

Data Visualisation on Mortality Rate and Air Pollution in South East Asia Countries between 1990-2019

COS30045

Data Visualisation

Semester 2, 2024

Presented to you by

**Mandy Si Yi LEE (102787473)**

**Fiona Teo** **(102789110)**

Contents

[Introduction 3](#_Toc183382850)

[Background and Motivation 3](#_Toc183382851)

[Visualisation Purpose 4](#_Toc183382852)

[Project Schedule 6](#_Toc183382853)

[Data 7](#_Toc183382854)

[Data Source (Proposal) 7](#_Toc183382855)

[Data Source (Final) 7](#_Toc183382856)

[Data Processing (Proposal) 7](#_Toc183382857)

[Data Processing (Progress) 8](#_Toc183382858)

[Primary Dataset Overview 8](#_Toc183382859)

[Data Processing Workflow 8](#_Toc183382860)

[Outputs and Results 9](#_Toc183382861)

[Data Processing (Final) 9](#_Toc183382862)

[Requirements 10](#_Toc183382863)

[Must-Have Features (Proposal) 10](#_Toc183382864)

[Overview Intro Page 10](#_Toc183382865)

[About Us Page 10](#_Toc183382866)

[Documentation Page 10](#_Toc183382867)

[Geopath Graph 10](#_Toc183382868)

[Must-Have Features (Final) 10](#_Toc183382869)

[Optional Features (Proposal) 11](#_Toc183382870)

[1. Bar Chart 11](#_Toc183382871)

[2. Stacked Area and Line Chart 11](#_Toc183382872)

[3. Ranking 11](#_Toc183382873)

[Optional Features (Final) 11](#_Toc183382874)

[Visualisation Design (Proposal) 12](#_Toc183382875)

[Visualisation Design (Progress) 12](#_Toc183382876)

[Visualisation Design (Final) 13](#_Toc183382877)

[Validation 13](#_Toc183382878)

[Conclusion 13](#_Toc183382879)

[References 14](#_Toc183382880)

[Appendix 15](#_Toc183382881)

# Introduction

## Background and Motivation

Air pollution is a pressing public health concern in Southeast Asia, significantly contributing to mortality rates across the region. Major pollutants, such as fine particulate matter (PM2.5), sulfur dioxide (SO2), and nitrogen dioxide (NO2), arise from industrial activities, vehicular emissions, and biomass burning (WHO 2023). Southeast Asia, with its rapid urbanization and economic growth, faces unique challenges in balancing development and environmental sustainability (OECD 2022).

While air pollution has long been recognized as a public health threat, its perception as a critical issue began to gain momentum in the early 2000s, spurred by increasing awareness campaigns and scientific evidence linking it to severe health outcomes. Governments, organizations, and advocacy groups have since taken steps to prioritize air quality management. However, the rapid pace of industrialization and modernization continues to correlate strongly with rising pollution levels. The region's dependence on fossil fuels, combined with practices like open burning, exacerbates exposure to harmful pollutants.

According to the WHO, outdoor air pollution caused approximately 650,000 deaths in Southeast Asia in 2019, with PM2.5 being the leading pollutant (WHO 2023). Beyond direct health outcomes, such as respiratory and cardiovascular diseases, air pollution also imposes broader societal and economic burdens, including reduced productivity and healthcare strain. Vulnerable groups, such as children and the elderly, suffer disproportionately, both physically and socioeconomically.

This project is designed to communicate the impact of air pollution to the general public, presenting the data in an accessible and compelling way. By visualizing the multifaceted effects of air pollution—using a combination of maps, charts, and interactive tools—we aim to foster awareness and inspire action.

Rather than focusing solely on premature deaths, we employ Disability-Adjusted Life Years (DALYs) as a metric to provide a more comprehensive picture. DALYs account for both years lost due to premature death and years lived with illness or disability caused by pollution-related diseases. This approach highlights that air pollution doesn't just affect us directly; it also impacts our quality of life indirectly through chronic illnesses, reduced productivity, and financial burdens.

Our goal is to empower the public with clear and relatable insights, enabling them to understand the trade-offs between economic growth and environmental health. By focusing on Southeast Asia—a region where air pollution and rapid industrialization intersect—we hope to inspire individuals, communities, and policymakers to advocate for sustainable solutions.

Through this visualization project, our goal is to bridge the gap between complex data and actionable insights, highlighting the urgent need for sustainable and effective policies. By exploring correlations between air pollution and public health across Southeast Asia, we aim to foster public awareness and encourage decision-makers to take proactive steps toward reducing pollution. We believe that data-driven insights can inspire meaningful change, helping balance economic development with environmental sustainability for the well-being of the region’s population.

