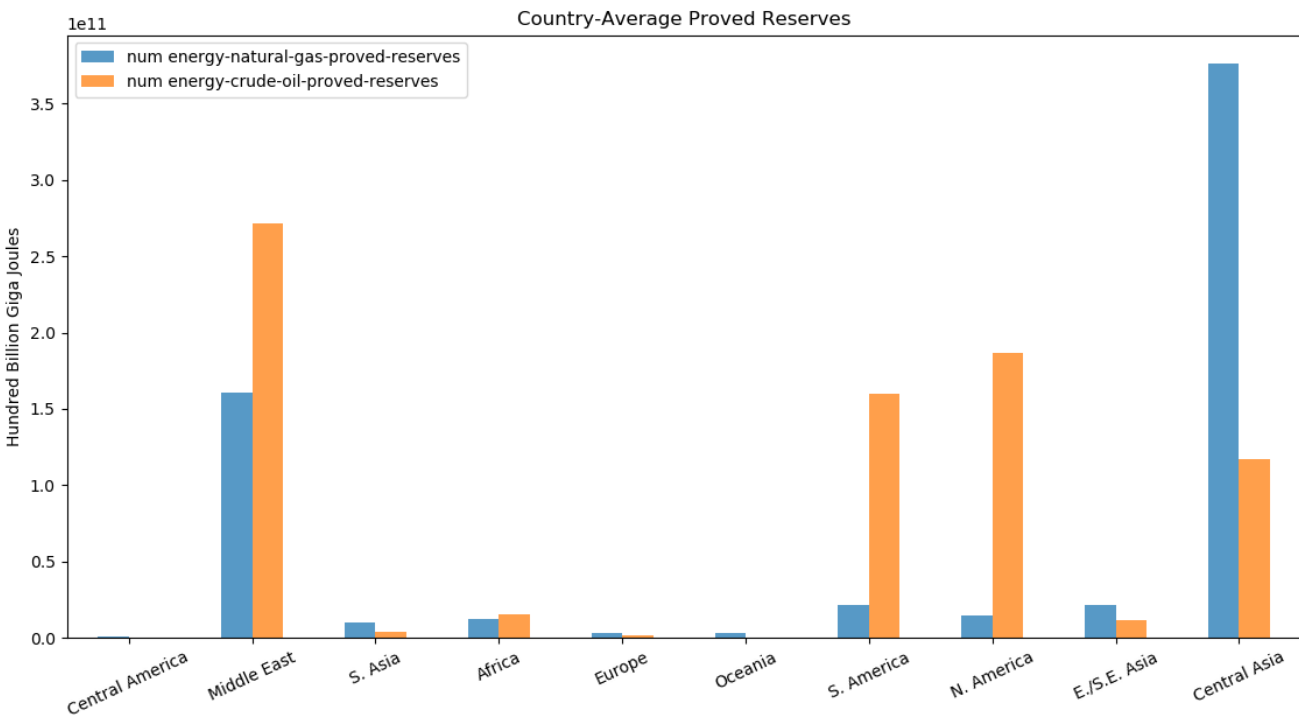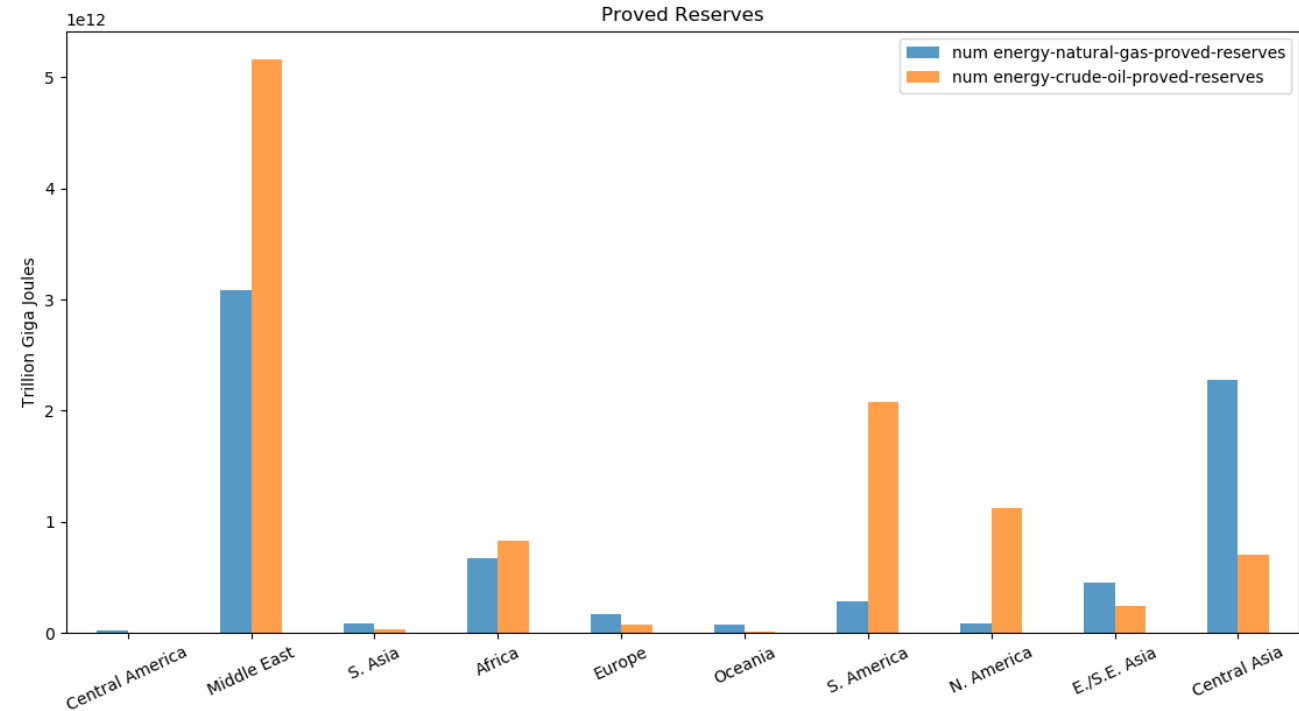
Resource Production
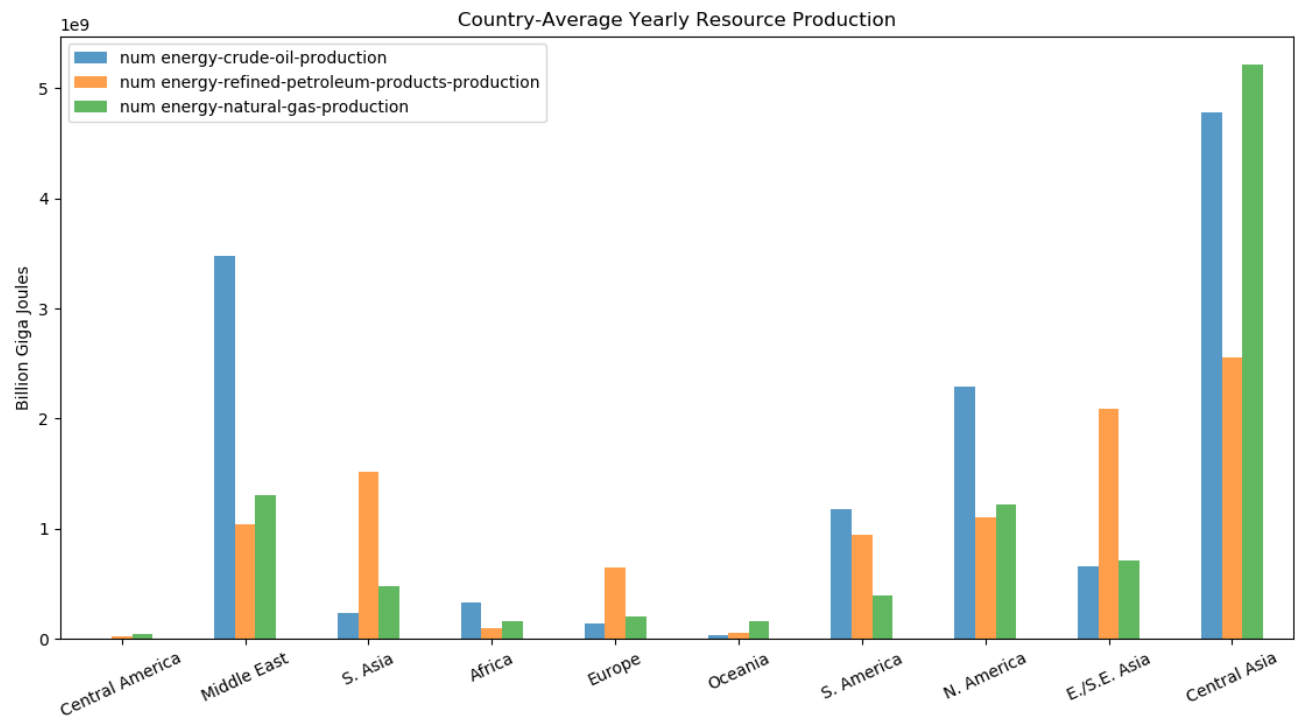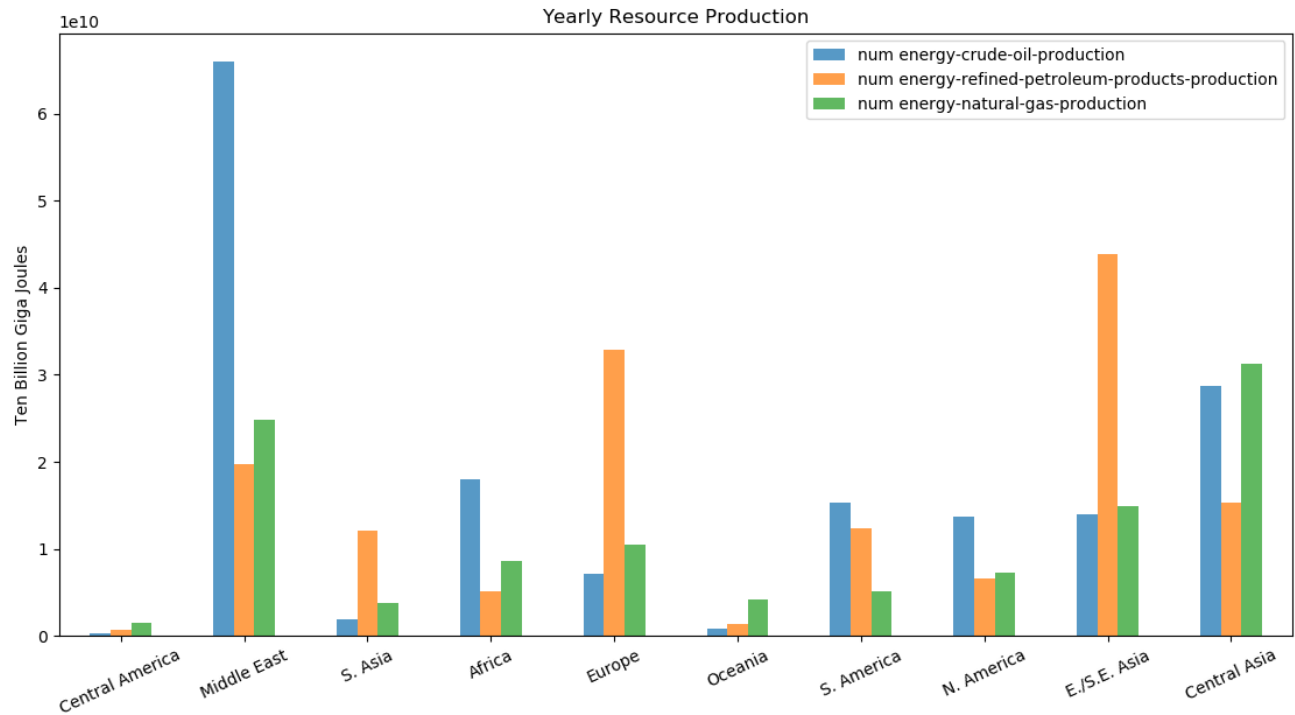
The Middle East outnumbers every other continent in terms of cumulative crude oil production. Interestingly, its refined petroleum products production is relatively low. Middle Eastern countries and private sector are more willing to simply extract crude oil and then export it rather than to invest in refining infrastructure. The profitability of this policy should be further examined. Europe has a high level of refined products, probably due to the technologically intense character of EU economies.

