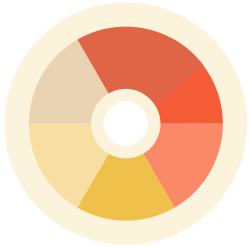
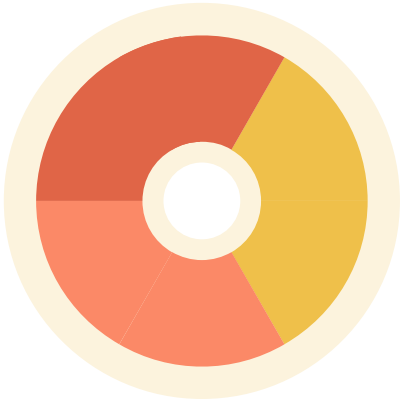




Air Pollution and Health

Visual Analytics Project

Andrea Franceschi 1709888
Riccardo Fiore 1705634

Introduction

- One of the biggest problems afflicting humanity today is air pollution, it refers to the contamination of the atmosphere by harmful chemicals or biological materials. According to the World's Worst Polluted Places by Blacksmith Institute in 2008, two of the worst pollution problems in the world are urban air quality and indoor air pollution. To solve the problem of air pollution, it's necessary to understand the issues and look for ways to counter it.

- Our project focuses on the processing and analysis of air pollution in its various chemical components and its impact on short- and long-term human health which was considered in relation to deaths due to respiratory diseases.

Goal and User Target

The target audience for this project is very broad as it has global relevance and can be used by different governments for a study of air pollution trends. This project could also be used by private companies and associations that protect the health of the world's population and of the planet.

GOAL

USER TARGET

Analysis of the relationship between the amount of air pollution and deaths related to it on a general scale. In particular, we want to assess the possible relationship and patterns between different types of emissions and different causes of death.

Database

Our database was composed by combining data from different sources.

- Global Health Data Exchange (GHDx): the world's most comprehensive catalog of surveys, censuses, vital statistics, and other health-related data.
- Emissions Database for Global Atmospheric Research (EDGAR): a joint project of the European Commission which provides global past and present day anthropogenic emissions of greenhouse gases and air pollutants by country and on spatial grid
- Some other like: The World Bank and Our World in Data

We selected the fields we thought would be interesting to use to conduct our analysis.

Data Pre-processing

The initial pre-processing took place on the various csv/xls files taken from the different sources and containing the various fields that formed our final database.
Having used different sources, we had to combine the different tables using the name of the various countries as a common value, which came in different formats and with inconsistencies in the name. Each initial file contained the data for all the years taken into consideration, the processing of this data was used to create one file for each year, all with the same format and the same fields

Once all the files with the source tables had been created, they were reprocessed with **PCA** algorithm in order to display an optimal projection on 2 axis of the majority of attributes. The attributes Total Cancer and Total deaths have been removed because their subsets more closely related to pollutants have been included in the **PCA** calculation. The result produced a set of coordinates of which only the 2 with the highest variance were added as columns to the database.

Data Structure

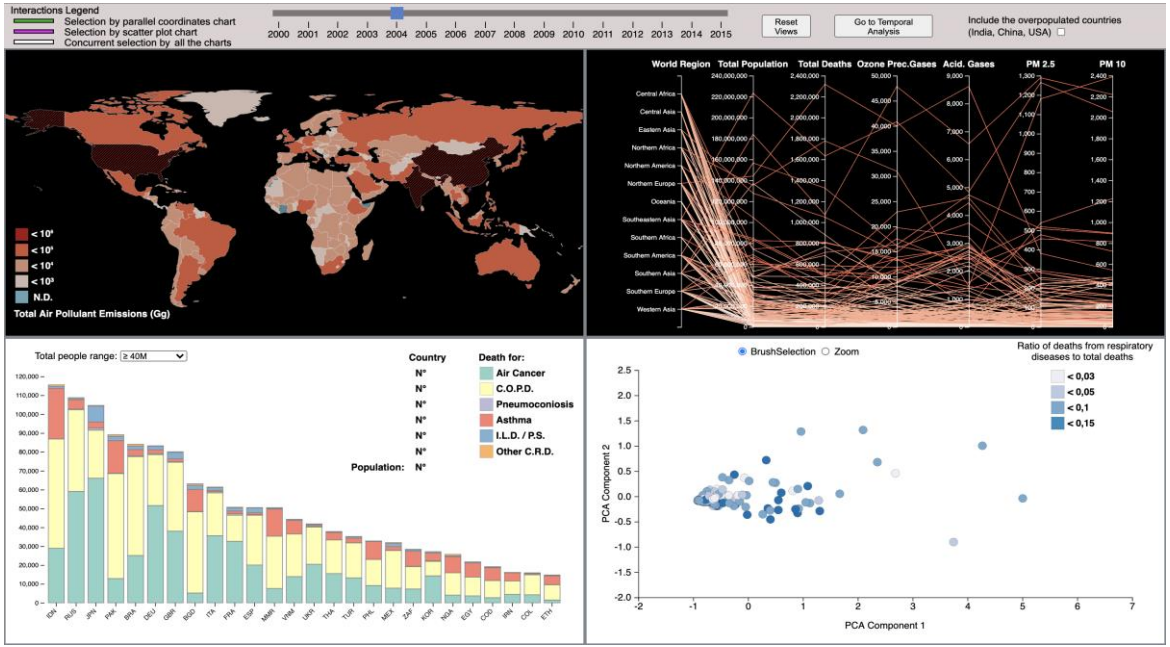
The fields we have chosen and entered in the database used for the visualisation are specified below

Country	All the world countries
Population	Total country population
Total Deaths	Total country deaths
CH4 (Gg)	Methane emission from all artificial causes
CO (Gg)	Carbone Monoxide emission from all artificial causes
NH3 (Gg)	Ammonia emission from all artificial causes
NMVOC (Gg)	Non-Methane Volatile Organic Compounds emission from all artificial causes
NOx (Gg)	Nitrogen oxides emission from all artificial causes
SO2 (Gg)	Sulfur Dioxide emission from all artificial causes
PM 2,5 (Gg)	Fine Particulate Matter of size ≤ 2,5 μm emissions
PM 10 (Gg)	Fine Particulate Matter of size ≤ 10 μm emissions
C.O.P.D deaths	Chronic Obstructive Pulmonary disease deaths
Pne. deaths	Pneumoconiosis deaths
Asth. deaths	Asthma deaths
I.L.D.P.S deaths	Interstitial lung disease and pulmonary sarcoidosis deaths
O.C.R.D deaths	Other chronic respiratory diseases deaths
Total Cancer deaths	All kind of cancer deaths
T.B.L Cancer deaths	Tracheal, bronchus and lung cancer deaths

Visualizations

Project Layout

In order to display all the details of our database through 4 different visualization dashboards, we have chosen a mixed representation mode with an additional header that allows us to navigate through the various years and access certain functionalities. As you can see the layout is divided with two main colours, on the black background we find the graphs concerning emissions, while on the white background we find the graphs concerning the number of related deaths.



Visualizations

Header Bar

At the top of the layout we find a header bar mainly used to manage certain interactions with also a legend about these interactions.

Legend

This is the legend showing how the perimeter is coloured depending on the graph from which the selection is triggered

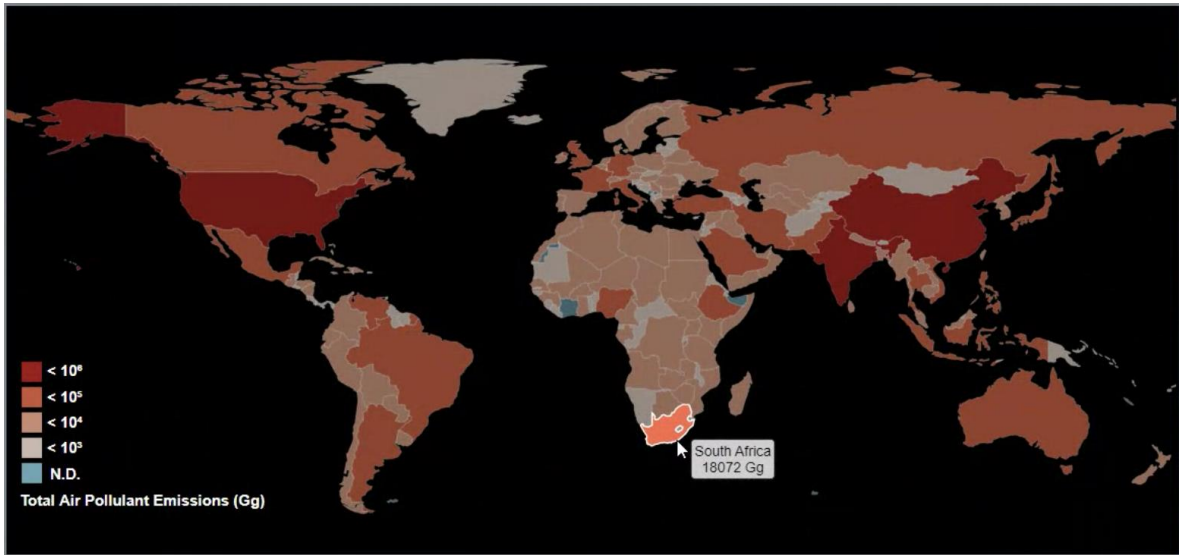
Time Selector

The purpose of this range bar is to scroll through the various years by changing the data display

Other Commands

This is where we find a number of additional commands that change the views or data displayed.

Visualizations



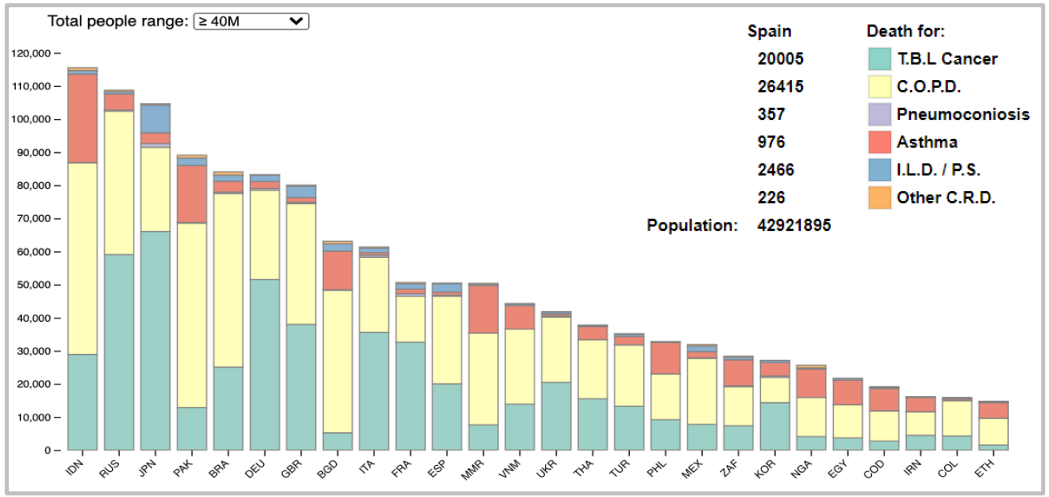
World Choropleth

The first visualization analysed is the world map Choropleth. The colouring of the various countries was done using a predefined sequential colour scale from white to red. In particular, white represents the countries with the lowest values of pollutant emissions and red the countries with the highest values. As can be seen from the legend at the bottom left, the ranges corresponding to the individual values increase exponentially as the difference in emissions between countries is high.
From the image below you can see the label that will appear when you move the mouse over the various countries and that will show in detail the total value of emissions of them.

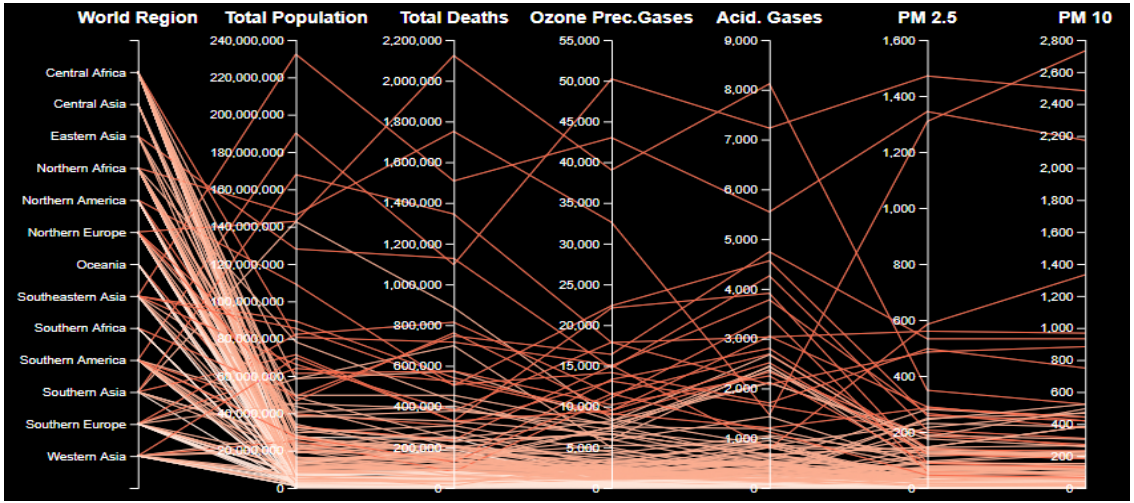
Visualizations

Bar Chart

It shows the total number of deaths from pollution-related causes, so the y-axis shows the number of deaths while the x-axis shows the initials of the names of the different countries. In particular, it can be seen that each bar is divided into different colours, each of which represents the percentage of deaths from each specific cause in relation to the sum of deaths from the different causes. The relationship between the colour and the cause of death can be seen in the legend at the top left. This legend also shows the specific value of deaths from each cause in the different countries, so it will vary according to the bar pointed to by the mouse. In order to allow a clear visualization to the user, the countries represented in this graph are divided into subgroups according to the total population range. This is done via a drop-down menu.



Visualizations



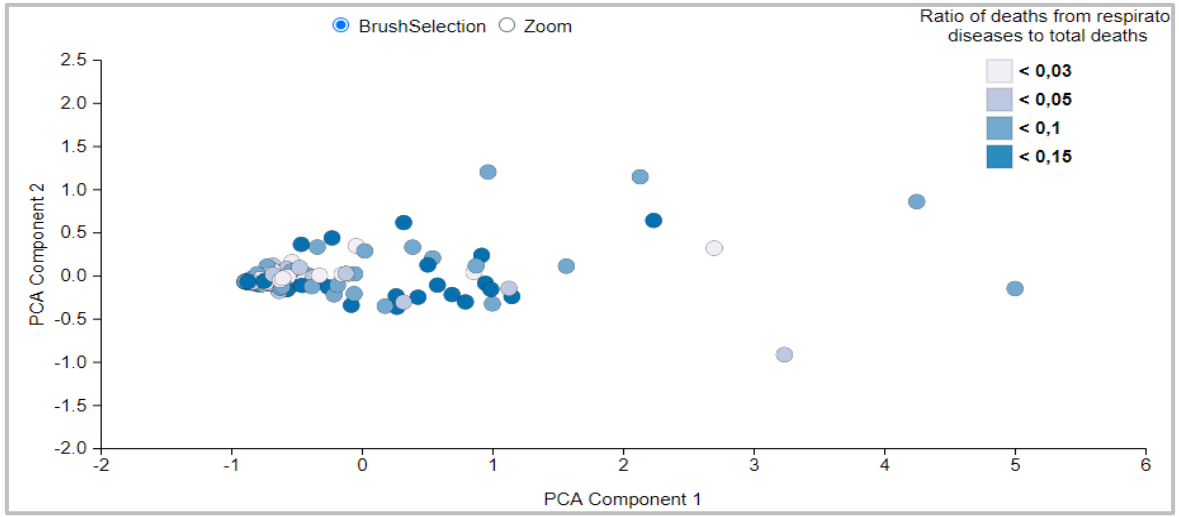
Parallel Coordinates

This graph gives a detailed view of the different components of emissions in relation also to the total annual number of deaths and the population of each country.
For an optimal visualization, we have decided to group each country into its respective World Region, these make up the first axis. Gaseous emissions are also grouped into two supergroups shown on different axes, Ozone precursor gases and Acidifying gases. Whereas fine particle emissions are shown on the axes separated by size. Each path is associated with a country and initially follows the same colouring as used in the World Choropleth.

Visualizations

Scatter Plot

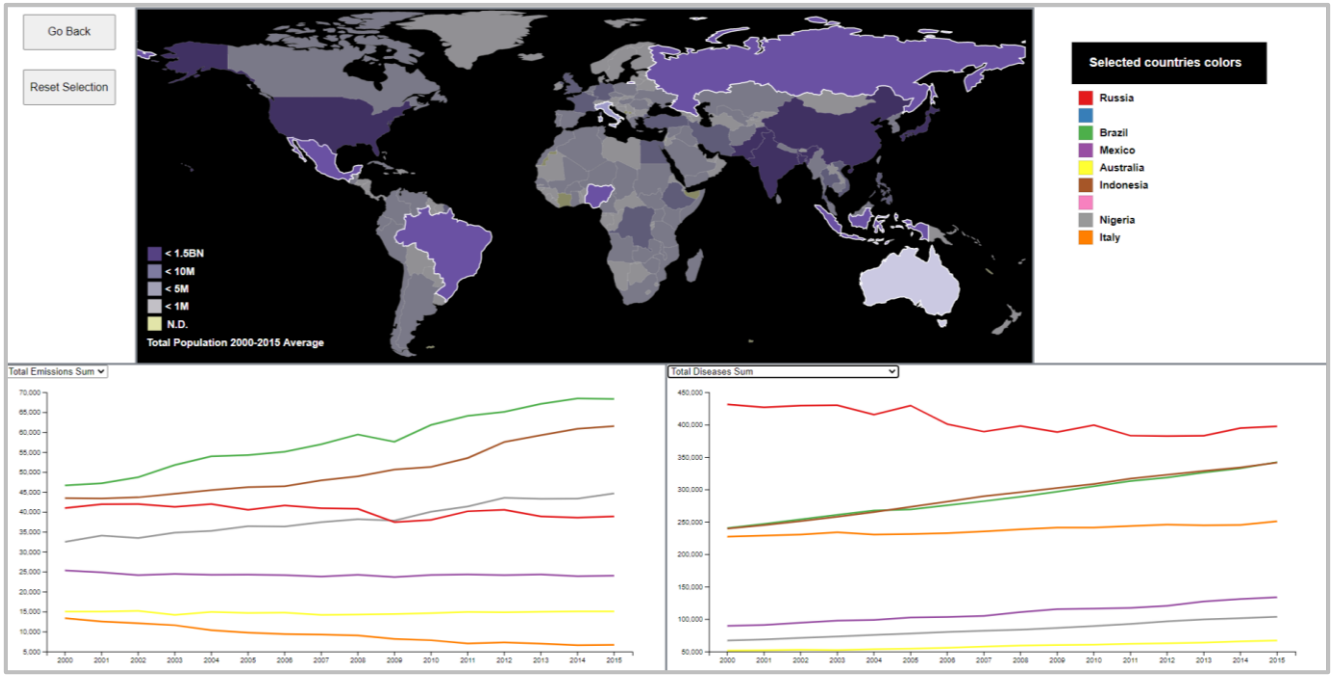
This graph shows the projection of all attributes on only two components calculated by PCA in the pre-processing phase. In particular, the x-axis shows the first component (maximum variance) as the most extended between the two axes. The y-axis shows the second component with the highest variance.
All the points on the graph represent a nation and their colouring was carried out on the basis of the value obtained from the ratio between the sum of deaths from pollution-related causes and the total number of deaths. The colours were obtained using a standard linear scale whose ranges are given in the legend at the top right. As the PCA algorithm preserves the Euclidean distance, the closer two points are, the closer the relationship between them is.



Visualizations

Comparison over the years

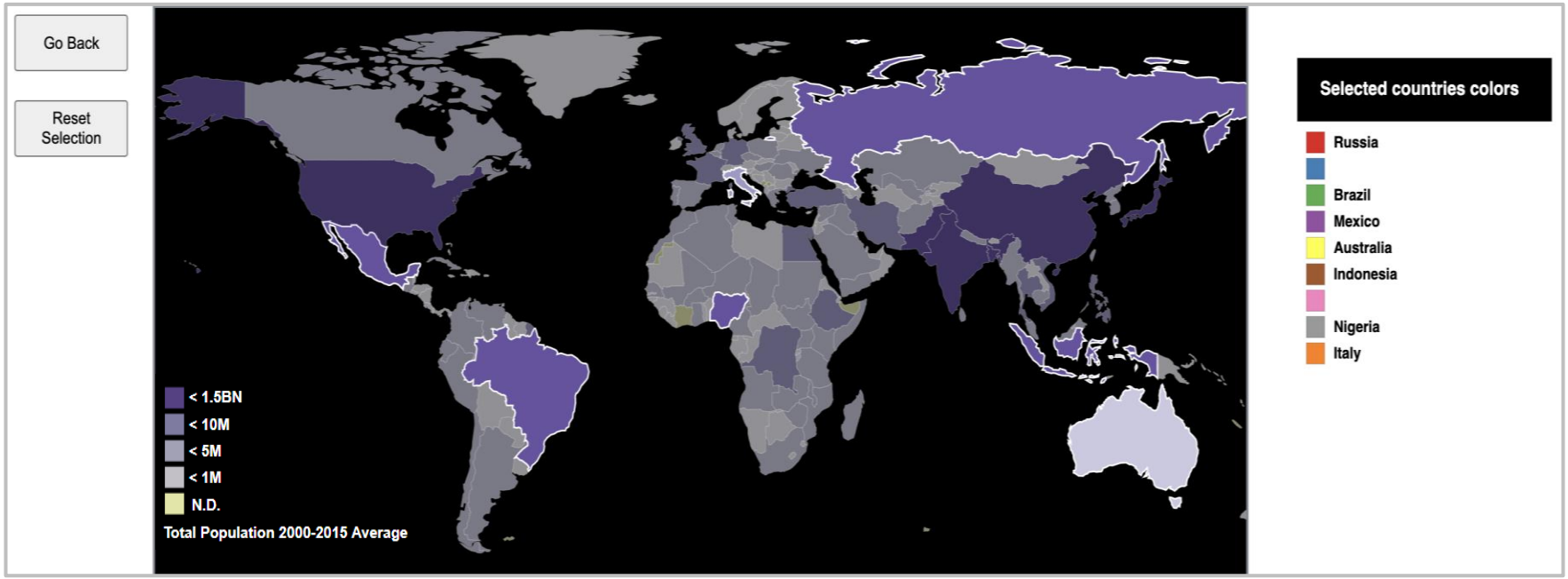
During the annual analysis, we noted that it would be interesting to be able to visualise and compare trends in emissions and related deaths over the years.



Visualizations

Comparison over the years

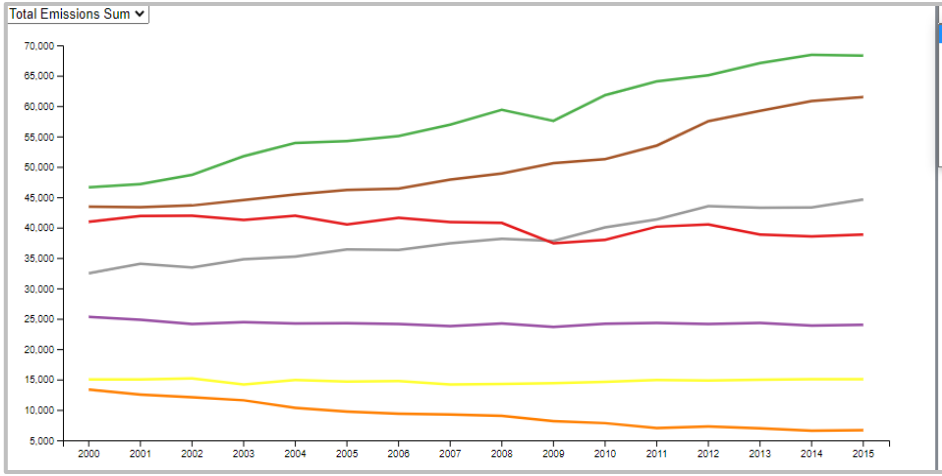
We have implemented an additional visualisation where the Choropleth shows the average total population in the years 2000-2015. There is a legend in the layout to associate each selected country with a colour. There are also two buttons to return to the previous layout and to reset selections.

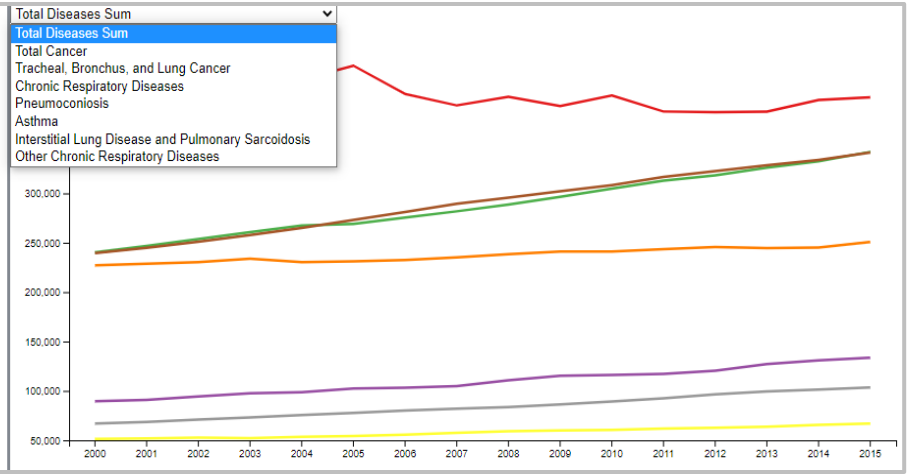


Visualizations

Comparison over the years

In this mixed layout, there are two line charts showing the trend over the years of emissions and deaths from related causes. In both cases we find a drop-down menu that allows the analysis of trends in the specific types of the two categories.
Each country selected on the map will be represented on these charts by a line whose colour represents the association with that country.





Go to demo…