Respiratory Health Impacts of Wildfire Smoke in Redmond, OR

Yash Manne | DATA 512 | 12/11/2023

I. Introduction

Situated in Oregon's High Desert Plateau, Redmond has evolved from a small-town community to a mid-sized city of over 30,000 residents within the past 20 years.[1] Recognized as the “HUB” of Central Oregon, Redmond is estimated to reach 50,000 residents by 2040.[1] In the face of the sudden expansion, the city has been mindful of navigating the growing pains and identifying potential future challenges.

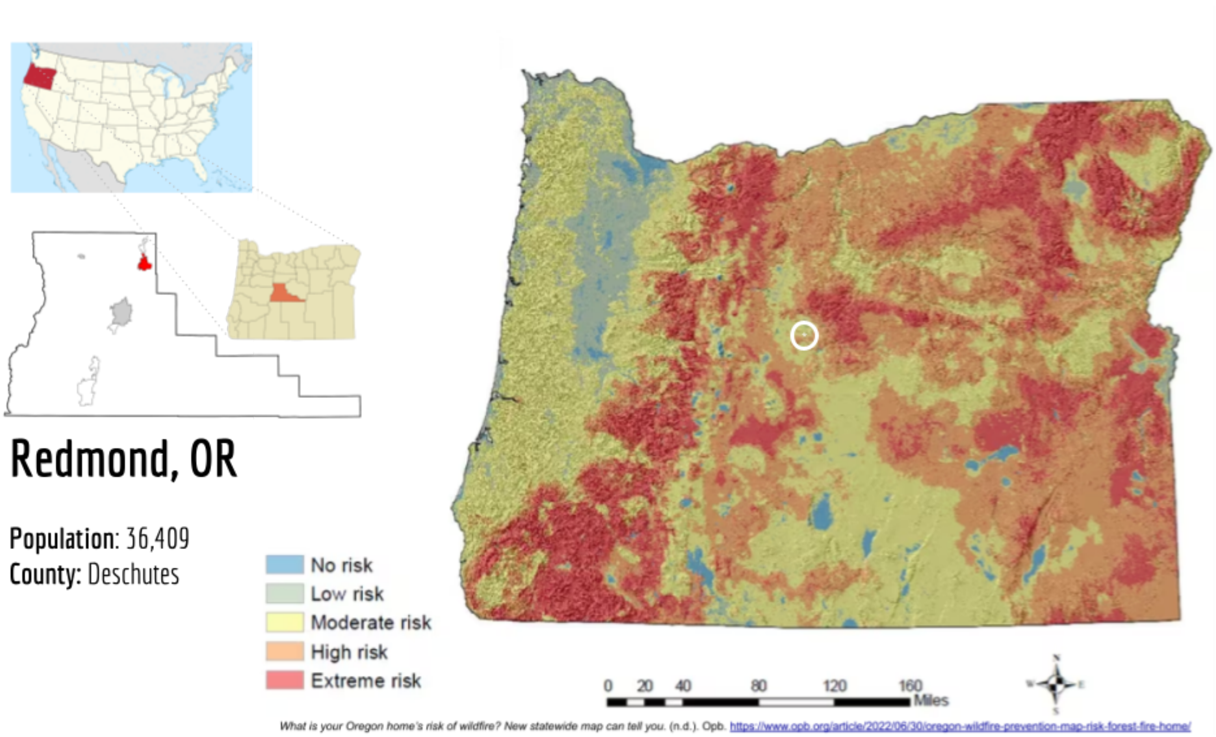

Figure 1: Overview of Redmond, OR

This graphic details the geographic location of Redmond, OR in the context of the United States, Oregon, and Deschutes County. Additionally, the graphic from the Oregon Wildfire Prevention Map highlights areas of high wildfire risk (orange-red regions) surrounding Redmond, OR (white circle).

One such challenge is the impact of wildfire smoke on the community. We see in **Figure 1** that the region surrounding Redmond faces an extreme risk of wildfires.[11] So, it is imperative to understand any health effects stemming from smoke exposure. Moreover, as Redmond's population rapidly increases, a better understanding is crucial to mitigate a potential future strain on the existing healthcare system with effects permeating to lower quality of life and other broad socioeconomic ramifications. This project focuses on analyzing the potential effects of wildfire smoke on public health and aims to propose measures to safeguard the well-being of both current and future residents. Specifically, I investigate a potential link between wildfire smoke and chronic respiratory illnesses among citizens.

From a scientific standpoint, this analysis contributes to the broader understanding of the health consequences associated with wildfire smoke exposure in a growing urban environment. Redmond's commitment to undertaking projects that benefit its residents makes this investigation particularly relevant and timely. The practical implications of this study lie in informing policymakers and city officials about potential health risks, thereby aiding in developing targeted strategies to mitigate the impact of wildfire smoke. Potential solutions include increasing public awareness, distributing N95 masks to vulnerable populations, and enhancing medical support during wildfire season.[5] With a state-wide reputation for effective project implementation, Redmond is positioned to address the challenges posed by wildfire smoke and enhance the health outcomes of its residents. By delving into this issue, I contribute valuable insights that support Redmond's dedication to the welfare of its community.

II. Background

II.I Background Research

The biggest component of wildfire smoke is particle pollution – specifically, fine particulate matter, known as PM2.5, which consists of tiny airborne particles with a diameter of 2.5 micrometers or smaller.[2] These particles can penetrate the respiratory system and potentially enter into circulation, posing significant health risks. According to the Environmental Protection Agency (EPA), it has been shown that brief exposures to fine particles, such as those found in wildfire smoke, are linked to a greater risk of worsening respiratory & cardiovascular diseases, ultimately resulting in premature mortality.[3] As such, people with chronic lung diseases such as asthma or chronic obstructive pulmonary disease (COPD) are at an elevated risk for fine particle-related health effects.[4]

It should be noted that all these studies are conducted at the national level and still need to be validated at the local level for Redmond.

II.II Research Questions

Through this analysis, I validate the link between wildfire smoke and chronic respiratory illnesses in Deschutes County by tracking hospitalizations due to asthma and

COPD from 2000 to 2021. Similarly, I investigate whether there is a tenable pattern between wildfire smoke exposure and premature mortality over the same timeframe. Additionally, I look for patterns between the county-level data and aggregated state-level data for incidence, prevalence, and mortality for chronic respiratory illnesses. In both cases, I model the metrics using a custom smoke estimate, which incorporates the size, distance, and fire type from all wildfires within 1250 miles of Redmond, OR. I then ultimately forecast the future impact of wildfires on each health metric over the next 25 years.

Specifically, I ask the following questions:

1. Does wildfire smoke have a significant effect on the city's respiratory health?
2. How has wildfire smoke intensity changed and how is this predicted to impact the city's respiratory health?

My initial hypotheses were that the increasing smoke levels over the past 20 years could be linked to an increase in respiratory illness occurrence, hospitalizations, and even mortality. Additionally, I expect each of these indicators to continue rising with the rising smoke levels in the coming years.

Based on the strength of the evidence and the severity of the effect size, I can then recommend the Redmond City Council pursue a select set of potential solutions mentioned above.

II.III Data

I accessed cleaned data of 135,061 US-based wildfires from the mid-1800s to 2021 collated by the US Geological Survey as the Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons) dataset. The data is provided under the public domain with citation.[12] Specifically, I used the GeoJSON data format stored under GeoJSON Files.zip. This folder contains both a raw merged dataset containing duplicates and a "combined", duplicate-free, dataset that comprises both wildfires and prescribed fires from the mid-1800s to the 2021 collated from 40 different original wildfire datasets. I use only the combined data stored as "USGS_Wildland_Fire_Combined_Dataset.json". This file contains 30 different attributes for each fire. The most relevant attributes include geospatial polygons denoting the fire's shape and location, the categorization of a fire on a discrete five-point scale ranging from known prescribed fire to known wildfire, and the year of the fire.