The averaged graph reveals more information on the matter. Central America appears to be the omnipotent region in terms of production. Its six member countries (Kyrgyzstan, Kazakhstan, Russia, Tajikistan, Turkmenistan, Uzbekistan) are all producers and some of them (Russia, Turkmenistan, Uzbekistan when it comes to natural gas) occupy top global positions. Middle Eastern production is once again unevenly distributed (since it becomes comparatively smaller in the averaged graph). Interestingly, refinement is also unequally distributed in Europe. North America and East/Southeast Asia remain relatively unchanged (except refinement in Southeast Asia) suggesting homogeneity of their members.

Yearly Imports

Country-Average Yearly Imports

Imports

NOTE: Regional Imports were calculated based on each individual country's imports. Thus, they also reflect intraregional trade.

East/Southeast Asia and Europe are the largest importers of crude oil, gas and refined products. Considering the production of these regions presented above, it can be assumed that in both cases refined products imports are intraregional. Moreover, considering East/Southeast Asia's larger production of crude oil and gas compared to Europe we can assume that European is the biggest importer of energy. America (Central, South and North) enjoys a high level of energy autonomy. Middle Eastern imports reflect intraregional trade. Central Asian countries are highly homogenous in regards to energy resources. This is probably the reason why imports are almost null while production (both cumulative and averaged) is high. There seems to be an absence of trade both inter and intra-regional.

The country-averaged data bring South Asia to the spotlight. South Asia has some of the smallest reserves and production worldwide, yet its members import the biggest amount of crude oil (averaged). South Asia is plagued by lack of regional cooperation and energy production potential and resources.

Key Take Away:

-Intraregional trade should be considered when trade is studied. It can provide an index of cooperation.

-South Asia is the most energy dependent region with the highest vulnerability followed be Europe.

-European energy demand is distributed between its many states. Its dependency to Russia for gas complicates the political realities. Yet, even though European energy security is heavily referenced in literature few focus on South Asian which looks grimmer.

