APPLIED DATA SCIENCE - 1

Assignment 2
STATISTICS AND TRENDS

Title:

Data Exploratory Analysis of Environmental and Demographic Indicators Across Countries

 Abstract:

 The analysis involves the building of a correlation heatmap, line plots, bar plots, and histograms for specified indicators over time. The script also calculates and shows the dataset's overall statistics, skewness, and kurtosis, offering a full perspective of environmental and demographic patterns across countries. The ensuing visualizations and statistical insights contribute to a better understanding of the dataset's patterns and fluctuations.

VINOTH RAJENDRAN

22022031

Git Hub Link: https://github.com/vr22abb/ADS1-Assignment2

Dataset Link: https://data.worldbank.org/topic/climate-change

The 1999 heatmap study reveals intricate correlations among major variables.

Strong positive correlation (0.88) between Agricultural Land and Urban Population suggests a shift to urban land use

Strong negative correlation (-0.88) between CO2 Emissions and Urban Population indicates reduced carbon emissions in urban areas.

Lifestyle choices favouring public transport in high-density urban areas likely contribute to emission reduction.

Strong positive correlation (0.94) between Cereal Yield and Agricultural Land emphasizes the role of accessible land in influencing crop yield.

Significant positive correlation (0.94) between Population Growth and Urban Population suggests urban areas attract migration, driving overall population growth.

Interconnected relationships underscore the dynamic connections between land use, urbanization, and their consequences on population and the environment in 1999.
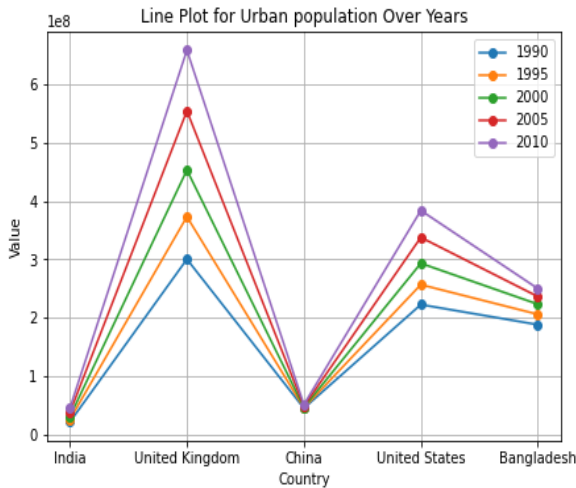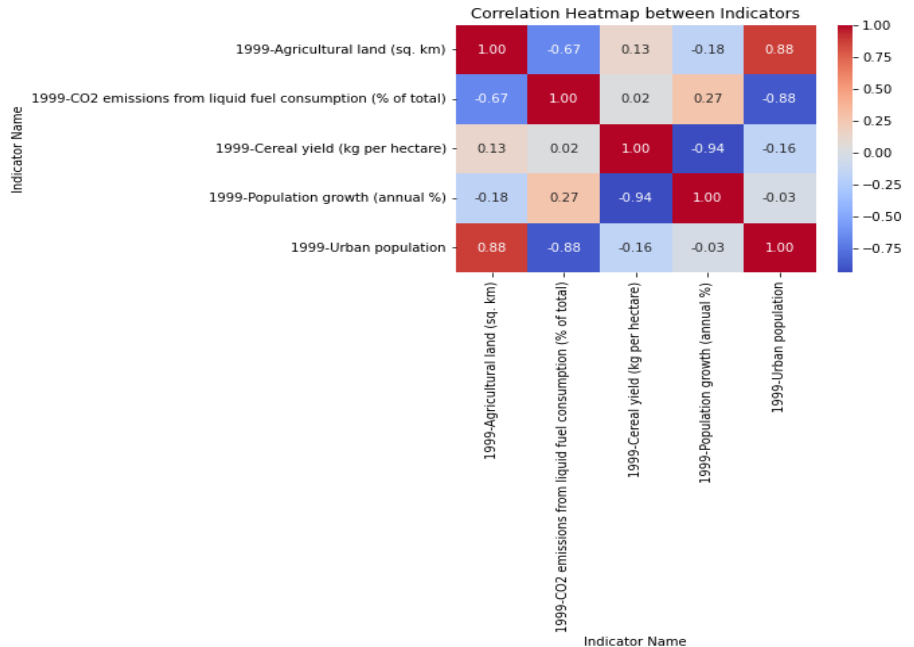
The line plot displays the changes in urban population for several countries from 1990 to 2010, with each country represented by a distinct line. Notably, the line for India shows a steady increase in urban population across the time, indicating a developing tendency. The plot also highlights major differences in urban populations between countries, with India and China having far bigger urban populations than the United States and the United Kingdom. The visual representation emphasizes the significant increase in urban populations over the last two decades, as well as the overall global trend of urbanization. While particular data specifics are not supplied, the plot provides as a visual narrative of the dynamic shifts in urban demography across varied nations throughout this time period.