A portion of this analysis required historical Air Quality Index (AQI) data for Redmond, Oregon, which is located in Deschutes County, during the fire season (May 1 - October 31) for each year from 1963 onwards. However, AQI data was only available from 1983 to 2023. The Air Quality Index is a measure designed to tell us how healthy the air is on any given day and is commonly used to track pollutants such as smog or smoke. Generally, a rating of 0-50 indicates healthy, clean air, while 500 is the highest tracked value for hazardous air. A thorough explanation of how AQI is calculated can be found in the EPA Technical Assistance Documentation.

For my analysis, I used the US Environmental Protection Agency (EPA) Air Quality Service (AQS) API to access this historical data.[13] The documentation for the API provides definitions of the different call parameters and examples of the various calls that can be made to the API. Additional information on the Air Quality System can be found in the EPA FAQ. Note that terms of use can be found here. All data accessed through the API lies in the public domain. Specifically, I used the maximal daily average sensor data for monitoring stations in Deschutes County, all of which were <17 miles from Redmond, OR. These daily max values were then averaged for the fire season to get the annual estimate for 1983-2023. There was no data available before 1983. Finding nearby monitoring stations requires the Federal Information Processing Series (FIPS) of the desired city, county, and state which was gathered from the US Census Bureau.

I further incorporate health data aggregated from three main sources:

1. Federal Reserve Economic Database (FRED)
2. Oregon Health Authority's Oregon Tracking Data Explorer
3. Institute for Health Metrics and Evaluation (IHME) at the University of Washington

FRED is an online database of 1000s of economic time series data aggregated from national, international, public, and private sources. It is maintained by the Research

Department at the Federal Reserve Bank of St. Louis. The terms of use for this database can be found here but generally, all data is meant for personal, non-commercial, or educational use. From this database, I accessed the annual Age-Adjusted Death Rate Data for Deschutes County from 1999 to 2020.[6] This data was aggregated from the Centers for Disease Control and Prevention (CDC) and more information can be found here. This specific data set was provided under the public domain with citation. The premature death rate is defined as the total number of deaths where the deceased is younger than 75 years of age per 100,000.

The Oregon Tracking Data Explorer is part of a larger National Environmental Public Health Tracking Network funded by the CDC.[7] As such, data from this source is provided under the public domain with citation. The database asks that users only use the data for statistical analysis and reporting purposes without attempting to learn or disclose the identities of individuals within the data. From this source, I downloaded two annual data sets:

1. Asthma: Hospitalizations and Emergency Department Visits
2. Chronic Obstructive Pulmonary Disorder (COPD): Hospitalizations

The first tracks the annual age-adjusted asthma-related hospitalization rate per 10,000 and the raw counts for adults older than 25 from 2000 to 2021. Additionally, I have access to the crude rates for 20-year age bins of 25-44, 45-64, and 65-84. I also have access to similar data for emergency department visits but due to low data history (2018—), I ignore it for my analysis. The second tracks the same metrics for COPD-related annual hospitalizations. Age-adjusted rates allow for fairer comparisons to be made between groups with different age distributions.[8]

The IHME data is provided under the IHME FREE-OF-CHARGE NON-COMMERCIAL USER AGREEMENT. The data provided by IHME arises from the Global Burden of Disease Study in 2019 shown in an interactive query tool.[9] From this query tool, I accessed the raw annual count and rate per 100,000 for the deaths, incidence, and prevalence of asthma, COPD, and all chronic respiratory illness diseases from 1990 to 2019 for the state of Oregon. Incidence monitors the number of new cases in a given year whereas prevalence tracks the total cases in a given year. In the data set, I have access to both the estimated metric as well as lower and upper bound values.

II.IV Model Selection

While there may be some existing models linking smoke to asthma hospitalizations, none are completely open-source and can be readily applied in the context of the limited data granularity I have for this project. Instead, we'll rely mainly on fundamental time series models like AutoRegressive Integrated Time Series (ARIMA) and Vector AutoRegression (VAR) models good for forecasting both univariate and multivariate time series, respectively.

III. Methodology

The following sections detail how I calculate my initial smoke estimate and how I use that to investigate a link to the respiratory health indicators.

III.I Wildfire Smoke Estimate

It is well-known that smoke from thousands of miles away can impact a city's air quality. [CITE] So, for my analysis, I filtered US fire data up to 1250 mi away. Additionally, I removed data before 1963 to avoid potentially bad estimates before the advent of satellite imaging. Note that the fires were filtered by their minimum distance as calculated by the geodesic distance between the GPS coordinates of the city center and the closest point on the perimeter of a wildfire. Specifically, the closest point was estimated by computing geodesic distances to all points on a wildfire perimeter before choosing the smallest value. From these successive states of filtering, I arrive at a final subset of 75,819 wildfires from an initial set of 135,601 US wildfires.

Typically, smoke is dependent on wind patterns over several days, the intensity of the fire, its duration, and the distance from the city. However, for this project, I only have access to the fire area & distance. Additionally, I can distinguish between the type of fire as a proxy for fire intensity: prescribed fires and true wildfires. Prescribed burns are conducted on days where weather conditions are optimal as a way to mitigate safety risks and the spread of smoke.[18] As such, prescribed fires can be assumed to contribute less to the smoke drifting over nearby cities than wildfires. It is also intuitive that larger fires near the city will contribute more to smoke quantity over a city than small fires further away. However, how much does factor – area & distance – contribute to the overall quantity?

Smoke is generated from incomplete combustion, denoted by the formula, ***Fuel + O_2 → CO_2 + H_2O + byproducts***.[19] This tells us that smoke is linearly proportional to the amount of fuel burned, which in turn is linearly proportional to the area burned. Meanwhile, the intensity of energy, force, or flux evenly radiated from a source follows an inverse-square law with distance as commonly observed with light.[20] Smoke can be modeled as flux originating from the fire and evenly radiating outward from the burned area. So, the smoke estimate of the city can be an inverse square of the distance between the city and the fire.

So, I calculate each fire's smoke contribution based on its size, distance, and fire type. I do this to account for the fact that there is more fuel in a larger area, that smoke diffuses over distance, and that the properties of a fire change based on whether or not it's controlled. Specifically, my final estimate for the smoke effect of a single fire is as follows:

$$s = \beta_0 + \beta_1 \frac{a}{d^2} + \beta_2 t ,$$

where ***s*** is the smoke experienced by the city due to a single fire, ***a*** is the area of a fire, ***d*** is the minimum geodesic distance from the city center to the closest boundary point of the fire, and $\boldsymbol{t}$ is an ordinal encoding of fire type ranging from known prescribed fires to known wildfires. Additionally, I set $\boldsymbol{\beta = [\beta_0, \beta_1, \beta_2]}$ as a vector denoting the set of tunable weights where $\boldsymbol{\beta_0}$ is the baseline amount of smoke present not attributed to wildfires, $\boldsymbol{\beta_1}$ is the tunable fire-dispersal weight, and $\boldsymbol{\beta_2}$ articulates the difference in baselines between the different fire types. Ideally, if I know the levels of the known quantity I aim to model, I can tune the weights further. Additionally, though area and distance might have different imperial units, I can ignore the conversion as this is something that can be tuned with $\boldsymbol{\beta}$. Currently, I use $\boldsymbol{\beta}$ values of [0, 1, 1].

To get an annual estimate, I sum the total smoke effects of all fires in a given year and divide by the number of days in the fire season. Summing the values approximates the total smoke intake by the city throughout the fire season. Dividing this by the number of days (184) in the fire season (May 1st through October 31st) could in turn approximate the average daily smoke quality for the city during the fire season. Note that this simplified average assumes that each fire's duration and inception were equal which is a fundamentally erroneous assumption but might make for a decent estimate.

