Visual Exploration of Large Earth Surface Temperature, Carbon Emissions and Sea Level Data Sets

Sudiksha Navik
sudiksha22@iitk.ac.in

Drashtant Singh Rathod
drashtants22@iitk.ac.in

Mannu Kr. Gaddhyan
mannukr22@iitk.ac.in

Susovan Patra
susovanp22@iitk.ac.in

Aditya Raj
adityaraj20@iitk.ac.in

Contents

1 Introduction

- Climate change refers to the long-term shifts in weather patterns that occur over decades or longer, resulting from changes in Earth's climate system caused primarily by human activities, such as the burning of fossil fuels, deforestation, and other land-use changes. It has significant impacts on ecosystems, natural resources, and human societies, including rising sea levels, more frequent and intense heatwaves, droughts, and extreme weather events such as hurricanes and floods.
- Climate Change is a complex and multifaceted phenomenon that is best understood through visualizations. Visualizing the change in climate over the years is essential for communicating the complex scientific information about the phenomenon in a more accessible and understandable way to various stakeholders. Visualization of climate data can help illustrate the current and future impacts of climate change, such as changes in temperature, precipitation, and sea level rise, which can aid policymakers, researchers, and the general public in developing informed decisions and taking action to mitigate and adapt to climate change. Effective visualization can also help raise awareness about climate change, fostering a greater sense of urgency for collective action.
- Our system strives to solve the problem of visualizing the complex and large datasets related to earth's climate. Through a combination of data visualization, mapping, and interactive features, our interface intends to provide a unique and engaging way to explore the impacts of climate change on our planet. We have mainly focused on three datasets:
  - Climate Change: Earth Surface Temperature Data
  - Carbon Dioxide Emissions of the World (1990-2018)
  - Global Sea Level Rise

2 Tasks

2.1 Data Pre-Processing

Data pre-processing refers to the process of preparing data for analysis by cleaning, transforming, and reducing the raw data into a more manageable format. It is a crucial step in data analysis, as the quality of the output depends largely on the quality of the input data.

We performed data pre-processing by following several steps:

2.1.1 Data Cleaning

This step involves removing irrelevant, duplicate, or inconsistent data, dealing with missing data, and correcting data errors and inconsistencies.

- Missing values: We handled missing values in our data in an efficient manner. For example, in temperature dataset, there were quite a few missing data for some cities. In order to handle this scenario, we averaged the data of its neighboring cities and filled in the values.

- Irrelevant features: We removed certain columns that were not useful for creating meaningful visualization.

2.1.2 Data Transformation

This step involves converting data into a more meaningful format, such as normalizing data, transforming categorical data into numerical data, or applying mathematical operations to data.

- We transformed the given data in order to suffice our requirements. For example, in the carbon emissions dataset, the data given was in the format:

Country	2018	2017	2016	2015...	...1991	1990

We transformed the data into the appropriate format:

Country	Year	CO2 Emissions

- We computed the average data file from the given data. Also in the original data, the temperatures were given month-wise so we calculated the annual temperatures out of it.

2.1.3 Data Analysis

This step refers to the process of systematically examining and interpreting data with the goal of extracting useful information and insights. It involves using various statistical and computational methods to identify patterns, relationships, and trends within a dataset.

- We analysed the attributes of our datasets and came up with the best suitable visualization for it.

3 Proposed Solution

We propose a standalone interface, built entirely from scratch using the **Plotly Dash**. It is a Python framework for building web applications that are focused on data visualization and analysis. Following are the reasons due to which we leaned towards the Dash Interface rather than creating visualizations over a web browser:

1. **Customization:** With Plotly Dash, we have complete control over the design and layout of our visualizations. We can create custom components, layouts, and styles to match your specific needs and branding. This level of customization is not possible with standard web browsers.
2. **Interactivity:** Plotly Dash allows us to create highly interactive visualizations that can respond to user input and update in real-time. This is very complex with static visualizations that are created using a web browser.
3. **Integration with Python:** Plotly Dash is built on top of Python, which is a popular language for data analysis and scientific computing. This means that we can use Python to manipulate and analyze our data, and then use Plotly Dash to create custom visualizations that are tailored to our specific needs.

4. **Performance:** Plotly Dash is designed to be highly performant, even when dealing with large and complex datasets. This is because Dash uses a reactive programming model that allows for efficient updates and rendering of visualizations.

5. **Sharing and Collaboration:** With Plotly Dash, we can easily share our visualizations with others and collaborate on projects in real-time. This is because Dash allows us to deploy our visualizations as standalone applications that can be accessed from anywhere with an internet connection.

3.1 Libraries Used

1. **Plotly:** Plotly provides a range of interactive visualization options that allow users to explore and analyze their data in real-time. These visualizations can be customized with various parameters and settings, and can be easily linked to data sources. We created several visualizations using the Plotly library, some of them being, Choropleth Maps, Heat Maps, Bar Charts, Line Charts, Box & Whiskers Plot, Area Chart, Map Boxes, Globes, Bubble Plots and more. The entire description of all the visualizations will be given in the next paragraph.

2. **Pandas:** Pandas is a Python library for data manipulation and analysis. It provides a range of functions and tools for working with structured data, including data frames and series. We used Pandas for loading and manipulating our the original datasets as per our requirements.

3. **Numpy:** NumPy is a Python library for numerical computing that provides powerful array manipulation and mathematical operations. It is a fundamental library for scientific computing in Python, providing support for mathematical operations on large, multi-dimensional arrays and matrices. We used Numpy for applying tranformations on our datasets.

4. **Json:** JSON stands for JavaScript Object Notation, and it is a lightweight data interchange format that is easy to read and write. It is based on a subset of the JavaScript programming language and is often used to transmit data between a server and a web application, as an alternative to XML. In order to read and load the GeoJson files for creating visulaizations of mapbox in the temperature datsets, Json is used.

5. **Matplotlib:** Matplotlib is a Python library for creating visualizations of data. It provides a range of functions and tools for creating a wide variety of plots and charts, including line plots, scatter plots, histograms, and more. We created a few visualizations of bar chart and line chart with the help of Matplotlib.

4 Results

We created a standalone interface visualizing the changes in Earth's surface temperature, carbon emissions and sea levels over the years.

4.1 Homepage

Figure 1

This is the Homepage of our interface. The page consists of a drop-down that lets us select what visualization we need:

1. Temperature
2. Carbon Emissions
3. Sea Level
4. Correlation

4.2 Visualization: Temperature

This dashboard consists of all the temperature visualizations that we created in order to gain insights from the temperature datset.

- **Calendar Heat Map:**
  - A calendar heat map is a visual representation of data using a calendar layout. It is a type of heatmap that displays data for each day of the year, usually in a rectangular grid. It is color-coded to represent different levels of a variable, in this case temperature.
  - The calendar Heat Map has three drop-downs, for selecting Country, City and Year. The y-axis represents twelve months and the x-axis represents days of a month.

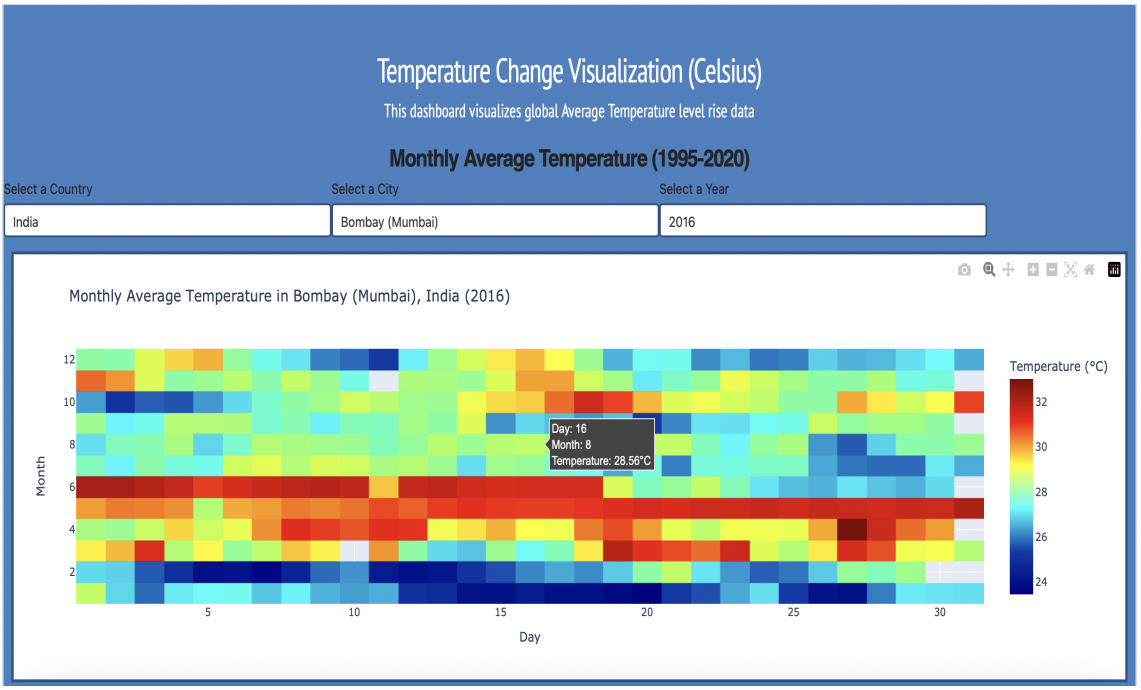


Figure 2

- **Choropleth Map:**
  - A choropleth map is a type of map that uses color to represent data values for geographic areas, such as countries, states, or counties.
  - Choropleth maps are created by dividing a geographic region into smaller units. The color scale used in a choropleth map is typically based on a gradient that ranges from light to dark colors, with lighter colors representing lower values and darker colors representing higher values.
  - We have given the years on the slider and the color-scale shows the average temperature of the corresponding country.

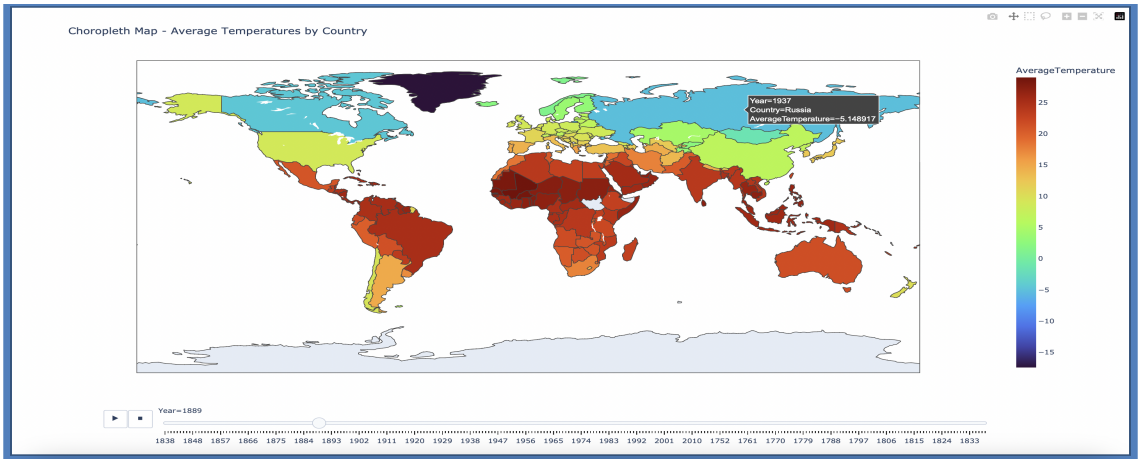


Figure 3

- **Earth Temperature Timeline:**

  The Earth Temperature Timeline is a visual representation of how the Earth's temperature has changed over time. It shows the fluctuations in temperature that have occurred over millions of years, as well as the rapid increase in temperature that has occurred in recent history. The plot shows the data between 1990 to 2020.

