# COVID-19 Death Rates and Air Quality Analysis

● ● ●

Brynn Hill, Clare Gibb, Anika Tripathi, Deetya Gupta, and David Nu
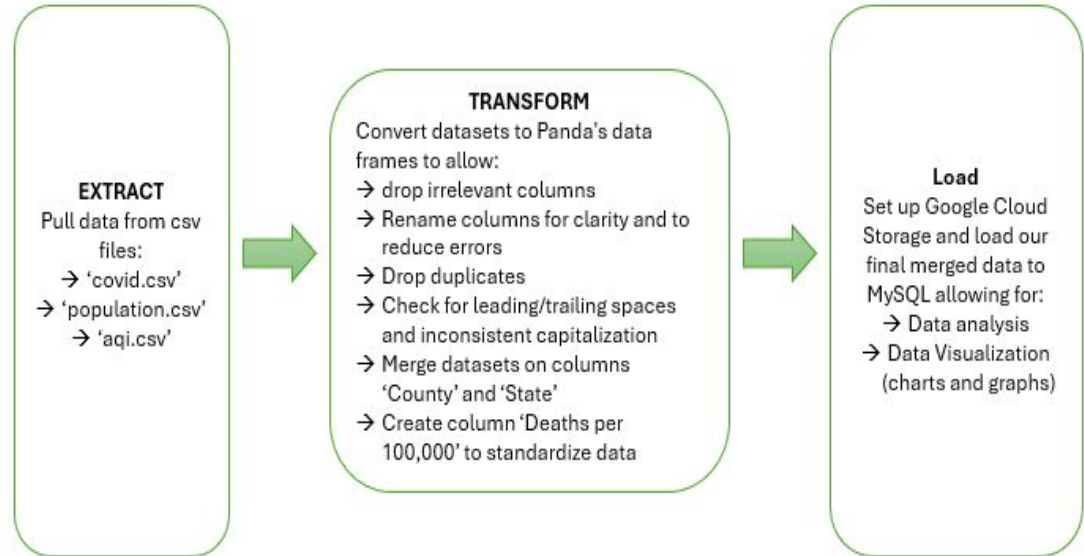
# Data Processing and Analyzing

- Question
  - Is there a correlation between COVID-19 Death Rates and the Median Air Quality Index across United State Counties during the year of 2020?
- Data Selection
  - Covid.csv (CDC)
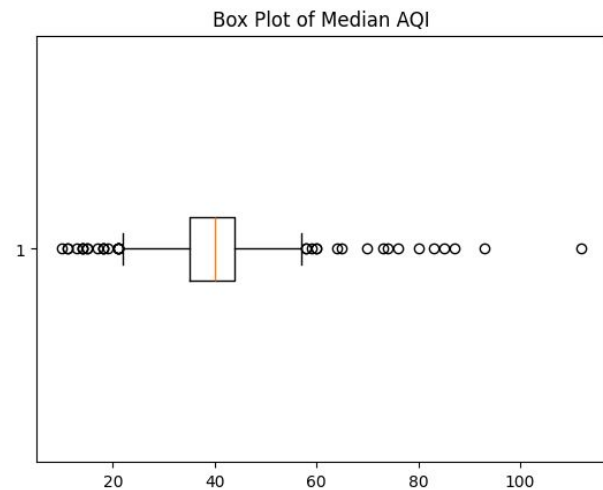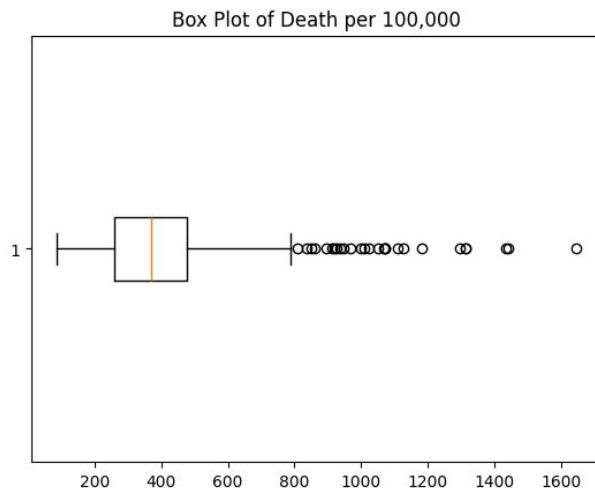  - population .csv (U.S. Census Bureau)
  - aqi.csv (EPA)
- Process
  - ETL Pipeline
  - Data Visualization
  - Basic Statistical Analysis