I validate my smoke estimates by comparing them with the daily average AQI data during the fire season. Specifically, monitoring stations in Deschutes County – all of which are less than 17 miles from Redmond's city center – were accessed to gather daily AQIs for every day in the fire season for both gaseous and particulate pollutants. The max AQI value for each day was chosen before an average across the fire season was computed to yield the final values. I chose the max because I don't have a good understanding of the exact proportion of each pollutant's contribution to smoke & that the overall reported AQI is the maximum value of the AQI for each subcategory, I'll make my estimate to be the highest AQI for any given day from any of the five stations near the city.

Once I validate my smoke estimate, I then use an AutoRegressive Integrated Moving Average (ARIMA) model for forecasting as it is the industry standard for time-series forecasting.[10] Specifically, I use Brendan Artley's code provided under the MIT License for an automated model fitting with optimal hyperparameters $\boldsymbol{p}$, $\boldsymbol{d}$, & $\boldsymbol{q}$ using the `auto_arima()` function within the `pmdarima` library.[17] I do this because I have limited experience with time series forecasting and this seemed like an effective tutorial for learning.

The parameter $\boldsymbol{p}$ denotes the order of the autoregressive model, $\boldsymbol{d}$ denotes the degree of differencing, and $\boldsymbol{q}$ denotes the order of the moving-average model. The function picks the optimal $\boldsymbol{d}$ values using the Augmented Dickey-Fuller Test to check for stationarity and uses Bayesian Information Criteria (BIC), a measure punishing the flexibility of a model to prevent overfitting, to select optimal values of $\boldsymbol{p}$ and $\boldsymbol{q}$. This model was chosen because it provides an in-built diagnostic plot like ordinary least squares while also being known for being more accurate. Additionally, it comes with the capability to show model uncertainty out of the box – indicated by the default 95% confidence intervals for model predictions.

III.II Linking Respiratory Health Indicators & Redmond's Smoke Index

I have access to respiratory health indicators tracking the rates of mortality, incidence, and prevalence of asthma, COPD, and all chronic respiratory illness diseases at the state level as rates per 100,000 values. Additionally, I have access to overall age-adjusted rates of hospitalizations for asthma and COPD at the county level provided as rates per 10,000. So, to begin, I scale all the rates to rates per 10,000 to allow comparison between metrics. Additionally, I join these various datasets with the wildfire data using Year as the primary key.

To establish potential links between these metrics and my annual smoke index, I look at the pairwise correlations between each measure. Specifically, I look at both the Pearson correlation coefficients and the more robust, Spearman correlation coefficients. The Pearson correlation coefficient tests rely on the assumption that the data is normally distributed without outliers and that there is a linear relationship between the two indicators.[15] Meanwhile, the Spearman correlation coefficient relaxes these assumptions, allowing for a more robust test.[16] For both tests, I use a significance level of 0.05.

To report values, I bin the correlation coefficients according to the following scale of the absolute coefficient: 0.0 – 0.3: Weak, 0.3–0.5: Moderate, 0.5–0.7: Strong, 0.7–1.0: Very Strong.

III.III Forecasting Respiratory Health Indicators

To forecast the respiratory health indicators, I use a Vector Autoregressive Moving Average with eXogenous regressors model (VARMAX) to model multivariate time series using my smoke forecast and time as the exogenous variables. I use a multivariate approach as many of the health indicators are understandably strongly linked. Unlike ARIMA, VARMAX allows us to model the effects of one of these indicators on other variables. Specifically, I forecast the future for sets of variables at a time as the model cannot handle more than five endogenous variables at a time due to the small number of data points. To forecast asthma rates, I consider the number of deaths, incidence, and prevalence at the state level and the overall age-adjusted hospitalization rate & crude hospitalization rates for those above 65 at the county level at once. I use the same corresponding variables to forecast COPD. Additionally, I also forecast the overall mortality, incidence, and prevalence rates for all chronic respiratory illnesses at the state level.

When building the model, I split the data into 80% train and 20% test and conducted a grid search for the best hyperparameter values of ***p*** & ***q*** using the Euclidean norm of the Root Mean Squared Error to track model fit. I don't use BIC as it is only valid when n is greater than the number of parameters in the model; in this case, the model has many parameters and very few data points. I take the Euclidean norm to ensure that the model predictions are valid for all endogenous variables and not just one at the expense of another.

III.IV Ethical Considerations

When designing my study, I remained cognizant of the fact that I was using protected health records of citizens and did not attempt to identify specific individuals or find information not readily available from a public, curated data source compiled with the legal consent of all individuals involved. As a result, I ensured all my data came from the US government with primary sources such as the Centers for Disease Control and the US Census Bureau. Both these organizations have devoted teams to ensure that all data is held accountable to their privacy and ethical standards. I also judiciously follow all the license agreements in place for the use of the data – my analysis is non-profit and open source. Additionally, my work is readily reproducible at my Git repository with a designated commit history to show where changes were made.

IV. Findings

IV.I Wildfire Patterns

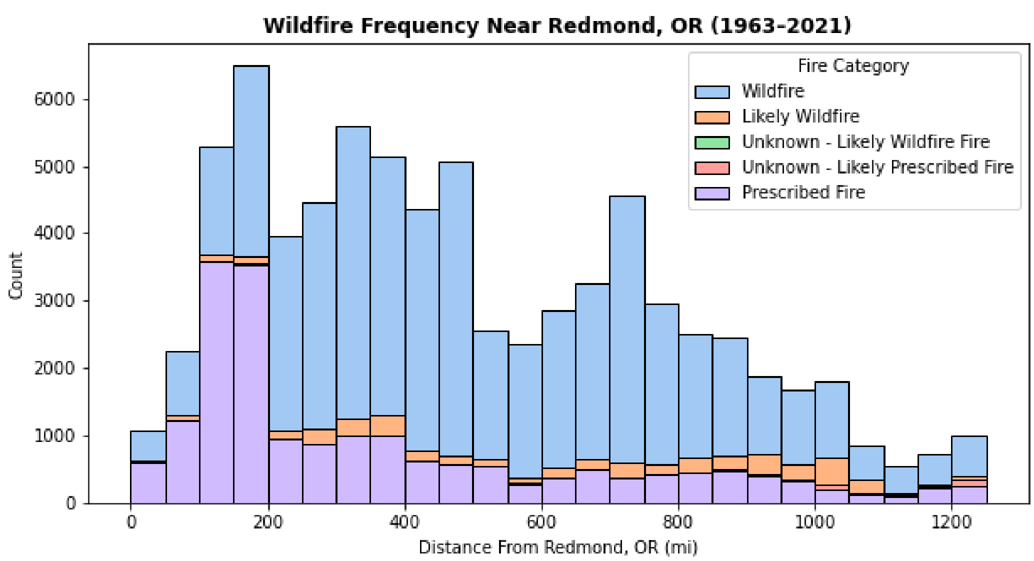

Figure 2: Wildfire Frequency Near Redmond, OR

This stacked histogram plot shows the number of fires occurring in the United States at every 50-mile increments up to 1,250 miles from Redmond, Oregon, between 1963 and 2021. The y-axis represents the wildfire count in each 50-mile bin while the x-axis denotes the minimum geodesic distance of fires from Redmond, OR. The histograms are stacked based on designated wildfire types. In total, the chart tracks 75,819 total fires, of which 52,305 are categorized as 'Wildfire', 18,995 as 'Prescribed Fire', 3,978 as 'Likely Wildfire', 451 as 'Unknown – Likely Prescribed Fire', and 90 as 'Unknown – Likely Wildfire'.

We see from **Figure 2** that the majority of the fires used for Redmond's smoke estimates lie between 100 and 500 miles away from the city center. Prescribed fires appear most frequently around 100–200 miles away before sharply decreasing in frequency as distance increases. Meanwhile, the frequency of true wildfires has a relatively prolonged spike between 200-500 miles, another spike around 700-800 miles, and otherwise low frequency at either end. This figure shows the vast amount of fires that can impact the air quality of Redmond. It should be noted that a potential reason for the decreasing number of fires further from Redmond can be attributed to the fact that international fires aren't accounted for in the data and a large portion of Canada lies within the 1250-mile range of the city center.

