Impact of Air Pollution on Chronic Respiratoy Diseases: An Exploratory Data Analysis of GBD Data

João Silva, Marco Pereira, Mariana Ribeiro

2025-10-22

Contents

1 Abstract

Este trabalho analisa os impactos da poluição do ar na saúde pública utilizando dados do Global Burden of Disease (GBD)(Bennitt et al., n.d.). Através de técnicas de preparação e análise exploratória de dados (EDA), são identificadas tendências, padrões e potenciais problemas de qualidade nos dados.

2 Introduction

A poluição do ar é um dos principais fatores de risco para doenças respiratórias e cardiovasculares. Este estudo visa explorar os dados do GBD para compreender melhor os efeitos da poluição do ar em diferentes países, faixas etárias e géneros.

3 Methodology

3.1 Data source

- Dados extraídos do GBD 2021
- Ferramenta utilizada: GBD Results Tool
- Link da pesquisa e ficheiro: 1990 + Global + SDI + Health System https://vizhub.healthdata.org/gbd-results?params=gbd-api-2021-permalink/4ea5715918446e5a6d9b154d62e0cc4a IHME-GBD_2021_DATA-4835a3dc-1.csv 2000 + Global + SDI + Health System https://vizhub.healthdata.org/gbd-results?params=gbd-api-2021-permalink/64781d061f111ef4af5ad17974b6eb98 IHME-GBD_2021_DATA-c56a3848-1.csv 2010 + Global + SDI + Health System https://vizhub.healthdata.org/gbd-results?params=gbd-api-2021-permalink/88ae5e347d197231fa598e3dfc8e219a IHME-GBD_2021_DATA-1923af35-1.csv 2020 + Global + SDI + Health System https://vizhub.healthdata.org/gbd-results?params=gbd-api-2021-permalink/2d070aac165f1d6a455d41df6c34e501 IHME-GBD_2021_DATA-d14075a8-1.csv 2021 + Global + SDI + Health System https://vizhub.healthdata.org/gbd-results?params=gbd-api-2021-permalink/3fad5e919cb9822381a5b56543978c2a IHME-GBD_2021_DATA-840155c6-1.csv 1990 + Países todos + Age (all + standardized) + Sex (Both) https://vizhub.healthdata.org/gbd-results?params=gbd-api-2021-permalink/8dcef8a43426d16927c72f4d2a96147c IHME-GBD_2021_DATA-7baf5a43-1.csv 2000 + Países todos + Age (all + standardized) + Sex (Both) https://vizhub.healthdata.org/gbd-results?params=gbd-api-2021-permalink/137d422bc7e27adb3442cd2a0c699368 IHME-GBD_2021_DATA-db69c1e8-1.csv 2010 + Países todos + Age (all + standardized) + Sex (Both) https://vizhub.healthdata.org/gbd-results?params=gbd-api-2021-permalink/be61825c6c4ec111b07e9d9847613cd0 IHME-GBD_2021_DATA-dcf93c30-1.csv 2020 + Países todos + Age (all + standardized) + Sex (Both) https://vizhub.healthdata.org/gbd-results?params=gbd-api-2021-permalink/972411dc3a931543fdecc15424db9019 IHME-GBD_2021_DATA-03b79351-1.csv 2021 + Países todos + Age (all + standardized) + Sex (Both) https://vizhub.healthdata.org/gbd-results?params=gbd-api-2021-permalink/337e92e538d09c5d3d80640ee324032e IHME-GBD_2021_DATA-ce582a59-1.csv Todos os anos + Países GBD exceto costum + Age standardized + Sex (Both) https://vizhub.healthdata.org/gbd-results?params=gbd-api-2021-permalink/0afce2e9094f891f22453111c6986ffb IHME-GBD_2021_DATA-382c4db5-1.csv
- Em termos de temas (features) temos: *Fatores de risco (poluição e subdivisões e tabagismo)* Causa (doenças respiratórias crónicas e suas subdivisões) *Sexo* Idade *Locais (regiões who, continentes (?), SDI e sistemas de saúde)* Anos (1990, 2000, 2010,2020 e 2021)

3.2 Data preparation

- Seleção de colunas relevantes (e.g., país, ano, sexo, idade, medida, causa)
- Tratamento de dados faltantes e outliers
- Verificação de formatos e consistência

```
## Rows: 438,015
## Columns: 11
## $ measure  <chr> "Deaths", "Deaths", "Deaths", "Deaths", "Deaths", "Deaths", "~
## $ location <chr> "Global", "Global", "Global", "Global", "Global", "Global", "~
## $ sex      <chr> "Male", "Female", "Both", "Male", "Female", "Both", "Male", "~
## $ age      <chr> "All ages", "All ages", "All ages", "All ages", "All ages", "~
## $ cause    <chr> "All causes", "All causes", "All causes", "All causes", "All ~
## $ rei      <chr> "Air pollution", "Air pollution", "Air pollution", "Air pollu~
## $ metric   <chr> "Number", "Number", "Number", "Rate", "Rate", "Rate", "Number~
## $ year     <int> 1990, 1990, 1990, 1990, 1990, 1990, 1990, 1990, 1990, 1990, 1~
## $ val      <dbl> 3.920050e+06, 3.420586e+06, 7.340636e+06, 1.459577e+02, 1.291~
## $ upper    <dbl> 4.542068e+06, 3.998895e+06, 8.540006e+06, 1.691177e+02, 1.510~
## $ lower    <dbl> 3.292876e+06, 2.809383e+06, 6.120949e+06, 1.226058e+02, 1.060~
```

```
##  [1] "measure"  "location" "sex"      "age"      "cause"    "rei"
##  [7] "metric"   "year"     "val"      "upper"    "lower"
```

