Air Pollution Leads To DEATH?

December 9, 2019
Tram Nghi Pham, Brandon Pratt, Siting Wang,
and Marta Wolfshorndl

Presentation Outline

- Problem Statement
- Data Overview
- Use Cases
- Demo of Map
- Design Components
- Github Repo Structure

Problem Statement

- Air pollution is an increasing concern across the globe as it has been linked to the development of respiratory illnesses, including asthma, cancer, and others
- An interactive platform for the public to visualize and predict the relationship between geographic respiratory death rates in the United States and air pollution does not exist.

Our Aim

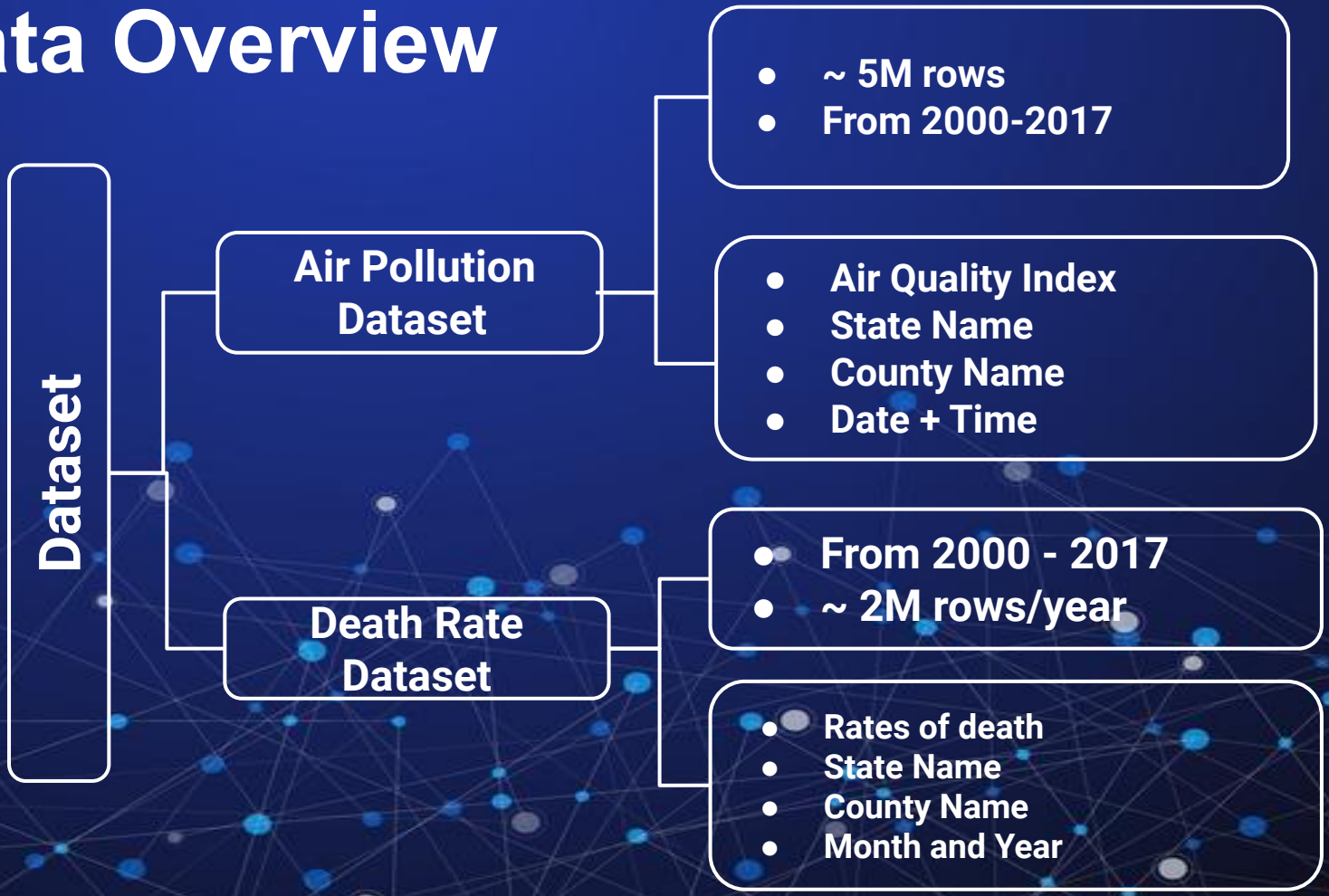
We aim to develop such a tool in order to illustrate:

- How the air quality index (AQI) correlates with geographic respiratory death rates in the United States
- If we could predict respiratory death rates given prior and current pollution levels.

