The Impact of Wildfires on Tulare County, CA

Data 512 – Course Project

Sagnik Ghosal

INTRODUCTION

The escalating occurrences of summer wildfires in the western United States, particularly in areas like Tulare County, CA, have prompted an urgent need for thorough analysis and proactive interventions. These fires, stemming from various factors like climate change and forestry policies, have released substantial smoke, profoundly impacting public health, tourism, property, and wider societal structures. Addressing this urgency requires predictive health planning, utilizing smoke-related respiratory health forecasts to establish preemptive strategies.

Tulare County, known for its agricultural richness, bears witness to the profound impact of wildfires on its vital farming sector [**Ref 1**]. These blazes disrupt not only residents' lives but also pose imminent threats to the agricultural landscape, endangering crop yields and livelihoods. The ripple effects of these wildfires permeate through the county's economic core, emphasizing the crucial need for proactive measures to protect lives and the region's agricultural sustenance.

Understanding the grave repercussions of smoke on Tulare County assumes a pivotal role in shaping policy decisions, guiding city planning, and effectively allocating healthcare resources. Alarming increases in hospitalizations and mortality rates underscore the immediate need for proactive measures and heightened community awareness. By translating extensive data into actionable policies, the community can be shielded from the respiratory risks induced by smoke.

This analysis doesn't solely benefit Tulare County's populace; it emerges as a guiding force for government officials and civic institutions, steering them toward implementing human-centered social and economic reforms. The insights derived from comprehensive data and visualizations wield substantial influence, capturing the attention of policymakers, city managers, and civic institutions. This, in turn, facilitates informed planning to mitigate future wildfire impacts and safeguard the lives of Tulare County residents.

Moreover, the urgency lies not only in predictive health planning but also in acknowledging the differential impact on the male and female population. Wildfire smoke, laden with particulate matter, poses varying health risks to men and women, underscoring the necessity for clear strategies tailored to combat these risks for both genders. Addressing these distinct vulnerabilities is imperative for effective mitigation and health management strategies.

BACKGROUND

Several seminal studies and articles have laid the foundation for understanding the impact of wildfire smoke on public health, particularly in regions like Tulare County, contributing significantly to the imperative and methodology of this current study. Research, such as that by Reid et al. [**Ref 2**], delineates the escalation in wildfire frequency and intensity due to climate change, resulting in heightened smoke exposure and emphasizing the urgent need for comprehensive analysis. The findings of Liu et al. underscore the intricate correlation between wildfire smoke and respiratory health outcomes, particularly highlighting the exacerbation of asthma and other respiratory conditions [**Ref 3**]. These studies, along with numerous others, have underscored the necessity for focused, data-driven research to elucidate the nuanced effects of wildfire smoke on vulnerable populations.

Moreover, studies like Masri et al. have emphasized the disproportionate impact of wildfire smoke on agricultural regions, mirroring the situation in Tulare County [**Ref 4**]. Such research elucidates not only the direct health implications but also the collateral damage to essential sectors like agriculture, aligning with the multifaceted concerns of this study beyond public health.

The influential work of Cascio in environmental health [**Ref 5**] has outlined the broader systemic impact of wildfire smoke, shedding light on its implications beyond immediate health risks. Cascio's work broadens the perspective, considering economic, social, and environmental impacts, thereby shaping the comprehensive approach adopted in this study to assess the holistic impact of wildfires on Tulare County.

Additionally, seminal research by Rappold et al. has demonstrated the acute health risks posed by wildfire smoke exposure, particularly concerning vulnerable groups such as the elderly and children [**Ref 6**]. This body of work resonates deeply with the concerns of this study, emphasizing the need for nuanced analysis to understand and mitigate the differential impacts on various demographics within Tulare County's population.

Collectively, these seminal studies have set the stage for this current investigation by delineating the intricacies of wildfire smoke exposure, its health ramifications, differential impacts on diverse demographics, and the broader societal consequences, thereby shaping the methodology and urgency of this study in addressing the multifaceted challenges faced by Tulare County.

In this research study, the following research questions and hypothesis have been raised:

1. What are the estimated smoke impacts on Tulare city for the last 60 years?
2. To what extent do certain respiratory diseases linked to smoke-related pollution (e.g., Chronic respiratory diseases, Chronic obstructive pulmonary disease, Asthma) exhibit higher mortality rates compared to diseases not directly related to smoke-related pollution (e.g., Pneumoconiosis, Coal workers pneumoconiosis, Silicosis, Asbestosis) in Tulare County?
3. How does fire smoke relate to the age-standardized mortality rate for respiratory diseases in Tulare County across different years?
4. Is there a noticeable difference in respiratory disease mortality rates between sexes due to variations in exposure to fire smoke in Tulare County?
5. How does the prevalence of smoke fires in Tulare County connect with rates of asthma-related hospitalizations across different age groups over time?

These five questions are crucial to understanding and addressing the well-being of Tulare County's residents. Firstly, exploring the estimated smoke impacts on Tulare City over the last 60 years provides historical context, enabling a comprehensive assessment of the prolonged effects of wildfires on the area. Secondly, investigating the mortality rates of specific respiratory diseases linked to smoke-related pollution compared to those not directly related allows for a targeted understanding of the health impacts, emphasizing the urgency of mitigating smoke-related illnesses. Thirdly, examining the relationship between fire smoke and age-standardized mortality rates for respiratory diseases across different years helps identify trends, aiding in proactive health planning. Additionally, assessing differences in respiratory disease mortality rates between sexes due to variations in smoke exposure sheds light on gender-specific vulnerabilities. Finally, analyzing the connection between smoke fires and asthma-related hospitalizations across various age groups over time provides insights into vulnerable demographics, guiding targeted healthcare interventions. These inquiries, rooted in the community's health concerns, serve as essential pillars for informed policymaking and proactive measures to safeguard the residents of Tulare County.

METHODOLOGY

Data Acquisition

The initial phase of our analysis revolved around sourcing appropriate datasets, ensuring they were non-copyrighted and accessible for our research purposes. It's crucial to acknowledge the limitations inherent in these datasets, which we detail further in the report's limitations section. The fire smoke estimate dataset was acquired from the USGS Wildland Fire Combined Dataset [**Ref 7**], compiled and structured by the US Geological Survey, available in GeoJSON format. For a comprehensive understanding of the health implications arising from fire smoke, we integrated two additional datasets. The first dataset, Mortalities from Respiratory Diseases in California Counties from 1980 to 2014 [**Ref 8**], and the second, Asthma Hospitalization Rates for California Counties from 2015 to 2020 [**Ref 9**], were sourced in CSV format from their respective websites. Detailed descriptions of these three datasets are outlined in the subsequent subsections.

Exploratory Data Analysis

In preparing for exploratory data analysis, several key steps were taken to ensure a robust foundation for statistical and visual examination. The initial focus centered on refining the dataset to exclusively encompass Tulare County, filtering out extraneous data from other counties using either county names or the FIPS code. Given the diverse sources of this data, a pivotal step involved consolidating these datasets to extract meaningful insights. The sole common column across all datasets was the "Year." Consequently, we merged these datasets based on the "Year" column. It's important to note that in cases where a particular year was present in one dataset but absent in another, the corresponding entry in the combined dataset was omitted for consistency and accuracy. Python 3, leveraging pandas, scipy, and matplotlib libraries, served as the primary analytical tools. Fire data was meticulously filtered based on proximity to Tulare within a 1250-mile radius and the year of occurrence, spanning from 1963 to 2023. To address inconsistencies in fire dates, an inclusive approach considered all fires within a given year rather than limiting the focus to specific calendar periods. Additionally, accounting for data anomalies, such as instances of curveRings, occurred, where the impact on overall results was deemed negligible. Through this process, 10 intermediate files were generated, offering detailed insights available on [**Ref 10**]. Notably, much of the mortality rate analysis employed age-standardization to mitigate discrepancies arising from varying age distributions among populations. This method, by adjusting for age structure disparities across populations, ensures a more precise and equitable comparison of mortality rates, transcending the potential skew introduced by varying age demographics and providing a more accurate representation of mortality rate differences across populations.

Developing Fire Estimate and ARIMA Forecasting Model

The process of creating the Fire Estimate and ARIMA Forecasting Model began with a correlation analysis, identifying crucial factors for the smoke estimation, as illustrated in Figure X. This smoke estimation involves multiple factors: fire type, fire size measured in GIS acres, and the distance from the fire source. Here's how each factor contributes:

- Fire types, like wildfires or prescribed fires, carry varying environmental impacts. Assigning weightages to these types accounts for their differing effects. For example, wildfires hold a higher weight of 1.0, while prescribed fires, controlled burns, receive a lower weight of 0.5.
- The fire's size, represented in GIS acres, significantly influences smoke production. Larger fires generate more smoke. The code adjusts for this by converting acres to square miles using a conversion factor (0.0015625).
- Distance from the fire source is critical. Smoke disperses and dilutes as it travels. Dividing the smoke impact by the distance offers a more accurate estimate based on proximity to the fire.
- To ensure visibility in plots, the smoke estimate was scaled by a factor of 10,000 to align with AQI values, though this scaling doesn't affect correlations.
- Summarizing smoke impact over time involves grouping data by year, calculating mean values for smoke impact, fire size, and distance. This consolidated data enables tracking yearly variations in smoke impact.

The formula used for the fire estimate is

Fire smoke estimate = $(Assigned\ Fire\ Type\ \ x\ GIS\ Acres\ x\ 15.625)/(Distance)$

Where, the weights for Assigned Fire Types are:

- Wildfire: 1.0,
- Unknown - Likely Wildfire: 0.9,
- Likely Wildfire: 0.9,
- Prescribed Fire: 0.5,
- Unknown - Likely Prescribed Fire: 0.4

Additionally, an ARIMA predictive model was developed for forecasting smoke estimates over the next 25 years (2024-2049). ARIMA models are standard for time series predictions. To convey prediction uncertainty, prediction intervals with a 95% confidence level are calculated. The ARIMA model parameters, represented as (p, d, q) where p is autoregressive, d is the degree of differencing, and q is the moving average, are selected via grid search to minimize the average RMSE over 5 folds.

