CSC343 Presentation
Ekagra & Connor

Air pollution and its consequences

https://www.psychologicalscience.org/news/releases/polluted-air-may-pollute-our-morality.html

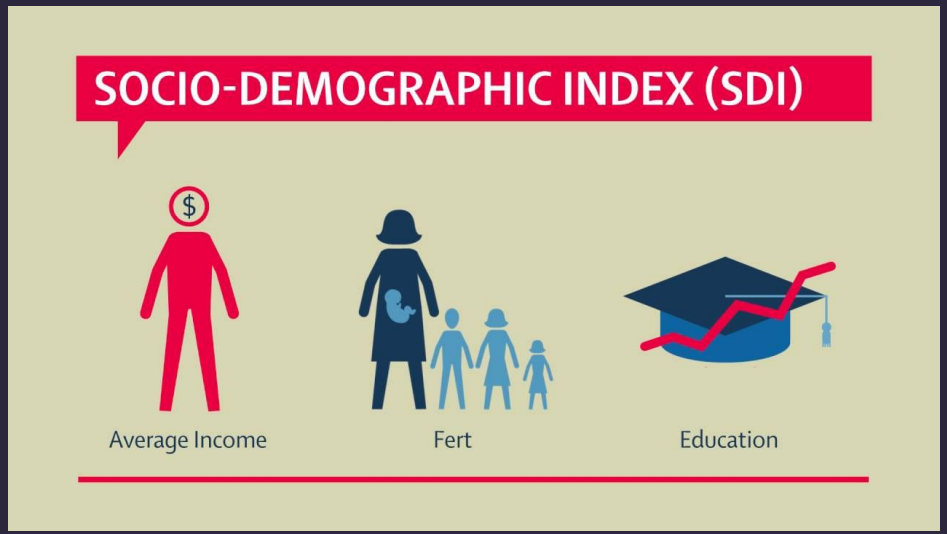
Domain

Our domain is CO2 emissions (or air pollution) and deaths caused by them, with respect to countries/entities. Entities could be groups that aren't countries, but consist of them (i.e a set of countries with a low socio-demographic index)

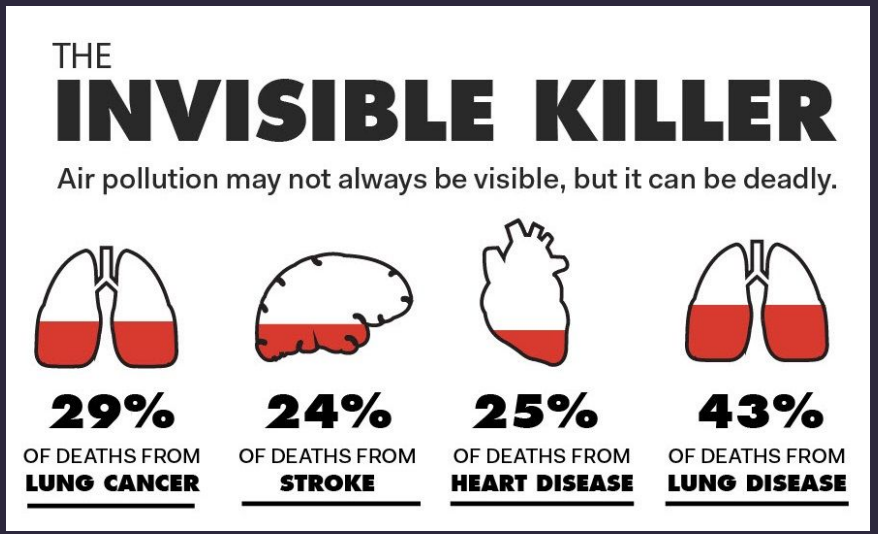
https://www.nrdc.org/stories/air-pollution-everything-you-need-know

Investigative Questions

1. Does the amount of emissions affect the amount of people dying from air pollution?
2. Does socio-demographic index have an effect on the number of deaths from air pollution?
3. Is there a general correlation between population and the amount of emissions?



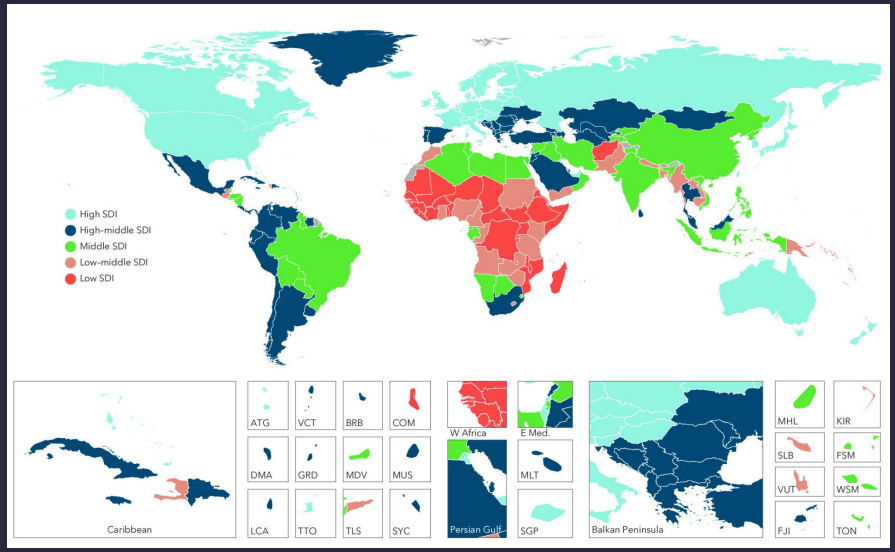
https://www.humanosphere.org/global-health/2016/10/a-new-way-to-measure-progress-global-health/