Data Overview

- United State Environmental Protection Agency (Air Pollution)
  https://aqs.epa.gov/aqsweb/airdata/download_files.html#AQI
- Centers for Disease and Control for Prevention (Death Rate)
  https://www.nber.org/data/vital-statistics-mortality-data-multiple-cause-of-death.html

Data Overview

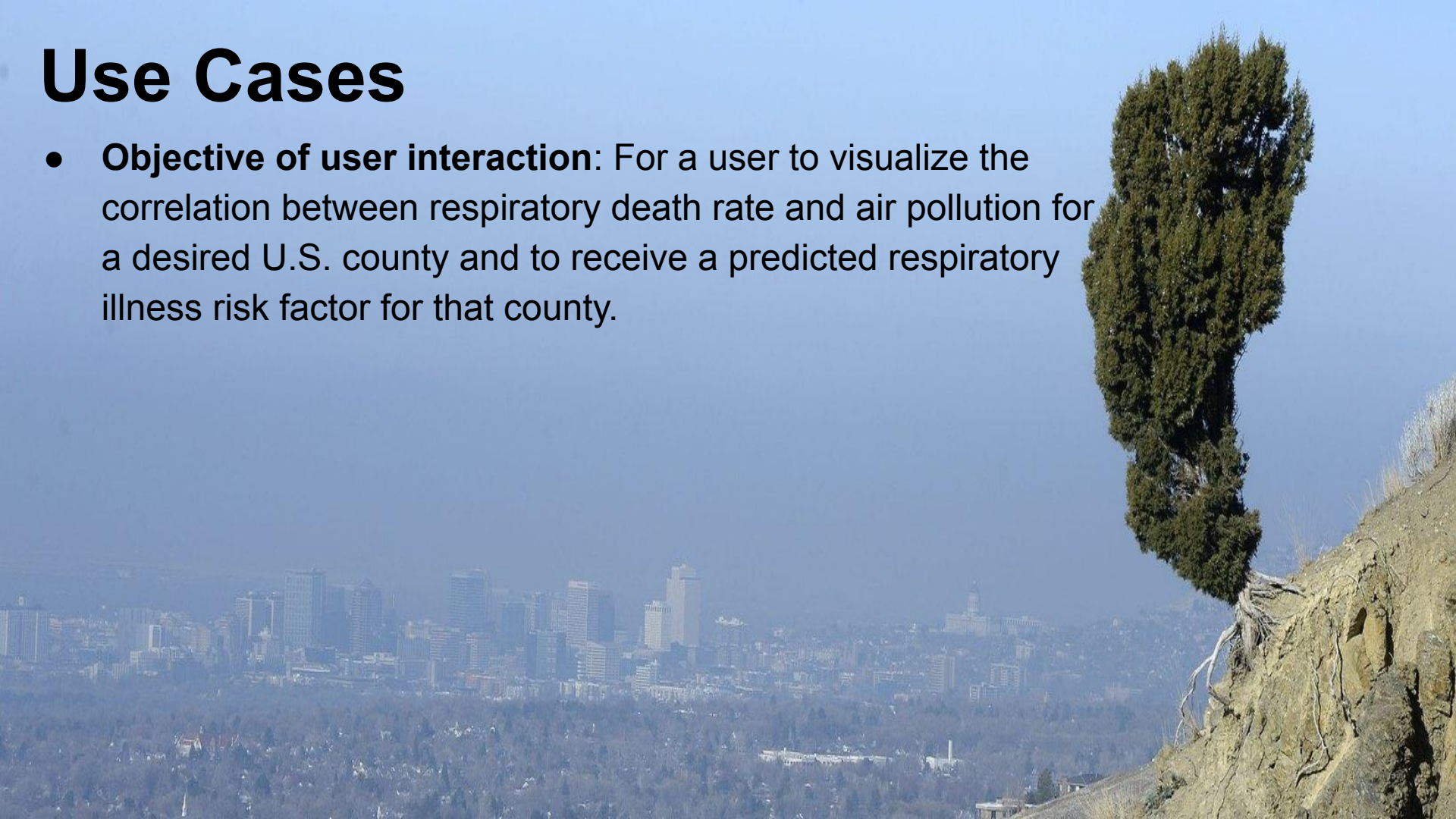


Data Limitations

- Air pollution and respiratory death data is only available for 1,086 counties out of the 3,007 U.S. counties.
- The data is only available for the past 15 years.
- Time resolution is different for the respiratory death (monthly) and the air pollution data (reported on every 3 days).
- Data files contained data for each year, and therefore, required concatenating the data across all years.
- County codes had to be converted into strings so that it contained a leading zero.
- Both data sets did not have the same geographical information required for relating the data to counties.
- Many NaN values

