DATA ARCHITECTURE

DATA - 039 - 008

STUDENT NAME: PRAISELIN LYDIA GLADSTON
STUDENT ID: 000965570

A REPORT ON GLOBAL WARMING DATA ANALYTIC SOLUTION

Table of Contents

About the Dataset 3

Purpose and Goals 4

Key Questions on the Data 4

Metrics and KPIs 5

- Metrics 5
- KPI (Key Performance Indicators) 5

Data Model 6

Power BI Dashboard 7

- Components 8
  - Slicers 8
  - Map 8
  - Measure 9
  - Line Graph 9
  - Scatter plots 10
  - Bar Chart 11

Results 11

Conclusion 12

About the Dataset

The dataset that is analyzed in this report is about Global Warming where there are data of 195 different countries from the year 1900 to 2023. The dataset was selected from Kaggle. It offers a comprehensive collection of environmental and socioeconomic indicators across the 195 countries.

We all know that global warming is the rise in earth's temperature as the result of emission of greenhouse gases like CO_2 and methane. The data does not just consist details of the emissions but also the other surrounding factors that contribute to the increase in temperature like deforestation, fossil fuel usage, industrial activity, energy consumption. These factors though are not the primary contributors for global warming, are enablers to the main cause.

This extensive dataset also encompasses various metrics, including temperature anomalies, population figures, GDP, renewable energy usage, sea level rise, Arctic ice extent, urbanization rates, extreme weather events, average rainfall, solar energy potential, waste management practices, per capita emissions, air pollution indices, biodiversity indices, ocean acidification levels, energy consumption per capita, policy scores, and average temperatures.

The columns and the units of data provided for them are:

Column Name	**Unit**
Temperature Anomaly	Degrees Celsius (ºC)
CO_2 Emissions	Metric Tons (Mt)
Population	Number of People
Forest Area	Percentage of Land Area (%)
GDP	US Dollars (Trillions)
Renewable Energy Usage	Percentage of Total Energy (%)
Methane Emissions	Metric Tons (Mt)
Sea Level Rise	Millimeters (mm)
Arctic Ice Extent	Million Square Kilometers (M Km²)
Urbanization	Percentage of Population (%)
Deforestation Rate	Percentage Change in Forest Cover (%)
Extreme Weather Events	Count (Number of Events)
Average Rainfall	Millimeters Per Year (mm)
Solar Energy Potential	Kilowatt-hours per Square Meter (kWh/m²)
Waste Management	Percentage of Waste Recycled (%)

Per Capita Emissions	Metric Tons of CO_2 per Person (Mt/person)
Industrial Activity	Industrial Production Index (Index Value)
Air Pollution Index	Index (AQI)
Biodiversity Index	Index Score (Scale Varies)
Ocean Acidification	pH Level
Fossil Fuel Usage	Tons of Oil Equivalent per Person (toe/person)
Energy Consumption Per Capita	Kilowatt-hours per Person (kWh/person)
Policy Score	Index Score (0 – 100)
Average Temperature	Degrees Celsius (ºC)

Purpose and Goals

The primary objective of analyzing this dataset is to gain a nuanced understanding of the multiple dimensions of global warming and climate change across different nations and time periods. By examining the interplay between environmental indicators (such as temperature anomalies, CO_2 emissions, and rise in sea level) and socioeconomic factors (including GDP, population, and urbanization), the analysis aims to:

- Identify temporal trends and patterns in climate-related variables.
- Assess the relationships and potential causal links between human activities and environmental changes.
- Evaluate the effectiveness of policy measures and renewable energy adoption in mitigating adverse climate impacts.
- Provide actionable insights for policymakers, researchers, and stakeholders to inform data-driven decisions and strategies addressing climate change challenges.

