Interactive Data Visualization for Exploring Health impacts from Social and Environmental Factors

Team181 - Andy Wang, Alex Stanley, Masakazu Nakayama, Paul Kim, Kaan Taner

Motivation

What is the problem?

Multi-method analysis of health impacts caused by air pollution.

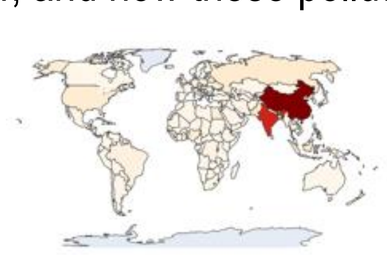
Why is it important and why should we care?

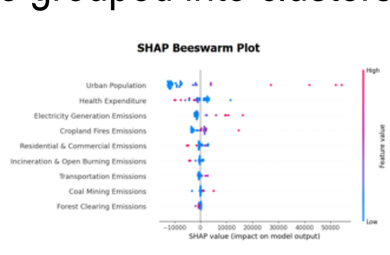
To formulate effective policies for our future, it is essential to conduct multi-method analysis to identify which countries or states are most severely affected by health impacts from air pollution and to determine which environmental and socio-economic factors have the greatest influence.

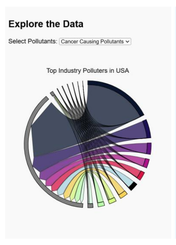
Approaches

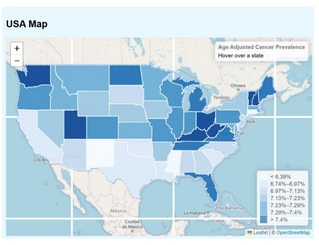
What are they?

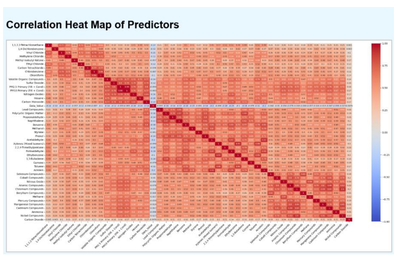
On a global scale, we analyzed which countries have the highest number of deaths caused by air pollution and which environmental and socio-economic factors have the greatest influence. On a domestic scale, we examined which states have higher cancer prevalence, which industries emit pollutants highly correlated with cancer, and how these pollutants are grouped into clusters.











How do they work?

Both global and domestic scales utilize choropleth maps to visualize regional characteristics. Additionally, by applying XAI methods to the regression predictions generated by boosting and stacking ensemble learning models, the contribution of individual features is visualized, enabling detailed factor analysis. For interactive exploration, chord diagrams are employed to allow deeper investigation into regional factors. A correlation matrix of pollutants is also provided a deeper understanding of the situation.

Why do we think they can effectively solve our problem?

It provides sufficient input for decision-making on selection and concentration necessary for policy formulation.

What is new in our approaches?

Our approach introduces new elements such as integrating analyses across multiple scales (global and domestic), utilizing XAI methods for deeper factor analysis, and incorporating interactive visualizations like chord diagrams.

Data

How did we get it?

Utilization of open data from WHO, the World Bank, and environmental organizations for global scale analytics. Domestic-scale analysis utilizing open data from the EPA and CDC.

What are its characteristics?

Various types of environmental and socio-economic factors related to health impacts caused by air pollution, observed simultaneously across 182 countries worldwide, as well as various air pollutants and industry data across all states in the United States.

Experiments and Results

How did we evaluate our approaches?

The global map clearly illustrates the severity of health impacts caused by air pollution. The interactive scatter plot visualizes several highly correlated factors, while the SHAP Beeswarm plot provides important insights into the structural background of health impacts from air pollution. The domestic map highlights clear contrasts in health impacts across states. By utilizing the interactive chord diagram integrated with the map, we can conduct in-depth exploratory analysis to identify which industries have the most significant impacts. Additionally, clustering through a correlation matrix and classifying each cluster using KNN allows for a deeper understanding of the pollutants themselves.

What are the results?

In global-scale analysis, it was shown that the positive correlation of urban population, the negative correlation of healthcare expenditure, and the positive correlation of power generation have a greater impact than incidents such as agricultural fires. **This suggests that governments need to focus not only on addressing direct incidents like agricultural fires but also on policies such as promoting decentralized societies, strengthening healthcare systems, and advancing clean energy adoption.** In domestic-scale analysis, it was found that in many states, fossil fuel power generation emits significant amounts of nitrogen oxides, which are linked to cancer development. **It suggests that moving away from fossil fuels is an urgent necessity, even in today's context of increasing electricity consumption driven by generative AI and other factors.**

How do our methods compare to other methods?

Traditional air pollution analyses often lack data diversity and exploratory approaches. Our method addresses these gaps and excels in providing visualizations that offer intuitive and insightful findings.