Impact of Emissions on Respiratory Disease

Exploring Europe's Emissions of Airborne Pollutants by Country
and the Correlation of Mortality Rate by Respiratory Disease

Team members:

1. Jillian Walter
2. Marina Connolly
3. Jennifer Giraldo Davila
4. Alexandra Martinez
5. Israa Adam

Research Overview:

Context:
Urbanization and economic growth in capital cities drive higher emissions from industries and transportation, impacting air quality and public health. Wealthier cities may produce more emissions but often invest in green technologies and stricter regulations. This complex relationship between economic prosperity, emissions, and respiratory health highlights the need to explore whether higher income levels mitigate or worsen these health impacts.

Hypothesis:
Higher income levels in European capital cities are correlated with increased emission rates due to industrial and transportation activities. However, wealthier cities may offset these emissions by adopting green technologies and implementing stricter environmental regulations, potentially reducing their overall environmental impact. As a result, while emissions may be higher in wealthier cities, we hypothesize that total emissions will not necessarily correlate with increased mortality rates from respiratory illnesses, reflecting the mitigating effects of economic resources and environmental policies.

Research Questions & Theoretical Framework:

1. How do emission levels vary among countries with different income levels in Europe?
2. What are the top emissions driving countries and which industries drive the most emissions?
3. Does location (latitude/longitude) impact emission amounts?
4. Which countries are most impacted by Respiratory illness in Europe? Which illnesses drive the highest mortality rate?
5. How do emission levels correlate with mortality rates due to air pollution related conditions?

About our data sources:

1. **Topline country data from World Bank Open Data**: API Source used to pull all European countries and their descriptive data around location (latitude and longitude), capital city, and income level (all falling into “high” in respect to all other countries in the world

2. **European Emissions Data from European Environment Agency:** downloadable CSV providing detailed data around Emissions by country in metric tons: total emissions data, top industry drivers, type of pollutants (i.e. CO2, N2O, etc.)

3. **Respiratory Health Data from World Health Organization**: provides country-level data on all respiratory illnesses and mortality rates

Data Cleaning Process:

Topline Country Data: Read URL through JSON format; Extract relevant variables (name, capital city, latitude, longitude, income level), filter for European countries, delete NaNs & format latitude/longitude to numeric

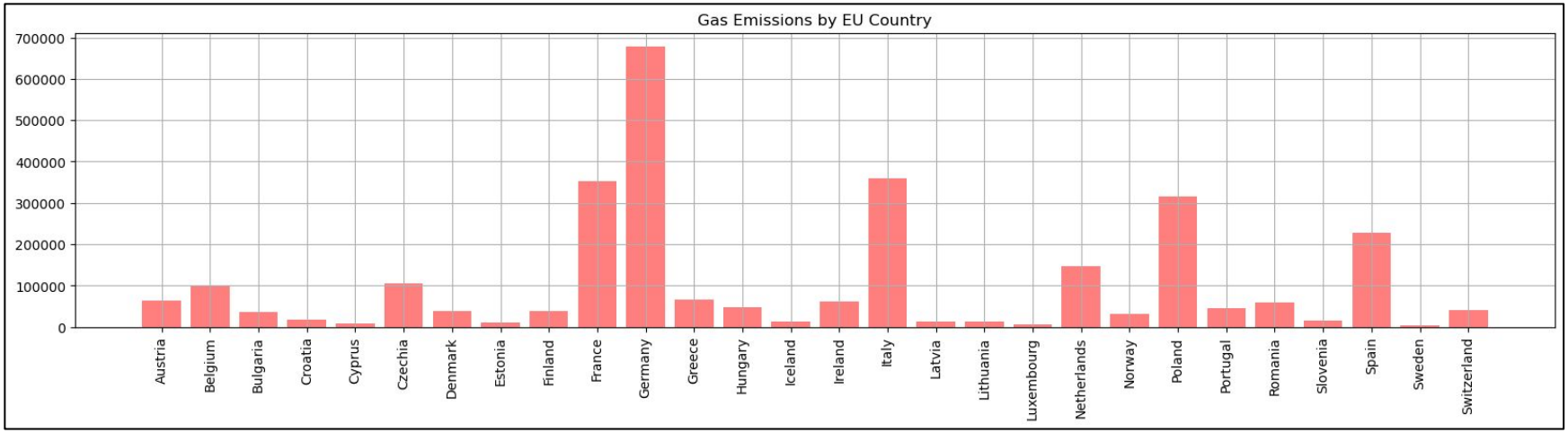
Emissions Data: Read CSV and extract relevant variables (country name, emissions, type of gas, type of industry); format variables into float (emissions) and strings (all other variables)

Respiratory Health Data: Read CSV and filtered for European data, pulling in Country Name, Mortality Rate, and Type of illness.

