Green House Gas Emissions - A Global Perspective

Kavya Kavuri
Rutgers University
Piscataway, NJ, USA
kavya.kavuri@rutgers.edu

Kavit Amit Shah
Rutgers University
Piscataway, NJ, USA
kas672@scarletmail.rutgers.edu

***Abstract*—** We aim to build an interactive data story board to visualize the change of Green House Gas Emissions over a time period of 50 years. The dataset is from multiple sources[1],[2],[3], but mainly from Kaggle [1]. On our dashboard we show information in hierarchy such as the countries and their sectors contributing to the emissions.

***Keywords*—** Green House Gases, Carbon dioxide, Methane, Nitrous oxide, choropleth map, carbon footprint, Pandas, MySQL, Amazon S3.

I. Project Description

We know that there is an overwhelming concern about climate change in majority of the populace. Children as young as fifteen are affright by the changes that are happening in the environment. Sea levels are rising and oceans are becoming warmer. Droughts and famines are longer, hurricanes are more intense. Our planetâs diversity of life is at risk from the changing climate. The root cause for this climate change is gaseous emissions that are trapping heat in the earth's atmosphere. So our focus is on exploring the components of the emitters to help the any responsible person or organization, mitigate the release of these gases. To achieve this, we start with accumulation and analysis of a dataset.

The purpose of this project is to provide an interactive data story to illustrate the information of the emission of Carbon dioxide and green house gases. The total gaseous emissions from all countries are shown using a choropleth map, and the breakup of their emissions into sectors like Energy, Electricity, Agriculture etc are shown using pie chharts. With the help of this interactive dashboard, any individual or government members or climate change activists can easily filter out the crucial sectors on which improvements need to be brought upon.

A. Stage1 - The requirements gathering stage.

For making useful visual interaction and analysis of the data it was most imperative to obtain an efficient dataset for this purpose[3] We obtained our dataset from the kaggle website. [1] This dataset primary consists of the emissions data for all the countries, for all years(1960-2020) and all the major sectors/industries. Also, for the purpose of our project, we would making use of the 9 important features out of this data, namely, Country Name, Country Code, Year, Name of the gas, Sector Name, Value of Emission, Region Name, Region Code, which have been normalised and stored into 3 different tables. The data we obtained is a time-variant data and would be at rest till new year data is available. We also created an entity-relationship diagram, for marking the various relationships our normalised tables would have and how to use those. We had obtained a total of 1.5 GB worth of data before preprocessing, with approximately 7M records.

After this, we used the python libraries pandas and seaborn, for processing the dataset, constructing the normalised tables. We refined various columns such as filtering only indicator names concerning to emissions data, separation of the name of the gas and the sector name from a complex indicator name given in the dataset. We also refined some rows of data which were having null values for some features such as the value of emissions. We mainly made use of the fundamental features of the pandas data frame along with some apply and filter functions the module provided.
Provided above (Fig.1) is the image of how our timeline and work distribution across coming weeks looks like, to ensure on time project completion. We have completed the project and dataset selection, and data processing steps and would be starting off with the development process.

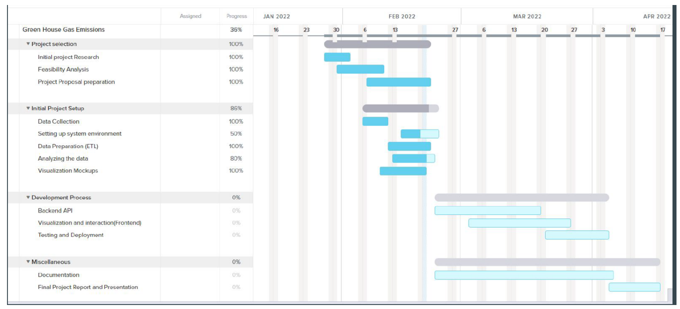

Fig. 1. Project Timeline and work distribution

B. Stage2 - The design stage.

- Short Textual Project Description: In our dashboard, we provide two types of information. Firstly, emission values (in Thousand Megatons) i.e., GHG emissions, CO2 emissions, Methane emissions, Nitrous Oxide emissions, secondly, sector-wise breakup for all these emission values. We provide aggregations like average, mean, percentage shares on all these values.
- Each view in the map is controlled by the same filters, i.e., range slider on years and multi-values drop-downs on countries and sectors. The purpose of doing this is to maintain coherent graphs all over the dashboard.
- We first identify and retrieve the important features from the dataset. Next we clean the data and impute the missing values. While doing the imputation, we faced with difficulties in converting Metric ton CO2 equivalent values to Kilotons, so we chose to convert kiloton values to Metric ton CO2 equivalents instead. This was done because the former method is not possible to implement.
- We then, normalize the tables and get three entities i.e., Emissions, Region, Country. Each region has multiple countries in it. Country has country name and its country code. We join these tables based on the required filters given by the user.
- We first take the dataset from Amazon S3 bucket and do the above process and put it in Amazon RDS MySQL database, which communicates with the flask application being hosted on Heroku server containing our Dash app. The Conceptual Flow

