

Part 1: Common Analysis

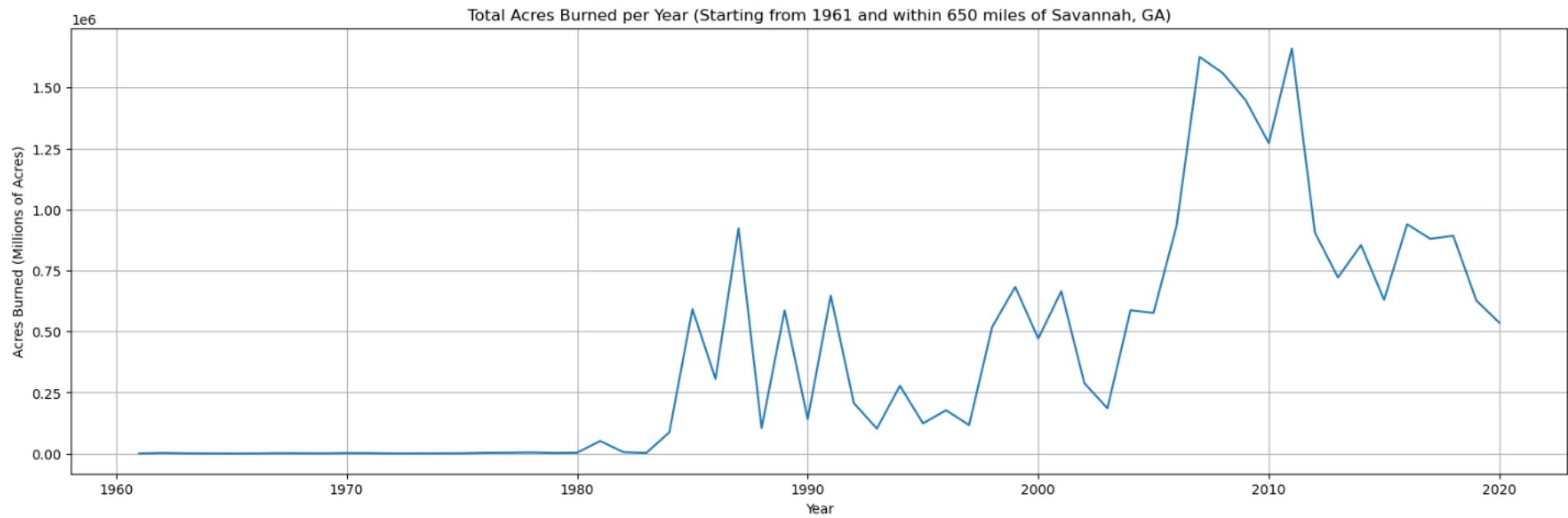
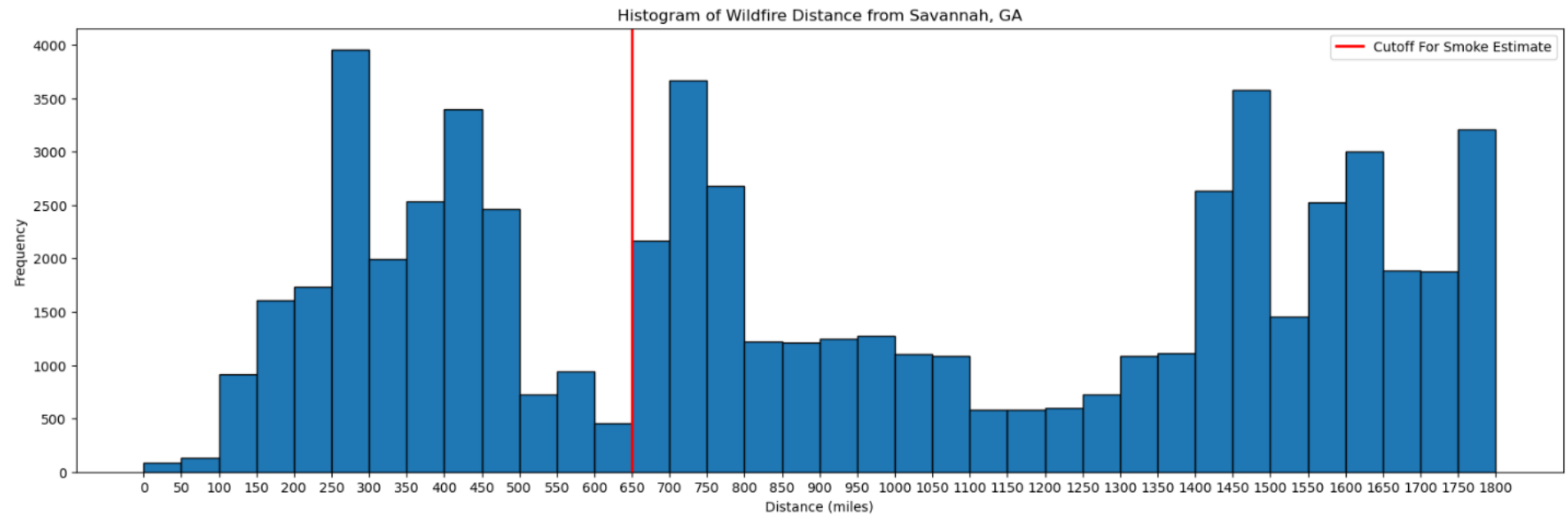
Assigned City: Savannah, GA