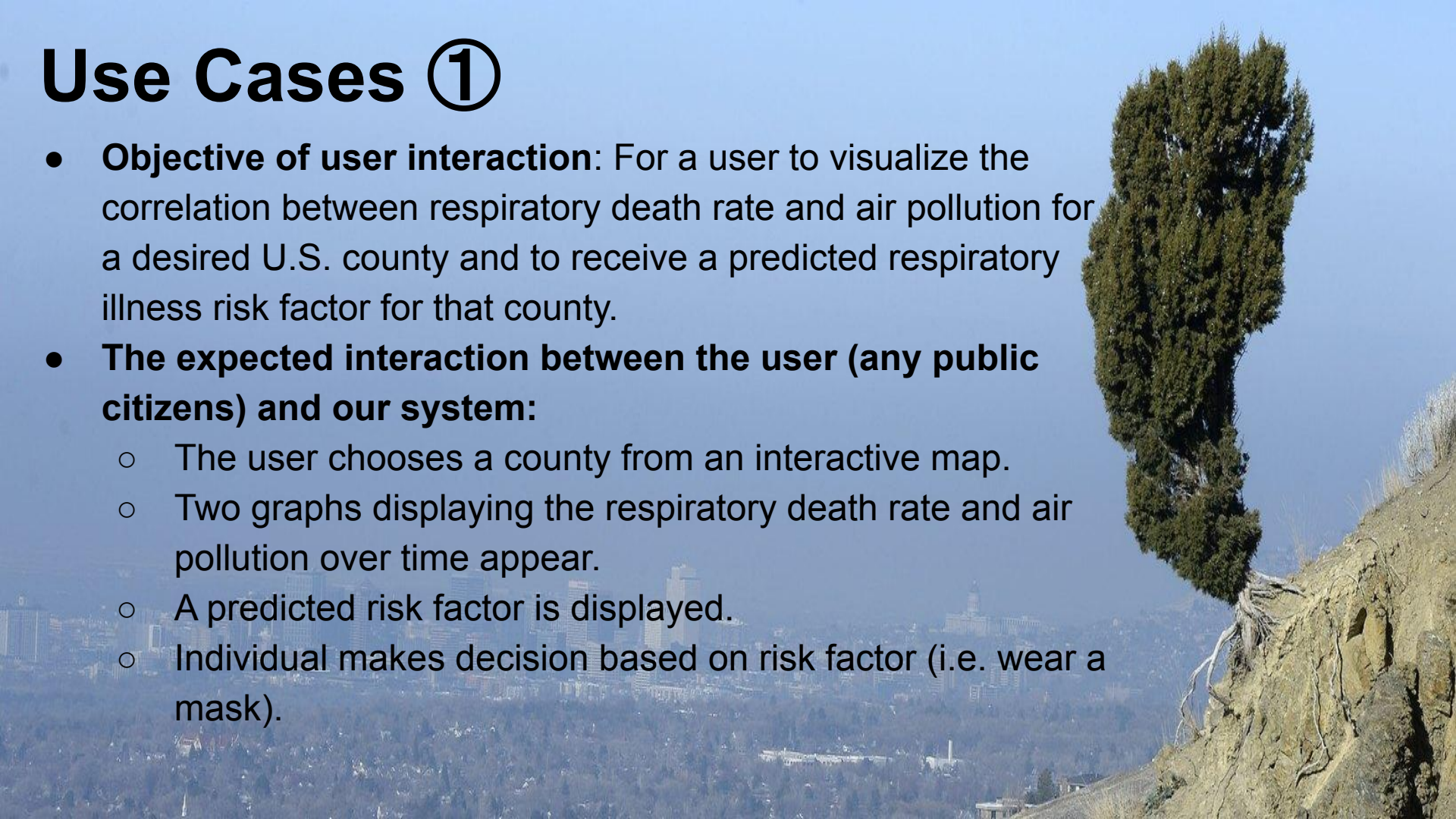
Use Cases

- **Objective of user interaction**: For a user to visualize the correlation between respiratory death rate and air pollution for a desired U.S. county and to receive a predicted respiratory illness risk factor for that county.



