Team 109 - Global Temperature Trends and Impact of CO2 and Deforestation on Temperature Change

Priyanka Bhardwaj, Gopi Chava, Imren Dinc, Phung Vy Luu, Sachin Sharma, Geneva Vezeau
*All team members have contributed a similar amount of effort.

1 INTRODUCTION

Climate change is one of the biggest threats in modern times, severely impacting Earth's ecosystem. The idea of climate change through global warming was first introduced in 1896 by Swedish scientist Svante Arrhenius [17]; however, it was not until 1988 that it became a worldwide topic of conversation as there was now proof that global warming was actively occurring [6]. Since then, it has been a "hot" topic for many researchers, with the majority of the work focused on high CO_2 emitting countries.

2 PROBLEM DEFINITION

After reviewing available literature, we identified that CO_2 and deforestation are two major factors impacting climate change. In this project, we will analyze temperature change for the last 250+ years and investigate effects of the two factors globally using machine learning (ML) and interactive visualization. We will use our analysis to predict the temperature change and impact of CO_2 emissions and deforestation for the next few decades based on historical datasets.

3 LITERATURE SURVEY

In this section, we summarize the studies done in climate change, deforestation, and CO_2 emissions. We started with review papers for each of our main topics. This helped us get more specific papers for our project's objective. These review papers gave us a broad idea of what has been studied in each of our main focuses so we could refine our study objectives.

Since the topic of climate change has been such a widely known issue, there have been many studies on its causes and effects and mitigation strategies. This helped us to get a starting point by building on what has been done in past studies. According to [11], the use of big data and ML methods enabled researchers to create specific solutions to climate change, opening our eyes to the use of big data and ML. [3] compared global climate change by using the latitudinal shift as a factor, predicting until 2050. This will be a good comparison for us as we are predicting until the same year, using a different method. [14]'s study shows an in depth climate shift for Europe; some of these ideas can be helpful when we are creating our models for European countries. Many recent studies were about ways to mitigate climate change, which involved impacts on the ecosystem [10] [15]; leading us to further research deforestation and its effects on climate change.

Several studies concluded that deforestation impacts climate change. Many of the papers mentioned the surface of the Earth cooling from the presence of forests or heating due to deforestation [4] [7] [13] [12]. [1]'s approach to predicting deforestation's impact on climate change used wet-bulb globe temperature estimate. However, this study was specifically for the Brazilian Amazon; while this worked for this region, it may not be useful for us when we expand to other countries. [16] investigates the impact of deforestation on climate change for different climates which will help us when we take a global approach.

Climate change has been linked to the greenhouse effect since its inception. CO_2 emissions are just one of the many gases that are associated with creating the greenhouse effect and causing climate change [19]. For CO_2 emissions studies, the majority of published papers are specific to a certain region; China [9], Iran [5], India [8], coastal areas [20] and developing ASEAN countries [2]. According to [19], CO_2 emission studies lack ML estimation and the global approach. In our project, we used several ML methods to evaluate performance. [18] applied the logarithmic mean Divisia index (LMDI). This identified major contributors to CO_2 emissions but does not relate to temperature change. [9]'s approach was to use demographic factors to predict CO_2 emissions while an innovative approach, this is difficult to use on a larger scale project like ours.

Through our research, we evaluated different ML approaches to model our data to reach our project objective.

4 PROPOSED METHOD

We will predict the temperature change for the next few years and the impact of CO_2 emission and forest area on temperature change. We evaluated multiple models that can fit the dataset and finalized two different approaches.

4.1 Intuition

The performance of the classifiers or models depend on dataset. Different type of datasets (i.e. time series, graph datasets, etc.) may require different models to fit. Therefore, in this study, we performed empirical study on different models to discover the best fit for our dataset. Below are the methods we chose to use for our two use-cases and why for our analysis on predicted temperature change.

4.1.1 SARIMAX. The first use-case is to predict the temperature based on the historical country-specific dataset. The global temperature is a time series dataset. ARMA and its variant models are common methods for time series datasets. We used *pmdarima* library to figure out which ARMA model fits the best based with our datasets and fine tuned the hyper-parameters to fit the country-specific datasets. The automated process calculated SARIMAX as the best fit. SARIMAX or Seasonal Autoregressive Integrated Moving Average eXogenous model is a combination of multiple approaches - Auto Regression and Moving Averages with seasonality and external factors. It also has a difference order (I) which is the number of transformations needed to make the data stationary. Below is the general forecasting equation:

$$\hat{y} = \mu + \phi_1 + y_{t-1} + \ldots + \phi p + y_{t-p} - \theta_1 \epsilon_{t-1} - \ldots - \theta_q \epsilon_{t-q} \tag{1}$$

where p, d, and q are autoregressive, integrated, moving average parameters, respectively.

