DSCI 5360 Data Visualization for Analytics

(Section 005 – Class Nbr 14366, BLB-015)

University of North Texas - Fall 2023

Final Project

Part – IV

Project Team – 7

Team Members

Sai Krishna Pallerla - 11705851

Dhanush Bijavara Mallikarjunappa - 11614243

Jyothi Lohith Kumar Mamidi - 11663097

Nirmala Aiswarya Dosibhatla - 11696616

Sahithi Reddy Bairi - 11723287

Snehalatha Reddy Anna Reddy- 11690996

BEYOND BORDERS: A VISUAL EXPLORATION OF COUNTRY STATISTICS

1. Introduction and Background:

The main driving elements behind the theme " Beyond Borders: A Visual Exploration of Country Statistics" that highlights worldwide insights are as follows, along with a few explanations for our decision:

- **Personal Awareness:** Many people are naturally curious about the globe and desire to keep up with current worldwide affairs to extend their horizons and knowledge.
- **Environmental Awareness:** Environmental challenges, including pollution, biodiversity loss, and climate change, frequently cut across country boundaries. For people concerned in sustainability and conservation, it is essential to understand global environmental trends.
- **Cultural Enrichment:** Exploring global perspectives can help you understand and learn about various cultures, languages, and traditions. This enriches the respective culture. It can promote tolerance and understanding between cultures.

We have been inspired from the research paper “Data Visualization to Explore the Countries Dataset for Pattern Creation” [5] to conduct an in-depth analysis of economic performance, Environmental Impact, Healthcare Accessibility, and price stability [1]. This analysis will offer a comprehensive perspective on the financial stability and growth prospects of each country [4].

The visualizations that need to be addressed are:

1. Which nations emerge as the top performers in GDP sum? What are the top 3 countries?
2. What trends do you observe in the yearly unemployment rates of the USA, China, and Japan? What do you observe from that?
3. What are your predictions for the yearly unemployment rates of the USA, China, and Japan beyond the year 2021?
4. How does the demographic composition, labor force participation rates, and unemployment trends?

5. What trends do you observe in the yearly Non-Renewable Energy production of the USA, China, and Japan? What are your thoughts on it?

6. What trends do you observe in the yearly Renewable Energy production of the USA, China, and Japan? What do you observe from that?

7. What trends do you observe in the yearly Green Energy production of the USA, China, and Japan? What do you observe from that?

8. Find various fuel productions in United States, China, Japan?

This project provides a multitude of advantages for a diverse set of stakeholders:

- **Holistic Understanding**: The study provides a comprehensive grasp of the selected countries' economic and environmental ecosystems. Stakeholders can understand the interconnection of numerous elements influencing a nation's well-being by investigating GDP, unemployment rates, energy production, and tax income.
- **Global Comparisons:** Global comparisons are possible through comparative analysis of the United States, China, and Japan. Understanding how these countries deal with economic issues and environmental concerns might help international collaboration and shared learning.
- **Engagement of Stakeholders:** The visualizations aid in successful communication with a wide range of stakeholders, including politicians, researchers, and members of the general public. Visual representations make difficult material easier to understand, increasing engagement and awareness.
- **Environmental Awareness:** The project develops environmental awareness by emphasizing the consequences of nonrenewable energy generation as well as trends in renewable energy uptake. This understanding is critical for instilling a commitment to sustainable practices and influencing environmental policies.
- **Adaptive Responses to Crises:** Insights on changes in unemployment rates, particularly during crises, assist countries to build adaptive responses. Decision-makers can implement preventative measures to lessen the impact of future economic downturns by analyzing previous patterns.

2. Datasets:

This dataset serves as a valuable resource for gaining a comprehensive understanding of diverse facets of nations across the globe. It enables researchers and analysts to delve into extensive data, facilitating deep insights and allowing for cross-country comparisons.

We discovered this dataset on the Kaggle website, which was collected and compiled by NIDULA ELGIRIYEWITHANA from various data sources before being uploaded to Kaggle [1]. The dataset comprises 195 columns and 35 rows, providing information for all 195 countries worldwide for the year 2023. Following are the variables used in our dataset:

- **Country**: Name of the country.
- **Density (P/Km2)**: Population density measured in persons per square kilometre.
- **Abbreviation**: Abbreviation or code representing the country.
- **Agricultural Land (%)**: Percentage of land area used for agricultural purposes.
- **Land Area (Km2)**: Total land area of the country in square kilometres.
- **Armed Forces Size**: Size of the armed forces in the country.
- **Birth Rate**: Number of births per 1,000 population per year.
- **Calling Code**: International calling code for the country.
- **Capital/Major City**: Name of the capital or major city.
- **CO2 Emissions**: Carbon dioxide emissions in tons.
- **CPI**: Consumer Price Index, a measure of inflation and purchasing power.
- **CPI Change (%)**: Percentage change in the Consumer Price Index compared to the previous year.
- **Currency Code**: Currency code used in the country.
- **Fertility Rate**: Average number of children born to a woman during her lifetime.
- **Forested Area (%)**: Percentage of land area covered by forests.
- **Gasoline Price**: Price of gasoline per Liter in local currency.
- **GDP**: Gross Domestic Product, the total value of goods and services produced in the country.
- **Gross Primary Education Enrolment (%)**: Gross enrolment ratio for primary education.
- **Gross Tertiary Education Enrolment (%)**: Gross enrolment ratio for tertiary education.
- **Infant Mortality**: Number of deaths per 1,000 live births before reaching one year of age.
- **Largest City**: Name of the country's largest city.
- **Life Expectancy**: Average number of years a newborn is expected to live.
- **Maternal Mortality Ratio**: Number of maternal deaths per 100,000 live births.
- **Minimum Wage**: Minimum wage level in local currency.
- **Official Language**: Official language(s) spoken in the country.
- **Out of Pocket Health Expenditure (%)**: Percentage of total health expenditure paid out-of-pocket by individuals.
- **Physicians per Thousand**: Number of physicians per thousand people.
- **Population**: Total population of the country.