measure: Deaths, DALYs (Disability-Adjusted Life Years), YLDs (Years Lived with Disability), YLLs (Years of Life Lost)
location: Global, High SDI, High-middle SDI, Middle SDI, Low-middle SDI, Low SDI, Advanced Health System, Basic Health System, Limited Health System, Minimal Health System, China, Democratic People's Republic of Korea, Taiwan, Cambodia, Indonesia, Malaysia, Lao People's Democratic Republic, Maldives, Myanmar, Philippines, Sri Lanka, Thailand, Timor-Leste, Viet Nam, Kiribati, Fiji, Micronesia (Federated States of), Marshall Islands, Papua New Guinea, Solomon Islands, Samoa, Tonga, Armenia, Vanuatu, Georgia, Azerbaijan, Kazakhstan, Tajikistan, Mongolia, Kyrgyzstan, Albania, Turkmenistan, Uzbekistan, Czechia, Bosnia and Herzegovina, Hungary, Bulgaria, Croatia, Poland, Romania, Montenegro, North Macedonia, Serbia, Slovakia, Belarus, Slovenia, Estonia, Russian Federation, Latvia, Ukraine, Republic of Moldova, Lithuania, Brunei Darussalam, Singapore, Andorra, Australia, New Zealand, Japan, Republic of Korea, Austria, Belgium, Cyprus, France, Germany, Iceland, Finland, Greece, Ireland, Denmark, Italy, Luxembourg, Malta, Norway, Netherlands, Israel, Spain, Portugal, United Kingdom, Argentina, Switzerland, Uruguay, Canada, Sweden, Chile, United States of America, Bahamas, Barbados, Antigua and Barbuda, Cuba, Belize, Dominica, Dominican Republic, Grenada, Guyana, Haiti, Jamaica, Saint Vincent and the Grenadines, Saint Lucia, Trinidad and Tobago, Suriname, Bolivia (Plurinational State of), Ecuador, Peru, Colombia, Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Venezuela (Bolivarian Republic of), Brazil, Paraguay, Algeria, Bahrain, Egypt, Iran (Islamic Republic of), Iraq, Jordan, Kuwait, Lebanon, Libya, Morocco, Palestine, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Tunisia, Türkiye, United Arab Emirates, Yemen, Afghanistan, Bangladesh, India, Bhutan, Nepal, Pakistan, Angola, Central African Republic, Congo, Democratic Republic of the Congo, Equatorial Guinea, Gabon, Burundi, Comoros, Djibouti, Eritrea, Ethiopia, Kenya, Madagascar, Malawi, Mauritius, Mozambique, Rwanda, Seychelles, Somalia, United Republic of Tanzania, Uganda, Zambia, Botswana, Lesotho, Namibia, South Africa, Eswatini, Zimbabwe, Benin, Burkina Faso, Cabo Verde, Cameroon, Chad, Côte d'Ivoire, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Sao Tome and Principe, Senegal, Sierra Leone, Togo, American Samoa, Bermuda, Cook Islands, Greenland, Guam, Monaco, Nauru, Niue, Northern Mariana Islands, Palau, Saint Kitts and Nevis, Puerto Rico, Tokelau, San Marino, United States Virgin Islands, Tuvalu, South Sudan, Sudan, Taiwan (Province of China), Turkey, East Asia, Southeast Asia, Oceania, Central Asia, Central Europe, Eastern Europe, High-income Asia Pacific, Australasia, Western Europe, Southern Latin America, High-income North America, Caribbean, Andean Latin America, Central Latin