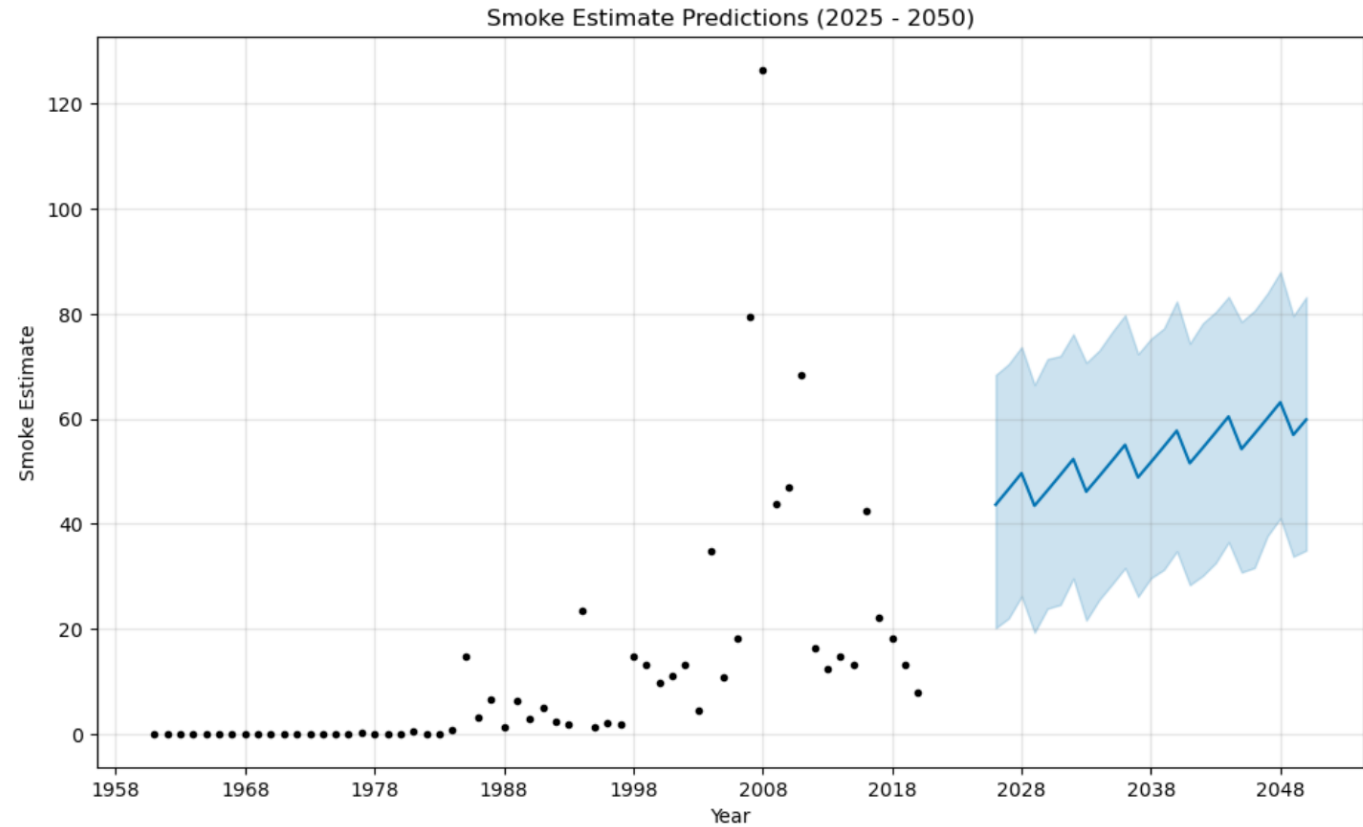
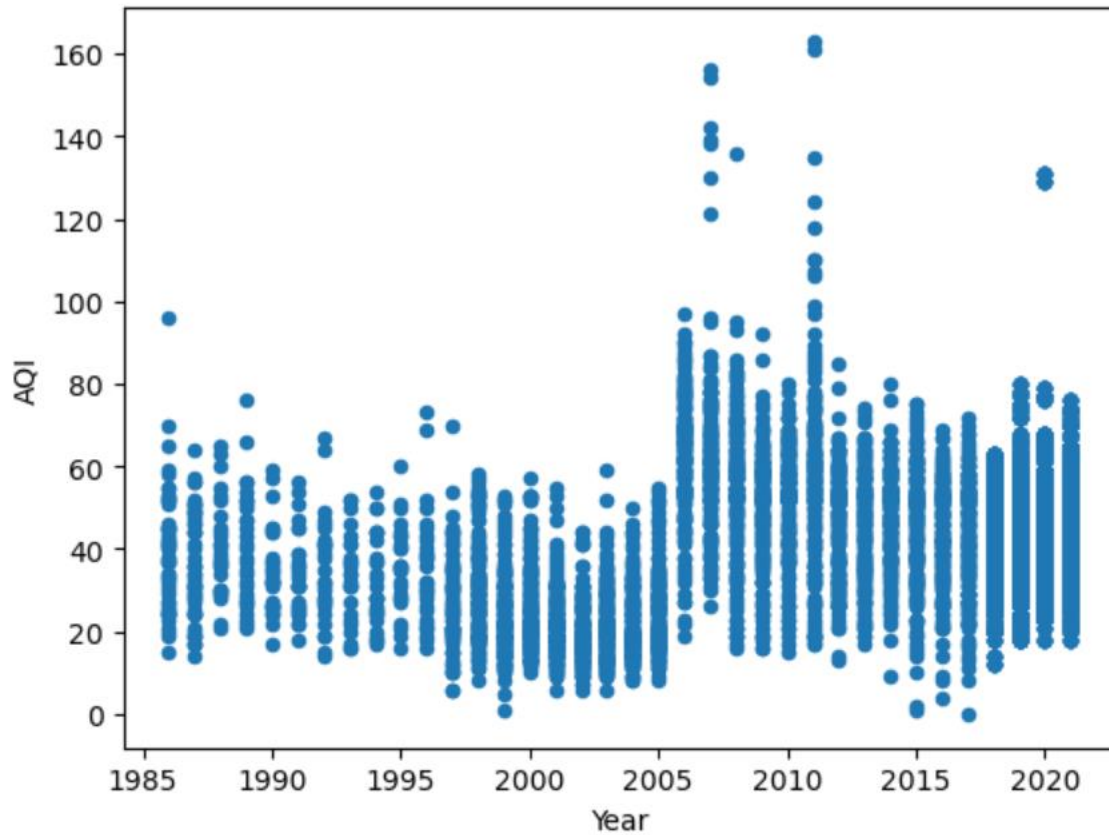
- Oldest City in Georgia
- Known for its natural beauty (parks, beaches, etc.)