Yearly Exports

Country-Average Yearly Exports
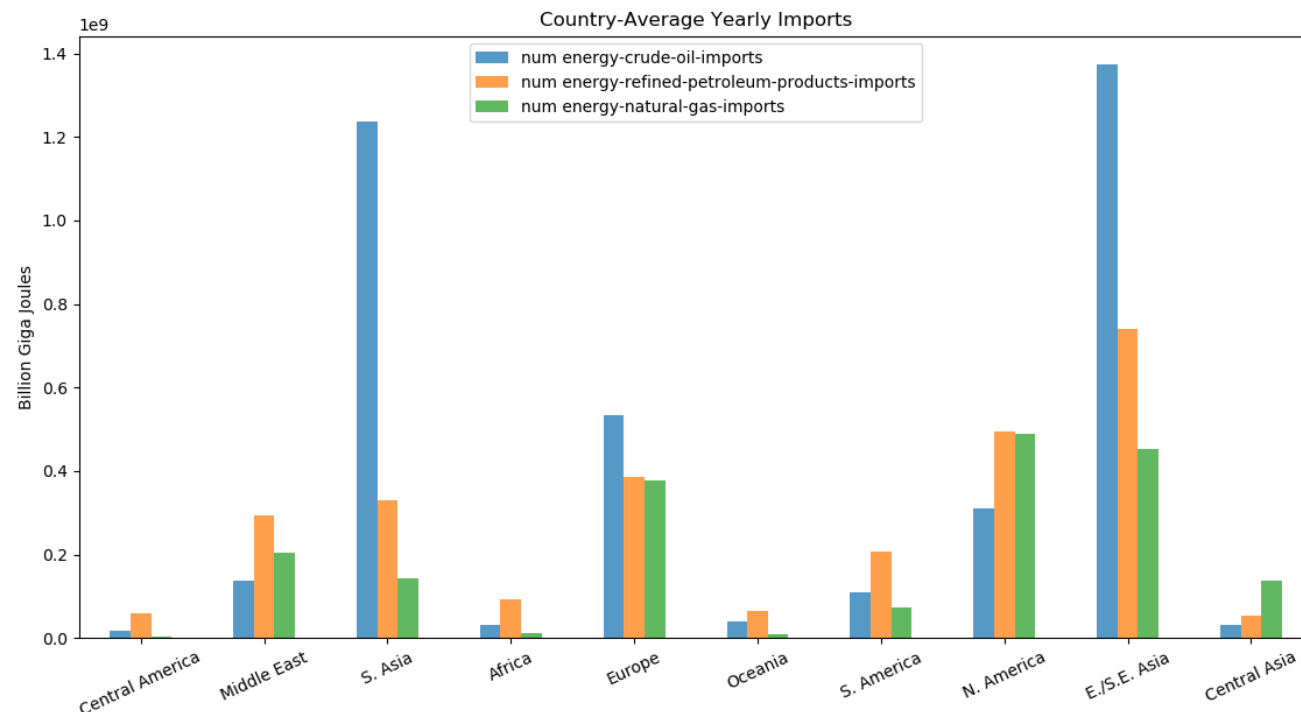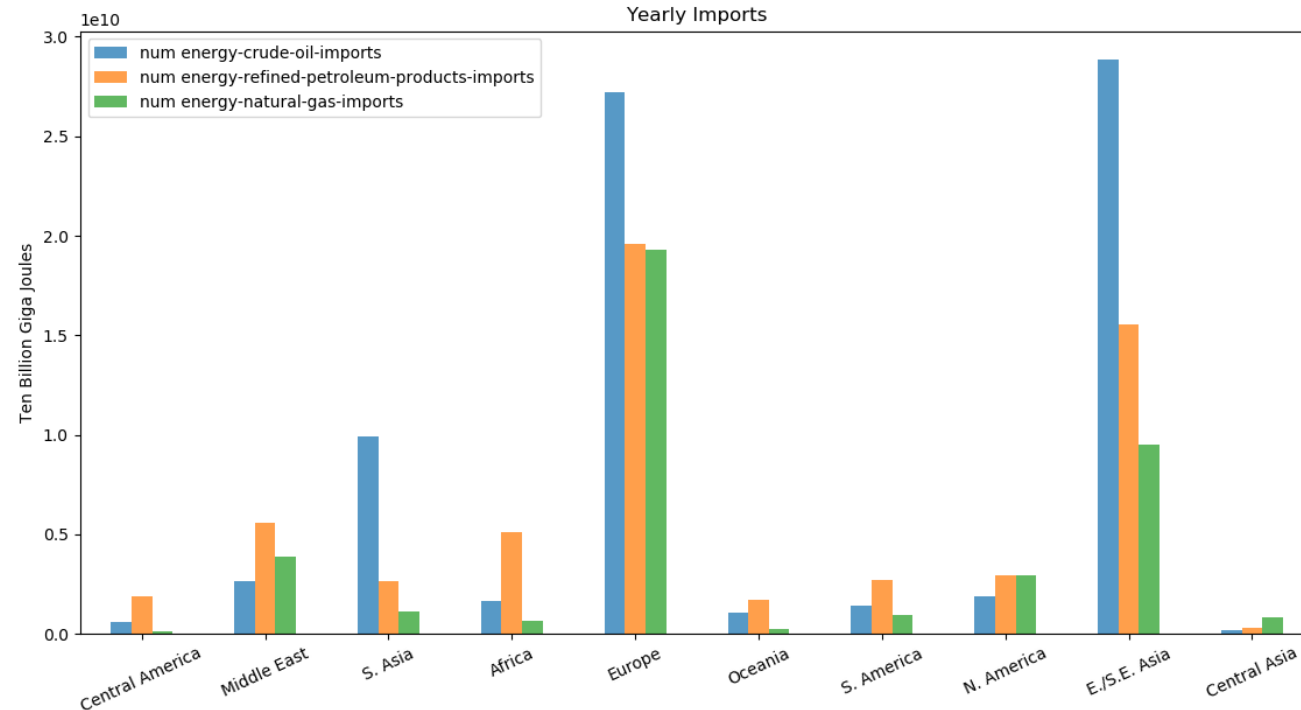
Exports

NOTE: Regional Exports were calculated based on each individual country's exports. Thus, they also reflect intraregional trade.

Middle East is the largest exporter of crude oil and Europe of refined petroleum products which are mostly distributed internally (EU). Africa is also an important exporter of crude oil which moves from poor oil producing countries to the rich West.

Country-averaged data suggest that Central Asian countries are collectively stronger exporters of natural gas and crude oil when compared to the Middle East where resources are more unevenly distributed. The same applies to Europe where refining capabilities are different between the industrialized north and the poorer Balkan countries.

Key Take Away:

-Intraregional trade should be considered when trade is studied. It can provide an index of cooperation.