- **Population: Labor Force Participation (%)**: Percentage of the population that is part of the labor force.
- **Tax Revenue (%)**: Tax revenue as a percentage of GDP.
- **Total Tax Rate**: Overall tax burden as a percentage of commercial profits.
- **Unemployment Rate**: Percentage of the labor force that is unemployed.
- **Urban Population**: Percentage of the population living in urban areas.
- **Latitude**: Latitude coordinate of the country's location.
- **Longitude**: Longitude coordinate of the country's location.

We also took the aid of another dataset named; World Energy Consumption compiled by Pralabh Poudel. This dataset comprises around 129 columns. Our complete Energy dataset is a collection of key metrics maintained by Our World in Data. It is updated regularly and includes data on energy consumption (primary energy, per capita, and growth rates), energy mix, electricity mix and other relevant metrics. Some of the key features of this dataset are mentioned below:

- **Biofuel consumption:** Primary energy consumption from biofuels, measured in terawatt-hours.
- **Coal production**: Coal production, measured in terawatt-hours.
- **Energy per capita:** Primary energy consumption per capita, measured in kilowatt-hours.
- **Gas production:** Gas production, measured in terawatt-hours.
- **Hydro consumption:** Primary energy consumption from hydropower, measured in terawatt-hours.
- **Nuclear consumption:** Primary energy consumption from nuclear power, measured in terawatt-hours.
- **Oil consumption:** Primary energy consumption from oil, measured in terawatt-hours.
- **Other Renewable consumption:** Primary energy consumption from other renewables, measured in terawatt-hours.
- **Solar consumption:** Primary energy consumption from solar, measured in terawatt-hours.
- **Wind consumption:** Primary energy consumption from wind, measured in terawatt-hours.

This comprehensive dataset offers a wealth of insights concerning countries across the globe, encompassing a wide range of indicators and attributes. With the help of these dataset attributes, we can generate graphs for different countries.

3. Data Stories:

Visualization-1: SYMBOL MAP

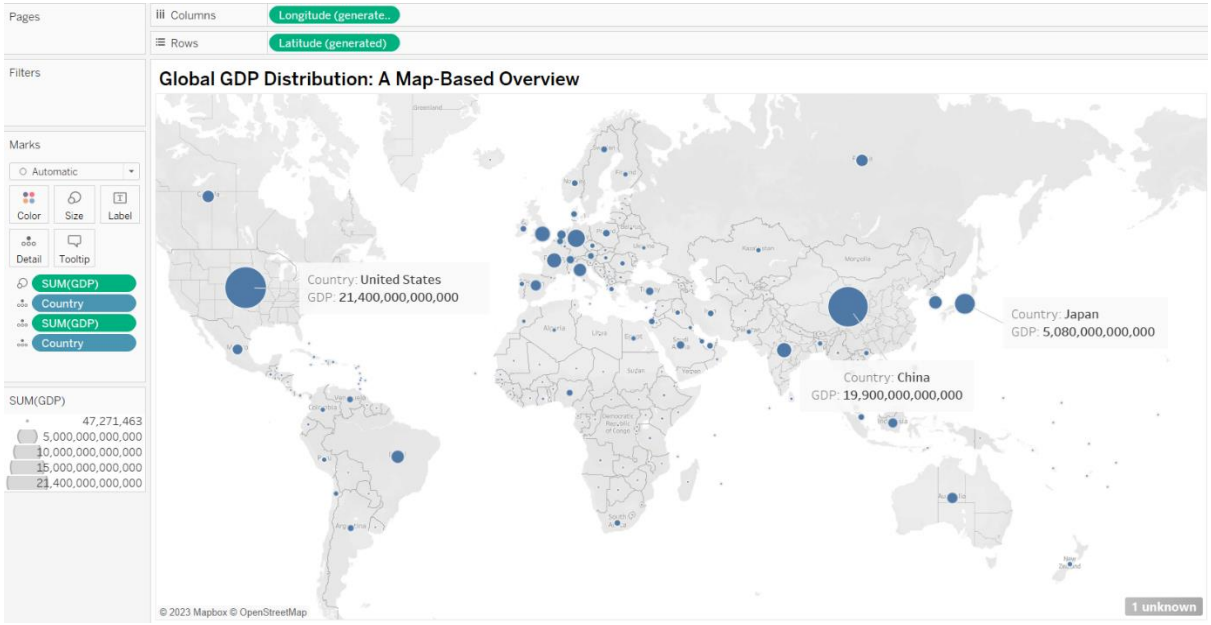


Figure 1

Above visualization (Figure 1) showcases the countries with the highest GDP. Here we can observe that top 3 countries with highest GDP is United States, China, Japan. Each data point is labeled with the country's name, and the size of the circles indicates the sum of GDP, making it easy to identify countries with higher GDP. A bigger size means higher GDPs, while smaller ones indicate lower GDPs. When you glance at the map, it's evident that the United States leads the world in GDP, followed closely by China and Japan. Other countries boasting substantial GDPs include Germany, India, and the United Kingdom. On the flip side, countries with smaller symbols and lower GDPs are mainly situated in Africa and Asia. This global GDP distribution symbol map is a valuable tool for getting a visual handle on how different countries are progressing economically worldwide. It helps spot countries with both high and low GDPs and keeps tabs on how GDP changes over time.