4.1.2 Artificial Neural Networks (ANN). The second use-case for our project is to predict the impact of CO_2 emission and forest area on temperature change. We use neural network for this use-case. Artificial neural networks (ANNs) are a type of machine learning algorithm that imitates the structure and function of the human brain. ANNs consist of interconnected nodes or "neurons" arranged in layers. Each neuron receives input signals, processes them, and produces an output signal as a result of an activation function. Finally, outputs are passed on to the next layer of neurons. ANNs are trained on a set of labeled data, adjusting the strength of connections between neurons to learn patterns and make predictions on new data. Specifically, in this study, we used Multilayer Perceptron (MLP) with linear activation function where $f(x) = x$ and x is the input of the neuron. MLP is a type of feed-forward neural network where the neurons are arranged in multiple layers, including an input layer, one or more hidden layers, and an output layer. We use ANN for regression. We use CO_2 emission and forest area data for each country as features for the model and country average temperature as the dependent variable.

4.2 Visualization

Data and facts alone are not sufficient to influence people's belief on climate change [3]. Since there are significant risks associated with climate change, it is necessary to use effective visualization to attract public attention. Therefore, we develop several visualizations to present our findings.

In our project, we use Tableau and D3 for interactive visualizations of the temperature data predicted using the two approaches outlined in the previous section. We implemented a few visualizations to present our experimental results from different perspectives.

First, we used the Tableau map layers feature along with an interactive date slider filter to change the year of interest to visualize the global temperatures forecast using the SARIMAX model. This visualization shows the global temperature change over time as the user moves the date slider forward. The data shown in this visualization contains both historical data used for model training as well as the temperature forecast by 2043. In order to emphasize the temperature change regionally, we developed another visualization using D3 with an auto-play feature for years from 1890 to 2043. This visualization has the ability to show temperature increases by quarter degree compared to a base year (1890). Therefore, users can see the temperature change over time regionally.

The second visualization is based on the temperature prediction from ANN models that use CO_2 emission and forest area as features. We use Tableau map layers along with the drop-down option for interactive selection of CO_2 emission and forest area changes which allow visualizing global temperature change based on chosen feature values.

4.3 Innovation

We identified during the literature survey that most of the studies are done for specific countries and regions since temperature shows different patterns for different parts of the earth. It is challenging to predict the temperature for all the countries in the world. In our project, we evaluate the impact of both factors on temperature trends in different countries globally with an interactive visualization tool. Our contributions can be summarized as follows:

(1) We predicted the temperature for most countries around the globe based on the available historical city level temperature data using a time series model.
(2) We created different models for each country by tuning the hyper-parameters based on the country-specific data, and then predicted temperature based on the country-specific models.
(3) We predicted the impact of CO_2 emission and forest area data on temperature for most of the countries using neural network.
(4) We trained country-specific models to evaluate the impact of CO_2 emission and forest area on temperature change.
(5) We created interactive visualizations to show country-specific temperature predictions and CO_2 emission and forest area impacts on temperature using a world map in Tableau.
(6) We developed an interactive visualization to show temperature increases by quarter using D3 with auto-play feature for years from 1890 to 2043.

Future Work: In this project, we only focus on predicting the temperature change for different countries across the globe to create awareness through interactive visualizations. We believe global view focusing on countries specific detail will create a lot of impact and awareness. The future work will include identifying the methods and processes to mitigate the impact of temperature change.

5 EXPERIMENTS & EVALUATION

In this section, we present experimental results collected from our dataset using the proposed method.

5.1 Experimental Questions

In our experiments, we aim to answer the following two questions:

(1) whether there is global trend in temperature changes,
(2) whether CO_2 emission and forest area have significant impact on climate change.

5.2 Dataset

In this paper, we used global temperatures, CO_2 emission, and forest area datasets as explained below.

5.2.1 Global Temperatures : This dataset has the temperature details in Celsius starting for mid 1700s until 2013 for most major cities for 150+ countries across the world. It has more than eight million records. The granularity of this data is one record per month per city. With this dataset, we calculate the mean of the monthly average temperature per year for each country. The Dataset is available at *Global Temperatures Dataset*.