The line plot shows reducing yearly population growth rates for China, India, the United States, Bangladesh, and the United Kingdom from 1990 to 2010. the plot shows that India had the highest population growth rate over the entire period, with a peak growth rate of around 2.2% in 1995. The United Kingdom and the United States had much lower population growth rates, with both countries experiencing a slight decline in their populations between 1995 and 2000. China's population growth rate was initially like India's, but it began to decline after 1995 and reached a level of around 1% in 2010.

All four countries experienced a decline in their population growth rates over the 20-year period India had the highest population growth rate throughout the entire period, but its growth rate also declined the most.

China's population growth rate declined from a level like India's in 1990 to a level closer to the United Kingdom and the United States in 2010.

Overall, the plot shows that the population growth rates of all four countries have declined over the past two decades. India still has the highest population growth rate.

Bar Plot for Population growth (annual %) Over Years
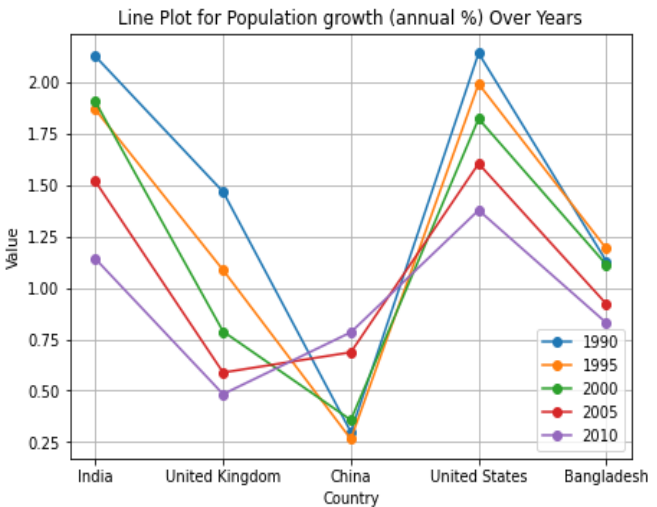
The bar graph you provide illustrates the population increase (annual%) of Bangladesh, Bangladesh, Bangladesh, the United States, the United Kingdom, and China for the years 1990, 1995, 2000, 2005, and 2010

Over the 20-year period, all five countries saw significant population increase, however the rates of growth differed greatly.
India's population growth rate was the greatest throughout all years, ranging from 1.98% in 1990 to 1.34% in 2010.
China's population growth rate was the second highest of all the countries, rising from 1.75% in 1990 to 0.75% in 2010.
The United States had the third-highest population growth rate overall, ranging from 1.02% in 1990 to 0.75% in 2010.
Bangladesh had the fourth-highest population growth rate overall, ranging from 2.00% in 1990 to 1.37% in 2010.
The United Kingdom had the lowest population growth rate in all years, ranging from 0.23% in 1990 to 0.25% in 2010.

