Fighting Fire with FIRES:

Risk Estimation Analysis for Wildfire Preparedness in California

2021-22 Modeling the Future Challenge

Team ID: 8935

Executive Summary

In 2018, the deadliest fire in recorded California history raged across the forests of Butte County. Known as the Camp Fire, it killed 86 people and destroyed 15,000 homes.[1] Wildfires are worsening: in the past decade, more than 12 million acres have been burned from wildfires in California.[2] Currently, more than 11 million Californians live in high-risk wildfire areas.[3] Fires endanger structures, air quality, recreation areas, water and watersheds, and human lives. While state and federal agencies spend billions of dollars fighting California wildfires, relatively little is spent on reducing California's wildfire hazard.[4]

To analyze strategies to mitigate the risks of wildfires, we used ArcGIS, a service that handles geographic information systems. We utilized California's Fire and Resource Assessment Program (FRAP) and California County Boundaries datasets to find where wildfires occurred in California and what caused them. We assessed that lightning and human causes such as illegal campfires or misused equipment resulted in the most acres of wildfire burning in California.

We also used the Fire Stations dataset from the United States Geological Survey (USGS) to locate fire stations in California. The United States Census data from 2020 allowed us to find population numbers by neighborhood, identify which areas are water, and pinpoint the number of occupied housing units in California. The California Farmland Mapping and Monitoring Program (FMMP) dataset contained the amount of farmland in each county in California.

The FRAP dataset allowed us to create a model that predicts the number of acres burned from wildfires in California in the future. We used quartic and linear regressions, both of which predicted steep increases in wildfire destruction in the future. Our ARIMA models found that the number of burned acres from wildfires would exceed 2.5 million acres in a single year in California, indicating we desperately need to implement wildfire risk mitigation strategies.

To protect against lightning-caused wildfires, we recommend surge protectors and lightning rod protection at high-risk locations. However, we found that protecting just one square kilometer of forest would require several hundred lightning rods. Instead, lightning rods should be used to protect power lines and residential houses. We recommend that homeowners spend the $1,300 that would cover lightning rod installation for an average house in California. Like lightning rods, we recognize other more standard fire mitigation strategies currently in place, such as home and fire insurance, fuel management, and prescribed burns, and thus omit those from our recommendations.

Using the Fire Stations dataset, we modeled the response time of fire stations with the Fire Station Reach Model. While fire stations should optimally be able to respond to a fire in under 15 minutes, many areas were unreachable in this time. As a result, we recommended more fire stations in areas we marked as "Cold Spots". We determined that adding 10 new fire stations in Cold Spots would cost $10 million and would be cost-effective fire protection, so we recommend building more than 10.

Further, another solution we considered was using Internet of Things (IoT) sensors to detect early sparks of wildfires. While sensors in the past have been limited by their range, sensors now have the potential to send information to each other as a network or via satellites. To cover 15% of California, a mere $1.7 million dollars would be required for materials. While upkeep and installation may be more expensive, we predict using IoT sensors would bring a net benefit of $1.07 billion in the first year.

Using the Fire Stations, Census, FRAP, and FMMP datasets, we found the value of the amount of loss each county in California would suffer from a wildfire. Using the FRAP data and the Fire Station Reach Model, we also created a metric for the chance each county might undergo a wildfire and its readiness to combat one. We then combined these two metrics to make the Fire Index Risk Estimator Score (FIRES), a comprehensive representation of fire risk in California. Armed with FIRES data, we identified a set of counties that are at high risk for fires but may be underprepared to fight them and recommended areas to place both fire stations and remote sensing networks. Kern, Fresno, and Tehama counties had the highest FIRES assessment. The California government can implement our suggestions to lower wildfire risk in the state's most vulnerable areas.

Background

Since 2012, more than 12 million acres of land have been burned in California by wildfires. To put that in perspective, the acre amount accounts for one in every eight acres of California land. Indeed, the situation is getting worse, as 9 of 10 of the largest California wildfires since 1932 have occurred in the past decade.[2]

The USDA Forest Service has relied on fire seasons to tamp out the burning. However, what was previously a four-month fire season blazing in the summertime now lasts six to eight months. It is not uncommon for fires to start even in the winter.[5] Worryingly, U.S. Forest Service Chief Randy Moore commented that, instead of fire seasons, California has "fire years". [2]

The 2018 wildfire season was the deadliest and most destructive ever recorded in California. More than 24,000 buildings were damaged or destroyed, and over 100 people lost their lives.[6] Wildfire damages in 2018 alone cost approximately $150 billion, causing losses of capital (19%), health costs (22%), and indirect losses to industry sectors (59%). Most indirect losses were outside of California as wildfires disrupted regional and national supply chains.[7]

Almost 10 million people own homes and businesses in wildland areas. Fires threaten residents' lives, structures, air quality, recreation areas, and water and watersheds. To address wildfires, the California government developed the California Fire Plan with four main pre-fire management techniques:[8]

1. Fuel breaks to stop wildfires in their tracks.
2. Wildfire Protection Zones around communities that buffer wildfires.
3. Forest maintenance for healthy forests.
4. Prescribed burns to reduce fire fuel, such as foliage and dead trees.

To identify areas in high risk, California has created Fire Hazard Severity Zones (FHSZ) map (Fig. 1). The map identifies areas of Moderate, High, and Very High risk. [9] However, California law does not have zones in local responsibility areas (LRAs) due to a lack of jurisdiction over them. Expanding the map would help local homes and businesses identify wildfire risk. Additionally, the current map does not reflect the shape of suburban neighborhoods, as some houses are in a Very High zone while others lack designation at all. Further criticisms of the map arise from its lack of detail, as the FHSZ map is not useful for assessing where the state should increase investments for wildfire resistance.[10]

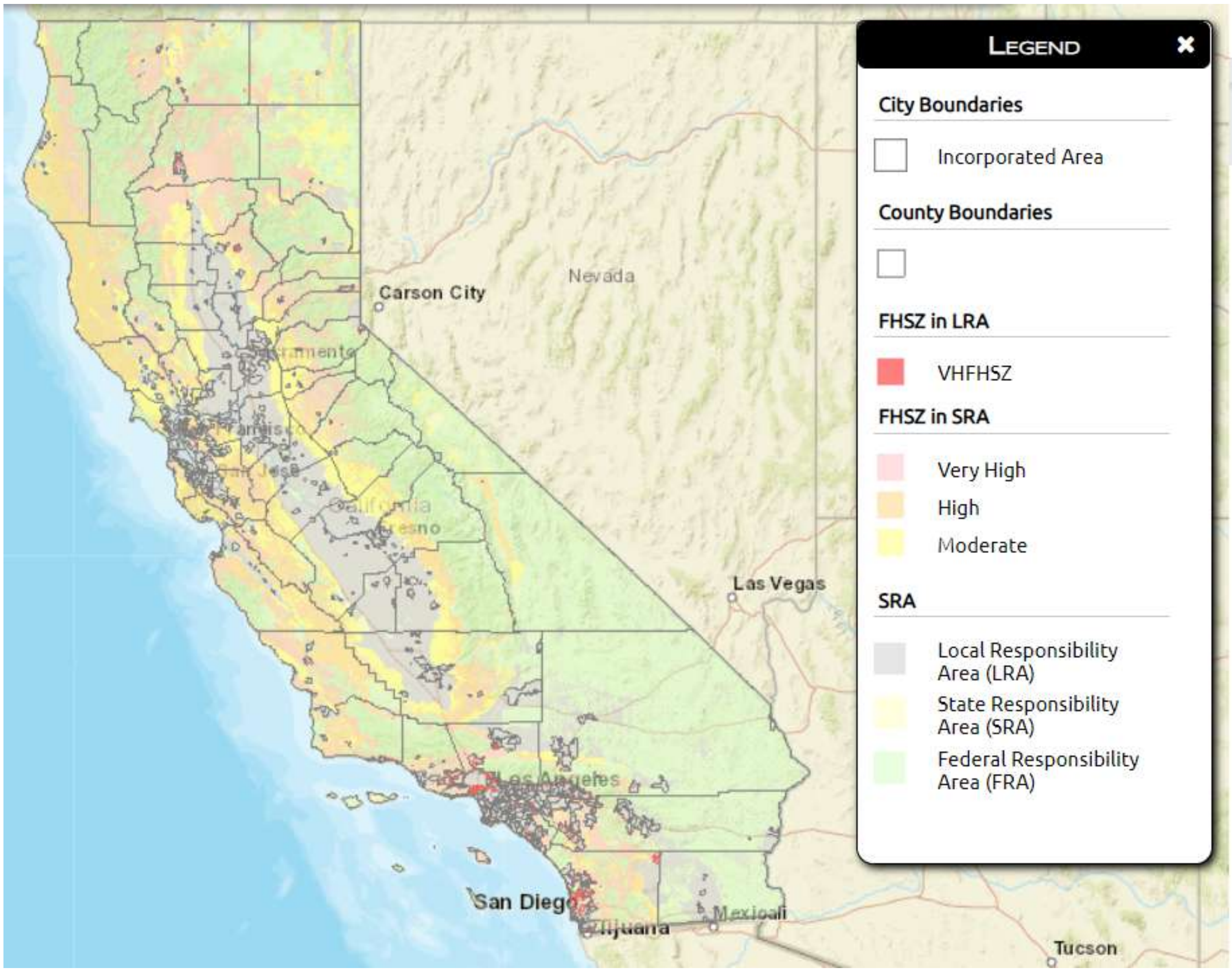

Figure 1. California's Fire Hazard Severity Zones (FHSZ) map.[9]

California has also sought to prevent wildfires through regulations on private properties. Since July 2021, California law requires all homes in High or Very High Fire Hazard Severity Zones (FHSZ) to have defensible space zones. Within a 100-foot perimeter of the home, people must cut their grass shorter than 4 inches, remove fallen leaves and branches, and limit combustible items. [11]

Problem Statement

In this study, we sought to model the locations and severity of historical California wildfires. Using available data, we planned to determine regions with the highest risks of wildfires. With California's existing actions in mind, our goal was to analyze strategies to mitigate the risks of wildfires.

Data Methodology

Most of our data is location-based. Due to the geographic nature of our data, we chose to use ArcGIS.

ArcGIS

ArcGIS is a service providing software and online geographic information system (GIS) that allows for easy handling of geographical data. Data are stored in categories called "layers", and ArcGIS creates complex data maps that can be rapidly visualized.

ArcGIS's parent company Esri has developed and maintains ArcGIS to create a simplified way to display and analyze data. Using ArcGIS, people can create maps with spatial data, which allows users to discover hidden geographical patterns. Then, predictive modeling lets companies and government agencies make better decisions with information about the future. With ArcGIS, people can share maps with others and create maps on any device.