Principle:

SPATIAL CONTEXT ENHANCES UNDERSTANDING:

Longitude and latitude maps provide a spatial context that improves our knowledge of global economic dynamics. The idea is that geographical placement allows viewers to associate economic strength with certain places, making it easier to spot and analyze patterns. This visualization technique uses spatial awareness to identify concentrations of economic power, allowing for a more intuitive understanding of GDP distribution.

Visualization-2: STACKED BAR CHART

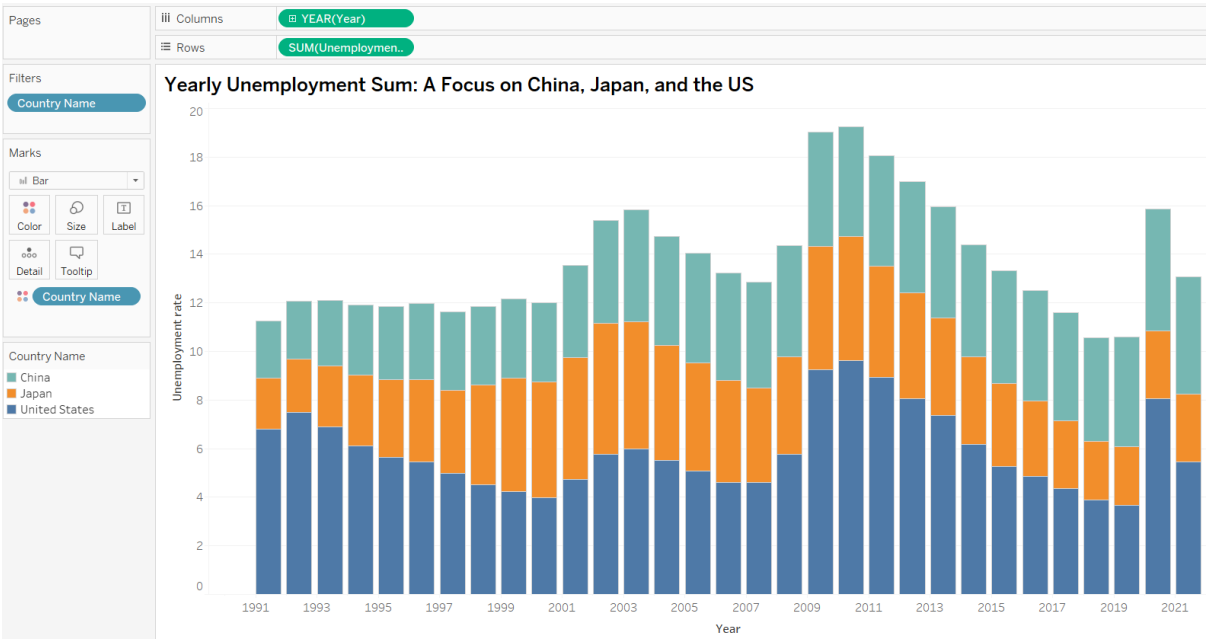


Figure 2

Above Stacked bar chart (Figure 2) visualizes the dynamics of the unemployment rate over time for the selected countries (United States, China, Japan). The cumulative impact of many factors contributing to changes in unemployment rates is represented by this chart in a stacked manner. Here we can observe that unemployment rates have increased significantly in 2009 and 2020. Wonder what happened in that Years? Here is what we should we know the high unemployment rate in 2009 was primarily a consequence of the global financial crisis that began in 2007 and intensified in 2008. Key factors contributing to high unemployment during this period include:

Financial Market Collapse, Global Recession, Business Failures and Layoffs, Housing Market Collapse, Consumer and Business Confidence The high unemployment rate in 2020 was largely a result of the COVID-19 pandemic, which had widespread economic implications.

Principle:

TEMPORAL LAYERING

The stacked bar chart visualization illustrating the unemployment rate over time employs the temporal layering technique. It produces a detailed, layered portrayal of the labor market's dynamics by stacking several elements that contribute to unemployment. This theory enables viewers to understand not just the overall trends in unemployment rates, but also the specific contributions of many circumstances. The layered layers provide an in-depth look at the varied nature of unemployment trends across time.

