



Dataset Nutrition Label

U.S. State Air Quality and Respiratory Deaths Dataset

About

This dataset contains data about the amount of respiratory deaths and air quality (pm10, pm2.5, and no2 concentration) per U.S. State per year (from 2014-2021).

Data Creation Range: 2014-2021

Created By: Andrew Chen, Aarfan Hussain,
Jack Scott

Content: csv file

Instances: 400

Sources:

CDC Chronic Lower Respiratory Disease
Mortality by State¹

https://www.cdc.gov/nchs/pressroom/sosmap/lung_disease_mortality/lung_disease.htm

WHO Ambient Air Quality Database

<https://www.who.int/data/gho/data/themes/air-pollution/who-air-quality-database>

Alert Count	4
Completeness	2
Amount of years	1
Scope	1
Provenance	0
Collection	1
Missing data	1
Description	0
Composition	1
COVID	1

Use Cases

Potential real world applications

- 1.) How are respiratory deaths and air quality related?
- 2.) How has air quality changed over time?
- 3.) How have respiratory deaths changed over time?
- 4.) Are there regions of the U.S. that have worse air quality?

Alerts:

Completeness:

Amount of years

Only covers years 2014 to 2021. By covering more years we could give a better analysis.

Scope

We only cover U.S. states, which generalizes urban and rural parts of each state and doesn't give as detailed of an analysis.

Collection:

Missing data

From the WHO dataset, there were some pm10, pm2.5, and no2 concentration values that were missing for some states during some years.

Composition:

COVID

Since during 2019-2022 COVID was a major source of respiratory deaths, this may skew data (unsure if CDC dataset takes COVID deaths into consideration).

¹ We ended up changing our previous dataset which was: Weekly Counts of Deaths by State from 2014 to 2019, from the CDC. We discovered that this dataset didn't have as many values as we hoped for, and we switched to a different dataset that covered the same topic, but had more values.