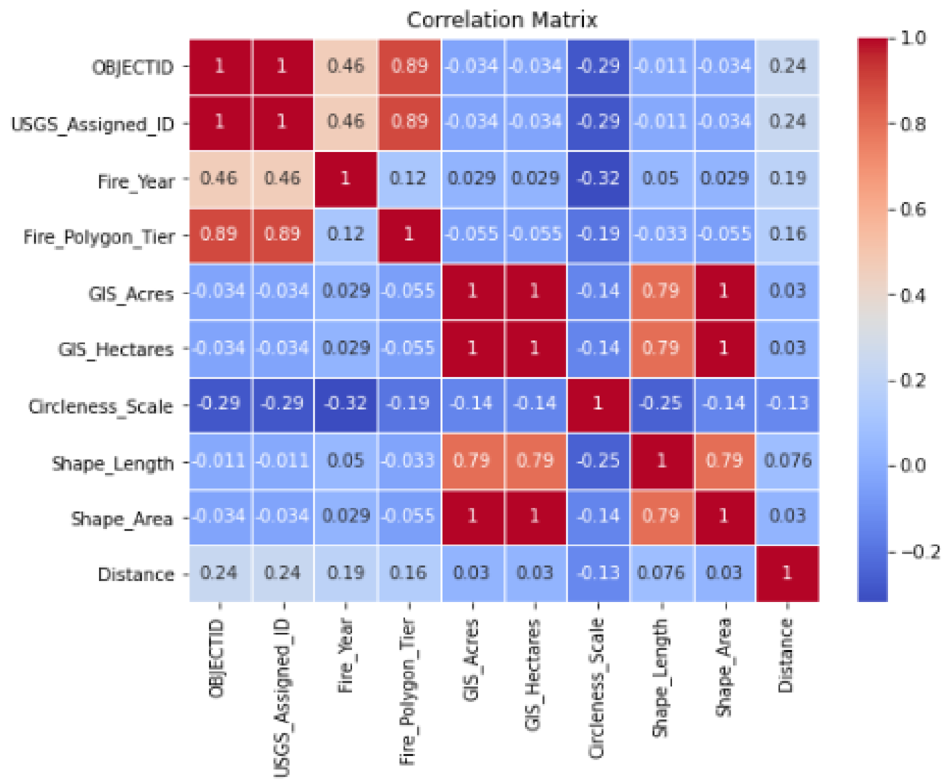

Fig 1: Heatmap representing the initial correlations

Statistical Analysis for Hypothesis Testing

The statistical analysis in this research sticks to a significance level of 0.05 (p-value). This level decides how sure we need to be before saying something is true or not. It helps us thoroughly check if there's a link between things, making sure our conclusions are strong. We used three statistical methods:

- Regression Analysis: This helps us see if there's a straight-line connection between things, like fire smoke and how many people die from respiratory problems.
- Correlation Analysis: This checks how strong the connection is between different things. It helps us understand how much one thing affects another.
- T-Test: This test checks if there's a big difference between how many men and women die from smoke-related asthma.

For the questions we looked at, here are the ideas we tested:

- Fire Smoke and Mortality Rates: We checked if smoke from fires has anything to do with how many people die from asthma or Chronic obstructive pulmonary disease (COPD). One idea was that there's no link between smoke and the number of deaths (null hypothesis). The other idea was that smoke does affect the number of deaths (alternative hypothesis).
- Gender and Asthma Deaths: We wanted to see if there's a big difference in how many men and women die from asthma caused by smoke. One thought was that there's no big difference (null hypothesis). The other idea was that there is a significant difference (alternative hypothesis).

This methodology places emphasis on interpreting statistical findings through a human-centered lens, aiming to comprehend and address real-world implications. By integrating statistical rigor with a focus on human health outcomes, this approach ensures that derived insights contribute to informed decision-making and effective policy formulations for the welfare of the community in Tulare County.

Data Visualization

In this study, a comprehensive methodology utilizing data visualization techniques, including bar plots and time series plots, has been employed. These visuals serve as the primary means to elucidate the relationship between wildfire events and health outcomes. Bar plots showcase comparative analyses between periods of wildfire occurrence and non-occurrence, displaying metrics such as hospital admissions due to respiratory issues or air quality indices during these periods. Additionally, time series plots track the temporal evolution of health indicators, such as asthma-related hospitalizations or mortality rates, juxtaposed against wildfire occurrences over time. These visualizations aim to make the findings more accessible and comprehensible to diverse audiences, facilitating a clear understanding of the impact of wildfire smoke on health in Tulare County.

Documentation for Reproducibility

To ensure human-centered data science is practiced, all the related Jupyter notebooks which can be found on [**Ref 10**] are well documented with comments and instructions. It also has relative address paths and intuitive variable names to ensure the code is reproducible and replicable. Links to all the data sources, models and literature surveys are also enlisted clearly. To help in facilitating further research, it is essential to make reproducibility easier. To avoid any copyright issues, the licenses, and terms of use of each dataset are also stated in this documentation.

FINDINGS

To effectively articulate the findings from this research project, we will be trying to answer the five questions (discussed previously) and then using visualizations, statistical analysis, we will try to find an answer to the specific question.

Question 1: What are the estimated smoke impacts on Tulare city for the last 60 years?

Solution: To answer this question, we will use three plots revealing compelling insights into the historical distribution of fires, the acres burned annually, and the correlation between fire smoke estimates and Air Quality Index (AQI) to estimate the impact smoke had on Tulare County over the last 60 years.

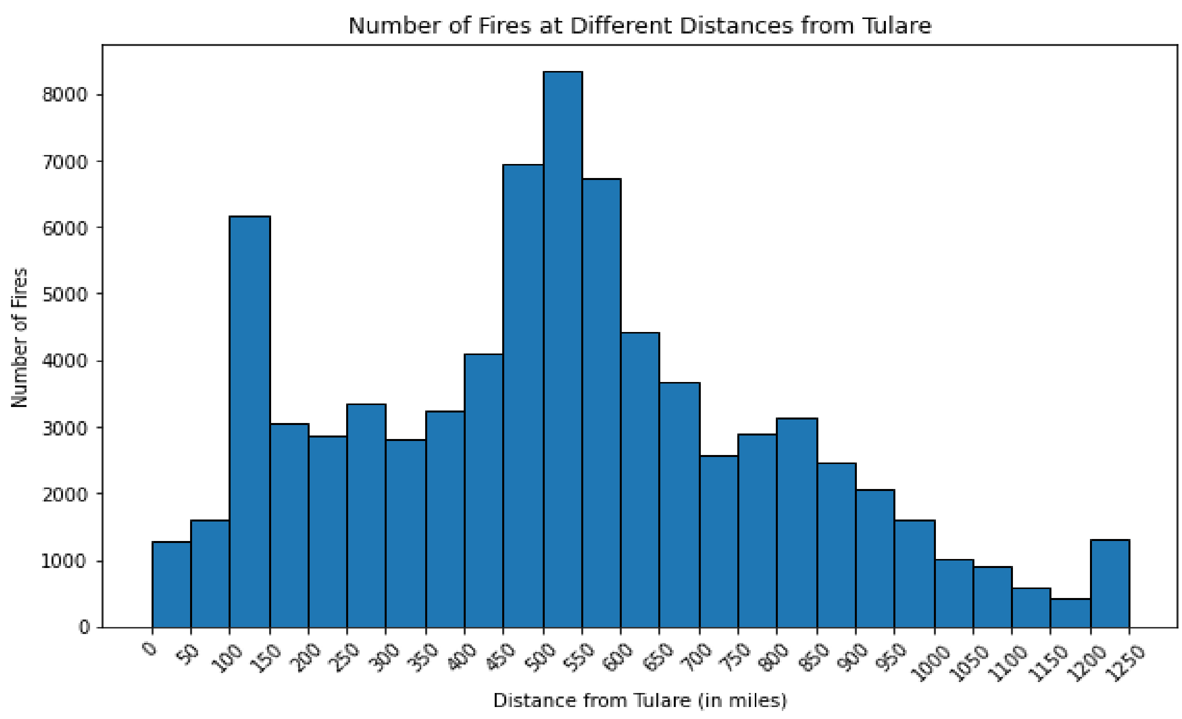

Fig 2: Plot of number of fires at different distances from Tulare County (max 1250 miles)

First, we will take a look at the Fire Distribution by Distance from Tulare. The histogram showcases the number of fires happening at different distances from Tulare, covering a span of up to 1250 miles. It provides valuable insights into the history of fires near Tulare from 1963 to 2023. On the chart, the y-axis displays the count of fires, while the x-axis measures distances from Tulare in miles. The data used here comes from the original fire dataset, focusing on fires within 1250 miles of Tulare between 1963 and 2023, using the "Distance" column to create the histogram. The bars in the histogram represent 50-mile increments, ranging from 0 to 1300 miles.

Looking at the plot, a few clear patterns emerge. The trend reveals fewer fires within the first 100 miles from Tulare. However, as the distance increases, the number of fires notably rises, especially between 100 to 450 miles, where counts range from 3000 to 4000 fires. The most significant spike happens between 450 and 600 miles from Tulare, with fires soaring to 7000 to 8000 incidents. Beyond 600 miles, the number of fires steadily declines, with fewer than 1000 fires recorded at distances of 1000 miles or more from Tulare. These patterns might be due to various reasons. The lower number of fires within the initial 100 miles could be influenced by the urban wildland interface (UWI), where fire activity is lower due to more people, better fire prevention, and quicker responses to fires. On the other hand, the peak in fires between 450 and 600 miles might be shaped by ecological factors such as vegetation, weather, and terrain. Some ecosystems within this range may be more prone to natural or lightning-caused fires, leading to more frequent incidents.

Next, let's take a look at the Annual Acres Burned in Proximity to Tulare. The time series plot shows the yearly total acres burned by fires within a 1250-mile radius from Tulare. It's a helpful tool to understand how wildfires have changed over time in this area. The x-axis shows the years, while the y-axis indicates the total acres burned each year. This clear layout makes it easy to see how wildfires relate to time. The data for this graph comes from the original fire dataset, filtered to include fires within 1250 miles of Tulare between 1963 and 2023. However, there's no data available after 2020 that meets these criteria. The dataset was organized by year, totaling the acres burned annually to create this graph. To make it clearer, the x-axis shows years in 5-year intervals.

