Name: Adrian Leong Tat Wei Student ID: 27030768

Tutor: Dr. Grace Ting Chai Wen, Applied Session Friday 2-4pm Malaysia

The Price of Progress: How the Pursuit of Income Fuels Pollution Mortality.

Contents

Introduction

Climate change brings vast and profound implications, yet pollution continues to pose a substantial and distinct threat. Rather than solely detailing the already well-documented harms of pollution, I wish to delve into its root causes and the systems that perpetuate it. Concerned by the potential for today's seemingly insignificant actions to fuel a future of the destruction of our own planet, and the lack of action by society, I aim to uncover the "where" and "what" of pollution's origins. By focusing on these underlying factors, I hope to provide a clearer understanding of what drives this unending treadmill of materialism.

Questions

1. How much is pollution killing us? Who suffers?

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

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 with 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.

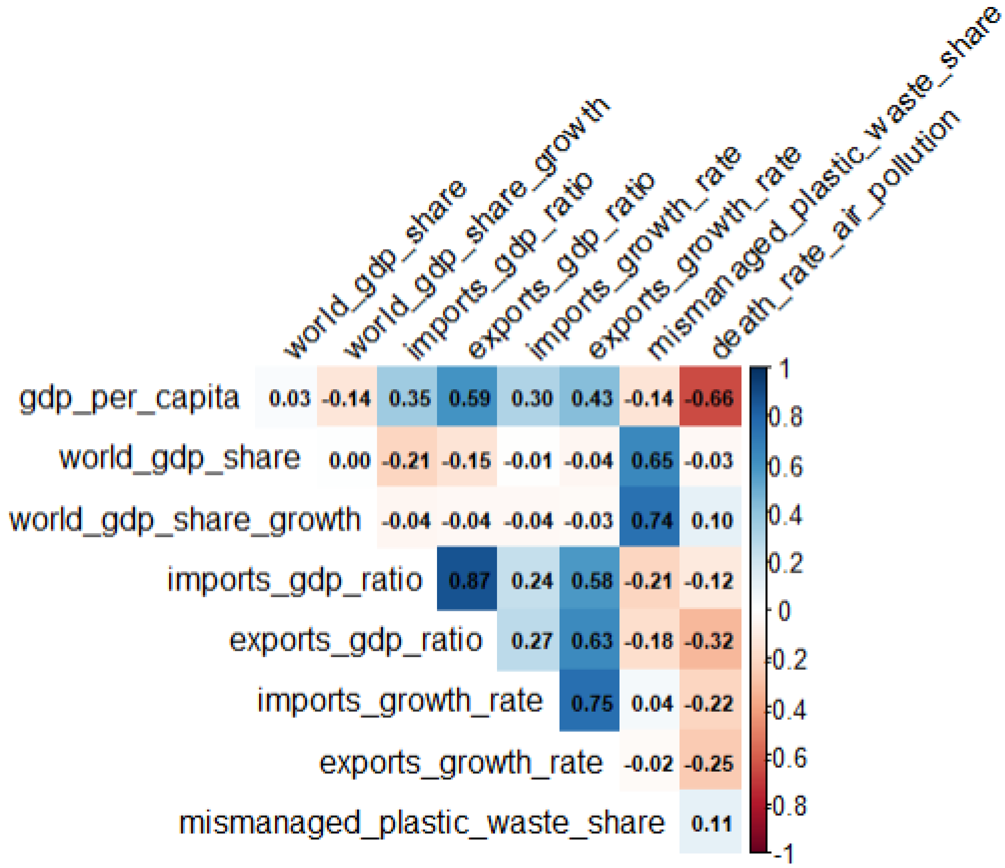
Many different data sources were used for this analysis. The data was merged by year and country with R. As the plastic waste mismanagement and the death rates by air pollution data is only available for 2019, the rate of change of GDP per capita, the percentage of world GDP per country, imports as percentage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

Data Exploration

Statistical Tests

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.

Multiple regression analyses were conducted in R to identify factors contributing to air pollution death rates and mismanaged plastic waste share.