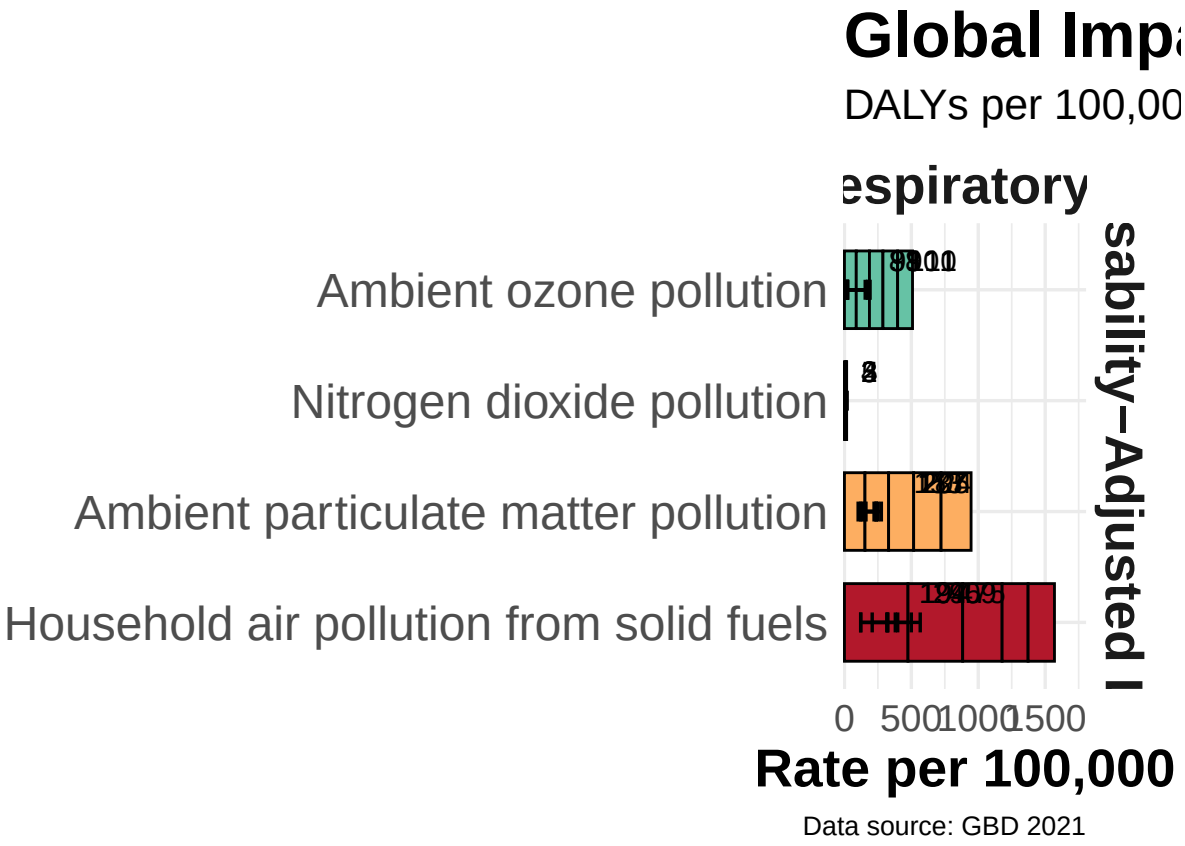
America, Tropical Latin America, North Africa and Middle East, South Asia, Central Sub-Saharan Africa, Eastern Sub-Saharan Africa, Southern Sub-Saharan Africa, Western Sub-Saharan Africa
sex: Male, Female, Both
age: All ages, 15-49 years, Age-standardized, 85-89 years, 90-94 years, 0-14 years, 50-74 years, 95+ years, 75-84 years
cause: All causes, Chronic respiratory diseases, Chronic obstructive pulmonary disease, Asthma
rei: Air pollution, Household air pollution from solid fuels, Ambient ozone pollution, Ambient particulate matter pollution, Particulate matter pollution, All risk factors, Nitrogen dioxide pollution
metric: Number, Rate
year: 1990, 2000, 2010, 2020, 2021

