

Air Pollution Leads To DEATH?

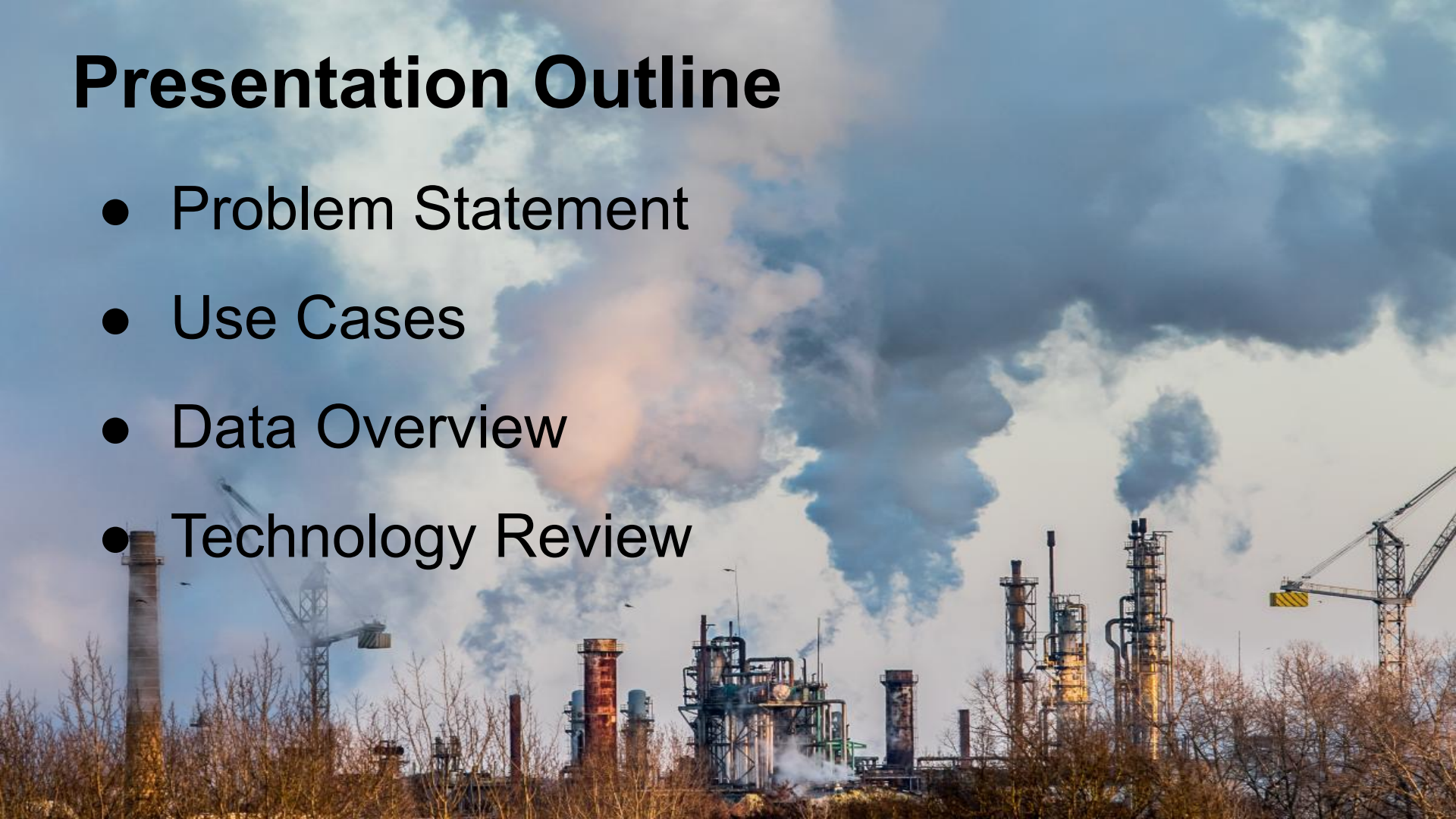
November 4, 2019

Tram Nghi Pham, Brandon Pratt, Siting Wang,
and Marta Wolfshorndl



Presentation Outline

- Problem Statement
- Use Cases
- Data Overview
- Technology Review



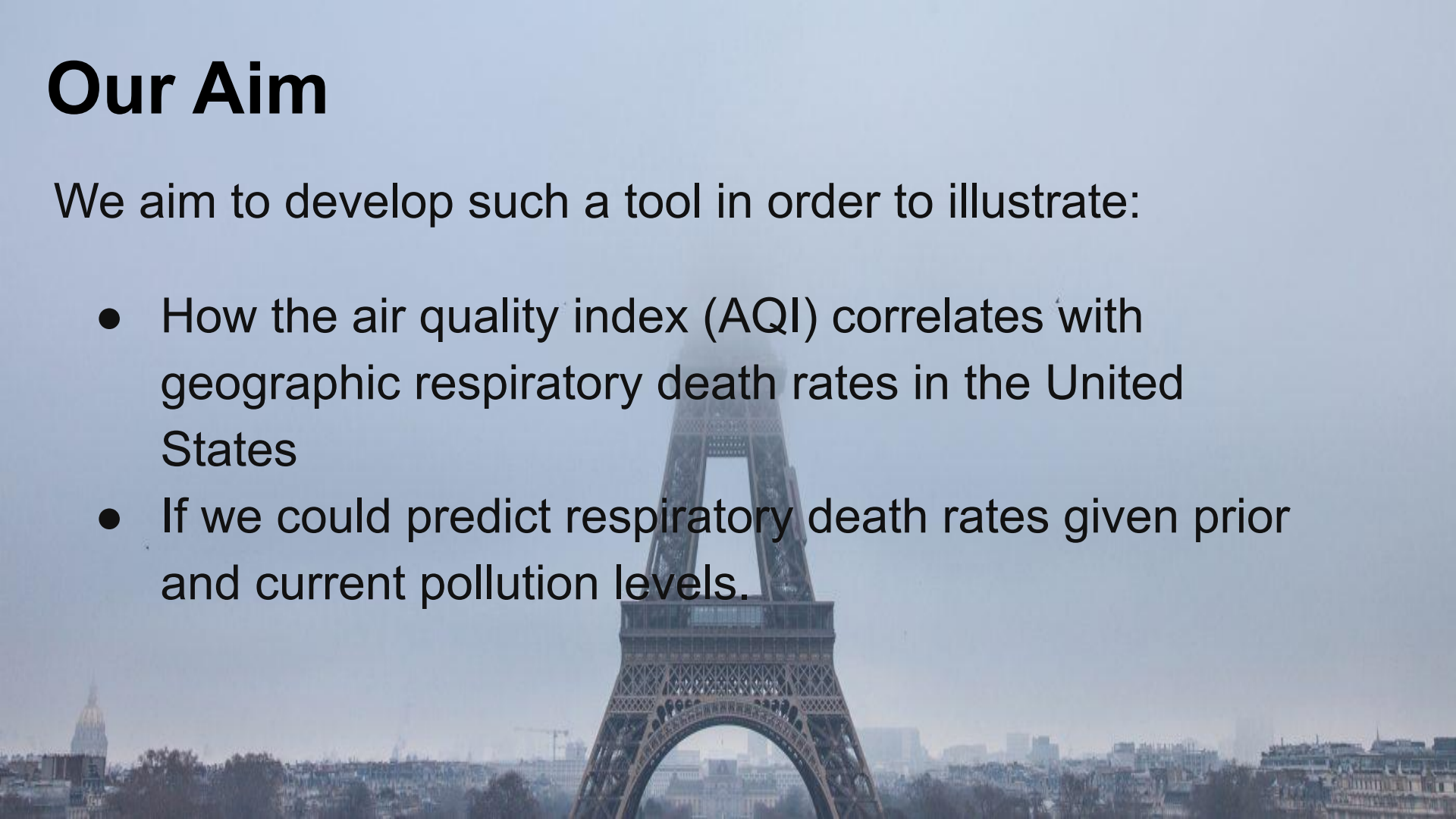
Problem Statement

- Air pollution is an increasing concern across the globe as it has been linked to the development of respiratory illness
- 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.



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. country and to receive a predicted respiratory illness risk factor for that county.