Air pollution multi-regression:

```
Call:
lm(formula = death_rate_air_pollution ~ income_group + imports_gdp_ratio +
    exports_gdp_ratio, data = combined_data)

Residuals:
     Min       1Q   Median       3Q      Max
-122.700  -20.007   -7.198   22.978  133.419

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      17.2215    10.4751   1.644   0.1025
income_groupLow income          174.7325    14.4147  12.122  < 2e-16 ***
income_groupLower middle income 127.2164    10.9603  11.607  < 2e-16 ***
income_groupUpper middle income  46.6070     9.8330   4.740 5.29e-06 ***
imports_gdp_ratio                 0.6805     0.3403   2.000   0.0475 *
exports_gdp_ratio                -0.4492     0.3162  -1.421   0.1576
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 43.5 on 137 degrees of freedom
  (70 observations deleted due to missingness)
Multiple R-squared:  0.6958,    Adjusted R-squared:  0.6847
F-statistic: 62.67 on 5 and 137 DF,  p-value: < 2.2e-16
```

Using income groups as classified by the World Bank, we find that income groups is a strong predictor of death rates attributed to air pollution. Low-income groups suffer the most, with lower-middle income being a runner-up, while higher income groups suffer much less. Being an import-based economy is statistically likely to increase deaths by air pollution, while being an export-based economy does the reverse, but the import-export composition of a country is in practice quite a small impact. The model is statistically significant with an adjusted R-squared value of 0.6847.

Mismanaged plastic waste share had a very weak correlation with death by air pollution. Using GDP per capita was also a much weaker correlation than using income groups – this resulted in an adjusted R-squared value of about 0.46. We hypothesize that in the case of plastic waste share, it is because the plastic waste is not normalized to the size of the country – the plastic waste data used here is percentage of mismanaged global plastic waste. For GDP per capita vs income group, it likely means that the many other factors the World Bank uses to classify a country’s income group beyond GDP per capita capture the factors that lead to health resilience much better.

```
Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                  1.405e+02  7.408e+00  18.971   <2e-16 ***
gdp_per_capita              -2.018e-03  2.052e-04  -9.837   <2e-16 ***
mismanaged_plastic_waste_share 4.631e-01  1.921e+00   0.241     0.81
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 57.25 on 137 degrees of freedom
  (73 observations deleted due to missingness)
Multiple R-squared:  0.421,     Adjusted R-squared:  0.4125
F-statistic: 49.81 on 2 and 137 DF,  p-value: < 2.2e-16
```

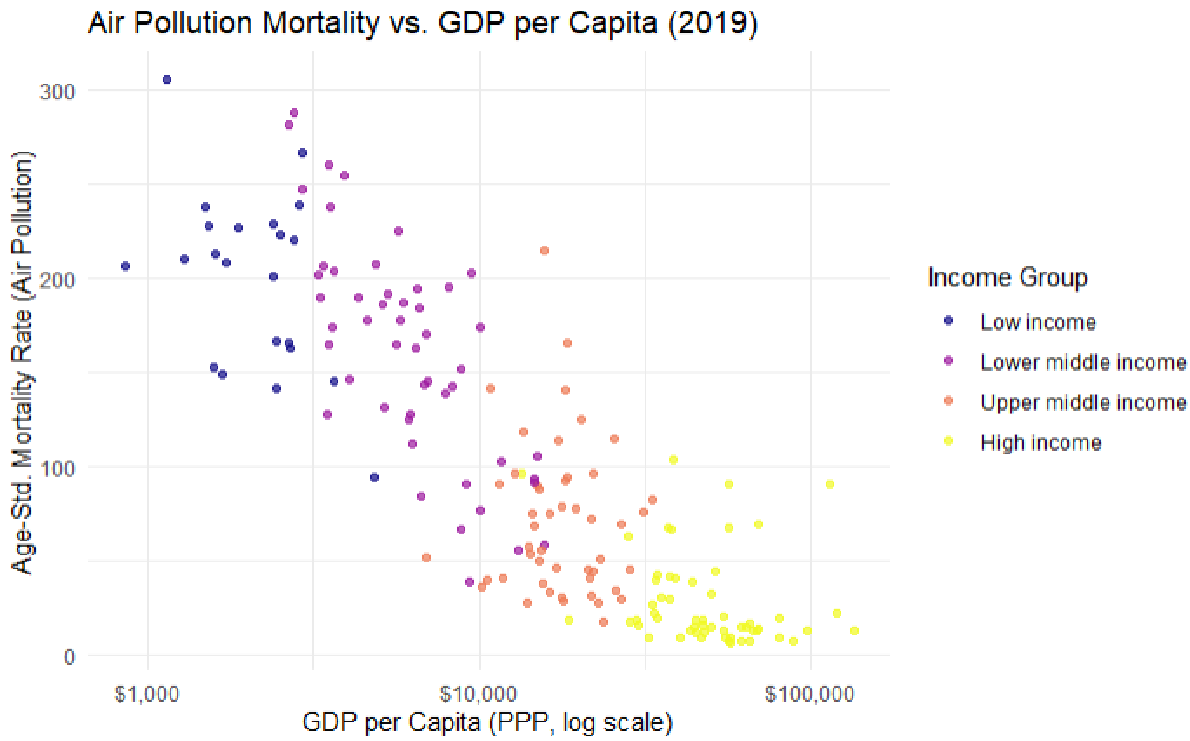
Mismanaged plastic waste share multi regression:

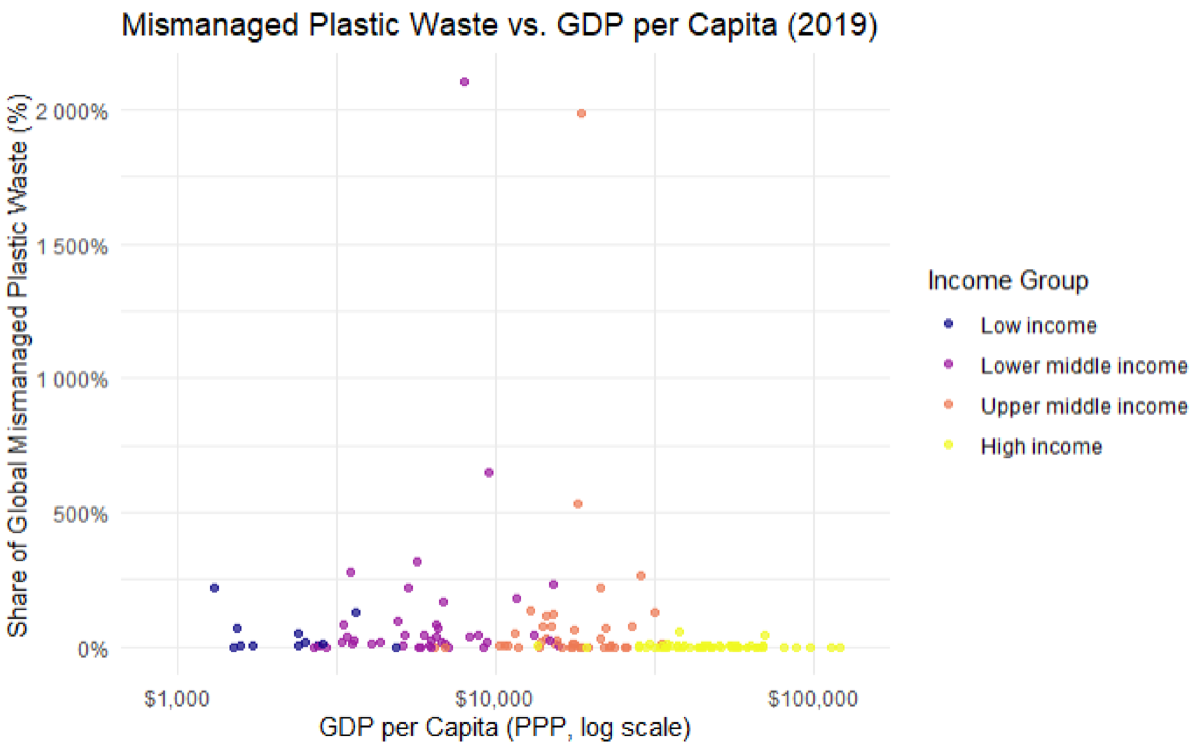
```
Call:
lm(formula = mismanaged_plastic_waste_share ~ income_group + 
    world_gdp_share + world_gdp_share_growth + imports_gdp_ratio + 
    exports_gdp_ratio, data = combined_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5417 -0.4964 -0.1280  0.1466 12.0861 

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)                      0.336414   0.434649   0.774   0.4406    
income_groupLow income           0.383978   0.645061   0.595   0.5529    
income_groupLower middle income  1.081578   0.409656   2.640   0.0095 ** 
income_groupUpper middle income  0.334810   0.380701   0.879   0.3811    
world_gdp_share                  0.451040   0.076070   5.929 3.64e-08 ***
world_gdp_share_growth          31.086938   3.400388   9.142 3.86e-15 ***
imports_gdp_ratio               -0.012975   0.010418  -1.245   0.2156    
exports_gdp_ratio                0.005417   0.010337   0.524   0.6013    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.511 on 109 degrees of freedom
  (96 observations deleted due to missingness)
Multiple R-squared:  0.7187,	Adjusted R-squared:  0.7006 
F-statistic: 39.78 on 7 and 109 DF,  p-value: < 2.2e-16
```

