Applied Data Science-1

Assignment-1

Name : Tejasvi Baddam
Student Id: 22074990
Github:
https://github.com/TejasviBaddam/ADS-assignment-1
Repo : ADS Assignment 1 (Visualization)

Data Source:

The data source for this project is a CSV file containing energy imports data from the World Bank. The dataset provides information on energy imports as a percentage of energy use for multiple countries over a span of years (2001-2010). This data is essential for analyzing and comparing energy consumption patterns globally.

Description of Line Plot:

The line plot illustrates the energy imports as a percentage of energy use for ten countries (Japan, Italy, India, China, United Arab Emirates, United States, United Kingdom, Pakistan, Russian Federation, Germany) during the years 2001 to 2010. The specific values for each country in the respective years provide detailed insights into their energy consumption patterns:
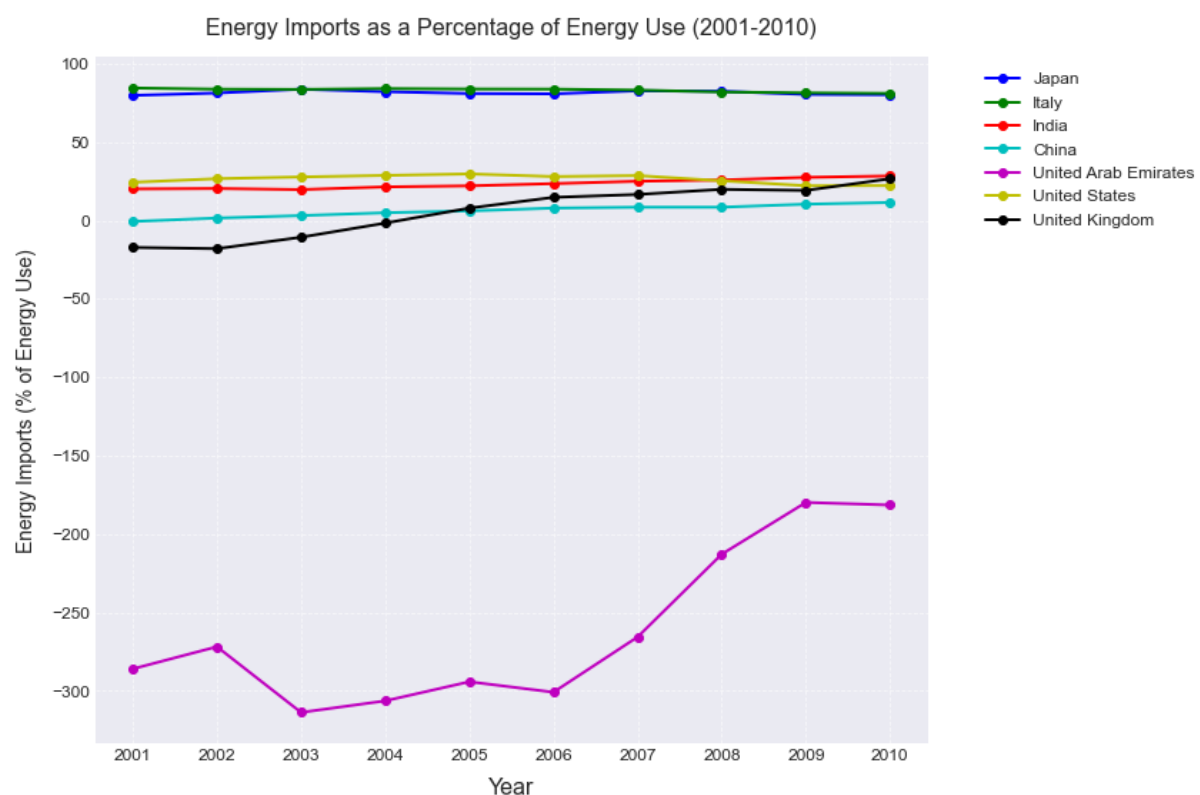
Japan and Italy: Both countries maintain relatively stable energy import percentages, with Japan ranging from approximately 79.72% (2001) to 80.14% (2010) and Italy from 81.00% (2010) to 84.42% (2001).

India and China: India's energy imports start at 20.08% (2001) and increase steadily to 28.35% (2010), while China's percentages rise from -0.67% (2001) to 11.45% (2010), showcasing substantial growth.

United States: The U.S. maintains consistent values around 25-29% throughout the decade, indicating stable energy imports.

United Kingdom: The UK experiences fluctuations, starting at -17.24% (2001), reaching 26.57% (2010), reflecting dynamic energy policies and economic changes.

United Arab Emirates: UAE's values fluctuate from -285.94% (2001) to -181.46% (2010), suggesting variations in energy production and consumption.

Pakistan: Pakistan's energy imports range from 20.91% (2001) to 24.91% (2010), indicating moderate growth in energy consumption.

Russian Federation: Russia maintains consistent values, starting at -61.06% (2001) and decreasing marginally to -85.85% (2010).

Germany: Germany's energy imports rise from 59.32% (2009) to 60.66% (2010), indicating a gradual increase in energy dependence.