Goals:

- Visualize wildfire proximity and size
- Create a smoke estimate model for the city
- Access AQI metrics for the city







- Smoke Estimate and Prophet Model on the Right
- AQI data on the Left

- Roughly similar shape with a peak between 2006 and 2010
- Predicted decrease in air quality through 2050

$$\text{Smoke Estimate} = \frac{\text{Fire Area}}{(\text{Distance})^2}$$

Part 2: Project Extension

Original Extension Plan

- Housing/Rental Data
- Explore effect of air quality on tourism
- Data was not good (missing, short time period, etc)

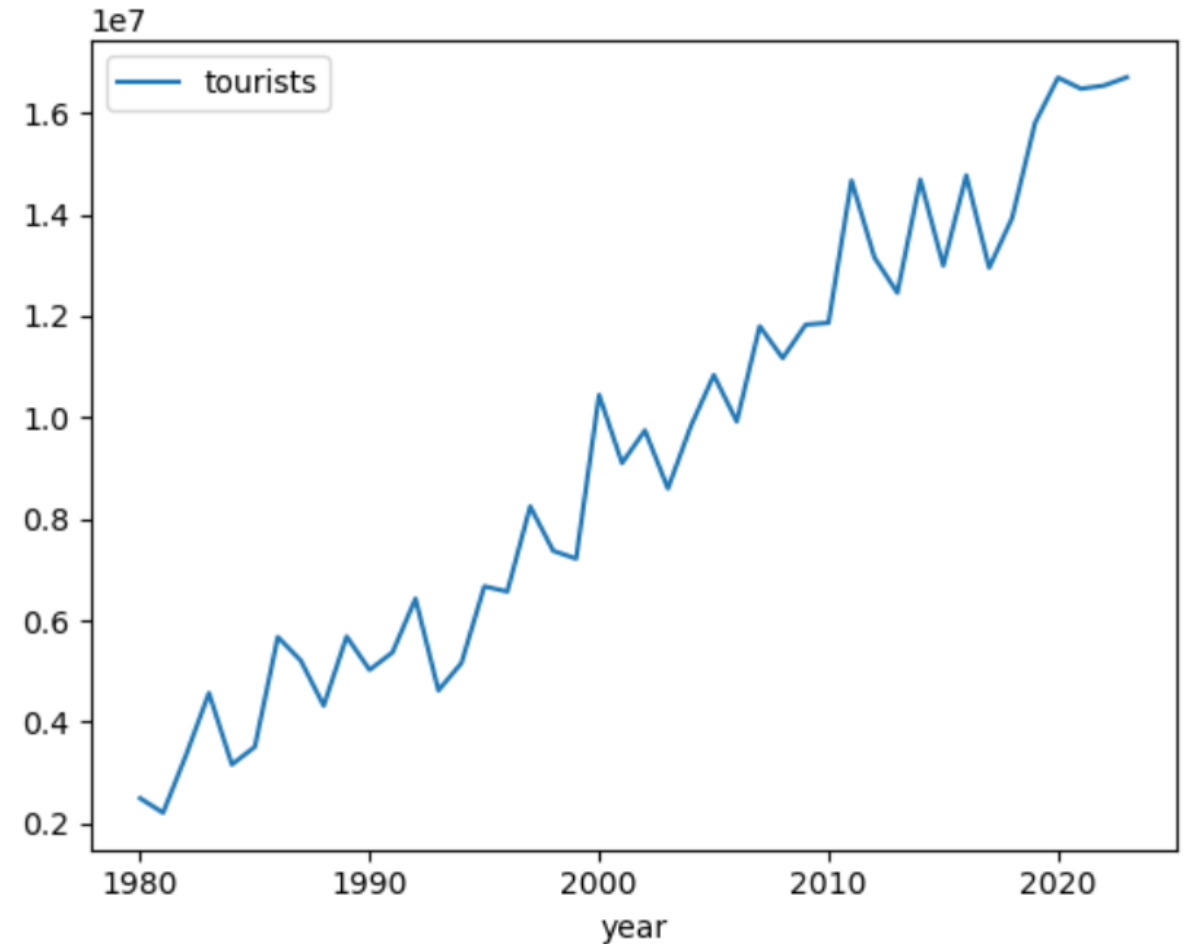
New Extension Plan

- Tourism data gathered by the city
- Extend the smoke estimate from part 1 to measure “human impact”