We chose to use ArcGIS so we could display our data in a meaningful and accessible way. With this software, we can efficiently access, analyze, and visualize wildfire data.

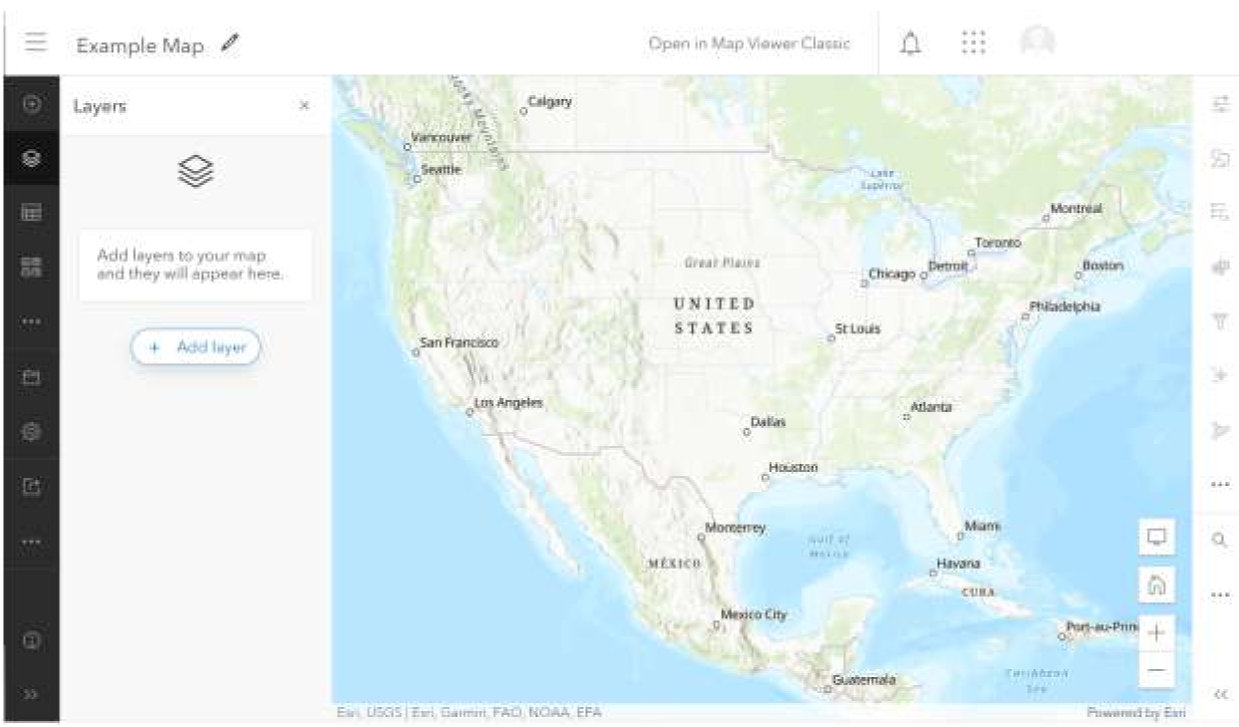

Figure 2. Example map in the software ArcGIS.

Datasets

To examine the historical trends of wildfires in California, we used the fire data from California's Fire and Resource Assessment Program (FRAP). The FRAP has data on wildfires from 1878 to 2020 and is maintained by an Esri ArcGIS File Geodatabase. Specifically, the layer we examined in FRAP was the fire perimeter GIS layer that includes California fires that burned greater than 10 acres and their perimeters. Thus, the FRAP data has the exact boundaries of the recorded wildfires in California for the past 150 years.

The agencies that contribute to the dataset are the Bureau of Land Management (BLM), CAL FIRE, National Park Service (NPS), and the United States Forest Service (USFS). All of these are government agencies that publicly release their data, making FRAP a reliable dataset. These agencies update the FRAP dataset annually to standardize and combine the fire perimeters into a single layer. They identify and remove duplicate fires, fill in gaps on data, and add the previous data to the history layer in ArcGIS.[12]

An additional important dataset we used was the California County Boundaries dataset, provided by the California Department of Forestry and Fire Protection. The California County Boundaries layer contains polygon data for the boundaries of the counties in California. The dataset represents the current counties in California as of June 2019.[13]

We also used 2020 Census data, which contains information in very small areas, like city blocks or neighborhood streets. From this Census data, we found California's total population, the number of occupied housing units in California, and the sections of the state that are land rather than water.[14] Additionally, we used data provided by the California Farmland Mapping and Monitoring Program (FMMP) to analyze data on the location of farmland in California.[15]

Vitally, we found data containing the exact location of fire stations in California. The Fire Stations dataset from the United States Geological Survey (USGS) includes manned fire stations and buildings from which a fire response occurs, like volunteer fire stations. The dataset also includes private and governmental entities and areas where fire stations are known to exist, like military bases and airports. Fire training facilities are not included unless they have an active fire station.[16]

One dataset we would have liked to have but couldn't find was data tracking the exact paths of the wildfires. With this information, we could have analyzed the trends of fire spread and seen what we can do to mitigate it, especially in high-risk areas where fires frequently occur. Another such dataset was historical data on government-issued fire weather watches and warnings, which we could have used to better understand which counties have a higher chance of undergoing a wildfire. Lastly, climate data with high geographical resolution could have helped inform our model when evaluating wildfire risk.

Historical Wildfire Trends

We started by examining the wildfires that occurred in California over the past 150 years using the FRAP dataset.

While data (like a wildfire) may have described multiple counties, we split the data to summarize each county individually for the Census, Farmland (FMMP), Fire Station (USGS), and Fire (FRAP) datasets. This new dataset made the data easier to compare, as only a few counties summarize California instead of using the incredibly granular data contained within the other datasets originally. The results are visible in Figure 3.

Figure 3. With data from the FRAP and California County Boundaries datasets, counties are colored by the amount of area in the county that was burned by wildfires from 1878-2020.

From our map (Fig. 3), we can see the geographical patterns apparent in wildfires. Valleys and deserts have very few wildfires, marked in lighter beige. Certain areas have significantly more wildfires than others, especially the counties in the top left and bottom left that are darkest.

To examine how fires have burned over time, we graphed the reported number of acres burned in California by year since 1878 (Fig. 4). We calculated the sums per year by using the FRAP dataset and importing it into Python. We used the Pandas, CSV, and ArcGIS libraries (see Appendix A). To clean the data, we removed years that were marked as “None” or simply unmarked as “ ”.

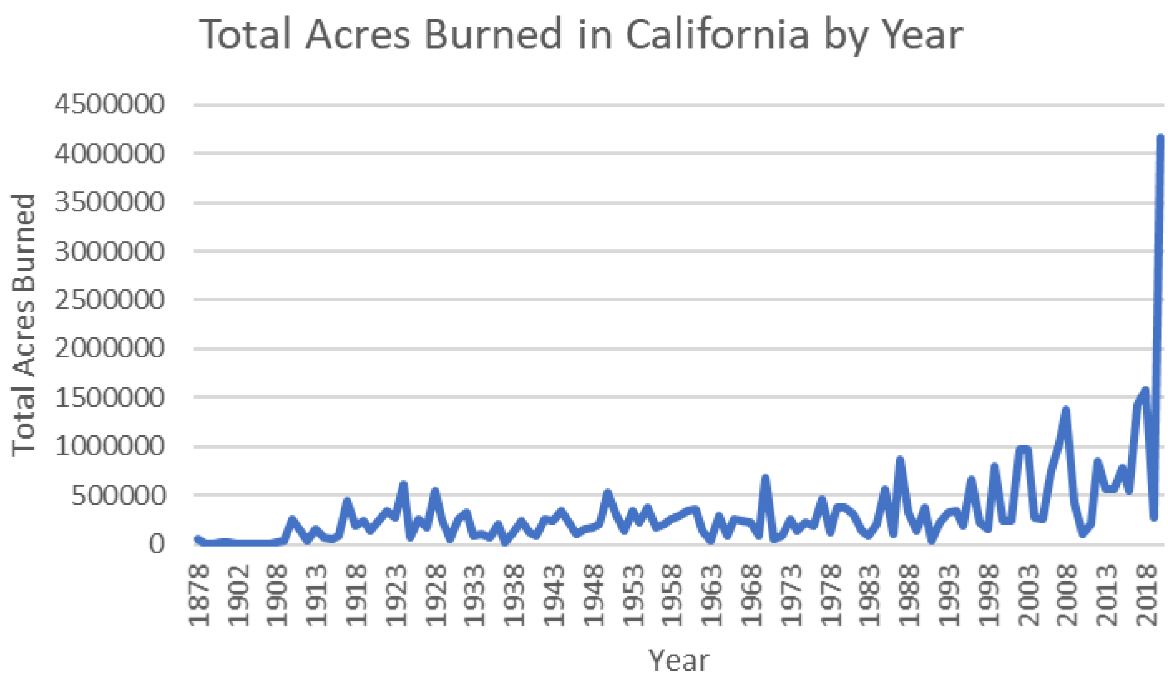

Figure 4. The total acres that fires burned in California from 1878 to 2020 from all causes of wildfires. The data comes from the FRAP.

From 1950 onwards, California has decreased the number of acres required to report a wildland fire to only 10. The criteria changed again in 2002, incorporating all fires greater than 10 acres.[12] However, the graph does not show a noticeable change from either of these changes (Fig. 4). The number of wildfires grew significantly in the early 2000s, and we see a general trend of wildfires burning more acres as time goes on. The year 2018 has the highest number of burned acres, which is consistent with reporting that 2018 was California's deadliest wildfire year on record.

Wildfire risk factors

California wildfires are caused primarily by 19 factors. From the FRAP dataset, we graphed the number of wildfires by their reported cause from the years 2000 to 2020. We decided on this range to look at the more modern data. Additionally, more recent data had better documentation about the causes, as many data points in the 19th and 20th centuries were marked "Unknown".

Wildfires caused by lightning strikes burned the most total acres of land in California, with almost 6 million acres burned (Fig. 5). Just in August 2020, a barrage of lightning in California caused more than 15,000 strikes in just a few days to make 600 fires, burning 2 million acres of land.