Merge all data into one dataframe, joining on Country Name

```
#WORLD BANK CLEANING
import json
import requests
import pandas as pd
import numpy as np
from pprint import pprint

url = "http://api.worldbank.org/v2/"
api_format = "json"

country_name = []
country_capital = []
latitude = []
longitude = []
income_level = []
region = []


page = 1
all_countries = []
for page in range(10):
    # Get country information in JSON format for each page
    response = requests.get(f"{url}countries?format={api_format}&page={page + 1}")
    # Check if there's data on this page; if not, break the loop
    if response[1] is None:
        break
    # Add countries from the current page to the list
    all_countries.extend(response[1])
    # page += 1  # Move to the next page
# Print all country information
#for country in all_countries:
    #print(country)

#Appending Data to Lists to be used in Dataframe

for country in all_countries:
    country_name.append(country['name'])
    country_capital.append(country['capitalCity'])
    latitude.append(country['latitude'])
    longitude.append(country['longitude'])
    income_level.append(country['incomeLevel']['value'])
    region.append(country['region']['value'])
```

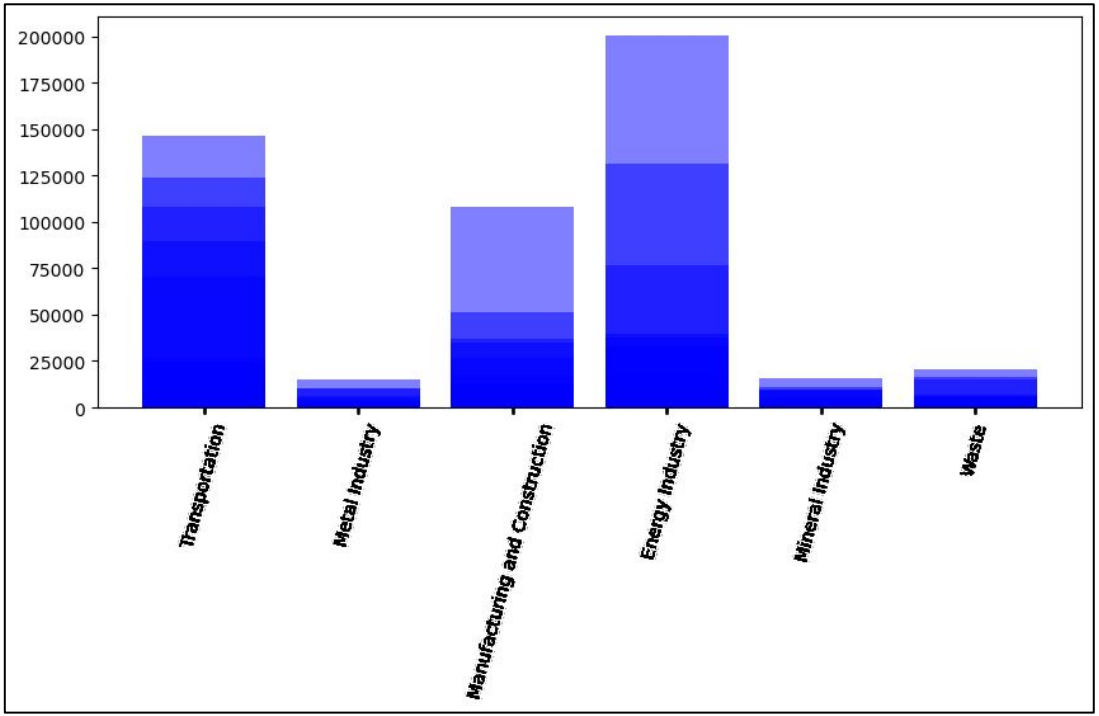
European Emissions Snapshot:



When looking at Annual Emissions by Country throughout the EU, Germany's emissions surpass nearly all other countries by at least double.