The analysis of mismanaged plastic waste share shows that the largest factor by far is the growth rate in the world GDP share. This far eclipses even the current world GDP share as a factor – a 1% share of the world GDP only correlates to contributing to 0.45% of the world's mismanaged plastic waste, but increasing in world GDP share by 1% in a year correlates to contributing to a staggering 31% of the world's mismanaged plastic waste. The lower-middle income group contributes notably more of the world's mismanaged plastic waste, but it is a small difference compared to its economic growth rate. A country's import/export composition is effectively irrelevant.

Visualizations

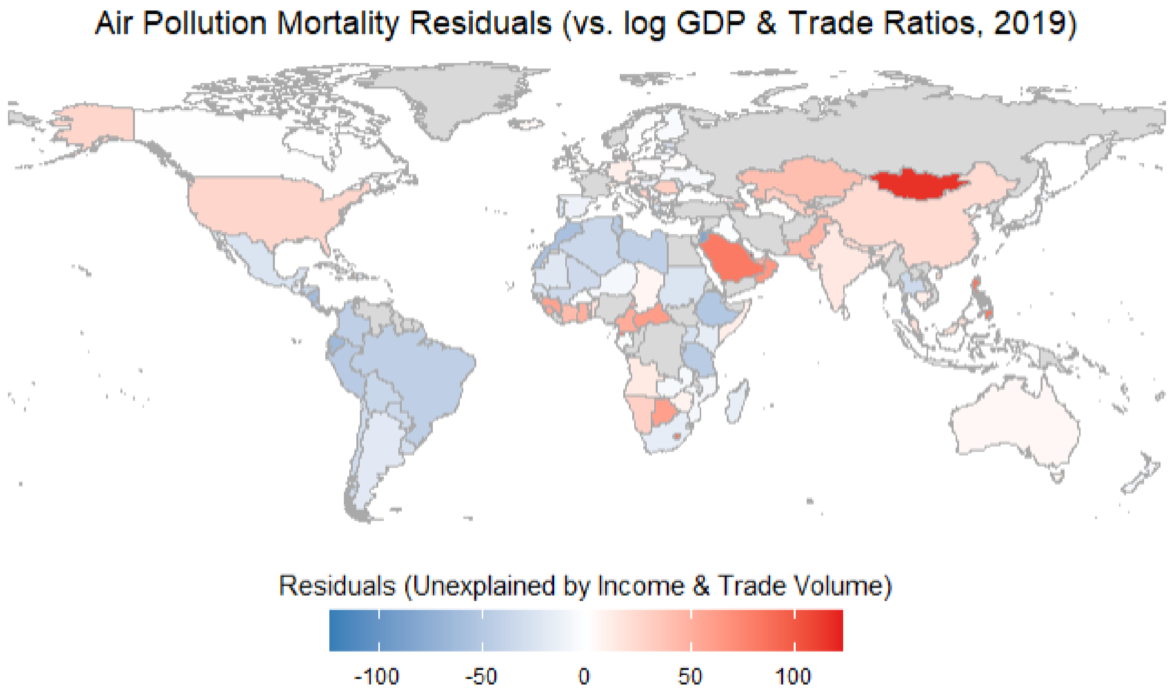
Having run the statistical tests above, it is time for visualizations. The first visualizations will be of the effect of GDP against air pollution mortality and mismanaged plastic waste share.