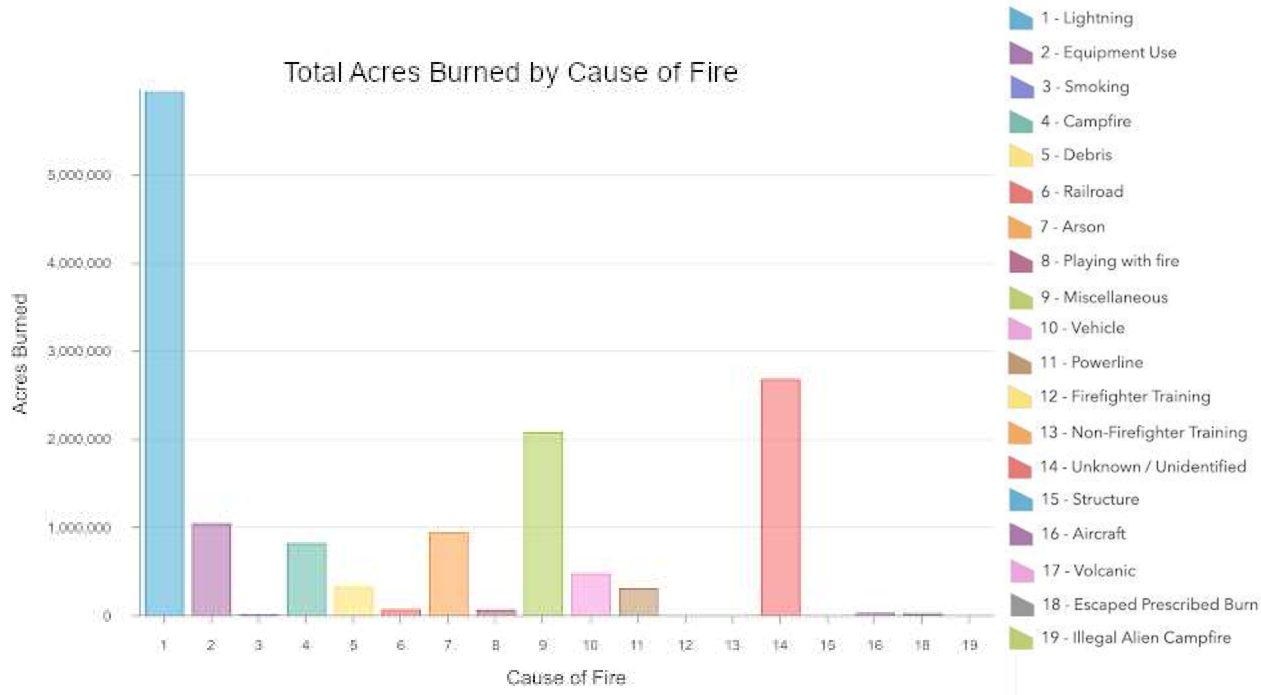

Figure 5. The total number of acres in California burned by each cause of wildfires, from 2000 to 2020. We used California's Fire and Resource Assessment Program (FRAP) dataset.

Worryingly, lightning strikes are increasing in severity and frequency due to climate change. Romps et al. predict that, for every one degree Celsius increase in temperature, the number of lightning strikes will grow by 12%. In the 21st century, the number will grow by 50%.[17]

Lightning is heavily unpredictable. It can strike as far as 20 miles away from the thunderstorm that generated them. During storms, lightning can strike plumbing and conduct electricity throughout the bathroom. Fortunately, cars are safe places for lightning storms as the electric current goes through the metal straight into the ground.[18]

Lightning causes wildfires when rain is absent in "dry" thunderstorms. Strong winds from the thunderstorm turn smoldering organic material into a raging fire.

The material changes from a flameless, smoldering state to a flaming combustion reaction. The combustible components of vegetation are largely cellulose, hemicellulose, and lignin. The wildfire releases smoke that consists of particulate matter, carbon monoxide, organic compounds, nitrogen oxides, and other gasses. Wildfires release particles as large as 2.5 µm (1×10^{-6} meters), which may damage lung function and irritate asthma when inhaled.[19]

$$\text{oxygen} + \text{fuel} \longrightarrow \text{carbon dioxide} + \text{water} + \text{heat}$$

The simplified reaction equation above explains the reactants and products involved in the combustion reaction to make fire. The exothermic reaction releases heat. However, the combustion can frequently be incomplete, releasing carbon monoxide instead of dioxide. Also, many intermediate molecules are involved in the process but are not shown. Having purer oxygen and more fuel drives the reaction towards the products, increasing the combustion and causing the fire to burn at higher temperatures above 1000 degrees Celsius.[20]

The second and third largest causes that burned the most total acres in California from 2000 to 2020 were marked as "Unknown/Identified" or "Miscellaneous" in the FRAP dataset (Fig. 4). These two categories accounted for 3.5 million burned acres. We found that the US Fire Service reported that in the last 10 years, over half (54%) of wildfires were caused by humans while the rest (46%) were started by lightning.[21] We assume that uncategorized causes are primarily human-caused wildfire incidents.

Equipment use has been the fourth-largest cause of wildfires, burning one million acres of California land (Fig. 4). Lawnmowers, weed-eaters, chain saws, tractors, and other yard work equipment can spark a wildfire, especially when the environment is dry and windy. The metal blades from lawn mowers can start a fire by striking rocks and creating sparks.[22]

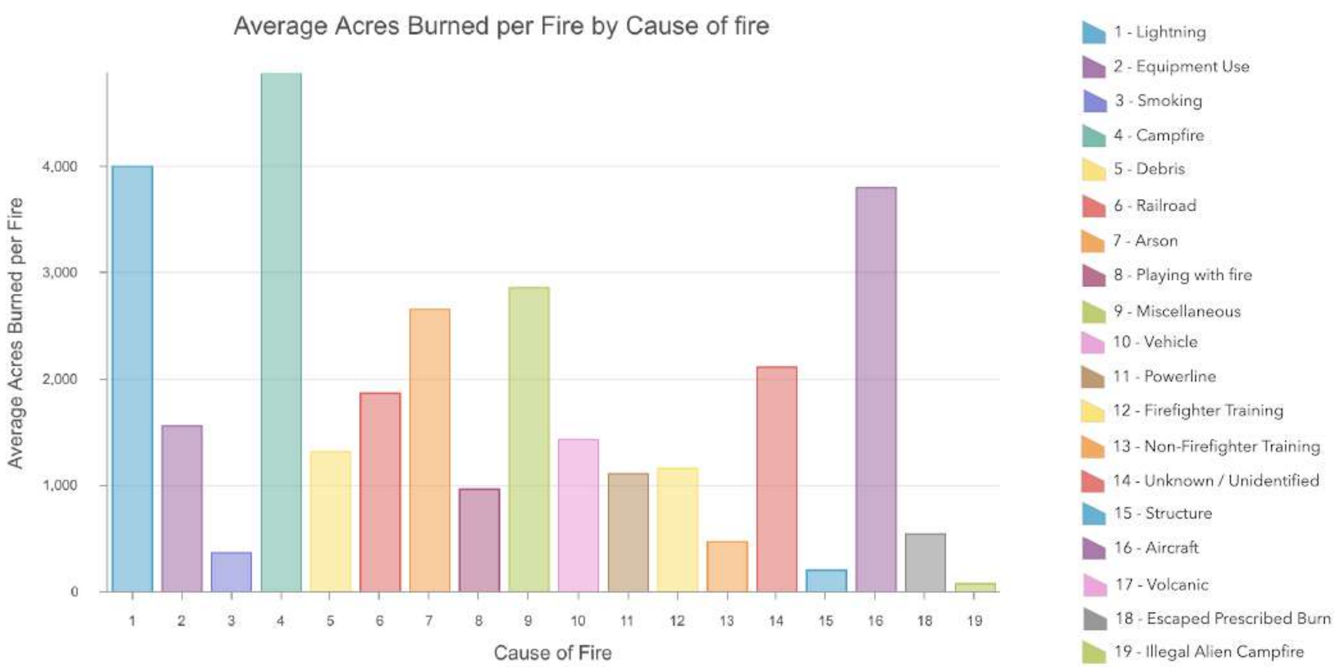

Figure 6. The number of average acres burned by each cause of fire, from 2000 to 2020. We used California's Fire and Resource Assessment Program (FRAP) dataset.

To assess how dangerous each cause of wildfire is, we found the average number of acres each fire cause burned in California from 2000 to 2020 from the FRAP dataset (Fig. 6). We created the chart on ArcGIS.

We found that campfires that started wildfires resulted in the most acres burned on average (Fig. 6). Indeed, campfires that spark fires can be so destructive because they are caused by human negligence, so they are difficult to predict ahead of time. Without an efficient response, the fire grows to cover many acres and firefighters can only focus on keeping the fire contained.

In 2016, an unattended and illegal campfire started the Soberanes Fire, the deadliest and most destructive fire that year. The fire destroyed 57 homes and killed one person. The fire grew to cover multiple acres before witnesses acquired cell phone service and were able to call the police.[23] To prevent such disasters, California state parks and the Bureau of Land Management often require permission in the form of a California Campfire Permit to start campfires on their land.[24]

Mathematical Modeling

After examining the wildfire data, we set out to identify the inefficiencies in the current California wildfire response system and identify how the wildfire risk will change over time. With these pieces of information in hand, we can create a wildfire protection analysis that is both efficient and adaptable.

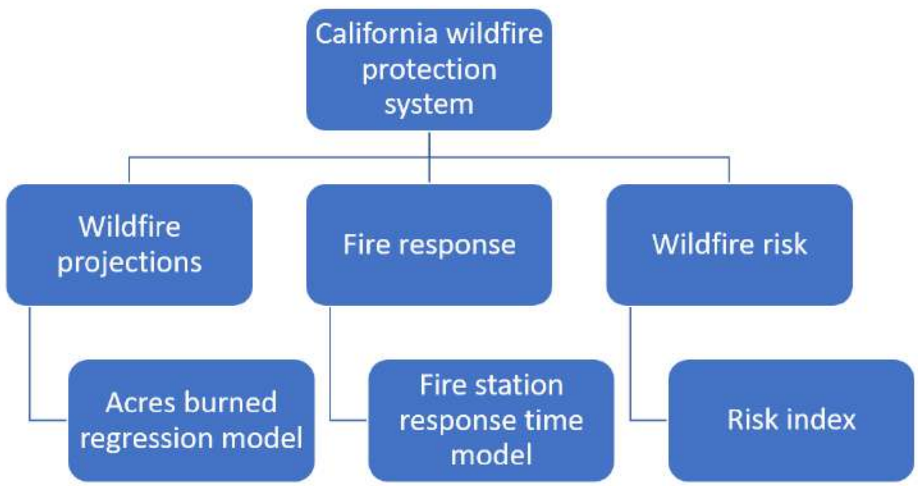

Figure 7. Flowchart describing our models.

The first step is to examine how wildfire activity figures to extend into the future. We performed a pair of regression analyses on data for historical fire extent and used them to understand the potential extent of fire damage in later years.

We also sought to explore the current state of fire response in California. Using ArcGIS, we plotted fire station locations and analyzed the parts of California within a certain travel time from a fire station.

Lastly, we combined a range of fire and Census data into an index for risk and computed this index for every county in the state. The index consists of two parts - a value calculation, which weighs factors such as the number of occupied residential buildings in the county and the amount of farmland, and a risk multiplier, which incorporates the fire history in the county and the accessibility of land in the county from existing fire stations. The index can be modeled to account for future trends, such as the increasing population in California.

Our models work together to create a detailed picture of the wildfire risk in the future. We provide an indication of current trends and response systems, and incorporate these figures to examine which precise areas are at the greatest risk.

Extrapolated Wildfire Activity

We aimed to use FRAP data to create a model for future wildfire activity in California. However, the dataset is missing many years of early historical data and contains sharp peaks and dips as seen in Figure 4. To account for these features, we began our analysis from the year 1916 until 2020 and performed a five-year rolling average on the yearly total acres burned, producing data points from 1918 to 2018. The resultant data was roughly constant to start, with a noticeable increase beginning at the start of the 21st century.

We initially attempted to fit this with an exponential model, but our results were poor with a low coefficient of determination (R^2). In response, we developed a quartic regression model that fit the 100-year data with a coefficient of determination $R^2 = 0.79$.

While the quartic model fits the data nicely, it does not provide great insight into the future, as it predicts a steep increase in acres burned (Fig. 8). Notably, the dataset contains roughly 70 years of mostly constant data followed by 30 years of data that depict increasing fire extent over time. In Figure 8 below, the quartic regression result is represented in red, and the data it extrapolates from is shown in blue.

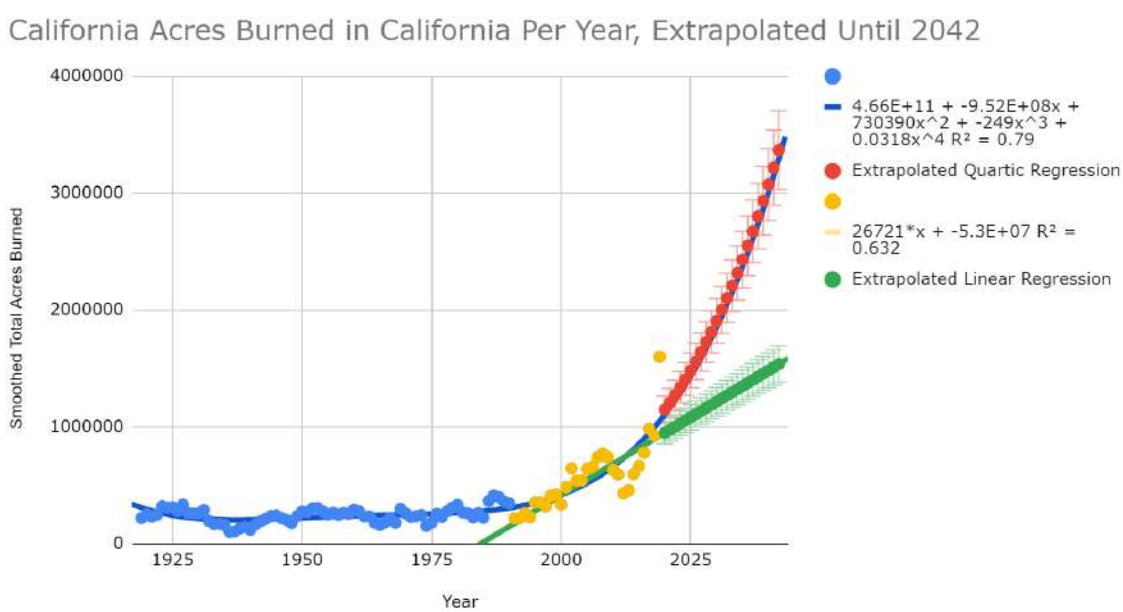

Figure 8. The projected number of burned acres in California extrapolated 20 years in the future to 2042, using data from 1918 to 2018 in the FRAP dataset.

To better model this increase, we isolated only the last 30 years of data and performed a simple linear regression. This regression also fits the data well, with a coefficient of determination $R^2 = 0.632$. In Figure 8, the linear regression result is represented in green, and the data it extrapolates from is shown in yellow.

Before discussing trends, it is worthwhile to note the appearance of the year 2018, the final year included on the graph. As the data are five-year rolling averages, the number for 2018 is an average of raw data from 2016 to 2020. In Figure 8, it appears as a yellow dot well above all the other dots. The final year in the FRAP dataset is 2020, and it was a catastrophic year for wildfire in California. More than four million acres of land burned throughout the year, a figure more than double the previous record. Additionally, wildfires in 2018 were the most destructive and deadliest on record in California.[25] As a result, the smoothed data for 2018 appears well out of line with the rest of the data. As 2020 is the last year we have on record, it is difficult to establish whether it is an outlier or an indication of continually increasing wildfire severity.

Overall, Figure 8 shows a very alarming trend. Both the polynomial model and the linear model display clear and dramatic increases over time in the extrapolated number of acres burned from fire. They indicate that the wildfire problem in California may get much worse in the coming years unless protection methods change markedly.

This model makes several assumptions about the state of wildfire conditions in California. First, we assume that the wildfire data can be modeled with a polynomial or linear equation, and that this modeling will be predictive. After both 1950 and 2002, the minimum acre threshold for wildfires in the dataset was lowered, which we assumed to have a negligible impact. Lastly, we assumed that the huge spike in wildfire activity in 2020 was not an outlier, but rather a piece of meaningful data. In reality, it is too soon to tell.

ARIMA Models

We continued to make predictions about the total acres burned in California by using autoregressive integrated moving average (ARIMA) models. The ARIMA model is defined by three parameters: p, d, and q, which account for the number of automatic regression terms, differences to make the model stationary, and moving average weights for error respectively.

We created ARIMA models because they don't assume that observations are all independent of each other, like linear regressions do. Instead, ARIMA models consider that observations are time dependent, where previous lags in the data are also used as predictors. As we are analyzing the number of acres burned over time, we found the ARIMA model to be appropriate.

To find the optimal parameters of p, d, and q, we used the pmdarima library in Python (see Appendix B). Then, we used the auto_arima function to find the parameters that minimize the Akaike's Information Criterion (AIC), an estimator for the error a mode will have for extrapolating data.[26] We ran this function on two sections of our dataset: number of acres fires have burned from 1920-2020 and from 2000-2020 to focus on long-term and short-term trends.

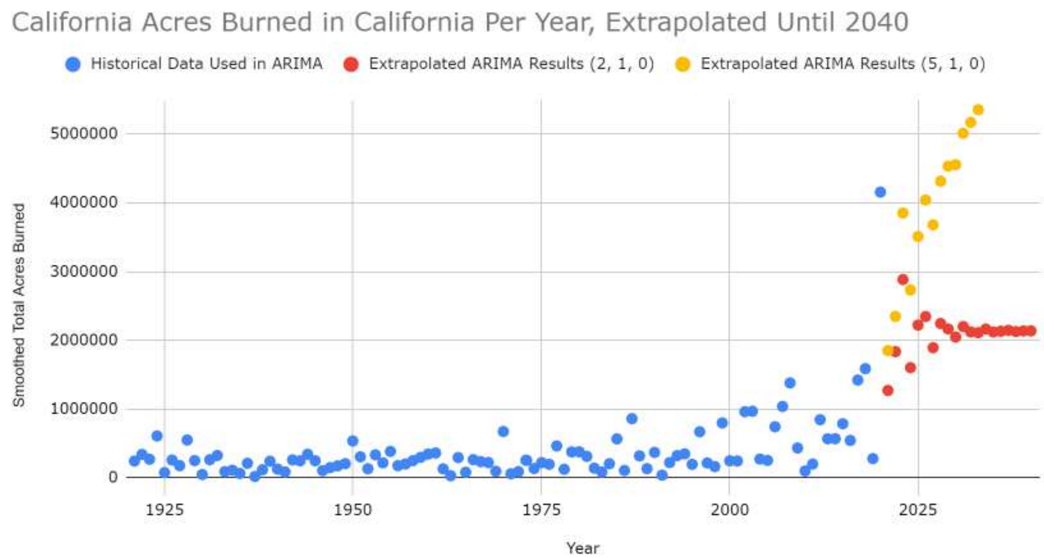

Figure 9. The predicted number of burned acres in California to 2040 based on FRAP data from 1920-2020. Two ARIMA models using different values of p, d, and q are shown.

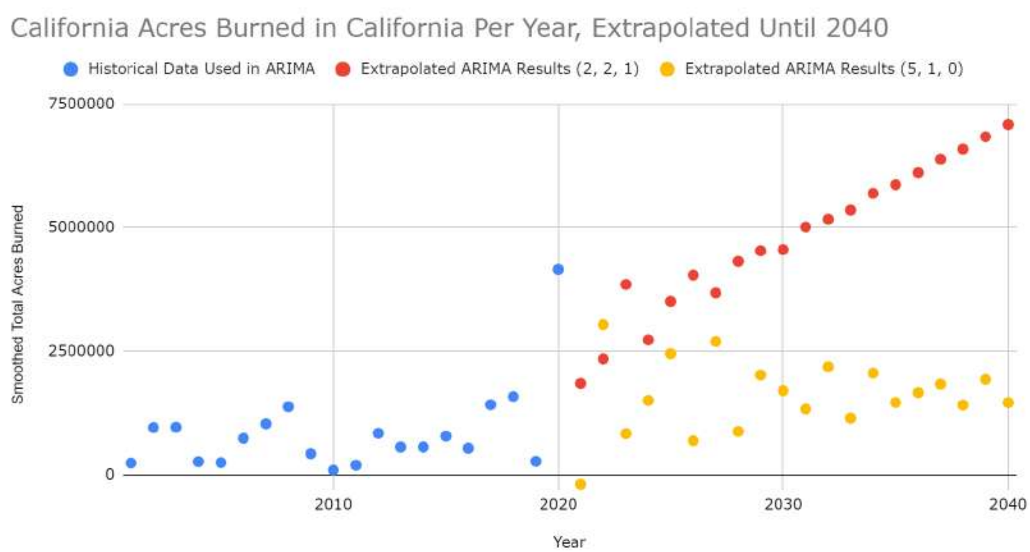

Figure 10. The predicted number of burned acres in California to 2040 based on FRAP data from 2000-2020. Two ARIMA models using different values of p, d, and q are shown.

The auto_arima function (Appendix B) displayed the optimal p, d, and q values for the ARIMA to be 2, 1, and 0 for the 1920 to 2020 dataset (Fig. 9), but gave 2, 2, and 1 for the 2000-2020 dataset (Fig. 10) where there is more variability to account for. We saw in our models that using a higher p value increased the variability in the results, seen by how the red line converged to a more linear pattern while the yellow points are scattered.

The ARIMA models assume that past performance will predict future outcomes. Should any large natural disasters or catastrophes occur, the model would not be accurate in those cases. Additionally, we assumed that the spike of burned acres in 2018 was accurate and not an outlier. Nevertheless, all the models we found predicted an increase in the number of acres burned in California, should current trends continue.

