**Project Title:** The Price of Progress: How the Pursuit of Income and Trade Fuels Pollution Mortality.

**Introduction**

Climate change is a long-running problem that by now is common knowledge, yet, society broadly sweeps under the rug. Aside from climate change's many problems, however, pollution also causes its own fair share of problems, a cost I want to investigate to bring more light to.

**Questions**

1. How much is pollution killing us, and how does it vary across income levels?

2. How does a country's income level influence its disposition to pollute?

3. How do trade dynamics conflate with waste management and pollution-related mortality?

**Data Wrangling and Checking**

The data sources used were:

1. GDP per capita, 2023. 7064 rows x 4 columns. Columns are country, country code, year of data, GDP per capita, PPP (adjusted for capita and purchasing power parity). https://ourworldindata.org/grapher/gdp-per-capita-worldbank

2. Pollution death rates, 2019. 197 rows x 4 columns. Columns are country, country code, year of data, age-standardized mortality rate attributed to household and ambient air pollution. This denotes the per-capita death rate by air pollution.  
https://ourworldindata.org/grapher/death-rate-household-and-ambient-air-pollution

3. Plastic waste mismanagement, 2019. 171 rows x 4 columns. Columns are country, country code, year of data, share of global mismanaged plastic waste. This denotes the global share percentage of mismanaged plastic waste.  
https://ourworldindata.org/grapher/share-of-global-mismanaged-plastic-waste

4. Imports as % of GDP, 2023. ~270 rows x ~50 columns. Represents imports of goods and services in % of GDP with each row containing the country, country code and an array of data for each year from 1960 to 2023.   
https://data.worldbank.org/indicator/NE.IMP.GNFS.ZS?view=map

5. Exports as % of GDP, 2023. ~270 rows x ~50 columns. Represents exports of goods and services in % of GDP with each row containing the country, country code and an array of data for each year from 1960 to 2023.  <https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS?view=map>

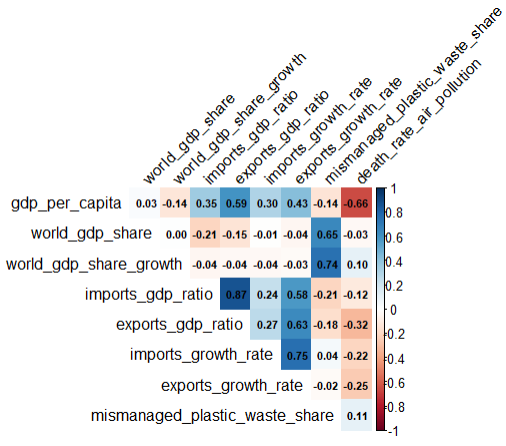
6. Population data, 2023. 18945 rows x 3 columns. Columns are country, year, population. Although not originally included, this is necessary to compute the global share of GDP from the GDP per capita, a metric I want to compare. The data is sourced from the same source of ourworldindata.org. <https://ourworldindata.org/explorers/population-and-demography?tab=table&indicator=Population&Sex=Both+sexes&Age=Total&Projection+scenario=None&country=CHN~IND~USA~IDN~PAK~NGA~BRA~JPN>

7. Percentage of world GDP, 2023. 7064 rows x 4 columns. Columns are country, country code, year, percentage of world GDP of the country. This data is computed from GDP per capita [1] and population data [6] using R to multiply the GDP per capita with their respective population for their respective years, and divided by the world GDP per capita multiplied by the world’s population for the year.

Data sources [4] and [5], the imports and exports as a percentage of the country’s GDP, were originally in wide format, with the import/export information being spread out across the row and a separate column being used for each year’s data. This had to be flattened into a long format for use with the rest of the data.

Many different data sources were used for this analysis. The data was merged by year and country with R. As the plastic waste mismanagement and the death rates by air pollution data is only available for 2019, the rate of change of GDP per capita, the percentage of world GDP per country, imports as percentage of GDP and exports as percentage of GDP were also computed to be used as growth rate variables to be compared and analysed against the pollution data.