-The Middle East is by far the biggest exporter of crude oil and central Asia of natural gas (European gas exports refer mostly to internal-EU trade).
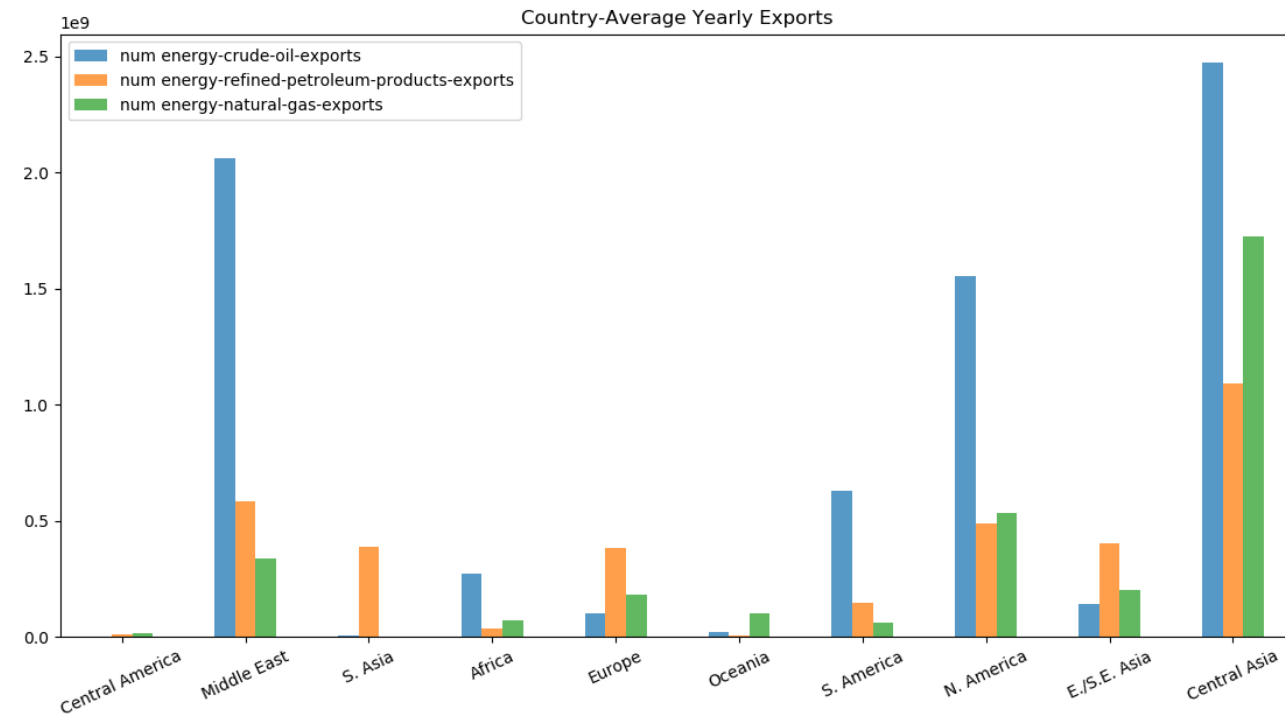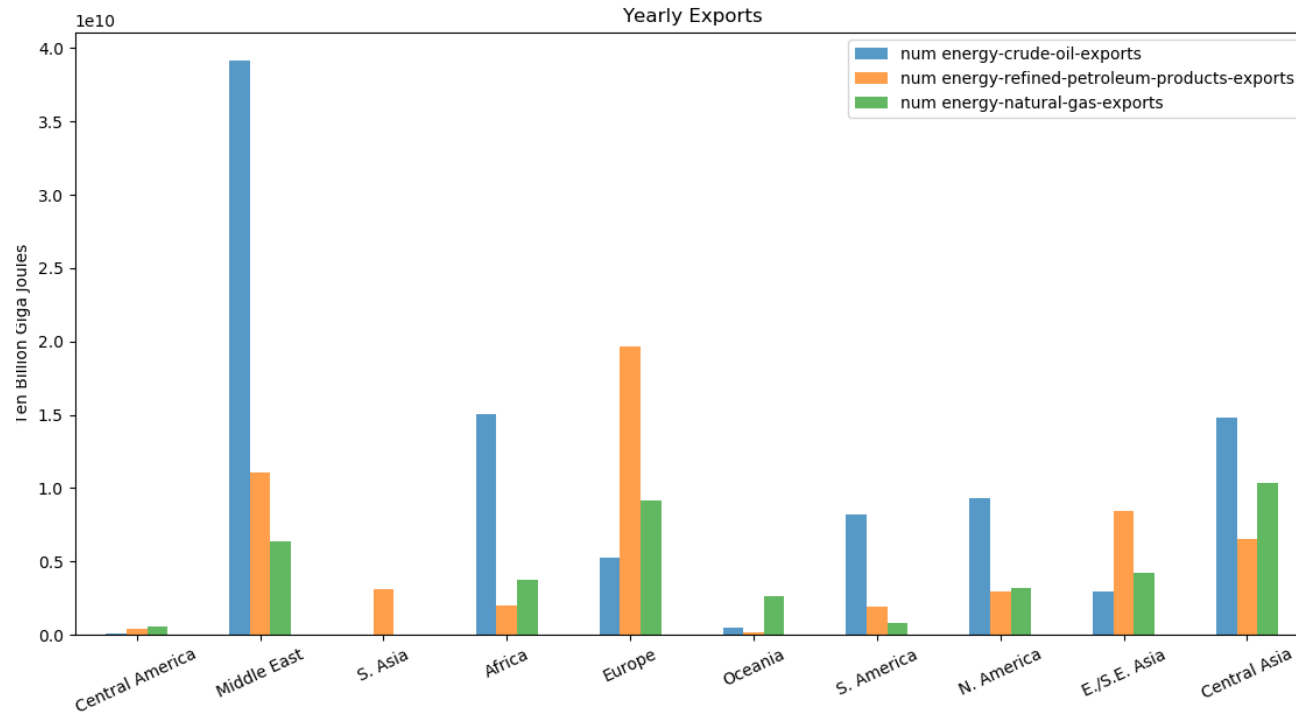
Electricity Production Sources

NOTE: The CIA World Factbook expresses energy production sources as a percentage of total energy production. Special attention should be given to the interpretation of the associated graphs which express **qualitative and not quantitative** data.

Europe is the only continent with considerable amount of nuclear energy. Central Asia, North American, South Asian and Middle Eastern countries also make use of nuclear sources but to a lesser degree (both cumulative and averaged). Fossil fuels remain the predominant source of electricity production in all continents. South Asia and South America both make extensive use of natural gas. Overall, Europe appears to be the most diversified continent and the greatest user of renewable sources. Moreover, the EU is willing to further expand its energy diversification and increase its renewable sources percentage while cutting fossil fuels.

The difference between country-averaged data and region cumulative shows the homogeneity between the countries of each region in regard to energy production. Middle East, South Asia and Central Asia seem to have the lowest homogeneity between their members in terms of energy production sources.

Key Take Away:

-These graphs express qualitative and not quantitative data.

-Europe is the most diversified region in terms of energy production.

-Middle East, South Asia, Central Asia have low homogeneity in their member states' electricity production sources reflecting a plethora of different economies, technological advancement and policies between their member states.