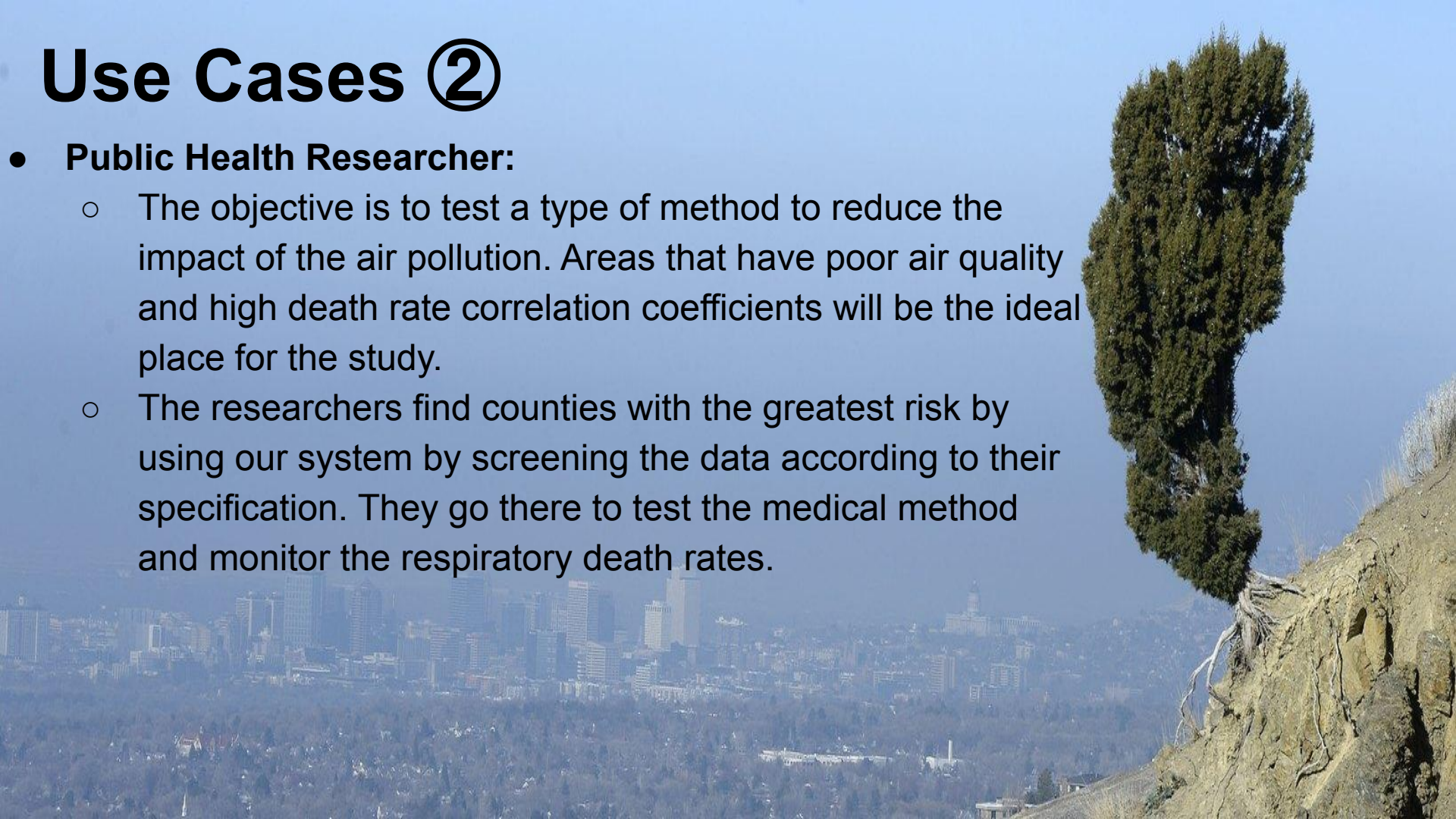
Use Cases ①

- **Objective of user interaction**: For a user to visualize the correlation between respiratory death rate and air pollution for a desired U.S. county and to receive a predicted respiratory illness risk factor for that county.
- **The expected interaction between the user (any public citizens) and our system:**
  - The user chooses a county from an interactive map.
  - Two graphs displaying the respiratory death rate and air pollution over time appear.
  - A predicted risk factor is displayed.
  - Individual makes decision based on risk factor (i.e. wear a mask).