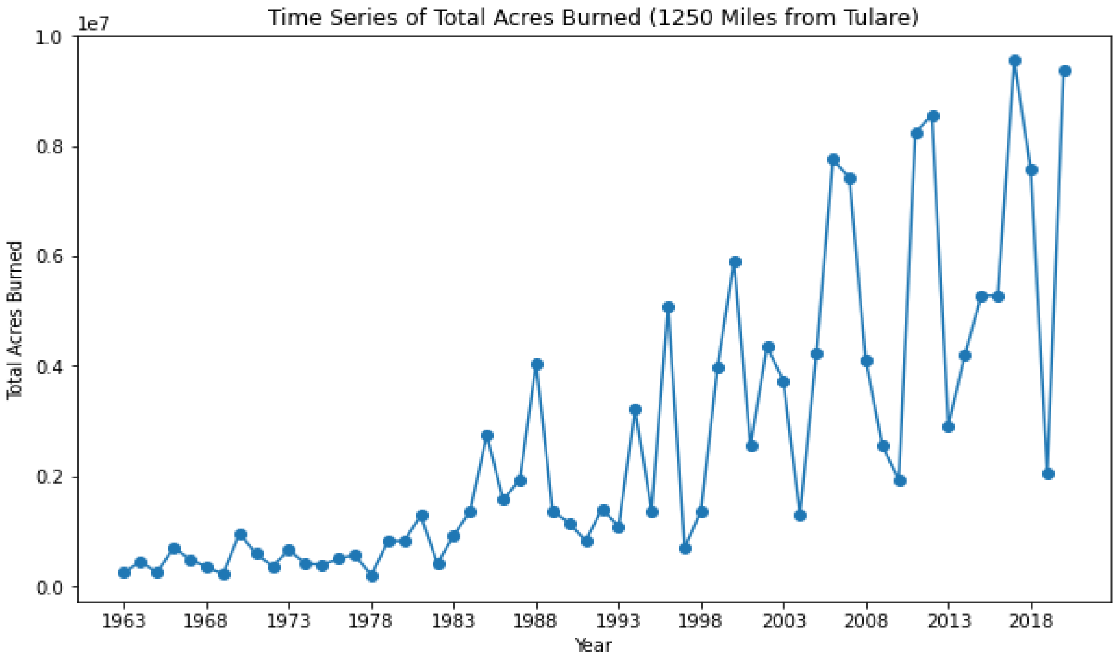

Fig 3: Time-series plot of the total acres burned (1250 miles from Tulare County)

Looking at the graph, a significant trend stands out: a sharp increase in burned acres over the years. For instance, the total acres burned in 2020 are nearly ten million times greater than those in 1963. While there are occasional fluctuations and temporary declines, the overall pattern is clear. Several reasons contribute to this concerning trend. Higher temperatures and extended periods of drought create conditions that make wildfires more likely and intense. The introduction of non-native plants can change ecosystems, making them more prone to fires. Additionally, as the population grows and urban areas expand, accidental ignitions and fires in areas where urban and wildlands meet (the wildland-urban interface) become more frequent. These factors combined contribute to the increasing impact of wildfires in the Tulare region.

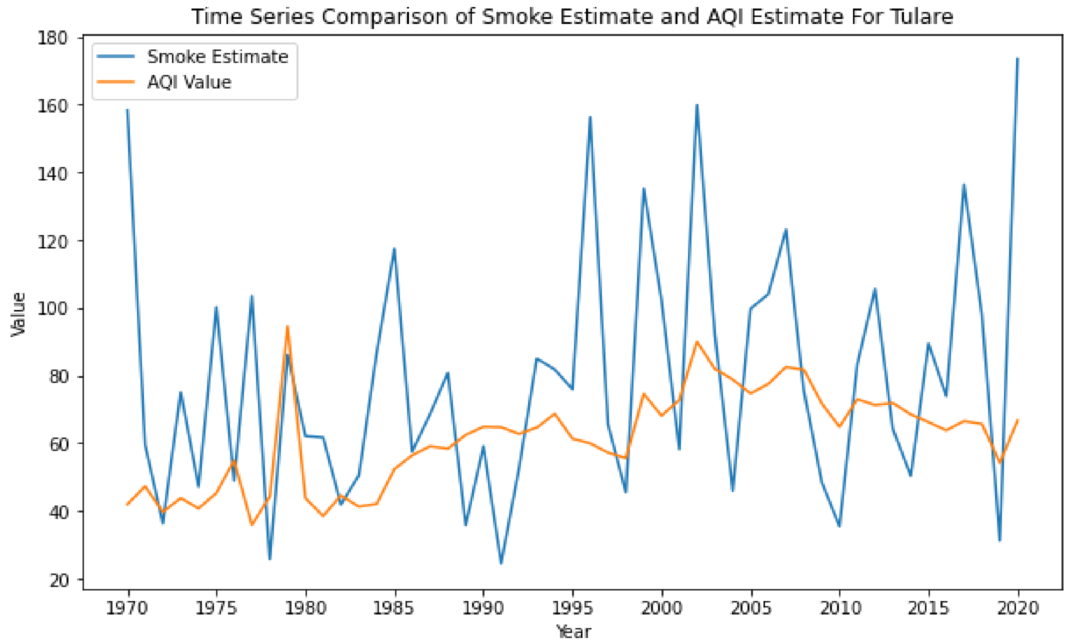

Fig 4: Time-series comparison of fire smoke estimate and the AQI estimate for Tulare County

Finally, we will take a look at Fire Smoke Estimate vs AQI in Tulare. The time-series plot shows the fire smoke estimate and AQI values over the years, with the y-axis indicating these values and the x-axis representing the years. Two lines are visible: the blue line represents the fire smoke estimate, calculated from fires within 1250 miles from Tulare between 1963 and 1970, considering factors like fire types, size, and distance from Tulare. Meanwhile, the orange line displays the AQI values for gaseous and particulate pollutants in Tulare County from 1970 to 2020, sourced from the US Environmental Protection Agency (EPA).

Two different sets of data were used for this visualization. The fire smoke estimate was derived from fires' characteristics within a specific distance from Tulare, while the AQI values were obtained from various pollutants' measurements in Tulare County. Notably, the fire smoke estimate spans from 1963 to 2020, while the EPA's AQI data covers 1970 to 2020. Thus, the graph presents data starting from 1970 for both estimates.
Observing the plot, no clear trend is visible for the fire smoke estimate. However, the AQI values demonstrate a slight upward trend, indicating a gradual increase over the years. According to AQI standards, values within 0-50 signify "good" air quality, while values above 50 fall into the "moderate" category. Additionally, there's a reported correlation value of 0.27 between the obtained AQI values and the fire estimate, suggesting a partial relationship between these variables.

Therefore, based on the three plots, we can understand how smoke from fires has affected Tulare City in the last 60 years. The first plot, showing the distribution of fires by distance from Tulare, highlighted that there were fewer fires close to the city but more between 100 and 600 miles away, especially peaking between 450 and 600 miles. This might be due to where people live and the types of plants and weather in those areas. The second plot, displaying the acres burned annually, revealed a huge increase in burned land from 1963 to 2020, likely because of hotter weather, long dry periods, new plants, and more people living in areas close to forests. Lastly, the third plot, comparing smoke estimates to air quality, didn't show a clear trend for smoke but indicated a small increase in air pollution. It suggests that the air quality has gotten a bit worse over time. These plots give us a sense of how smoke impacts Tulare City, showing changes in fire patterns, burned areas, and air quality over the years.

We've implemented a predictive model using fire data and smoke estimates for Tulare County to forecast the trend of fire smoke for the upcoming 25 years (2024-2049). Employing an ARIMA model, a widely used technique for time series prediction, we've ensured that our predictions incorporate a 95% confidence level, accounting for appropriate levels of uncertainty. In ARIMA models, the 'order' denoted as (p, d, q) determines the autoregressive (AR), differencing, and moving average (MA) components, respectively. Our approach involved a grid search across specified (p, d, q) values to minimize the average Root Mean Square Error (RMSE) over 5 folds, selecting the combination yielding the lowest error. The plot derived from this predictive model indicates a sustained or potentially worsening trajectory for fire smoke estimates in Tulare County for the future years. This forecast raises concerns about the ongoing or potential exacerbation of fire smoke impact in the region.

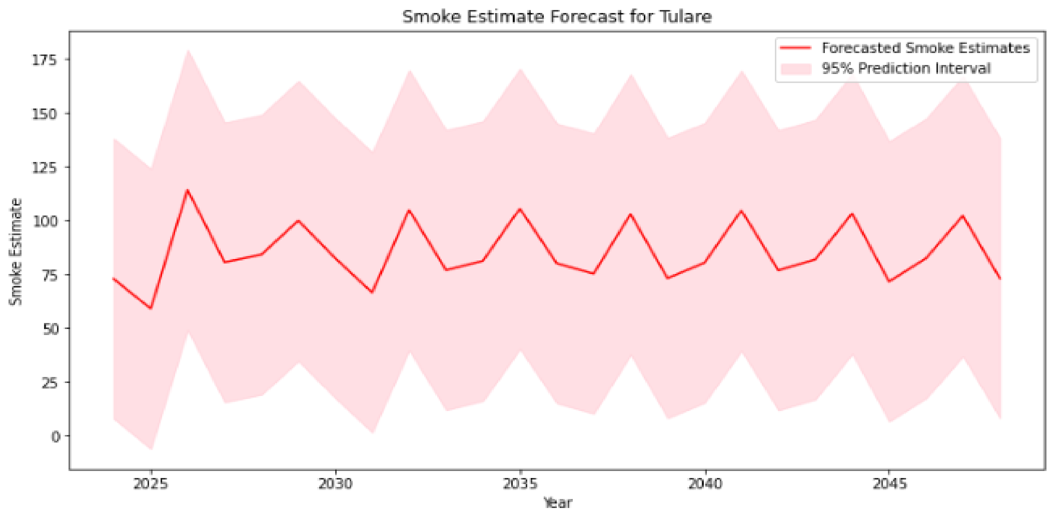

Fig 5: Fire smoke estimate forecast for Tulare County for the next 25 years

Question 2: To what extent do certain respiratory diseases linked to smoke-related pollution (e.g., Chronic respiratory diseases, Chronic obstructive pulmonary disease, Asthma) exhibit higher mortality rates compared to diseases not directly related to smoke-related pollution (e.g., Pneumoconiosis, Coal workers pneumoconiosis, Silicosis, Asbestosis) in Tulare County?