Key Questions on the Data

Based on the purpose of our analysis, we should be asking questions over the dataset. By deciding on what we are expecting over the dataset, we can decide on what visualization is required to be shown in our dashboard. The key questions we asked for our analysis are:

1. **How has global warming progressed over time?**
    - Track changes in **temperature anomaly** and **sea level rise** over different years.

- Analyze long-term trends in **CO_2 and methane emissions** to understand their contribution.

2. **Which country contributes more towards global warming?**
   - Identify the top contributors to **CO_2 emissions** and **methane emissions** globally.
   - Examine country-wise variations in **temperature anomalies** and **sea level rise**.
3. **What are the factors and their correlation to an increase in global warming?**
   - Assess relationships between **CO_2 emissions**, **methane emissions**, and **temperature anomaly**.
   - Investigate the impact of **average rainfall** on temperature variations.
4. **Does economic growth (GDP) and urbanization impact renewable energy adoption?**
   - Compare **GDP growth** and **urbanization trends** with shifts in **CO_2 emissions**.
   - Examine if countries with high economic growth are adopting more renewable energy sources.

Metrics and KPIs

To effectively monitor and evaluate progress toward the analysis goals, the following metrics and Key Performance Indicators (KPIs) have been established:

Metrics

The dataset consists of 23 columns other than Country and Year which are all measurements that are used to track various aspects of the environment. All the measurements used here can be used as a metric.

KPI (Key Performance Indicators)

Though all the KPIs are metrics, not all metrics are eligible to be KPIs. The KPIs that are chosen to be the indicators are:

1. **Sea Level Rise**

   This indicator how much the glaciers have melted resulting in the increase in sea levels. If sea level rises, it means global temperature is higher than the historical average. It is measured in millimeters per year. If the sea level is maintained about in its average level, it means the year has very less effect of global warming.

2. **Average Rainfall**

   Global warming disrupts the water cycle, leading to changes in rainfall patterns across the world. Some regions experience heavier rainfall and flooding, while others face droughts and reduced precipitation. These changes occur due to increased temperatures affecting evaporation, moisture content, and atmospheric circulation. Observing a region's average rainfall can indicate the magnitude of the effect of global warming in it.

3. **Temperature Anomaly**

   Temperature anomaly is a key indicator of global warming because it measures deviations from the historical average temperature, revealing overall climate trends. A positive anomaly indicates warming, while a negative anomaly suggests cooling.

4. **CO_2 Emissions**

   Carbon dioxide (CO_2) emissions are a primary driver of global warming because they trap heat in the Earth's atmosphere, intensifying the greenhouse effect. Human activities, especially fossil fuel combustion and deforestation, have significantly increased CO_2 levels since the Industrial Revolution. The direct correlation between rising CO_2 concentrations and global temperature increases is well-documented, as higher emissions lead to more heat retention, resulting in climate shifts.

5. **Methane Emissions**

   Methane (CH_4) is a potent greenhouse gas that contributes significantly to global warming, despite being present in smaller quantities than CO_2. It has a much higher heat-trapping potential — over 25 times more effective than CO_2 over a 100-year period. As methane emissions rise, they accelerate atmospheric warming.

Data Model

The dataset is organized as a flat table, where each record corresponds to a unique recording of data of a country in a year. The primary components of the data model include:

- **CountryID**: A unique identifier for each country.
- **Year**: The specific year of the recorded data.
- **Environmental Indicators**: Columns capturing data on temperature anomalies, CO_2 emissions, methane emissions, sea level rise, Arctic ice extent,

deforestation rates, extreme weather events, average rainfall, solar energy potential, air pollution index, biodiversity index, ocean acidification, and average temperature.

- **Socioeconomic Indicators**: Columns detailing population, forest area, GDP, renewable energy usage, urbanization, industrial activity, fossil fuel usage, energy consumption per capita, waste management practices, and policy scores.