Dataset: Savannah Tourism Estimates

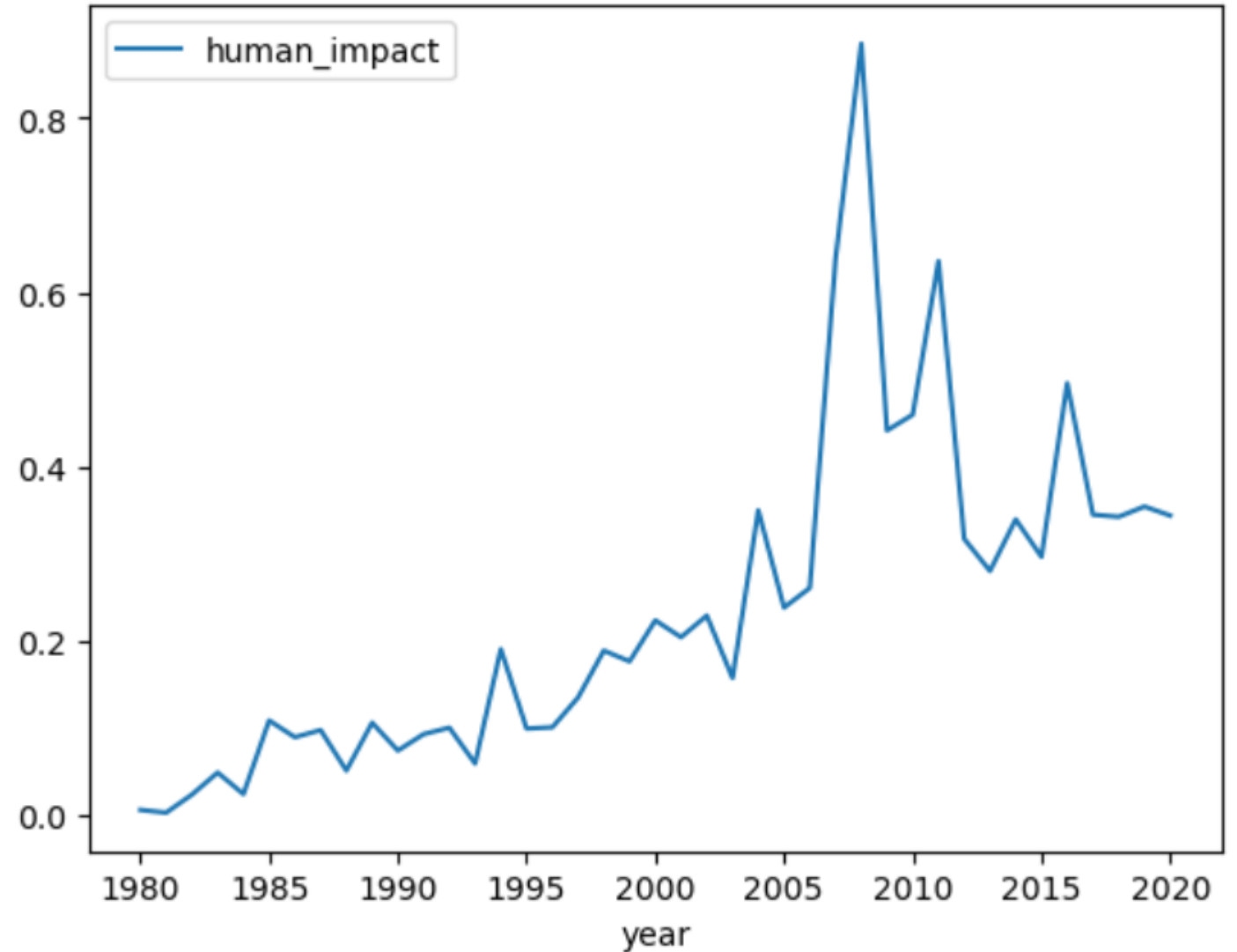
- Gathered by the city, available by request
- Columns:
 - Year: 1980 - 2023
 - Tourists: 2 million – 17 million
 - Overnight-Visit: Max 1-day visitors
 - Day-Visit: More than 1-day visitors
- Metric:
 - “Human Impact Estimate”

```
df_tourism = pd.read_csv("tourist_data.csv")  
df_tourism.plot(x='year', y='tourists', kind='line')  
plt.show()
```