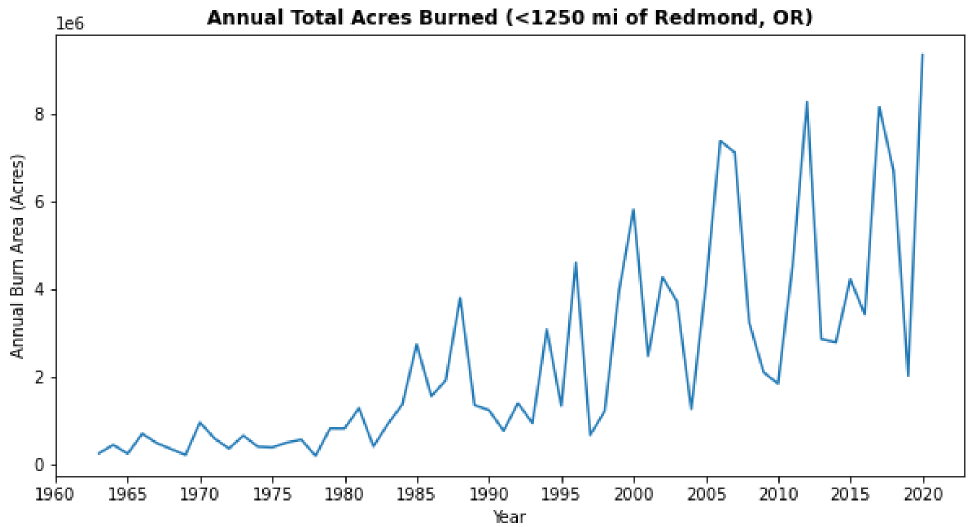

Figure 3: **Annual Cumulative Acres Burned by Wildfires**

This time series plot shows the annual cumulative acres burned by wildfires occurring in the United States within 1,250 miles of Redmond, Oregon between 1963 and 2021. The y-axis represents the total acres burned each year in millions of acres while the x-axis denotes the time in years.

The graph shows that the wildfire damage generally increased over time with a periodicity of 5 to 7 years for the good and bad years of wildfire damage near Redmond, OR. There are also significant peaks in 2020, 2017–18, 2012, and 2006–07. These peaks can be linked to the 2020 Bay Area Fire, the 2018 Camp Fire, and the 2017 Tubbs Fire, three of the worst wildfires in US recorded history.[14] It should also be noted that a potential reason for the sharp increase in burned acres per year could be that more fires are simply observed and recorded due to advances in technology & a rise in the public interest. However, we can presume from the large oscillations in data beginning in 1985 that subsequent fire data is more reliably tracked.

We create my estimate of Redmond's annual smoke index from these fires and validate it using the AQI data (**Figure 4**). We noticed that these values were not an exact match; the Spearman rank coefficient test shows that the correlation is not significant though the Pearson correlation coefficient is approximately 0.43. However, we see visually that the peaks and valleys of both the AQI and smoke estimates match and that both have values within the ranges of 20–60 for years after 1985 except for the outlying smoke index value for 2020.

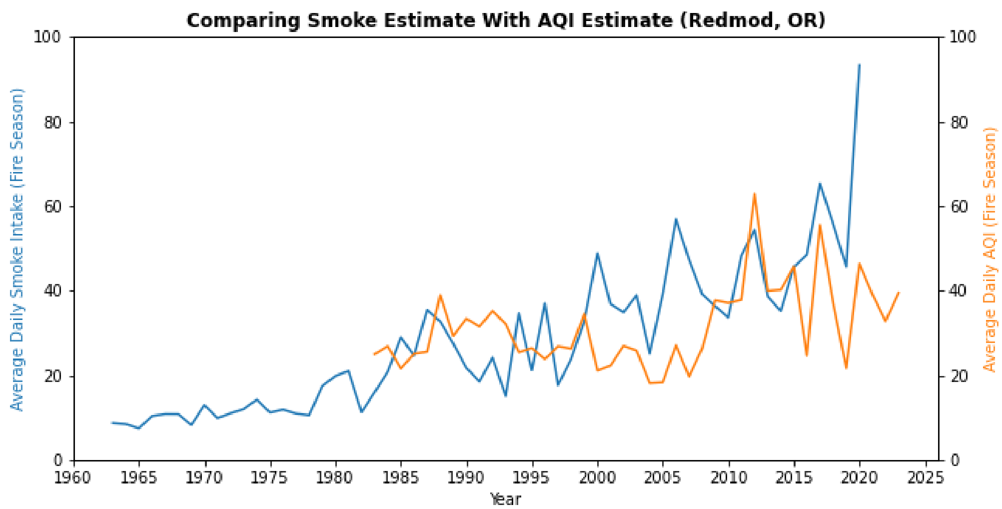

Figure 4: A Comparison of Custom Smoke Estimate & Air Quality Index

This chart compares the average custom daily smoke estimate (blue) with the average daily Air Quality Index (orange) during the fire season, May 1st - October 31st, for each year that either index is available for Redmond, OR. The average daily smoke intake metric was generated solely from the filtered subset of 75,819 wildfires between 1963 and 2021 near Redmond, OR, from an initial data set of 135,061 US wildfires aggregated by the US Geological Survey. AQI data was only available from 1983 to 2023.

Thus, I can be reasonably confident about the smoke estimate and proceed with forecasting ahead until 2050 (**Figure 5a**). From the ARIMA forecast, we see that the annual smoke levels for Redmond are expected to steadily increase with minor oscillations for seasonality. The 95% confidence interval, depicted as a gray range, shows that there is a great deal of uncertainty for the forecast. The predicted rates for 2050 could be as low as 60, signifying a plateau, and as high as 140, pointing to a sharp increase! As such, we must be prepared for either of these eventualities.

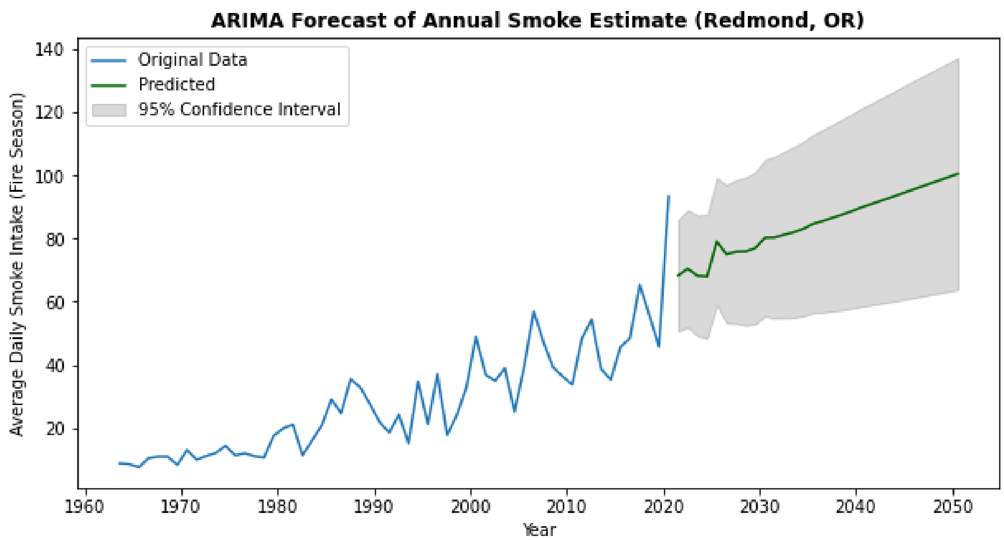

Figure 5a: Forecast of Annual Smoke Index for Redmond, OR

This chart details the historical increase in the estimated annual smoke index for Redmond, OR (blue) as well as the projected continued increase for the next 30 years (green). Additionally, the forecast has a 95% confidence interval (grey).

