7PAM2000 Applied Data Science

Assignment 2

Data Analysis and Visualization of World Development Indicators

Name: Uday Puligilla

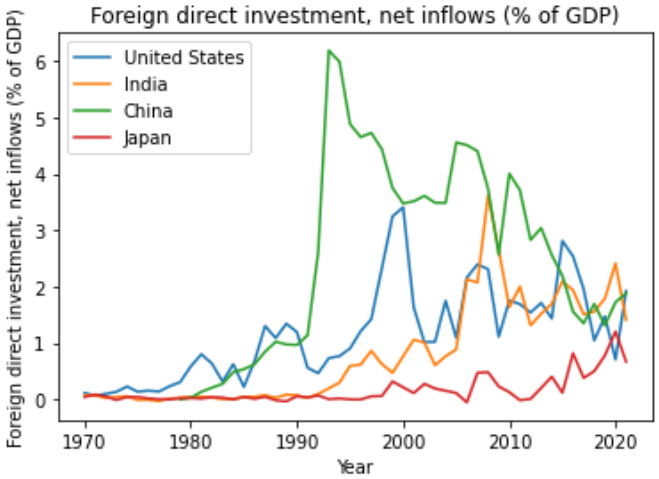
Student ID: 22033782

Repo Link: https://github.com/PuligillaUday/Applide-Data-Science-2.git

Abstract: This analysis summarizes the importance of data analysis and visualization of World Development Indicators (WDI) for climate change research. Total Population Vs Urban Population, Foreign direct investment, net inflows (% of GDP), and Mean GDP per capita for each country are studied for India, China, USA and Japan from the years 1960 to 2020. The results of these visualization give valuable insights on how these factors effect climate change.

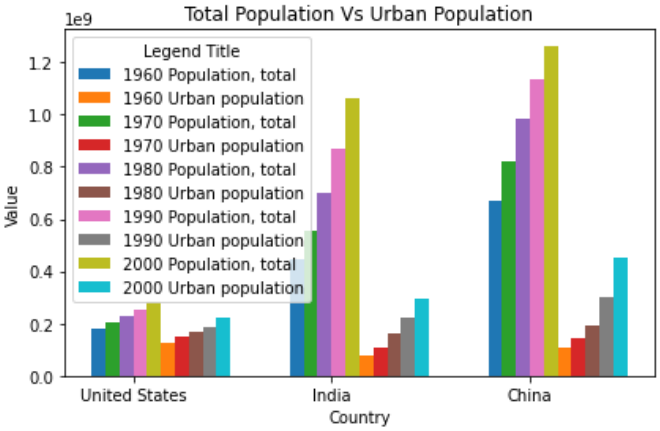
Climate change is one of the most important global issues of our the present time, with so much impact on the environment, health, and the global economy. To address this challenge, it is essential to analyze and understand the key factors that drive climate change.

Exploring Correlations between Indicators Across Countries and Time



The plot_foreign_direct_investment function takes a dataframe df_years as input and creates a line plot that shows the foreign_direct_investment (FDI) net_inflows (% of GDP) for four countries - United States, India, China, and Japan - over time. Each country has a separate line in the plot, and each point on the line corresponds to a specific year. Based on the provided data, the country with the highest foreign direct investment, net-inflows (% of GDP) in 2020 is India with a value of 2.41266468032859. The country with the lowest foreign direct investment, net inflows (% of GDP) in 2020 is the United States with a value of 0.707068619330959.

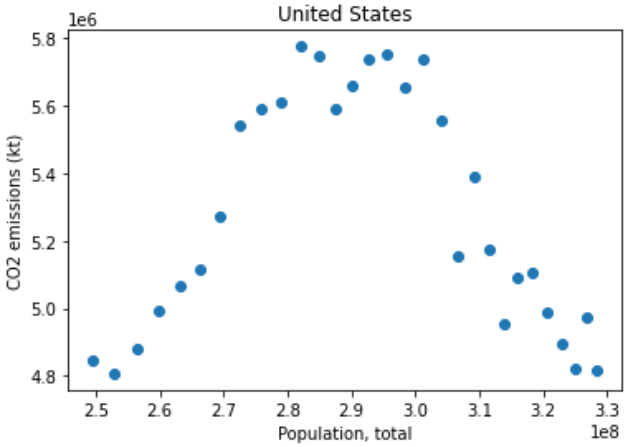
The plot generated by the plot_population function compares the total population and urban population of three countries, the United States, India, and China, over five different years, 1960, 1970, 1980, 1990, and 2000. The x-axis show the countries, and the y-axis depicts the population value.