We found the model that looked the most predictive was based on the data from 2000 and onwards, with a p, d, and q of 5, 1, and 0 (Fig. 10, yellow). Unlike the red lines that converged to lines, the yellow points included variability that was like previous years. Then, because we used more recent data, the predictions look more normal and predictive.

In all the ARIMA models, the predicted number of acres burned exceeded 2.5 million acres in a single year. This is concerning when the highest number of acres burned in a single year has never exceeded 2 million, showing that we need to implement mitigation strategies for the risks of wildfires to prevent them in the future.

Fire Station Reach Model

We modeled the places a fire station would be able to reach within a certain amount of time using the Fire Stations dataset from the USGS (Fig. 11). The USGS dataset includes all types of active fire stations, from those in airports to volunteer fire departments. The areas shown in purple change depending on what time one indicates. For example, 11a shows the areas a fire station could reach in 15 minutes, while 11g shows the areas a fire station could reach in 30 minutes. Our model can show any response time (down to seconds).

Figure 11a

Figure 11b

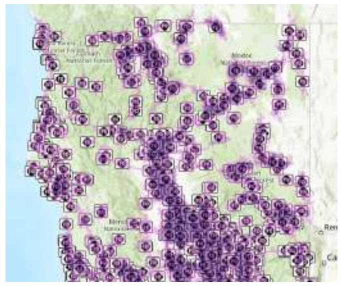

Figure 11c

Figure 11d

Figure 11e

Figure 11f

Figure 11g

Figure 11h

Figure 11. Northern California areas are shaded based on response time from a fire station. Figures 11a and 11b show areas with arrival times of 15 minutes or less, Figures 11c and 11d show arrival times of 20 minutes or less, Figures 11e and 11f have arrival times of 25 minutes or less, and Figure 11g and 11h show arrival times of 30 minutes or less. Figures 11b, 11d, 11f, and 11h include labels for fire stations with "FD". Green areas are those that are unreachable from fire stations within that time.

Our model, which we named the Fire Station Reach Model, shows the places that are protected by fire stations in case a wildfire occurs. We calculated the areas using Dijkstra's algorithm based on road paths and distance from all California fire stations, as provided by the USGS dataset, in ArcGIS. The Fire Station Reach Model can also take into account different times of day using a traffic dataset, so that the middle of the night does not give the same distances as peak rush hour.

We found the optimal fire station response time to be less than 15 minutes. Indeed, the National Fire Protection Association (NFPA) recommends a response time of less than 560 seconds (9 minutes and 20 seconds) for an immediately life-threatening fire incident.[27] Thus, we showed areas that could be reached in less than 15 minutes in Figure 11, though our model can show any requested time.

In the Fire Station Reach Model, we made the following assumptions:

1. We assumed there were no roadblocks or other barriers slowing down fire response time, like car accidents or fallen trees.
2. We assumed traffic is constant, though our model can use the traffic dataset in ArcGIS depending on what time is requested (e.g., "7 a.m.").
3. We assumed California only responds to fires from active fire stations.
4. We assumed all the fire stations functioned normally, without issues like lack of personnel or equipment.

The Fire Station Reach Model responds to trends as it is easily updatable with more fire stations. If more fire stations are added in the future, the model can import that information.

FIRES (Fire Index Risk Estimator Score)

We wanted to find the areas that would be most at risk to a wildfire. Risk is calculated by multiplying the expected loss by the probability of the loss happening. We organized the Fire Index Risk Estimator Score (FIRES) by county using the California Counties Boundary dataset. We used counties as they have the ability to make local laws and ordinances changing their approaches to wildfires. We also accounted for the size of counties, so larger counties do not have a larger risk just due to size.

Assumptions

Although the FIRES model incorporates many different risk factors to wildfires and potential damage areas, its framework makes many assumptions.

1. We assumed that the value of farmland does not change across California
2. We assumed that the value of land does not change across California.
3. We assumed that there is no variation (e.g. age) within California's population that will significantly impact the risk from one region to another.
4. We assumed that the value of housing units does not change across California.
5. We assumed that the damage done to the water area by wildfires is negligible.
6. We assumed that factors other than the ones included in our model do not impact the wildfire risk.

Expected Loss

We first chose the variables most important to a loss function. Fires have a large impact on California's agriculture industry and the Californian population. Thus, we included Census data containing statistics on the total population and number of occupied housing units to help calculate the loss it would have on people.

Since fires have destructive impacts on the environment, we included the percentage of the California county that was farmland, as provided by the FMMP dataset. Additionally, we used Census data to find the percentage of land in a county. Some counties contained large bodies of water, and we did not wish to include the water in the area burned calculations.

We found the expected loss with the following formula:

$$\frac{Land\ Area}{Land\ Area + Water\ Area} + \frac{1}{10}\sqrt{ln(Population\ +\ 1)\ \cdot\ ln(Occupied\ Housing\ Units\ +\ 1)} + \frac{Farm\ Area}{500}.$$

Burn Chance

Besides expected loss, the FIRES model incorporates the probability of the loss happening, which we called Burn Chance.

We found the Burn Chance with the following formula:

$$\frac{Area\ Burned\ by\ Fires - Area\ within\ 15\ min.\ of\ a\ Fire\ Station}{2 \cdot Total\ Area} - ln(\frac{Fire\ Stations}{Total\ Area})$$

We first found the percentage of acres burned in each county from the FRAP dataset. Then, we incorporated the results of our Fire Station Reach Model, where we used the percentage of area in each county that had a response time of fire station of less than 15 minutes.

We combined these two values with the following formula to create an index for wildfire risk.

$$Wildfire\ Risk\ =\ Expected\ Loss\ \cdot\ Burn\ Chance$$

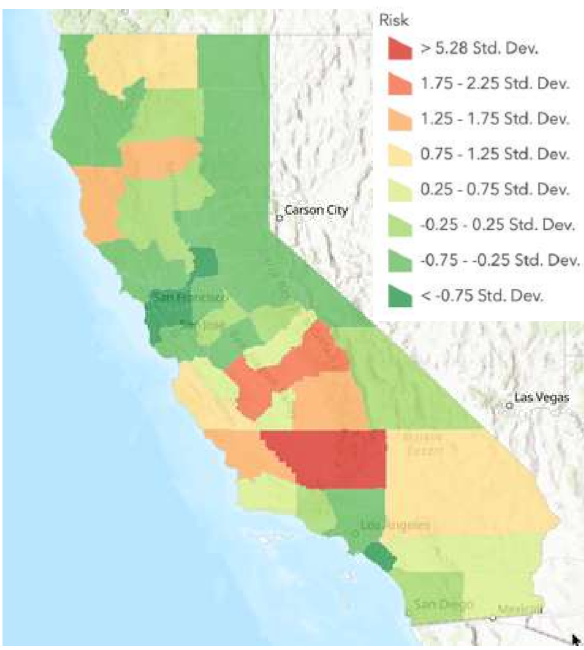

Figure 12. We used the Fire Station Reach Model and the California County Boundaries, FMMP, Census, and FRAP datasets to calculate the FIRES model. Then, we mapped the risk in terms of standard deviations to analyze with maps that had the highest risk.

Our FIRES model suggests particular locations with the most dangerous combination of wildfire likelihood and expected loss from wildfires. The counties are, in least to greatest order of impact, Tulare County, Mendocino County, Tehama County, Fresno County, and Kern County. Farmland-dense Kern County has the greatest risk by a large margin. Kern and Fresno counties also have relatively high populations and therefore larger potential cost in damage to housing units. All of these counties are split between vulnerable forest areas and valuable farmland, and many are fairly large, so a potential fire has more potential for damage.

Table 1.

Five Lowest and Highest Risk Counties

County Name	**Population**	**Area (sq. mi)**	**Expected Loss**	**Burn Chance**	**Risk Z-Score**	**FIRES Score**
Kern	909,039	8,163	198	75	5.24	🔥🔥🔥🔥🔥
Fresno	1,008,604	6,010	100	69	2.11	🔥🔥🔥🔥
Tehama	65,835	2,958	74	82	1.73	🔥🔥🔥🔥
Mendocino	91,487	3,512	83	68	1.64	🔥🔥🔥🔥
Tulare	473,100	4,843	67	78	1.44	🔥🔥🔥🔥
...	...	...	...	...	...	...
Contra Costa	1,165,896	803	23	3	-0.80	🔥
Alameda	1,682,348	822	23	2	-0.80	🔥
Orange	3,186,601	799	22	2	-0.81	🔥
San Mateo	764,070	552	10	5	-0.90	🔥
San Francisco	873,818	106	1	-34	-1.22	🔥

The five lowest and highest risk counties, displayed with their Expected Loss and Burn Chance scores, both normalized to an outlier-proof 0-100 scale. We also included their Z-Scores for risk, as well as their risk score out of 5🔥. The full table appears in Appendix C.

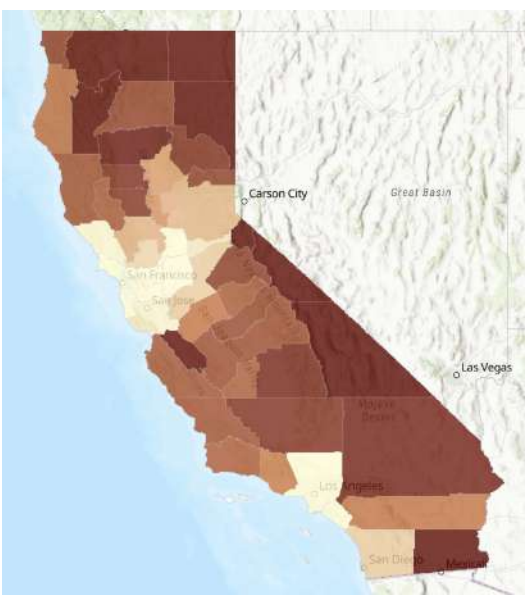

Figure 13. Changes in FIRES model due to wildfire damage over the next 50 years, considering projected population increase (darker indicates more change).

The FIRES model also responds to trends to account for the change in the index over time. It assumes that the projected population increase in California over the next fifty years is the only change in its inputs. The California Department of Finance predicted this population change to be 0.5% per year.[28] The model could also change with other variables, such as land area if bodies of water dried up or farmland if it was developed into urban land.

Recommendations

Our different investigations into the causes and greatest potential losses lead to four main categories of recommendations to mitigate the future damage from California wildfires, which we have shown will very likely be significantly more than it has been historically. We recommend strategies to mitigate the impact of lightning strikes from our investigation of fire causes. Our geographic risk model leads to recommendations for potential new locations for fire stations and sensor networks, two effective large-scale approaches for quickly reacting to new wildfires, especially in remote areas. Finally, we discuss the age-old strategy of fire insurance for homeowners themselves to alleviate their risk.