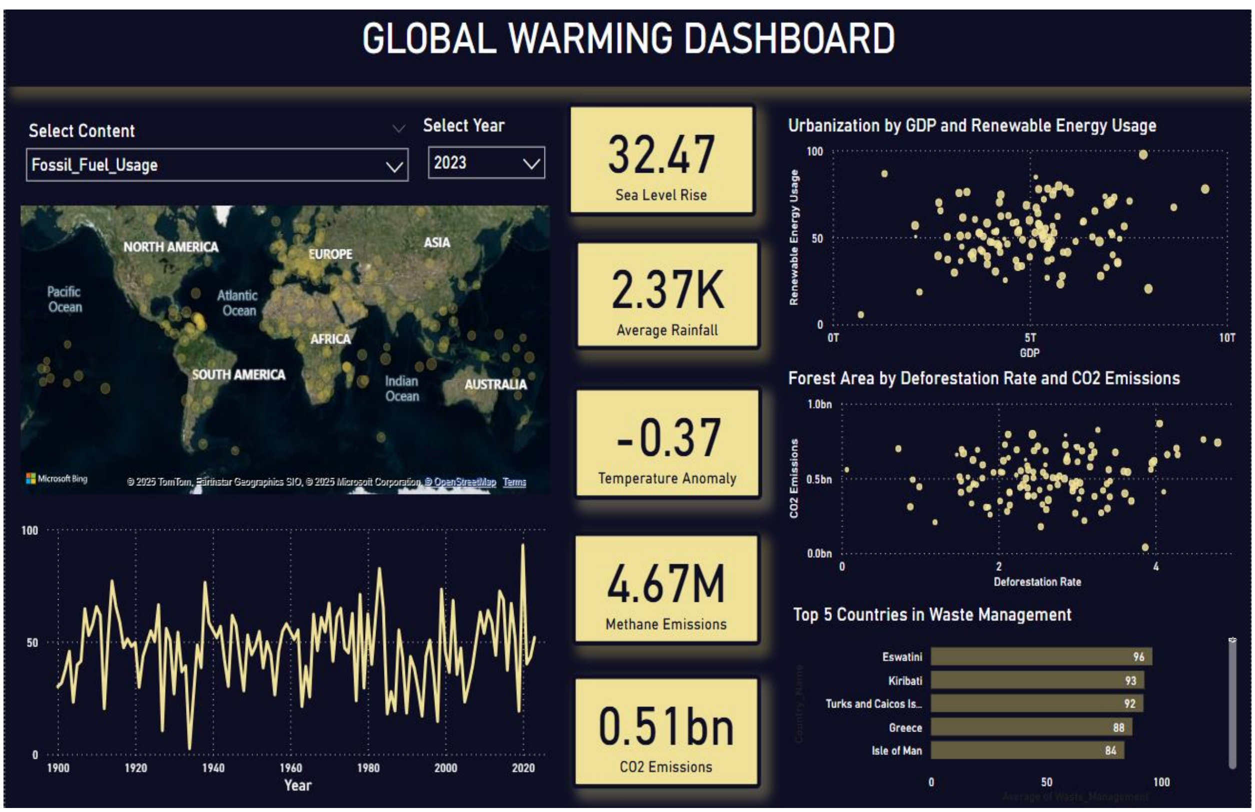
This structure facilitates multifactorial analyses, enabling the exploration of relationships between environmental changes and socioeconomic factors across different temporal and spatial scales.

The only entity in the dataset is global warming and therefore there isn't an entity relationship to be shown in a data model.