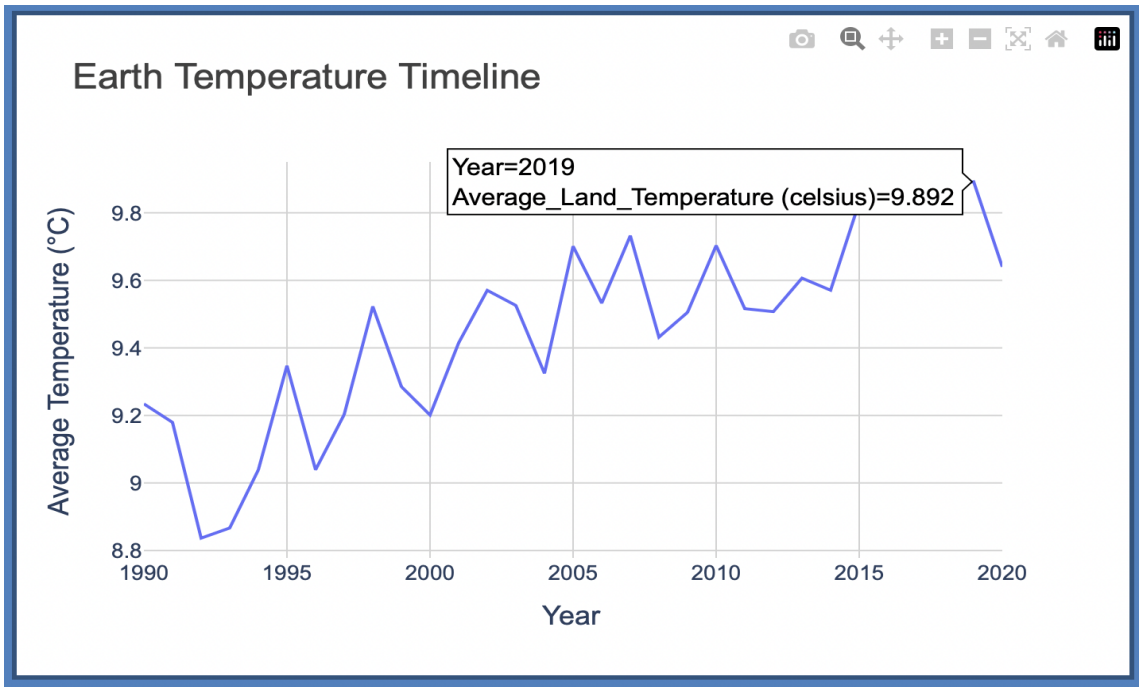


Figure 4

- **3-D Globe Representation:**

  A 3-D globe representation of average temperature is a visual representation of the Earth's temperature that displays the average temperature across the globe in three dimensions. The temperature data is mapped onto a globe using a color scale, with warmer temperatures represented by red or orange colors and cooler temperatures represented by blue colors.

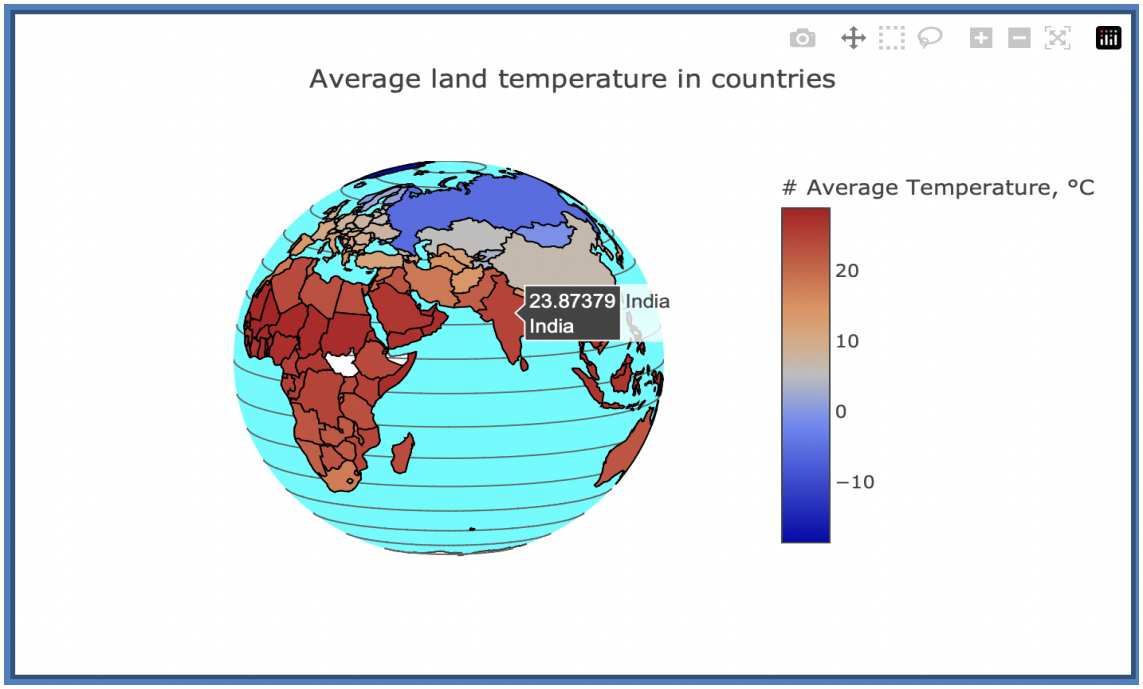


Figure 5

- **Heat Map:**

  This heat map of average temperature is a visual representation of how temperatures vary across different cities using a color scale. It helps to identify patterns and trends in temperature across different regions and seasons, and can be used to compare temperature data between cities.

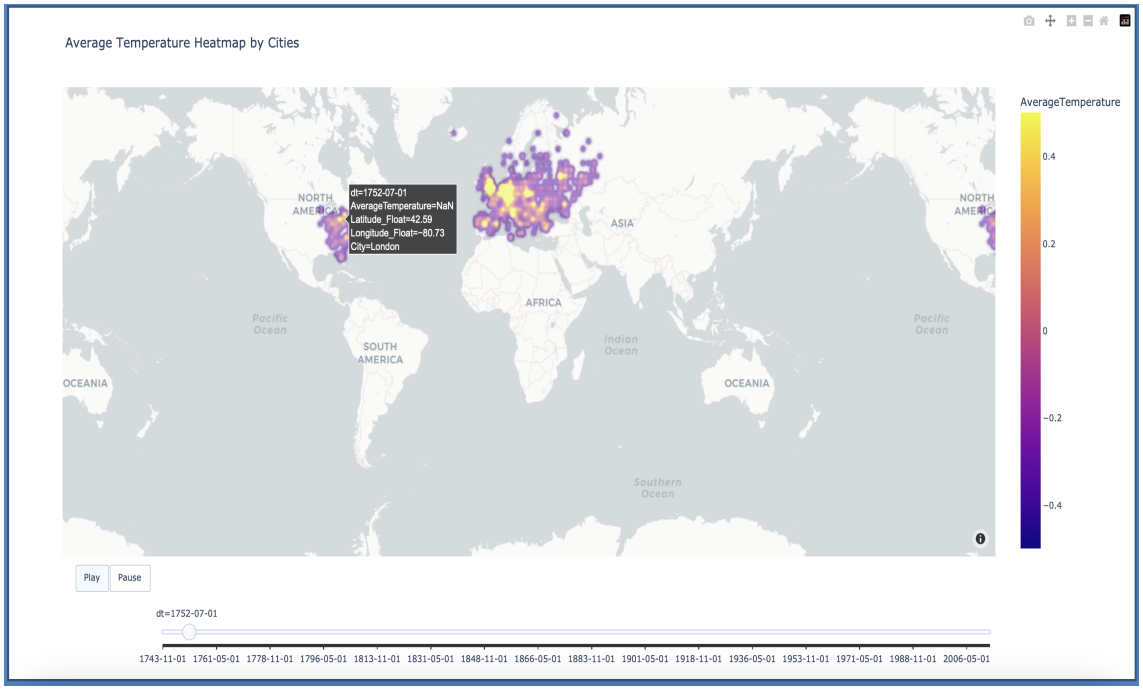


Figure 6

- **Map Box:**

  It is a map representation of temperature on the basis of states, averaged over since 1880 to 2015. The temperatures are decided according to the colorscale. We had the data of six countries:

  - India
  - China
  - Brazil
  - Canada
  - Russia
  - USA

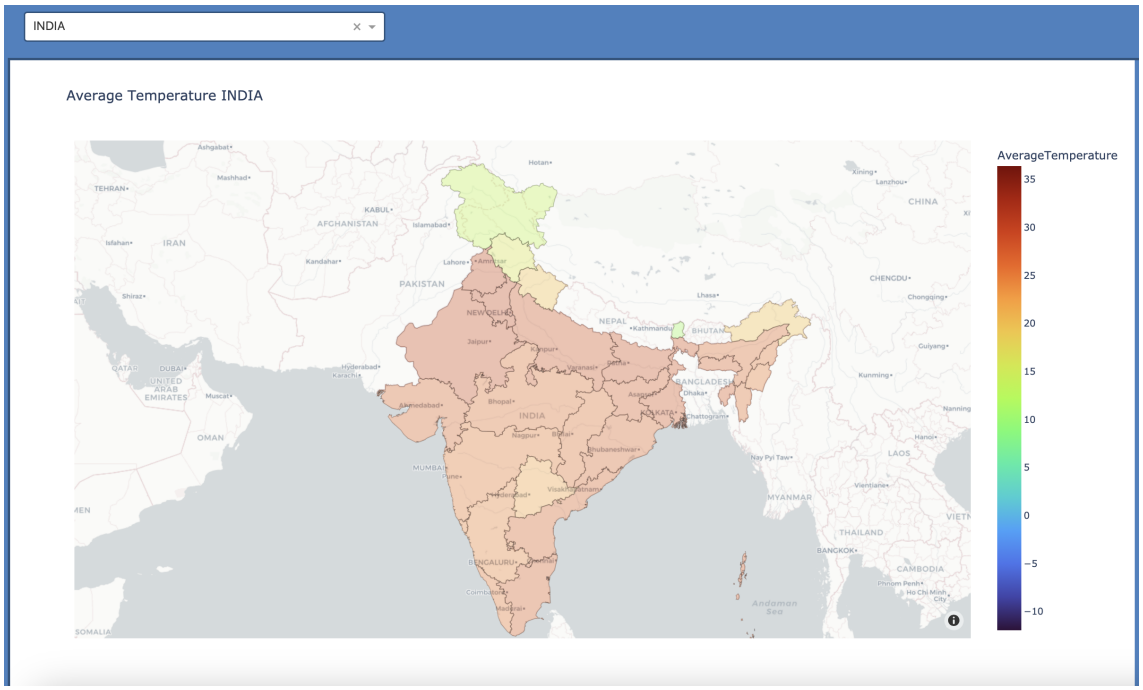


Figure 7

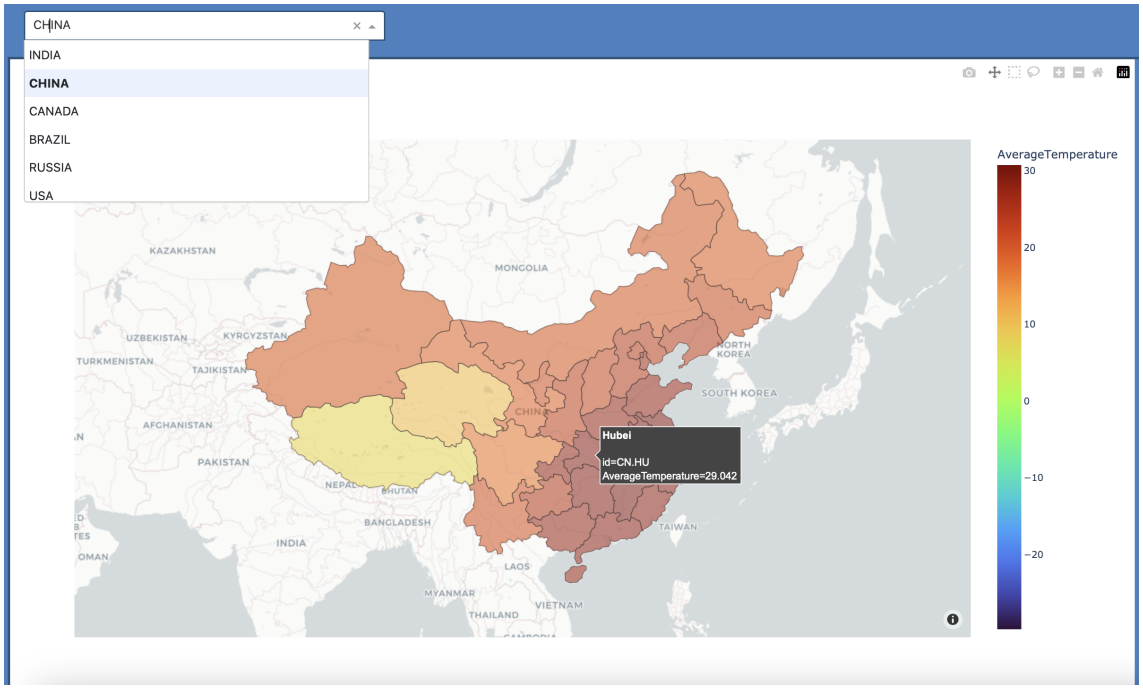


Figure 8

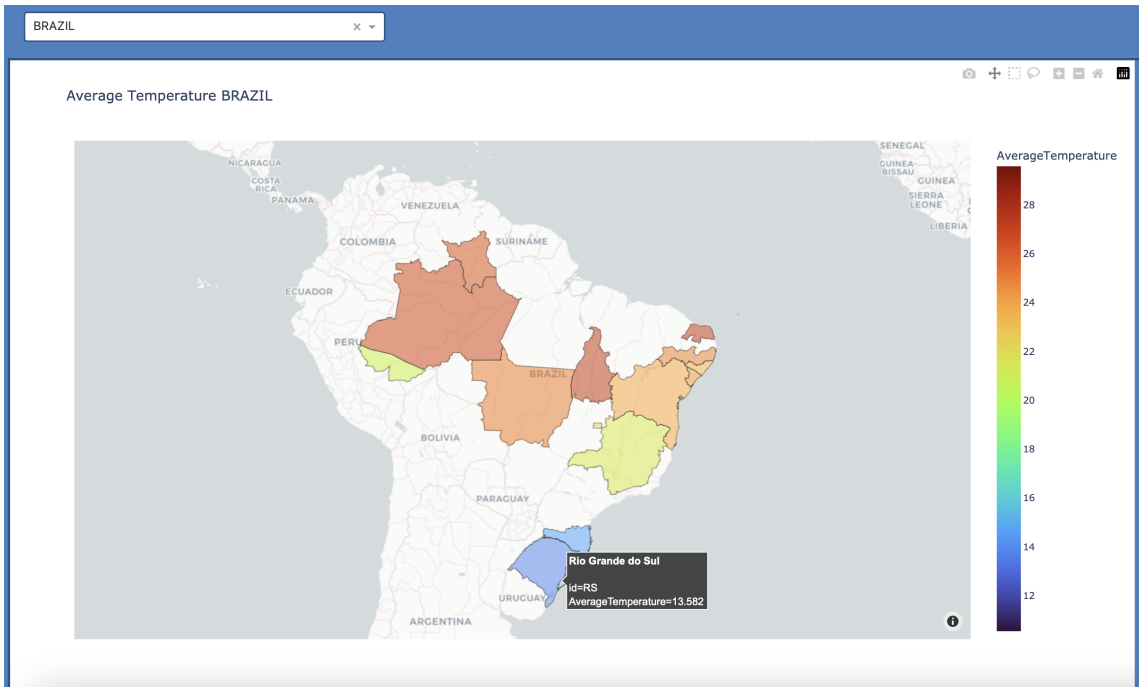


Figure 9

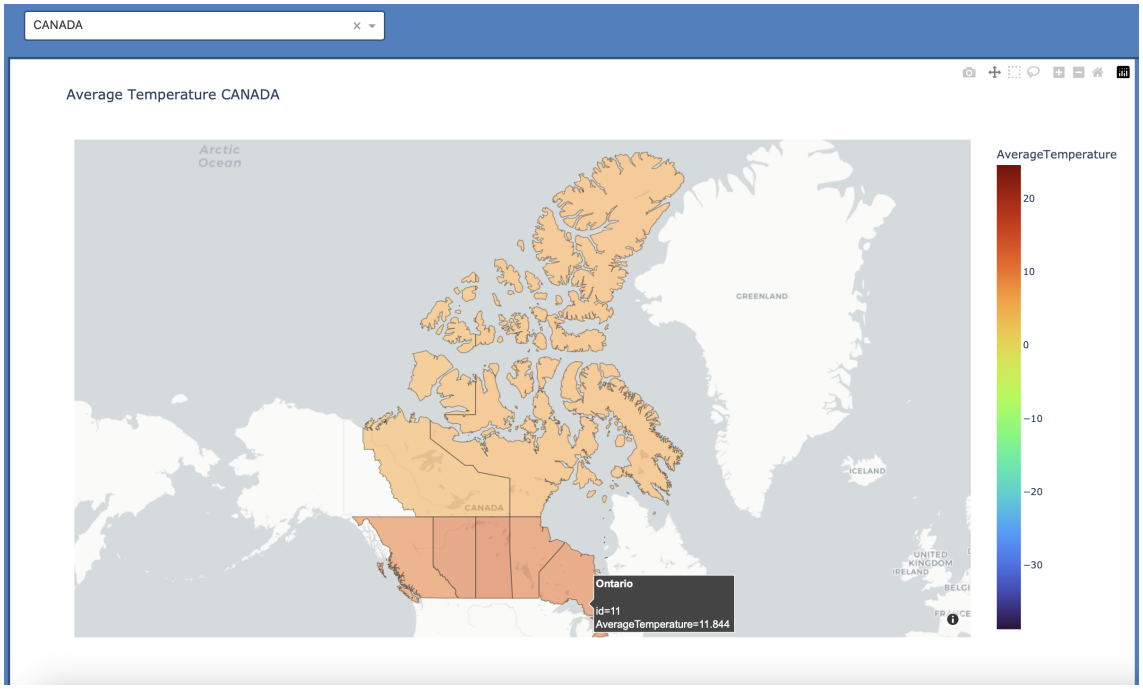


Figure 10

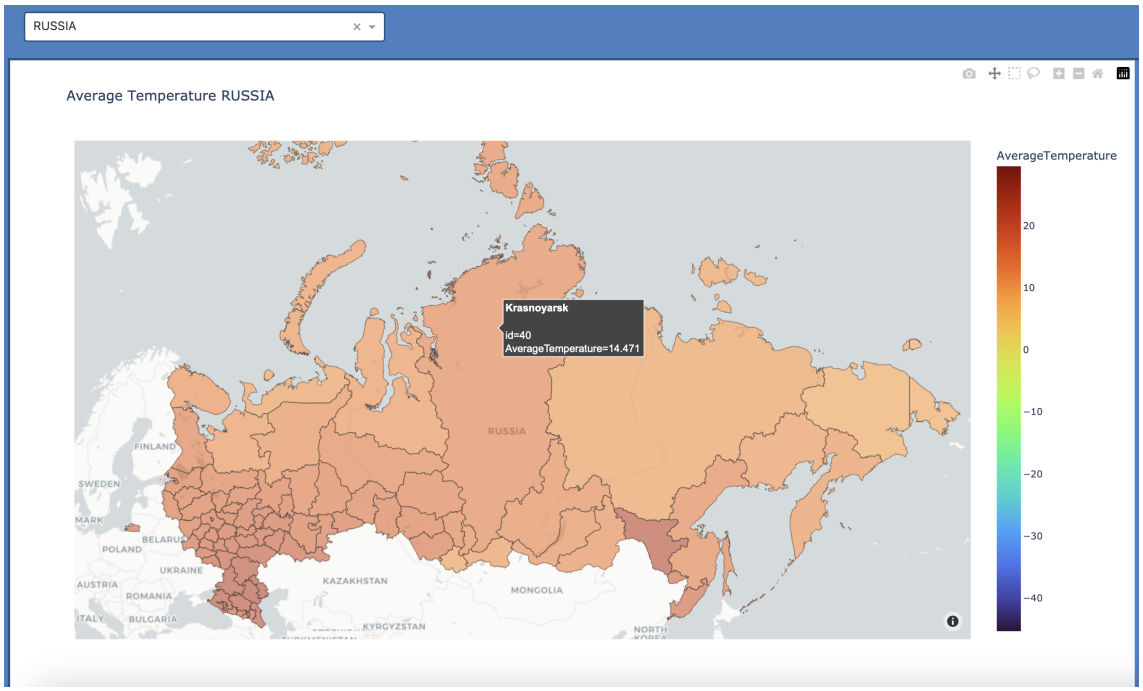


Figure 11

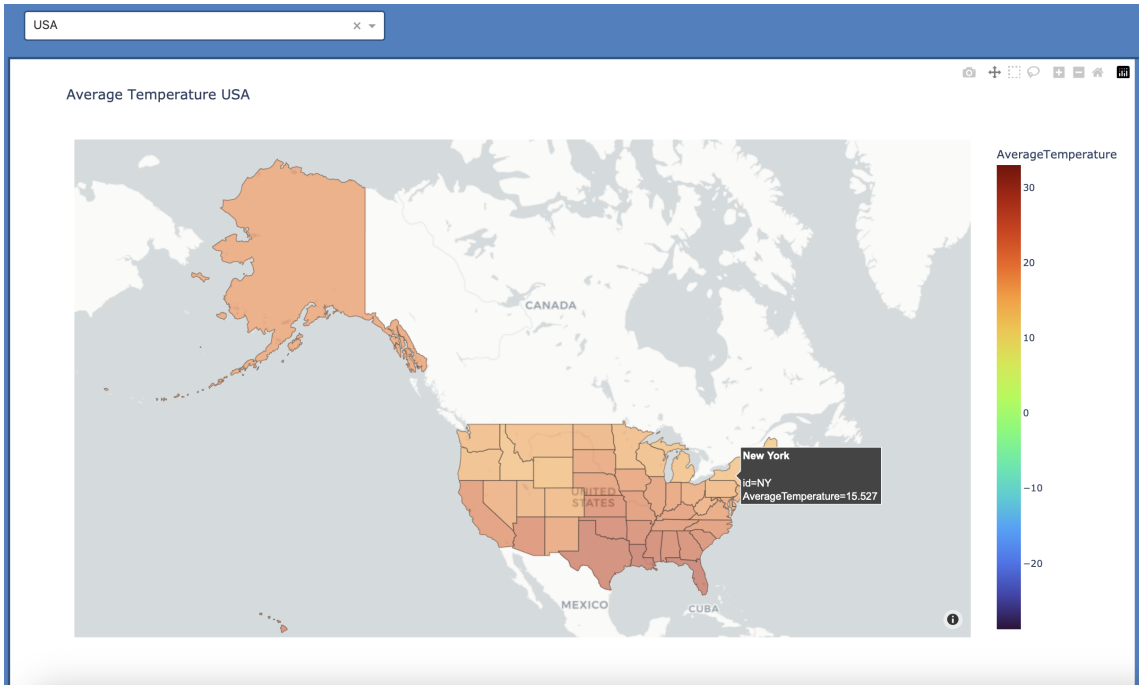


Figure 12

- **Line Charts:**

  We created line charts showing average temperatures of continents from 1994 to 2019. We can turn any of the graph on/off as per our wish.

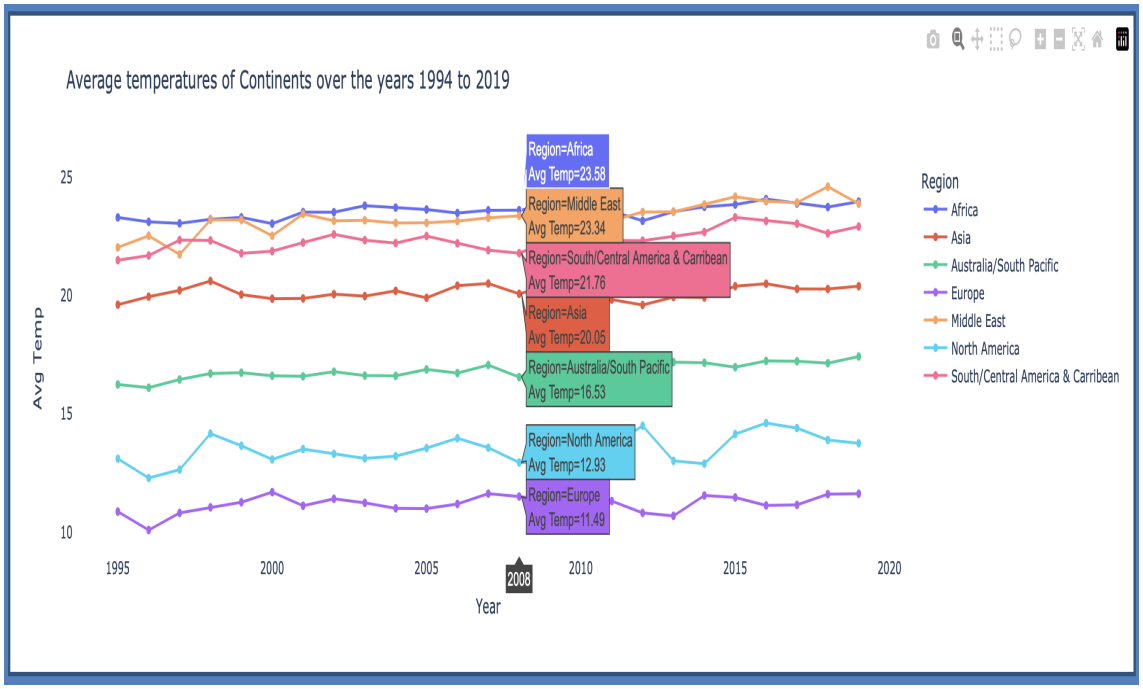


Figure 13

- **Globe-2D:**
  - This globe representation shows the how much the average temperature of countries has risen since 1825 to 2018.
  - We can see the dots of different sizes in the visualization, showing the difference value. Also the overall average temperature is shown according to the colorscale.