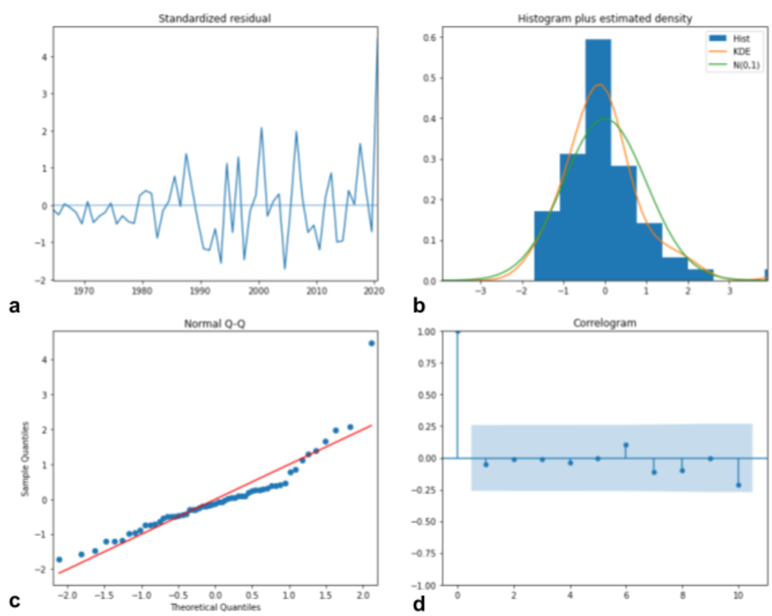

Figure 5b: Diagnosis Plot of Fire Forecast

*This chart details the diagnostic plot for the ARIMA forecast detailing that the residuals don't have a consistent pattern except for an outlying factor in 2020 (**a**) and the model doesn't have any self-correlation with various lags (**d**). Still, they aren't completely normally distributed (**b, c**) so the forecast should be taken with a grain of salt.*

IV.II Respiratory Health Indicator Patterns

If we look at the respiratory health indicators at the county level, we see that generally, the county's hospitalization rates for asthma and COPD seem to decrease over time (**Figure 6a–b**). However, it is important to consider that the hospitalization rates of more susceptible groups – those above 65+ – show greater variation over the years that may be linked to rising smoke levels. From **Figure 6c** we can see that the number of people dying before the age of 75 has dropped steadily. This likely points to the fact that there are other factors at play like advancements in medicine & better infrastructure leading to longer lives for citizens. It should be noted that **Figure 6a–b** only tracks hospitalizations when asthma or COPD is the primary diagnosis. If either condition increased individual susceptibility to other diseases like COVID-19, this isn't tracked but would be useful to know as the eventual hospitalization is indirectly caused by asthma/COPD.

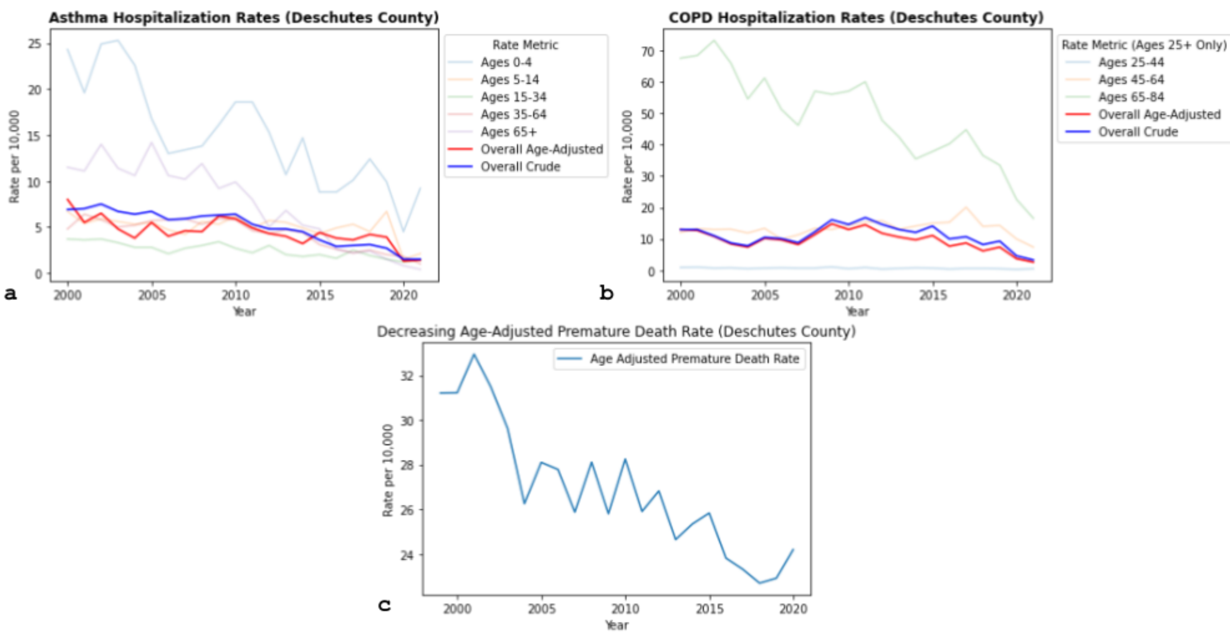

Figure 6: County-Level Hospitalization Rates

*This figure tracks the county hospitalization rates due to asthma (**a**) and COPD (**b**) as well as the premature death rate (**c**) from 2000 to 2020 In both **a** & **b**, the overall age-adjusted rates are shown in red while the overall crude rates are shown in blue. The other age-specific lines are shown faintly. the county hospitalization rates for asthma (**a**), county hospitalization rates for COPD (**b**), and the county age-adjusted premature*

Similarly, if we look at the state–level metrics, we see that the overall mortality rates for COPD and all chronic respiratory illnesses have steadily increased over time (**Figure 7a**). This indicates that advances in medicine haven't been able to effectively treat COPD as of yet and COPD makes up a vast majority of chronic respiratory illness deaths. These alarming estimates should be evaluated cautiously as the 95% confidence intervals are fairly large and both rates might be stable (this falls within the interval). Meanwhile, the already low mortality rate for asthma has dropped steeply in line with the decrease in overall premature mortality rates.

