Relation between Climate Change and Disease

Robin Jose Raju

July 2024

1 Introduction

In recent decades, we have been witnessing an alarming increase in emergence of new diseases in human beings. Climate change could be a significant contributor for this trend along with many other physical factors. A major factor for climate change is the emissions of several gases which are observed to be a significant contributors to climate change and environmental pollution. European Parliament classified 7 gases as the main contributors for global warming in the article "Climate change: the greenhouse gases causing global warming" and those are : Methane (CH4), Carbon Dioxide (CO2), Nitrous Oxide(N2O), Sulfur Hexafluoride (SF6), Nitrogen Trifluoride (NF3), Hydrofluorocarbons (HFC) and Perflurocarbons (PFC).

The main agenda of the project was to see if there is any relation between the emission rate of the above mentioned gases and number of new diseases reported. After data cleaning and transformation, I have used the resulting dataset to plot 2 graphs which maps Averaged Emissions and Number of Diseases on y-axis of the graphs and year on x-axis for both graphs. The graph which plots the emission to year follows a negative trend for most of the gases while the diseases to year graph is scattered. The final result is that I couldn't find a relation between these 2 factors as even though the average emissions decline, the number of diseases reported are scattered and doesn't follow any trend.

2 Used Data

I have used 2 datasets: one contains the reported emission of gases by different categories in the country and the other with reported new diseases in a country. Both these datasets contains the values that are obtained in an yearly manner.

2.1 PRIMAP-crf[1]

All countries are required to report their domestic emissions to United Nations Framework Convention on Climate Change (UNFCCC) in the Common reporting Format(CRF) on annual basis. This data is formatted to meet the IPCC 2006 guidelines and organised in a table containing all available countries and their corresponding greenhouse gas emissions.

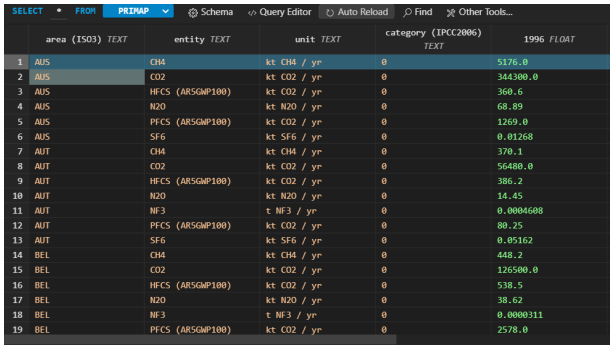
The above data is combined into a dataset which gives a detailed view into the gases emitted by each available countries, along with amount, category from which the gas was emitted, and the year. This dataset contains emission data from 1986 to 2019.

The original data, which is freely available from UNFCCC website, is obtained and modified to meet the IPCC2006 guidelines. This helps in more standardized datasets and formats for helping future usage. The PRIMAP-crf dataset is a tabular data in a .csv file which is formatted consistently with the PRIMAP2 interchange format.

Cleaning & Transformations

Since we only have data of diseases from 1996, we could also use the emission data from 1996. To begin the cleaning process, I started with filtering the data rows with category equal to 0 which is the total emission of the country, it is what we need rather than emissions from different categories in the country. From the resulting dataset, we only select those data rows which contain emissions of previously mentioned 7 gases. As the next step, I have transformed all 'NULL' values by replacing it with the minimum emission value of that gas by the country. In the last step, I transformed all the emis-

sion values to kilotons/year scale to have a uniform scaling between the data.



SELECT * FROM PRIMAP — Schema | Query Editor | Auto Reload | Find | Other Tools...

	area (ISO3) TEXT	entity TEXT	unit TEXT	category (IPCC2006) TEXT	1996 FLOAT
1	AUS	CH4	kt CH4 / yr	0	5176.0
2	AUS	CO2	kt CO2 / yr	0	344300.0
3	AUS	HFCS (AR5GWP100)	kt CO2 / yr	0	360.6
4	AUS	N2O	kt N2O / yr	0	68.89
5	AUS	PFCS (AR5GWP100)	kt CO2 / yr	0	1269.0
6	AUS	SF6	kt SF6 / yr	0	0.01268
7	AUT	CH4	kt CH4 / yr	0	370.1
8	AUT	CO2	kt CO2 / yr	0	56480.0
9	AUT	HFCS (AR5GWP100)	kt CO2 / yr	0	386.2
10	AUT	N2O	kt N2O / yr	0	14.45
11	AUT	NF3	t NF3 / yr	0	0.0004608
12	AUT	PFCS (AR5GWP100)	kt CO2 / yr	0	80.25
13	AUT	SF6	kt SF6 / yr	0	0.05162
14	BEL	CH4	kt CH4 / yr	0	448.2
15	BEL	CO2	kt CO2 / yr	0	126500.0
16	BEL	HFCS (AR5GWP100)	kt CO2 / yr	0	538.5
17	BEL	N2O	kt N2O / yr	0	38.62
18	BEL	NF3	t NF3 / yr	0	0.0000311
19	BEL	PFCS (AR5GWP100)	kt CO2 / yr	0	2578.0

PRIMAP dataset after processing

2.2 A global dataset of pandemic and epidemic-prone disease outbreaks[2]

This dataset contains new infectious diseases outbreaks collected from the Diseases Outbreak News(DONs). This dataset is product of a paper in which the researchers collected news about infectious disease outbreaks from DONs and Coronavirus Dashboard produced by World Heath Organization. The dataset contains information on diseases occurred over the period from 1996 to 2022 in 233 countries and territories around the world but we are using only the data till 2019 as the PRIMAP-crf contains data till 2019. The researchers have classified Africa, America and Asia as hot spots since they noticed high incidences of outbreaks in these regions.

The paper produces different layouts of the dataset from which 'Outbreaks.csv' is the one that we are interested in as they have all the new outbreaks and can be easily integrated to the dataset of domestic emissions.

Since the data is obtained directly from the WHO, modification on it is necessary to adapt the dataset for different use cases and the researchers have divided the datasets into different subsets of the main dataset. This helps us to choose the data that is relevant for the project and neglect other unwanted data.

From different subsets of the dataset, I choose the Outbreaks subset, which contains all required information like country code based on ISO 3166, year of the occurrence of the outbreak, name according to ICD-10 etc.

Cleaning & Transformations

I have modified the dataset for a simpler and easier representation of the data, which will help in removing all unwanted columns that do not contribute significantly to the project and also truncated the data to contain only data from 1996 to 2019. All naming of diseases and country names follow the standard codes like ICD-10, ISO-3166 etc.

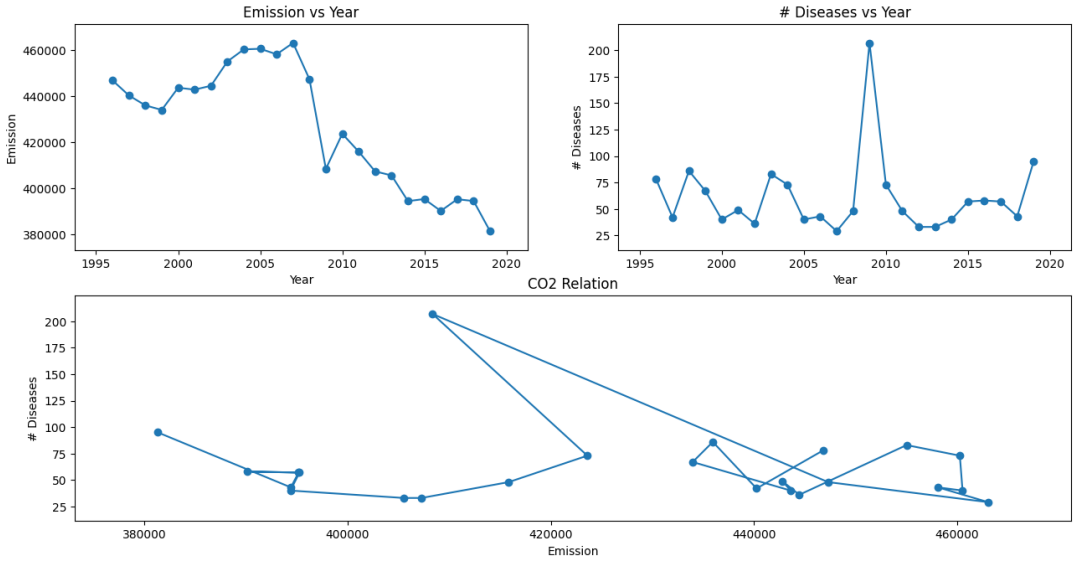
The 2 datasets that are used in this project are covered under an Open Data CC BY 4.0 license, which can be confirmed by going to the respective metadata links: PRIMAP-crf & A global dataset of pandemic- and epidemic-prone disease outbreaks. The CC BY 4.0 license lets users to share and adapt the dataset to meet the requirements of the project.

3 Analysis

The main agenda of the project is to find out if there is any relation between the emission of major gases that contribute to climate change and number of new diseases reported. For research, I have started with getting each gases, from the dataset outputted by the data pipeline, one by one and finding the average value of emission for that particular gas from 1996 to 2019. After this step, we have 23 values for each 7 gases which represents the averaged global emission of that gas, yearly from 1996 to 2019. We observed a downward trend for Methane (CH4), Sulfur Hexafluoride (SF6) and Perflurocarbons (PFC), upward trend for Hydrofluorocarbons (HFC), dipping and rising to almost stable level for Nitrous Oxide(N2O) and rising then dipping trend for Carbon Dioxide (CO2) and Nitrogen Trifluoride (NF3). From this, I noticed that most of the gases are following a downward trend, so I summarised that the emission rate of these gases are diminishing. I have plotted different graphs which shows the trends of each gas with emissions and year on x and y axis respectively.

On diseases dataset, I calculated the total number of new diseases reported in each years and plotted them in a graph against years. The highest number of new diseases were reported on the year of 2009 with 205 new disease reported around the globe. The graph doesn't follow any trend over the years, so we

cannot find any relation between the arise of new disease to emissions of gases.



Graph of Carbon Dioxide (CO2

The above figure, is of Carbon Dioxide (CO2) gas, which shows 3 graphs : the top-left graph shows the average gas emission to year with downward trend, the top-right graph is the one which maps number of new diseases to years. The bottom graph is mapped with number of new diseases on y-axis and average emissions on x-axis, this graph is to show the relation of emitted gas to number of new diseases. There are 6 other graphs which shows the mapping of all other 6 gases in the same manner.

The final analysis is that there isn't any noticeable direct relation between emission rate of these gases to the number of new diseases. This maybe due to many other factors like natural calamities or diseases spread by other means other than air. There is possibility of a delayed effect of high emission causing diseases in a later time, meaning the after effect of high emission of gases leading to new diseases is delayed. This could be the case with Carbon Dioxide (CO2) which hit the peak in 2007 and the highest number of new disease at 205 around the world in 2009 but this needs more detailed study and is out of the scope of this project.

4 Conclusion

To wrap up the project, I summarized that this basic setup to find the relation between the emission rate of gases and number of new diseases reported is not enough to find a solid result. The project doesn't show any relation between them but there are chances of delayed effect as mentioned above with Carbon Dioxide (CO2). There could also be many other causes for the number of new disease like natural calamities or animals spreading the disease etc.

All the data related to this project is openly available for anyone to utilise, modify or share. The code for the data pipeline and for the graphs can also be found on the public github repository which also allows you to use, modify and share as mentioned in Open Data CC BY 4.0 license.

References

[1] Gütschow, M. L. Jefferyand A. Günther., *PRIMAP-crf: UNFCCC CRF data in IPCC categories (PRIMAP-crf-2021-v1)*, Zenodo, Apr. 27, 2021. doi: 10.5281/zenodo.4723476.

[2] Torres Munguía, Juan Armando (2022). *A global dataset of pandemic- and epidemic-prone disease outbreaks*, figshare. Dataset. https://doi.org/10.6084/m9.figshare.17207183.v2