Lightning

We found that lightning burned the most acres of all wildfire causes in California historically. Lightning rods are one of the most commonly used lightning protection devices. They consist of a metallic (usually copper) rod affixed to the highest point of a building and connected to a cable. When lightning strikes the building, the rod guides the charge into the soil to dissipate harmlessly. While lightning rods are very effective for houses and small structures, they offer much less protection for tall structures, such as trees.[29]

A lightning rod more than 30 meters taller than the surrounding forest provides lightning protection in a 30-meter radius. Thus, protecting just one square kilometer of forest would require several hundred lightning rods. The sheer quantity of lightning rods required to protect forests makes it a costly and difficult-to-maintain strategy against lightning.[29]

Proper lightning rods also provide surge protection for power, telephone, and cable lines. Since faulty power lines can also start wildfires, lightning rods would account for two causes of wildfires at once. Thus, lightning rods should be used to protect residential homes and historical sites.[30]

The cost of installing a copper lightning rod in the western United States is $0.80 per square foot of floor in a residential building (Table 1).[31] With 14 million houses in California,[32] and each house an average of 1,625 square feet,[33] the total cost of installing lightning protection would be $18,200,000,000 or more than $18 billion.

Fortunately, for an average house, a lightning rod installation costs only $1,300. Thus, installing a lightning rod is a feasible behavior change inhabitants can implement to prevent lightning from causing their homes to catch fire.

Table 2

Lightning Protection Installation Cost Estimates

LIGHTNING PROTECTION INSTALLATION COST ESTIMATES	RESIDENTIAL BUILDING		LOW-RISE BUILDING		5-STORY BUILDING	
	Aluminum	Copper	Aluminum	Copper	Aluminum	Copper
NORTHEAST						
$/Sq.Ft. of Roof	$1.56	$1.58	$0.74	$0.89	$1.75	$1.89
$/Sq.Ft. of Floor	$0.94	$0.95	$0.54	$0.65	$0.35	$0.38
SOUTH						
$/Sq.Ft. of Roof	$0.98	$1.10	$0.42	$0.50	$1.16	$1.33
$/Sq.Ft. of Floor	$0.59	$0.66	$0.31	$0.37	$0.23	$0.27
MIDWEST						
$/Sq.Ft. of Roof	$0.88	$1.06	$0.78	$1.02	$1.45	$1.82
$/Sq.Ft. of Floor	$0.53	$0.64	$0.58	$0.75	$0.29	$0.37
WEST						
$/Sq.Ft. of Roof	$1.60	$1.77	$0.88	$1.04	$1.46	$1.61
$/Sq.Ft. of Floor	$0.96	$1.06	$0.65	$0.76	$0.29	$0.32
NATIONAL						
$/Sq.Ft. of Roof	$1.18	$1.34	$0.65	$0.78	$1.44	$1.64
$/Sq.Ft. of Floor	$0.71	$0.80	$0.48	$0.58	$0.29	$0.33

Cost of protecting sitework, such as trees, is not included.

Note. Reprinted from "Lightning Protection Installation Cost Study" by Chusid, M., (2015). Retrieved from https://ecle.biz/coststudy/.[31]

Also, merely installing surge protectors costs an affordable $100 per house, which amounts to $1.4 million for the residential homes in California.[34] This could prevent fires caused by power lines, which burned a quarter-million acres from 2000 to 2020 in California (Fig. 5).

Fire Stations

Building new fire stations in particularly exposed regions is another effective measure to mitigate damage from wildfires. A fire station is a structure that stores firefighting technology, such as fire engines, protective equipment, and fire hoses. Fire stations respond to emergency calls, dispatch on-site staff, and provide working and living space for firefighters.

As California grows hotter and drier and wildfires become increasingly common, more funding is available for wildfire prevention projects. For example, President Biden's bipartisan infrastructure bill from November 2021 included $3.3 billion for "reducing the risk of wildfire". Among other wildfire-related concerns, the bill increases the pay for federal firefighters.[35][36]

With the support of the federal government and local communities, there are opportunities for larger and more expensive fire protection efforts. In March 2022, the San Francisco Fire Department completed the two-year-long construction of a $50 million floating fire station specifically for fires caused in the case of an earthquake.[37] Evident from the massive project, there is a particular interest in fire stations.

Fortunately, the construction cost of a fire station typically ranges from $500,000 to $1 million, a figure easily offset by protecting against the billions of dollars in losses from large wildfires.[38][39] We therefore recommend new fire stations to be built in California, which can be potentially covered by Governor Newsom's 2022 Five-Year Infrastructure Plan, where nearly $1 billion are allotted especially for wildfire protection.[40]

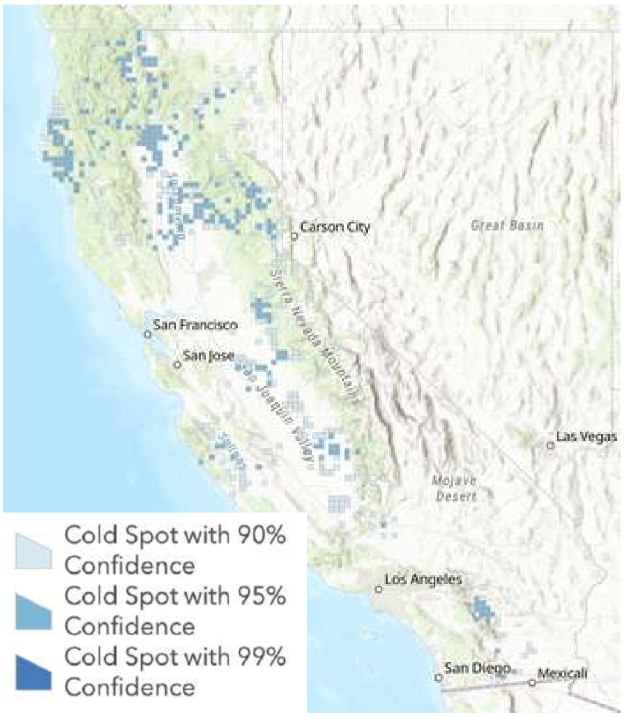

Figure 14. Gridlock of California with statistically significant areas marked as Cold Spots. We used the results of the Fire Station Reach Model (Fig. 11) and the Fire Stations dataset from the USGS.

Using the Getis-Ord Gi* statistic on ArcGIS[41] (Fig. 15) and the results of the Fire Station Reach Model, we created a gridlock of California that shows the best places to add more fire stations, marked in blue (Fig. 14).

Cold spots with 99% confidence are the places we recommend adding more fire stations, especially in northern California and southeast of Los Angeles. These blue locations would be optimal for the construction of new fire stations, as they would have the greatest impact on previously unprotected areas.

Fire stations are important because they modify the outcomes of wildfires. Since fire stations also directly benefit environmental protection and are increasingly being used as community centers,[42] their main costs are financial. Based on Figure 14, we should add approximately 10 new fire stations in the blue areas, which would cost $10 million at the higher range.

The Getis-Ord local statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j} \right)^2 \right]}{n-1}}} \tag{1}$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{2}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\bar{X}\right)^2} \tag{3}$$

The G_i^* statistic is a z-score so no further calculations are required.

Figure 15. Getis-Ord Gi* statistic formula for features (in our case, fire stations) in the dataset. The results are statistically significant when the local sum of an area is different from its neighbors. The link to the graphic: https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm[41]

IoT Sensors

A different, state-of-the-art strategy for wildfire detection would be to deploy sensors in vulnerable areas of California's wildland. Connected with broadband wireless networks or specialized satellites, thermal and visual sensors in the right place provide near-real-time warnings for fire sparks. Sensors measure smoke, CO_2 levels, humidity, and other factors to determine that a fire has ignited. Placing sensors in the wildland would allow for the detection of wildfires in their earliest stages.[43]

To get information back to fire stations, sensors can use the Internet of Things (IoT). The IoT describes the communication of physical devices, typically over the internet or a wireless network. Historically, IoT sensor systems have been unfeasible because the Internet of Things has been dependent on short-range Wi-Fi and Bluetooth. However, with the emergence of extremely power-efficient Arduino microcontrollers and improved wireless technologies like Zigbee, it is possible to collect data from a network of sensors scattered throughout the wildland.[44] Rather than costly and potentially manned towers, it may even be possible to connect the sensors through satellites.[43]

Each sensor has a range of up to 2 square kilometers. Due to communication with satellites, a maximum of 32,400 sensors can report each minute. Multiplying the two numbers, sensors alone can cover 64,800 square kilometers. Thus, 15.3% of California's total area of 423,960 km^2 could be covered with granular, near-real-time wildfire warnings.[45][43] This sensing could even be fine-tuned based on changing wildfire risk data at each location.[46]

At around $50 per sensor,[46] the coverage described above would cost less than $1.7 million. Simulations suggest that this cost, combined with the cost of the carbon emissions from the deployment of the sensors, would be greatly outweighed by the damages saved from potential burn time reduction based on CAL FIRE historical data. Indeed, the impact of using IoT sensors would result in a net benefit of approximately 1.07 billion USD in the first year.[43]

The FIRES model suggests particular California counties with the greatest risk, or multiplied wildfire likelihood and expected loss from wildfires, where IoT networks would make the largest impact: Tulare, Mendocino, Tehama, Fresno, and Kern counties, in increasing order of impact.

Insurance

Another effective strategy for people in high-risk wildfire areas is to invest in home insurance. Insurance is a contract or policy where a person or group receives financial protection from an insurance company. People typically pay a certain small amount monthly so that, if an emergency happens, the insurance company covers the larger losses.

Most standard homeowners insurance policies cover fire damage, including that from wildfires. One popular type of insurance is dwelling coverage, which covers the cost of rebuilding or replacing the physical structure of the home. However, the limit the insurance covers will depend on how much of your property you insure beforehand (e.g., 60% of a property).[47]

Unfortunately, some insurance companies have been dropping homeowners who live in areas affected by major wildfires. California imposed a one-year ban on those insurance companies. While insurers are looking to reduce their risks of wildfires, it may be a challenge in wildfire areas where fewer insurers make their coverage available.[48]

The average home insurance premium is $1,224 a year in California.[49] Still, the price will depend on one's area and home size. With 14 million homes in California,[32] insuring every home will cost $17,136,000,000. Ultimately, this massive cost will fall on private homeowners, which further justifies the overall cheaper prevention and mitigation strategies above.