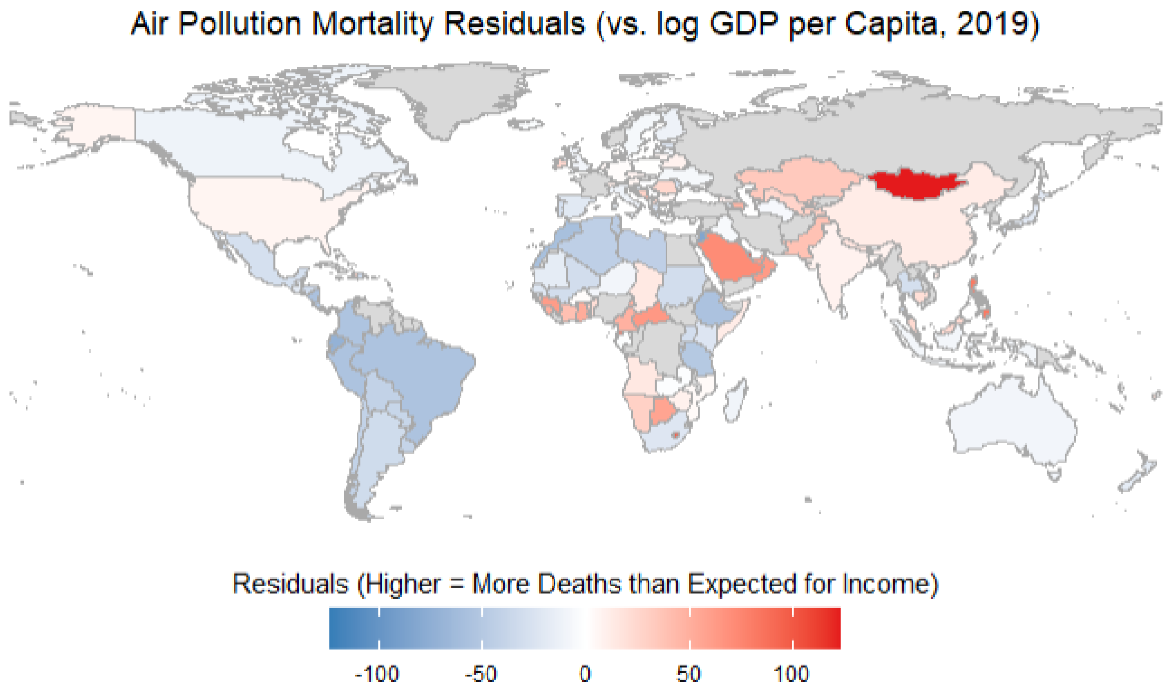




As seen, air pollution mortality very clearly goes down with GDP per capita. It is much less clear when it comes to plastic waste, though. This would be because the plastic waste scale is on a global scale, while the GDP scale is on a local scale (relative to its population).

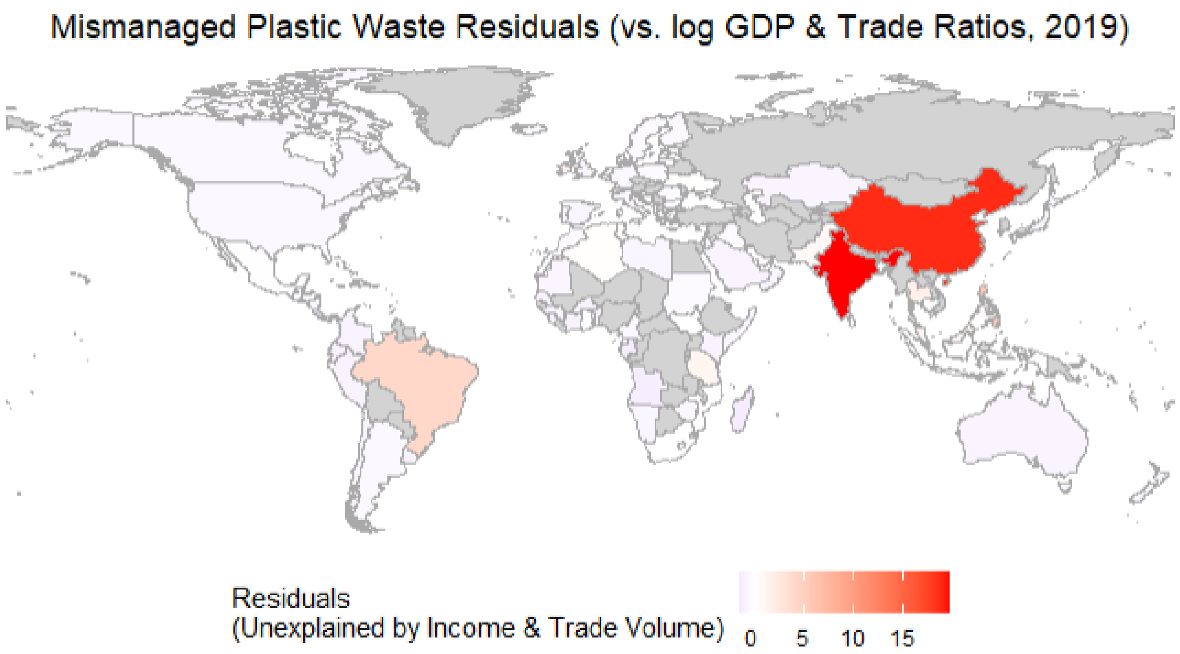
We next want to visualize how directly death rates by air pollution are correlated to GDP per capita and the import/export ratios of countries.