Use Cases ②

- **Public Health Researcher:**
  - The objective is to test a type of method to reduce the impact of the air pollution. Areas that have poor air quality and high death rate correlation coefficients will be the ideal place for the study.
  - The researchers find counties with the greatest risk by using our system by screening the data according to their specification. They go there to test the medical method and monitor the respiratory death rates.



Demo of Interactive Map

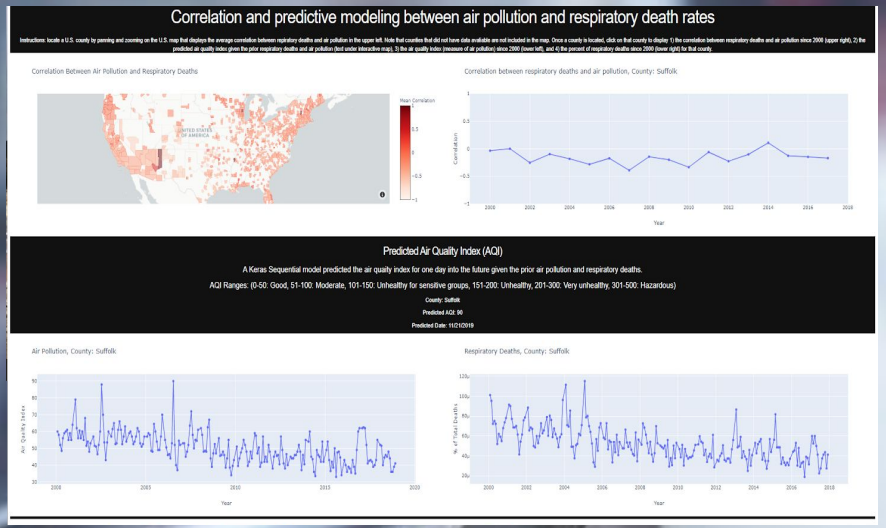
Go to the Interactive_Map Directory:

```
Brandon Pratt@DESKTOP-AHVPLNH MINGW64 ~/Desktop/CSE583/Final_Project/CSE583_FinalProject/air_pollution_death_rate_relate
d/Interactive_Map (master)
$ ls
app.py  AQI_sample.csv  combined_air_data_2000_2019.csv  deathrate_countydata.csv  export_cor.csv  interactive_map.py
(base)
```

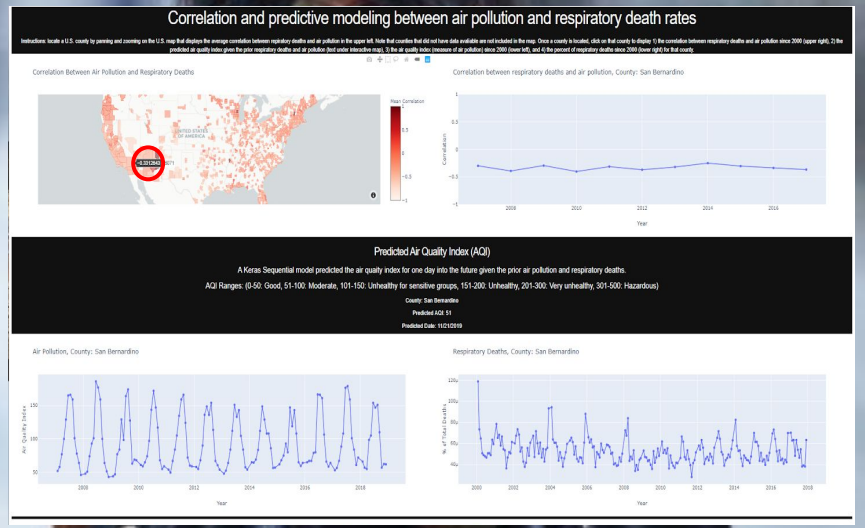
In the command terminal run “python interactive_map.py”:

```
Brandon Pratt@DESKTOP-AHVPLNH MINGW64 ~/Desktop/CSE583/Final_Project/CSE583_FinalProject/air_pollution_death_rate_relate
d/Interactive_Map (master)
$ python interactive_map.py
 * Serving Flask app "interactive_map" (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
 * Debug mode: off
 * Running on http://127.0.0.1:8050/ (Press CTRL+C to quit)
```