References

1. Associated Press. (2019, August 8). *Camp fire death toll rises to 86 after man dies of burn injuries*. Los Angeles Times. https://www.latimes.com/california/story/2019-08-08/camp-fire-death-toll-rises-to-86-after-man-dies-of-burn-injuries
2. Castronuovo, C. (2021, October 1). *1 in 8 acres in California has burned in last decade*. The Hill. https://thehill.com/policy/equilibrium-sustainability/574928-1-in-8-acres-in-california-has-burned-in-last-decade
3. Arcuni, P. (2021, June 22). *Millions of California homes are in fire-prone areas. Researchers say it's time to reimagine where people live*. KQED. https://www.kqed.org/science/1975443/millions-of-calif-homes-are-in-fire-prone-areas-researchers-say-its-time-to-reimagine-where-they-live
4. Boxall, B. (2020, December 23). *Billions of dollars spent on fighting California wildfires, but little on prevention*. Los Angeles Times. https://www.latimes.com/environment/story/2020-12-23/billions-spent-fighting-california-wildfires-little-on-prevention
5. Schweizer, D. (2021, July 29). *Wildfires in All Seasons?* United States Department of Agriculture. https://www.usda.gov/media/blog/2019/06/27/wildfires-all-seasons
6. Jiang, Tanqiu et al. (2021, December 15). *From static to dynamic prediction: Wildfire risk assessment based on multiple environmental factors.* 2021 IEEE International Conference on Big Data (Big Data). Retrieved from https://doi.org/10.1109/bigdata52589.2021.9672044.
7. Wang, D., Guan, D., Zhu, S., Kinnon, M. M., Geng, G., Zhang, Q., Zheng, H., Lei, T., Shao, S., Gong, P., & Davis, S. J. (2020). Economic footprint of california wildfires in 2018. *Nature Sustainability*, *4*(3), 252-260. https://doi.org/10.1038/s41893-020-00646-7
8. *California Fire Plan*. (n.d.). California Department of Forestry and Fire Protection. Retrieved March 11, 2022, from https://www.fire.ca.gov/media/4934/fireplan.pdf
9. *Fire Hazard Severity Zones*. (n.d.). Office of the State Fire Marshal. https://osfm.fire.ca.gov/divisions/community-wildfire-preparedness-and-mitigation/wildfire-preparedness/fire-hazard-severity-zones/
10. *California wildfire public policy: Mapping the wildfire hazard*. (2021) Insurance Institute for Business & Home Safety. https://ibhs.org/wp-content/uploads/wpmembers/files/IBHS_Wildfire-Mapping-and-Building-Codes_2021.pdf
11. *Defensible Space*. (n.d.). California Department of Forestry and Fire Protection. https://www.fire.ca.gov/dspace/
12. *Fire Perimeters*. (n.d.). California Department of Forestry and Fire Protection's Fire and Resource Assessment Program. https://frap.fire.ca.gov/frap-projects/fire-perimeters/
13. California Department of Forestry and Fire Protection. (2019, October 16). *California County Boundaries*. ArcGIS Online. https://services1.arcgis.com/jUJYIo9tSA7EHvfZ/arcgis/rest/services/California_County_Boundaries/FeatureServer

14. U.S. Census Bureau. (2020, August 12). *USA Census 2020 Redistricting Blocks.* ArcGIS Online. https://services1.arcgis.com/jUJYIo9tSA7EHvfZ/ArcGIS/rest/services/California_County_Boundaries/FeatureServer/
15. Farmland Mapping and Monitoring Program. (2014, January 7). *CA Farmland Mapping and Monitoring Program.* ArcGIS Online. https://services.arcgis.com/jDGuO8tYggdCCnUJ/arcgis/rest/services/CAFarmlandMappingandMonitoringProgram/FeatureServer
16. United States Geological Survey. (2020, April 20). *Fire Stations*. ArcGIS Online. https://services1.arcgis.com/Hp6G80Pky0om7QvQ/arcgis/rest/services/Fire_Station/FeatureServer
17. Cart, J. (2021, September 24). *Lightning could spark more California fires as world warms*. CalMatters. https://calmatters.org/environment/2021/09/california-fires-lightning/
18. Rivera, T. (2021, June 17). *Lightning can strike 20 miles away from the storm...that's Kernersville to Greensboro!* WFMY News 2. https://www.wfmynews2.com/article/news/local/2-wants-to-know/lightning-can-strike-20-miles-away-from-the-stormthats-kernersville-to-greensboro-myths-vs-facts/83-e58076c5-a70c-4451-9ee6-917077ef4bdc
19. Brunning, A. (2018, January 15). *Periodic graphics: The chemistry of wildfires*. Chemical & Engineering News, American Chemical Society. https://cen.acs.org/articles/96/i3/Periodic-graphics-chemistry-wildfires.html
20. *What is fire? Reaction that converts fuel and oxygen into carbon dioxide and water*. (n.d.). New Scientist. https://www.newscientist.com/question/what-is-fire/
21. *Safe and Effective Wildfire Response*. (n.d.). Forest Service, U.S. Department of Agriculture. https://www.fs.usda.gov/managing-land/fire/response
22. *Equipment Use*. (n.d.). CAL FIRE Ready for Wildfire. https://www.readyforwildfire.org/prevent-wildfire/equipment-use/
23. Larson, A. (2016, August 5). *Soberanes Fire caused by illegal unattended campfire, Cal Fire says*. KSBW. https://www.ksbw.com/article/soberanes-fire-caused-by-illegal-unattended-campfire-cal-fire-says-1/1297793
24. *Passes & Permits*. (n.d.). Forest Service U.S. Department of Agriculture. https://www.fs.usda.gov/detail/sequoia/passes-permits/
25. *2018 California Wildfires*. (2018). U.S. Census Bureau. https://www.census.gov/topics/preparedness/events/wildfires/2018-ca-wildfires.html
26. Zajic, A. (2019, December 27). *Introduction to AIC — Akaike information criterion.* Towards Data Science. https://towardsdatascience.com/introduction-to-aic-akaike-information-criterion-9c9ba1c96ced
27. *Fire Response Time*. (n.d.). DC Fire and Emergency Medical Services Department. https://fems.dc.gov/page/fire-response-time
28. Schwarm, W. (2020, June 11). *Recent trends and projections.* California Department of Finance. https://scag.ca.gov/sites/main/files/file-attachments/walterschwarm.pdf

29. The Editors of Encyclopaedia Britannica (2010, August 13). *Lightning rod.* Encyclopedia Britannica. https://www.britannica.com/technology/lightning-rod
30. Miley, M. (2021, May 19). *How do lightning rods work — and do I need them?* Hunker. https://www.hunker.com/13764754/how-do-lightning-rods-work-and-do-i-need-them
31. Chusid, M. (2015, July). *Lightning Protection Installation Cost Study*. East Coast Lightning Equipment. https://ecle.biz/coststudy/
32. U. S. Census Bureau. (2021, July 1). *California QuickFacts.* United States Department of Commerce. https://www.census.gov/quickfacts/fact/table/CA/PST045221
33. Rider, R. (2020, August 30). *Compared to other states, CA has uber-high median home prices — coupled with much smaller lots and smaller houses*. San Diego Rostra. https://sdrostra.com/compared-to-other-states-ca-has-uber-high-median-home-prices-coupled-with-much-smaller-lots-and-smaller-houses/
34. HomeAdvisor. (2022, January 10). Install Lightning Protection. HomeAdvisor. https://www.homeadvisor.com/cost/environmental-safety/install-lightning-protection/
35. Dunsavage, J. (2021, November 23). *Federal legislation dramatically increases funding for wildfire prevention, mitigation*. Resilience Accelerator Insurance Information Institute. https://resilience.iii.org/resilience-blog/wildfires/federal-legislation-dramatically-increases-funding-for-wildfire-prevention-mitigation/
36. Infrastructure Investment and Jobs Act, H.R. 3684, 117th Cong. (Nov. 15, 2021). https://www.congress.gov/bill/117th-congress/house-bill/3684
37. DiFeliciantonio, C. (2022, March 10). CA Floating Fire Station Opens for Operations. Firehouse. https://www.firehouse.com/stations/news/21259817/san-francisco-floating-fireboat-station-35-opens-for-operations
38. Wilmoth, J. (2017, September 25). *Why do fire stations cost so much?* Firehouse. https://www.firehouse.com/stations/article/12369981/fire-station-costs-architects-explain-why-fire-station-construction-planning-has-high-costs
39. Querolo, N., & Sullivan, B. K. (2019, October 28). *California fire damage estimated at $25.4 Billion*. Bloomberg. https://www.bloomberg.com/news/articles/2019-10-28/california-fire-damages-already-at-25-4-billion-and-counting
40. Newsom, G. (2021). California Five-Year Infrastructure Plan. California Budget. https://www.ebudget.ca.gov/2022-Infrastructure-Plan.pdf
41. *How Hot Spot Analysis (Getis-Ord Gi*) works*. (n.d.). ArcGIS Pro. https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm
42. Pedersen, J., Frank, T., & Bergstrom, B. (n.d.). *Nine major trends shaping modern fire station design*. SEH Inc. https://www.sehinc.com/news/nine-major-trends-shaping-modern-fire-station-design
43. Liu, H.-H., Chang, R. Y., Chen, Y.-Y., & Fu, I.-K. (2021). Sensor-Based Satellite IoT for Early Wildfire Detection. *2021 IEEE Globecom Workshops*. https://doi.org/10.1109/gcwkshps52748.2021.9682098

44. Ross, A. (2021). *Qualitative study of the perception of wildfire risk; IoT sensor system for wildfire detection* [Unpublished bachelor's thesis]. University of Virginia School of Engineering and Applied Science. https://doi.org/10.18130/z6kw-pn20
45. *State Area Measurements and Internal Point Coordinates*. (2010). U.S. Census Bureau. https://www.census.gov/geographies/reference-files/2010/geo/state-area.html
46. Tsipis, A., Papamichail, A., Angelis, I., Koufoudakis, G., Tsoumanis, G., & Oikonomou, K. (2020). *An alertness-adjustable cloud/fog IoT solution for timely environmental monitoring based on wildfire risk forecasting*. Energies, *13*(14), 3693. https://doi.org/10.3390/en13143693
47. Severino, C. (2021, June 3). *What you need to know about wildfire insurance in California*. Bankrate. https://www.bankrate.com/insurance/homeowners-insurance/wildfire-insurance-california/
48. Fields, S. (2021, September 21). *California orders insurers not to drop homeowners in wildfire-prone areas*. Marketplace. https://www.marketplace.org/2021/09/21/california-orders-insurers-not-to-drop-homeowners-in-wildfire-prone-areas/
49. Sibor, D. (2021, July 28). *The best homeowners' insurance in California*. NerdWallet. https://www.nerdwallet.com/article/insurance/california-home-insurance

Appendix A

Figure 7 ArcGIS Notebook in Python

In [116]:
```
from arcgis.gis import GIS
import pandas as pd

gis = GIS("home")
```

In [117]:
```
wildfire_perimeter = gis.content.get('6fd0d8d6f47d414da7bcb1dcd0539999')
wildfire_perimeter
```

Out[117]:

[California Fire Perimeters 1878 - 2020]

This layer contains the fire perimeters from the previous calendar year, and those dating back to 1878, for California. Perimeters are sourced from the Fire and Resource Assessment Program (FRAP) and are updated after the end of each calendar year.