Ports and Terminals
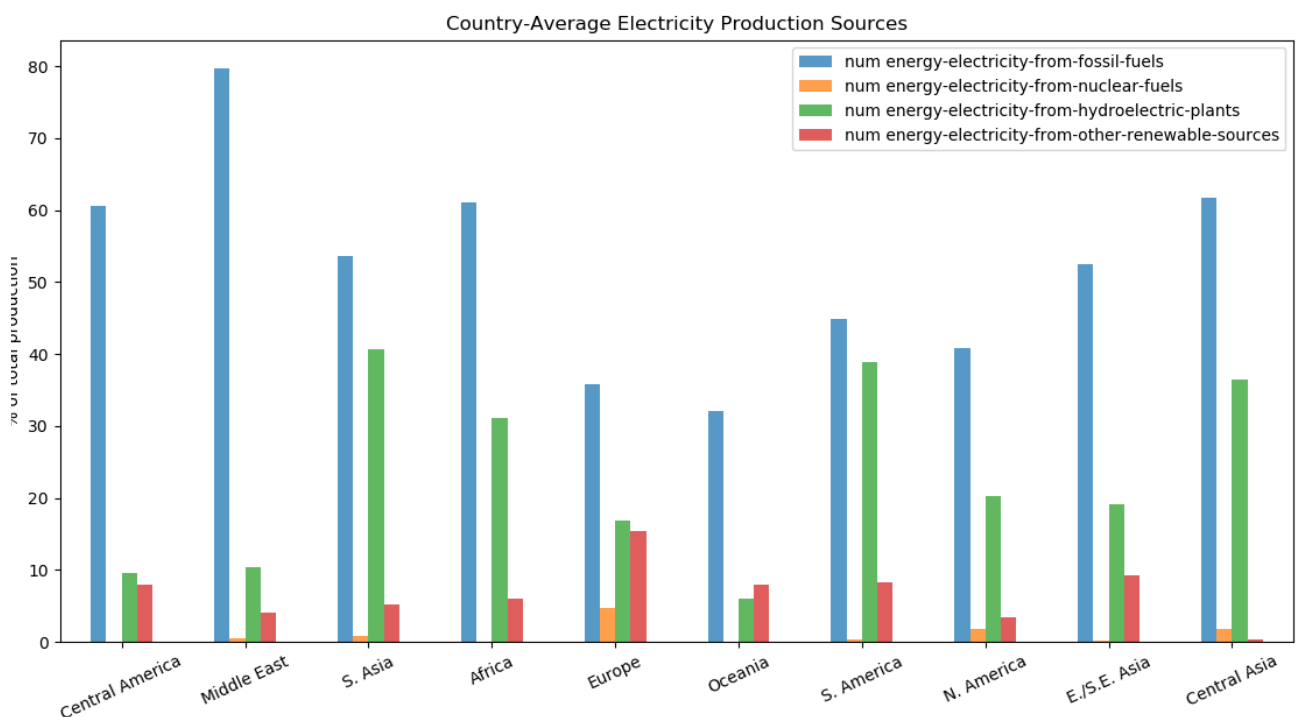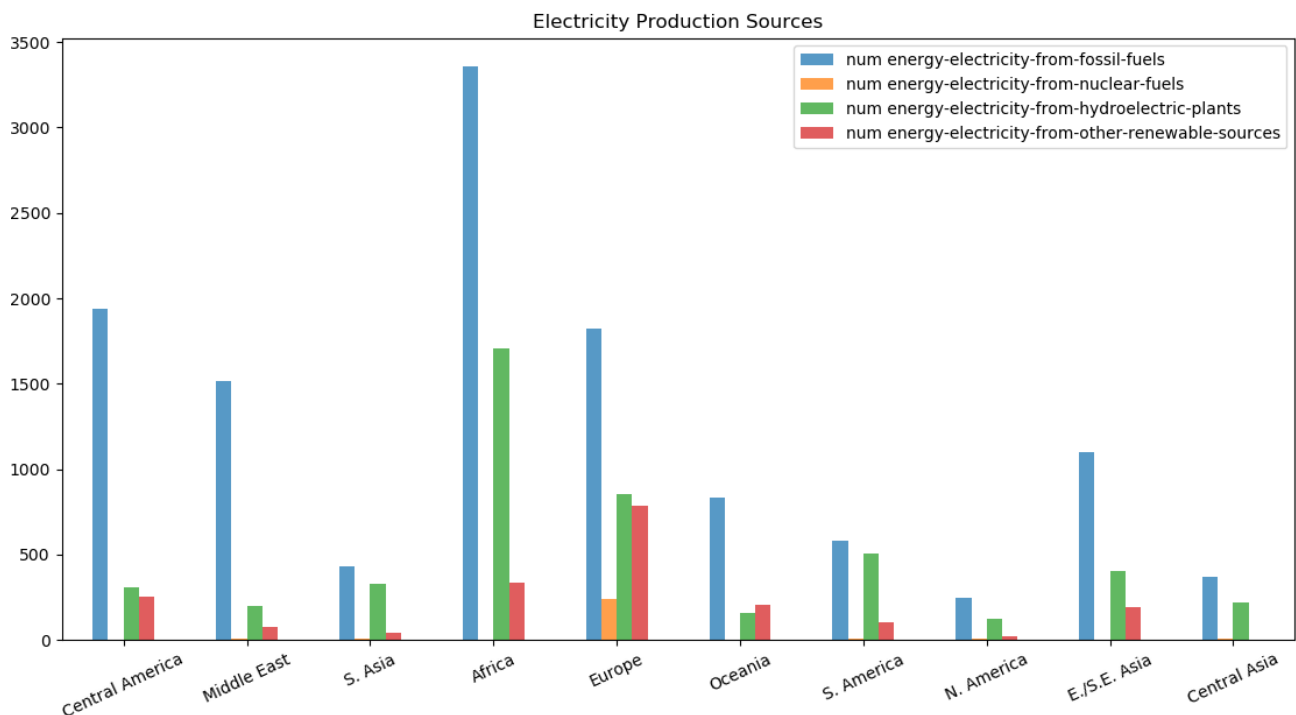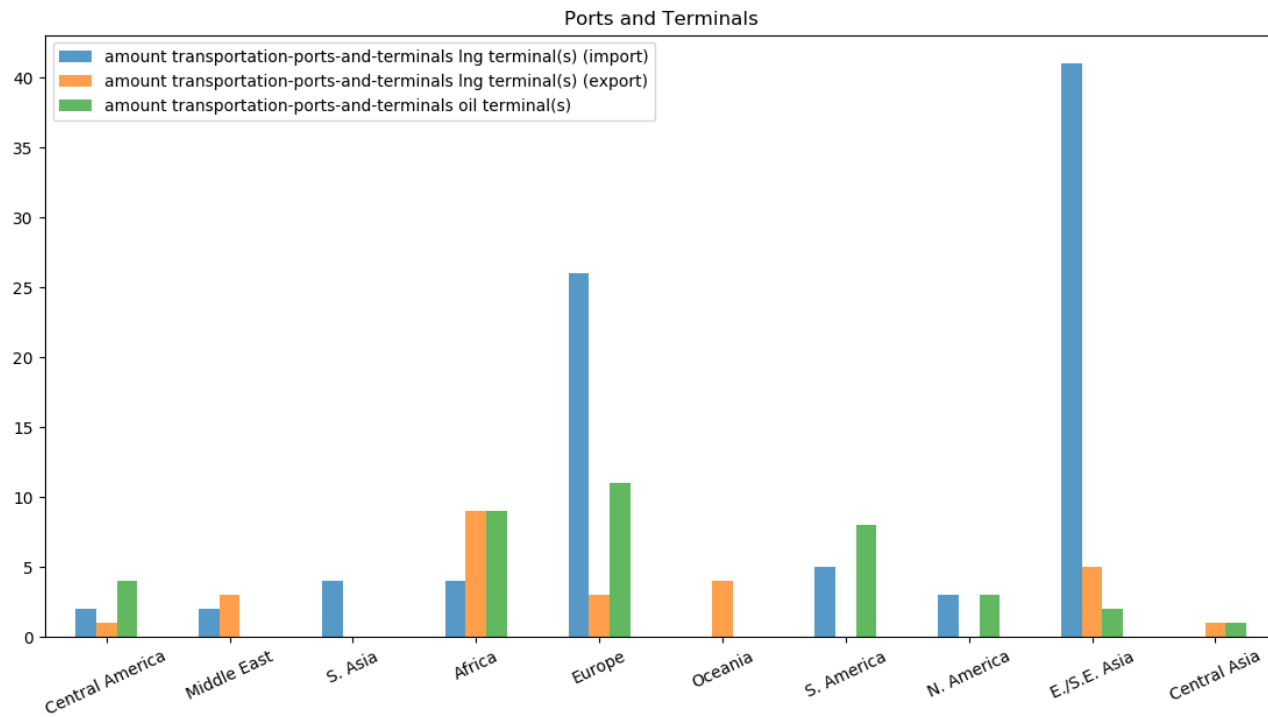
Ports and Terminals

NOTE: The number of ports and terminals is subjective since it does not explain the volume of resources transferred through each. It merely shows the extend of existing infrastructure.

Interestingly, both Europe and East/Southeast Asia are not major gas producers yet they possess the most ports and terminals. This can be attributed to their geographic location which serves as a hub for global trade. The export-oriented economies of Africa (due to low consumption and lack of strong markets) justifies the number of terminals covering both oil and LNG. The Middle East has an extremely small number of terminals due to its geography (pipelines favored), the lack of intraregional trade and also because of its proximity with Europe. American regions are more oil intense.