France, Italy, Poland also contribute the most emissions, while Malta, Luxembourg and Sweden contribute the least.

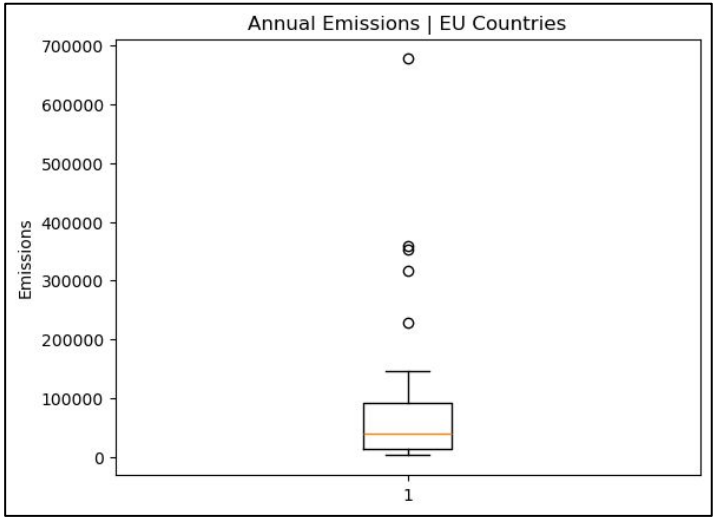
European Emissions Snapshot:



Upon investigating further, Energy, Transportation, Manufacturing/Construction and Waste appear to be driving the majority of emissions in Europe.

European Emissions Snapshot:

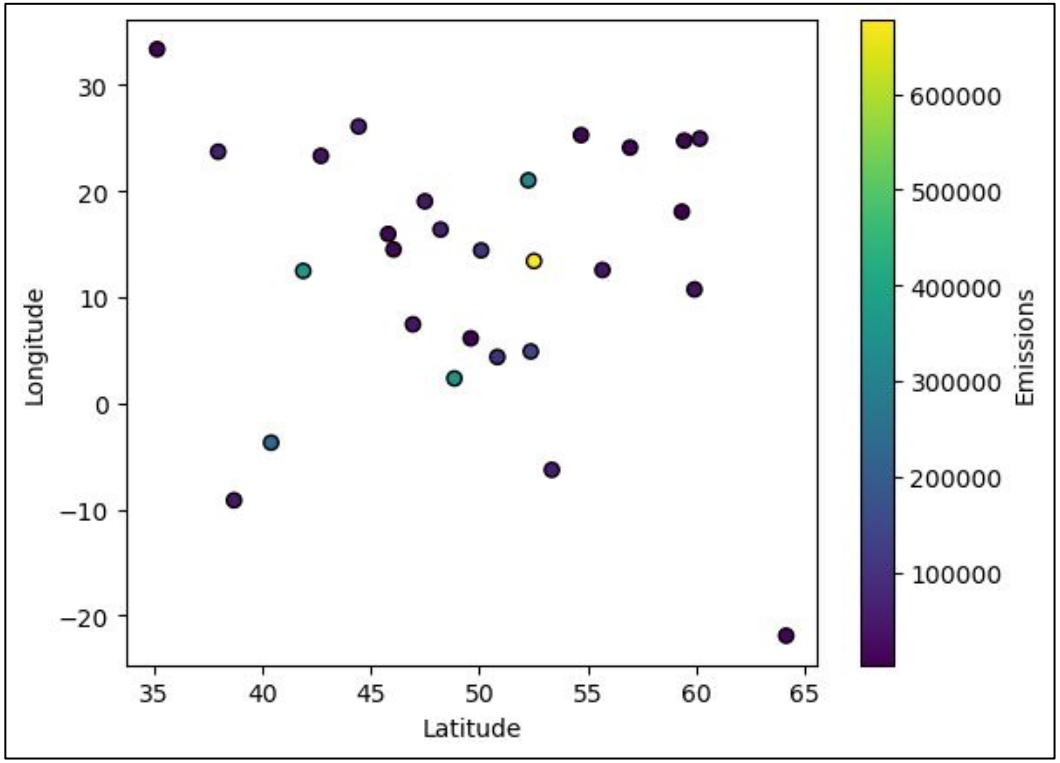
```
total_emissions_df["Emissions"].describe()

count        30.000000
mean      98180.436000
std      149335.753844
min        2131.300000
25%       13071.942500
50%       40137.315000
75%       90645.272500
max      677567.520000
Name: Emissions, dtype: float64
```



Emission amounts **vary greatly by European Country**; the data has a **mean** net annual emissions by country at 98.2k metric tons, but a range of
675k tons and standard deviation of 145k tons.