Chart and the ER diagram of our data looks like the following:

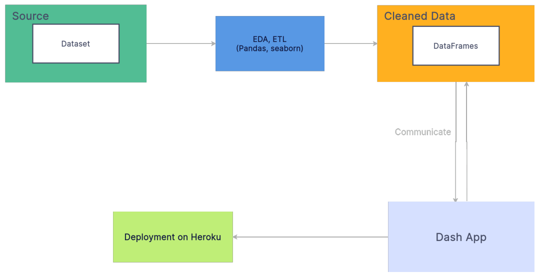

Fig. 2. Conceptual Flowchart

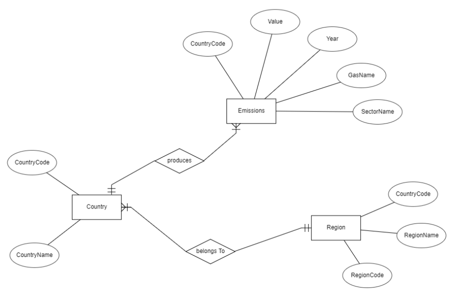

Fig. 3. ER Diagram.

C. *Stage3 - The implementation stage.*

- This web application primarily uses *dash* app to handle views, back-end computations and view-updates.
- The raw dataset had multiple csv files with around seven million records. The data manipulation was done using the Pandas library. All the data that is used for the app is stored in the Pandas DataFrames. The files are loaded into the DataFrames as soon as the app is loaded. The only hierarchies that were present in the raw data were Country and Year. Regions hierarchy was added on top of country level by means of aggregation.
- Views in the dashboard are implemented using dash's *appCallBack* functions that trigger whenever the value in the filters are changed. HTML, CSS and Python are used for producing the *figures* and *div* elements containing graphs and text in the dashboard.
- Dash package with Python language is used for this project. Dash automatically loads the database whenever application starts running.

The steps involved in the implementation are as follows:

- Downloaded the dataset as csv files from the sources provided in the references [1],[3].
- Preprocess each csv file by removing unwanted features and reorganising the required columns.
- Join the relevant files to create a master DataFrame, from which all the *figures* will be generated.
- Perform exploratory data analysis and see how the which countries are the highest emitters and which ones are the least. Survey which factors are the highest contributors in the highest gas-emitting countries.
- Generate aggregations on the data like mean gasesous emissions and percebtage division on sectors for each country.
- Generate graphs to visualize these computations and set triggers which update the views according to the user-applied filters.

D. *Stage4 - User interface.*

The User Interface was developed using Dash framework. The interactive graphs and plots are generated using Plotly express. After the exploratory data analysis, we surveyed different visualizations that could accurately represent our interesting findings. Based on this we created each graph in each section. The following libraries were used to perform exploratory data analysis and create graphs:

- Dash
- Plotly Express
- Numpy
- Pandas
- Matplotlib
- Raceplotly

There are mainly 3 types of gases in our dataset, namely, Carbon Dioxide, Methane and Nitrous Oxide. We give an option to visualize each of them individually or all together as well. We display a choropleth map which shows which country has the highest emission based on the filter selected. When a user hovers over the country, we show the Average emission of each gas in that country over the selected timeline. Alongside this, we should all the other essential information such as name of the country, country code, emission from all sectors for that country. The color scale is encoded using the emission values of each country. If the country is towards green color, it means the country is emitting less amount when compared to the red ones. This can be better understood with a visual representation, the following is the Choropleth map in our dashboard.

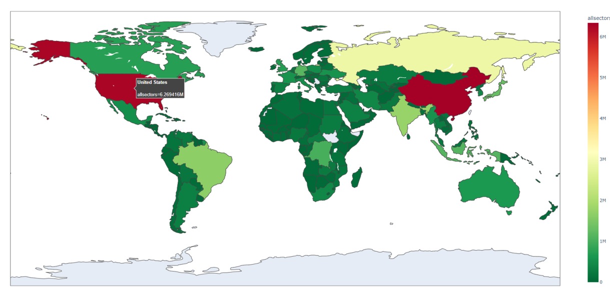

Fig. 4. Choropleth Map

In the map above, we can preemptively process the visual and say that the countries USA, China are clearly the top emitters overall. Here we can also select only a few countries and see how they are doing in a detailed manner using the line plots. In the next graph, we see the line plots for various countries, which show the trends in gaseous emissions of the selected gas. Again, we can clearly see that for all the countries, there is an increasing trend of gaseous emissions until 2010s and there is a slow decrease thereafter.