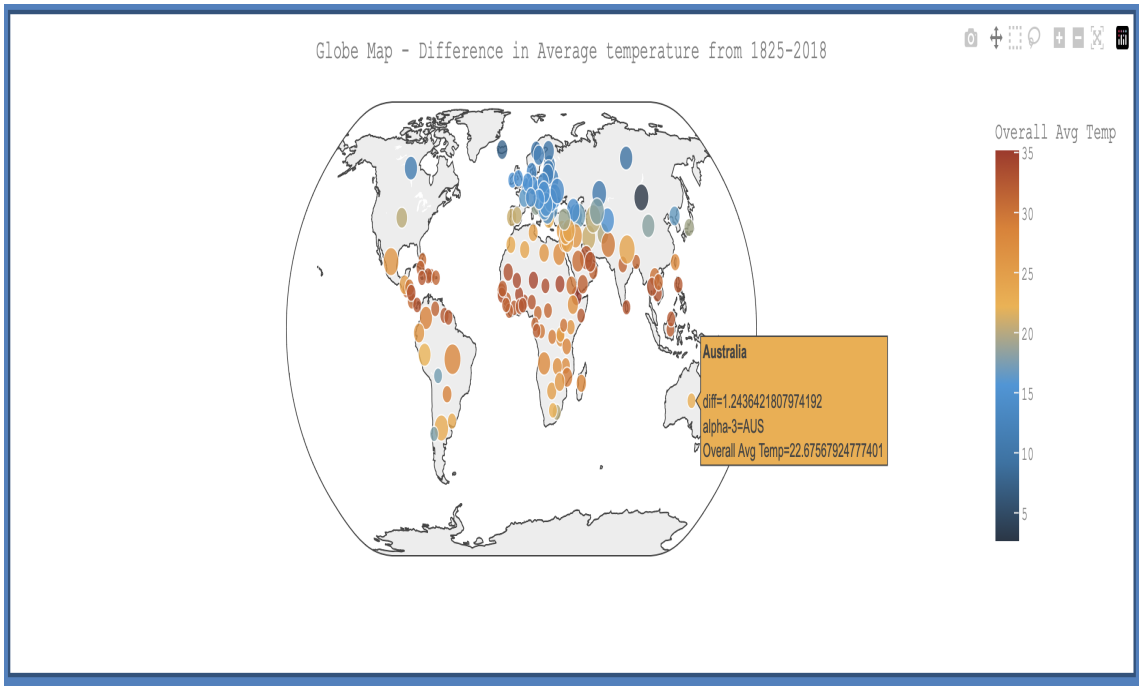


Figure 14

- **Difference Bar Chart:**

  The difference bar chart is another representation of the above 2-D Globe. It shows the ranks of the countries experiencing the difference in temperature from highest to lowest. From the picture, we can conclude that Brazil has experienced the highest raise in temperature since 1825 to 2018.

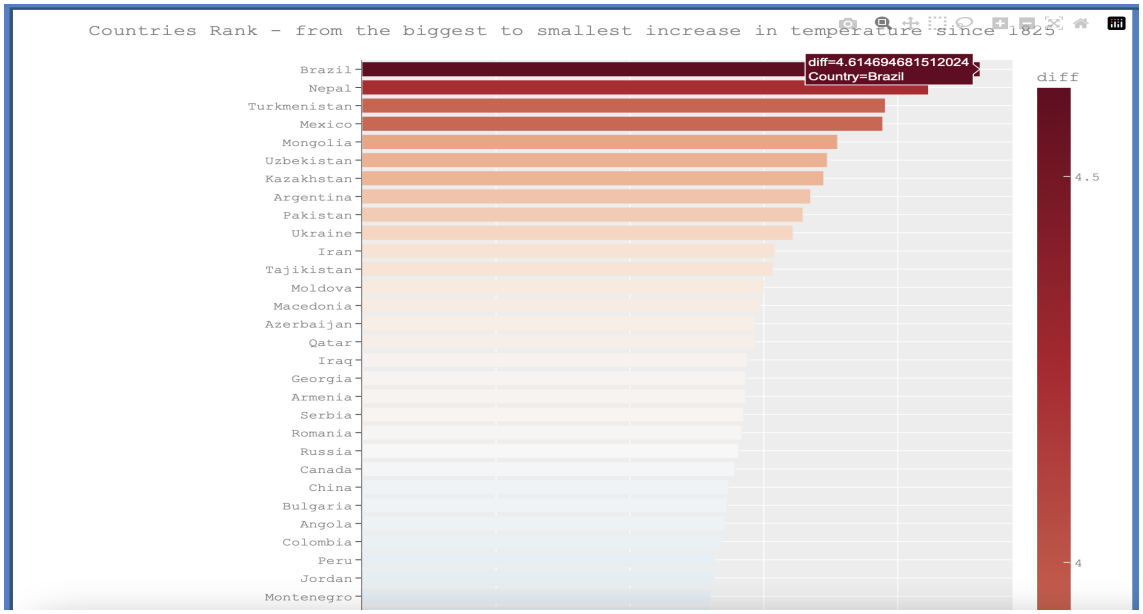


Figure 15

4.3 Visualization: Carbon Emissions

This dashboard consists of plots, figures and animations displaying the amount of carbon emissions by countries.

- **Scatter Plot:**
  - A scatterplot is a type of graph that is used to display the relationship between two quantitative variables. It is a set of points that are plotted on a two-dimensional plane, with each point representing a pair of values from the two variables.
  - We have the years from 1990 to 2020 on the x-axis and the amount of carbon emissions in MTCO2e on the y-axis.
  - We can select the desired countries for which we want to see the emissions. We have given a drop-down for selecting the countries and can also de-select a country.

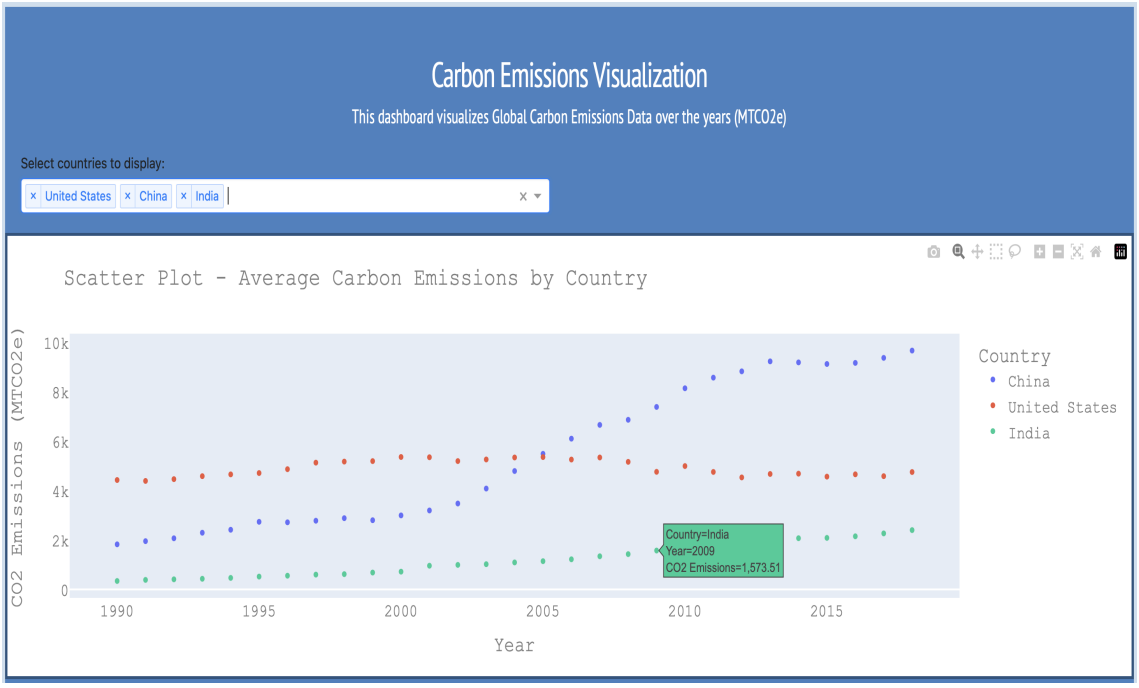


Figure 16

- **Bar Chart Race:** This bar chart displays a "race" between the top ten carbon emitting countries. It is an animation how, say China, beat USA to become the world's top carbon emitting countries over the years.

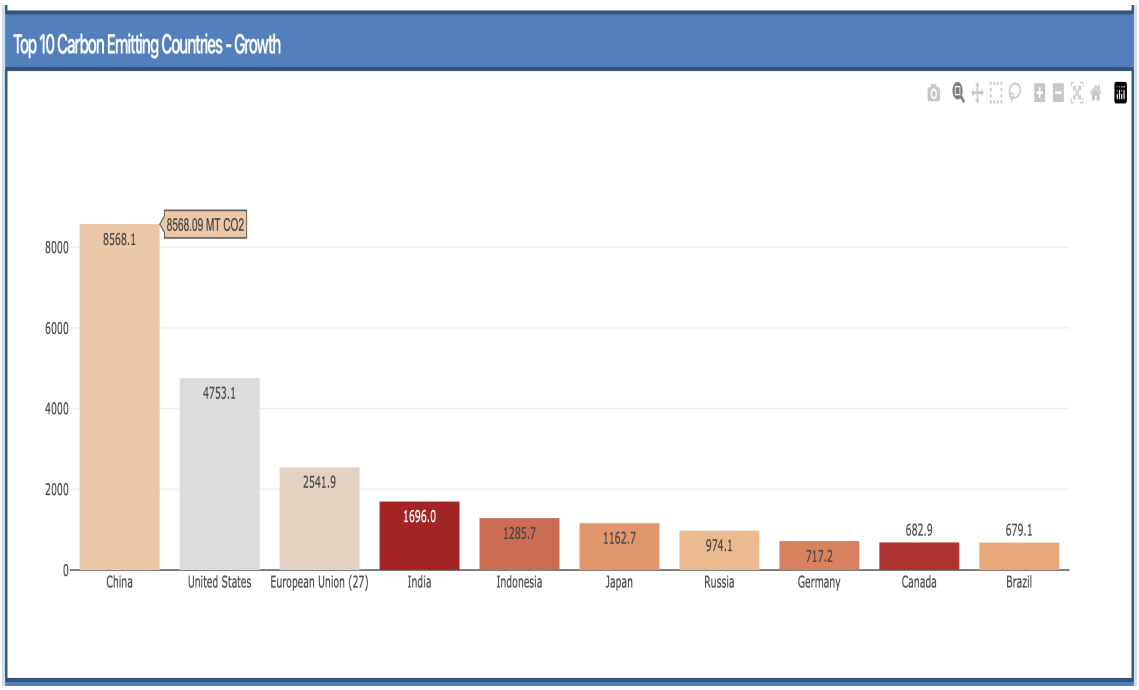


Figure 17

- **Bar Chart:**
  - This bar chart shows the top five carbon emitting countries in the world from 1990 to 2018.
  - The thickness of the boxes in the chart shows the amount of carbon content in that particular year.

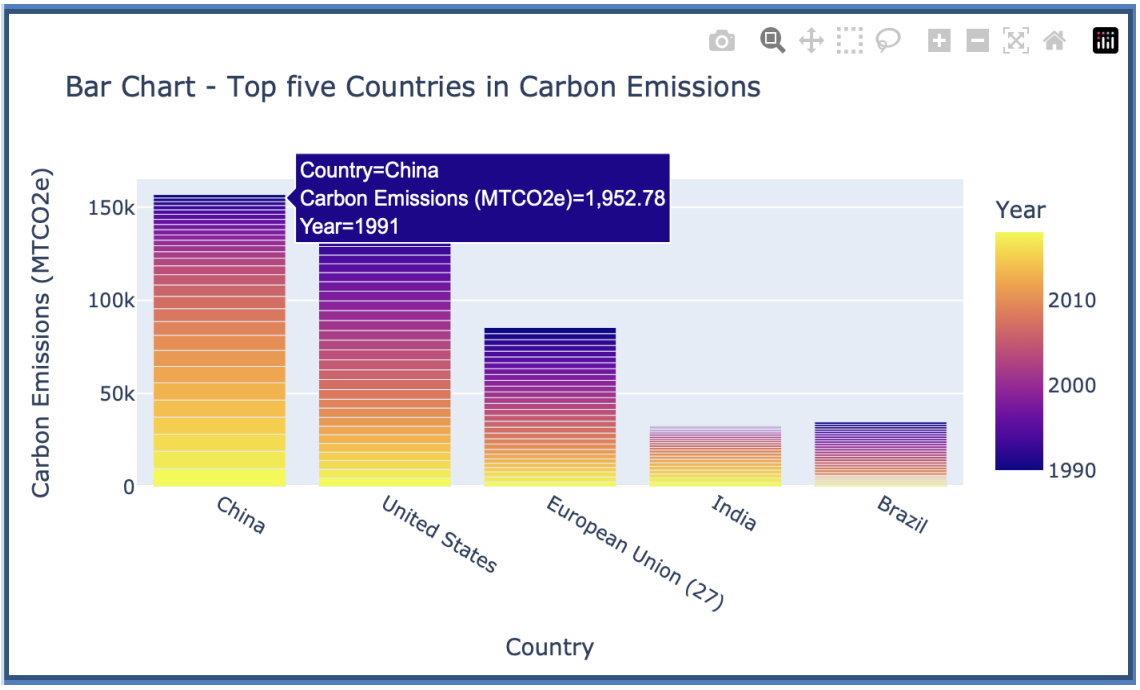


Figure 18

- **Bar Chart:**
  - This bar chart shows the bottom five carbon emitting countries in the world from 1990 to 2018.
  - The thickness of the boxes in the chart shows the amount of carbon content in that particular year.
  - As we can see that the carbon emission values are in negative, this shows that the country is not emitting any carbon content in the environment, instead it is removing the carbon content from the environment by green practices such as reforestation.