Solution: Based on [**Ref 11**], our focus on smoke pollution-related diseases centers on Asthma and COPD, while our attention shifts to Pneumoconiosis, Coal workers pneumoconiosis, Silicosis, and Asbestosis for non-smoke pollution-related diseases, as these are the only diseases represented in our dataset. Our visualization, a bar plot, illustrates the average mortality rates for smoke-pollution related diseases (in blue) and non-smoke pollution related diseases (in green). Notably, the mortality rate for non-smoke pollution related diseases stands at 0.8810, while for smoke-pollution related diseases, it drastically escalates to 26.47, indicating a staggering 30-fold difference. This stark contrast underscores the heightened fatality associated with smoke-pollution related diseases. Moreover, this supports the predictions made by the ML model, indicating an anticipated increase in smoke-related incidents over time, signifying a potential surge in deaths attributed to smoke-pollution related diseases. Therefore, in the rest of the analysis, we will consider only the smoke-pollution related diseases.

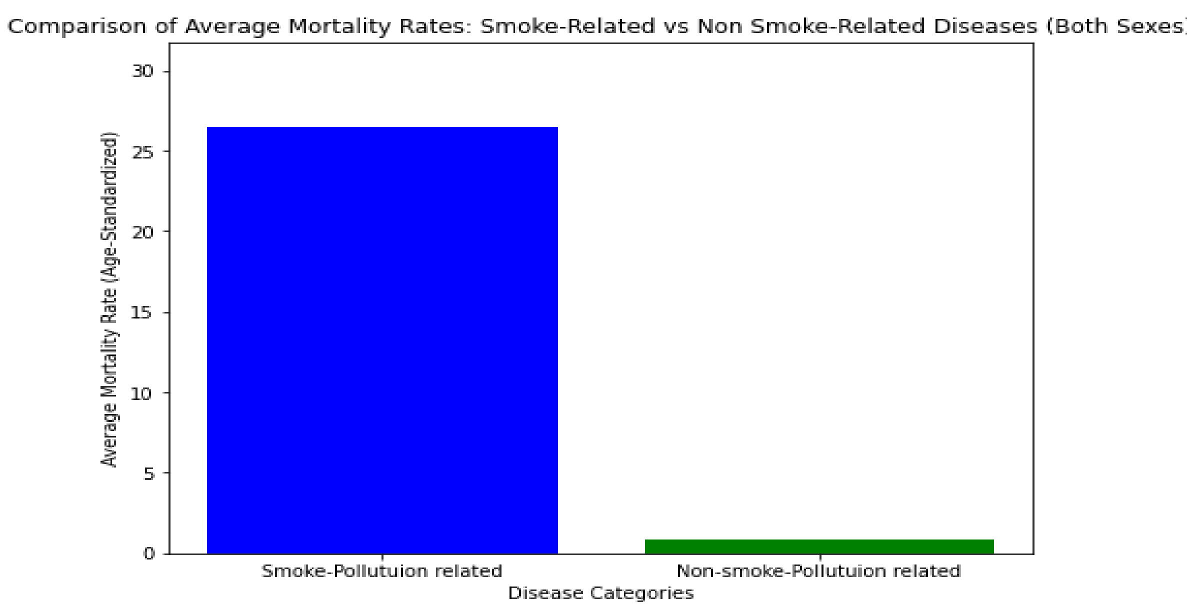

Fig 6: Bar plot comparing the average mortality rates for smoke-related and non-smoke related diseases for both the sexes

Question 3: How does fire smoke relate to the age-standardized mortality rate for respiratory diseases in Tulare County across different years?

Solution: We will individually focus on the impact of mortality rate from Asthma and COPD and the fire smoke estimate.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     mx   R-squared:                       0.227
Model:                            OLS   Adj. R-squared:                  0.204
Method:                 Least Squares   F-statistic:                     9.688
Date:                Tue, 05 Dec 2023   Prob (F-statistic):            0.00382
Time:                        02:15:22   Log-Likelihood:                -34.709
No. Observations:                  35   AIC:                             73.42
Df Residuals:                      33   BIC:                             76.53
Df Model:                           1
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const              2.2591      0.290      7.792      0.000       1.669       2.849
Smoke_estimate     0.0108      0.003      3.113      0.004       0.004       0.018
==============================================================================
Omnibus:                        3.168   Durbin-Watson:                   0.485
Prob(Omnibus):                  0.205   Jarque-Bera (JB):                2.363
Skew:                          -0.635   Prob(JB):                        0.307
Kurtosis:                       3.078   Cond. No.                         214.
==============================================================================
```

Fig 7: Regression analysis results for age-standardized mortality rate from asthma and the fire smoke estimate

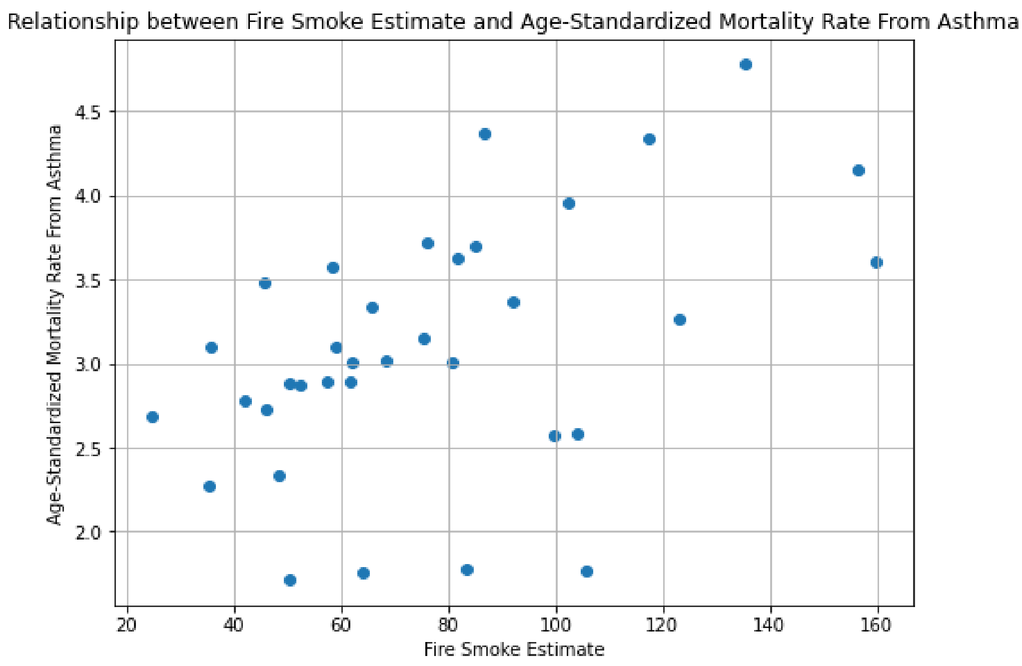

Fig 8: Correlation analysis between age-standardized mortality rate from asthma and fire smoke estimate

Through regression analysis, we explored the connection between the smoke estimate and age-standardized mortality rates from Asthma. The regression plot indicates that a rise in the fire smoke estimate leads to a 0.0108-unit increase in Asthma's age-standardized mortality rate. Additionally, the correlation plot reinforces this relationship, showing a strong correlation coefficient of 0.47639 between the two variables. Importantly, the regression model summary underscores a statistically significant association, emphasizing the relationship between fire smoke estimate and Asthma's mortality rate (p-value = 0.004). These findings accentuate the detrimental impact of wildfire smoke on Asthma, illuminating its role as a critical factor influencing population health.

We will now look at how mortality rate from COPD compares with the fire smoke activity and if there is any correlation between the two.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     mx   R-squared:                       0.003
Model:                            OLS   Adj. R-squared:                 -0.027
Method:                 Least Squares   F-statistic:                   0.09477
Date:                Tue, 05 Dec 2023   Prob (F-statistic):              0.760
Time:                        02:15:22   Log-Likelihood:                -26.649
No. Observations:                  35   AIC:                             57.30
Df Residuals:                      33   BIC:                             60.41
Df Model:                           1
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const              2.7009      0.230     11.728      0.000       2.232       3.169
Smoke_estimate    -0.0008      0.003     -0.308      0.760      -0.006       0.005
==============================================================================
Omnibus:                        9.833   Durbin-Watson:                   0.034
Prob(Omnibus):                  0.007   Jarque-Bera (JB):                3.959
Skew:                          -0.547   Prob(JB):                        0.138
Kurtosis:                       1.767   Cond. No.                         214.
==============================================================================
```

Fig 9: Regression analysis results for age-standardized mortality rate from COPD and the fire smoke estimate

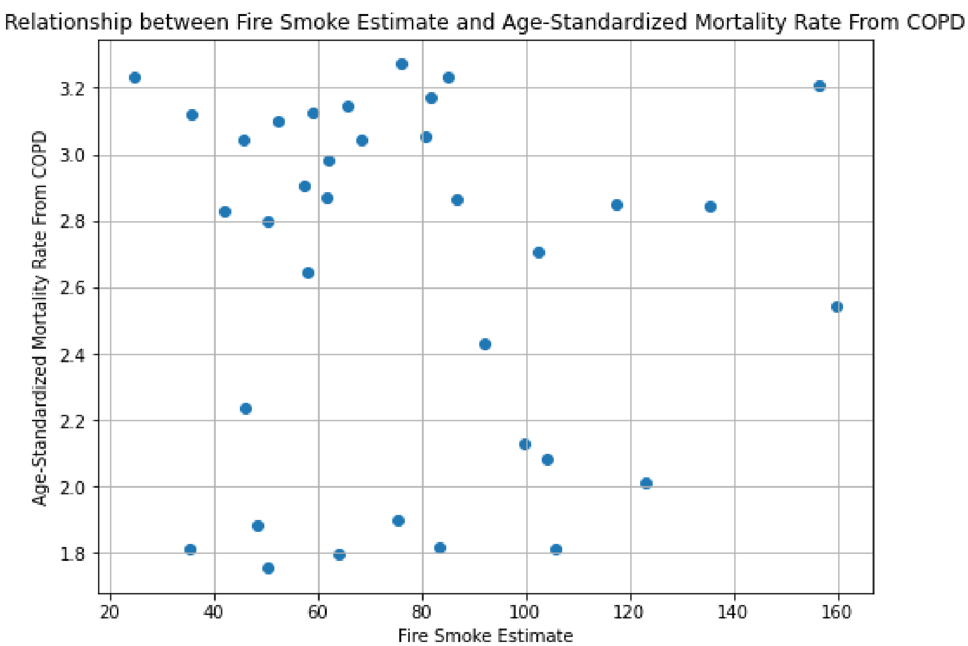

Fig 10: Correlation analysis between age-standardized mortality rate from COPD and fire smoke estimate

After conducting a regression analysis comparing smoke estimates with age-standardized mortality rates from COPD, the results indicated a minimal effect—a unit increase in fire smoke estimate correlated with a 0.0008-unit decrease in COPD mortality rates. The correlation plot between these variables displayed a weak association, with a reported correlation coefficient of -0.05351. Further scrutiny through the regression model summary revealed a statistically insignificant relationship between fire smoke estimate and COPD mortality rates, with a high p-value of 0.760. Consequently, the available data doesn't definitively link COPD to fire smoke activity.