The top 5 max contributors are actually outliers (Germany, Italy, France, Poland and Spain), while the remaining 25 EU Countries fall between 2.1k and 125k emissions.

European Emissions Snapshot:



Emissions amounts appear to vary across latitude and longitude in Europe, indicating that there is no relationship between location in Europe and emissions amounts.

To prove this, we ran a correlation analysis:

- Relationship between Longitude and Emissions:

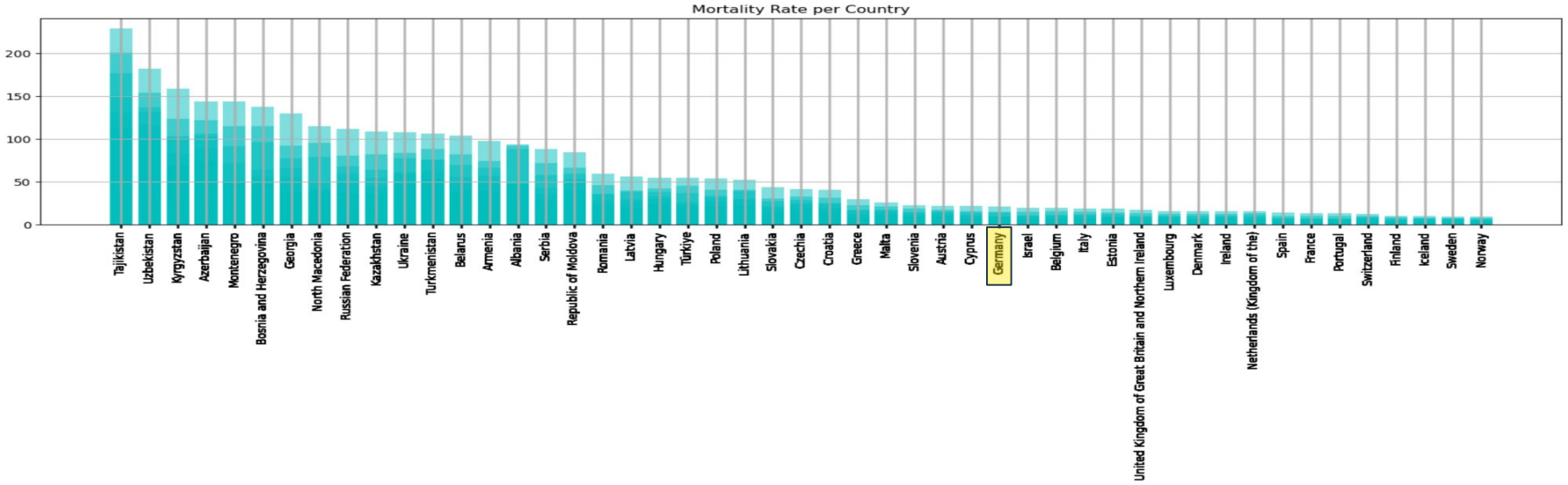
```
(statistic=-0.1110119121562874, pvalue=0.5738563932723585)
```

- Relationship between Latitude and Emissions:

```
(statistic=-0.10241977849720856, pvalue=0.6040307324703169)
```

Given the minimal correlation between location and emissions, we decided to instead investigate the relationship between emissions and residents' health by country