3.3 Tools

- Linguagem: R
- Ambiente: RStudio
- Pacotes: `tidyverse`

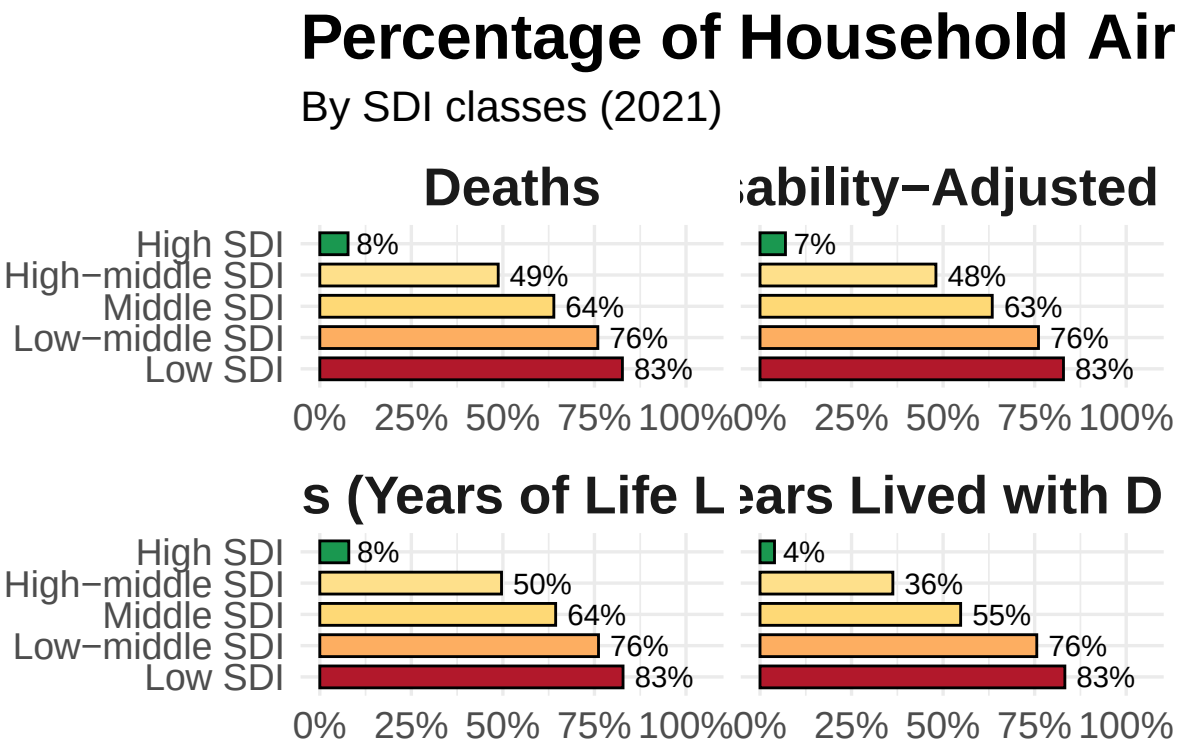
4 Results

4.1 Impact of different types of polluents



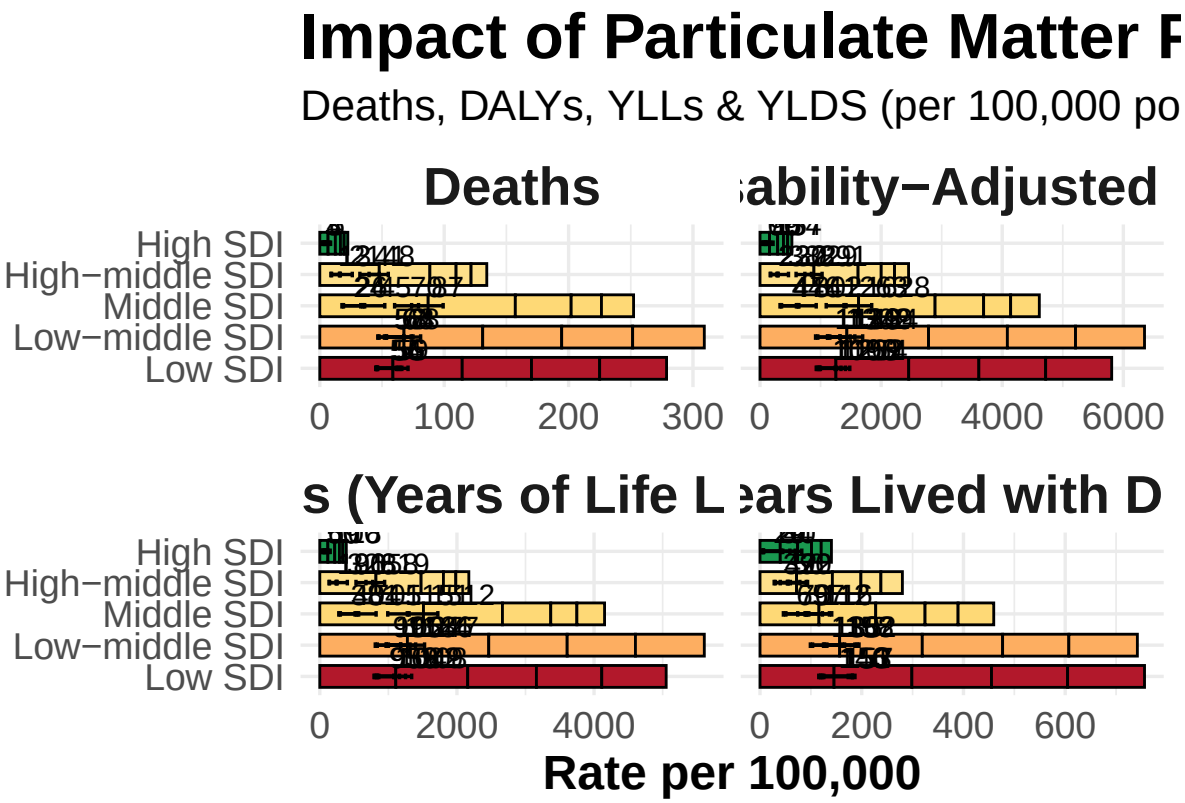
4.2 Impact on different Socio-demographic Index