Therefore, subsequent analysis will solely concentrate on Asthma, given the ambiguity regarding COPD's association with fire smoke based on our findings.

Question 4: Is there a noticeable difference in respiratory disease mortality rates between sexes due to variations in exposure to fire smoke in Tulare County?

Solution: To determine if there's a disproportionate impact on either the male or female population, we'll conduct separate analyses for each gender. Starting with the examination of the male population and its correlation with asthma, we'll delve into regression and statistical analyses.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     mx   R-squared:                       0.276
Model:                            OLS   Adj. R-squared:                  0.254
Method:                 Least Squares   F-statistic:                     12.61
Date:                Tue, 05 Dec 2023   Prob (F-statistic):            0.00118
Time:                        02:15:23   Log-Likelihood:                -41.824
No. Observations:                  35   AIC:                             87.65
Df Residuals:                      33   BIC:                             90.76
Df Model:                           1
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const              1.9981      0.355      5.624      0.000       1.275       2.721
Smoke_estimate     0.0151      0.004      3.550      0.001       0.006       0.024
==============================================================================
Omnibus:                        4.438   Durbin-Watson:                   0.691
Prob(Omnibus):                  0.109   Jarque-Bera (JB):                3.127
Skew:                          -0.694   Prob(JB):                        0.209
Kurtosis:                       3.464   Cond. No.                         214.
==============================================================================
```

Fig 11: Regression analysis results for age-standardized mortality rate for male population from asthma and the fire smoke estimate

Based on our analysis, we conducted a regression study linking the smoke estimate and the age-standardized mortality rate of males due to Asthma. Our regression model summary revealed that for every unit increase in the fire smoke estimate, there is a 0.0151-unit rise in the age-standardized mortality rate of males from Asthma. Additionally, our correlation plot emphasized a robust association between these variables, indicated by a correlation coefficient of 0.5257. The regression model summary underscored a significant relationship between the fire smoke estimate and the age-standardized mortality rate of males from Asthma, with a notable p-value of 0.001. These findings highlight a considerable impact on the male population concerning Asthma-related mortality rates.

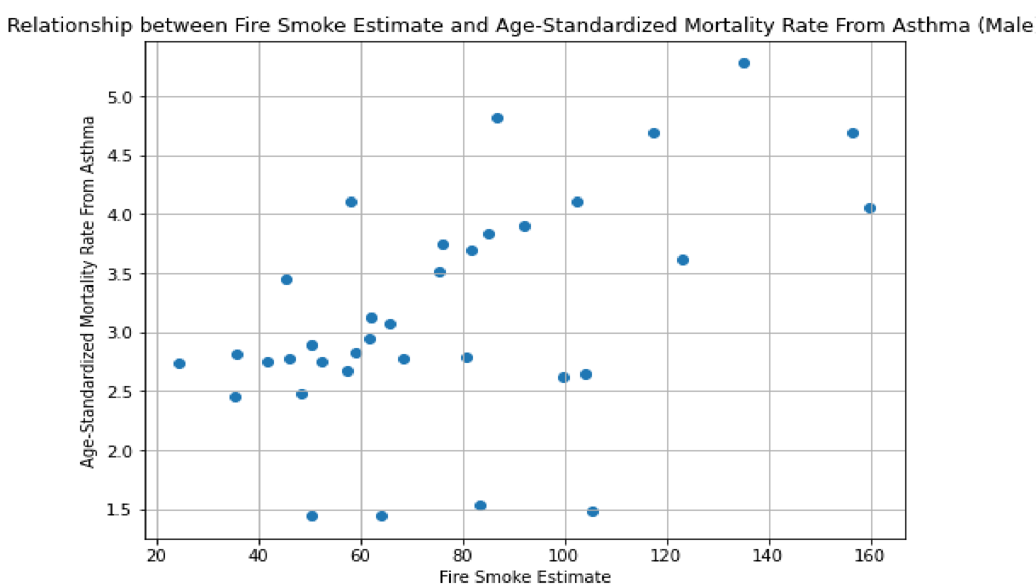

Fig 12: Correlation analysis between age-standardized mortality rate for male population from asthma and fire smoke estimate

Next, we will perform a similar analysis on the female population.

Our analysis involved a regression study linking the smoke estimate and the age-standardized mortality rate among females due to asthma. The regression model summary demonstrated that as the fire smoke estimate increases by one unit, there's a corresponding 0.0065-unit rise in the age-standardized mortality rate among females suffering from asthma. Moreover, the correlation plot displayed a noticeable connection between these variables, indicating a robust correlation with a coefficient value of 0.3532. The regression model summary underscored the statistical significance of the relationship, revealing a noteworthy association between the fire smoke estimate and the age-standardized mortality rate of females affected by asthma, indicated by a p-value of 0.037. These findings emphasize the substantial impact of asthma on the female population, especially concerning age-standardized mortality rates, attributing a significant role to fire smoke estimates in understanding this relationship.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                     mx   R-squared:                       0.125
Model:                            OLS   Adj. R-squared:                  0.098
Method:                 Least Squares   F-statistic:                     4.705
Date:                Tue, 05 Dec 2023   Prob (F-statistic):             0.0374
Time:                        02:15:23   Log-Likelihood:                -29.632
No. Observations:                  35   AIC:                             63.26
Df Residuals:                      33   BIC:                             66.37
Df Model:                           1
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const              2.5213      0.251     10.054      0.000       2.011       3.032
Smoke_estimate     0.0065      0.003      2.169      0.037       0.000       0.013
==============================================================================
Omnibus:                        3.203   Durbin-Watson:                   0.331
Prob(Omnibus):                  0.202   Jarque-Bera (JB):                1.854
Skew:                          -0.305   Prob(JB):                        0.396
Kurtosis:                       2.052   Cond. No.                         214.
==============================================================================
```

Fig 13: Regression analysis results for age-standardized mortality rate for female population from asthma and the fire smoke estimate

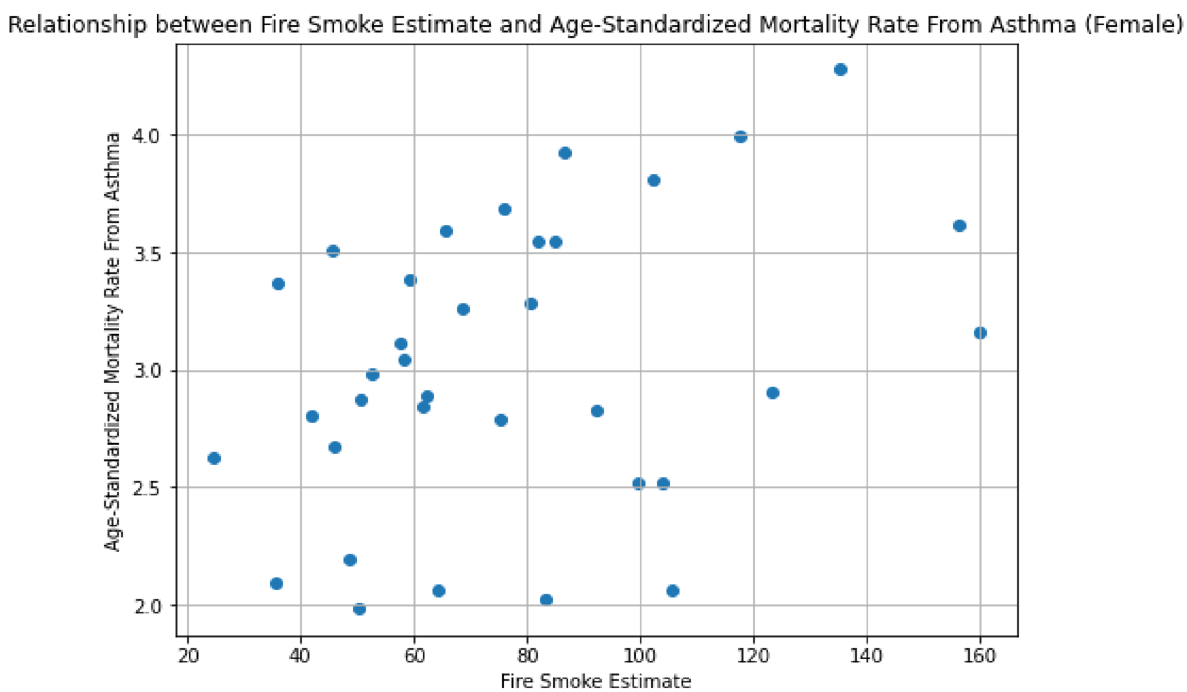

Fig 14: Correlation analysis between age-standardized mortality rate for female population from asthma and fire smoke estimate

Although we have found correlations between the mortality rates of male and female populations with the fire smoke estimate, however, the question “Is there a difference between the two” still remains largely unanswered. To answer this, we plot a time series plot, a bar plot and perform a statistical test.