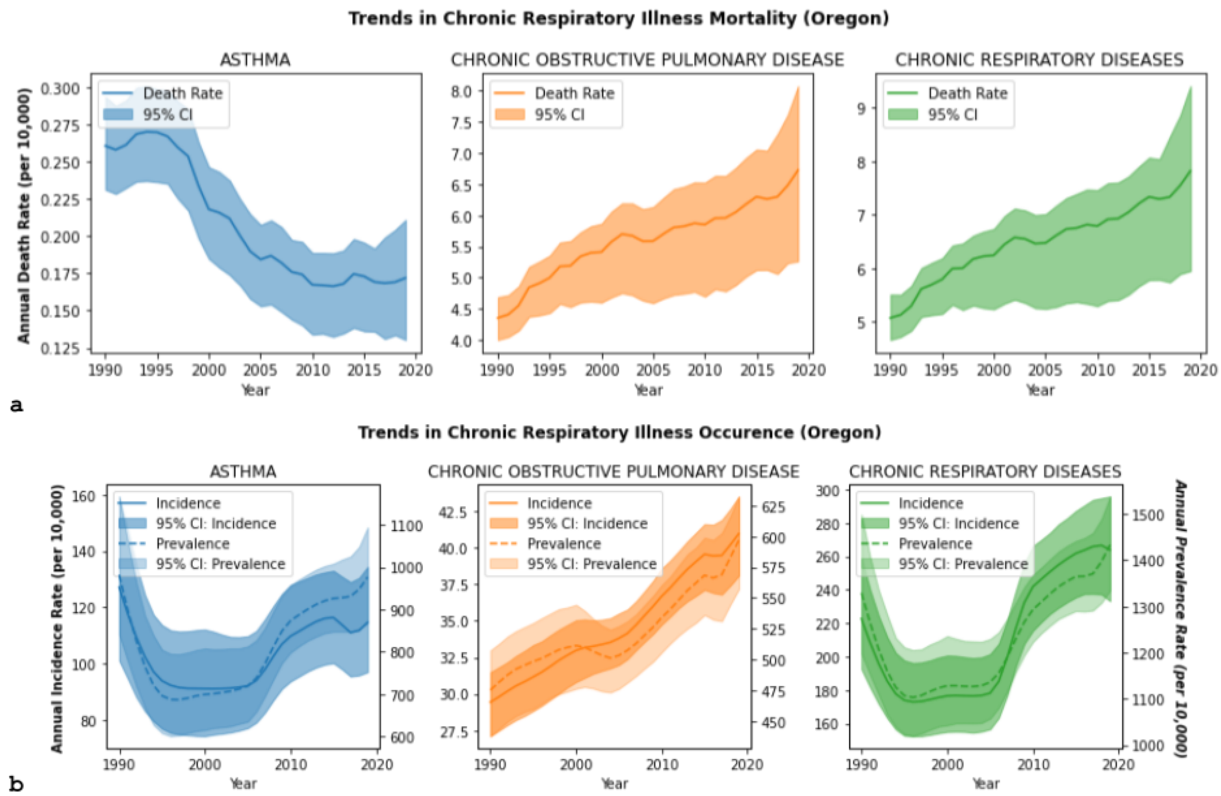

Figure 7: State-Level Disease Occurrence and Mortality Rates

*This figure shows the rates of mortality (**a**) and incidence and prevalence (**b**) for asthma (blue), COPD (orange), and all chronic respiratory illnesses (green) from 1990 – 2020 in Oregon. All rates are reported at rates per 10,000. Additionally, each metric contains a confidence interval tracking the uncertainty of the estimate provided by the IHME. In **b**, the rates of incidence (solid line) and prevalence (dotted line) are overlayed on dual-axis plots to highlight the similarities between the two.*

On the other hand, in **Figure 7b**, we notice that both the incidence and prevalence rates closely mirror one another for asthma, COPD, and all chronic respiratory illnesses. Going forward, I consider both variables in tandem. Expectedly, the prevalence and incidence rates of all chronic respiratory illnesses closely mirror that of asthma as new asthma cases make up more than 50% of new chronic respiratory illnesses and new asthma cases make up nearly 80% of existing chronic respiratory illness cases. From this chart, we notice that COPD occurrence has steadily increased from 1990 to present-day while asthma occurrence dropped steeply from 1990 to 2000 before increasing in 2020 back to levels comparable to 1990.

IV.III Tenable Links

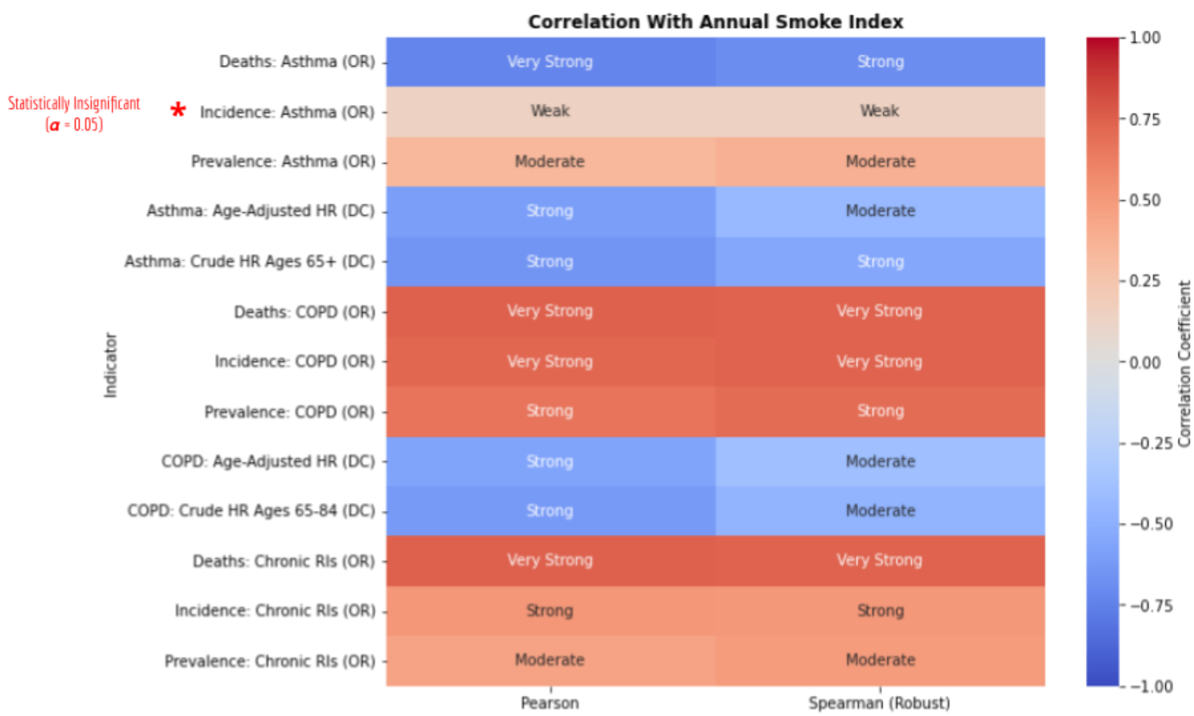

Figure 8: Correlation Tests For Respiratory Health Indicators With Annual Smoke Index

This heat plot shows the pairwise correlation with the annual smoke index of Redmond, OR for various respiratory health indicators depicted on the y-axis for both Pearson and Spearman correlation tests (x-axis). Each cell contains the value on the strength of the correlation while the coloring follows a diverging colormap denoting the true correlation coefficient. "OR" in the indicators denotes that it is a state-wide metric for Oregon while "DC" denotes that it is a county-level metric for Deschutes County. Note that only the correlation between state-wide asthma incidence rates and the annual smoke index of Redmond is not statistically significant at a 0.05 significance level.

From the correlation heat map (**Figure 8**), it is easy to see that nearly all respiratory health indicators have significant correlations with the annual smoke index for Redmond, OR. The only exception is the state-wide asthma incidence rates. Interestingly, I find that all correlations aren't positive; the asthma mortality rate and both asthma and COPD hospitalization rates are negatively correlated! This goes against what I expected and is likely a result of other confounding variables such as advancements in medical infrastructure and cultural awareness.