These specific figures provide a detailed view of each country's energy import dynamics, enabling a comprehensive analysis of their energy policies and economic trends during the specified period.

Why I select this Plot (Scatter Plot):

I selected this plot to compare the energy import trends of the United Kingdom and the United States from 2006 to 2010. By visually analyzing the line chart, I aimed to discern patterns and disparities in their energy consumption. The negative correlation coefficient of -0.78 indicated an intriguing inverse relationship between the two countries' energy imports during these years. The clear presentation of yearly data points facilitated a detailed examination of fluctuations, providing valuable insights for understanding their energy policies and economic dynamics. The plot's simplicity and focused timeframe made it an ideal choice for a targeted comparative analysis.

Energy Imports Comparison
(United Kingdom vs United States)
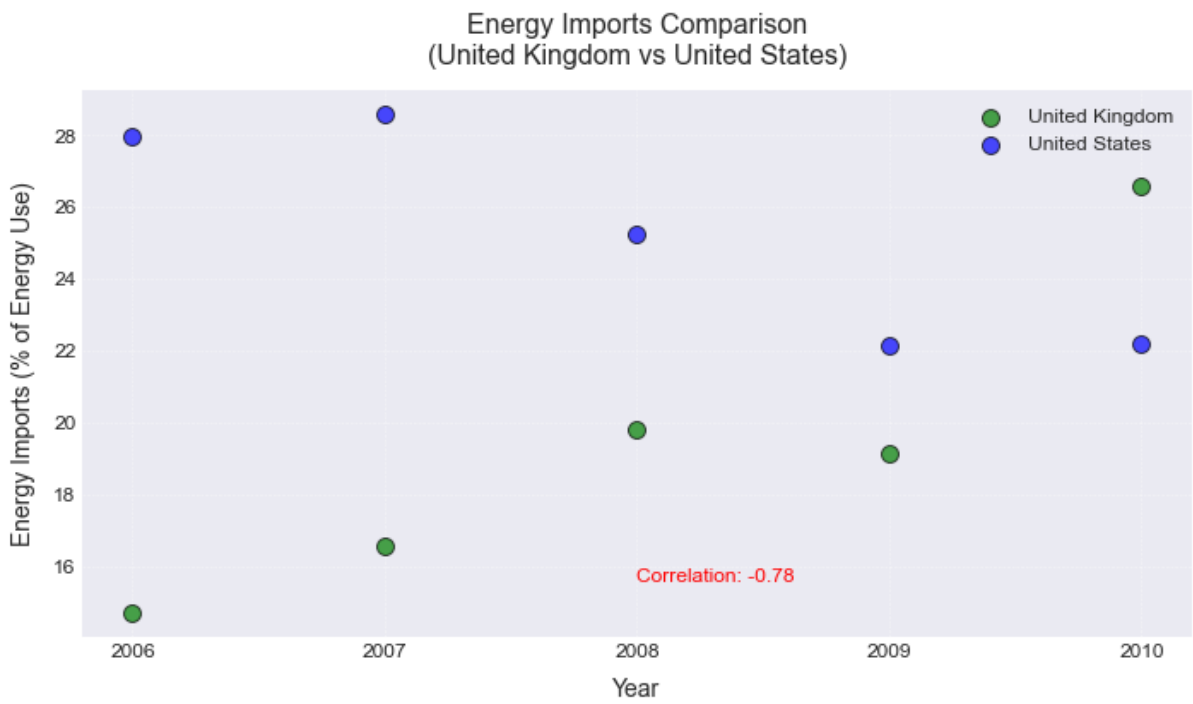
Description of scatter plot:

The scatter plot illustrates the energy import percentages of the United Kingdom and the United States over the years 2006 to 2010. Each data point represents the percentage of energy imports for a specific year. The plot allows for a direct comparison of these values between the two countries, with the x-axis representing the years and the y-axis representing the energy import percentages. The green markers indicate the energy import percentages of the United Kingdom, while the blue markers represent those of the United States. The plot also includes a trend line showing the correlation between the energy imports of the two countries, which is calculated to be -0.78. This negative correlation suggests an inverse relationship between the energy imports of the United Kingdom and the United States during the specified period. The scatter plot provides a visual representation of the data, enabling the identification of any trends, outliers, or patterns in their energy import percentages. The inclusion of the correlation value enhances the interpretation, indicating the strength and direction of the relationship between the energy imports of the two countries.