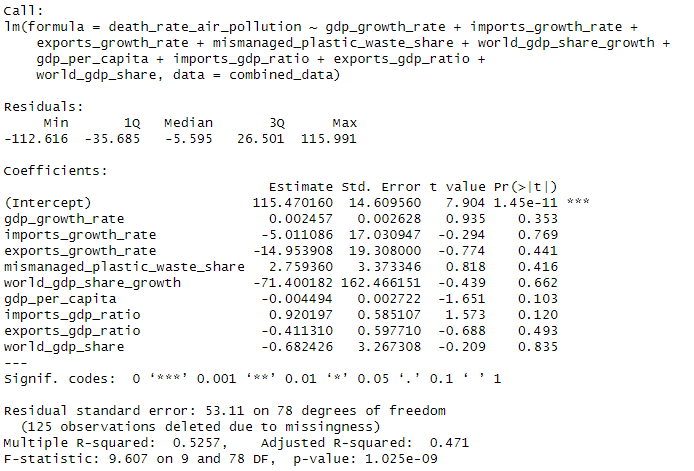
Many NA values were found in the data. They were removed/ignored to avoid making assumptions about them, as I do not want to misrepresent them. The names of the variables were also not consistent i.e. “Country name” could be “Country”, “Country Name”, “Entity”. They were identified by column order after manual inspection of the files.

A correlation matrix was run in R to compare these variables. The results are:  


At first glance, this suggests that richer countries have lower air pollution death rates, and that countries which contribute a larger portion of the world GDP also contribute a larger portion of the world’s plastic waste. However, richer countries are known to have better healthcare and living environments which directly contribute to lower death rates as a whole, so further analysis is needed.

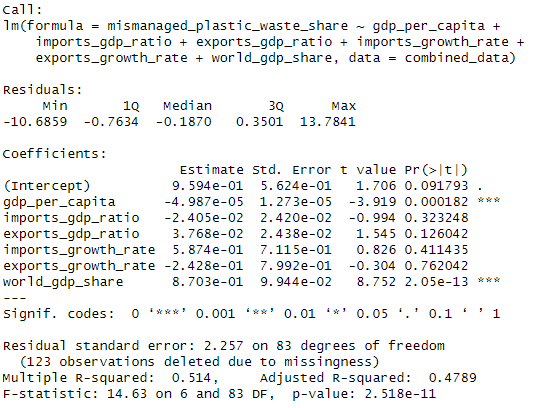
A multiple regression analysis was also run in R to identify contributing factors for air pollution death rates and mismanaged plastic waste share.

Air pollution multi-regression:



The analysis concludes that the overall model is statistically significant (F-statistic p < 0.001) and explains about 47.1% of the variance in air-polluted death rates. However, none of the individual factors are statistically significant (PR (>|t|) < 0.05). GDP per capita is marginally correlated with lower death rates, and a country importing a larger percentage of their GDP is marginally correlated with higher death rates after controlling for the listed factors, but ultimately, air pollution death rate is a multicollinear problem with more factors than identified here.

Mismanaged plastic waste share multi regression:



The analysis concludes that the overall model is statistically significant (F-statistic p < 0.001) and explains about 47.9% of the variance in air-polluted death rates. GDP per capita is correlated with slightly less global mismanaged plastic waste share, but when considering global GDP share instead, it becomes a very strong correlation. This highlights that richer countries are marginally more efficient, but contribute vastly more in total to global plastic waste mismanagement. There is also a marginally significant correlation of export-based economies contributing to global plastic waste mismanagement, which hints that plastic waste mismanagement could occur in countries which are selling the contributing products to other countries. Ultimately, more data is once again needed to identify the missing contributing variables

**Data Exploration**

**Conclusion**

**References**

I used Google Gemini to brainstorm and find relevant, free and public data.