## Visualisation Purpose

With the raw data provided by OECD, the vast volume of data on mortality rates globally since 1990 till 2019 can be overwhelming and difficult to interpret without data processing made. Presenting basic statistics on mortality rates or pollution levels in summary tables or purely on premature deaths does not adequately convey the depth of the issue. Therefore, the primary objective of our visualization is to leverage real-world data, transforming it into an engaging, interactive format that enables stakeholders to better understand the complexities of air pollution and mortality trends in Southeast Asia.

Through our visualizations, complex datasets are distilled into intuitive and accessible graphics, allowing users to quickly grasp important patterns and insights. Users can engage with the data interactively, filtering for specific years and age, and hovering in on areas of interest to examine trends in detail.

The purpose of these visualizations is to clearly communicate the relationship between air pollution and mortality, highlighting the impact of ambient particulate matter and the effectiveness of policies in mitigating these effects. Specifically, our visualizations will help users answer key questions, such as:

**Data Trends and Comparisons:**

* How have mortality rates for different age groups changed over time in Southeast Asia?
* Which age group is most affected by ambient particulate matter in Southeast Asia?
* Are there significant differences in mortality patterns among the Southeast Asian countries?
* Which year saw the highest overall mortality rate in Southeast Asia?
* What is the trend in mortality rates among children compared to the elderly?

**Regional Analysis:**

* How does the mortality rate in urbanized countries like Singapore compare to less developed countries in the region?
* Is there a noticeable difference in mortality rates between mainland Southeast Asia and island nations?
* Which countries show a significant decline or increase in mortality rates over the years?

**Impact of Risk Factor:**

* What factors might contribute to the variations in mortality rates across countries?
* Are countries with higher urbanization levels experiencing higher mortality rates due to ambient particulate matter?
* Which age group is most affected by ambient particulate matter in Southeast Asia?
* Are there significant differences in mortality patterns among the Southeast Asian countries?
* Which year saw the highest overall mortality rate in Southeast Asia?
* What is the trend in mortality rates among children compared to the elderly?

Due to time constraints, not all the questions will be answer via the visualization. Therefore,

the visualization delivered by the team allow to answer the following questions:

* What is the correlation between air pollution levels and mortality rates across Southeast Asian countries from 1990 to 2019?
* How do air quality standards and regional policies affect public health outcomes?
* Which countries in Southeast Asia face the greatest public health risks from air pollution, and which populations are most vulnerable?
* How have mortality rates for different age groups changed over time in Southeast Asia?

By transforming data into a more digestible and interactive format, we aim to foster better decision-making and raise awareness about the urgent need for sustainable policies in Southeast Asia’s fight against air pollution.

## Project Schedule

Week 4: Discussed and chose a topic for our project, that is “Air Pollution Mortality in Southeast Asia between 1990-2019”.

Week 5: Gather information that is related to our topic into our shared drive and have further discussion with the lecturer regarding our topic

Week 6: Visualise the data we have collected through software like Microsoft Excels and let the lecturer view our ideas and datas gathered.

Tuition week: Discuss on the UI design and style of the project website.

Week 7: Start out the coding of the website through github shared repository and update on the lecturer regarding our process.

Week 8: At least 30% of the website must be done and update on the lecturer regarding our process.

Week 9: At least 65% of the website must be done and update on the lecturer regarding our process.

Week 10: At least 85% of the website must be done and update on the lecturer regarding our process.

Week 11: Final touch of the website must be done and update on the lecturer regarding our process.

# Data

## Data Source (Proposal)

The data was collected from OECD iLibrary, specifically from OECD Environment Statistics, which is the main requirement for this particular assignment. The datasets came in Comma-separated values (CSV) file. Due to the size of the file, the team needs to manually extract the data of the countries we are focusing on into another csv data as it was impossible to open the raw file due to device processing constraints. This dataset included various types of attributes such as categorical attributes for countries, interval attributes for year and quantitative attributes for values.