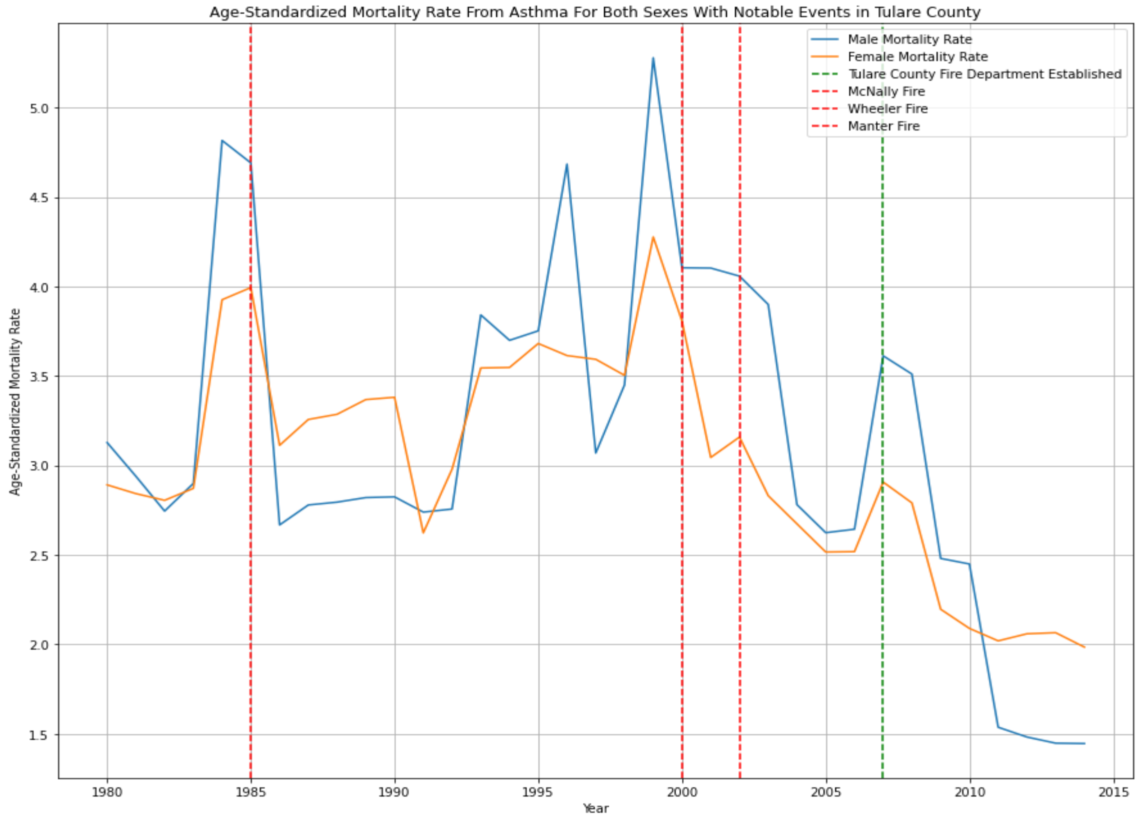

Fig 15: Age-standardized mortality rate from asthma for both sexes with notable events

The time series plot illustrates the age-standardized mortality rates for males and females from 1980 to 2014. The blue line indicates male mortality, while the golden yellow line represents female mortality. Both populations exhibit similar peaks and valleys, although males tend to display slightly higher mortality rates. Contextual markers, including major wildfires in Tulare County and the establishment of the fire department, offer insight into potential influences on these rates. The red dashed lines correspond to three significant wildfires in the vicinity, coinciding with notable spikes in mortality for both genders. Additionally, the green dashed line signifies the establishment of Tulare County's fire department in 2007, aligning with a consistent decline in mortality rates for both male and female populations—an encouraging trend observed since then.

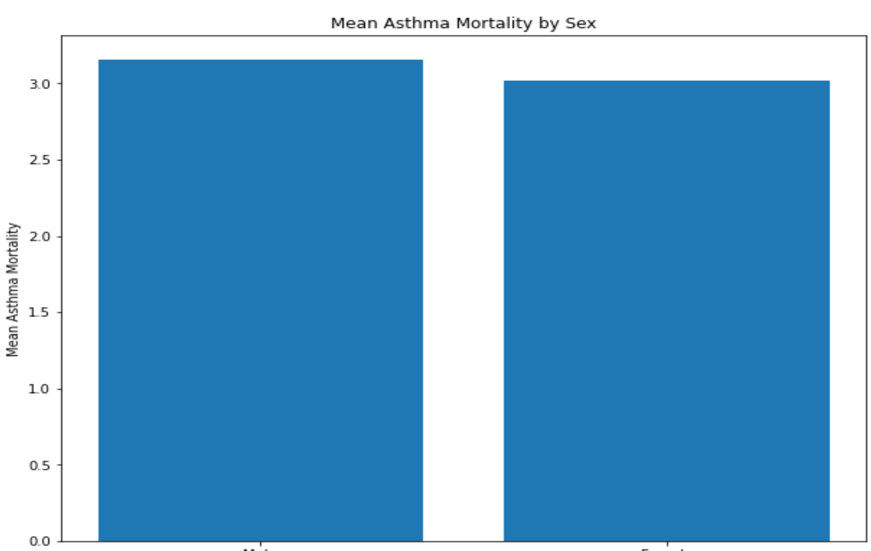

Fig 16: Mean asthma mortality rate by sex

When comparing the mean asthma mortality rates between males and females using a bar chart, it appeared that both genders exhibited nearly equivalent rates, albeit with a slightly higher mean for males. To validate this observation, a t-test was conducted, revealing a t-statistic value of 0.7152 and a corresponding p-value of 0.4769. These results indicate insufficient evidence to assert a significant difference in smoke-related asthma deaths between genders. Therefore, statistically, there doesn't seem to be a noteworthy disparity in asthma mortality rates attributable to smoke exposure between males and females.

Question 5: How does the prevalence of smoke fires in Tulare County connect with rates of asthma-related hospitalizations across different age groups over time?

Solution: We aim to assess if the effects of fire smoke effect differ across age groups. Specifically, we'll examine asthma-related hospitalizations in various age brackets for this analysis.

Fig 17: Plot of year vs smoke estimate and year vs number of hospitalizations

After conducting an initial analysis of our data, we generated two plots that offer valuable insights. Over the years from 2015 to 2020, we observed a rising trend in the fire smoke estimate. Surprisingly, during this period, there's a downward trend in the number of hospitalizations across all age groups. This contrast between an increase in fire smoke and a decrease in hospitalizations is remarkably positive.

Based on our analysis of three age groups—0-17 years, 18-64 years, and 65+ years—we constructed a bar plot showcasing the average number of hospitalizations within these groups (see Figure 18). From this visual representation, it's evident that the mean hospitalizations for the 0-17 years category are approximately two to four times higher than those for both the 18-64 years and 65+ years groups. Surprisingly, the second most affected demographic appears to be the 18-64 years bracket. This disparity emphasizes a disproportionate impact of Asthma on children and teenagers, signaling an urgent need for improved measures and monitoring to address this concerning trend and ensure better outcomes for younger age groups in the future.

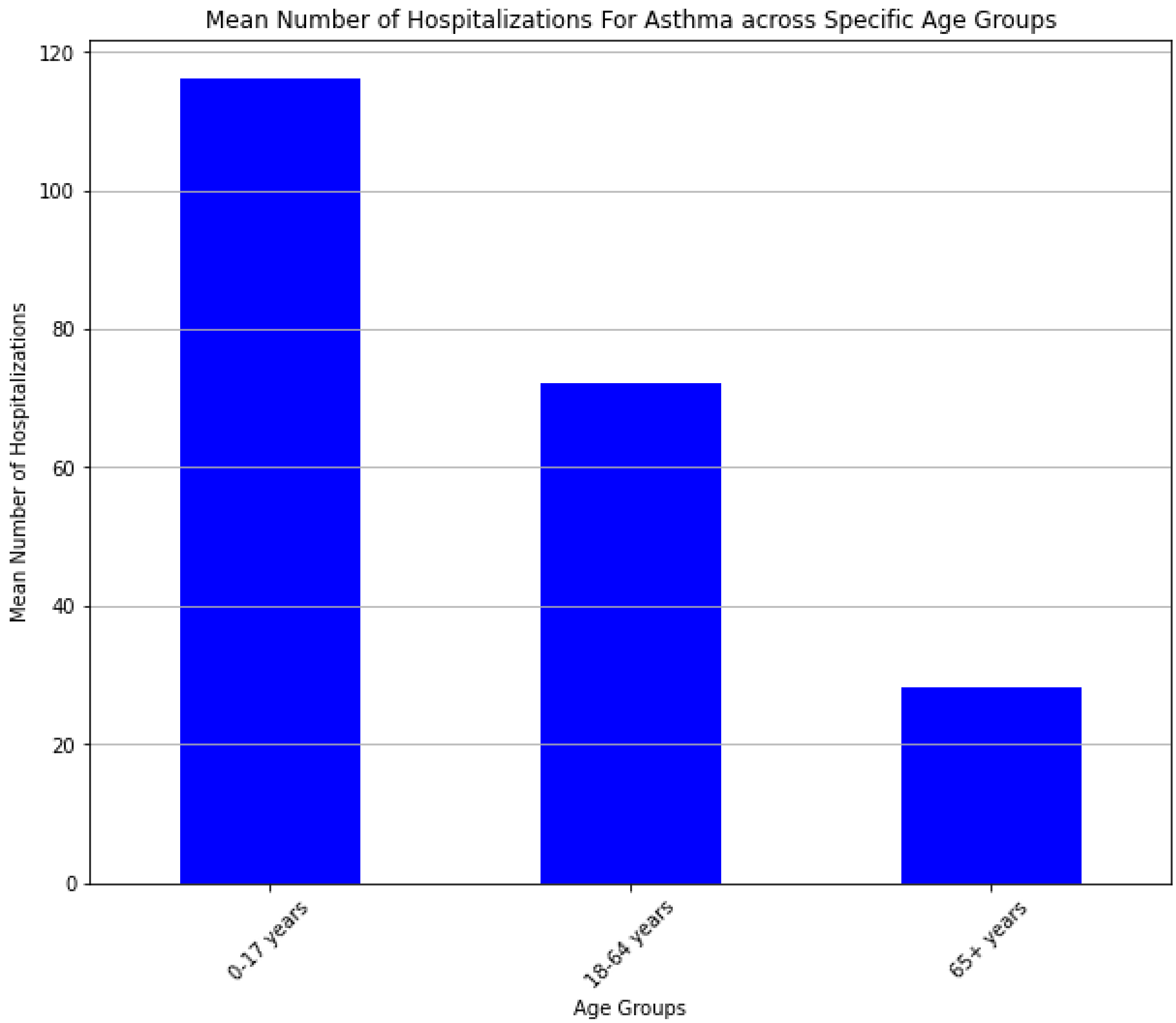

Fig 18: Plot of mean number of hospitalizations for asthma across specific age groups