4.2.1 Percentage of Household Air Pollution contribution on particulate matter pollution

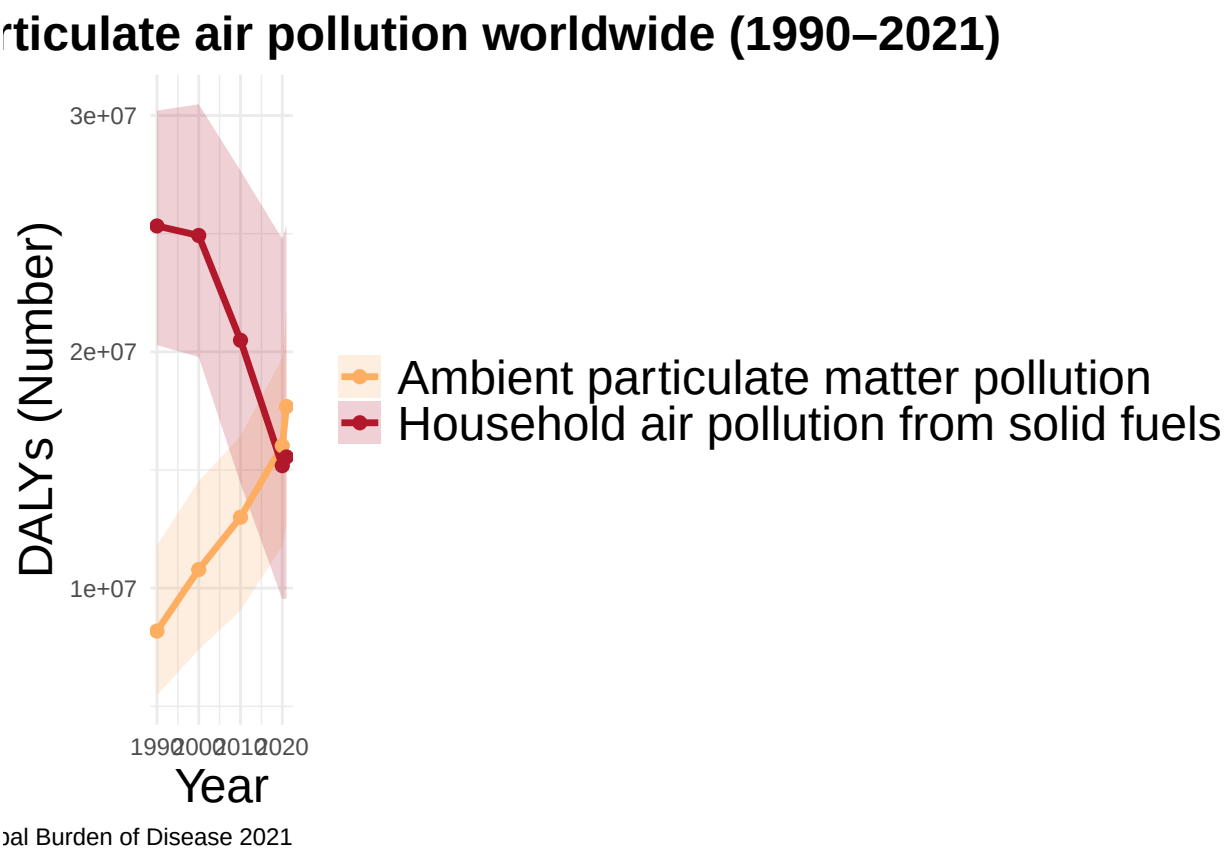


Source: GBD 2021

4.2.2 Impact of Particulate Matter Pollution by SDI Classes (2021)

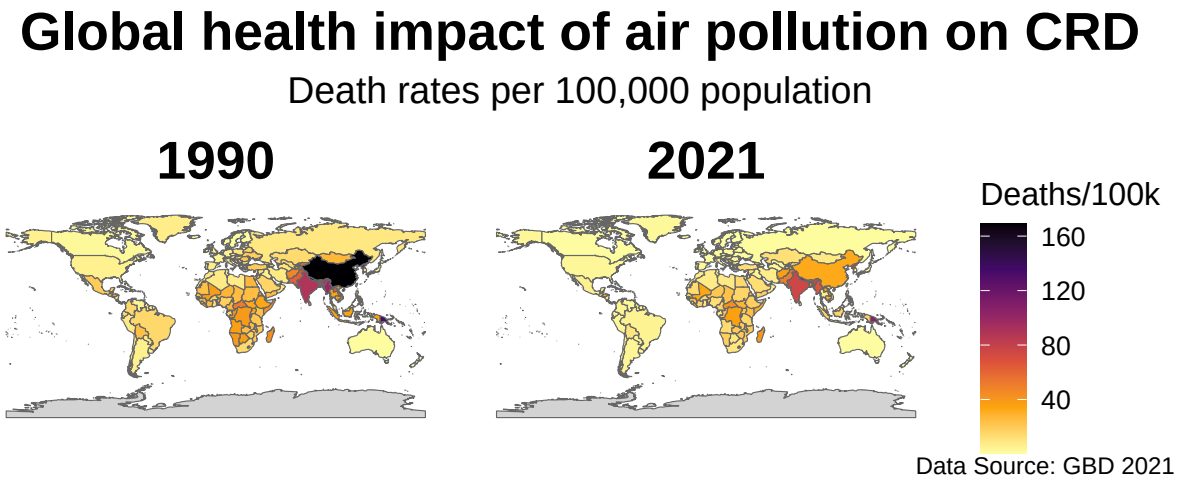