Use Cases

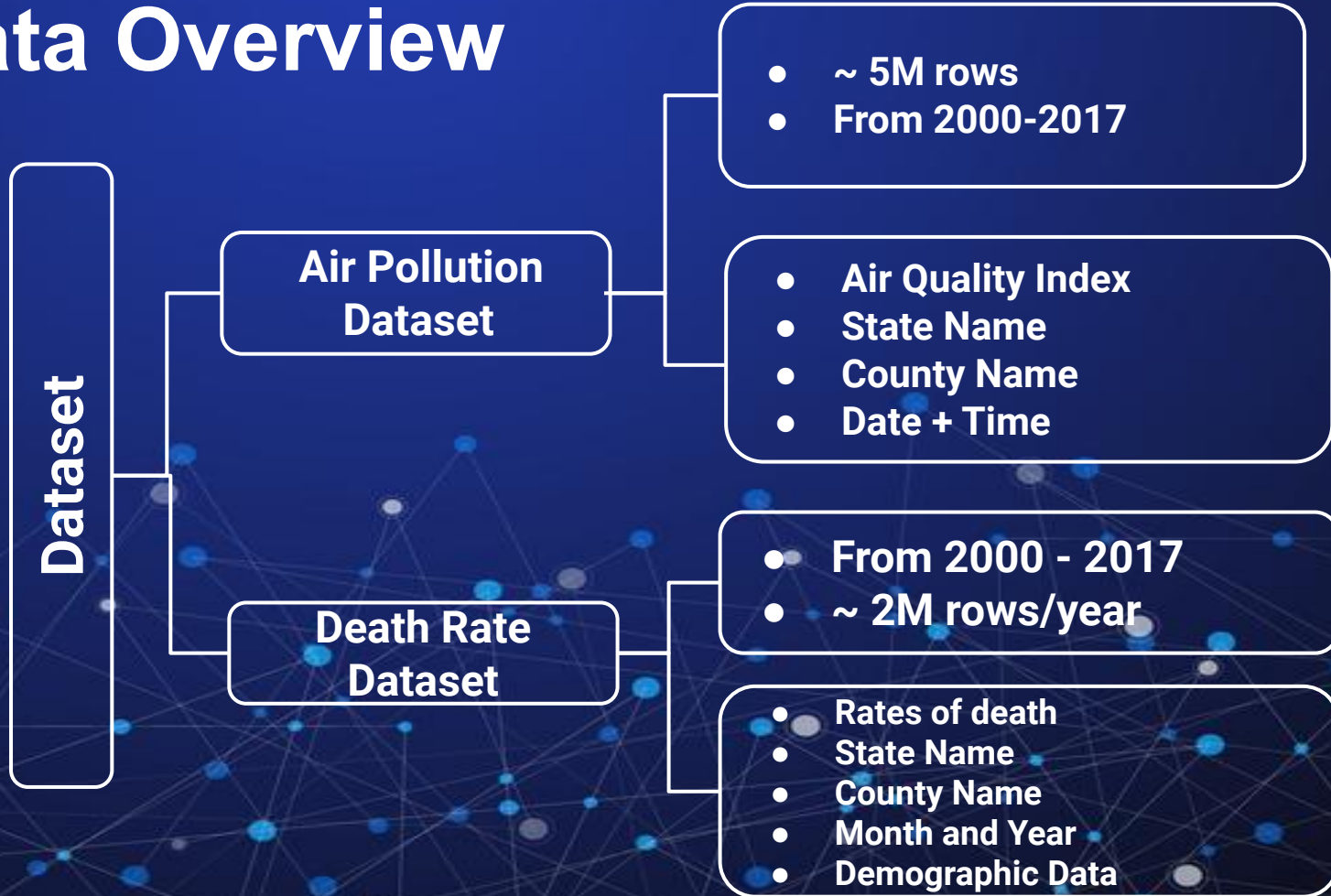
- **Objective of user interaction:** For a user to visualize the correlation between respiratory death rate and air pollution for a desired U.S. country and to receive a predicted respiratory illness risk factor for that county.
- **The expected interaction between the user (any public citizens) and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