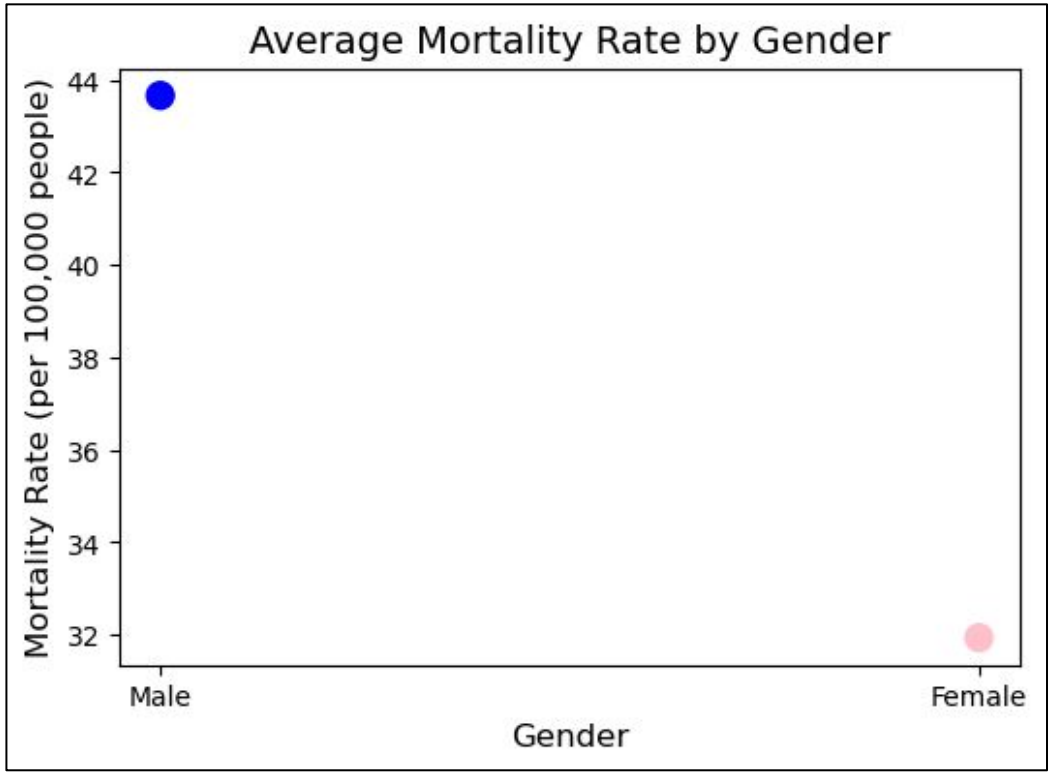
European Health Data Snapshot



Taijikistan had the highest mortality rate from Emissions of all European countries.

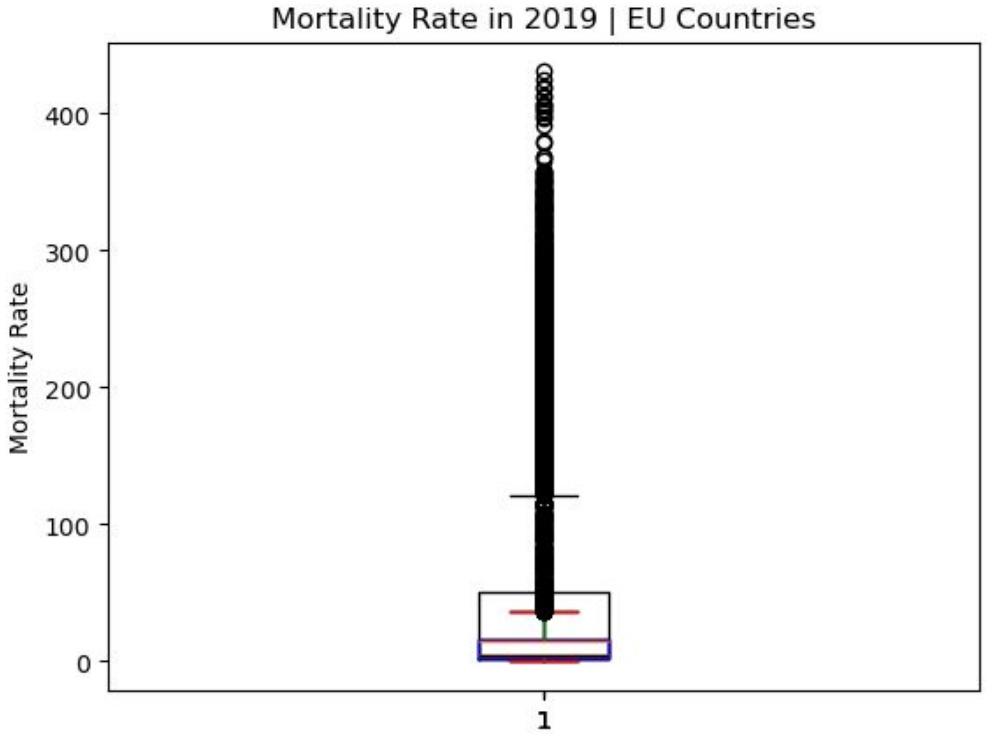
Even though our previous analysis showed Germany has the highest emission rate, the country has one of the lowest mortality rates due to emissions