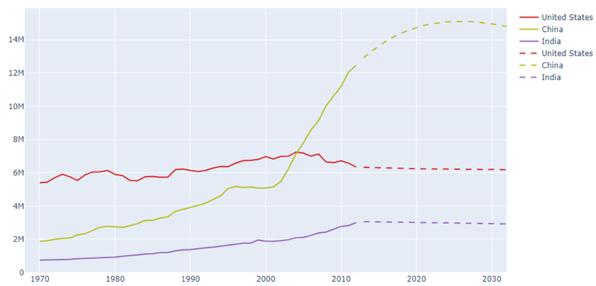

Fig. 5. Line plots of the selected countries

We can further dissect the main cause for the emissions in each country. We can visualize this using a sunburst chart. The sunburst chart is very effective in showing the division of the total gaseous emissions. Once again, using the sunburst we can immediately point out that the liquid fuel consumption is the largest sector contributing to the gaseous emissions. The following part of our dashboard explains the aforementioned details in a clearer manner.

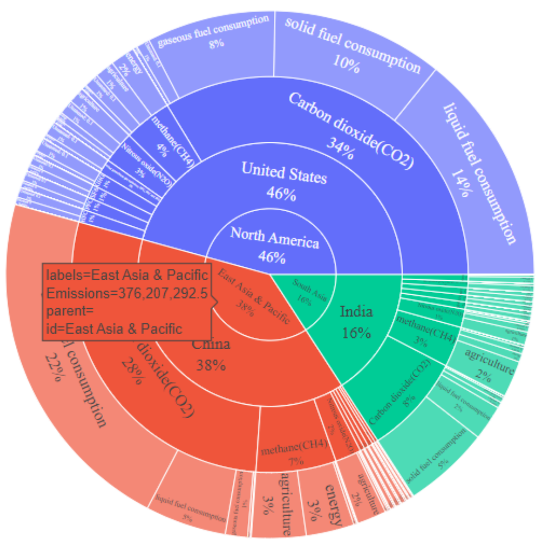

Fig. 6. Sunburst Chart to see sector-wise distribution

We created a race-chart which is a racing bar graph which shows how the top 10 countries has been changing over time. Using this visualization, we can point out a new detail that even though China has been a top emitter for a long time, after 2009 it has quickly reduced the emissions and has been on a downward trend since. The following is a snapshot of the race-chart.

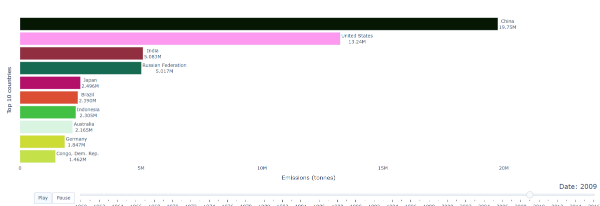

Fig. 7. Snapshot of Race-Chart to visualize top 10 polluters over a timeline of 1960 to 2018

We also created a Treemap or a Hierarchy Tree, which shows distribution of gaseous emissions by region. The regions are East AsiaPacific, Europe and Central Asia, North America, Sun-Saharan Africa, Latin America Cariibean, South Asia, Middle East North Africa.

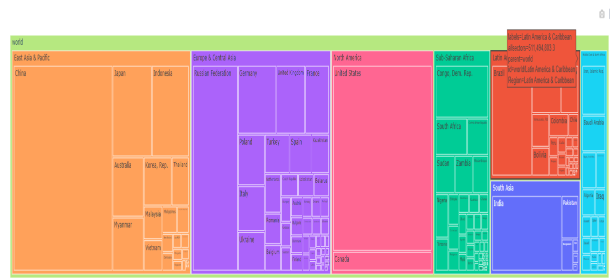

Fig. 8. Snapshot of Treemap to visualize emissions based on regions

II. Project Highlights.

Our dashboard has presented with numerous useful and interesting information about Green House Gas emissions. They are as follows:

1) Our dataset is a large set of assimilated data provided by Kaggle, Global Methane Initiative(GMI) and Our World Data, which sum up to a total of 4 GB.
2) Aggregations, matrics computed was effectively presented, following the guidelines of Gestalt Principles.
3) The core-principles of building this dashboard was to transition smoothly between different levels of visualisation complexity, effective visualization of large information and coordination of multiple views.
4) After using the dashboard, one can definitely remember that industrial usage liquid fuels is mainly causing the maximum pollution in any country.

III. Future work.

- Obtaining a state wise data for each country, for each year, and also for all the sectors, to gain even more depth into what is happening at the local level, and how regulatory measures could be taken by the counties as well.
- Obtaining more information on effects on gaseous emissions will help us see better and stronger correlation of gaseous emissions and its effects.

References

[1] M. HUI. (2020) World ghg emissions analysis dataset. [Online]. Available: https://www.kaggle.com/manchunhui/world-co2-emissions-analysis/data

[2] G. M. Initiative(GMI). (2020) Methane gas emissions. [Online]. Available: https://www.globalmethane.org/methane-emissions-data.aspx

[3] H. Ritchie and M. Roser. (2020) Coâ and greenhouse gas emissions. [Online]. Available: https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions