Air Pollution and Health
Visual Analytics

Sapienza University of Rome

Andrea Franceschi 1709888
Riccardo Fiore 1705634

2020 March 02

Abstract: Air pollution refers to the release of pollutants into the air that are detrimental to human health and the planet as a whole, it is one of the most serious problems in the world it refers to the contamination of the atmosphere not only by harmful chemicals but also by biological material. Air pollution can cause long-term and short-term health effects. It's found that the elderly and young children are more affected by air pollution. Short-term health effects are reduced to irritations, allergic reactions and infections, while long-term health effects include lung cancer, several types of respiratory disease, brain damage, liver damage, kidney damage and heart disease. Unfortunately, the air pollution also affects the world around us, indeed it causes damage to crop, animals, forests, and bodies of water. It also contributes to the depletion of the ozone layer, which protects the Earth from the sun's UV rays. Another negative effect of air pollution is the formation of acid rain, which harms trees, soils, rivers, and wildlife. Some of the other environmental effects of air pollution are haze, eutrophication, and global climate change.

Introduction

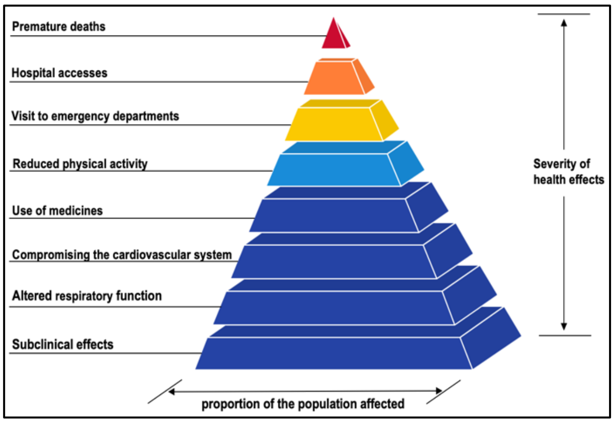
The rapid deterioration of the air quality in the world has aroused much attention from the world population and in particular from organisations involved in safeguard the Earth. As many cities around the world become more congested, concerns increase over the level of urban air pollution being generated and in particular its impact on localised human health. Human activities have been highlighted as the major causes of air pollution.

To support a larger population, there's a need for energy production, transportation, and industries, resulting in the emission of harmful chemicals into the atmosphere.

Gaseous components and atmospheric particulate matter are commonly used criteria for evaluating air quality.

The degree of adverse health effects depends on the size and composition of the particles. PM2.5 and PM10 are defined as particles with diameters of 2.5 µm or less and 10 µm or less, respectively; these parameters are usually measured using air quality index (AQI). Several publications affirm that asthma, COPD, lung cancer and respiratory infections all seem to be exacerbated due to exposure to a variety of environmental air pollutants with the greatest effects due to particulates matter, ozone and nitrogen oxides.

Below is a pyramid showing the overview of health impacts resulting from exposure to air pollution. It illustrates the inverse relationship between the severity of the health effects and the proportion of people affected by them.



In relation to this overall overview of the health impacts of air pollution, our project takes into account all deaths from the causes contracted by the largest number of people. In particular, the first two pyramid's groups with the highest proportion of people affected are considered.
We wanted to analyse the correlation between harmful emissions that contribute to air pollution and their impact on human health, through an analysis over the time period 2000-2015 it is possible to establish new insights into how air pollution is associated with increasing respiratory diseases and mortality rates.

Database

The used database is the result of a combination of several sources:

- *Global Health Data Exchange*, the world's most comprehensive catalog of surveys, vital statistics, and other health-related data from which we have taken the number of deaths differentiated by respiratory diseases.

- *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.

- *World bank* from which we have taken the total population for each country in the time range of our interest.

- *Our World in Data* from which we have taken the total deaths.

Pre-processing

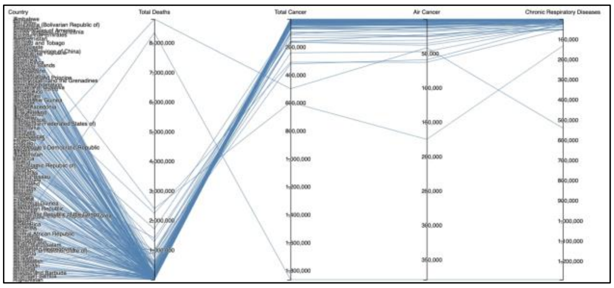
The csv/xsl files obtained from the above sources were pre-processed to obtain a final database. The presence of different sources led to typing inconsistencies between the files, so we had to combine the different tables using the name of the various countries as a common value. Each initial file contained data for all the years considered, the processing of this data served to create a file for each year, all with the same format and the same fields
The next step was the application of PCA algorithm to the resulting database, to display an optimal projection on 2 axes of the majority of attributes. Among all the attributes in our database only 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. So, at the end of pre-processing we have the following columns:

- ***Country***: all the world countries
- ***Population***: total country population
- ***Total***: total country death

- ***CH4***: Methane emission from all artificial causes
- ***CO***: Carbone Monoxide emission from all artificial causes
- ***NH3***: Ammonia emission from all artificial causes
- ***NMVOC***: Non-Methane Volatile Organic Compounds emission from all artificial causes
- ***NOx***: Nitrogen oxides emission from all artificial causes
- ***SO2***: Sulphur Dioxide emission from all artificial causes
- ***PM2,5***: Fine Particulate Matter of size ≤ 2,5 μm emissions
- ***PM10***: 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
- ***Air Cancer deaths***: Tracheal, bronchus and lung cancer deaths

Prototype

During the development of our project, we were sometimes faced with the decision to change the approach to the data, for example the initial idea was to represent on the Parallel Coordinates all the countries, but this led to an unclear visualization. Therefore, we decided to group each country in its respective world region. We also thought of representing on each axis of the Parallel Coordinates the number of deaths due to different causes, but it did not allow us to establish useful insights due to the excessive number of attributes and therefore of axes. We thought it would be more useful to extrapolate from this chart the possible correlation between the total number of deaths for each country and its emissions of particles and/or gaseous components.



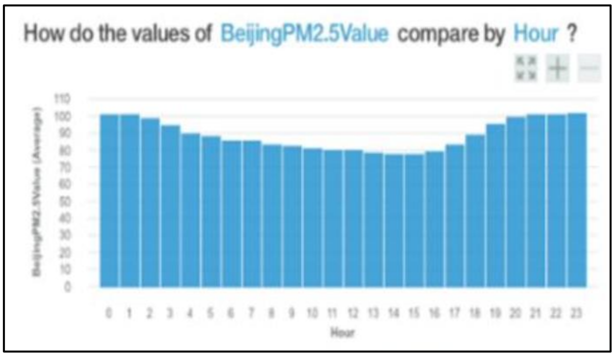
In the management of user interaction and the various charts we had to make decisions about coloring. Initially when a country was selected on the map it was illuminated in white color, but with the development of the interactions of the charts with the map we realized that a single color to indicate the selected country was not enough, it could create confusion when we had more than one chart interacting with the map. Then we

used a different color to indicate what kind of chart the selection came from.
The problem related to the large number of countries also emerged during the management of the Bar Chart, in order to have a clearer and more efficient visualization the countries represented in this graph are divided into subgroups according to the total population range, this is done via a drop-down menu.
In relation to the Bar Chart, the initial idea to have on it a one-way interaction, but when we used the prototype to establish useful insights, we found it was difficult to understand the relationship between emissions and deaths from related causes. Indeed, selecting countries with high emissions from the other charts did not give us a direct feedback on the Bar Chart that contained the information on deaths from the various causes. Therefore, we decided that the selection of countries from the various graphs on the bar chart should illuminate the corresponding bar. Although, for the reasons explained above, the latter will have to be looked for in the various population intervals that are easily identifiable thanks to the relative axis of the total population present in the Parallel Coordinates chart.