Visualization-3: TIMELINE SERIES INCLUDING FORECASTING:

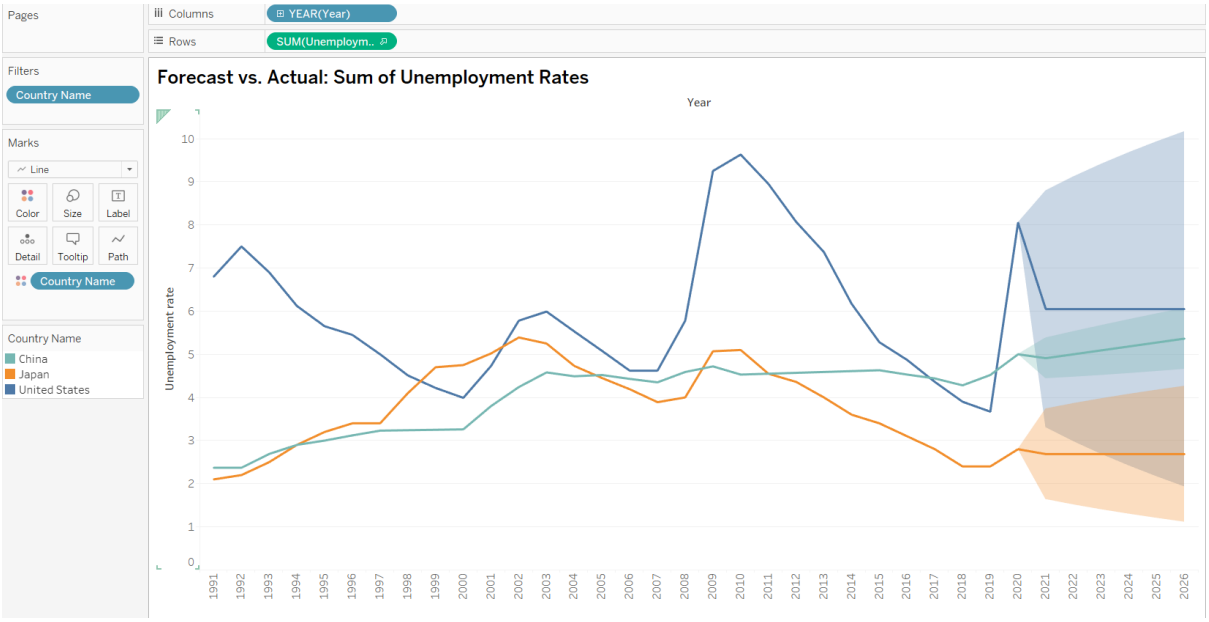


Figure 3

Above line viz (Figure 3) presents a comprehensive view of unemployment rates and its forecast for various countries. This chart presents the unemployment rate trends of Japan (red line), China (green line), and the United States (blue line) from 1991 to 2026. The y-axis displays the unemployment rate. Japan takes the lead with the lowest and most consistent unemployment rate among the trio, showcasing an adept use of its labor market. In China, things are a bit like a roller coaster. The country has a strong economy, but every now and then,

there are bumps in the road with more people being without jobs. This shows that even in a thriving economy, it can be tricky to make sure everyone has a job all the time. The United States displays a commendable trend, steadily reducing unemployment, painting a promising outlook for the future.

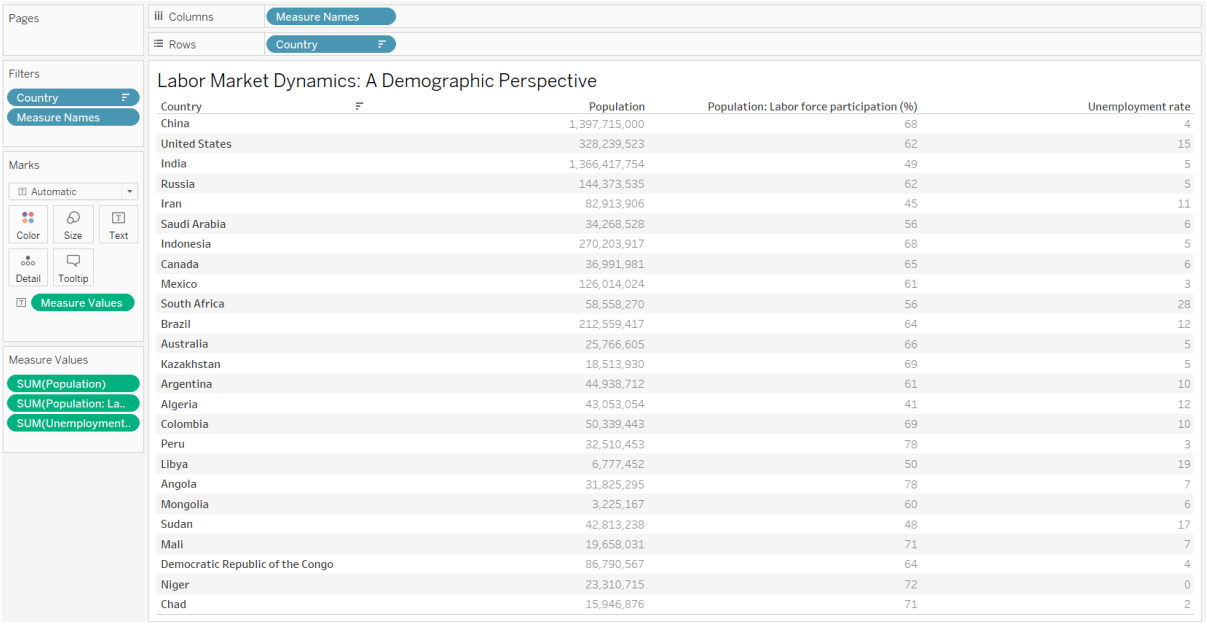
Overall, this visualization reveals differing unemployment trends among these major economies. While Japan maintains stability, China grapples with fluctuations, and the U.S. shows a positive trajectory.