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



Model Evaluation

Death Rate and Air Pollution Correlation

- Pearson Correlation
- Spearman Correlation



Modeling Evaluation

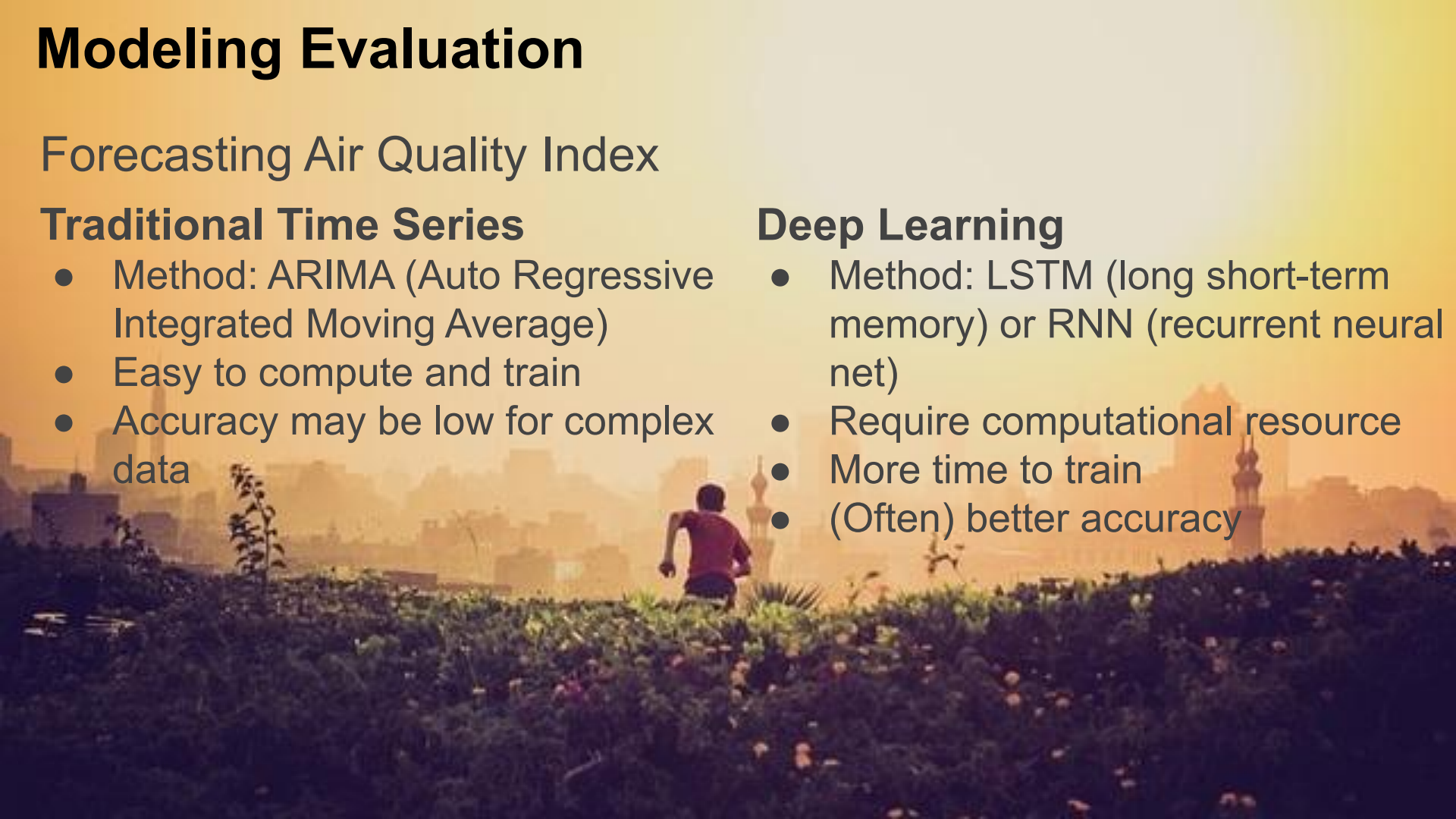
Forecasting Air Quality Index

Traditional Time Series

- Method: ARIMA (Auto Regressive Integrated Moving Average)
- Easy to compute and train
- Accuracy may be low for complex data

Deep Learning

- Method: LSTM (long short-term memory) or RNN (recurrent neural net)
- Require computational resource
- More time to train
- (Often) better accuracy