The country-averaged data do not challenge or change the landscape painted by the cumulative graph.
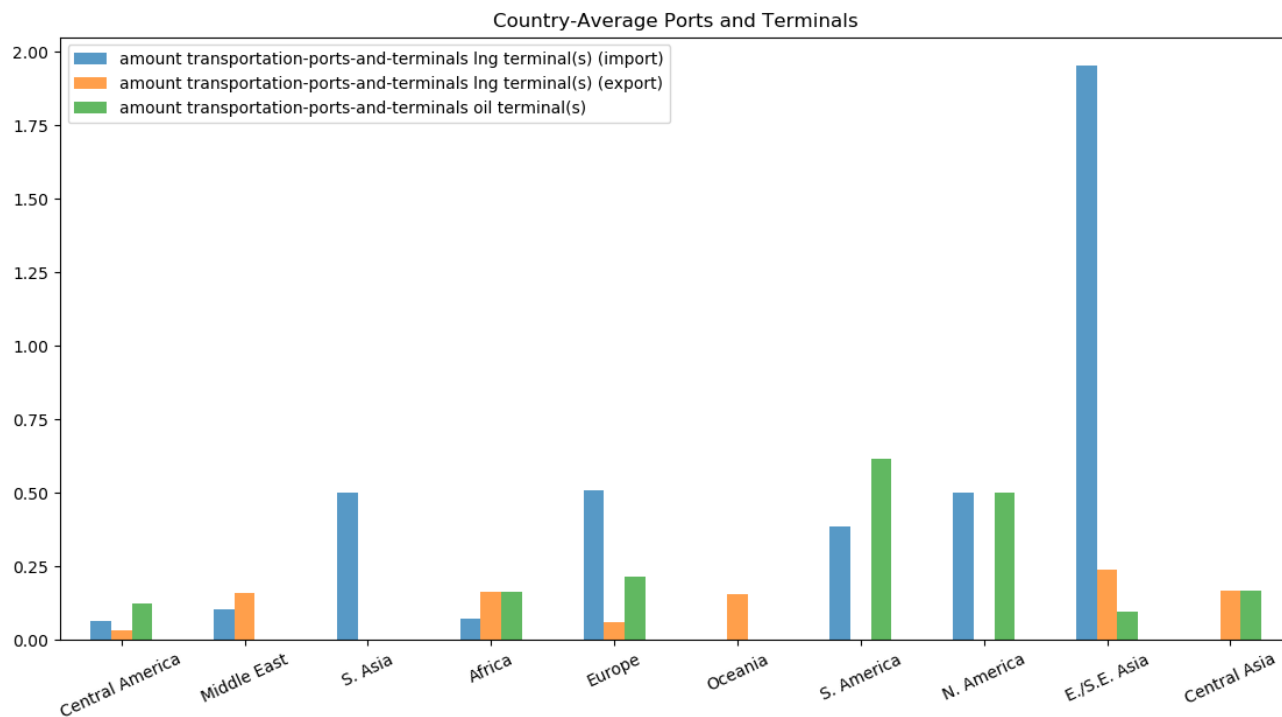
Pipelines

NOTE: Pipeline types are classified differently in various countries. Caution in the interpretation of data is advised. Moreover, pipeline length does not necessarily reflect the volume of resources transferred.

Europe has the lengthiest gas pipeline infrastructure followed by Central Asia. Gas pipelines outnumber oil globally. LPG pipelines are extensive only in North America and in great lengths compared to other pipelines.

The Country-Average data suggest that pipelines in Europe are distributed into few countries (hence the discrepancy between cumulative length and average). This is not the case for Central and North America where pipelines cover most of the countries.
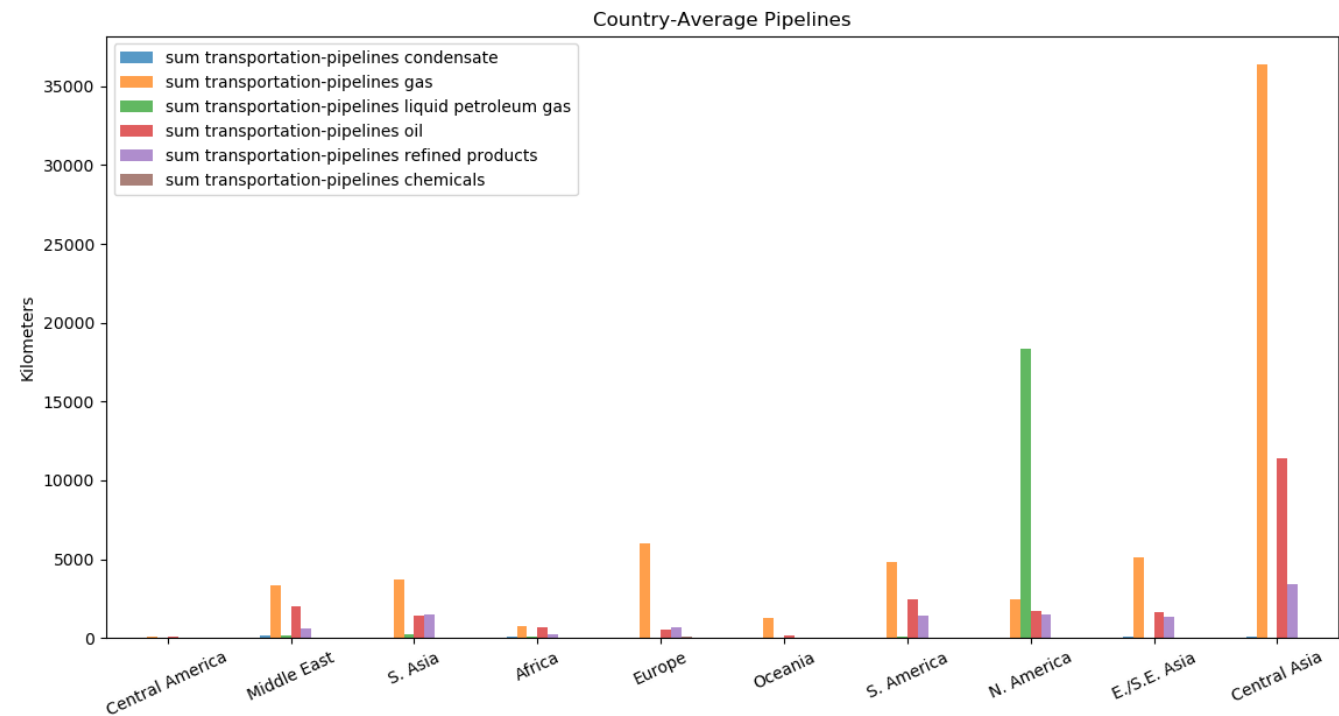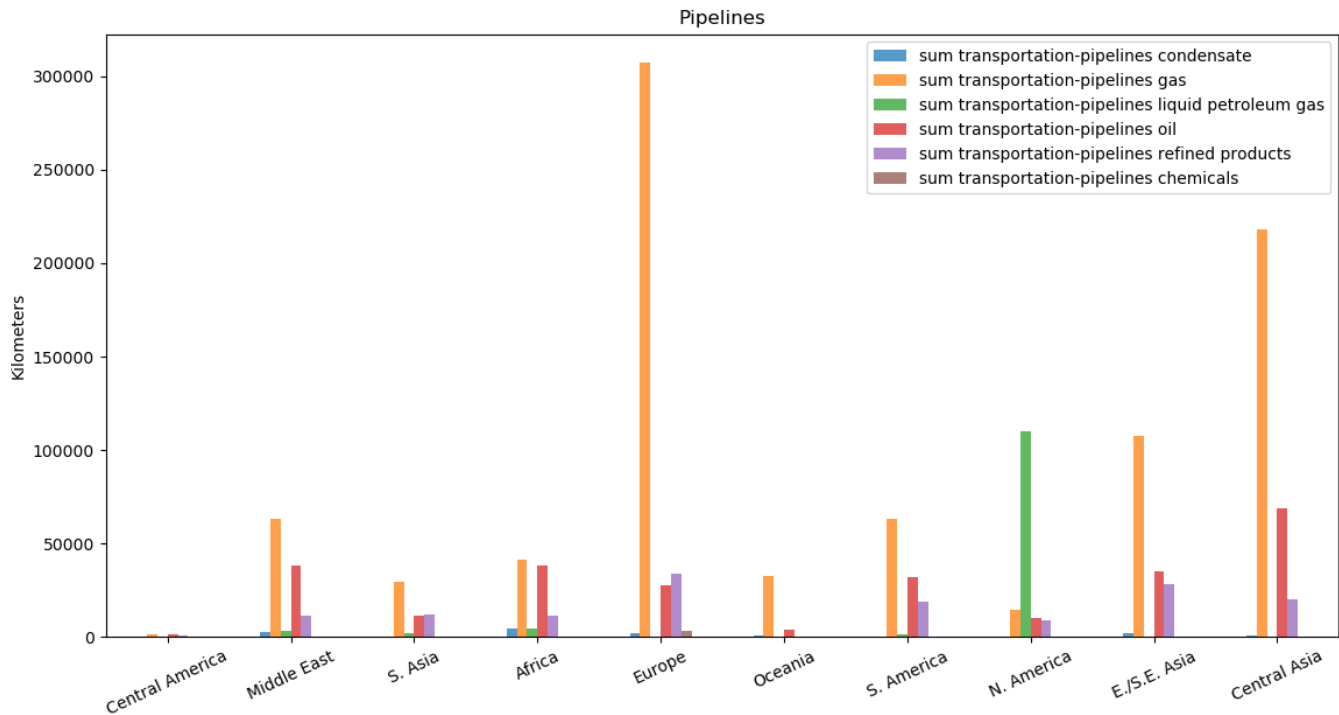
Table 3 - Clustering of Regions (Reserves, Production)