Power BI Dashboard

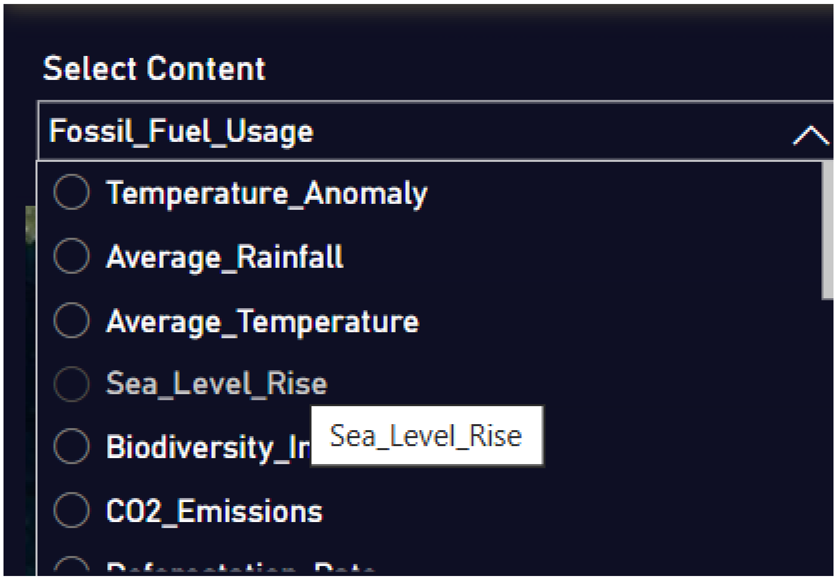
After the thorough preparation of the questions and identifying KPIs for the same, we can move on to designing the dashboard for our analysis.

Below is the overview of the dashboard:



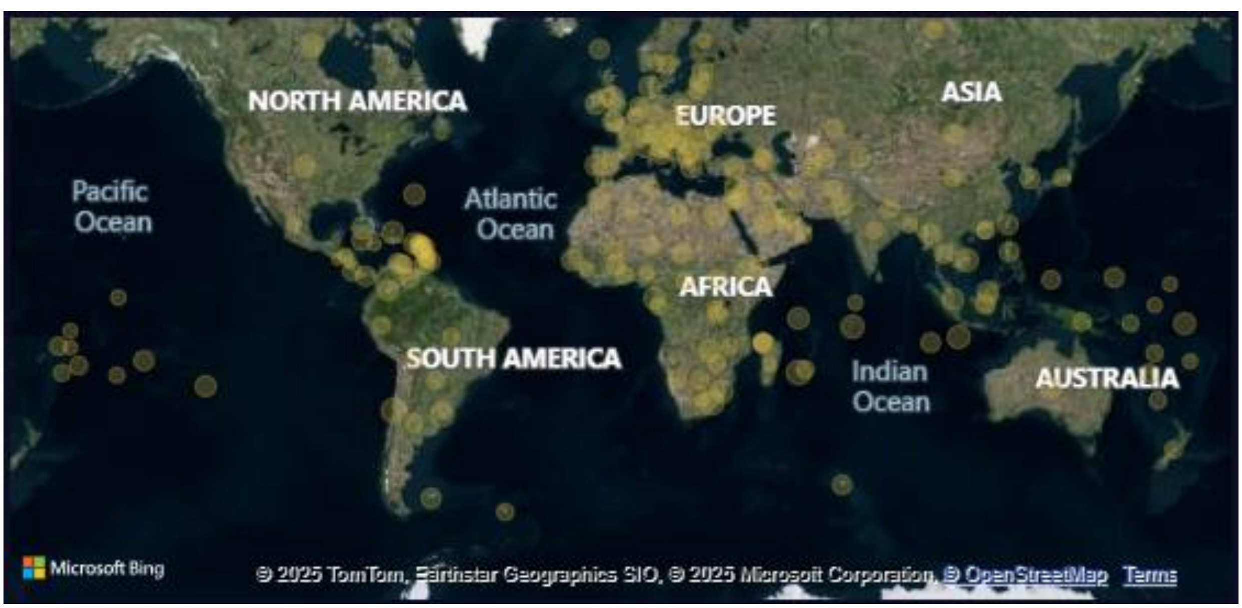
Components

Slicers



The slicers used here are to filter out data based on the values selected. The first slicer to select content is used to select the metric which we want to analyze. This filters out one metric, and its trend will be shown in a line graph that will be explained below. The second slicer is used to filter out a year. This is used to filter data in all other components used in the dashboard.