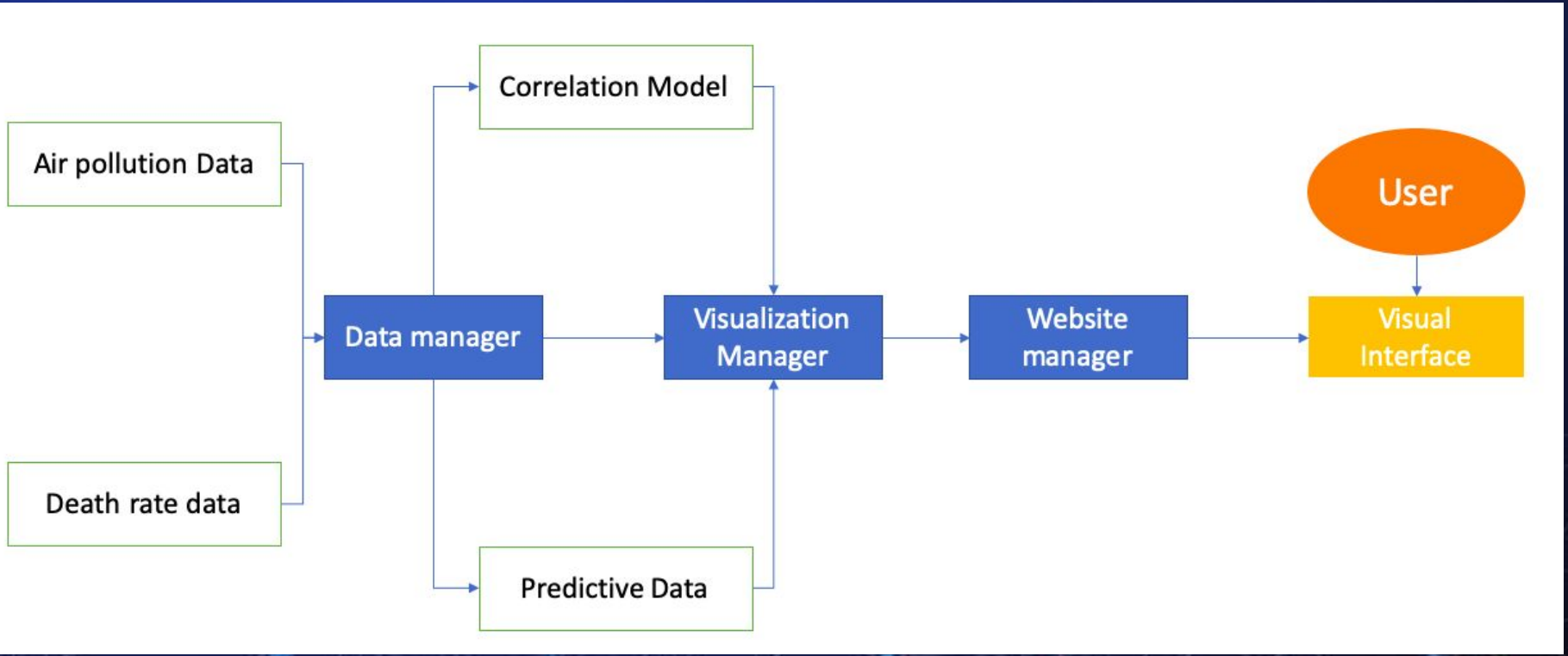
Copy html link into the browser:



Click on a U.S. county on the interactive map in the upper left to display the associated time series data:



Design Components Flowchart



Design Components

Data Manager: Our data is stored on Github, and imported into Python modules. All data needed to be concatenated and was loaded into the python module as a Pandas Dataframe.

Visualization Manager: The analyzed and raw time series data is presented in an interactive U.S. map using the Dash and Plotly libraries. Clicking on a U.S. county displays: 1) the correlation between respiratory deaths and air pollution since 2000 (upper right), 2) the predicted air quality index given the prior respiratory deaths and air pollution (text under interactive map), 3) the air quality index (measure of air pollution) since 2000 (lower left), and 4) the percent of respiratory deaths since 2000 (lower right) for that county.