Principle:

FORECASTING FOR INFORMED DECISION-MAKING

The aim here is to use forecasting to make more informed decisions. By broadening the visualization beyond historical data, stakeholders may forecast likely future patterns in unemployment rates. This forecasting principle provides decision-makers with insights into possible difficulties and possibilities, allowing for proactive policy adjustments and resource allocation.

Visualization-4: TABLE CHART



Country	Population	Population: Labor force participation (%)	Unemployment rate
China	1,397,715,000	68	4
United States	328,239,523	62	15
India	1,366,417,754	49	5
Russia	144,373,535	62	5
Iran	82,913,906	45	11
Saudi Arabia	34,268,528	56	6
Indonesia	270,203,917	68	5
Canada	36,991,981	65	6
Mexico	126,014,024	61	3
South Africa	58,558,270	56	28
Brazil	212,559,417	64	12
Australia	25,766,605	66	5
Kazakhstan	18,513,930	69	5
Argentina	44,938,712	61	10
Algeria	43,053,054	41	12
Colombia	50,339,443	69	10
Peru	32,510,453	78	3
Libya	6,777,452	50	19
Angola	31,825,295	78	7
Mongolia	3,225,167	60	6
Sudan	42,813,238	48	17
Mali	19,658,031	71	7
Democratic Republic of the Congo	86,790,567	64	4
Niger	23,310,715	72	0
Chad	15,946,876	71	2

Figure 4

The table chart describes about the complex interplay of population size, labor force participation, and unemployment rates in 25 of the world's 195 countries. The percentage of the population that is actively employed provides extra information about worker involvement. A strong and efficient labor market may be reflected in high labor-force participation rates.

Principle:

DEMOGRAPHIC SYNTHESIS FOR HOLISTIC INSIGHT

The premise underlying the visualization "Labor Market Dynamics: A Demographic Perspective" is "Demographic Synthesis for Holistic Insight." This principle recognizes the inherent link between population dynamics and labor market changes and underlines the necessity to combine demographic data for a more thorough understanding of workforce issues.

Visualization-5: BOX & WHISKER PLOT

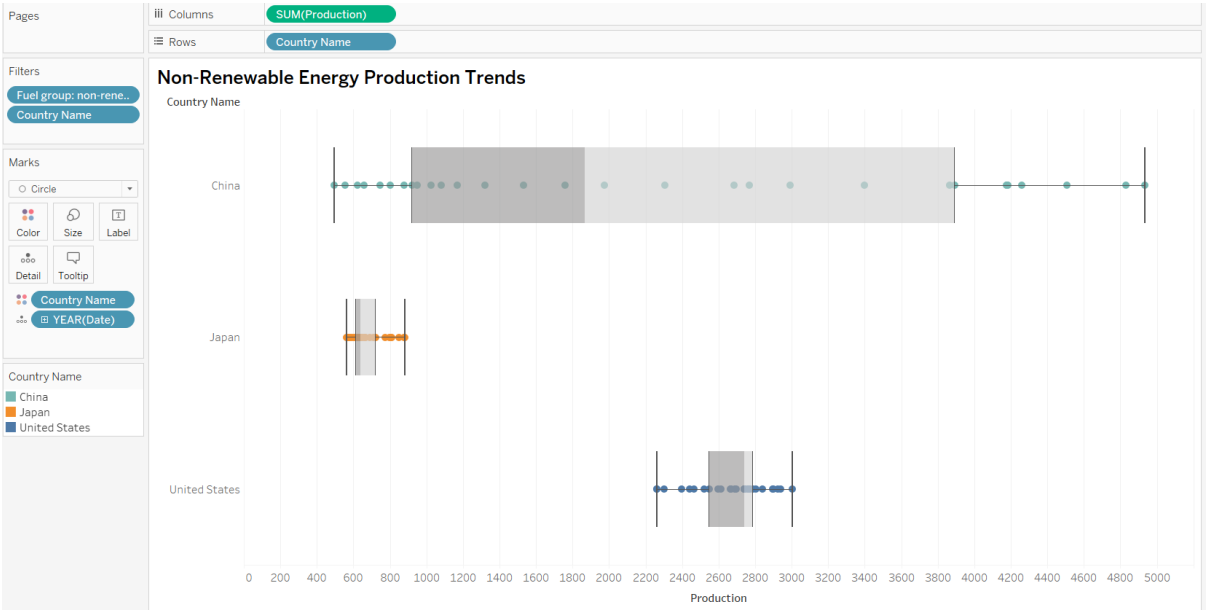


Figure 5

Here the visualized box and whisker plot investigates the sum of non-renewable energy production across the countries—United States, China, and Japan. This visualization aimed to discern patterns, variations, and outliers in the production of non-renewable energy sources over time. Here we can observe that China's massive non-renewable energy output in 2021, there is a major change in environmental sustainability and energy policy which leads to

environmental damages. Luxian Earthquake is the example for this. We can also observe Japan uses very less fuel resources than the other both. United States also decreasing its usage of fuel resources from 2007.