**EXTRACT**
Pull data from csv files:
→ 'covid.csv'
→ 'population.csv'
→ 'aqi.csv'

**TRANSFORM**
Convert datasets to Panda's data frames to allow:
→ drop irrelevant columns
→ Rename columns for clarity and to reduce errors
→ Drop duplicates
→ Check for leading/trailing spaces and inconsistent capitalization
→ Merge datasets on columns 'County' and 'State'
→ Create column 'Deaths per 100,000' to standardize data

**Load**
Set up Google Cloud Storage and load our final merged data to MySQL allowing for:
→ Data analysis
→ Data Visualization (charts and graphs)

# Statistical Support

- The Pearson Correlation Coefficient
  - Calculated to be -0.0004108
  - Weak/non-linear relationship
- Outliers
  - Many for both datasets

# Summary of Findings

- Figure 1
  - AQI trends do not line up with death rate consistently
- Figure 2
  - 30-50 median AQI range
  - 100-600 death rate range
  - Outliers!
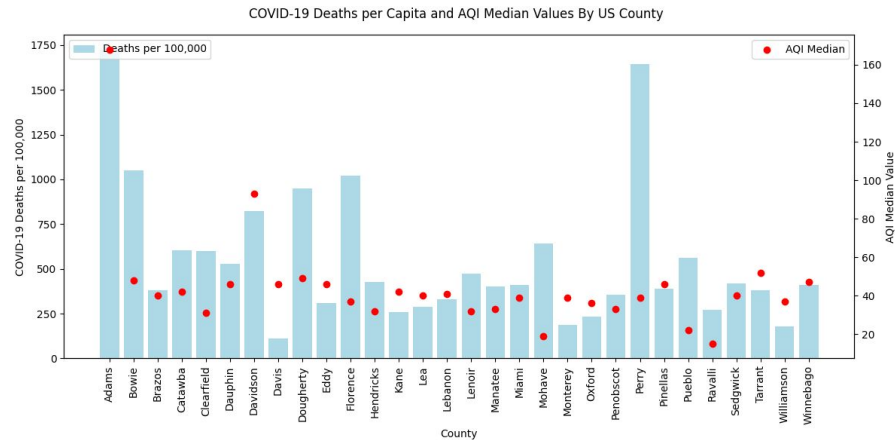- No substantial correlation
- Other factors?



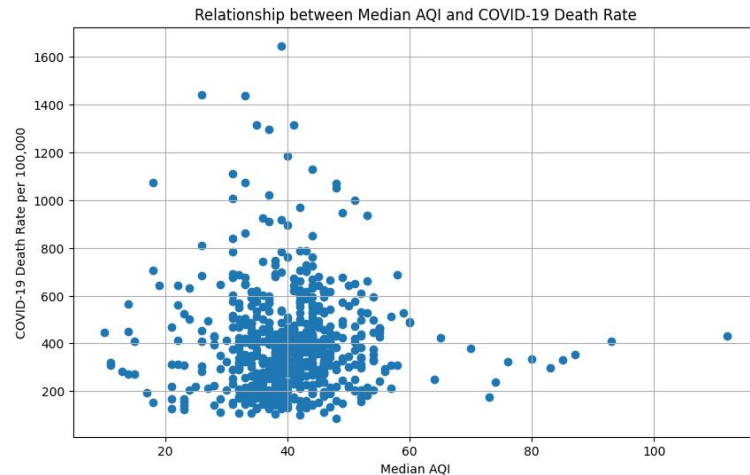Figure 1: COVID-19 Deaths per Capita and Median AQI Values Across US Counties During 2020



Figure 2: Relationship Between Median AQI and COVID-19 Deaths During 2020

# Challenges, Key Insights, and Takeaways

- Technical Difficulties
    - Data selection
    - Name and format inconsistencies
    - Initializing JSON keys for Google Cloud Storage
- Skills
    - Designing and implementing ETL pipelines
    - Data visualization
    - Cloud storage
    - Team coordination and communication
- For Future Projects
    - Allocate more time to data cleaning
    - Analyze additional factors apart from API (age, socioeconomic status, etc.)