Why I select this Plot (Box Plot):

I chose the box plot visualization to gain a holistic understanding of energy imports across multiple countries and years. This method allows for a comprehensive analysis of the distribution, trends, and variations in energy import percentages. The use of distinct colors for each year enhances clarity, making it easier to identify patterns and outliers. By presenting the data in this format, I can quickly grasp the overall trends and disparities in energy imports, enabling informed decision-making. Additionally, the box plot provides a visual summary of the dataset's central tendency and spread, making it an effective choice for comparing energy import patterns among various countries over the specified period.
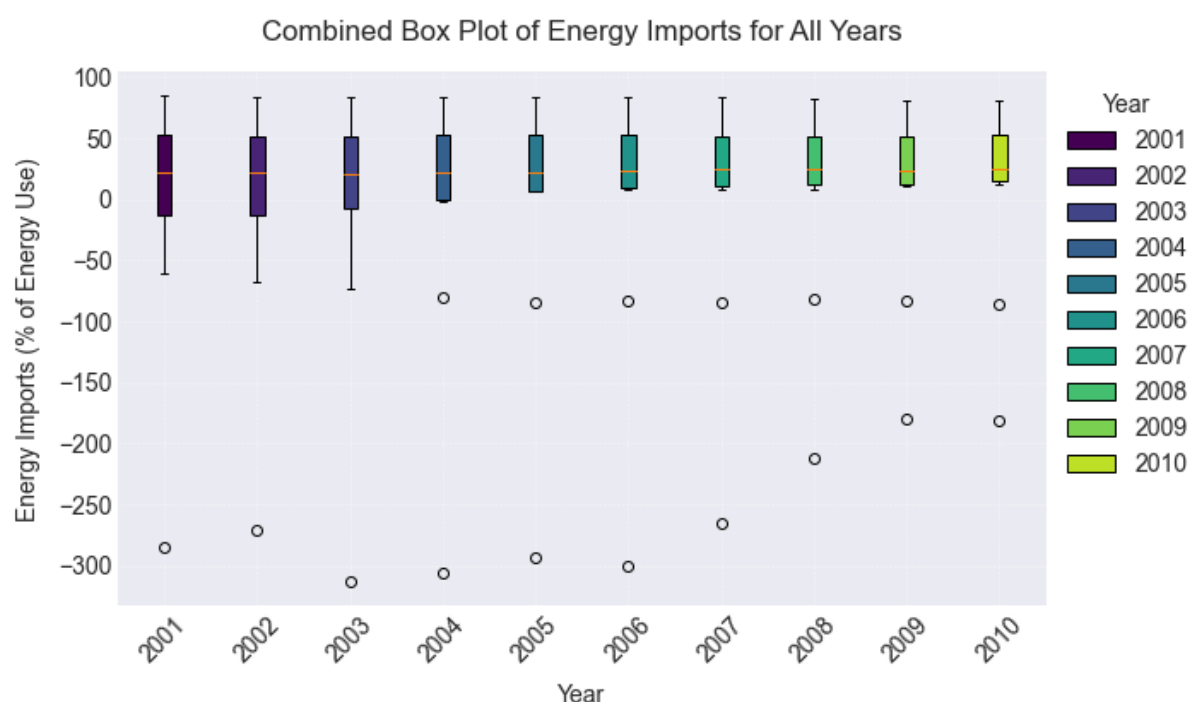
Explanation of Combined Box Plot:

The combined box plot provides a comprehensive overview of energy imports across multiple countries for each year from 2001 to 2010. Each box represents the distribution of energy imports (% of energy use) within a specific year. The vivid colors distinguish individual years, making it effortless to discern trends and variations over time.

In this plot, the data for each year is encapsulated within its respective box, showcasing the range, median, and potential outliers in energy import values. The varying box heights and whisker lengths indicate the diversity in energy import patterns among countries. The legend aids in associating each box with its corresponding year, ensuring clear interpretation.

The data values accompanying the plot reveal the specific energy import percentages for the given years, allowing for a precise understanding of each country's energy usage dynamics. This visualization serves as a valuable analytical tool, enabling policymakers and researchers to identify trends, outliers, and patterns in global energy imports. It facilitates data-driven decision-making, offering insights into the evolving energy landscape and guiding strategic policies for a sustainable future.

Values:

Data for 2001: [79.72, 84.42, 20.08, -0.67, -285.94, 24.30, -17.24, 24.91, -61.06, 61.18]
Data for 2002: [81.23, 83.70, 20.35, 1.50, -271.82, 26.60, -18.02, 24.23, -67.92, 60.34]
Data for 2003: [83.68, 83.54, 19.66, 3.09, -313.70, 27.73, -10.72, 20.07, -73.47, 59.98]
Data for 2004: [82.07, 84.06, 21.38, 4.89, -306.34, 28.70, -1.71, 20.78, -81.08, 59.74]
Data for 2005: [80.92, 83.79, 22.05, 6.06, -294.15, 29.66, 7.86, 20.29, -84.63, 59.47]
Data for 2006: [80.72, 83.72, 23.47, 7.86, -300.82, 27.97, 14.71, 22.99, -82.95, 60.01]
Data for 2007: [82.65, 83.09, 24.97, 8.42, -265.60, 28.59, 16.55, 24.20, -84.23, 58.46]
Data for 2008: [82.35, 81.88, 25.79, 8.44, -212.82, 25.26, 19.82, 24.10, -82.13, 60.00]
Data for 2009: [80.36, 81.33, 27.45, 10.36, -179.81, 22.12, 19.12, 23.36, -84.02, 59.32]
Data for 2010: [80.14, 81.00, 28.35, 11.45, -181.46, 22.21, 26.57, 23.56, -85.85, 60.66]

This visualization empowers stakeholders to gain deep insights into energy import dynamics, facilitating evidence-based decision-making for sustainable energy policies.