IV.IV Respiratory Health Forecasts

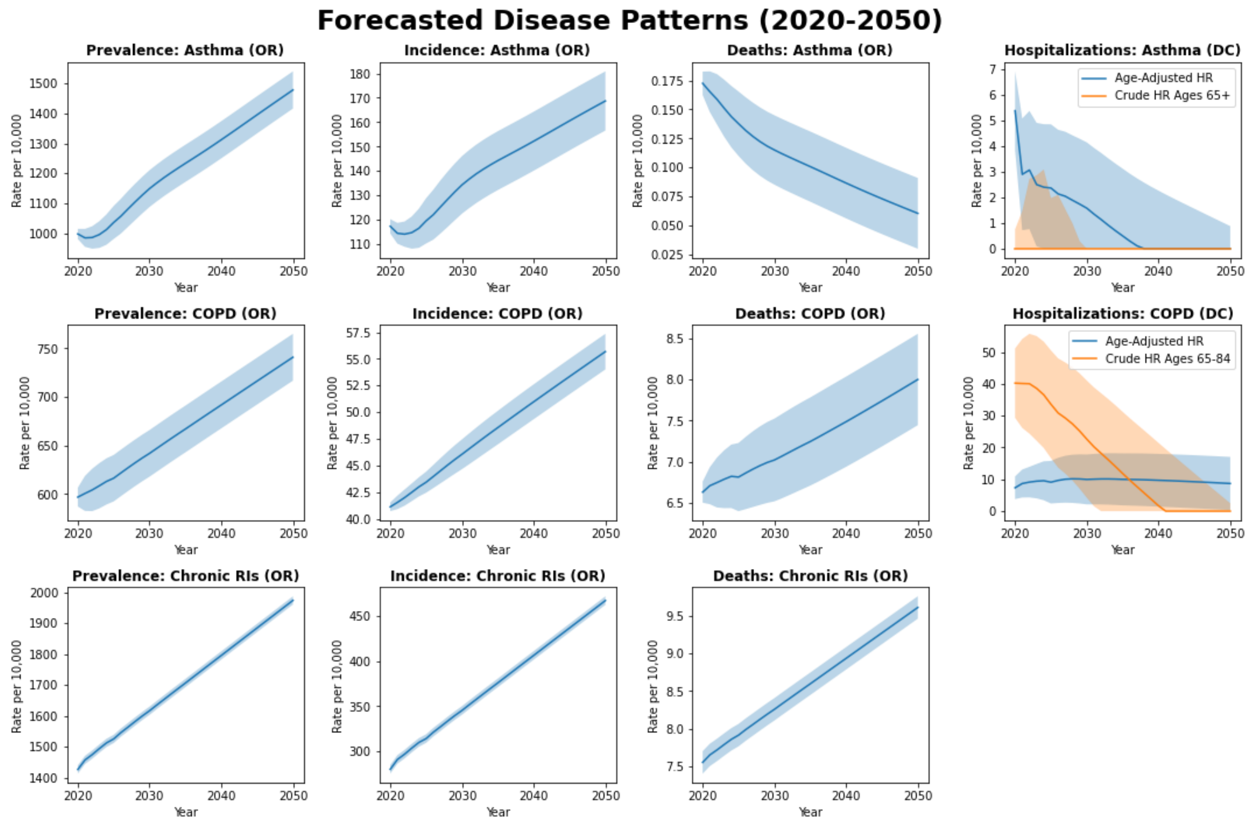

Figure 9: Forecasted Rates for Respiratory Health Indicators

This figure shows the forecasted rates of prevalence, incidence, mortality, and hospitalizations (columns 1-4, respectively) for asthma, COPD, and all chronic respiratory illnesses (rows 1-3, respectively). All rates are reported as the rate per 10,000. Only the hospitalization rates are at the county level (DC) while all others are at the state level (OR). All metrics have the predictions in solid lines while a semi-transparent region denotes the 95% confidence interval of the predictions. All metrics in blue denote the rates for all people in the region while the plots for hospitalization rates have an additional series in orange denoting the crude rates for those above 65.

From the VARMAX forecast of each respiratory health indicator, we note that all metrics are expected to increase significantly, except the asthma mortality rate and both asthma and COPD hospitalization rates. Interestingly, the forecast doesn't show much oscillation from the expected seasonality of the data. I expect this is due to the low data resolution – only 20 annual data points of each metric from 2000 to 2020 were used in the model. Additionally, the model only uses the estimated means for each year of each metric – as shown earlier, some of these estimates have incredibly large confidence intervals. As such, the forecast must again be taken with a grain of salt.

V. Discussion & Implications

My analysis points to an urgent need for proactive measures to address the growing intersection of Redmond's population expansion and the escalating risks of wildfires, as indicated in **Figure 1** and **Figure 5a**. The validated link between wildfire smoke and respiratory health (**Figure 8**) underscores the potential for a substantial strain on healthcare systems, threatening individual well-being and the overall public health landscape. In fact, in **Figure 9**, we see that the overall rates of respiratory illness occurrence are projected to increase substantially over the next 30 years. This is particularly critical given the projected increase in healthcare costs and the potential for economic challenges that may also loom ahead as a growing city.

The city council must take immediate action and prioritize more frequent tracking of health indicators to gain a nuanced understanding of the short and long-term healthcare impacts associated with wildfires. This enhanced data collection will enable the formulation of a comprehensive and effective healthcare response strategy. My current analysis was unable to find growing hospitalization numbers but the low granularity of the data doesn't allow us to find the signal. Instead, I recommend the city council capitalize on an existing Oregon Facilities Use ESSENCE for Syndromic Surveillance program that already tracks hospitalization rates and daily AQI in Oregon. Redmond already has several AQI monitoring stations and all that is required is to liaise with Carol A. Trenga, MS, PhD at the Oregon Health Authority to set up a dashboard for the city of Redmond.

Furthermore, the city council should consider implementing measures to raise awareness among residents about the potential health risks of wildfire smoke. Oregon already has an excellent flyer that can be readily mailed to citizens. Additionally, I recommend that the city provide citizens with suitable masks and kits well in advance of the wildfire season as a way to mitigate potential health triggers. N-95 masks currently go for less than $70 for a 1000 so it would be fairly cheap to provide each of the 30,000 residents with masks in times of need. Should supplies be scarce, I would recommend sending these supplies to sensitive groups such as individuals with existing respiratory illnesses. However, this may require further deliberation to ensure the privacy of citizens. The state of Oregon also provides air filtration systems for at-risk populations. It would be fairly easy for the council to direct citizens to the request form.

Considering the urgency of these issues, the city council, along with relevant stakeholders, should work swiftly to establish and communicate these measures to the public. Given the potential for wildfires and their associated health impacts, a concrete plan should be developed within the next few months to ensure that the community is adequately prepared for the upcoming wildfire season beginning in May.

After 2-3 years of monitoring the daily hospitalization cases, the council should reconvene to update their approach and consider more expensive solutions.

V.I Reflection

In the interest of the FATE principles – fairness, accountability, transparency, and explainability – I don't include any personally identifiable information (PII) information such as race, sex, or age in the model. Additionally, I picked open-source, well-documented, models like ARIMA and VARMAX that have been readily used in the industry for decades and have transparent and explainable components. Arguably, using ordinary least squares regression models would've been more explainable but because ARIMA and VAR are built off of base regression models, they don't sacrifice much explainability while increasing the model performance. Additionally, my approach has processes in place to easily update predictions with new data; I modularized the smoke estimate away from the health indicators to easily update it with new information as needed. Similarly, any new health indicators can be easily joined with the smoke estimates to improve the model.

In the interest of comprehensibility to a layman audience, I generated figures – except **Figure 5b** – that are readily understandable by all viewers with a high school education. Specifically, in **Figure 7**, I make sure to bin the correlation coefficients into ordinal values ranging from "weak" to "very strong" to allow users to take actionable insights without getting bogged down by numbers. Meanwhile, **Figure 5b** is geared towards other data scientists to highlight potential limitations in my approach.

VI. Limitations

My analysis is severely limited in scope due to the inherent complexity of the task and the limitations in resources & time. To begin, I must contend with the fact that there are a series of confounding factors that all contribute to the trends in asthma & COPD prevalence, incidence, and hospitalizations. Advancements in medical care, medical accessibility, and changes in societal behavior concerning exercise, nutrition, and smoking all play a part in impacting the trends of asthma, COPD, and other chronic respiratory diseases. These are just a few of the external factors – it becomes increasingly complex trying to account for all of them.

Moreover, the reliance on historical data introduces a dependency on the consistency and reliability of records, which may be subject to variations in reporting practices, data collection methodologies, and technological advancements over time. Incomplete or inconsistent datasets may limit the depth and reliability of my findings, presenting another potential hurdle in achieving comprehensive insights. In fact, the medical coding for reporting the primary cause of hospitalization changed on October 1st, 2015 for both Asthma & COPD in Oregon's county-level health tracker. This can potentially confound the seeming decrease in hospitalization rates for both diseases after 2015. Additionally, the set of medical diagnosis codes attributed to each disease differs between Oregon's Tracking Data Explorer and the IHME dataset – thereby impacting the validity of drawing comparisons between the two datasets.