Principle:

DISTRIBUTION ANALYSIS FOR VARIABILITY UNDERSTANDING

The box and whisker graphic for the sum of non-renewable energy output follows the distribution analysis concept. This visualization style enables viewers to comprehend the fluctuation in nonrenewable energy output across time. Examining the distribution of data points and detecting outliers provides a full picture of production trends and emphasizes unusual years, driving conversations on environmental sustainability and policy improvements.

Visualization-6: BAR CHART USING INSERT SHEETS FEATURE ON TOOLTIP:

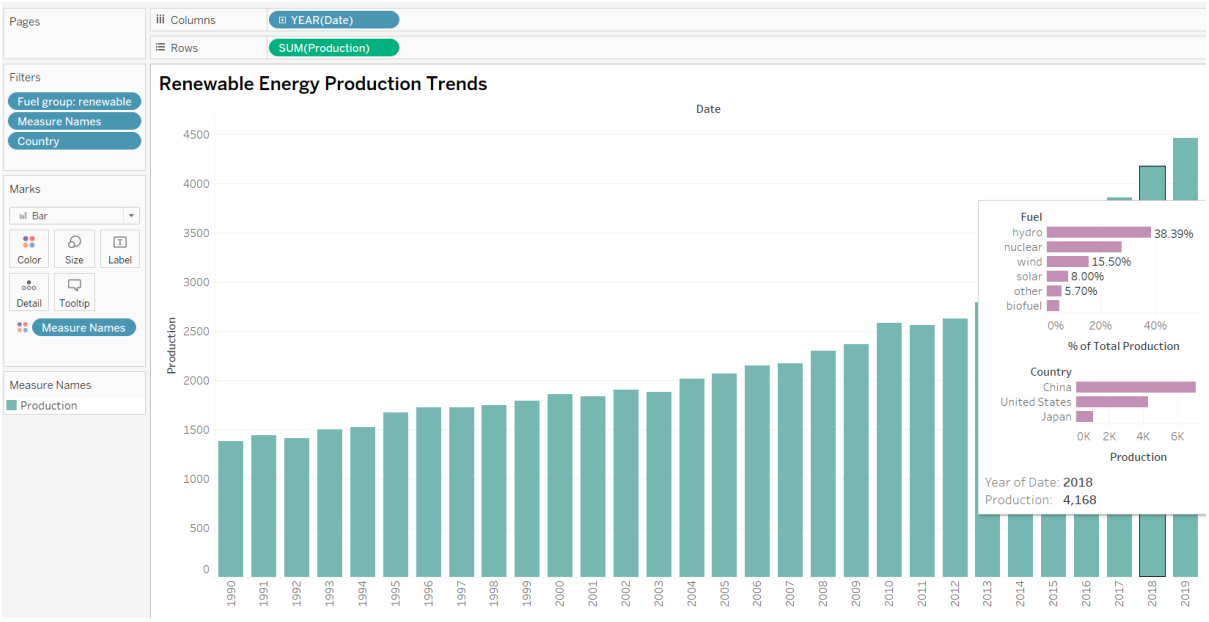


Figure 6

The above visualization (Figure 5) offers a comparative perspective on countries, showcasing their Production of renewable and non-renewable energy sources. Look at this bar chart named "Renewable Energy Production Trend." It gives us a snapshot of how much renewable energy each country produced in the year 1990 - 2019. China leads the pack in renewable energy production, churning out a whopping 1,321 terawatt-hours (TWh) in 2020. That's more than twice what the United States produced, putting them in second place with 935 TWh. Brazil,

Germany, and India also make a strong showing, each producing over 450 TWh of renewable energy.

We have also utilized the insert sheets feature in the tooltips to delve deeper into the details of every year by the fuel that is highly consumed and also differentiated by the country that is highly consuming these resources. This chart really emphasizes how important renewable energy is becoming worldwide. As countries work towards their climate goals, renewable energy is becoming a big player in global energy trends. It's like a growing trend where more and more countries are turning to cleaner energy sources.

Principle:

INTERACTIVITY FOR GRANULAR INSIGHT:

The usage of tooltips in a bar chart adheres to the interaction principle, improving the user experience and giving a path to granular knowledge. Tooltips are dynamic information overlays that display when users hold their cursor over particular bars and provide detailed data points. Users can utilize this interaction to investigate individual data values, compare bar heights, and obtain a more nuanced view of the dataset. The interactivity principle assures that the visualization is not static, but rather engages people in an adventurous journey, promoting a fuller understanding of the underlying information.