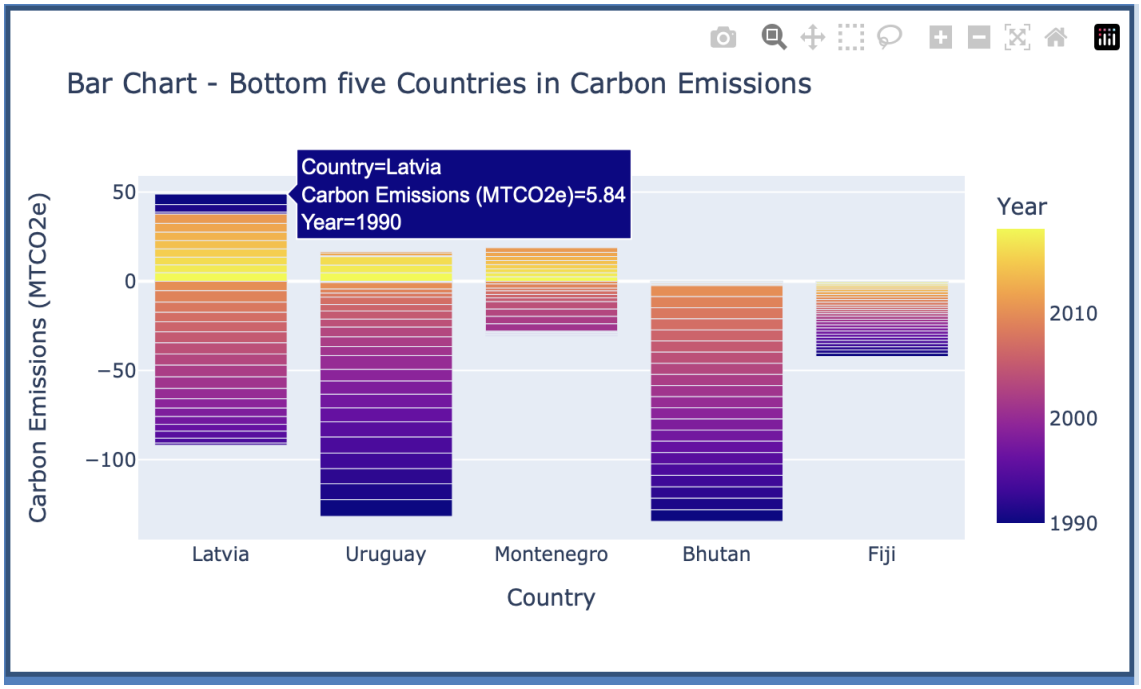


Figure 19

- **Bubble Plot:**
  - The Bubble Plot is an animation showing the growth in carbon emissions of different continents over the years.
  - On the x-axis, we have the amount of carbon emissions and on the y-axis, the years from 1990 to 2018.

Figure 20

- **Line Chart:**
  - This line chart shows the top five carbon emitting countries in the world from 1990 to 2018.
  - We can toggle the graphs on or off as per our requirements.

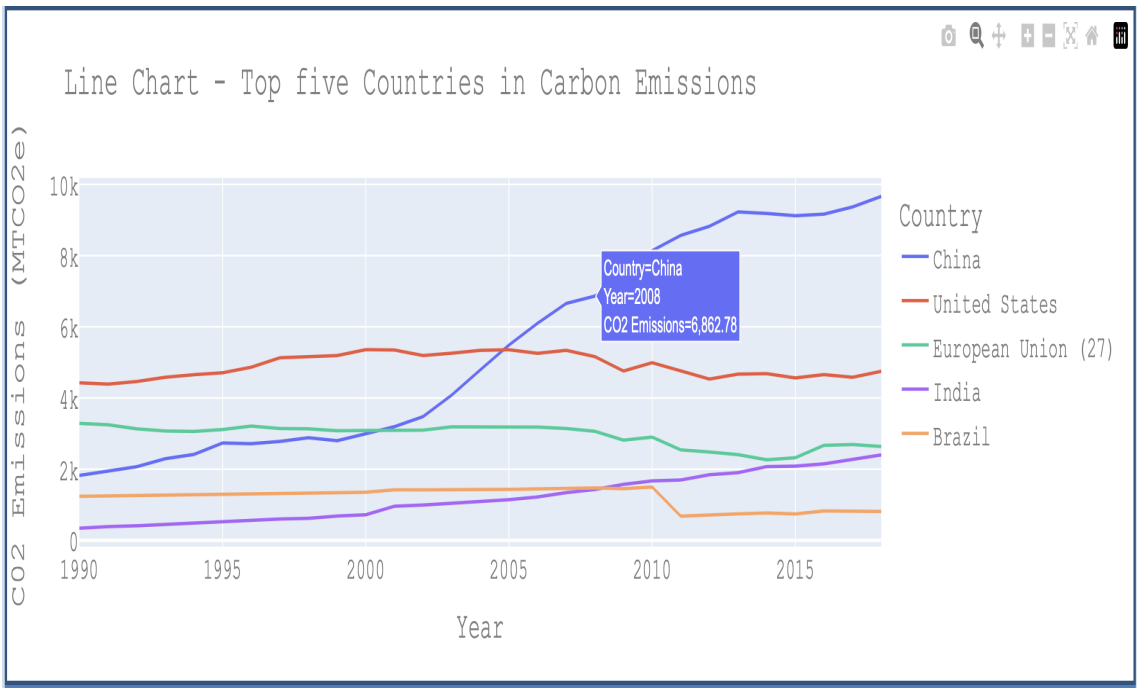


Figure 21

- **Line Chart:**
  - This line chart shows the bottom five carbon emitting countries in the world from 1990 to 2018.
  - We can toggle the graphs on or off as per our requirements.

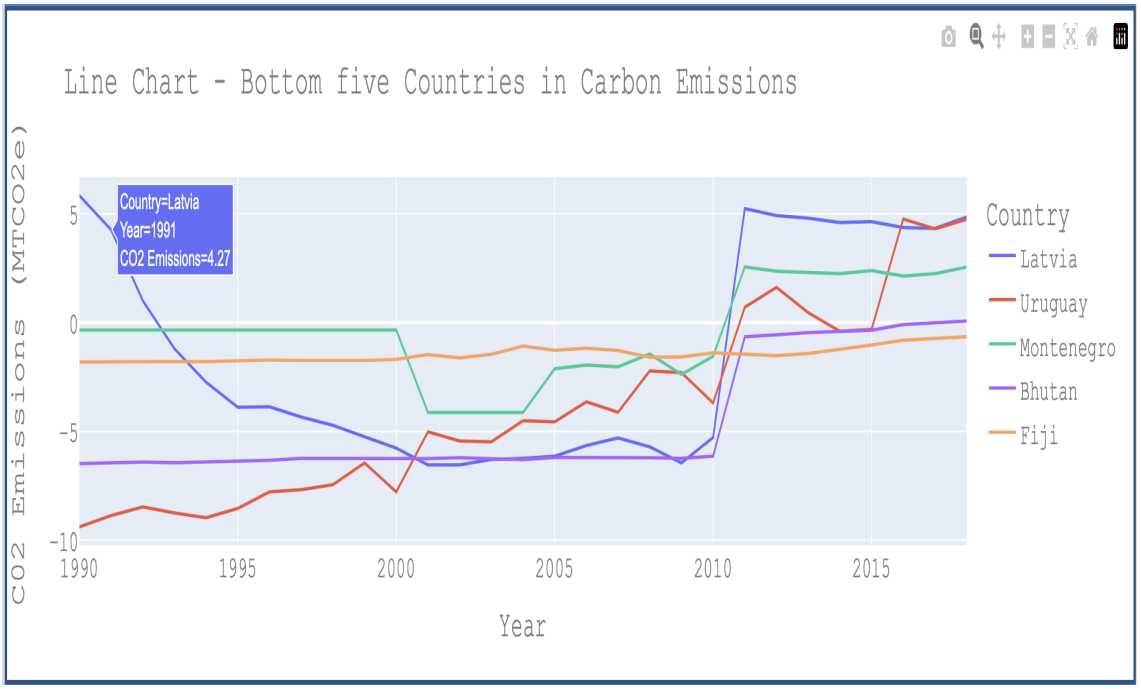


Figure 22

- **Choropleth Map:**
  - The choropleth map shows the carbon content that the countries have emitted over the years 1990 to 2018 according to the colorscale.
  - The years are given in the slider.

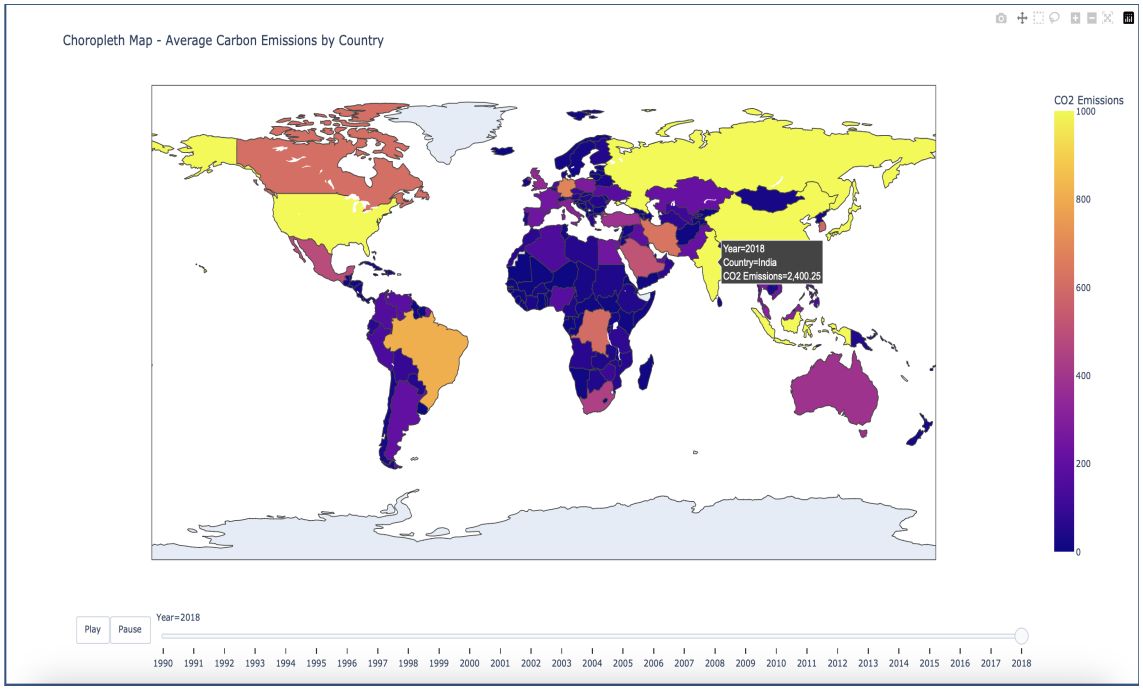


Figure 23

- **Heat Map:**
  - The Heap map shows the carbon content that the countries have emitted over the years 1990 to 2018 according to the colorscale.
  - The years are given in the slider.

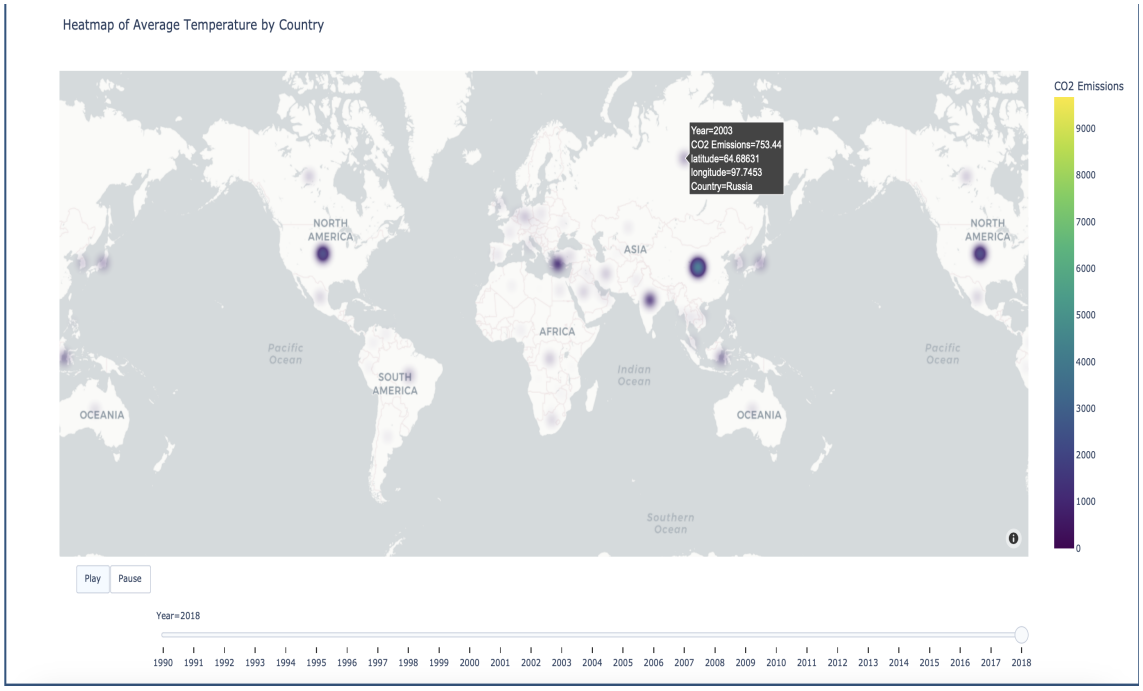


Figure 24

4.4 Visualization: Sea Level

This dashboard consists of plots and figures displaying the global rise in sea level. We will see that the values of sea level will be in negatives, this is because that the data is measured keeping the mean sea level (MSL) as the reference. So if the global sea level is below the mean sea level, the value is considered to be negative.