DISCUSSION

This research study presents a promising opportunity to tackle the urgent challenge of wildfire smoke's impact on public health, specifically within Tulare County. The findings uncovered here carry substantial significance for the county's well-being, shedding light on the significant effects of wildfire smoke on respiratory health. This emphasizes the immediate need for action to reduce health risks. The detailed connections found between wildfire smoke and respiratory issues, notably Asthma, highlight the pressing need to address these concerns promptly. City officials, such as the council, city manager, and mayor, can use this valuable information to shape policies prioritizing fire prevention, emergency responses, and healthcare improvements. Strengthening rules and readiness plans becomes crucial going forward. Equally important is translating these research discoveries into understandable information for the public. By using clear visuals and transparent documentation, the study can empower communities to grasp the risks linked to wildfire smoke and advocate for their health needs. For Tulare County officials, these insights offer guidance for long-term planning, emergency readiness, and allocating resources effectively. The projected trends in fire smoke estimates serve as a vital roadmap for resource allocation, emergency training, and infrastructure enhancements.

For the city council, city manager, mayor, and residents, a multi-faceted approach is essential:

- They should consider revising existing policies to integrate more stringent measures for wildfire prevention, enhance emergency response protocols, and allocate resources for better healthcare facilities.
- Collaborative efforts involving city officials and residents are vital to launching educational campaigns.

- The city council should allocate resources for infrastructure improvements, including better ventilation systems in public buildings, especially in schools and hospitals, to minimize the impact of poor air quality.
- Residents should actively participate in community forums, engage in emergency preparedness workshops, and collaborate with local authorities to develop neighborhood-level response plans.

A concrete plan incorporating policy revisions, public health campaigns, infrastructure enhancements, and community engagement should be formulated within the next 6 to 12 months to ensure timely intervention.

Throughout this project, human-centered data science principles have been at the forefront. Every aspect of the study was geared towards understanding the tangible impact on human lives. From analyzing health outcomes to forecasting future trends, the aim was always to translate data insights into actionable solutions that directly benefit the community. The use of data visualization techniques and clear documentation was deliberate. These efforts aimed to make complex data more understandable and accessible to policymakers, officials, and residents, fostering informed decision-making and community engagement. The study not only identified issues but aimed to empower communities by providing insights that foster collaboration between residents, local authorities, and policymakers. The emphasis on collaborative efforts is key to effecting meaningful change.

LIMITATIONS

There are several limitations to this study in terms of the assumptions made and the collection of data. The following are some of the limitations encountered while performing this analysis:

1. **Data Limitations**: The reliance on existing datasets limits the scope and depth of analysis, potentially overlooking critical factors such as household incomes or other nuances not captured in the available data.
2. **Temporal Constraints**: The study's focus on a specific timeframe (1963-2023) may overlook long-term historical trends or emerging patterns beyond this duration, affecting the comprehensiveness of the findings.
3. **Geographical Scope**: For most of the analysis we consider mostly Tulare County or nearby places. However, it is possible that wildfires in other regions might also have a big factor on the effect of wildfire smoke in Tulare County. SO it is important to consider the cross-county or regional influences.
4. **Data Merging and Omission**: The merging process based on the "Year" column could result in data loss due to inconsistencies or missing entries in different datasets for specific years, affecting the accuracy of the consolidated dataset.
5. **Assumption in Fire Impact Estimation**: The fire smoke estimate model's assumptions regarding fire types, size, and distance might oversimplify the complex nature of smoke dispersion, potentially leading to inaccuracies in estimations.
6. **Limited Disease Representation**: The study primarily focuses on Asthma and COPD, potentially neglecting other respiratory conditions or health issues indirectly impacted by wildfire smoke.
7. **Causality and Correlation**: While the study identifies correlations between fire smoke estimates and health outcomes, establishing a direct cause-effect relationship requires more extensive experimental or longitudinal studies.
8. **Population Dynamics**: The analysis doesn’t explicitly account for population demographic shifts, which can influence health outcomes independently of fire smoke exposure.
9. **Model Predictive Reliability**: While the ARIMA forecasting model predicts smoke estimates, its accuracy might diminish as the prediction period extends further into the future, impacting the reliability of long-term projections.
10. **Gender and Age-Specific Analysis**: The study’s segmentation into gender and age groups might oversimplify complexities within these demographics, potentially masking nuanced variations in health impacts.

CONCLUSION

The study reveals interesting results from the analyses conducted on the following research questions and hypotheses.

What are the estimated smoke impacts on Tulare city for the last 60 years?

The analysis of fire distribution, burned acres, and the correlation between fire smoke and Air Quality Index over the past 60 years in Tulare City indicates a significant increase in fires between 100 and 600 miles from the city, peaking notably between 450 and 600 miles. This escalation in burned acres from 1963 to 2020 likely stems from hotter climates, prolonged dry spells, changing flora, and increased human presence near forests. Furthermore, a predictive model forecasts a sustained or potentially worsening trajectory for fire smoke in Tulare County from 2024 to 2049, raising concerns about the persisting impact of fire smoke in the region.

To what extent do certain respiratory diseases linked to smoke-related pollution (e.g., Chronic respiratory diseases, Chronic obstructive pulmonary disease, Asthma) exhibit higher mortality rates compared to diseases not directly related to smoke-related pollution (e.g., Pneumoconiosis, Coal workers pneumoconiosis, Silicosis, Asbestosis) in Tulare County?

Smoke-related respiratory diseases like Asthma and Chronic obstructive pulmonary disease (COPD) exhibit significantly higher mortality rates, averaging 30 times greater, compared to diseases not directly linked to smoke-related pollution such as Pneumoconiosis, Coal workers pneumoconiosis, Silicosis, and Asbestosis in Tulare County. This stark difference emphasizes the elevated fatality associated with smoke-related respiratory illnesses, prompting the analysis to focus solely on exploring these smoke-pollution related diseases further.

How does fire smoke relate to the age-standardized mortality rate for respiratory diseases in Tulare County across different years?

In this research study, we analyzed the relation between fire smoke estimate and the mortalities happening from asthma and COPD dieseases. Based on the regression, correlation and statistical analysis, we conclude that although there is a definitive link between fire smoke activity and mortality rates from asthma, there is none for COPD in the Tulare County region.

Is there a noticeable difference in respiratory disease mortality rates between sexes due to variations in exposure to fire smoke in Tulare County?

Based on the analysis we performed in this study, fire smoke seems to impact Asthma-related mortality rates for both genders in Tulare County, with regression analyses showing associations between smoke and mortality. For males, a unit increase in fire smoke correlates with a 0.0151-unit rise in Asthma-related deaths, while for females, it results in a 0.0065-unit increase. Yet, examining mortality rates from 1980 to 2014 reveals similar trends between genders, with slightly higher peaks for males. A statistical t-test found no significant difference in smoke-related Asthma deaths between males and females, suggesting no notable gender-based disparity in smoke-related Asthma mortality rates.

How does the prevalence of smoke fires in Tulare County connect with rates of asthma-related hospitalizations across different age groups over time?

The analysis of fire smoke estimates and asthma-related hospitalizations across age groups unveils intriguing trends. Examining the age-specific impact reveals a strikingly higher average hospitalization rate for the 0-17 year group, significantly surpassing both the 18-64 years and 65+ years categories. This pattern underscores a disproportionate effect on children and teenagers, signaling an urgent need for enhanced strategies and vigilant monitoring to address asthma's impact on these younger age cohorts in the future.

This analysis demonstrates the significance of human-centered data science in Tulare County, examining the impact of fire smoke on respiratory health. It highlights the long-term effects on respiratory health, particularly among younger age groups with asthma. The study emphasizes the role of predictive modeling in forecasting health trends, urging proactive measures to address escalating respiratory issues. By integrating data insights with health priorities, it informs interventions and policies for better outcomes. These findings, derived from quantitative and qualitative data, offer valuable guidance for implementing reforms to improve public health in Tulare County.

REFERENCES

Ref 1: “Tulare County.” Wikipedia, Wikimedia Foundation, 10 October 2023, https://en.wikipedia.org/wiki/Tulare_County,_California

Ref 2: Reid, C.E., Brauer, M., Johnston, F.H., Jerrett, M., Balmes, J.R. and Elliott, C.T., 2016. Critical review of health impacts of wildfire smoke exposure. Environmental health perspectives, 124(9), pp.1334-1343.

Ref 3: Liu, J.C., Wilson, A., Mickley, L.J., Dominici, F., Ebisu, K., Wang, Y., Sulprizio, M.P., Peng, R.D., Yue, X., Son, J.Y. and Anderson, G.B., 2017. Wildfire-specific fine particulate matter and risk of hospital admissions in urban and rural counties. *Epidemiology (Cambridge, Mass.)*, *28*(1), p.77.

Ref 4: Masri, S., Scaduto, E., Jin, Y. and Wu, J., 2021. Disproportionate impacts of wildfires among elderly and low-income communities in California from 2000–2020. *International journal of environmental research and public health*, *18*(8), p.3921.

Ref 5: Cascio, W.E., 2018. Wildland fire smoke and human health. *Science of the total environment*, *624*, pp.586-595.

Ref 6: Rappold, A.G., Reyes, J., Pouliot, G., Cascio, W.E. and Diaz-Sanchez, D., 2017. Community vulnerability to health impacts of wildland fire smoke exposure. *Environmental Science & Technology*, *51*(12), pp.6674-6682.

Ref 7: 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

Ref 8: Institute for Health Metrics and Evaluation (IHME). United States Chronic Respiratory Disease Mortality Rates by County 1980-2014. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), 2017.