5.2.2 CO_2 Emission: This dataset has CO_2 emissions measured by metric tons per capita starting from 1990 until 2019 for all countries. The dataset has one record per country with each column showing one recorded year. The data is available at CO_2 *Emission*

5.2.3 Forest Area: This dataset has the forest area details in hectares at regional and country level from 1990 until 2020. The granularity of this dataset is at country and year level. To use this dataset, we filter only country data and remove any record that does not have area data. The dataset is available at *Forest Area*

5.3 Data Analysis and Preprocessing

This section summarizes the data analysis and preprocessing that we perform before running our experiments.

5.3.1 Temperature Prediction: The global temperature dataset has temperature for major cities from November 1743 until September 2013, but a few countries do not have data from 1743. In addition, some countries have missing temperature data for a few months. In our experiments, we dropped all null temperature records and kept the cities with data available for all twelve months. During our experiments, we identified that there are a lot of differences in temperature ranges for different countries. Some cities in a country might have completely extreme temperatures that differ from the rest of the country (i.e. Alaska vs Texas). Figure 1 shows the temperature box plot of some major countries.

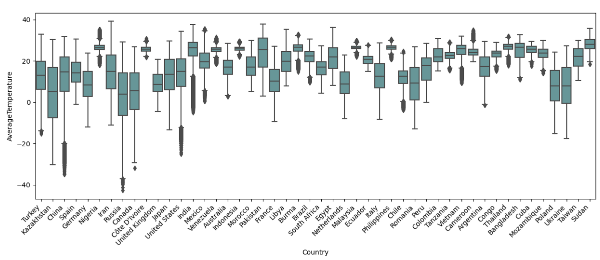


Figure 1: Countries Temperature Difference

As it can be seen from the figure, the temperature shows different distributions in different countries, making it challenging to create one model that can be used to predict the temperature for all countries. Therefore, we train our models for each country in the dataset to obtain the best model fit. There are multiple records for the same month for different cities for a specific country, which cannot be used in time series modeling. Therefore, we calculated the yearly mean temperature at country level to prepare the time series data.

We further analyzed the data to check if the time series data is correlated for the model using autocorrelation for a few major countries. According to our analysis, the data is fairly correlated between 5-30 years. We identified the best ARIMA model and hyper-parameters to fit the country-specific data using *pmdarima* library to get the best model fit.

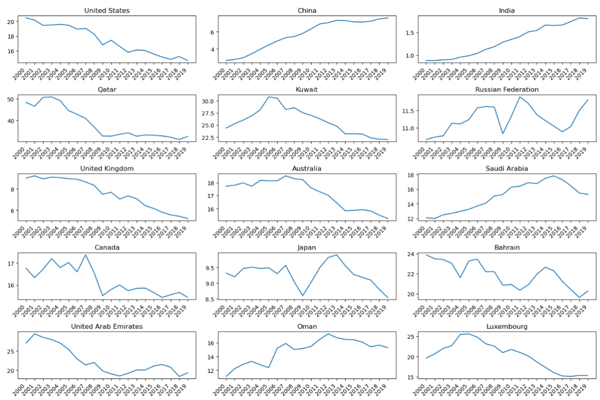


Figure 2: CO2 Emission for Major Countries

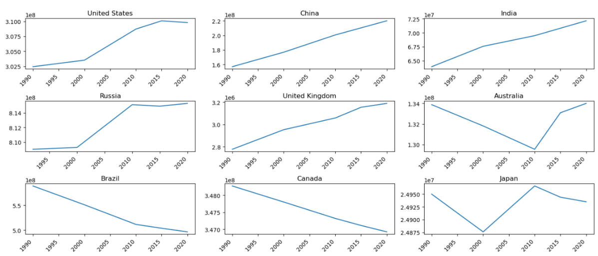


Figure 3: Forest Area for Major Countries

5.3.2 Impact of CO_2 Emission and Forest Area: We used three datasets (Global Temperature, CO_2 emission and Forest area) to evaluate the impact of CO_2 emission and deforestation. The CO_2 dataset has emission values for all years in different columns, so we transposed it to show years in rows for all countries to merge with other datasets on country and year. We also dropped all null values before merging the datasets.