The chosen dataset that will be used in visualisation to answer the questions:

[Air quality and health: Mortality and welfare cost from exposure to air pollution (Edition 2020)](file:///C:\q\data\EXP_MORSC-2020-1-EN-20210323T100329.csv)

## Data Source (Final)

The finalized datasets for the visualization project on mortality rates and air pollution trends in Southeast Asian countries over the past 29 years (1990–2019) have been compiled and thoroughly verified. During the data processing stage, unnecessary attributes unrelated to the core research questions were removed to streamline the analysis. The data was sourced from the following references:

(Dataset: Air quality and health: Mortality and welfare cost from exposure to air pollution (Edition 2020))

|  |  |  |
| --- | --- | --- |
| Attribute | Description | Type |
| Country(s) | Name of the country. | Categorical (Text) |
| Year(s) | Year for data collection. | Ordinal (Number) |
| Risk | Type of risk factor (e.g., Ambient Particulate Matter). | Categorical (Text) |
| Variable | Indicator used (e.g., DALYs). | Categorical (Text) |
| Value | Numerical value of the indicator. | Quantitative/Continuous (Numerical) |
| Unit | Measurement unit (e.g., Number). | Categorical (Text) |
| Sex | Gender category (e.g., Both Male & Female). | Categorical (Text) |
| Age | Age group (e.g., More than 64 years old). | Categorical (Text) |

## Data Processing (Proposal)

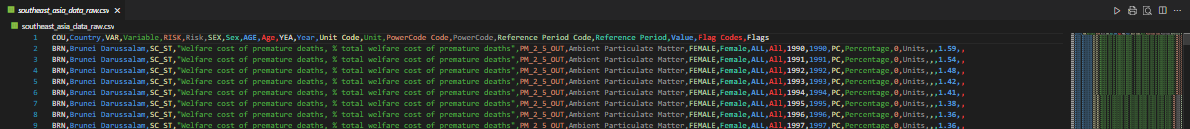
As the original document titled “Air quality and health: Mortality and welfare cost from exposure to air pollution (Edition 2020)” by the OECD Environment Statistics database contains extensive global data spanning from 1990 to 2020, including all data of different mortality caused by other factors beside ambient particulate matter, it was too large to be opened on the current device. To resolve this issue, multiple sources from online forums, including StackOverflow and CSV Explorer, were referred for potential solutions. In order to manage and process the data effectively, based on the advices on StackOverflow, the Python data processing library Pandas will be used to extract relevant data and save it into a more manageable CSV file for further analysis.

## Data Processing (Progress)

In this section, we detail the methodology employed to process the diverse datasets collected from the OECD for this project. Our primary data source is the OECD dataset, which provides extensive information in CSV format, covering a range of years from 1990 to 2020. This serves as the foundation for analyzing the impact of air pollution on mortality rates in Southeast Asia.

### Primary Dataset Overview

The OECD dataset encompasses global data, including mortality caused by other factors. For this project, we focused on extracting the data relevant to Southeast Asian countries, that are Brunei Darussalam, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand, Timor-Leste, and Vietnam.



From the available risk types in the dataset, we specifically filtered data related to Ambient Particulate Matter—a key pollutant significantly affecting mortality rates. The extracted data is processed on a yearly basis, ensuring precise temporal tracking of mortality trends.

### Data Processing Workflow

To transform the raw dataset into a structured and usable format, we employed Python's pandas library, ensuring efficient and streamlined data processing. Given the dataset's extensive size, we loaded the first 3 rows of the raw data file to locate the header. Relevant columns, including Country, Year, Risk, Value, Age, and Sex, were selected for analysis, while rows with missing data were excluded to maintain data integrity.

A computer screen shot of a computer code

Description automatically generated A screenshot of a computer

Description automatically generated

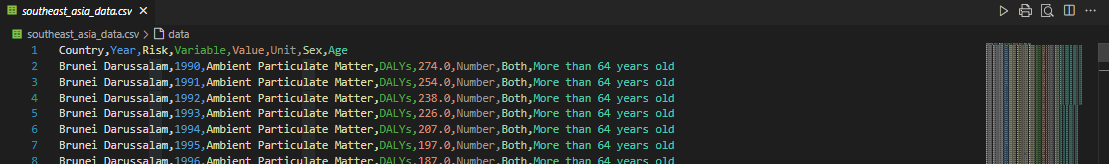
The filtering process specifically focused on Southeast Asian countries, retaining records where the sex category was "Both" and the risk type was Ambient Particulate Matter. The data was further organized by sorting it based on predefined age categories, ensuring clarity and consistency in the analysis. Finally, the mortality values were averaged by country and year, resulting in aggregated insights that preserved the regional specificity of the data. This meticulous approach facilitated the generation of meaningful and actionable findings. It was loaded iteratively in chunks of 4,000 rows to optimize memory usage. This approach allowed for seamless handling of large amounts of data without overwhelming system resources.

A screenshot of a computer program

Description automatically generated

### Outputs and Results

The final processed data highlights the annual average mortality rates due to Ambient Particulate Matter for each Southeast Asian country between 1990 and 2019. The results were outputted both as a formatted preview and saved to a CSV file for visualization and further analysis.