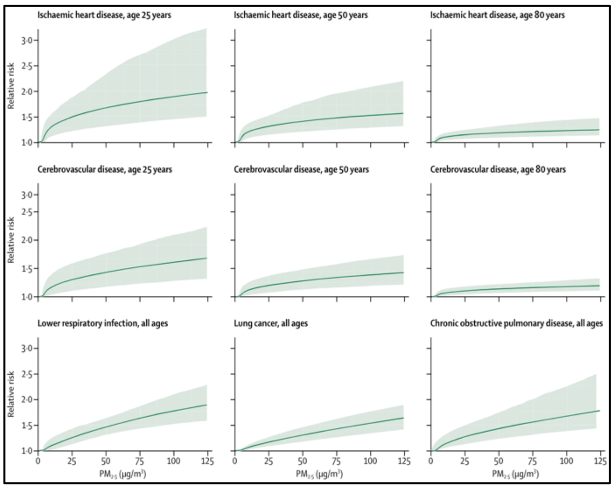
Related Work

Air pollution is one of the biggest problems afflicting humanity today, therefore, from some years to now, it has been the subject of interest of many scientific publications. Many experts have tried to propose efficient visualization methods to represent the information contained in air pollution data in order to verify hypotheses and obtain useful findings.
A very interesting publication is *Visual Analysis for Exploring the Relation between Air Pollution, Environmental Factors and Respiratory Diseases*, a study conducted on the city of Beijing in China. In this paper is used the advanced visualization tool based on cloud and artificial intelligence "Watson Analytics", several datasets on PM2.5 air pollution, environments and public health in Beijing were collected and analyzed in order to discover the intrinsic relationships between these factors. Its Data set also contains air pollution levels (PM2.5, SO2, CO2, NO2, O3) based on the Air Quality Index (AQI), meteorological data (temperature and relative humidity) and emergency room visits data for Beijing in the time period 2013-2015.
The analysis produced by Watson Analytics tool demonstrated a strong correlation between the three parameters: PM2.5 and other pollutants, environment factors such as season and temperature, and emergency room visits (ERV) related to respiratory diseases. This analysis showed that the highest number of respiratory diseases occur when the temperature is lowest, and the air quality is most hazardous. It showed that January, February, March and especially the months of November and December, experience increased numbers of ERVs for respiratory diseases. Similarly, the concentrations of PM2.5 and CO, NO2, and SO2 are also highest during these months
Although the subject matter is the same, this publication shows some differences from our project. A preliminary difference is the sample period: while the

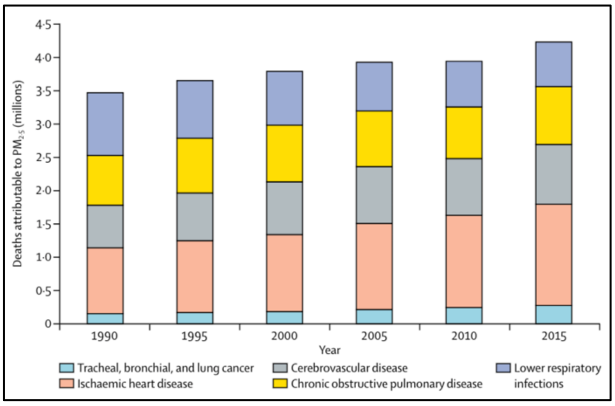
publication focuses on PM2.5 emissions in the hours, months and seasons of the time period 2013-2015, our project focuses on a global view for each year from 2000 to 2015. We consider death data, data on deaths from related causes and in particular data on emissions of Ozone precursor gases, Acidifying gases and Fine Particulate Matter (PM10 and PM2.5).
An analysis of air pollution in relation to temperature and season of the year is given by the publication, for example a Bar Chart is used to compare PM2.5 levels in Beijing in relation to daily hours.



Subsequently, the publication focuses on analyzing the correlation between the PM2.5 emission and the Respiratory disease incidence trends.
Although the sampling scale is different from our project in some cases the same insights emerged. In our case, by selecting countries in a given population range and with different amounts of emissions, it can be observed that in most cases the amount of emissions is directly proportional to the number of deaths from related causes.
Another interesting publication is *Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015*, it is reiterated that exposure to ambient air pollution increases morbidity and mortality, it is a major contributor to the global disease burden. In particular, it explores spatial and temporal trends in mortality and disease burden attributable to ambient air pollution from 1990 to 2015 at global, regional and national levels.
Population-weighted annual mean concentrations of PM2,5 and tropospheric ozone are the two indicators used to quantify exposure to air pollution.
A risk estimation is calculated in according to the burden attributable to PM2-5 for ischemic heart disease (IHD), cerebrovascular disease (ischemic stroke and hemorrhagic stroke), lung cancer, chronic obstructive pulmonary disease (COPD), and lower respiratory infections (LRI), and the burden attributable to ozone for COPD are analysed.



The publication also estimates trends in population-weighted average mass concentrations of particles with aerodynamic diameters below 2,5µm.
A Bar Chart is used to report the deaths attributable to ambient particulate matter pollution by year and cause.



A comparison with our project shows some similarities, such as the analysis of PM 2.5 emissions and deaths from diseases attributable to them.
The detailed publication uses curves showing the central integrated exposure-response estimates and their 95% uncertainty intervals.
One difference is the use of the Bar Chart, in the publication it is used to show deaths attributable to PM2.5 exposure at 5-year intervals. Our project, on the other hand, uses this chart to show the number of deaths per country, in particular the percentage of deaths due to related causes is displayed.
The results of this publication and our project seem to coincide in some ways, in both cases showing the close relationship between exposure to air pollution and the increase in deaths from related causes.

References

- https://d3js.org/
- https://pdfs.semanticscholar.org/dd2b/d84ca080a06e896eb89c00cd2c2324e5ef9a.pdf
- https://reader.elsevier.com/reader/sd/pii/S0140673617305056
- https://edgar.jrc.ec.europa.eu/overview.php?v=50_AP
- http://ghdx.healthdata.org/
- https://data.worldbank.org/
- https://ourworldindata.org/