

WORLD ENERGY CONSUMPTION AND POPULATION

Gagana Uday Kumar WOV796

ABOUT THE DATA:

Population Data

World Energy Consumption Data(1995-2021) whole data is in exajoules.

GOAL OF STUDY:

- To see if population and energy consumption are correlated.
- To understand energy consumed per person (renewable/non-renewable) by country.

METHODOLOGY:

- First, Path prompt was created to set library directory. Datasets were placed in "P:/Project".

```
1 Libname Kingg "&path";
```

- Then, a program was created with the above code and placed into a separate "autoexec" process flow to automatically prompt the user to set the library path.

Import and summary of data:

- Imported population data and world energy consumption data.
- Summary statistics output for population data and world energy consumption data tables before cleaning showed country names are not standardized across both tables and it can be found in 'import and summary of data process flow' in egp file.

Data Cleaning process:

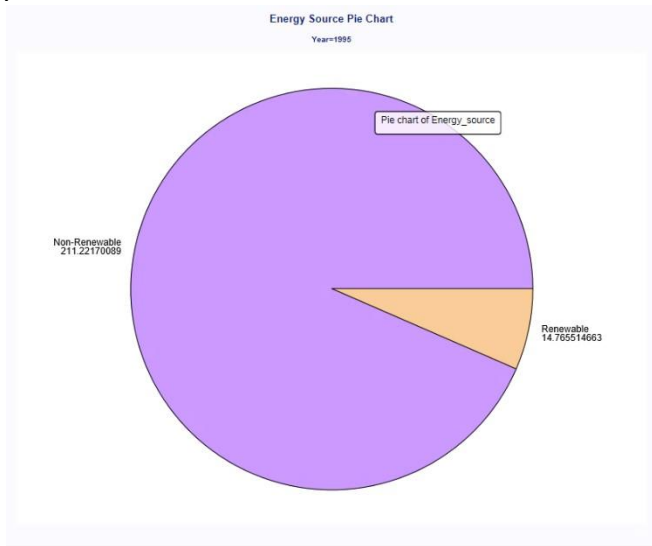
Considering the above issues data was cleaned and following were the changes that were made:

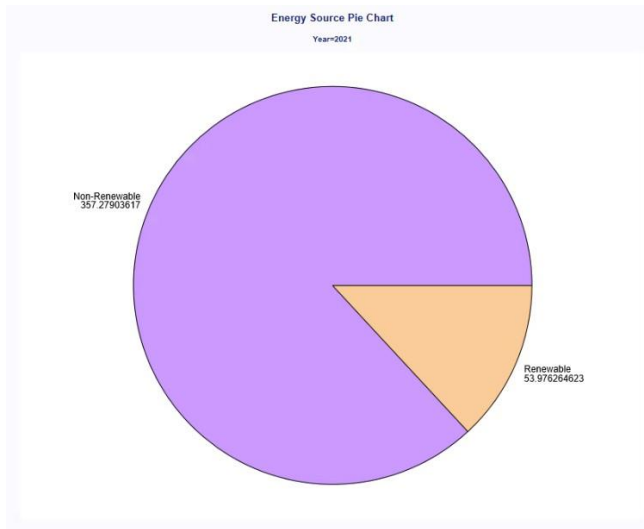
- Created a new table named 'population' by selecting only the required columns and misspelled countries were corrected.
- Selected top 20 most populated countries from 1995-2021 and created a table because world energy consumption data was very large for sas to handle.
- Changed 0 to null because while calculating the average '0' was getting added.
- Changing the country names to maintain the unique key variables in both tables.
- "Nigeria", "Ethiopia", and "Congo, Dem. Rep." were deleted from POPULATION table because their data was missing in world energy consumption table. This left us with total of 17 countries.

- Summary statistics output for POPULATION table and ENERGY_CONSUMPTION_TOP20 table showed that data was cleaned successfully and grouped by years.
- Used functions like **CREATE, UPDATE AND DELETE** it can be seen in the 'data cleaning, verification, summary of data process flow' in egp file.

Advance PROC SQL and Prompt:

- ENERGY_CONSUMPTION_TOP20 table was updated to include new variable "Energy_source." Then according to the energy type, each source was categorized to be either "Renewable" or "Non-Renewable."
- This was merged with the Population table via **inner join** on the common variables "Country" and "Year". The results are grouped by country and ordered in descending order.
- Merged_Population_Energy table was used to create pie charts grouped by year using the **pie chart task** to see the gradual increase of energy consumption from 1995 to 2021. The chart split depending on the Energy_source, "Renewable" or "Nonrenewable" as you see below.





The total energy increased in exajoules from 14.77 to 53.98 can be seen from 1995 to 2021. So, it can be said that we are moving towards a greener and cleaner future.

Then to know the country that uses the maximum amount of energy resources we used the **subquery** and the output shows that “United States” tops the chart this can be found in Advanced Proc SQL process flow in egp file.

As the goal of the study was **to check if population and energy consumption are correlated**, we used **Dual Axis Time Series Graph** like below.