## Data Processing (Final)

# Requirements

## Must-Have Features (Proposal)

### Overview Intro Page

This website will feature an introductory page designed to provide users with a comprehensive overview of the project on Air Pollution Mortality in Southeast Asia between 1990-2019. The page will include the project title, a brief introduction to the significance of air pollution as a pressing environmental and health issue in Southeast Asia, and the project's aims and objectives. The objective is to offer a clear and engaging entry point, setting the context for the data visualizations and analyses presented on the website.

### About Us Page

The "About Us" page will highlight the development team behind the project. Each member’s profile will feature a photograph, a concise biography outlining their expertise and contributions, and a description of their role in the project. This personal touch is designed to foster trust and credibility with users by showcasing the team's qualifications, dedication, and the collaborative effort that went into creating the project.

### Documentation Page

To ensure academic rigor and transparency, the documentation page will provide detailed information about the data sources utilized in this project. It will include citations for the primary dataset from the OECD Environment Statistics database and supplementary sources such as GeoJSON data from GitHub. Additionally, it will outline the methodologies used for data collection, processing, and analysis. This page will serve as a reliable reference for users, substantiating the credibility of the data visualizations and insights presented on the website.

### Geopath Graph

A geopath graph will be employed to depict the mortality rate from air pollution across Southeast Asia. This dynamic visualisation will use the colour depth on different countries based on the number of mortality. The deeper the depth of colour, the higher the casualties.

This graph will offer valuable insights into the mortality caused by air pollution, illustrating how serious the air pollution is towards one’s health. A time slider will enhance the geopath graph, allowing users to observe changes in mortality rate intensities over the years.

## Must-Have Features (Final)

## Optional Features (Proposal)

### Bar Chart

The bar chart will visualize the annual mortality rates due to air pollution for each Southeast Asian country. Each bar will represent a specific country, with the height of the bar indicating the number of deaths attributable to ambient particulate matter (PM2.5) for a given year. The chart will include options to filter or group data by age categories through "Less than 15 years old", "15 to 64 years old", "More than 64 years old" and "All" to show how different demographic groups are affected.

To enhance interactivity, users can toggle between viewing mortality rates for individual years or viewing trends over the entire 1990-2019 period. The bar chart will also allow for side-by-side comparisons between countries, highlighting disparities in mortality rates and the effectiveness of air quality policies across the region. This visual tool aims to make the impact of air pollution on health more accessible and comprehensible.

### Stacked Area and Line Chart

The visualizations will include a stacked area chart to show mortality trends in Southeast Asia due to air pollution, broken down into categories such as age groups or other demographics. This will provide a clear understanding of how mortality has shifted over time. Complementing this will be a line chart that displays changes in pollution levels (e.g., PM2.5 concentrations) across the same time frame. Together, these tools will help identify correlations between air pollution and mortality trends.

### Ranking

A ranking of mortality rates due to air pollution in Southeast Asian countries over the years will provide valuable insights into how each country has been impacted. By comparing mortality rates, stakeholders can identify countries with the highest burdens and evaluate the effectiveness of regional policies and interventions in mitigating the effects of air pollution.

## Optional Features (Final)

# 

# Visualisation Design (Proposal)

A white board with writing on it

Description automatically generated

A white line on a white background

Description automatically generated

A map of asia with different countries/regions

Description automatically generated

A white board with writing on it

Description automatically generated

# Visualisation Design (Progress)

A screenshot of a computer screen

Description automatically generated

# Visualisation Design (Final)

# Validation

# Conclusion

# References

---will be cleaned by the end of project book for us to find references easily

---Background

Koplitz, SN, Jacob, DJ, Sulprizio, MP, Myllyvirta, L & Reid, C 2021, 'Burden of disease from rising coal-fired power plant emissions in Southeast Asia', Environmental Science & Technology, vol. 55, no. 15, pp. 10150–10160, viewed 17 November 2024, https://doi.org/10.xxxx.

Organisation for Economic Co-operation and Development (OECD) 2022, The economic cost of air pollution in Southeast Asia, OECD Publishing, viewed 17 November 2024, https://www.oecd.org/southeast-asia.

World Health Organization (WHO) 2023, Air pollution and public health: Southeast Asia overview, viewed 17 November 2024, <https://www.who.int/regions/searo>.

--Data processing

https://stackoverflow.com/questions/11622652/large-persistent-dataframe-in-pandas/12193309#12193309

# Appendix