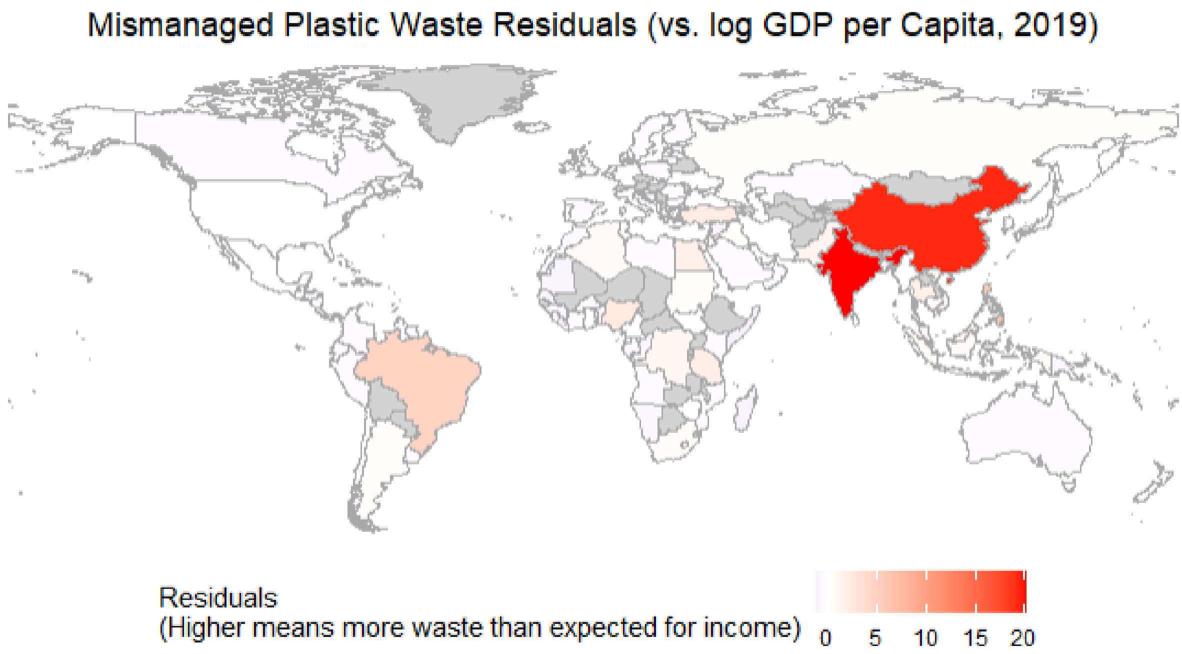




As seen from these visualizations, Mongolia is a notable outlier for having many more deaths than their economy would suggest, but for the rest of the world they align quite closely with a log scaled GDP per capita. Trade is quite a small factor, and doesn't help to explain much more just yet.

We also want to visualize how directly global mismanaged plastic waste share are correlated to GDP per capita and the import/export ratios of countries.