European Health Data Snapshot



The average mortality rate shows us that men are mostly affected by respiratory illnesses and lead mortality rate in all European countries

European Health Data Snapshot



The average mortality rate across European Countries is 16%, but has a high range when factoring in EU countries vs. non-EU Countries.

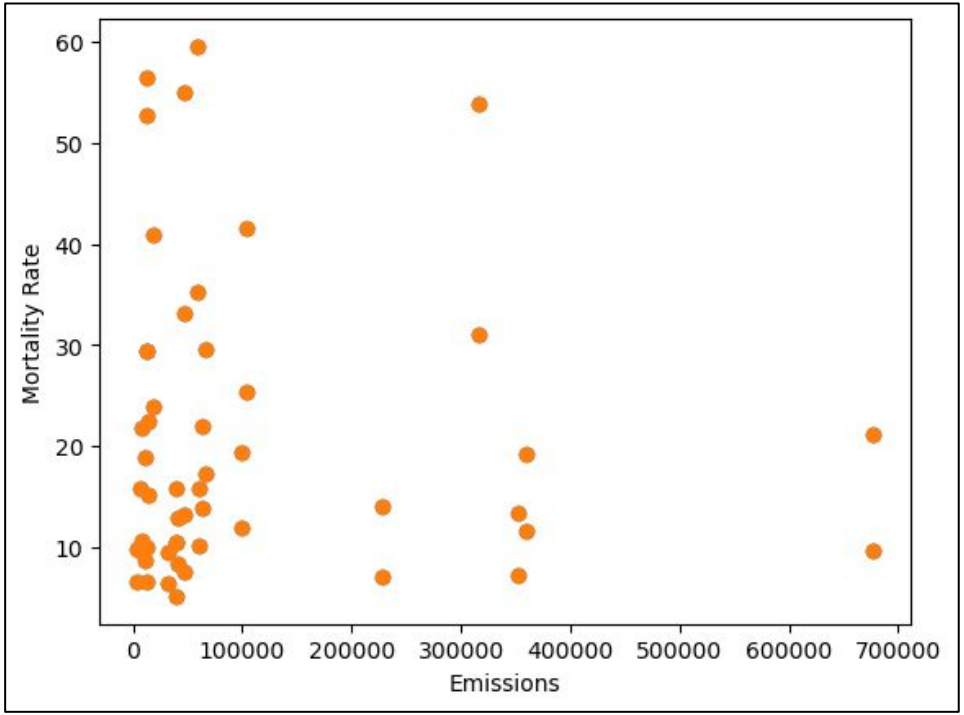
```
sort_mortality["Rate"].describe()

count    882.000000
mean      16.505578
std       29.493625
min        0.100000
25%        1.462500
50%        4.045000
75%       15.457500
max      229.500000
Name: Rate, dtype: float64
```

Statistical Analysis:

Ultimately, we found that there is a low correlation between Emission amounts and Mortality Rate, causing us to accept our hypothesis (high P-value).

A low statistic (correlation) shows that respiratory diseases do not increase when emissions increase.

Other external factors that could be contributing to respiratory illness mortality rates may include income, diet, drug/alcohol/smoking use and lifestyle, **not country-wide emissions**.



Pearson test between emissions and mortality rate:
statistic=-0.047799167255435135
pvalue=0.7364952211473059)

Conclusion:

Emissions and Mortality Rates by Country vary greatly throughout Europe.

Location in Europe (Latitude/Longitude) does not correlate with an increase in Emissions.

An increase in Emissions does not correlate with an increase in Respiratory mortality rate.

Additional external factors and analyses would need to be considered to explain respiratory mortality rate.

Questions?