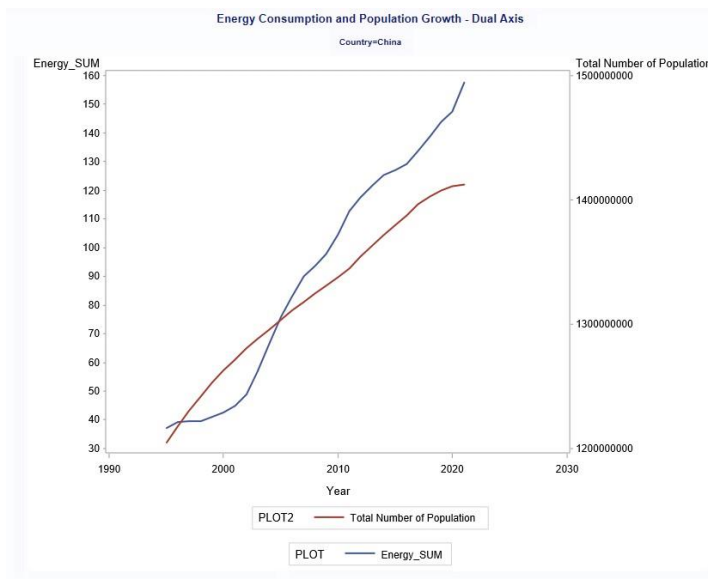


Fig: China graph as example.

From the dual axis time series graphs it was observed that the Population and Energy Consumption were correlated for 13 out of 17 countries which indicates that when we see

change in population similar trend is observed in energy consumption. This accounted for 76.47% of data.

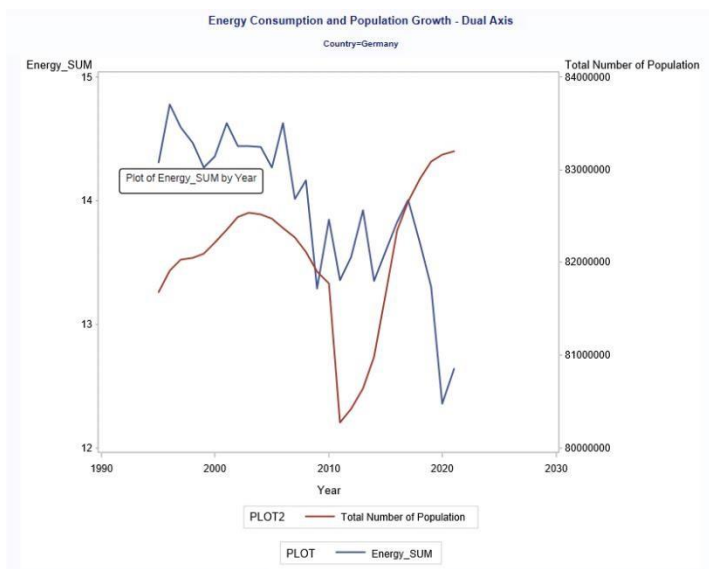


Fig: Germany graph for example

From the dual axis time series graphs it was observed that Population and Energy Consumption were not correlated for 4 out of 17 countries which indicates that when we see change in population opposite trend is observed in energy consumption i.e, when population increase energy consumption decreased or vice versa. This accounted for 23.53% of data.

Later, we were interested in knowing which country consumes the highest amount of energy for each type. Following were the results of **summary table**:

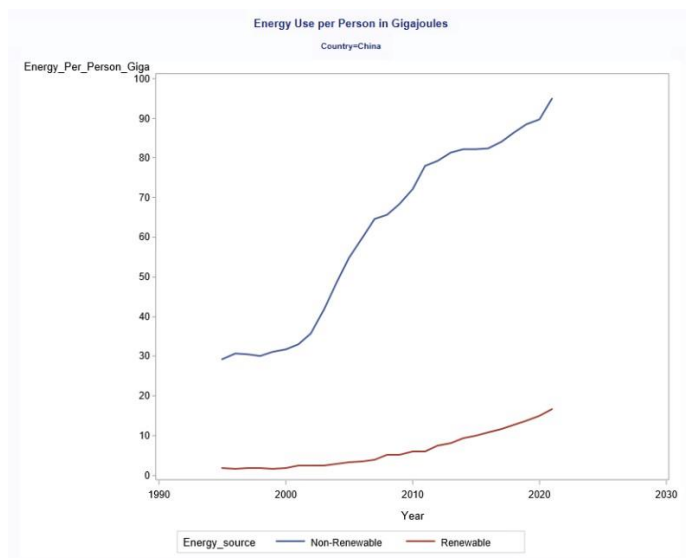
Total Energy Consumption by Country and Type										
	Biofuel	Coal	Hydro-electric	Natural gas	Nuclear energy	Oil	Other renewables	Solar	Wind	Energy Consumption per Country
Bangladesh		1.28	0.21	17.01		5.83	0.00	0.03	0.00	24.35
Brazil	13.17	15.98	94.07	23.07	3.08	113.35	8.52	0.37	3.47	275.08
China	1.52	1617.91	174.45	119.18	28.77	461.27	10.99	11.90	30.71	2456.70
Germany	2.09	88.38	5.54	83.03	34.25	138.10	8.36	4.25	12.50	376.50
India	0.98	314.40	29.71	39.82	6.39	172.26	4.15	2.48	5.73	575.92
Indonesia	1.33	40.82	3.65	37.32		70.92	2.75	0.01	0.02	156.81
Iran		1.59	3.27	123.81	0.46	85.67	0.00	0.02	0.06	214.89
Japan		122.34	22.21	91.91	51.48	260.13	6.70	4.72	0.83	560.32
Mexico	0.12	11.37	8.16	58.47	2.63	98.35	1.91	0.32	1.14	182.45
Pakistan		5.95	7.46	29.65	0.96	23.11	0.11	0.07	0.15	67.46
Philippines		8.87	2.26	2.45		19.33	2.67	0.07	0.08	35.73
Russian Federation		108.90	47.34	394.95	41.91	159.67	0.10	0.07	0.05	752.98
Thailand	0.86	15.19	1.67	34.58		53.56	1.33	0.33	0.14	107.66
Turkey		32.91	12.88	30.16		40.69	0.94	0.43	1.68	119.69
United States	20.71	510.06	76.27	652.38	211.50	991.31	22.11	7.13	28.68	2520.14
Vietnam		18.98	9.97	5.79		17.68	0.02	0.40	0.05	52.88
Energy Consumption per Type	40.78	2914.93	499.10	1743.58	381.42	2711.22	70.66	32.58	85.29	8479.55