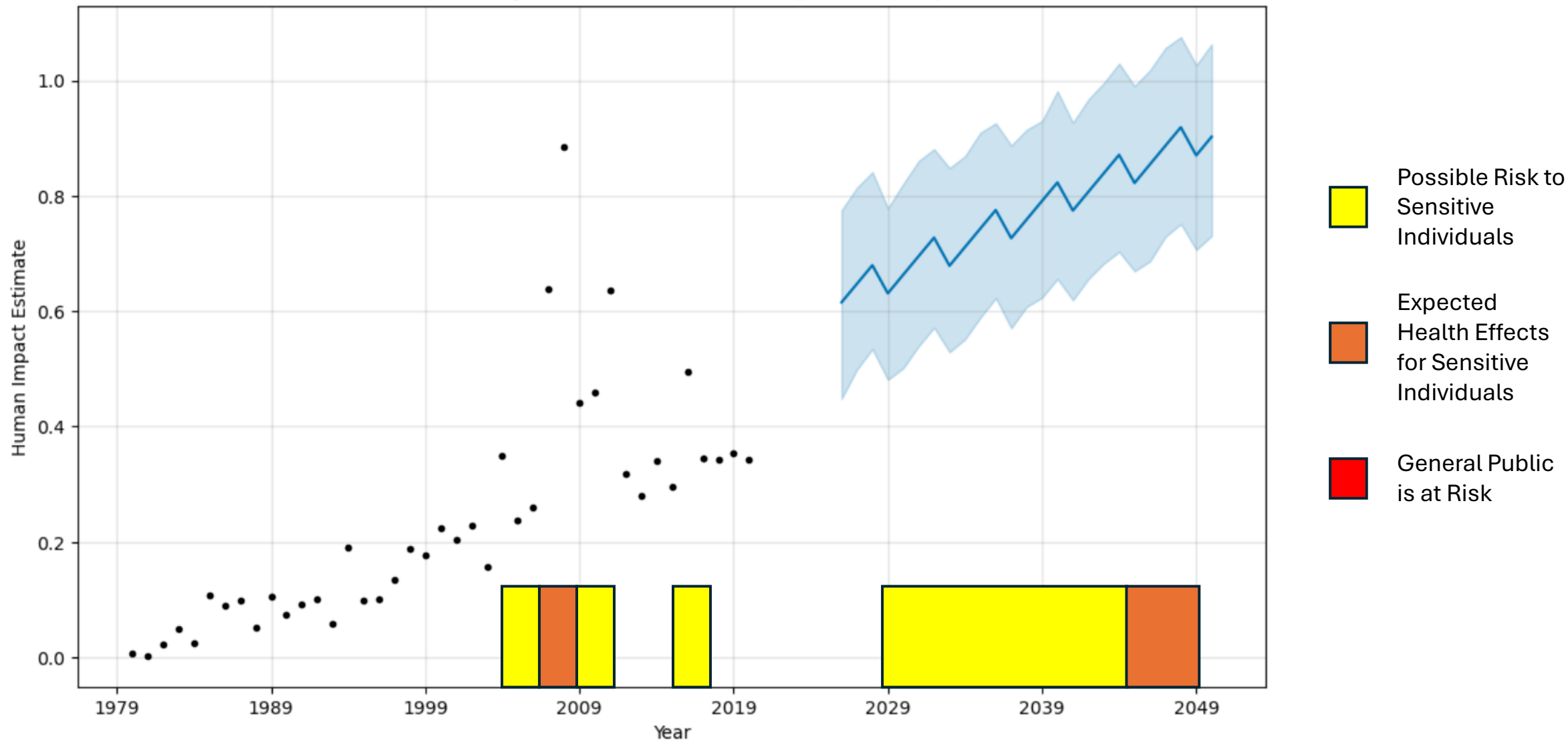


Metric and Model Creation

- Normalize the component data
 - Smoke Estimate: ~ 0 to 170
 - Tourism: ~ 2 mil to 17 mil
- Weights: [0.7, 0.3]
- Linear combination



Human Impact Estimate Predictions (2025 - 2050)



Takeaways and Impact

- Worsening air quality is a threat to the increasing tourism rate
- Smoke estimate reaches “moderate” by late 2030’s
- Human Impact metric reaches 1.0 (high) around 2050

Call To Action:

- City Level: Green infrastructure, Energy Policy, Health Campaigns
- State Level: Wildfire Management, Improved Monitoring