Design Components

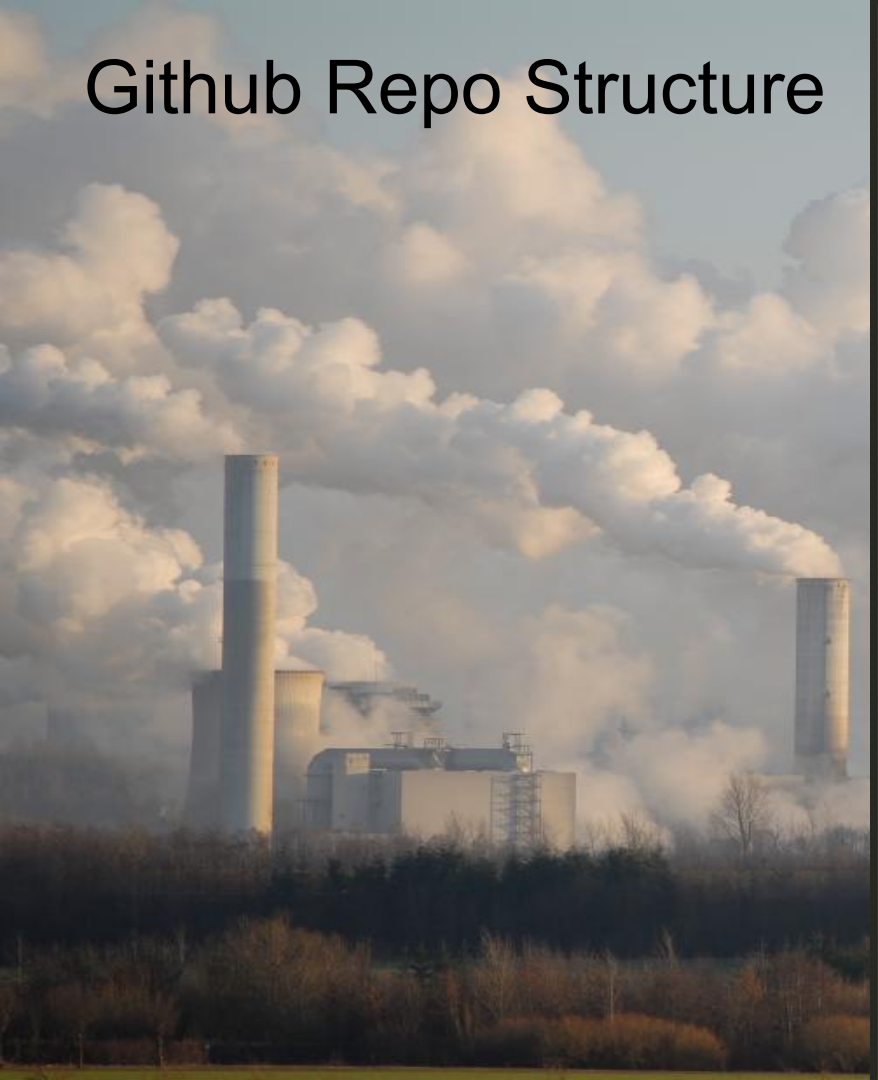
Website manager: we implement the Plotly and Dash python libraries to visualize the U.S. map showing the the correlations of the respiratory death rates and air pollution across counties.

Mathematical model for the correlation between respiratory death rates and air pollution: The air quality index (AQI) and the respiratory death rates data for each county is first loaded into python as a pandas dataframes and then correlated using the scipy.signal.correlate() function in the scipy python package.

Sequential Keras predictive model for predicting the future AQI.

Interactions: the interactions are between the user and the interface, the interface and the webpage, the interface and the mathematical model, the interface and the data manager, and the model and the data manager.

Github Repo Structure



```
CSE583_FinalProject/
  |- README.md
  |- Death_Air_Pollution/
     |- Data/
        |-Respiratory_Death/
            |- ...
        |-Air_Pollution/
            |- ...
     |- Scripts/
        |-Model_AQI/
            |- ...
        |-Correlation/
            |- ...
        |-Respiratory_Death/
            |- ...
     |- Tests/
        |- ...
     |- Trained_AQI_Model/
        |- ...
     |- Interactive_App/
        |- interactive_map.py
        |- associated_data/
            |- ...
  |- examples
     |- User_Guide
  |- logos
     |- image.png
  |- doc/
     |- FunctionalSpec
     |- ComponentSpec
     |- TechnologyReview
     |-Final presentation
  |- setup.py
  |- LICENSE
  |- requirements.txt
  |- environment.yml
```