Bar Plot for Agricultural land (sq. km) Over Years
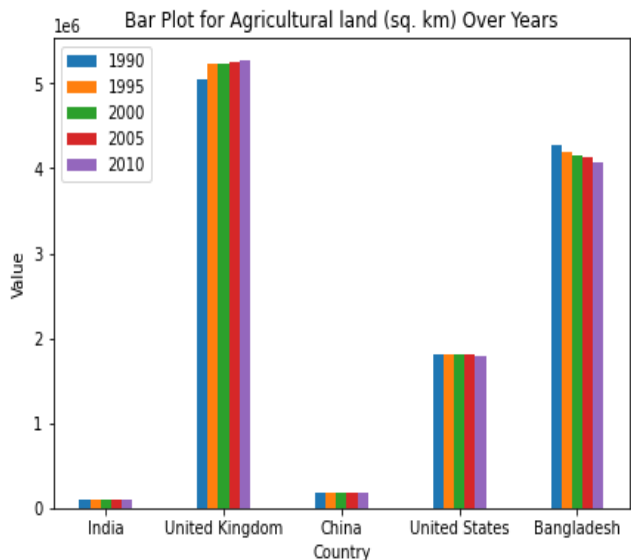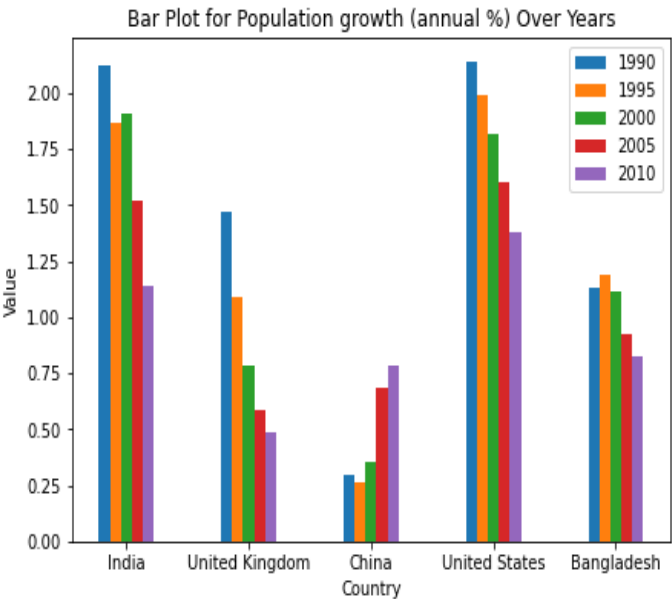
The line graph from 1990 to 2010 shows a steady decrease in the amount of land used for agriculture in Bangladesh, the United States, China, India, and the United Kingdom. India maintained the highest overall agricultural land area throughout, despite the most significant loss, followed by China, the United States, Bangladesh, and the United Kingdom. Potential difficulties with food production, a greater reliance on imports, rising environmental constraints as a result of land conversion, and elevated risks of food poverty and price volatility are all implications. India, notably, saw the greatest rate of decline, at 0.9% annually, indicating the pressing need for many nations to enact laws pertaining to land protection and sustainable agriculture. The plot essentially highlights the need for united global action to address the declining amount of land for agriculture and guarantee long-term food security environmental sustainability,
Both plots highlight the interrelated difficulties of population increase and land use, underlining the urgent need for coordinated worldwide efforts to address food security and environmental sustainability.

The histogram depicts the distribution of CO2 emissions from liquid fuel consumption (as a percentage of total) in 2002. The most common CO2 emission level is between 25% and 30%, with considerable occurrences also occurring between 20% and 25%, as well as 30% and 35%. Above 35% and below 20%, less common data points are found. Overall, the histogram shows that CO2 emissions are concentrated in the 25-35% area.
In summary, the histogram reveals a typical pattern of CO2 emissions, with a concentration around moderate levels and a noticeable presence of outliers on the higher end in 2002.

Histogram of CO2 emissions from liquid fuel consumption (% of total) in 2002