Group (number of regions)	Regions	Comments
A. (6)	Central America, S. Asia, Africa, Oceania, S. America, N. America	-Small to Medium oil reserves / Small to none gas reserves -Small to Medium production of crude oil / Small to Medium production of refined products / Small production of natural gas Generally, oil-oriented regions.
B. (1)	Middle East	-Biggest oil reserves / Biggest gas reserves -Highest production of crude oil / Medium production of refined products / High production of natural gas. Strongest overall reserves, most extensive production, relatively small production refined petroleum products (discussed above in the respective section)
C (2)	Europe, E./S.E. Asia	-Small oil reserves / Small gas reserves -Small production of crude oil / High production of refined products / Small to Medium production of natural gas. Biggest portion of production and reserves is attributed to East/Southeast Asia. Both of the regions are connected by their very high production of refined petroleum products and comparatively medium (for East/Southeast Asia) to small (for Europe) reserves and production. Moreover, considering the country composition of E./S.E. Asia we cannot avoid observing the technological discrepancy between them and of European countries. Does technological advancement and quality of human capital affect refined petroleum product production? Further research is suggested on the matter.
D. (1)	Central Asia	-Small oil reserves / Big gas reserves -Medium production of crude oil / Medium to Small production of refined products / Highest production of natural gas. Considering the size of reserves and the volume of production, it can be assumed that Central Asian countries make full use of their reserves by maintaining medium to high production across all types of energy resources securing big amounts of income. The outcome of such a production-intensive approach should be studied on the long term by considering the rates of production comparatively to the existing proved reserve amounts. Is there an imminent danger of resource depletion? What would be the timeframe?

Table 4 - Clustering of Regions (Reserves, Production, Imports, Exports)

Group (number of regions)	Regions	Comments
A. (6)	Central America, S. Asia, Africa, Oceania, S. America, N. America	All regions in this group are characterized by medium to low reserves, productions, imports and exports.
B. (1)	Middle East	The exact opposite of Europe. Massive reserves and production (except refined petroleum products) and a fully exports-oriented economy.
C (1)	Europe	Probably the most import intense region. Europe imports the biggest amounts of most resources and only produces and exports refined petroleum products, important part of which end up in EU (developed intraregional) markets.
D. (1)	E./S.E. Asia	Produces the highest number of refined products but has to import big percentage of crude oil. Most of the produced refined products are circulated within the region reducing the overall exports to a medium-low level. Similar is the case for natural gas.
E. (1)	Central Asia	The most production intensive region considering its reserves and the relative production. The Middle East may have higher oil production but it also has considerably larger oil reserves. Central Asia is clearly exports-oriented (medium volume of exports but almost zero imports). Aspiring countries pursuing fast economic growth.

Table 5 - Clustering of Regions (Reserves, Production, Imports, Exports, Electricity Production Sources)

Group (number of regions)	Regions	Comments
A. (5)	Central America, S. Asia, Oceania, S. America, N. America	Balanced regions in terms of reserves, production, imports, exports and electricity production sources. These regions cover the whole low to medium spectrum of the aforementioned factors. No outstanding values or extremes.
B. (1)	Middle East	Biggest oil and gas reserves, highest oil production and very large gas production. Totally exports-oriented. Electricity production reliant on fossil fuels with very limited renewable sources, hydroelectric and nuclear power.
C. (1)	Africa	The most heavily reliant region on fossil fuels and hydroelectric power for electricity generation. High import and export rates of same resources probably indicate intensive intraregional trade and/or logistics hub for global trade.
D. (1)	Europe	The biggest user of renewable sources and nuclear power. Important producer of refined petroleum products, commercial hub for all resources with a plethora of ports, terminals and pipelines.
E. (1)	E./S.E. Asia	Electricity production distributed across various sources. Important producer of refined petroleum products. Imports oriented. Similar profile to Europe but clustered separately due to lack of exports.
F. (1)	Central Asia	Primarily centered around the production and export of natural gas along with medium amounts of crude oil. Balanced production of electricity.

Country Clustering

Table 6 - Clustering of Countries (Reserves, Production, Imports, Exports, Electricity Production Sources)

Group (Number of countries)	Regions	Comments (based on descriptive statistics)
A. (159)	Aruba, Antigua And Barbuda, United Arab Emirates, Afghanistan, Algeria, Azerbaijan, Armenia, American Samoa, Argentina, Australia, Austria, Bahrain, Barbados, Botswana, Bermuda, Bahamas, The, Bangladesh, Belize, Bosnia And Herzegovina, Bolivia, Burma, Benin, Belarus, Solomon Islands, Brunei, Cambodia, Chad, Sri Lanka, Congo, Republic Of The, Chile, Cayman Islands, Cameroon, Comoros, Central African Republic, Cuba, Cabo Verde, Cook Islands, Cyprus, Denmark, Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, Ireland, Equatorial Guinea, Estonia, Eritrea, El Salvador, Czechia, Fiji, Falkland Islands (Islas Malvinas), Faroe Islands, French Polynesia, Gambia, The, Gabon, Georgia, Ghana, Gibraltar, Grenada, Greenland, Guam, Greece, Guatemala, Guinea, Guyana, Haiti, Hong Kong, Honduras, Croatia, Hungary, Indonesia, Israel, "Cote D'Ivoire", Iraq, Jamaica, Jordan, Kenya, Korea, North, Kiribati, Kuwait, Kosovo, Kazakhstan, Lebanon, Latvia, Lithuania, Liberia, Libya, Madagascar, Macau, Montserrat, Macedonia, Mali, Morocco, Mauritius, Mauritania, Malta, Oman, Maldives, Mexico, Malaysia, New Caledonia, Niue, Niger, Vanuatu, Nigeria, Nauru, Suriname, Nicaragua, New Zealand, South Sudan, Peru, Pakistan, Poland, Panama, Portugal, Papua New Guinea, Guinea-Bissau, Serbia, Romania, Philippines, Puerto Rico, Rwanda, Saint Pierre And Miquelon, Saint Kitts And Nevis, Seychelles, South Africa, Senegal, Saint Helena, Ascension, And Tristan Da Cunha, Sierra Leone, Somalia, Saint Lucia, Sudan, Syria, Trinidad And Tobago, Thailand, Turks And Caicos Islands, Tonga, Togo, Sao Tome And Principe, Tunisia, Turkey, Taiwan, Turkmenistan, Tanzania, Ukraine, Burkina Faso, Uruguay, Uzbekistan, Saint Vincent And The Grenadines, British Virgin Islands, Vietnam, Virgin Islands, Namibia, West Bank, Samoa, Eswatini, Yemen, Zimbabwe	Countries heavily reliant on fossil fuels for electricity production. Lack of diversity regarding sources. Low to medium production of resources and imports/exports.
B. (22)	Albania, Angola, Brazil, Bhutan, Burundi, Congo, Democratic Republic of The, Colombia, Costa Rica, Ethiopia, Iceland, Kyrgyzstan, Laos, Lesotho, Malawi, Montenegro, Mozambique, Norway, Nepal, Paraguay, Tajikistan, Uganda, Zambia	Extensive use of hydroelectric power. Medium production of resources. Mostly exports-oriented.
C. (8)	Belgium, Bulgaria, Slovakia, Luxembourg, Spain, Sweden, Timor-Leste, United Kingdom	Strong renewable source electricity production sector. Production mostly of refined petroleum products. Imports-centered economies.
D. (1)	Canada	Large crude oil reserves and extensive use of hydroelectric power. Strong exports and production of most resources.
E. (1)	China	The biggest producer of petroleum refined products. A huge manufacturing sector which requires great amounts of imports. Also, considerable amounts of resource production. Electricity comes mainly from fossil fuels.
F. (1)	France	The biggest user of nuclear power for electricity. Heavily imports oriented with small production of resources and few exports.
G. (2)	Germany, Italy	Major natural gas importers (and oil to a lesser degree). Extensive use of renewable sources. Producers and of refined petroleum products which also constitutes their major export.
H. (1)	India	Heavily reliant on fossil fuels and crude oil imports. Produces and exports a considerable amount of refined petroleum products. Generally, imports-