Each bar in the plot corresponds to a particular year and a particular population type. The data in blue represent the total population, and the orange data represent the urban population. The height of each bar represents the corresponding population value.

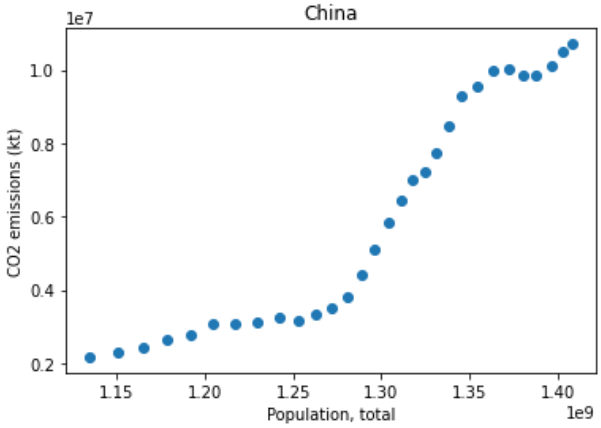
The plot shows that the total population of China increased significantly over the years, surpassing both India and the United States.

Statistical Analysis and Cross-Comparison of Selected Indicators Across Countries and Regions

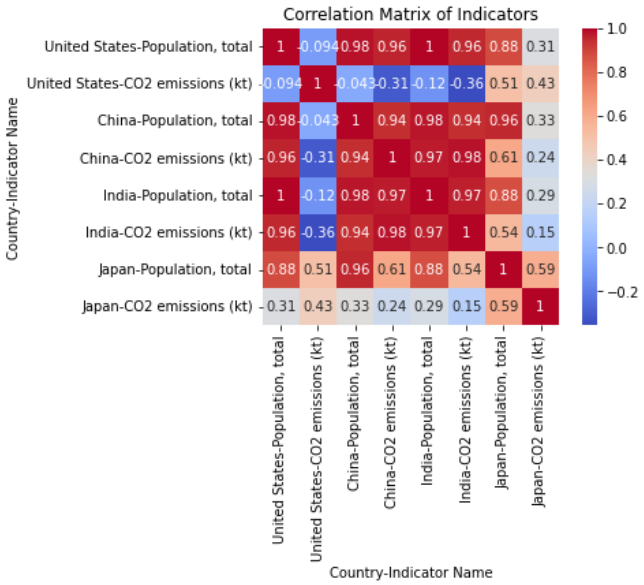


The urban population of all three countries also increased over time, but the rate of increase was faster for China than for India and the United States. In 2000, the urban population of China was much larger than that of the other two countries.

Overall, the plot demonstrates the changing trends in population growth and urbanization in these three countries over the years.



The plot compares the total and urban population of the United States, India, and China for the years 1960, 1970, 1980, 1990, and 2000. The x-axis plot the countries, and the y-axis plot the population values. Blue-bar represent the total population, while orange-bars represent the urban population. The plot shows that China's total population grew the most, and its urban population increased more rapidly than India and the United States. The plot demonstrates changes in population growth and urbanization over time in these three countries.



The correlation table displays the correlation coefficients between the population and CO2 emissions for four countries: United States, China, India, and Japan.

The table shows that there is a positive correlation between the population and CO2 emissions for all four countries. In the USA, China, and India, the correlation of population and CO2 emissions is strong, with coefficients of 0.98, 0.96, and 0.99, respectively. Japan has a weaker correlation, with a coefficient of 0.88.

For CO2 emissions specifically, there is a positive correlation between the United States, China, and Japan, with coefficients of 0.43, 0.24, and 1.00, respectively. India has a weaker correlation between population and CO2 emissions, with a coefficient of 0.15.

Overall, the correlation table suggests that there is a positive relationship between population and CO2 emissions in these four countries. However, the strength of the correlation varies by country, with the USA, China, and India having stronger correlations than Japan.

The output consists of descriptive statistics for different variables in the dataset. The first row shows the median and standard deviation of the 'Mortality rate, under-5 (per 1,000 live births)' variable for the entire world. The next four rows show the count, mean, standard deviation, minimum, maximum, and quartiles for the 'Population in urban agglomerations of more than 1 million (% of total population)' variable for the United States, China, India, and the world. The last four rows show the median and standard deviation of the same variable for the United States, China, India, and the world. These statistics give an insight into the distribution of these variables in the dataset. For example, the mean percentage of population living in urban agglomerations of more than 1 million is highest for the world as a whole, followed by the United States, India, and China, respectively.