Figure 2 shows the CO_2 emissions for major countries by 2019. The figure shows most of the countries are decreasing their CO_2 emissions. However, CO_2 emissions of China, India and Russia are increasing over the time.

Figure 3 shows the increase and decrease in forest area for major countries. It shows that forest area for most of the countries is increasing except Brazil and Canada.

5.4 Results

In this section, we present our experimental results for the selected models: 1) SARIMAX and 2) artificial neural networks. We used the three datasets described in Section 5.2 in our experiments to evaluate our method. To measure performance of the

Country \Metric	Train MSE	Train MAPE	Test MSE	Test MAPE
USA	3.700	0.096	0.200	0.023
China	2.573	0.095	0.157	0.023
India	5.012	0.020	0.236	0.016

Table 1: SARIMAX Model Error Summary

Country \Metric	Train MSE	Train MAPE	Test MSE	Test MAPE
USA	5.08	0.11	4.26	0.12
China	3.99	0.11	4.08	0.12
India	8.77	0.10	6.85	0.09

Table 2: ANN Model Error Summary

models, we split the data into two groups: 80% training and 20% testing. The following subsections will present results of our models and visualization.

5.4.1 SARIMAX. We trained ∼159 SARIMAX models for temperature prediction using global historical time series temperature data. Table 1 shows the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) metrics of the training and testing sets of the top three countries. Figure 4 shows the SARIMAX model fit for USA as an example.

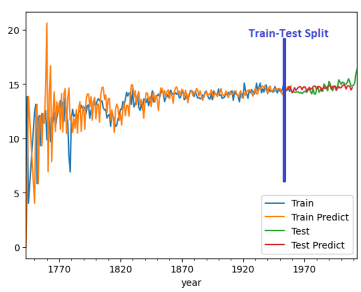


Figure 4: US Temperature Train Test Model Fit

5.4.2 Artificial Neural Networks. As we mentioned in Section 4.1.2, we trained an artificial neural network to predict the impact of deforestation and CO_2 emission on climate change for ∼140 countries globally. We prepared the final dataset to predict the impact of CO_2 emissions and forest area on temperature by joining all three datasets on year and country. CO_2 emission and forest area values are the features, and temperature is the dependent variable for the ANN model. Our neural network model has two hidden layers, each of which is followed by a dropout layer to avoid overfitting. First and second hidden layers have 128 and 64 neurons, respectively. Our model learns a total of 8,785 parameters.

Table 2 shows the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) metrics of the training and testing sets of the top three countries for the ANN model. Figure 5 shows a scatter plot of the actual and predicted samples for USA as an example.

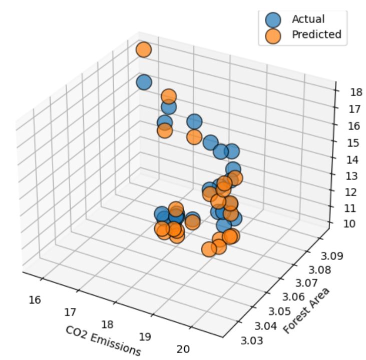
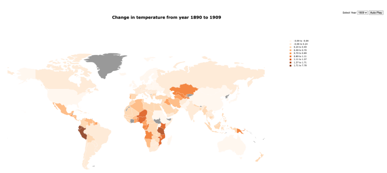


Figure 5: ANN US Actual vs Predictions Plot

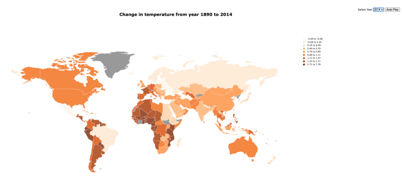
5.4.3 Visualization. Once the experimental results are collected, the right visualization becomes crucial for readers to understand information. In this section, we will present our visualization results for our climate dataset.

As we explained in Section 4.2, we developed two main visualizations in this project. Figure 6 shows the snapshots of our global temperature change visualizations which is implemented in D3. When we collect temperature forecast until 2043, we visualize the temperature increase by quarter degrees from 1890 to 2043 with an auto-play feature. The benefit of this visualization is to give an understanding of how fast temperature is increasing in different regions over time. Figure 6-a shows our baseline temperatures when the first temperature data is available for many regions. Figure 6-b and Figure 6-c show temperature changes by quarter degrees based on our baseline temperatures in 2014 and 2043. When a user auto-plays this visualization, the user can see how temperature will change in the future.