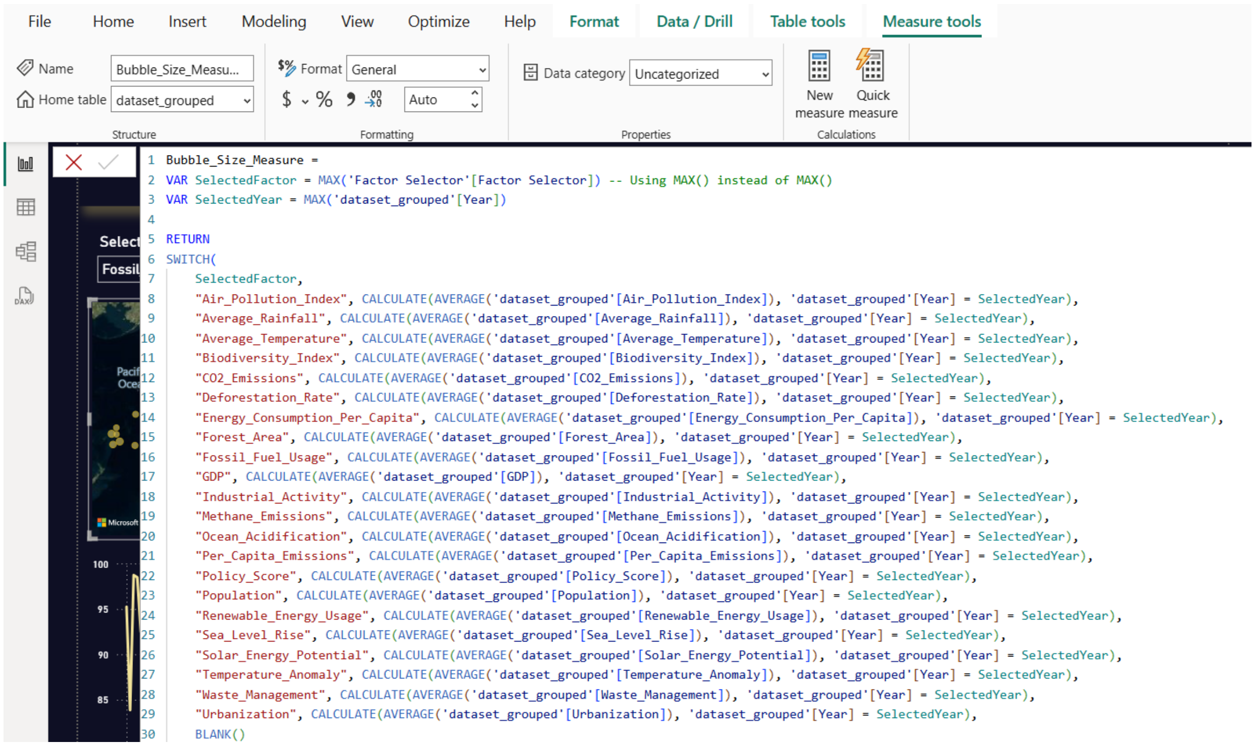
Map



The map is plot over the countries using the Country field in the dataset. The Country field is made plottable by setting the data category of that field to ‘Country’ in Column Tools. The bubble over the counties determines the magnitude or intensity of the content selected from the first slicer in the year selected in the second slicer. For this we defined a new measure called ‘Bubble Size Measure’.