MSL is the average height of the ocean's surface, calculated as the midpoint between high and low tides over a long period of time. When we talk about sea level being negative or positive, we are usually referring to the deviation of the local sea level from the mean sea level.

- **Bar Chart:**
  - The bar chart shows the rise in global sea level from the years 1880 to 2020. We have years on the x-axis and the value of sea level in mm on the y-axis.

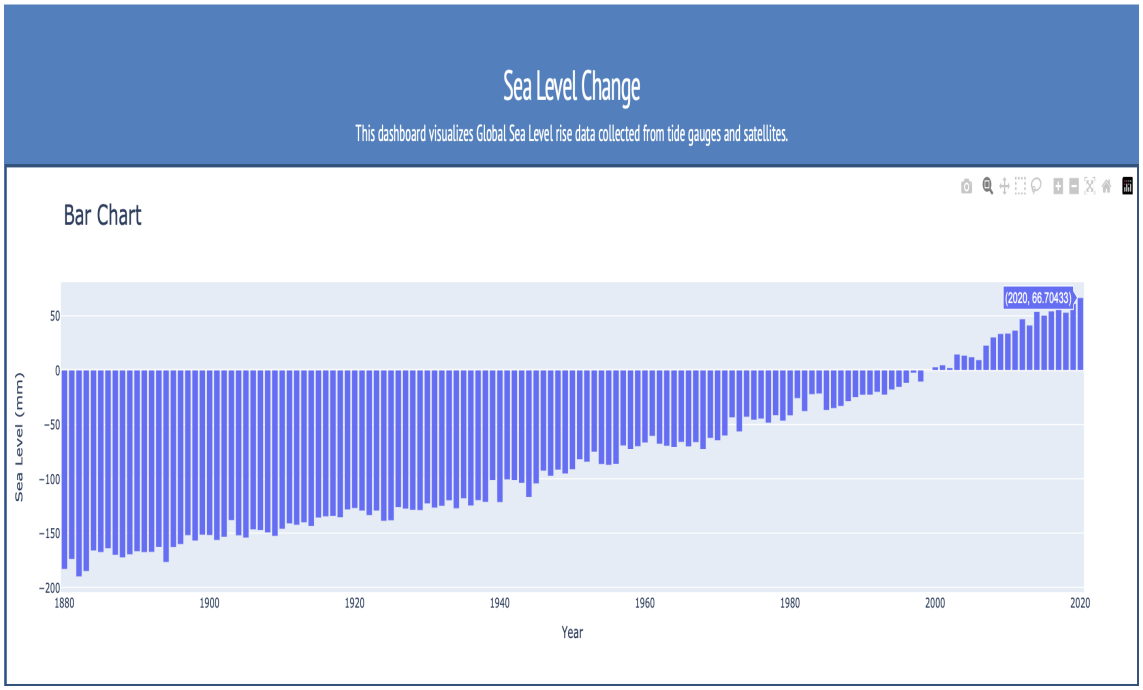


Figure 25

- **Scatter Plot:**
  - The scatterplot shows the rise in global sea level from the years 1880 to 2020. We have years on the x-axis and the value of sea level in mm on the y-axis.

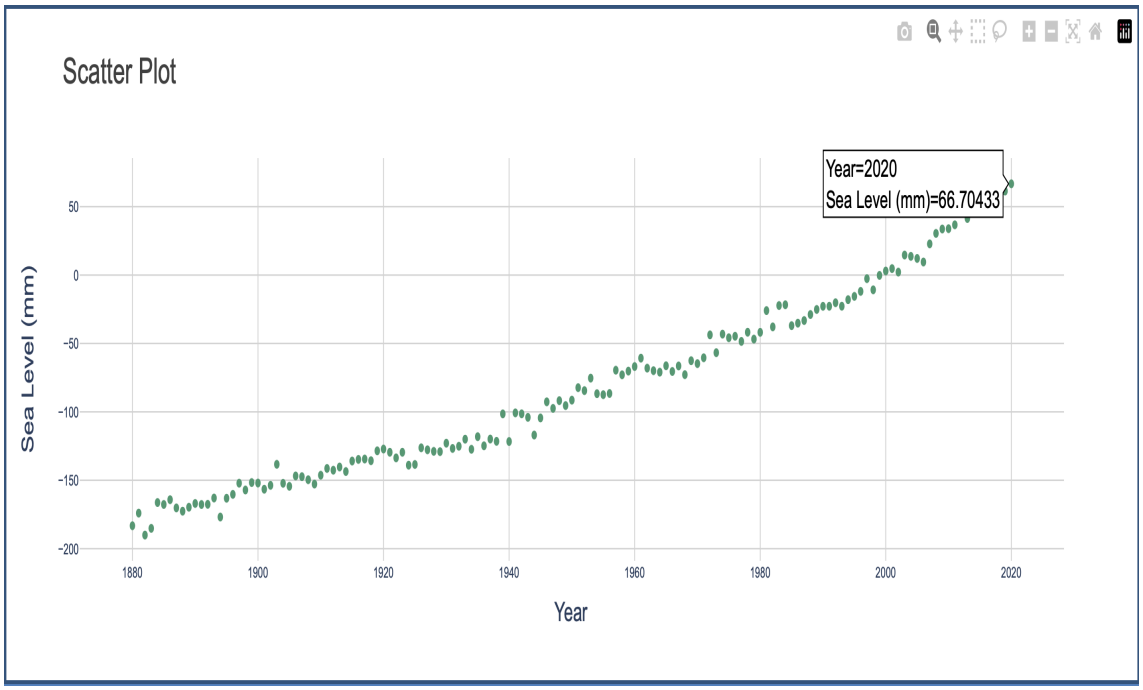


Figure 26

- **Box and Whiskers Plot:**
  - A box and whisker plot, also known as a box plot, is a type of graph that is used to display the distribution of a set of quantitative data. It is based on the quartiles of the data and consists of a rectangular box and two whiskers extending from the box.
  - The box in the plot represents the middle 50% of the data, with the top and bottom of the box representing the third quartile (Q3) and first quartile (Q1) respectively. The median (Q2) is represented by a horizontal line inside the box.

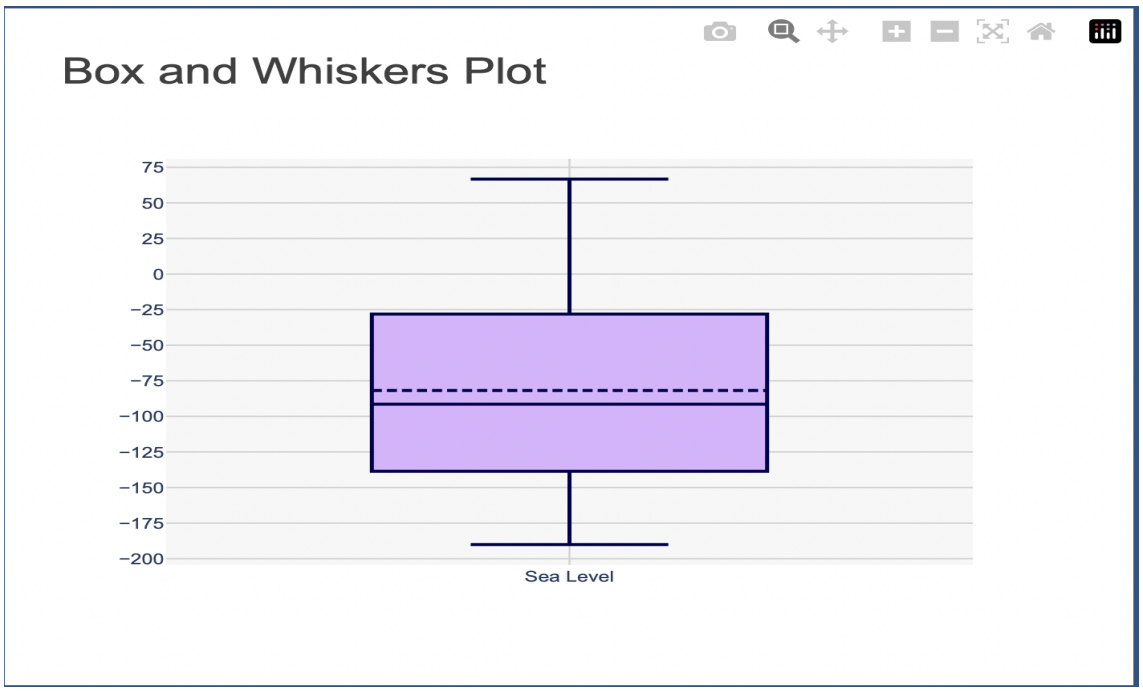


Figure 27

  - When we hover on the plot, we see all the information about the data: Q1, Median, Q3, Lower fence and the Upper fence.

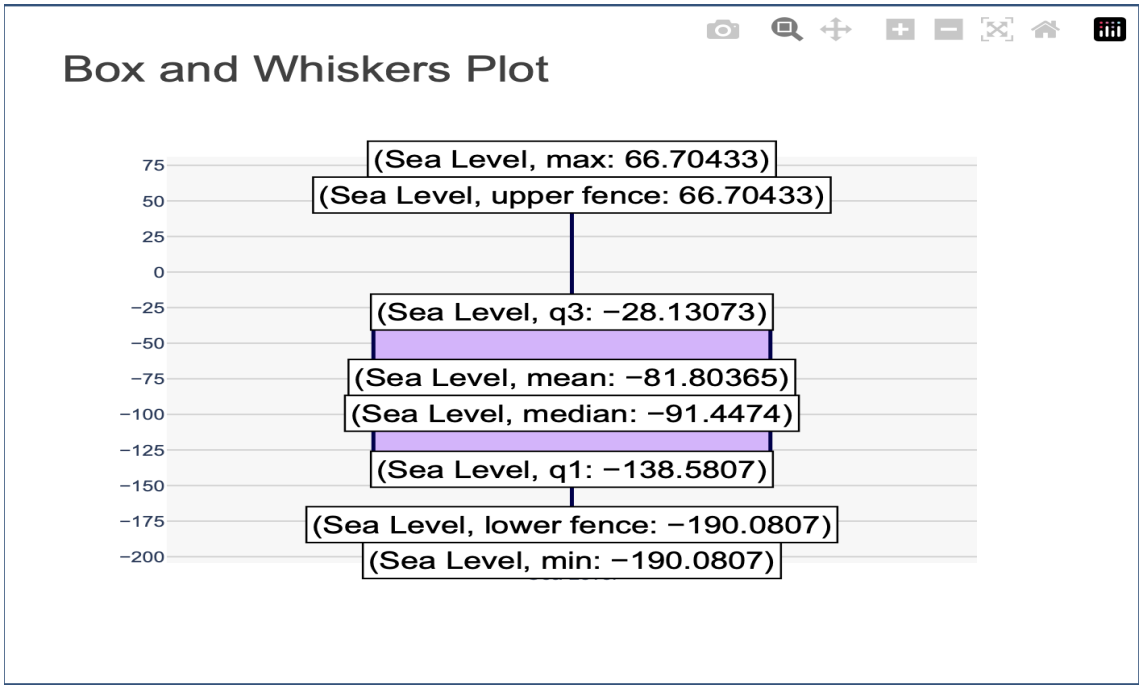


Figure 28

- **Area Chart:**
  - An area chart is a type of graph that displays data using a series of points connected by a line, with the area between the line and the x-axis filled with color or shading. Area charts are useful for visualizing trends and changes in data over time.
  - The x-axis represents time in years, while the y-axis represents the sea level. The area between the line and the x-axis represents the magnitude of the sea level at a given time, and the filled area makes it easier to see the magnitude of the data at different time points.