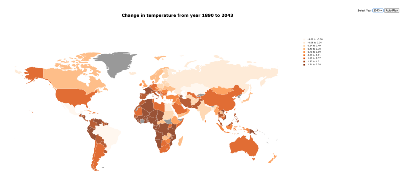
Figure 7-a and Figure 7-b show the Tableau visualizations for temperature change between our



(a) Heatmap for 1980

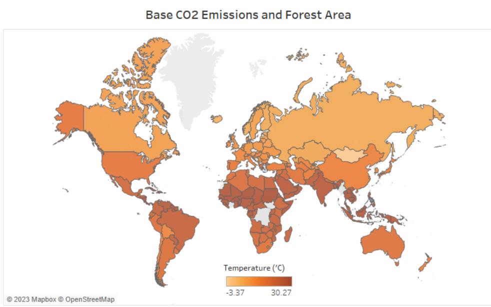


(b) Heatmap for 2000

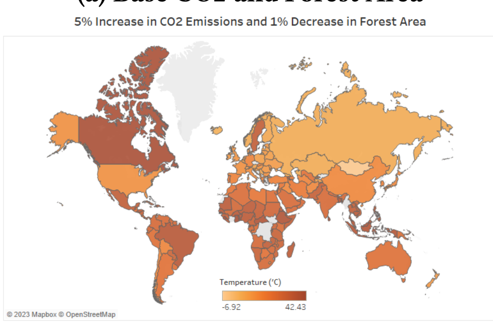


(c) Heatmap for 2043

Figure 6: D3 Visualization for Global Temperature Change



(a) Base CO2 and Forest Area



(b) Changed CO2 and Forest Area

Figure 7: Tableau Visualization for Temperature Prediction

base case values for CO_2 emissions and forest area against five percent increase in CO_2 and a one percent decrease in forest area. As seen in the figure, the temperature has increased in Canada. The user can visualize different what-if scenarios with our tool and can have a better understanding of impact of these two factors on climate change.

6 CONCLUSION & DISCUSSION

In this project, we investigated global temperature trends and impacts of CO_2 emission and deforestation on climate change for the next few decades. For global temperature trends, we trained 159 SARIMAX models. In order to evaluate the impacts of CO_2 emission and deforestation on climate change, we trained 140 different ANN models. Finally, we used Tableau and D3 to visualize our results to highlight the significance of our findings. In summary, we have reached the following results by this study:

(1) Our first visualization (i.e. Figure 6) shows that the temperature is getting warmer substantially over time globally.
(2) Since temperature is changing slightly over time, traditional visualization methods cannot capture the impact of the change. Therefore, we visualize the temperature increase by quarter degrees from 1890 to 2043 with an auto-play feature to highlight the impact.
(3) According to our experiments, the deforestation and CO_2 emission negatively affect climate change. Our second (i.e. Figure 7) visualization shows this impact percentage-wise. This visualization lets users visualize different scenarios by changing these two factors.
(4) During our data analysis, we noticed that the temperature fluctuates from time to time and location to location. Therefore, it is very challenging to train a single model that fits global data. Therefore, we trained our models country-wise. Our SARIMAX models' training took ~24 hours.
(5) According to our experiments, we cannot conclude whether CO_2 emission and deforestation are correlated to each other due to a lack of data.
(6) When we join all three datasets, the training data for ANN was reduced significantly since we took the intersection of three datasets. This caused some fitting problems for some of our ANN models.
(7) We could not find a reliable model for some countries due to a lack of data. Therefore, we had to exclude these countries from our analysis.

REFERENCES

[1] Beatriz Fátima Alves de Oliveira, Marcus J Bottino, Paulo Nobre, and Carlos A Nobre. 2021. Deforestation and climate change are projected to increase heat stress risk in the Brazilian Amazon. *Communications Earth & Environment* 2, 1 (2021), 207.

[2] NAM Azmin, Z Ahmad, R Mahmood, ASM Zahari, and H Hendar. 2022. The Dynamic Linkages between CO2 Emissions, Energy Consumption and Economic Factors in ASEAN Countries. In *IOP Conference Series: Earth and Environmental Science*, Vol. 1102. IOP Publishing, 012038.

[3] Jean-Francois Bastin, Emily Clark, Thomas Elliott, Simon Hart, Johan Van Den Hoogen, Iris Hordijk, Haozhi Ma, Sabiha Majumder, Gabriele Manoli, Julia Maschler, et al. 2019. Understanding climate change from a global analysis of city analogues. *PloS one* 14, 7 (2019), e0217592.