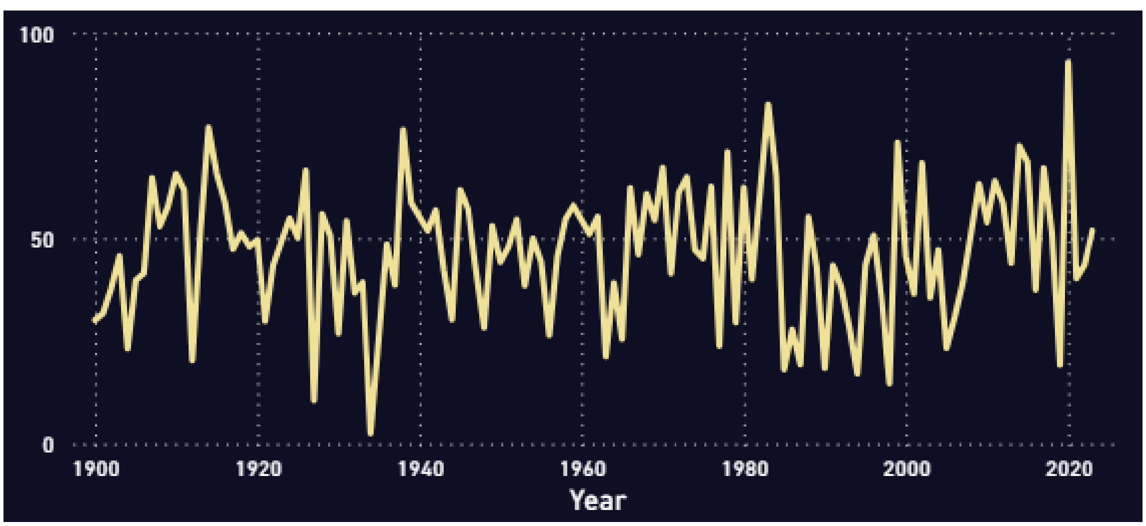
Measure

The bubble size measure is defined as follows:



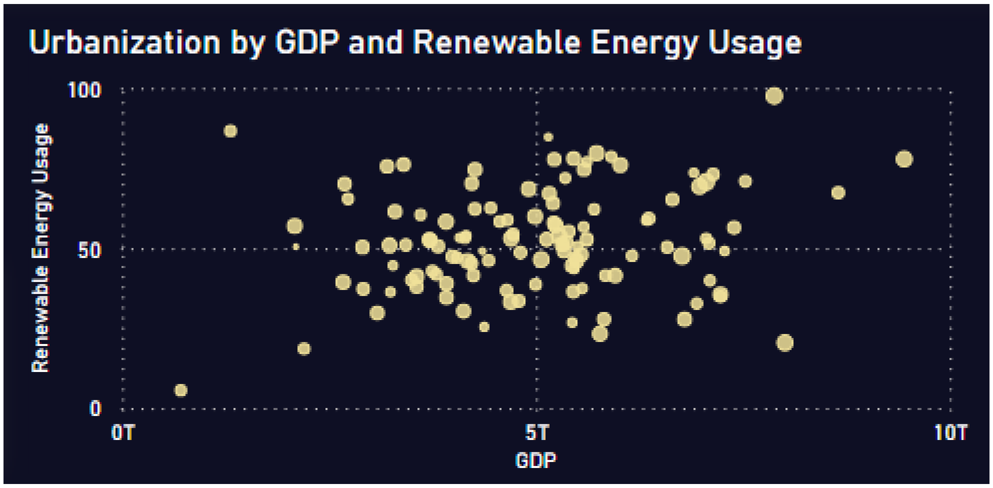

```
Bubble_Size_Measure =
VAR SelectedFactor = MAX('Factor Selector'[Factor Selector]) -- Using MAX() instead of MAX()
VAR SelectedYear = MAX('dataset_grouped'[Year])

RETURN
SWITCH(
    SelectedFactor,
    "Air_Pollution_Index", CALCULATE(AVERAGE('dataset_grouped'[Air_Pollution_Index]), 'dataset_grouped'[Year] = SelectedYear),
    "Average_Rainfall", CALCULATE(AVERAGE('dataset_grouped'[Average_Rainfall]), 'dataset_grouped'[Year] = SelectedYear),
    "Average_Temperature", CALCULATE(AVERAGE('dataset_grouped'[Average_Temperature]), 'dataset_grouped'[Year] = SelectedYear),
    "Biodiversity_Index", CALCULATE(AVERAGE('dataset_grouped'[Biodiversity_Index]), 'dataset_grouped'[Year] = SelectedYear),
    "CO2_Emissions", CALCULATE(AVERAGE('dataset_grouped'[CO2_Emissions]), 'dataset_grouped'[Year] = SelectedYear),
    "Deforestation_Rate", CALCULATE(AVERAGE('dataset_grouped'[Deforestation_Rate]), 'dataset_grouped'[Year] = SelectedYear),
    "Energy_Consumption_Per_Capita", CALCULATE(AVERAGE('dataset_grouped'[Energy_Consumption_Per_Capita]), 'dataset_grouped'[Year] = SelectedYear),
    "Forest_Area", CALCULATE(AVERAGE('dataset_grouped'[Forest_Area]), 'dataset_grouped'[Year] = SelectedYear),
    "Fossil_Fuel_Usage", CALCULATE(AVERAGE('dataset_grouped'[Fossil_Fuel_Usage]), 'dataset_grouped'[Year] = SelectedYear),
    "GDP", CALCULATE(AVERAGE('dataset_grouped'[GDP]), 'dataset_grouped'[Year] = SelectedYear),
    "Industrial_Activity", CALCULATE(AVERAGE('dataset_grouped'[Industrial_Activity]), 'dataset_grouped'[Year] = SelectedYear),
    "Methane_Emissions", CALCULATE(AVERAGE('dataset_grouped'[Methane_Emissions]), 'dataset_grouped'[Year] = SelectedYear),
    "Ocean_Acidification", CALCULATE(AVERAGE('dataset_grouped'[Ocean_Acidification]), 'dataset_grouped'[Year] = SelectedYear),
    "Per_Capita_Emissions", CALCULATE(AVERAGE('dataset_grouped'[Per_Capita_Emissions]), 'dataset_grouped'[Year] = SelectedYear),
    "Policy_Score", CALCULATE(AVERAGE('dataset_grouped'[Policy_Score]), 'dataset_grouped'[Year] = SelectedYear),
    "Population", CALCULATE(AVERAGE('dataset_grouped'[Population]), 'dataset_grouped'[Year] = SelectedYear),
    "Renewable_Energy_Usage", CALCULATE(AVERAGE('dataset_grouped'[Renewable_Energy_Usage]), 'dataset_grouped'[Year] = SelectedYear),
    "Sea_Level_Rise", CALCULATE(AVERAGE('dataset_grouped'[Sea_Level_Rise]), 'dataset_grouped'[Year] = SelectedYear),
    "Solar_Energy_Potential", CALCULATE(AVERAGE('dataset_grouped'[Solar_Energy_Potential]), 'dataset_grouped'[Year] = SelectedYear),
    "Temperature_Anomaly", CALCULATE(AVERAGE('dataset_grouped'[Temperature_Anomaly]), 'dataset_grouped'[Year] = SelectedYear),
    "Waste_Management", CALCULATE(AVERAGE('dataset_grouped'[Waste_Management]), 'dataset_grouped'[Year] = SelectedYear),
    "Urbanization", CALCULATE(AVERAGE('dataset_grouped'[Urbanization]), 'dataset_grouped'[Year] = SelectedYear),
    BLANK()
```