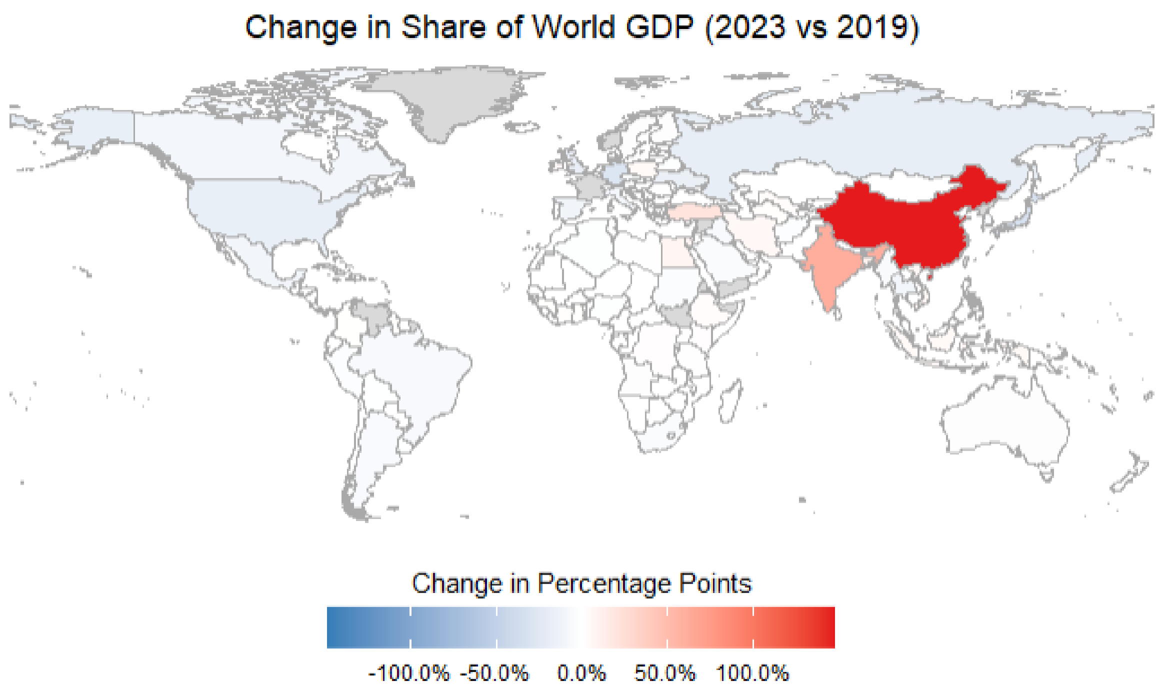




As these visualizations show, a very large proportion of the plastic waste share is attribute to China and India, far eclipsing every other country as notable outliers.

Conclusion

From the statistical tests run earlier, the percentage change in the share of the world's GDP is by far the strongest predictor of global share in mismanaged plastic waste. As such, we would like to identify factually which countries are the growing economies to watch out for growing mismanaged plastic waste.



This visualization points very clearly to China as the biggest outlier, with India as a runner-up. A notable conflating variable that was heard of, but not studied in this project, is that it has been said that many countries have been exporting their waste to China, which would disproportionately increase the waste production in China. However, studying that would vastly increase the difficulty of the study, and so it was left out.

To answer the original questions as asked in the introduction,

1. How much is pollution killing us? Who suffers?
Pollution is disproportionately killing the lower-income countries globally. It's likely that this is because they have less healthcare infrastructure to handle its effects.

2. How does a country's income level influence its disposition to pollute?
Interestingly, it is also lower-income countries that disproportionately mismanage waste. A notable plausible cause would be that richer countries may export their waste to lower-income countries, which would hide them as a root cause.

3. How do trade dynamics conflate with waste management and pollution-related mortality?
Not much at all – trade dynamics are scarcely a factor at all.

Reflection

I did not have all the data sources that could've been relevant. A more comprehensive dataset could have allowed for a more detailed analysis. Unfortunately, even in the initial stages it was difficult to find more comprehensive datasets available for free.

China and India are massive outliers for this. I am unsure how to account for them as they are such large proportions of the world too.

It was surprising how little plastic waste management mattered to influence death rates by air pollution. I already knew that the income level of a country vastly affected its life expectancy, but it was astounding to see it in practice.

I did not get to identify how geographical proximity to polluting countries affects death rates by air pollution. This would likely be very difficult to do.

References

GDP per capita. (2025, January 15). Our World in Data. https://ourworldindata.org/grapher/gdp-per-capita-worldbank

Death rate attributed to household and ambient air pollution. (n.d.). Our World in Data. https://ourworldindata.org/grapher/death-rate-household-and-ambient-air-pollution

Share of global mismanaged plastic waste. (n.d.). Our World in Data. https://ourworldindata.org/grapher/share-of-global-mismanaged-plastic-waste

World Bank Open Data. (n.d.). World Bank Open Data. https://data.worldbank.org/indicator/NE.IMP.GNFS.ZS?view=map

World Bank Open Data. (n.d.-b). World Bank Open Data. https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS?view=map

Population & Demography Data Explorer. (n.d.). Our World in Data. https://ourworldindata.org/explorers/population-and-demography

I used Google Gemini to brainstorm and find relevant, free and public data. I also used it to help me do the R coding.