https://world-heart-federation.org/news/air-pollution-and-cardiovascular-disease-a-window-of-opportunity/

Results

- The average number of deaths due to air pollution of lower SDI categories are higher than SDI categories further up on the scale (The average computed factors in every country from that SDI category):
    - We observed the decreases in the average number of deaths while moving up each category of the SDI indicators (Low, Low-middle, Middle, Middle-high, High). We saw that for the most part, the average number of deaths would decrease by ~20% when comparing consecutive categories ie. (Low to Low-middle). There was one outlier however, and that was when we saw the average number of deaths decrease by ~66%, when comparing Middle-high to High.



http://www.healthdata.org/acting-data/new-way-measuring-development-helps-assess-health-system-performance

```
 low_to_lowmiddle_sdi_percent_change | lowmiddle_to_middle_sdi_percent_change | middle_to_middlehigh_sdi_percent_change | middlehigh_to_high_sdi_percent_change
-------------------------------------+----------------------------------------+-----------------------------------------+---------------------------------------
          -21.07101280558789289900 |           -22.56637168141592920400 |            -19.61904761904761904800 |           -66.11374407582938388600
(1 row)
```

- There is a moderate correlation between an increase in population and an increase in emissions:
  - To be exact, for any two particular consecutive years, when there is an increase in population in a country, there is generally an increase in the amount of emissions in that country (more than 75% of the time).

https://chiefexecutive.net/companies-often-grow-faster-people/

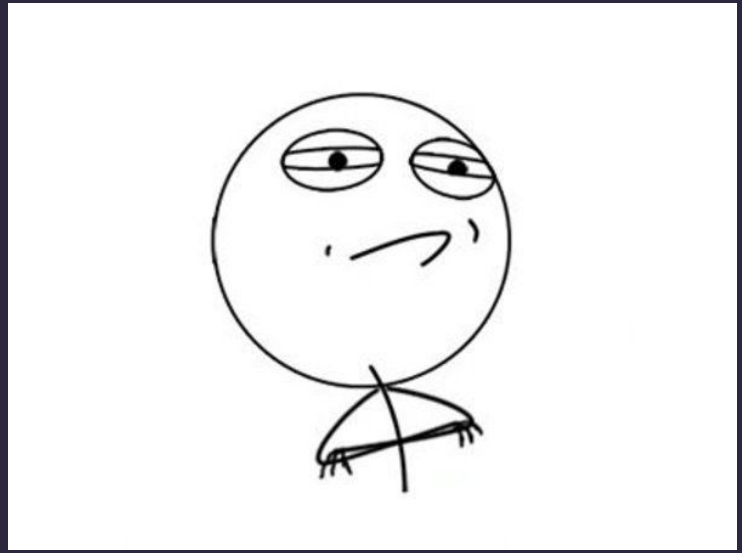
https://elemental.medium.com/a-surprise-surge-in-air-pollution-may-be-causing-more-coronavirus-complications-c444e82a659e

Challenges

Our dataset containing deaths due to air pollution only considered deaths per 100,000, so in order to get the actual number of deaths due to air pollution for a given entity, we had to use a population dataset so that we could compute the total number of deaths in a given country in a particular year.

It was a challenge for us to ensure that we had a large enough year range for each of our countries from our datasets, so that we could come to reasonable conclusions about our investigative questions. This stemmed from needing all our datasets to contain similar countries data about a similar number of (or larger) years (i.e. if we believed that checking a country's emissions from 1990 to 2017 was a large enough range to determine rate of change among emissions, we needed to ensure that all our datasets had that country and that range).

Another challenge we had was finding a way to remove things from our data that didn't fall under the category of country or SDI rank in our deaths due to air pollution dataset and our emissions dataset. This included the rows that displayed data about high, middle, and low income countries.



Lessons

We also learned of ways to compare data across consecutive/various years.

We learned that when working with multiple sources of data, it's important to consider that sometimes your data will be bottlenecked by whichever dataset contains the least information.

In Phase 2, we learned how to cleanup data in a few different ways. For example, some data we manually removed from our datasets, while other data was removed through the use of SQL queries in order to get the rows which we needed for our tables.

https://www.channelfutures.com/best-practices/14-lessons-learned-in-14-years-as-a-business-owner

Questions?