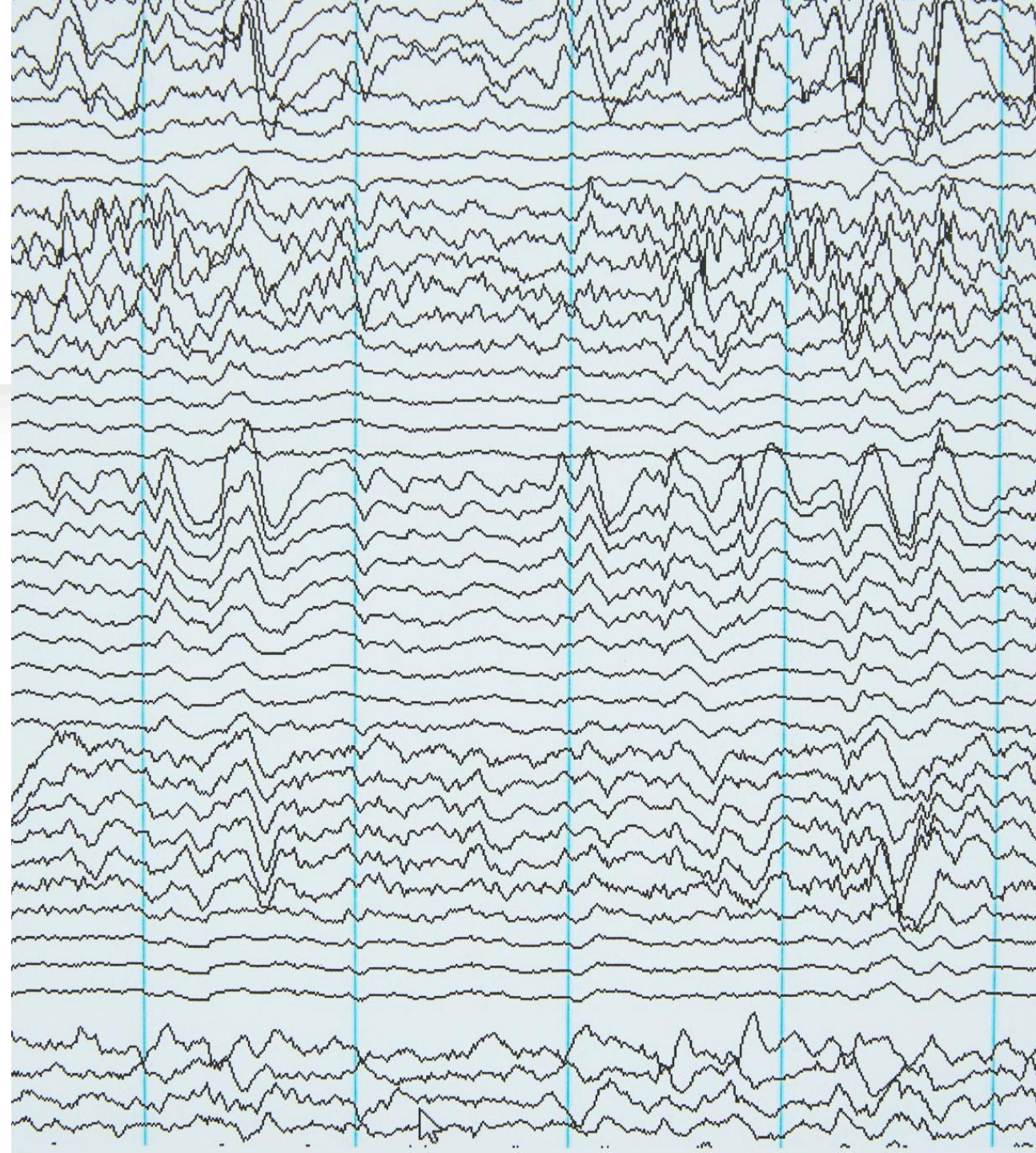
Limitations

Main Limitation:

- Bad air quality will affect tourism numbers
- Variables are not independent, so model is slightly inaccurate

Smaller Limitations:

- Smoke estimate data only through 2020
- “Length of stay” is binary rather than a gradient
- Wind patterns play a large impact in fires that are “far away”



Project Reflection



Data Acquisition is hard

Paywalls

Missing data



Experience with “real world” data science



Creating a well documented and reproducible project