4.3 DALYs attributable to household vs ambient particulate air pollution worldwide (1990–2021)

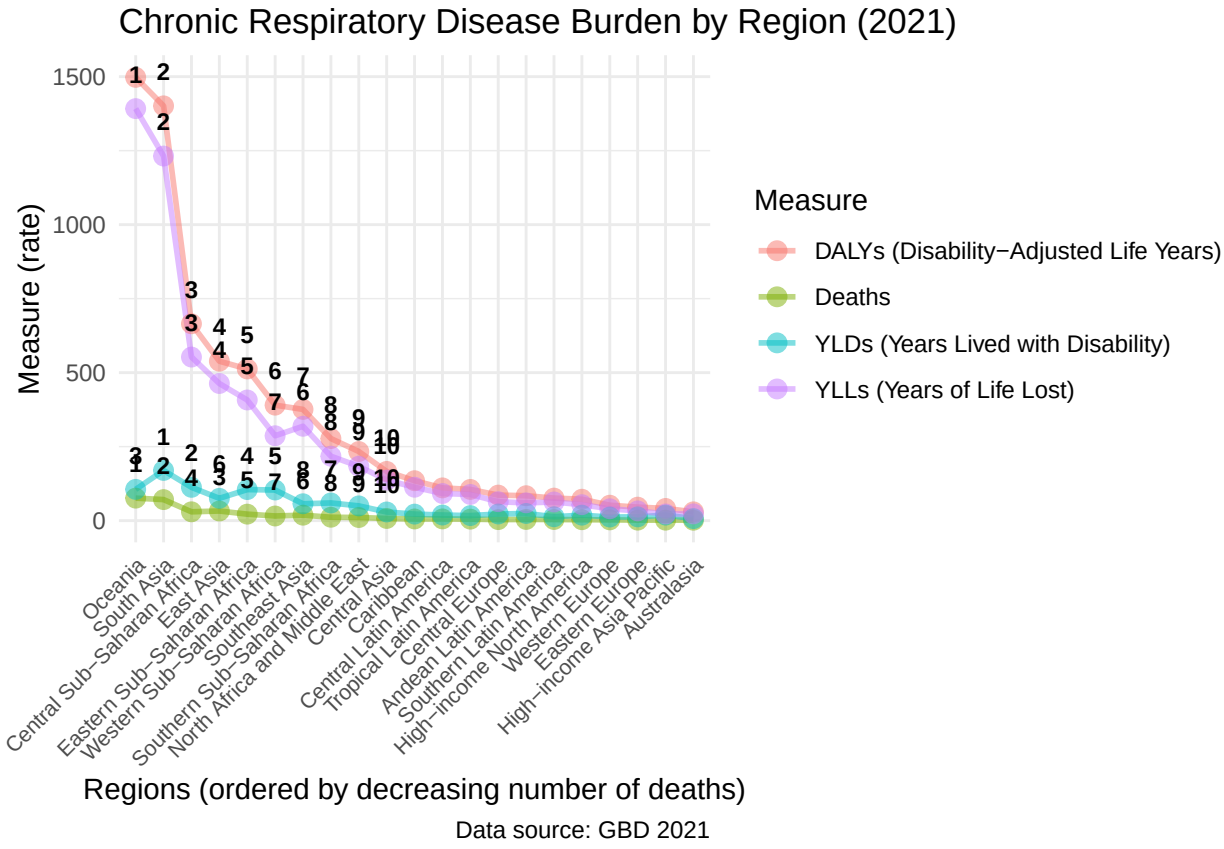


4.4 Geographical Distribution

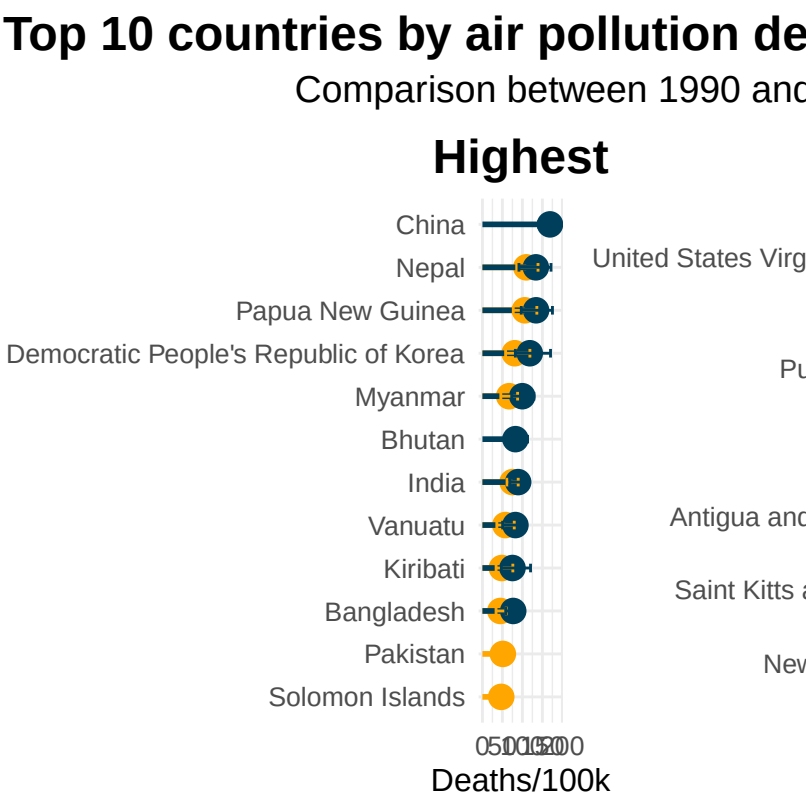
4.4.1 Global health impact of air pollution on CRD in 1990 and 2021