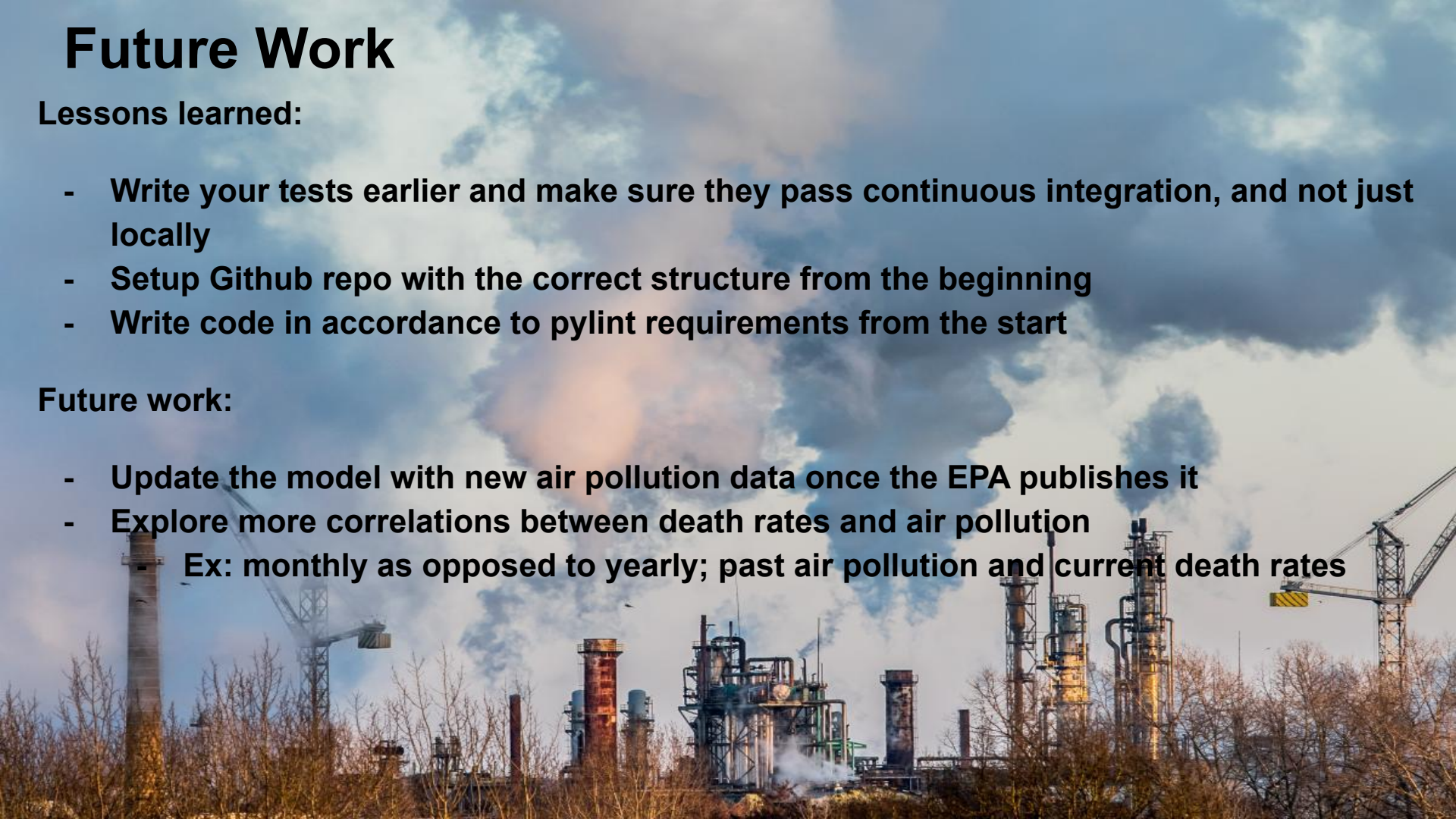
Future Work

Lessons learned:

- **Write your tests earlier and make sure they pass continuous integration, and not just locally**
- **Setup Github repo with the correct structure from the beginning**
- **Write code in accordance to pylint requirements from the start**

Future work:

- **Update the model with new air pollution data once the EPA publishes it**
- **Explore more correlations between death rates and air pollution**
  - **Ex: monthly as opposed to yearly; past air pollution and current death rates**



Thank You!

Question?