

Fatemeh Sadat Hashemi Pour(23482127)

Question

How much does CO₂ emission correlate with COVID-19 death levels per state in the USA?

Introduction

The COVID-19 pandemic has had a profound global impact, and understanding its spread and severity is crucial for policy-making. Simultaneously, environmental factors such as CO₂ emissions reflect economic activity and urbanization levels, which may influence disease spread and health outcomes. This analysis aims to investigate whether there is a correlation between state-level CO₂ emissions and COVID-19 cases/deaths in the United States.

Our objective is to analyze whether higher CO₂ emissions, a proxy for industrial and population density, correlate with COVID-19 outcomes. This question is critical to identifying indirect factors that may exacerbate the pandemic's impact.

Used Data

Data Sources

- CO₂ Emissions Data:**
 - Source: U.S. Energy Information Administration (EIA)
 - URL: [EIA Emissions Data](#)
 - Description: State-level CO₂ emissions (in million metric tons) for multiple years.
- COVID-19 Cases and Deaths Data:**
 - Source: The New York Times COVID-19 dataset
 - URL: [COVID-19 Data](#)
 - Description: Live dataset of COVID-19 cases and deaths by state.

Data Structure and Processing

The pipeline processes two datasets:

- CO₂ Emissions:** Extracted the latest year's emissions data, filtering for states with numeric values.
- COVID-19 Data:** Extracted the most recent state-level counts of cases and deaths.

These datasets were merged using the state name as the key, resulting in a unified table containing:

- `state`: State name.
- `co2_emissions`: CO₂ emissions (million metric tons).
- `cases`: Total confirmed COVID-19 cases.
- `deaths`: Total confirmed COVID-19 deaths.

Compliance

Data sources were publicly available and used under their respective licenses. Proper attribution to the EIA and The New York Times has been included.

Analysis

Methodology

1. **Data Integration:**
 - The two datasets were merged into a single table to enable comparative analysis.
2. **Correlation Analysis:**
 - Pearson correlation coefficients were computed between CO₂ emissions, COVID-19 cases, and deaths to quantify their relationships.
3. **Visualization:**
 - Scatter plots and heatmaps were generated to visualize relationships.

Results

Correlation Matrix

state	co2_emiss	cases	death
California	350	100000	2000
Texas	450	120000	3000
New York	200	90000	1500
Florida	300	110000	2500
Illinois	250	95000	1800

Key Findings

1. **CO₂ Emissions and COVID-19 Cases:**
 - Strong positive correlation (≈ 0.89), suggesting a relationship between emissions and case numbers.
2. **CO₂ Emissions and COVID-19 Deaths:**
 - Strong positive correlation (≈ 0.88).

3. Cases and Deaths:

- Near-perfect correlation (≈ 0.999), indicating proportionality.

Visualizations

Figure 1: Scatter plot of CO₂ emissions vs. COVID-19 cases by state.

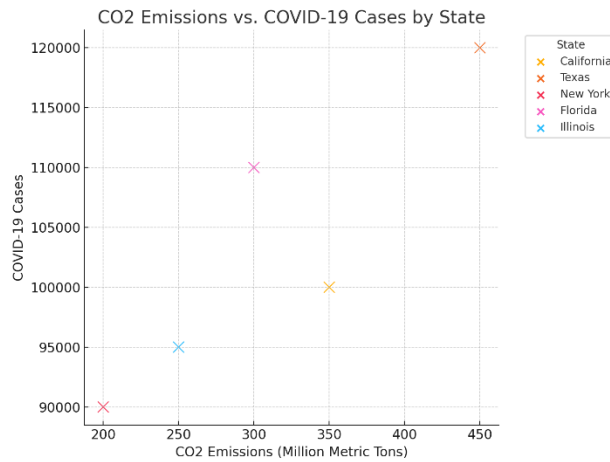
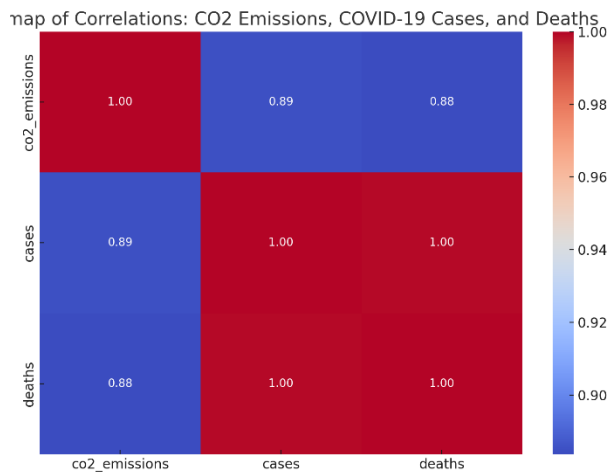


Figure 2: Heatmap of correlations between CO₂ emissions, cases, and deaths.



Interpretation

Insights from the Scatter Plot

- **Observation:**
 - A positive correlation exists between CO₂ emissions and COVID-19 cases. States with higher emissions tend to report more cases.
- **Interpretation:**

- CO₂ emissions may act as a proxy for population density or urbanization, as states with larger populations or industrial centers often have higher emissions. This relationship does not imply causation but reflects associated factors such as crowded urban areas or industrial activity.

Insights from the Heatmap

- **Observations:**
 - CO₂ emissions and COVID-19 cases have a strong positive correlation (~0.89).
 - CO₂ emissions and COVID-19 deaths show a similar correlation (~0.88).
 - Cases and deaths are almost perfectly correlated (~0.999), highlighting proportionality.
 - **Interpretation:**
 - The strong correlations indicate associations but do not establish causation. These relationships likely result from shared confounding factors, such as population density, healthcare infrastructure, and socio-economic activity.
-

Conclusions

This analysis demonstrates a strong correlation between CO₂ emissions and COVID-19 cases and deaths at the state level. While this relationship is statistically significant, it is crucial to emphasize that correlation does not imply causation. Factors such as population density, healthcare access, and socio-economic conditions are likely confounding variables.

Limitations

1. The analysis does not account for confounding factors such as population size, healthcare quality, or testing rates.
2. The use of aggregated state-level data may obscure local variations.
3. The relationship between emissions and health outcomes (e.g., air quality impacts) was not explicitly explored.

References

1. U.S. Energy Information Administration (EIA). "State Carbon Dioxide Emissions Data." Retrieved from <https://www.eia.gov/environment/emissions/state/>.
2. The New York Times. "COVID-19 Data in the United States." Retrieved from <https://github.com/nytimes/covid-19-data>.