Ref 9: Healthcare Information Division, Department of Health Care Access and Information [Distributor: California Breathing Asthma Program, California Department of Public Health]. 2019. AsthmaHospitalizationRates-by-County-CA-CDPH.csv

Ref 10: “data-512-Wildfires_Impact_on_Tulare_County”, https://github.com/sagnikgh1899/data-512-Wildfires_Impact_on_Tulare_County

Ref 11: Sears, M.R., 2015. Smoking, asthma, chronic airflow obstruction and COPD. *European Respiratory Journal*, *45*(3), pp.586-588.

DATA SOURCES

1. ***USGS Wildland Fire Combined Dataset***: This dataset is in JSON format. The common analysis research question is based on one specific dataset which can be found at Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons). This dataset was collected and aggregated by the US Geological Survey. The dataset is relatively well documented. Fire polygons are available in ArcGIS and GeoJSON formats. We have been assigned one US city that will form the basis for our individual analysis. We can find our individual US city assignment from this Google spreadsheet.

License: [USGS Copyrights and Credits]

Table 1: Description of USGS Wildland Fire Combined Dataset

NAME	DTYPE	DESCRIPTION
OBJECTID	Integer	Unique identification for the polygon and it's attributes
USGS_Assigned_ID	Integer	Assigned unique identification for the polygon and it's attributes. Used to provide consistency if parts of the dataset are exported or the OBJECTID is otherwise changed
Assigned_Fire_Type	String	Based on the fire polygon(s) used to create this fire feature what is the type assigned to this fire? If more than one type was assigned to a combined polygon, the assigned fire type was assigned in the following order of dominance: Wildfire, Likely Wildfire, Unknown - Likely Wildfire, Prescribed Fire, Unknown - Likely Prescribed Fire
Fire_Year	Integer	The calendar year when the dataset creators determined the fire occurred
Fire_Polygon_Tier	Integer	The tier from which the fire polygon was generated. One or more polygons within the tier could be combined to create the fire polygon
Fire_Attribute_Tiers	String	All fire tiers that contributed attributes to the fire feature. A list of all tiers where a polygon intersects the current fire perimeter in space and time
GIS_Acres	Float	The GIS calculated acres of the fire polygon calculated by using the Calculate Geometry tool in ArcGIS Pro
GIS_Hectares	Float	The GIS calculated hectares of the fire polygon calculated by using the Calculate Geometry tool in ArcGIS Pro
Source_Datasets	String	All of the original source datasets that contributed to either the polygon or the attributes. Each dataset has the number of polygons contributed listed in parentheses after the dataset name
Listed_Fire_Types	String	Each fire type listed in the fires from the merged dataset that intersect this polygon in space and year. The number of features that contributed the specific fire type are in parentheses after the fire type
Listed_Fire_Names	String	Each fire name listed in the fires from the merged dataset that intersect this polygon in space and year. The number of features that contributed the specific fire name are in parentheses after the fire name

Listed_Fire_Codes	String	Each fire code listed in the fires from the merged dataset that intersect this polygon in space and year. The number of features that contributed the specific fire code are in parentheses after the fire code
Listed_Fire_IDs	String	Each fire type listed in the IDs from the merged dataset that intersect this polygon in space and year. The number of features that contributed the specific fire ID are in parentheses after the fire ID
Listed_Fire_IRWIN_IDs	String	Each fire IRWIN ID listed in the fires from the merged dataset that intersect this polygon in space and year. The number of features that contributed the specific fire IRWIN ID are in parentheses after the fire IRWIN ID
Listed_Fire_Dates	String	Each fire date listed in the fires from the merged dataset that intersect this polygon in space and year. The number of features that contributed the specific fire date are in parentheses after the fire date
Listed_Fire_Causes	String	Each fire cause listed in the fires from the merged dataset that intersect this polygon in space and year. The number of features that contributed the specific fire cause are in parentheses after the fire cause
Listed_Fire_Cause_Class	String	Each fire cause class listed in the fires from the merged dataset that intersect this polygon in space and year. The number of features that contributed the specific fire cause class are in parentheses after the fire cause class
Listed_Rx_Reported_Acres	String	Each prescribed fire reported acres listed in the fires from the merged dataset that intersect this polygon in space and year. The number of features that contributed the specific reported acres are in parentheses after the reported acres
Listed_Map_Digitize_Methods	String	Each fire digitization method listed in the fires from the merged dataset that intersect this polygon in space and year. The number of features that contributed the specific fire digitization method are in parentheses after the fire digitization method
Listed_Notes	String	Each fire notes listed in the fires from the merged dataset that intersect this polygon in space and year. The number of features that contributed the specific fire notes are in parentheses after the fire note
Processing_Notes	String	Indicates that the attribute data were altered during the processing and a new attribute was indicated. It will also explain the rationale for the change. Each polygon that had an attribute changed will be listed along with a count, in

		parentheses indicating how many polygons had the change made to them
Wildfire_Notice	String	A notice present in every field that indicates the quality of the wildfire data in this dataset
Prescribed_Burn_Notice	String	A notice present in every field that indicates the quality of the prescribed burn data in this dataset
Wildfire_and_Rx_Flag	String	A text flag field indicating that the attributes from the various sources indicate that the fire was both a wildfire and a prescribed fire. This could indicate an error in assigning the fire type, a misassignment of the fire type, or that there were actually two fires that occurred in this area in the same year, one a wildfire and one a prescribed burn
Overlap_Within_1_or_2_Flag	String	An ArcGIS Tabulate Intersection Tool was used to identify areas that burned with >10% overlap of the current fire within 1 or 2 years of the current burn. Each fire that met that criteria was included in this attribute including it's ID, year burned, percent overlap, and acres
Circleness_Scale	Float	A measure of a polygon's similarity to a true circle. calculated using the Shape_Length and Shape_Area fields. Circle-ness = 4pi(Shape_Area/(Shape_Length * Shape_Length)). As the number approaches 1, the polygon becomes more circular
Circle_Flag	String	Any Circle circle-ness values >=0.98 are flagged with a 1. The remaining values are null. 1 indicates that the polygon is very circle-like and is likely incorrect. However, other values that are not flagged may still be quite circular and incorrect
Exclude_From_Summary_Rasters	String	Some fires in this dataset appear to be buffered circles. These were kept in the dataset to show location and approximate area. However a decision was made to exclude circular fires larger than 1 acre in size from the summary raster calculations. This field indicates whether the fire was excluded from ('Yes') or included in ('No') the summary raster calculations
Shape_Length	Float	Automatically calculated perimeter length in meters
Shape_Area	Float	Automatically calculated polygon area in square meters

2. ***Mortalities from 1980 to 2014 from Respiratory Diseases in California Counties***: This dataset is in CSV format and can be downloaded from the Institute for Health Metrics and Evaluation (IHME) website. The dataset is maintained and hosted by IHME. It provides us the mortalities from 1980 to 2014 from Respiratory Diseases in all California Counties. This dataset comprehensively records data on causes, years, sexes, posterior mean estimates, 2.5th percentile estimates, and 97.5th percentile estimates. It

delineates the age-standardized mortality rate (deaths per 100,000 population) across different sexes and for both sexes combined, spanning the years 1980-2014 for all counties within California. Focused on Tulare County, the dataset underwent filtering using FIPS=6107, resulting in 1050 entries across 16 columns. The dataset includes both string-based descriptors (measure_name, location_name, cause_name, sex, age_name, metric) and integer-based identifiers (measure_id, location_id, FIPS, cause_id, sex_id, age_id, year_id, mx, lower, upper). It's worth noting that the dataset stands clean, devoid of any null values across all columns.

License: [IHME FREE-OF-CHARGE NON-COMMERCIAL USER AGREEMENT]

Table 2: Description of Mortalities from 1980 to 2014 from Respiratory Diseases in California Counties

NAME	DTYPE	DESCRIPTION
measure_id	Integer	Unique numeric identifier for the measure generated
measure_name	String	The measure (indicator) of the estimate
location_id	Integer	Unique numeric identifier for the location generated
location_name	String	Location of the estimate
FIPS	Integer	The Federal Information Processing Standards (FIPS) code, a unique identifier for states and counties in the United States
cause_id	Integer	Unique numeric identifier for the cause of disease or injury generated
cause_name	String	Cause of disease or injury of the estimate
sex_id	Integer	Unique numeric identifier for the sex generated
sex	String	Gender for the estimate
age_id	Integer	Unique numeric identifier for the age group generated
age_name	String	Age group estimated
year_id	Integer	Time period of estimate
metric	String	Metric/unit of measure for the estimate
mx	Float	Posterior mean estimate
lower	Float	2.5% percentile estimate
upper	Float	97.5% percentile estimate

3. ***Asthma Hospitalization Rates For California Counties from 2015 to 2020***: This dataset is also in CSV format and can be downloaded from the California Health and Human Services Open Data Portal (CalHHS) website. The dataset is maintained and hosted by CalHHS. This comprehensive dataset encompasses counts and rates (per 10,000 residents) of asthma-related hospitalizations across all counties in California. The dataset is organized by age groups (all ages, 0-17, 18-64, 65+) and racial/ethnic categories (white, black, Hispanic, Asian/Pacific Islander, and American Indian/Alaskan Native). Sourced

from the Department of Health Care Access and Information Patient Discharge Data, the information is further refined by filtering exclusively for Tulare County, resulting in 76 entries and 9 columns.

Table 3: Description of Asthma Hospitalization Rates for California Counties (2015 – 2020)

NAME	DTYPE	DESCRIPTION
_id	Integer	Unique numeric identifier for the specific entry
COUNTY	String	County of residence
YEAR	Integer	Year of hospitalization (discharge)
STRATA	String	General demographic category under which responses have been stratified
STRATA NAME	String	Specific demographic type under which responses have been stratified
AGE GROUP	String	All ages, 0-17 years, 18-64 years, and 65+ years
NUMBER OF HOSPITALIZATIONS	String	Number of hospitalizations for asthma. Events are counted per hospitalization and not per person
AGE-ADJUSTED HOSPITALIZATION RATE	Float	Calculated by dividing the number of asthma hospitalizations by the estimated population in that county and age group, age-adjusting to the 2000 U.S. Census, and multiplying by 10,000
COMMENT	String	Any specific notes related to the particular entry