4.4.2 Country ranking by all measures in 2021

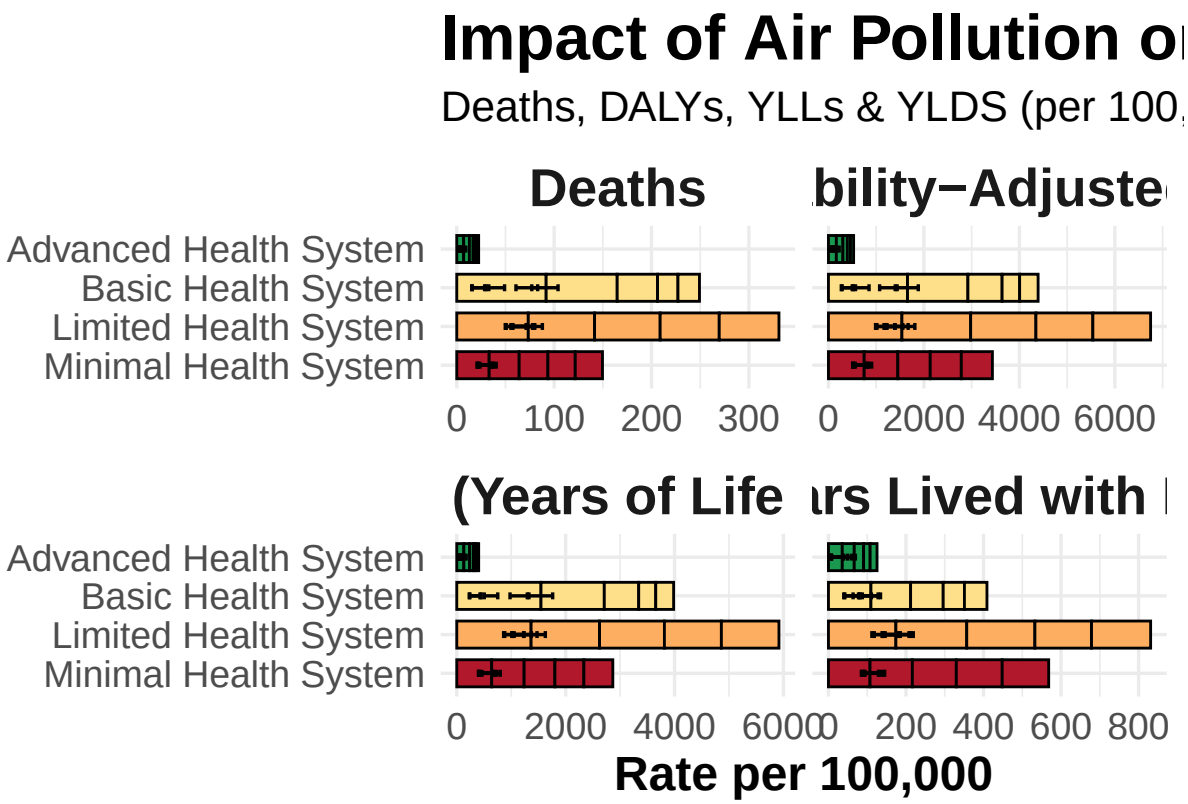


4.4.3 Top 10 countries by air pollution death rates on CRD



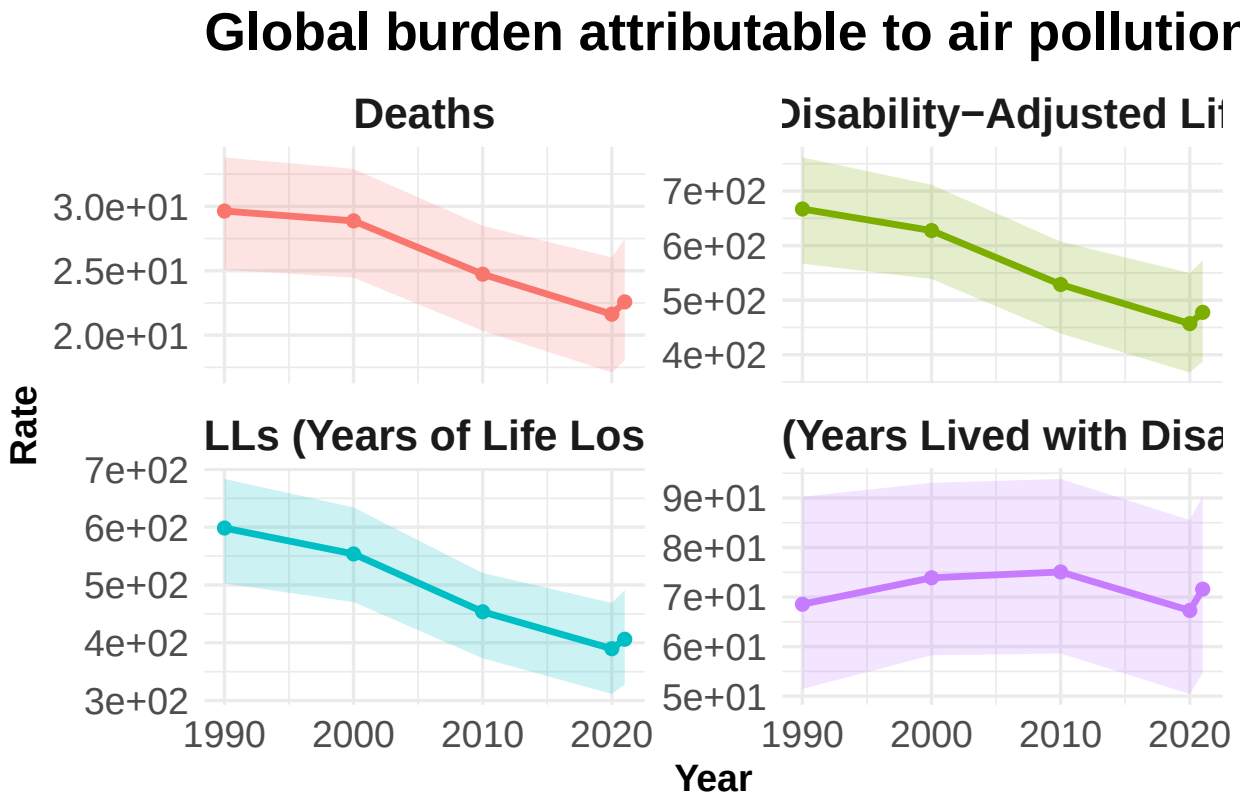
4.4.3.1 Comparison between 1990 and 2021