Technical Tools Evaluations

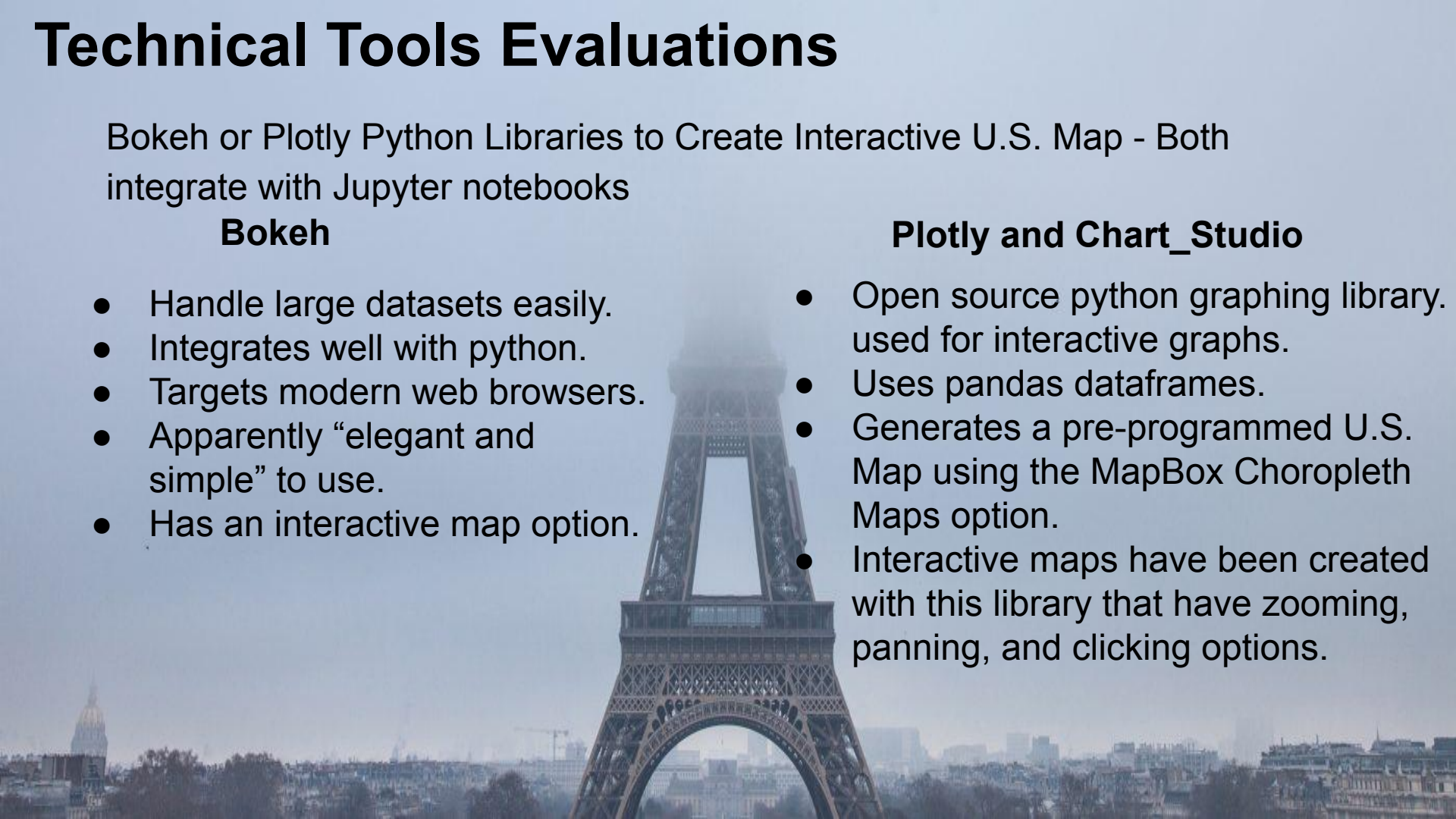
Bokeh or Plotly Python Libraries to Create Interactive U.S. Map - Both integrate with Jupyter notebooks

Bokeh

- Handle large datasets easily.
- Integrates well with python.
- Targets modern web browsers.
- Apparently “elegant and simple” to use.
- Has an interactive map option.