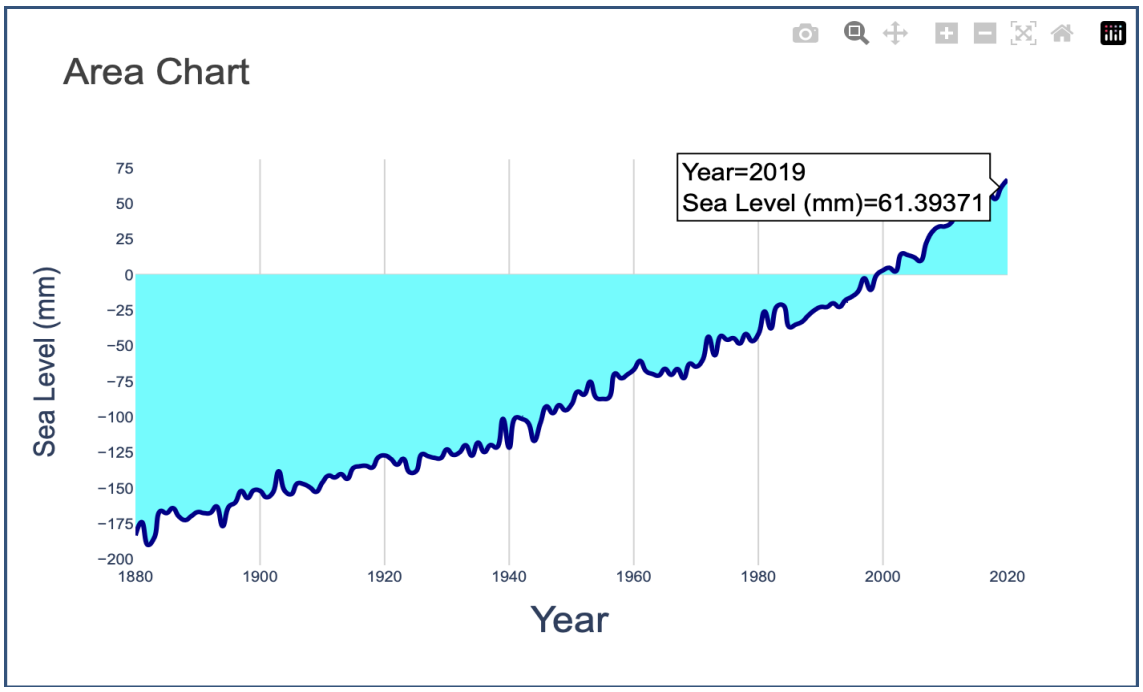


Figure 29

- **Line Chart:**

  The line chart shows the rise in global sea level from the years 1880 to 2020. We have years on the x-axis and the value of sea level in mm on the y-axis.

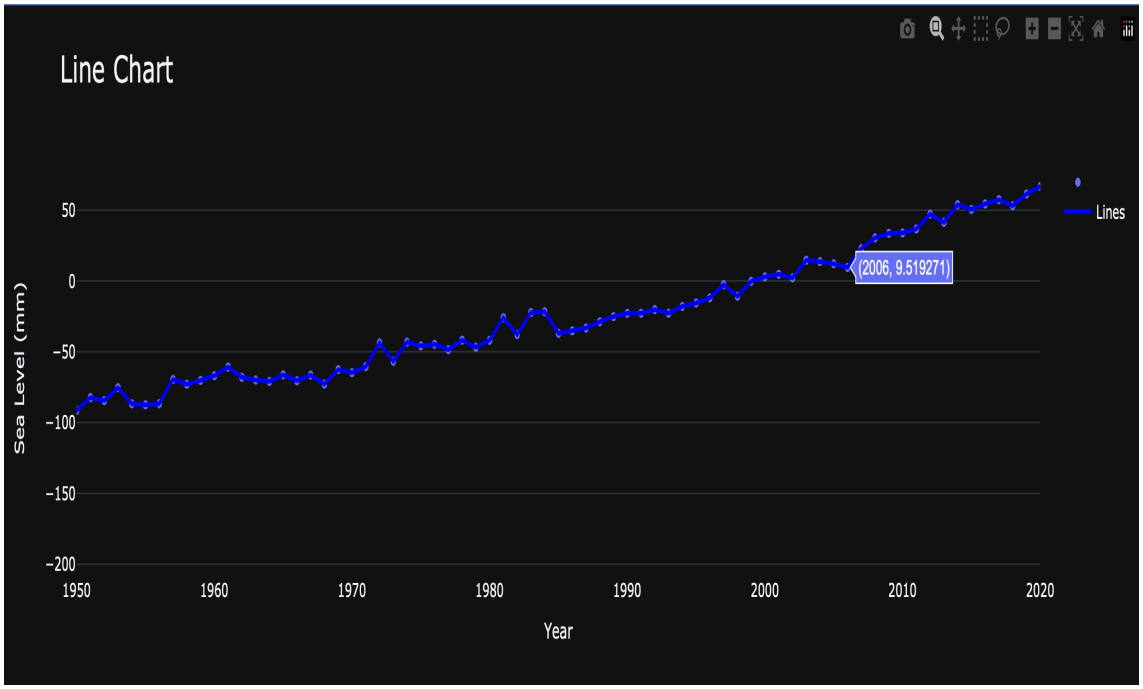


Figure 30

4.5 Correlation

This dashboard visualizes the correlation between the change in temperature, carbon emission and sea level. It consists of plots, figures and animations containing all the three variables in a single plot. We can toggle the plots on/off as per our requirements.

- **Stacked Bar Chart:**

  The bar chart visualizes the correlation between the change in average land temperature and the global temperature in a single plot.

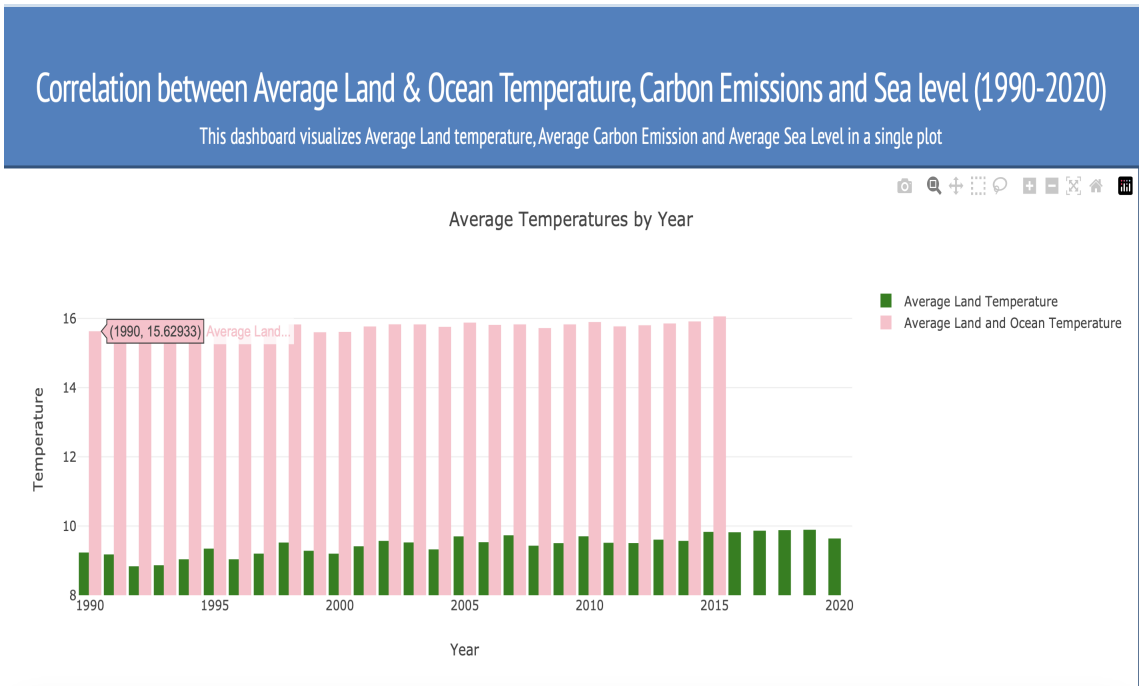


Figure 31

- **Line Chart:**

  The line chart visualizes the correlation between the change in average land temperature, carbon emissions and the global sea level in a single plot.

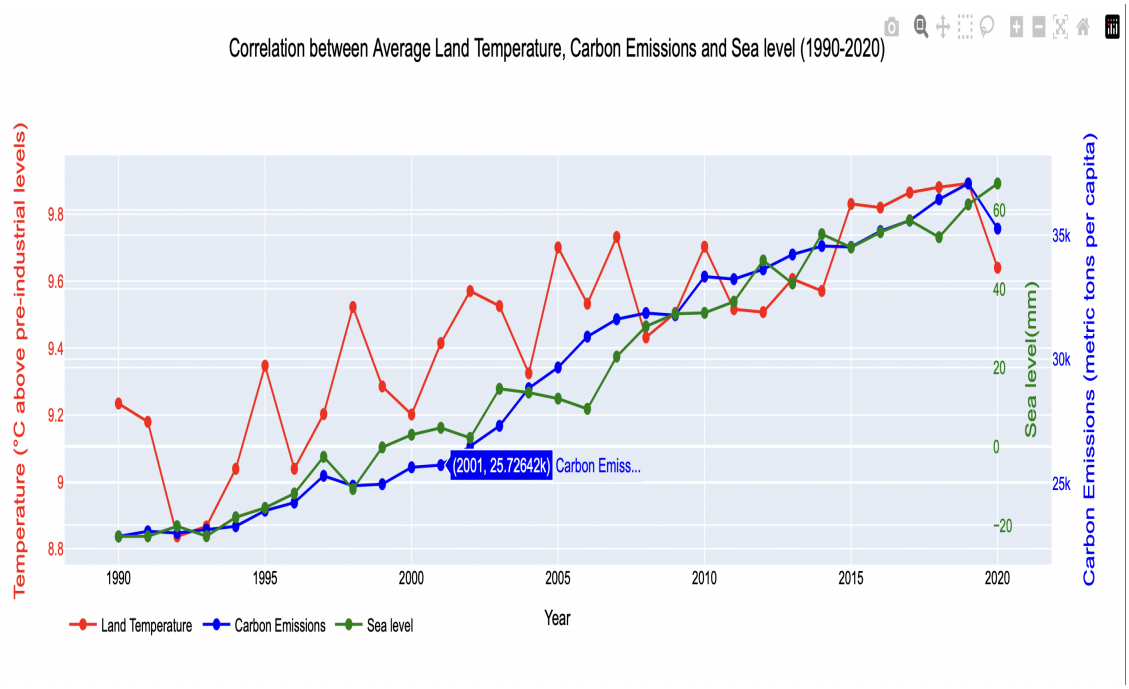


Figure 32

- **Stacked Bar Chart:**

  The bar chart visualizes the correlation between the change in average land temperature and the global temperature and average carbon emissions in a single plot.

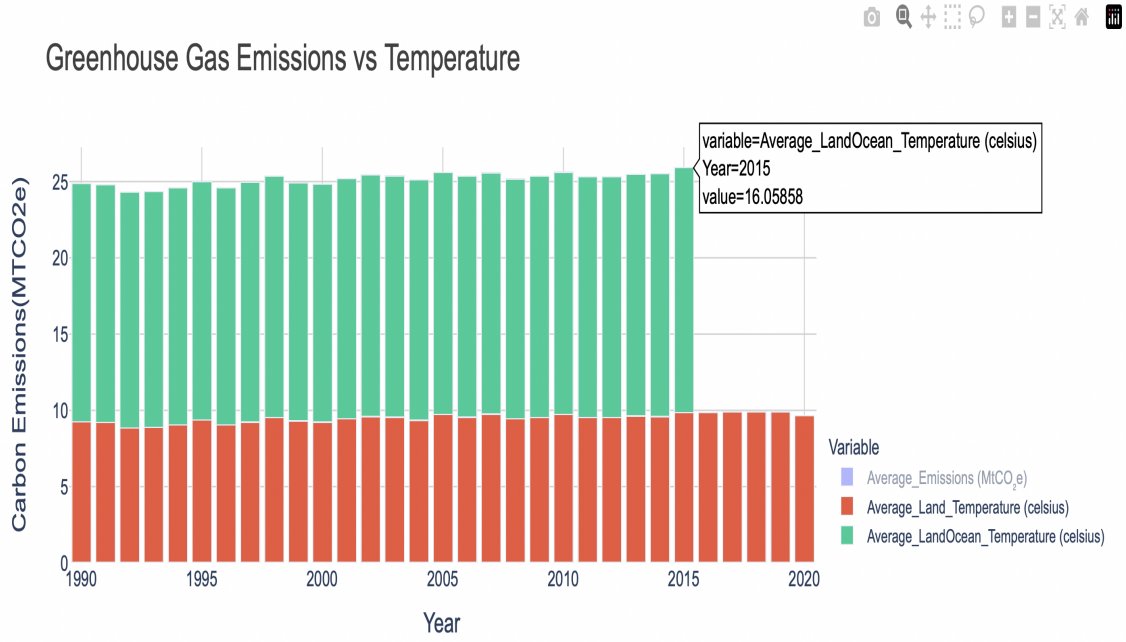


Figure 33

- **Line Chart:**

  The line chart visualizes the correlation between the change in average land and ocean temperature, carbon emissions and the global sea level in a single plot.