Visualization-7: AREA CHART

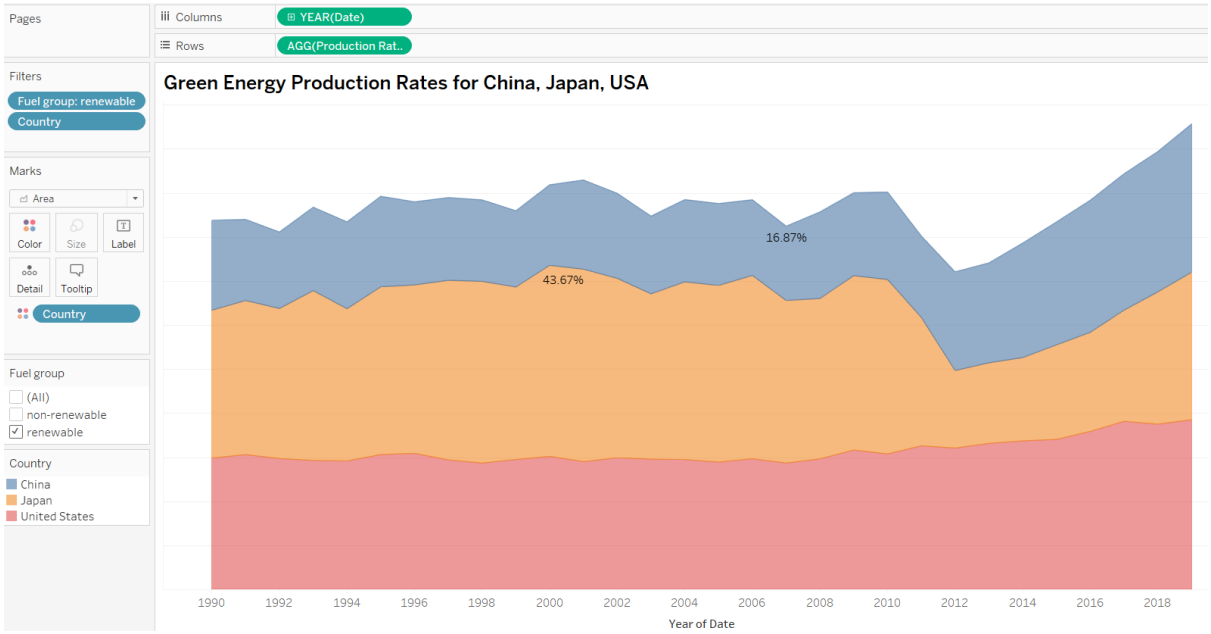


Figure 7

Above Area chart (Figure 6) visualizes the fluctuation in green energy production rates across China, Japan, and the USA from 1990 - 2020. China consistently led in green energy production showing a steady increase from 2012 to 2020. Japan experienced fluctuations in its green energy production rate, with a notable dip in 2012 compared to 2010. The USA witnessed a gradual rise in green energy production throughout the given period.

China stands out as the dominant force in green energy production among these three nations. Japan's journey in green energy is marked by more ups and downs compared to China and the USA. The USA has shown consistent growth in green energy production, although there were slight setbacks in recent years.

Principle:

CHRONOLOGICAL REPRESENTATION FOR TRENDS IDENTIFICATION:

The area chart illustrating renewable energy output over time adheres to the principle of chronological depiction. Viewers can spot trends and patterns in renewable energy output by viewing the data in chronological sequence. This approach provides a temporal knowledge of how renewable energy generation has progressed, which aids in debates regarding the transition to sustainable energy sources.