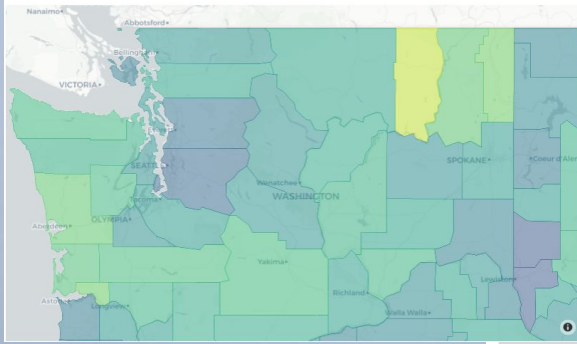
Plotly and Chart_Studio

- Open source python graphing library used for interactive graphs.
- Uses pandas dataframes.
- Generates a pre-programmed U.S. Map using the MapBox Choropleth Maps option.
- Interactive maps have been created with this library that have zooming, panning, and clicking options.

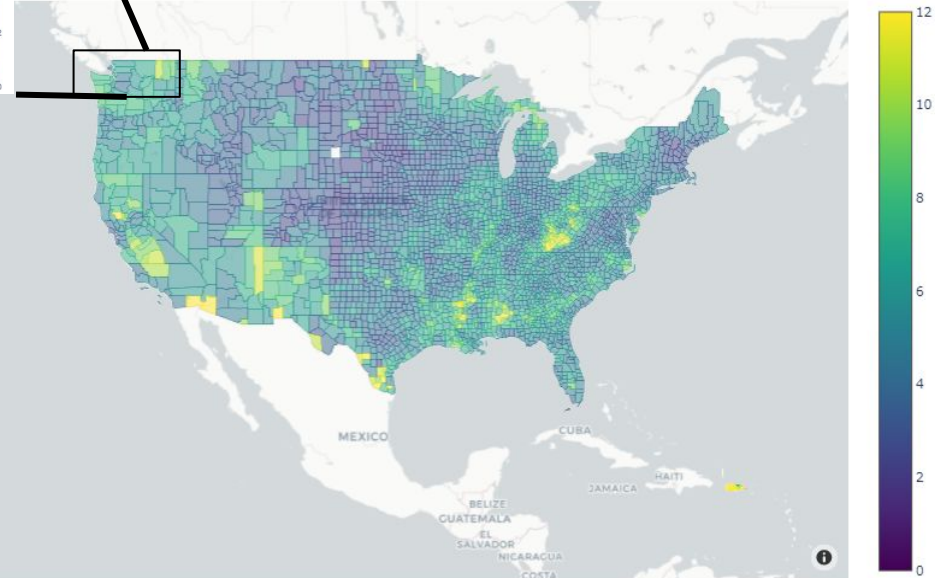


Plotly and Chart_Studio

Panning and Zooming in on Washington counties



Interactive map of the U.S. showing unemployment rates across U.S. counties.



Tutorial code used to make map:

```
# Import python Libraries
from urllib.request import urlopen
import json
with urlopen('https://raw.githubusercontent.com/plotly/datasets/master/geojson-counties-fips.json') as response:
    counties = json.load(response)

import pandas as pd
df = pd.read_csv('https://raw.githubusercontent.com/plotly/datasets/master/fips-unemp-16.csv',
                 dtype={"fips": str})

import plotly.graph_objects as go

fig = go.Figure(go.Choroplethmapbox(geojson=counties, locations=df.fips, z=df.unemp,
                                   colorscale="Viridis", zmin=0, zmax=12,
                                   marker_opacity=0.5, marker_line_width=0))

fig.update_layout(mapbox_style="carto-positron",
                  mapbox_zoom=3, mapbox_center = {"lat": 37.0902, "lon": -95.7129})
fig.update_layout(margin={"r":0,"t":0,"l":0,"b":0})
fig.show()
```