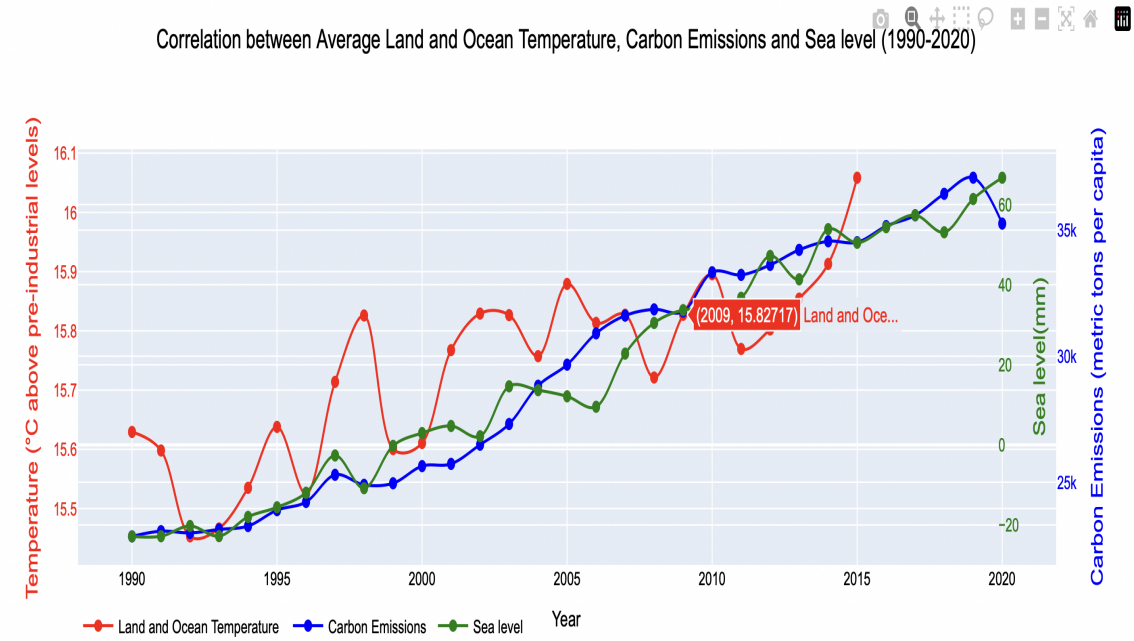


Figure 34

- **Scatter Plot:**

  The scatter plot visualizes the correlation between the change in average land temperature, carbon emissions and the global sea level in a single plot.

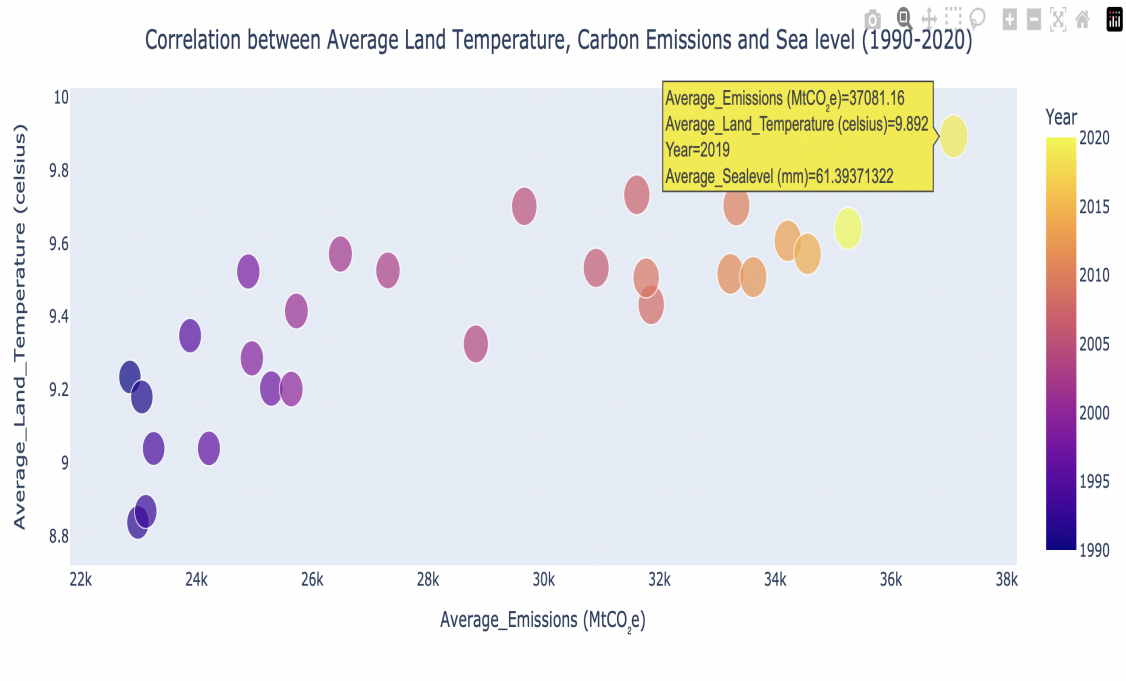
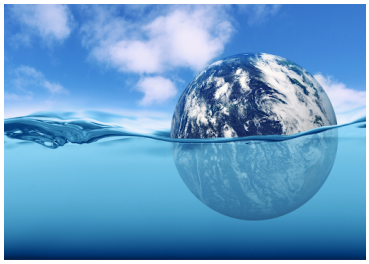


Figure 35

5 Scope

Climate visualization is a powerful tool that can help us better understand the complex patterns and processes that shape our planet's climate. With the help of cutting-edge data visualization techniques and advanced computer models, we can now create highly detailed and accurate depictions of climate patterns and trends on a global scale. These visualizations can be used to inform scientific research, climate policy, and public awareness campaigns.



- **Temperature:**
  - **Climate research:** Visualizing temperature data can help climate scientists better understand long-term temperature trends and patterns on a global scale. This can help them identify changes in temperature that may be linked to climate change and other factors.
  - **Agriculture:** The data visualization can help farmers and agricultural researchers monitor crop growth and predict yields. This can be especially useful in areas where temperatures vary widely throughout the growing season.

- **Public health:** Visualizing temperature can help public health officials monitor the spread of diseases that are affected by temperature, such as Lyme disease or West Nile virus. This can help them identify areas where outbreaks are more likely and develop targeted prevention strategies.
- **Energy management:** The analysis of temperature data can help energy companies and utilities better manage energy consumption and production. This can include monitoring temperature fluctuations in different regions to optimize the distribution of energy resources.
- **Environment monitoring:** The visualization of temperature data can help environmental scientists monitor changes in temperature and weather patterns over time. This can help them identify areas where changes in temperature are affecting natural habitats, such as coral reefs or forests, and develop strategies to protect these areas.

- **Carbon Emissions:**
  - **Identify trends and patterns:** Carbon emissions data can be analyzed and visualized to identify trends and patterns over time. This can help researchers and policymakers understand the drivers of emissions, and track progress in reducing them.
  - **Informing policy decisions:** Visualization of carbon emissions can help policymakers understand the impact of different policies on emissions levels. For example, visualizing the impact of a carbon tax or cap-and-trade system can help policymakers determine the most effective approach for reducing emissions.
  - **Tracking progress:** Carbon emissions visualizations can be used to track progress towards emission reduction goals, such as those set forth in the Paris Agreement. This can help stakeholders monitor progress and adjust strategies as needed to achieve these goals.

- **Sea Level:**
  - **Coastal planning:** Sea level rise poses a significant threat to coastal communities around the world. Visualizing sea level can help coastal planners and engineers understand the extent of the risk and develop strategies for adapting to rising sea levels. For example, visualizations can help identify areas that are vulnerable to flooding, evaluate the effectiveness of different flood protection measures, and prioritize investments in infrastructure and land use planning.
  - **Spreading awareness:** Visualizing sea level can also help raise public awareness about the impacts of climate change on coastal communities. By creating engaging and accessible visualizations, we can help people understand the potential consequences of rising sea levels, such as increased flooding, erosion, and storm surges. This can encourage people to take action to reduce their own carbon footprint, support climate policies, and advocate for the protection of vulnerable communities.

6 Conclusion

- We visualized the change in earth's temperature, carbon emission and sea level and studied the correlation between them.
- The change in these variables are interconnected. One is cause and the other is effect.
- As we emit a lot of carbon content in the environment, temperature of the environment increases. This is due to the trapping of heat coming from sun by carbon-dioxide. Due to the rise in temperature, the gobal sea level rises. So, carbon emissions are the cause and rise in global sea level is the effect.
- Ultimately, rise in average temperature of the earth is referred as Global Warming.
- **Some more insights:**
  - Climate change is real! We often talk about summers in the current year being hotter than the previous years. We can actually feel this gradual rise in the temperature every year.
  - Human activities are driving this change, particularly the burning of fossil fuels like coal, oil, and gas, have been releasing large amounts of greenhouse gases, including carbon dioxide, into the atmosphere. These gases trap heat from the sun and prevent it from escaping into space, causing the Earth's temperature to rise.
  - The climate change impacts are global and varied. Some parts of the globe are affected more than the other, for example, Brazil has faced the highest temperature rise since 1885.
  - Actions must be taken to stop the temperature rise and mitigate the impacts of climate change. This requires a multi-faceted approach that involves both reducing greenhouse gas emissions and adapting to the changes that are already underway. Some of the key actions that can be taken include:
    * **Transitioning to renewable energy sources:** This involves reducing our reliance on fossil fuels and increasing the use of renewable energy sources, such as solar, wind, and hydropower.
    * **Improving energy efficiency:** This involves using energy more efficiently in buildings, transportation, and industry to reduce the amount of energy needed to power our societies.
    * **Reducing emissions from transportation:** This involves promoting the use of low-emission vehicles, public transportation, and active transportation modes such as biking and walking.
    * **Promoting carbon capture and storage:** This involves capturing carbon dioxide emissions from power plants and other industrial processes and storing them underground or in other ways.
    * **Adopting sustainable agriculture and forestry practices:** This involves reducing deforestation and promoting reforestation, as well as reducing emissions from agriculture and promoting sustainable farming practices.

* **Adapting to the changes that are already underway:** This involves preparing for and responding to the impacts of climate change, such as sea level rise, increased flooding, and more frequent and severe weather events.

7 Link to source code

- Click here to see our github repository: Repository
- The repository consists of complete python code of the interface along with the datasets. A few data files are missing from the github because of the size constraints. Missing files include:
  - GlobalLandTemperaturesByState.csv (30.8 MB)
  - GlobalLandTemperaturesByCity.csv (532.8 MB)
  - city_temperature_2.csv (140.6 MB)
  - city_temperature.csv (140.6 MB)
  - UpdatedCity_temperatures.csv (605.1 MB)
- The file Integration_dash.ipynb is for running the interface.

8 Project Log

Member	Contribution
Drashtant Singh Rathod	Filtered temperature datasets, created temperature visualizations: Heatmap, Maps of 6 countries, Choropleth Map, Correlation of temperature, sea level and carbon emissions, Correlation matrix, Bar Chart Race, Integrated visualizations to the plotly dash interface, Styled the interface by applying css style components.
Mannu Kumar Gaddhyan	Filtered historical_emissions.csv, created all the visualizations of Carbon emission dataset: Scatter Plot, Choropleth Map, Two Bar Graphs and two Line Charts of top 5 and bottom 5 countries in carbon emission and Heat Map.
Sudiksha Navik	Filtered Global_sea_level_rise.csv, created average_dataset.csv, created visualizations of sea level: Bar Graph, Scatter Plot, Box and Whiskers Plot, Line Chart, Area Chart, Bubble plot, Correlation between Avg. Land temp and Global temperature, Heat Map, Earth's temperature timeline, Integrated visualizations to the plotly dash interface.
Susovan Patra	Filtered temperature datasets, created temperature visualizations: Globe representation, Choropleth Map, Rank of countries a/c to temperature difference, Line Charts based on continents, Scatter Geo Plot, Calendar Heat Map, Scatter Plot Correlation.
Aditya Raj	Filtered temperature dataset, created correlation graph between average land and ocean temperature, sea level and global carbon emissions, line graphs and handled missing and duplicated data in the used datasets.

References

[1] Plotly Dash. *https://dash.plotly.com/*. Accessed: 2023-03-19.

[2] Python Dashboard using Plotly Dash. *https://medium.com/geekculture/python-dashboard-using-plotly-dash-example-1-b9cc94039b5e*. Accessed: 2023-03-19.

[3] Climate Change: Earth Surface Temperature Data. *https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data*. Accessed: 2023-03-08.

[4] Carbon Dioxide Emissions of the World(1990-2018). *https://www.kaggle.com/datasets/ankanhore545/carbon-dioxide-emissions-of-the-world*. Accessed: 2023-03-08.

[5] Global Sea Level Rise. *https://www.kaggle.com/datasets/jarredpriester/global-sea-level-rise*. Accessed: 2023-03-08.

[1] [2] [3] [4] [5]