[15]

			oriented economy heavily reliant on crude oil.
I.	**(1)**	Iran	Heavily reliant on fossil fuels for electricity production. Some of the largest crude oil reserves translated into high volumes of production and exports (same applies to gas but at lower levels). Minimal imports.
J.	(1)	Japan	Very large imports of natural gas and medium imports of crude oil. Strong electricity production from fossil fuels. Very high imports combined with minimal exports. Very small production of resources, except refined petroleum products which move to a low-medium level, mostly related to technological goods.
K.	**(1)**	Netherlands	The most imports-exports active country due to its role as a major hub for international maritime trade. Small production of resources, heavy reliance on fossil fuels for electricity production.
L.	(1)	Qatar	Totally reliant on fossil fuels for electricity production. Important production and exports of natural gas and oil. Oil production and exports are comparatively smaller than those of gas. Big reserves of both resources. Absence of imports.
M.	**(1)**	Russia	Balanced electricity production sources. Immense natural gas reserves and small to medium crude oil reserves. Maximum possible production of both resources. The biggest exports of natural gas and refined petroleum products. The intense production of both oil and gas regardless of reserves (oil reserves are not that large to justify such an extensive production) raises serious questions regarding the viability and sustainability of this model on long run. Almost null imports.
N.	(1)	Saudi Arabia	Totally reliant on fossil fuels, the biggest exporter of crude oil. Some of the biggest crude oil reserves. Important exports of refined petroleum products. Small natural gas reserves and production. Zero imports (except few refined petroleum products).
O.	**(1)**	Singapore	Zero reserves and production of gas and oil. Total reliance on fossil fuels for electricity production. These fossil fuel demands combined with the absence of oil production and reserves leads to massive refined petroleum products imports which mostly reflect products of natural gas which are used for electricity generation. Overall, Singapore imports massive amounts of refined petroleum products which uses for electricity generation and further refinement/manufacturing in order to export them again. There is a total lack of reserves, imports, exports and production of oil.
P.	(1)	Venezuela	The largest reserves of crude oil. Production of electricity is distributed between fossil fuels and hydroelectric power. Very Low production and exports of crude oil compared to existing reserves. Political unrest and oil type (high extraction costs) explain the very low production and exports. Almost minimal imports.

*Countries with incomplete data removed.

The study of all data presented above can lead to the following conclusions:

1. Energy, in the context of International Relations studies, is a very complex phenomenon. The understanding and description of the economics and politics of energy not only is data intensive but also requires technical knowledge on resources, mining techniques and relevant costs. Furthermore, national policies and private multi national corporations further perplex the study. Multidisciplinary knowledge is required.
2. The private sector and other institutions work intensively on forecasting the future development of energy related indexes and data with Machine Learning (and Deep Learning). Yet, the academia, especially International Relations, have done little to nothing in using Machine Learning tools to describe and research energy. What could Data Science do for the matter? Can unsupervised machine learning methods accurately describe the phenomenon? The academic and political study of energy has nothing to lose in trying new methods.
3. There are important discrepancies between countries of the same region. This is reflected in most graphs by the total change of landscape between region-cumulative and country-averaged data.
4. The world consists of the following region groups in regards to energy:
    a. Group A (Central America, S. Asia, Oceania, S. America, N. America): Regions which are relatively autonomous. They produce decent amounts of resources without being considered powerhouses. Equilibrium of imports/exports. Probably the most energy secure and stable regions.
    b. Group B (Middle East): Focused on crude oil, still has strong natural gas potential. Biggest reserves, very high production which can be further increased. Lack of internal trade, many global exports. The highest potential for political influencing (already a reality) but also for long term resource production.
    c. Group C (Central Asia): The Middle East of natural gas. Massive natural gas reserves and medium crude oil ones. Very high production of gas and high production of oil. An unsustainable but strong short-term development strategy. Many exports, mostly towards Europe (Russian gas). Need for investment in non-energy sectors. Short term development at the expense of sustainability.
    d. Group D (Europe): Diversified electricity production, biggest importer of energy resources. Minimal reserves and production except refined petroleum products which mostly come from the manufacturing sector and are circulated internally. Global commercial hub with extensive infrastructure (pipelines, ports etc.). High cost of energy and low security. The EU is aware of all these and is increasing investment and reliance on renewable sources.
    e. Group E (East/Southeast Asia): Similar profile to Europe but without exports. Probably the most energy dependent region of the world, combined with relatively low living conditions and lack of regional trade. A track similar to this of Europe would be ideal. Diversification could start with hydroelectric which is cheaper than renewable.
    f. Africa*: Africa proved to be the most versatile region being clustered differently every time new features were added. It presents characteristics from both Group A and Group D but at the same time important differences with both. It has a decent production of resources like Group A but very active imports and exports. At the same time, its intense trade looks similar to this of Europe but there is a great differentiation in the selection of electricity production sources with Africa being extremely fossil-fuel reliant. Very high potential for investment and development. Economies with low demands, central geographic location and variety of countries.
5. The clustering of countries distinguished Canada, China, France, Germany, Italy, India, Iran, Japan, Netherlands, Qatar, Russia, Saudi Arabia, Singapore and Venezuela These countries constitute interesting case studies. Further analysis on them is strongly suggested.

Panagiotis Podiotis is an International Relations (BA) graduate, Hellenic Army reserve officer, Red Cross volunteer and US Consulate General (Thessaloniki, GR) ex-intern. He aims to promote more practical and data-centered approaches for the study and policy implementation within IR. For the accomplishment of this purpose, he has converted the CIA World Factbook in dataset format (https://arxiv.org/abs/2010.05640) in order to offer a pool of data for relative data science projects.

References

CIA. (2019, January 04). *Publications: Download*. Retrieved August 25, 2019, from https://www.cia.gov/library/publications/resources/the-world-factbook/index.html

Eric, J., Travis, O., Pearu, P., & others, a. (2001). SciPy: Open Source Scientific Tools for Python.

Hey, T., Tansley, S., & Tolle, K. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Redmond: Microsoft Research.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering, vol.9, n.3*, pp. 90-95.

JetBrains, s. (2000). *JetBrains: Pycharm*. Retrieved August 29, 2019, from https://www.jetbrains.com/pycharm/

Podiotis, P. (2020, October 3). *Towards International Relations Data Science: Mining the CIA World Factbook.* Retrieved from https://arxiv.org/abs/2010.05640

Python Foundation, S. (2001). *Python*. Retrieved August 20, 2019, from https://www.python.org/

scikit-learn developers. (2020, March 04). *scikit-learn.* Retrieved April 01, 2020, from https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

Shearer, C. (2000). The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, pp. 13-22.

Unver, H. A. (2018). *Computational International Relations. What Can Programming, Coding and Internet Research Do for the Discipline?* Paper.

Wes McKinney and the Pandas Development Team. (2020, March 18). *pandas: powerful Python data analysis Release 1.0.3.* Retrieved February 05, 2020, from Pandas: https://pandas.pydata.org/docs/pandas.pdf