Bokeh

```
In [38]: from bokeh.io import show
from bokeh.models import LogColorMapper
from bokeh.palettes import Viridis6 as palette
from bokeh.plotting import figure

from bokeh.sampledata.us_counties import data as counties
from bokeh.sampledata.unemployment import data as unemployment

palette.reverse()

counties = {
    code: county for code, county in counties.items() if county["state"] in ["tx", 'sc', 'sd', 'tn', 'tx',
    'ut', 'vt', 'va', 'wa', 'wv', 'wi', 'wy', 'al', 'id', 'il', 'in', 'ky', 'la', 'me', 'md', 'ma', 'mi', 'mn', 'ms', 'mo', 'mt', 'ne', 'nv', 'nh'
    ]
}

county_xs = [county["lons"] for county in counties.values()]
county_ys = [county["lats"] for county in counties.values()]

county_names = [county["name"] for county in counties.values()]
county_rates = [unemployment[county_id] for county_id in counties]
color_mapper = LogColorMapper(palette=palette)

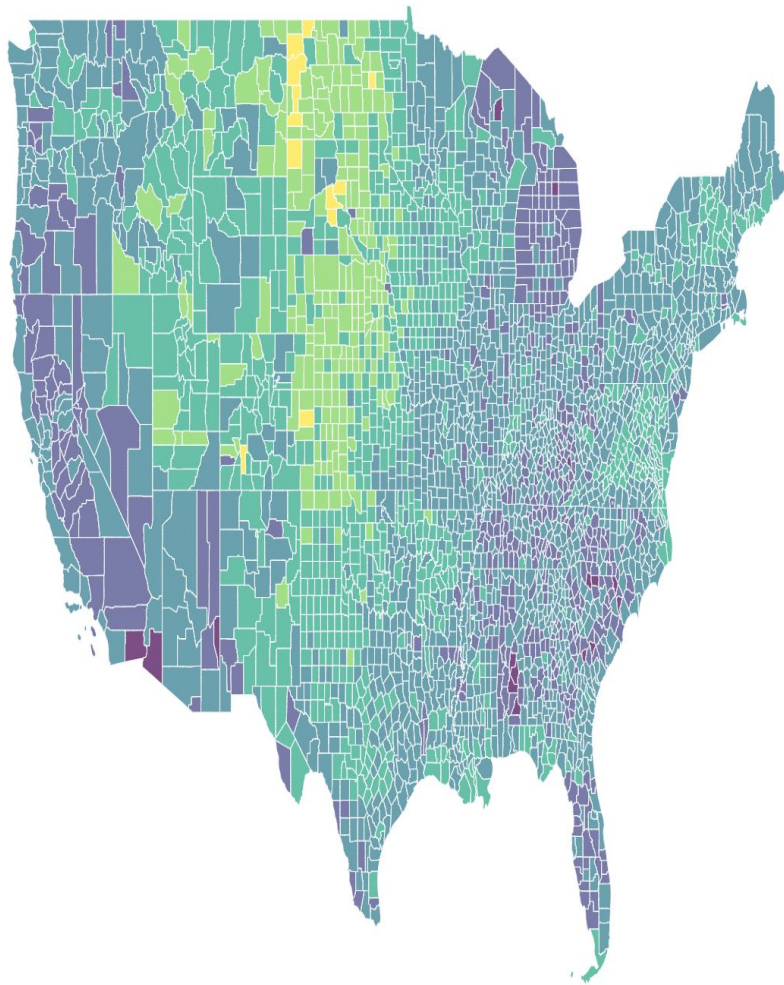
data=dict(
    x=county_xs,
    y=county_ys,
    name=county_names,
    rate=county_rates,
)

TOOLS = "pan,wheel_zoom,reset,hover,save"

p = figure(
    title="Texas Unemployment, 2009", tools=TOOLS,
    x_axis_location=None, y_axis_location=None,
    tooltips=[
        ("Name", "@name"), ("Unemployment rate", "@rate%"), ("Long, Lat", "($x, $y)")
    ], x_range=(100,100), y_range=(100,100))
p.grid.grid_line_color = None
p.hover.point_policy = "follow_mouse"

p.patches('x', 'y', source=data,
    fill_color={'field': 'rate', 'transform': color_mapper},
    fill_alpha=0.7, line_color="white", line_width=0.5)

show(p)
```



Bokeh Versus Plotly

Bokeh

Pros:

- Generates interactive map.
- Generates figure URL.
- Interactive tool properties are adjustable.
- Able to link to Google Maps.

Cons:

- Documentation is limited for creating interactive maps.
- Requires a lot of lines of code to generate the map.
- Doesn't use the pandas library directly.
- Difficult to learn.



Plotly and Chart_Studio

Pros:

- Easy to use and requires little code to generate the interactive U.S. map.
- Interactive features (i.e. panning, zooming, and clicking) are easy to implement.
- Works well with pandas library.
- The API documentation is clear and understandable.

Cons:

- Must pay to generate URL for interactive map: **PlotlyRequestError**: This file is too big! Your current subscription is limited to 524.288 KB uploads. For more information, please visit: <https://go.plot.ly/get-pricing>.



Thank You!

Question?