Line Graph

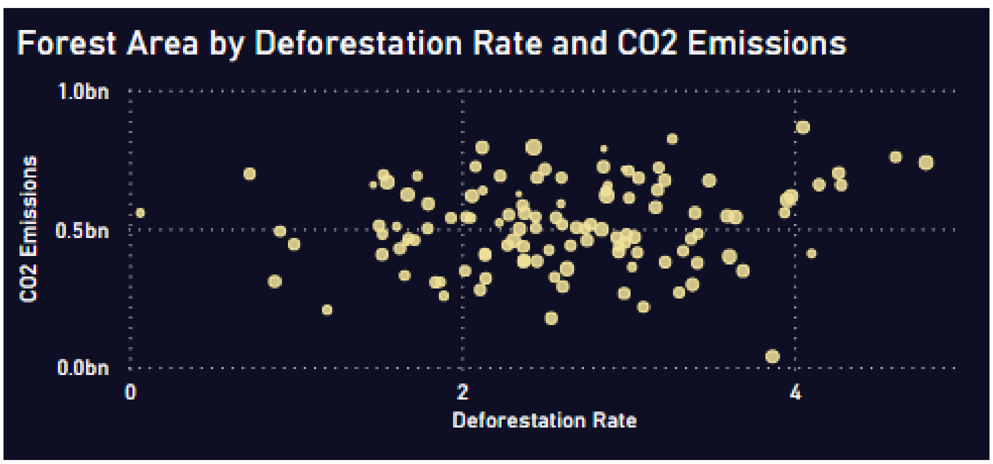


The line graph represents the trend of the content selected in the first slicer over the years 1900 to 2023. This line graph represents the fossil fuel usage of a country. It is sensible to notice why the fossil fuel usage does not grow since the population would grow over the years, but the graph is averaged out at 50 Tons if Oil Equivalent (toe). This is because the values in our dataset represent toe per person.

Scatter plots



This plot shows the correlation between GDP and the renewable energy usage of a country (selected from Map) in all the years recorded. The bubbles in the graph represent the urbanization index in that country in that specific year. The values seem to be averaged out because the renewable energy usage percentage is normalized based on the population of that country. Maintaining an average means the country has been improving in adaptation of renewable energy as the GDP grows.



This plot shows the correlation between deforestation rate and CO_2 emissions in the atmosphere. As the country faces deforestation, the trees would not absorb the CO_2 in the atmosphere. The bubbles represent the forest area in that country in that specific year. The values here seem to be averaged out because the deforestation rate is the percentage change in the forest cover (normalized based on the forest area).

Bar Chart

This graph represents the top 5 countries, in the year selected in second slicer, that had good waste management percentage. Countries with good waste management have control over the emission of gases like methane into the atmosphere.

Results

Upon analyzing the dataset, several key findings emerge:

- **Rising Temperature Anomalies**: There is a clear upward trend in global temperature anomalies, with significant increases observed post-1970s. This period aligns with industrial expansion and increased greenhouse gas emissions.
- **Correlation Between Deforestation and CO_2 Emissions**: A positive correlation exists between deforestation rates and CO_2 emissions, suggesting that countries experiencing higher deforestation also report elevated CO_2 levels, likely due to reduced carbon sequestration capacity.
- **Impact of Renewable Energy on Emissions**: Countries with higher percentages of renewable energy usage tend to have lower per capita CO_2 emissions, indicating the effectiveness of renewable energy adoption in mitigating carbon footprints.
- **Policy Effectiveness**: Nations with higher policy scores often exhibit favorable trends in environmental indicators, such as reduced air pollution indices and stabilized biodiversity indices, underscoring the role of robust environmental policies in driving positive outcomes.

Conclusion

The comprehensive analysis of the "Global Warming Dataset: 195 Countries (1900-2023)" underscores the intricate interplay between human activities and environmental changes. The observed trends highlight the critical need for concerted global efforts in adopting sustainable practices, enforcing stringent environmental policies, and promoting renewable energy to mitigate the adverse impacts of climate change. Policymakers and stakeholders are encouraged to leverage these insights to formulate data-driven strategies that address the multifaceted challenges posed by global warming.

Identifying the causes and taking appropriate measures is necessary to create a better future for the next generations. This analysis has been a motivating factor for the same and clarified how data can affect or be helpful in improvising an existing environmental issue.