[4] David Ellison, Cindy E Morris, Bruno Locatelli, Douglas Sheil, Jane Cohen, Daniel Murdiyarso, Victoria Gutierrez, Meine Van Noordwijk, Irena F Creed, Jan Pokorny, et al. 2017. Trees, forests and water: Cool insights for a hot world. *Global environmental change* 43 (2017), 51–61.

[5] Seyed Mohsen Hosseini, Amirali Saifoddin, Reza Shirmohammadi, and Alireza Aslani. 2019. Forecasting of CO2 emissions in Iran based on time series and regression analysis. *Energy Reports* 5 (2019), 619–631.

[6] Peter Jackson. 2007. From Stockholm to Kyoto: A Brief History of Climate Change. Retrieved March 31, 2023 from https://www.un.org/en/chronicle/article/stockholm-kyoto-brief-history-climate-change#:~:text=In%201988%2C%20global%20warming%20and,public%20debate%20and%20political%20agenda

[7] Alexander Koch and Jed O Kaplan. 2022. Tropical forest restoration under future climate change. *Nature Climate Change* 12, 3 (2022), 279–283.

[8] Surbhi Kumari and Sunil Kumar Singh. 2022. Machine learning-based time series models for effective CO2 emission prediction in India. *Environmental Science and Pollution Research* (2022), 1–16.

[9] Shijie Li and Chunshan Zhou. 2019. What are the impacts of demographic structure on CO2 emissions? A regional analysis in China via heterogeneous panel estimates. *Science of the Total Environment* 650 (2019), 2021–2031.

[10] Yadvinder Malhi, Janet Franklin, Nathalie Seddon, Martin Solan, Monica G Turner, Christopher B Field, and Nancy Knowlton. 2020. Climate change and ecosystems: Threats, opportunities and solutions. , 20190104 pages.

[11] Felix Creutzig Nikola Milojevic-Dupont. 2021. Machine learning for geographically differentiated climate change mitigation in urban areas. *Sustainable Cities and Society* 64 (2021).

[12] Álvaro Salazar Oswaldo Maillard, Roberto Vides-Almonacid and Daniel M. Larrea-Alcazar. 2022. Effect of Deforestation on Land Surface Temperature in the Chiquitania Region, Bolivia. *Land* 12, 1 (2022), 2.

[13] Jayme A Prevedello, Gisele R Winck, Marcelo M Weber, Elizabeth Nichols, and Barry Sinervo. 2019. Impacts of forestation and deforestation on local temperature across the globe. *PloS one* 14, 3 (2019), e0213368.

[14] Guillaume Rohat, Stéphane Goyette, and Johannes Flacke. 2018. Characterization of European cities' climate shift–an exploratory study based on climate analogues. *International Journal of Climate Change Strategies and Management* 10, 3 (2018), 428–452.

[15] Haroon Sajjad, Pankaj Kumar, Md Masroor, Md Hibjur Rahaman, Sufia Rehman, Raihan Ahmed, and Mehebub Sahana. 2022. Forest vulnerability to climate change: A review for future research framework. *Forests* 13, 6 (2022), 917.

[16] Catherine E Scott, Sarah Anne Monks, DV Spracklen, SR Arnold, PM Forster, A Rap, M Äijälä, P Artaxo, KS Carslaw, MP Chipperfield, et al. 2018. Impact on short-lived climate forcers increases projected warming due to deforestation. *Nature communications* 9, 1 (2018), 157.

[17] Holly Shaftel. 2023. How Do We Know Climate Change is Real? Retrieved March 31, 2023 from https://climate.nasa.gov/evidence/#:~:text=In%201896%2C%20a%20seminal%20paper,Earth's%20atmosphere%20to%20global%20warming

[18] Saeed Solaymani. 2019. CO2 emissions patterns in 7 top carbon emitter economies: The case of transport sector. , 989–1001 pages.

[19] Lebunu Hewage Udara Willhelm Abeydeera, Jayantha Wadu Mesthrige, and Tharushi Imalka Samarasinghalage. 2019. Global research on carbon emissions: A scientometric review. *Sustainability* 11, 14 (2019), 3972.

[20] Kelvin O Yoro and Michael O Daramola. 2020. CO2 emission sources, greenhouse gases, and the global warming effect. In *Advances in carbon capture*. Elsevier, 3–28.