4.5 Impact of Air Pollution on CRD by Health System Type (2021)



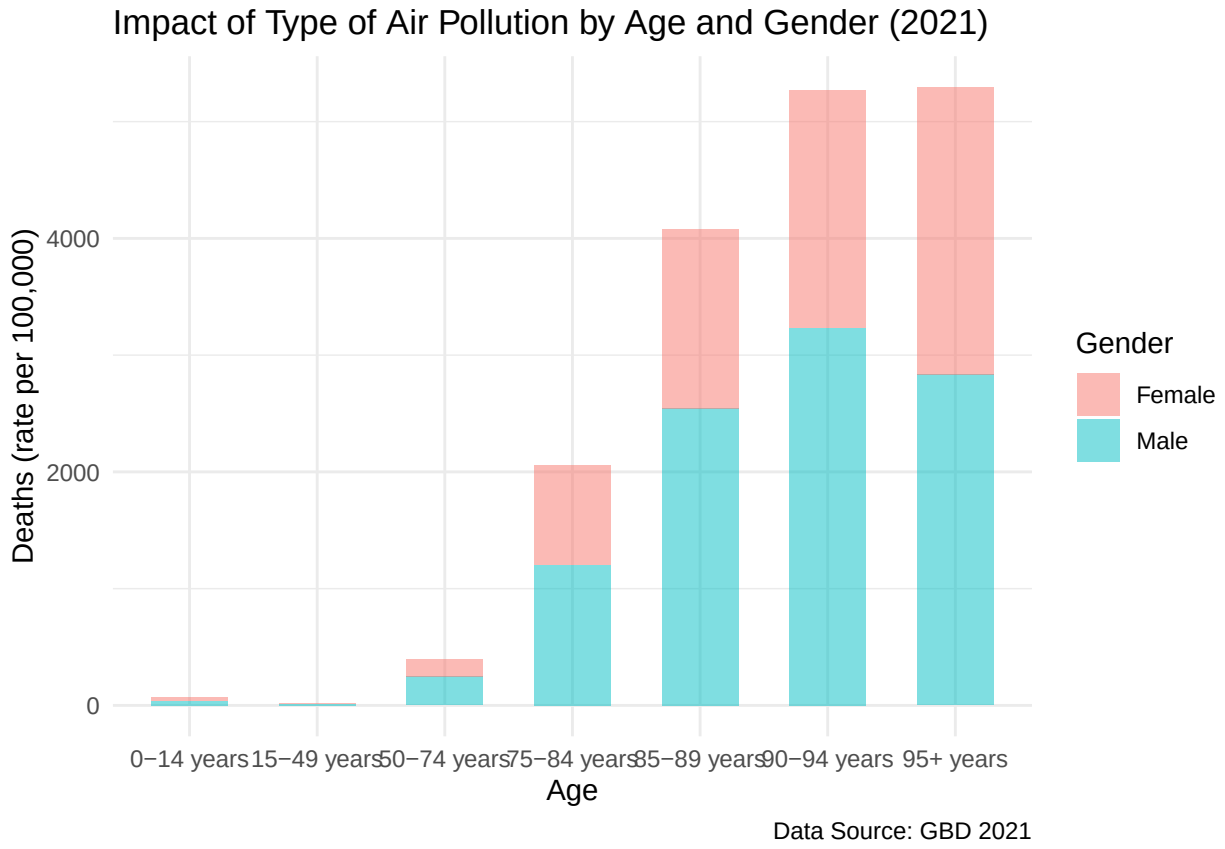
Data Source: GBD 2021

4.6 Temporal trends



Data Source: GBD 2021

4.7 Age and Gender



5 Discussão

Interpretação dos resultados Comparação com fontes externas (e.g., WHO, dados governamentais) *Limitações dos dados e da análise

6 Conclusão

Principais conclusões sobre o impacto da poluição do ar Sugestões para futuras investigações

7 Referências

Anexos

Bennitt, F. B., Wozniak, S., Causey, K., Spearman, S., Okereke, C., Garcia, V., Hashmeh, N., Ashbaugh, C., Abdelkader, A., Abdoun, M., Abdurebi, M. J., Abedi, A., Zuñiga, R. A. A., Aboagye, R. G., Abubakar, B., Abu-Zaid, A., Adane, M. M., Adegboye, O. A., Adekanmbi, V., … Burkart, K. (n.d.). *Global, regional, and national burden of household air pollution,*

1990–2021: A systematic analysis for the global burden of disease study 2021. Retrieved October 1, 2025, from https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)02840-X/fulltext?uuid/x3duuid%3A40008a0f-2266-4637-8209-d6041e2790e4