Visualization-8: PACKED BUBBLE CHART

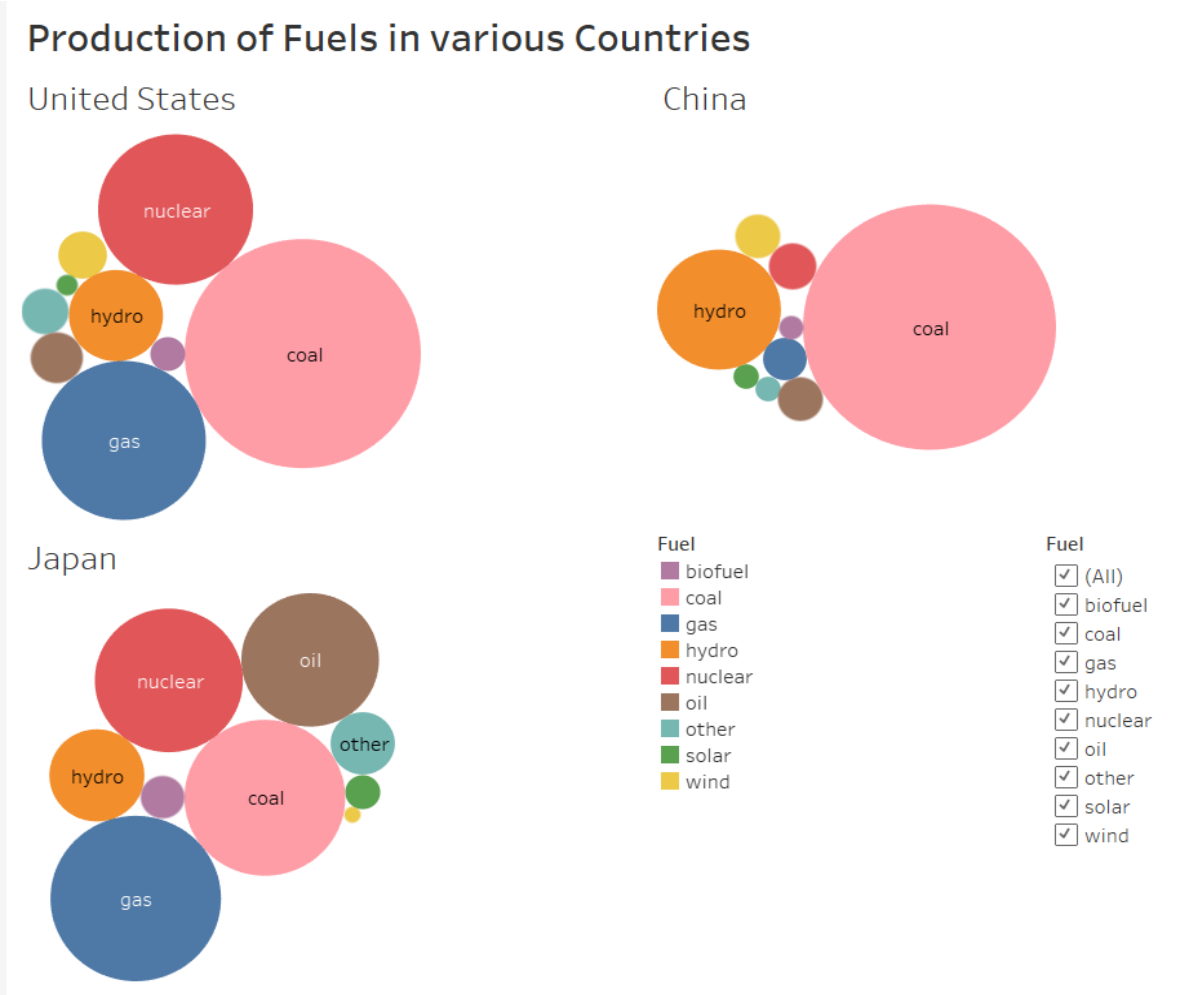


Figure 8

Above area chart (Figure 7) presents a comprehensive view of forested areas in various countries alongside their total land area. The variables Forested Area (%), Land Area (Km²), and Country are employed in this Area chart. Each country is represented by a shaded area, and the vertical axis showcases both forested and total land areas. Additionally, an average line provides a benchmark for comparing individual country data to the overall average. This visualization offers an effective means to understand the distribution of forested areas relative to land size, facilitating insights for environmentalists, policymakers, and researchers interested in forest conservation and land use around the world.

Principle:

VISUAL HIERARCHY FOR COMPARATIVE ANALYSIS:

The packed bubble charts for renewable energy production follow the visual hierarchy principle. This visualization technique creates a hierarchy by adjusting the size and color of bubbles, allowing viewers to readily compare and contrast different fuel kinds and production numbers. This method promotes efficient comparison analysis, which aids in identifying dominating renewable energy sources.

4. Summary and Conclusions:

ECONOMIC INSIGHTS: The GDP map showed the world's economic powerhouses, with the United States, China, and Japan at the top. Visualizations of unemployment rates revealed peaks in 2009 and 2020, highlighting the global impact of economic crises and the COVID-19 epidemic. Projected unemployment trends beyond 2021 provide insight into future economic recovery and obstacles. Found about countries having Tax rate%<5 and countries having zero % rate.

ENVIRONMENTAL INSIGHTS: Analysis of nonrenewable energy production revealed China's significant output, increasing concerns about environmental sustainability. Renewable energy production patterns in the selected countries demonstrated a positive shift toward sustainable energy sources. Packed bubble charts gave a detailed knowledge of fuel production composition, highlighting leading fuel sources and the diversity of energy mixes.

This research has provided a thorough examination of the selected countries (United states, China and Japan) economic and environmental aspects. The visualizations are effective decision-making tools, providing useful insights that can inform policies, tactics, and future directions. The interaction of economic and environmental forces highlights the importance of integrated solutions that balance growth with environmental stewardship. This initiative invites stakeholders, politicians, and researchers to join us on a path of continual discovery, adaptation, and collaboration in crafting a world where economic vibrancy and environmental stewardship coexist.

5. References:

1. NIDULA ELGIRIYEWITHANA, Global Country Information Dataset 2023, Kaggle (https://www.kaggle.com/datasets/nelgiriyewithana/countries-of-the-world-2023)

2. Global Country Information Dataset 2023, GigaSheet (https://www.gigasheet.com/sample-data/global-country-information-dataset-2023-283f8#:~:text=The%20table%20provides%20comprehensive%20data,and%20comparing%20global%20country%20information)

3. FERNANDO LASSO, Countries of the World, Kaggle (https://www.kaggle.com/datasets/fernandol/countries-of-the-world)

4. Daniboy370, WORLD DATA by country (2020), Kaggle (https://www.kaggle.com/datasets/daniboy370/world-data-by-country-2020)

5. Shakir khan, Data Visualization to Explore the Countries Dataset for Pattern Creation, Research Gate (https://www.researchgate.net/publication/356813621_Data_Visualization_to_Explore_the_Countries_Dataset_for_Pattern_Creation)

6. Edward Tufte, The Visual Display of Quantitative Information (https://faculty.salisbury.edu/~jtanderson/teaching/cosc311/fa21/files/tufte.pdf)

7. Pralabh Poudel, World Energy Consumption (https://www.kaggle.com/datasets/pralabhpoudel/world-energy-consumption)

6. Contributions:

Sai Krishna Pallerla - Data Collection, Data Visualization, Data Mapping

Dhanush Bijavara Mallikarjunappa - Data Collection, Data Visualization, Data Exploration

Jyothi Lohith Kumar Mamidi - Data Collection, Data Visualization, Data Trends

Nirmala Aiswarya Dosibhatla - Data Collection, Data Visualization, Data Preprocessing

Sahithi Reddy Bairi- Data Collection, Data Visualization, Storytelling

Snehalatha Reddy Anna Reddy - Data Collection, Data Visualization, Decision Making