Typically, smoke is dependent on wind patterns over several days, the intensity of the fire, its duration, and the distance from the city. However, for the sake of this assignment, I only have access to the fire area, distance, and type. So, my smoke model is fundamentally flawed. My initial smoke estimate was also generated without being refined with a well-known target variable. AQI can be used to proxy the smoke estimate for a day but can't be the end-all-be-all for validating a smoke estimate from distance, area, and fire type – which in and of itself cannot model the inherent complexity of smoke drift from 1000s of miles away. With a flawed baseline, any subsequent comparisons should be taken with a grain of salt. Furthermore, the data set only comprises US national wildfires yet portions of Canada and Mexico lie within the designated 1250 mi of Redmond, OR. These international wildfires most definitely contribute to Redmond's smoke levels and should be accounted for.

I also only have access to fires during the fire season (May 1st through October 31st) without any information about the duration of each fire. As such, I'm forced to aggregate each fire's smoke estimate and average each contribution by the number of days in a fire season. This average assumes that each fire's duration and inception were equal which is a fundamentally erroneous assumption and should be corrected in the future.

Lastly, my analysis stems from an observational study of historical data. At most, I can present potential associations between predictor and outcome variables but I cannot suppose causation from my analysis. Though there is a significant association between the annual smoke index and the health indicators, I can't judge the validity of the effect size without considering all the uncertainties of estimating the metrics themselves. A thorough, valid, deep statistical analysis of the task at hand lies beyond the scope of a month-long school project. Instead, I recommend that the city council hire knowledgeable statisticians with domain expertise to follow up on my analysis.

Additionally, I face limitations in my use of ARIMA and VARMAX to forecast annual smoke index and respiratory health indicators, respectively. Specifically, the size of my data set is extremely small – only 20 annual data points from 2000 to 2020 for the VARMAX forecasting. As such, any forecasts are highly susceptible to outliers like those in 2020. Some may even be influenced by poor data collection during the global pandemic. Additionally, with low-resolution data sampling (1 per year), it's hard to capture the anomalies presented from wildfires – these often show themselves at the daily or weekly level. Furthermore, VARMAX makes an underlying assumption of the stationarity of the data. This condition likely isn't met and any tests used for the data can't be trusted as the data is simply too small to reasonably check.

VII. Conclusion

This project focused on the healthcare impact of wildfires on Redmond, OR, Deschutes County, and the greater Oregon area. Specifically, I investigated the respiratory health of citizens by tracking the county-level premature mortality rate, asthma-related hospitalizations, and COPD-related hospitalizations from 2000–2021. Additionally, I investigated the prevalence, incidence, and mortality of chronic respiratory illnesses at the state level. By combining data from both the state & country levels, I approximated the impact on the city itself without city-specific data resolution.

I found that the rates of occurrence for asthma, COPD, and all chronic respiratory conditions have increased over the past 30 years and are expected to continue increasing significantly over the next 30 years. This follows what I expected and it is in line with my initial expectation that more smoke exposure will lead to more respiratory illnesses. Surprisingly, I see that the county's premature mortality rate and asthma/COPD hospitalization rates have steadily dropped and are projected to drop in the future. I hypothesize that this is likely due to advances in modern medicine & a better medical infrastructure in Redmond. In my analysis, I validate a link between the respiratory health of a city and the smoke effect of wildfires but can't validate the effect size.

So, I recommend that the city council more frequently track health indicators to better understand and predict both short and long-term healthcare impacts arising from wildfires. Additionally, I recommend raising awareness and providing citizens with suitable masks & kits in preparation for the wildfire season. This project is merely a starting point and I look forward to continue working with the Redmond City Council & other data scientists to create a healthier Redmond community.

VIII. References

1. Community Profile. (n.d.). *City of Redmond*. Retrieved November 17, 2023 from https://www.redmondoregon.gov/our-community/community-profile
2. Why is Wildfire Smoke a Health Concern? (n.d.). *U.S. Environmental Protection Agency.* Retrieved November 17, 2023, from https://www.epa.gov/wildfire-smoke-course/why-wildfire-smoke-health-concern
3. Health Effects Attributed to Wildfire Smoke. (n.d.). *U.S. Environmental Protection Agency*. Retrieved November 17, 2023, from https://www.epa.gov/wildfire-smoke-course/health-effects-attributed-wildfire-smoke
4. Which Populations Experience Greater Risks of Adverse Health Effects Resulting from Wildfire Smoke? (n.d.). *U.S. Environmental Protection Agency*. Retrieved November 17, 2023, from https://www.epa.gov/wildfire-smoke-course/which-populations-experience-greater-risks-adverse-health-effects-resulting
5. How to protect urban lives, health and property from wildfire. (n.d.). *C40 Knowledge Community.* Retrieved November 17, 2023, from https://www.c40knowledgehub.org/s/article/How-to-protect-urban-lives-health-and-property-from-wildfire?language=en_US
6. Centers for Disease Control and Prevention, Age-Adjusted Premature Death Rate for Deschutes County, OR [CDC20N2UAA041017], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CDC20N2UAA041017, November 16, 2023.
7. McGeehin MA, Qualters JR, Niskar AS. National Environmental Public Health Tracking Program: bridging the information gap. Environ Health Perspect. 2004;14:1409–1413.
8. Crude and Age-Adjusted Rate (n.d.). *Health.mo.gov.* Retrieved November 17, 2023, from https://health.mo.gov/data/documentation/crude-aarate.php#:~:text=Rates%20for%20specific%20age%20groups
9. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2020. Available from https://vizhub.healthdata.org/gbd-results/
10. Bajaj, A. (2021, May 11). *ARIMA & SARIMA: Real-World Time Series Forecasting*. Neptune.ai. Retrieved November 17, 2023, from https://neptune.ai/blog/arima-sarima-real-world-time-series-forecasting-guide
11. *What is your Oregon home's risk of wildfire? New statewide map can tell you*. (n.d.). Opb. https://www.opb.org/article/2022/06/30/oregon-wildfire-prevention-map-risk-forest-fire-home/
12. Welty, J.L., and Jeffries, M.I., 2021, Combined wildland fire datasets for the United States and certain territories, 1800s-Present: U.S. Geological Survey data release, https://doi.org/10.5066/P9ZXGFY3.
13. US Environmental Protection Agency. Air Quality System Data Mart [internet database] available via https://www.epa.gov/outdoor-air-quality-data. Accessed November 08, 2023.
14. Earth.Org. (2022, September 15). *15 largest wildfires in US history*. Earth.Org. https://earth.org/worst-wildfires-in-us-history/
15. Scribbr. (n.d.). Pearson Correlation Coefficient. Retrieved from https://www.scribbr.com/statistics/pearson-correlation-coefficient/
16. Scribbr. (n.d.). Correlation Coefficient: A Guide. Retrieved from https://www.scribbr.com/statistics/correlation-coefficient/#spearmans-rho
17. Towards Data Science. (n.d.). Time Series Forecasting with ARIMA, SARIMA, and SARIMAX. Retrieved from https://towardsdatascience.com/time-series-forecasting-with-arima-sarima-and-sarimax-ee61099e78f6
18. Oklahoma State University Extension. (n.d.). The Best Time of Year to Conduct Prescribed Burns. Retrieved from https://extension.okstate.edu/fact-sheets/the-best-time-of-year-to-conduct-prescribed-burns.html
19. Science Learning Hub. (n.d.). What is Smoke? Retrieved from https://www.sciencelearn.org.nz/resources/748-what-is-smoke
20. Wikipedia. (n.d.). Inverse-square law. Retrieved from https://en.wikipedia.org/wiki/Inverse-square_law

IX. Data Source Links

Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons)

EPA AQS API

Asthma Hospitalizations and Emergency Department Visits

Chronic Obstructive Pulmonary Disorder (COPD) Hospitalizations

IHME Query Tool for Oregon Respiratory Illness Incidence, Prevalence, & Mortality Rates

Age-Adjusted Death Rate Data