From the above table we found that “China” and “US” were the highest energy consumers in the world from (1995 to 2021).

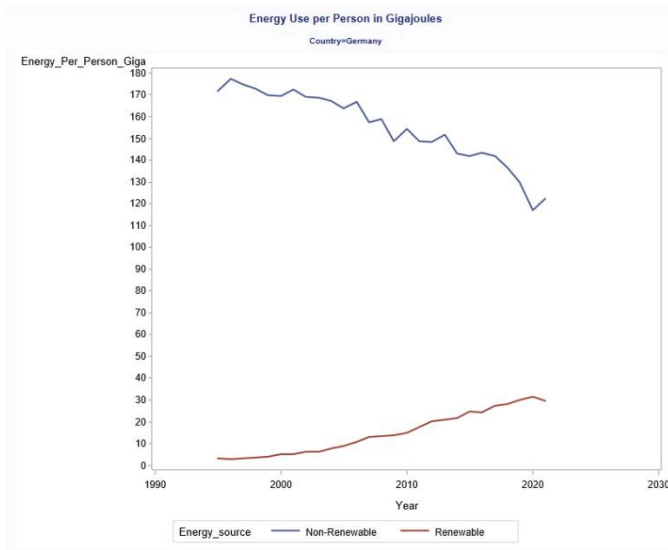
The most energy type used was “Coal” and “Oil” from 1995 to 2021. Highlighted total Figures in exajoules.

Next, we wanted to compare the average of the renewable energy consumption by top countries in all years with the entire data of renewable energy to see which country surpassed this average and in which year using **inline view**. The same process was followed for nonrenewable energy. This can be found in ‘advance PROC SQL and prompt process flow’ in epg file.

Further, to understand energy consumed per person (renewable/non-renewable) by country we created a new table named ‘Total_energy_country_with_source’ to see the per person energy consumed in each year of each Country (converted from Exajoules to Giga joules) by energy source combining the ‘Total_energy_country table’ and ‘Population table’. Following are the line graphs used to visualize the results of above table:



13 out of 17 countries shared this kind of pattern: Increasing non-renewable energy use, while very slow to no-growth in renewable energy use. This accounted for 76.47% of data. Figures were expressed in gigajoules.



4 out of 17 countries shared this kind of pattern: Decreasing non-renewable energy use, while increasing renewable energy use. This accounted for 23.53% of data. Figures were expressed in gigajoules.

Additionally, to this we created 'Year_Range_Prompt' to be used on TOTAL_ENERGY_COUNTRY_WITH_SOURCE table. Output from the query shows Year_Range_Prompt is working correctly by selecting only the years indicated. This can be found in 'advance PROC SQL and prompt process flow' in epg file.

CONCLUSION:

- **To see if population and energy consumption are correlated.**
- 76.47% indicates correlation.
- 23.53% indicates lack of correlation.
- **To understand energy consumed per person (renewable/non-renewable) by country.**
- 76.47% indicates increasing non-renewable energy usage while renewable energy usage runs constant.
- 23.53% indicates declining non-renewable energy usage while renewable energy usage increases.

RECOMMENDATIONS:

- For more in-depth research, all countries can be considered from both datasets instead of just 17 countries.
- New variable for clean energy can be added (renewable energy does not always mean clean energy)
- Climate can also be taken into consideration for increase in energy consumption.

REFERENCES:

- WORLD ENERGY CONSUMPTION:
<https://www.kaggle.com/datasets/nirmalprasad/world-energy-consumption>
- WORLD POPULATION: <https://data.worldbank.org/indicator/SP.POP.TOTL>