Feature Layer Collection by juliah_esri
Last Modified: August 03, 2021
5 comments, 1,944,614 views

In [118]:
```
# print the layers contained in the Feature Layer Collection
wildfire_perimeter.layers
```

Out[118]:
```
[<FeatureLayer url:"https://services.arcgis.com/jIL9msH9OI208GCb/arcgis/rest/services/California_Fire_Perimeters_1878_2019/Feat
ureServer/0">, <FeatureLayer url:"https://services.arcgis.com/jIL9msH9OI208GCb/arcgis/rest/services/California_Fire_Perimeters_
1878_2019/FeatureServer/1">, <FeatureLayer url:"https://services.arcgis.com/jIL9msH9OI208GCb/arcgis/rest/services/California_Fi
re_Perimeters_1878_2019/FeatureServer/2">]
```

In [119]:
```
# select the second layer containing the burned areas
wildfire_sdf = pd.DataFrame.spatial.from_layer(
    wildfire_perimeter.layers[1])

# print the first 5 rows of the DataFrame
wildfire_sdf.head()
```

Out[119]:

	AGENCY	ALARM_DATE	CAUSE	COMMENTS	CONT_DATE	C_METHOD	FIRE_NAME	FIRE_NUM	GIS_ACRES	INC_NUM	OBJECTID	OBJECTIVE	REPORT_
0	CDF	2020-06-18	11.0		2020-06-23	1.0	NELSON	None	109.60250	00013212	21440	1.0	1
1	CDF	2020-06-01	2.0		2020-06-04	1.0	AMORUSO	None	685.58500	00011799	21441	1.0	67
2	CDF	2020-08-10	14.0		2020-03-01	1.0	ATHENS	None	27.30048	00018493	21442	1.0	[illegible]
3	CDF	2020-03-31	9.0		2020-04-01	1.0	FLEMING	None	12.93155	00007619	21443	1.0	[illegible]
4	CDF	2020-04-14	18.0		2020-04-19	1.0	MELANESE	None	10.31596	00008471	21444	1.0	[illegible]

In [120]:
```
acres_and_year = pd.DataFrame(wildfire_sdf, columns=['YEAR_', 'GIS_ACRES'])
acres_and_year.head()
```

Out[120]:

	YEAR_	GIS_ACRES
0	2020	109.60250
1	2020	685.58500
2	2020	27.30048
3	2020	12.93155
4	2020	10.31596

In [121]:
```
# add together all the burned acres by year, dropping incorrect and incomplete data
sums = acres_and_year.groupby('YEAR_')['GIS_ACRES'].sum()
sums.drop(sums.index[('YEAR_'==None)], axis=0, inplace=True)
sums = sums.iloc[1:]
sums.head()
```

Out[121]:
```
YEAR_
1878     59468.88000
1895       628.11360
1896      7531.71700
1898     28244.19970
1900     24528.50192
Name: GIS_ACRES, dtype: float64
```

In [122]:
```
# export the data to graph
export = pd.DataFrame({'YEAR_':sums.index, 'list':sums.values})
export.to_csv("sum_acres.csv")
```

Appendix B

Figures 9 and 10 (ARIMA Models) Google Colaboratory in Python

```
# evaluate an ARIMA model using a walk-forward validation
from pandas import read_csv
import pandas as pd
import pmdarima as pm

df = pd.read_csv('acres.csv', names=['value'], header=0)

model = pm.auto_arima(df.value, start_p=1, start_q=1,
                      test='adf',       # use adftest to find optimal 'd'
                      max_p=5, max_q=5, # maximum p and q
                      m=1,              # frequency of series
                      d=None,           # let model determine 'd'
                      seasonal=False,   # No Seasonality
                      start_P=0,
                      D=0,
                      trace=True,
                      error_action='ignore',
                      suppress_warnings=True,
                      stepwise=True)

print(model.summary())
```

```
Performing stepwise search to minimize aic
 ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=2878.181, Time=0.10 sec
 ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=2915.025, Time=0.02 sec
 ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=2888.609, Time=0.05 sec
 ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=2888.633, Time=0.07 sec
 ARIMA(0,1,0)(0,0,0)[0]             : AIC=2913.657, Time=0.02 sec
 ARIMA(2,1,1)(0,0,0)[0] intercept   : AIC=2877.415, Time=0.14 sec
 ARIMA(2,1,0)(0,0,0)[0] intercept   : AIC=2876.282, Time=0.09 sec
 ARIMA(3,1,0)(0,0,0)[0] intercept   : AIC=2878.052, Time=0.12 sec
 ARIMA(3,1,1)(0,0,0)[0] intercept   : AIC=2879.403, Time=0.27 sec
 ARIMA(2,1,0)(0,0,0)[0]             : AIC=2875.775, Time=0.17 sec
 ARIMA(1,1,0)(0,0,0)[0]             : AIC=2887.530, Time=0.07 sec
 ARIMA(3,1,0)(0,0,0)[0]             : AIC=2877.625, Time=0.30 sec
 ARIMA(2,1,1)(0,0,0)[0]             : AIC=2876.995, Time=0.32 sec
 ARIMA(1,1,1)(0,0,0)[0]             : AIC=2876.185, Time=0.09 sec
 ARIMA(3,1,1)(0,0,0)[0]             : AIC=2878.744, Time=0.18 sec

Best model:  ARIMA(2,1,0)(0,0,0)[0]
Total fit time: 2.059 seconds
```

```python
# evaluate an ARIMA model using a walk-forward validation
from pandas import read_csv
import pandas as pd
import pmdarima as pm

df = pd.read_csv('acres.csv', names=['value'], header=0)

model = pm.auto_arima(df.value[80:], start_p=1, start_q=1,
                      test='adf',       # use adftest to find optimal 'd'
                      max_p=5, max_q=5, # maximum p and q
                      m=1,              # frequency of series
                      d=None,           # let model determine 'd'
                      seasonal=False,   # No Seasonality
                      start_P=0,
                      D=0,
                      trace=True,
                      error_action='ignore',
                      suppress_warnings=True,
                      stepwise=True)

print(model.summary())
```

```
Performing stepwise search to minimize aic
 ARIMA(1,2,1)(0,0,0)[0] intercept   : AIC=589.594, Time=0.14 sec
 ARIMA(0,2,0)(0,0,0)[0] intercept   : AIC=595.074, Time=0.03 sec
 ARIMA(1,2,0)(0,0,0)[0] intercept   : AIC=591.963, Time=0.04 sec
 ARIMA(0,2,1)(0,0,0)[0] intercept   : AIC=592.443, Time=0.08 sec
 ARIMA(0,2,0)(0,0,0)[0]             : AIC=593.586, Time=0.03 sec
 ARIMA(2,2,1)(0,0,0)[0] intercept   : AIC=587.267, Time=0.25 sec
 ARIMA(2,2,0)(0,0,0)[0] intercept   : AIC=inf, Time=0.16 sec
 ARIMA(3,2,1)(0,0,0)[0] intercept   : AIC=589.260, Time=0.55 sec
 ARIMA(2,2,2)(0,0,0)[0] intercept   : AIC=587.584, Time=0.43 sec
 ARIMA(1,2,2)(0,0,0)[0] intercept   : AIC=591.023, Time=0.41 sec
 ARIMA(3,2,0)(0,0,0)[0] intercept   : AIC=inf, Time=0.36 sec
 ARIMA(3,2,2)(0,0,0)[0] intercept   : AIC=589.123, Time=0.56 sec
 ARIMA(2,2,1)(0,0,0)[0]             : AIC=584.214, Time=0.16 sec
 ARIMA(1,2,1)(0,0,0)[0]             : AIC=585.009, Time=0.08 sec
 ARIMA(2,2,0)(0,0,0)[0]             : AIC=inf, Time=0.35 sec
 ARIMA(3,2,1)(0,0,0)[0]             : AIC=586.141, Time=0.18 sec
 ARIMA(2,2,2)(0,0,0)[0]             : AIC=584.518, Time=0.24 sec
 ARIMA(1,2,0)(0,0,0)[0]             : AIC=590.166, Time=0.04 sec
 ARIMA(1,2,2)(0,0,0)[0]             : AIC=585.324, Time=0.47 sec
 ARIMA(3,2,0)(0,0,0)[0]             : AIC=inf, Time=0.37 sec
 ARIMA(3,2,2)(0,0,0)[0]             : AIC=inf, Time=0.47 sec

Best model:  ARIMA(2,2,1)(0,0,0)[0]
Total fit time: 5.507 seconds
```

```
import csv

series = read_csv('acres.csv', header=0, index_col=0, parse_dates=True,
squeeze=True, date_parser=parser)
X = series.values

history = [x for x in series.values]
predictions = list()

model = ARIMA(history, order=(2, 1,0))
recent_model = ARIMA(history[80:], order=(2, 2, 1))

model_fit = model.fit()
recent_model = recent_model.fit()

predictions = model_fit.forecast(steps=20)
recent_predictions = recent_model.forecast(steps=20)

f = open('ARIMA_predictions.csv', 'w')
f_r = open('ARIMA_recent_predictions2.csv', 'w')

writer = csv.writer(f)
writer_r = csv.writer(f_r)
writer.writerow(["Years", "Acres Burned"])
writer_r.writerow(["Years", "Acres Burned"])

for x in range(len(predictions)):
  writer.writerow([2021+x,predictions[x]])
  writer_r.writerow([2021+x, recent_predictions[x]])

f.close()
f_r.close()
```

ARIMA_predictions.csv ×

1 to 10 of 20 entries Filter

Years	Acres Burned
2021	1273252.0263697514
2022	1837766.7725359686
2023	2887346.1178806946
2024	1605272.048709259
2025	2223099.6245552097
2026	2348628.636081449
2027	1895243.7108298936
2028	2247659.0464719
2029	2167728.273426415
2030	2049775.2641822104

Show 10 per page 1 2

Appendix C

Table 1 FIRES Model Values for All Counties

County Name	Population	Area	Expected Loss	Burn Chance	Risk Z-Score	Risk Score
Alameda	1,682,348	822	23	2	-0.80	🔥
Alpine	1,203	741	0	92	-0.52	🔥🔥
Amador	40,476	606	12	30	-0.66	🔥🔥
Butte	211,626	1,677	32	50	-0.03	🔥🔥
Calaveras	45,290	1,036	4	37	-0.75	🔥🔥
Colusa	21,839	1,157	25	74	0.12	🔥🔥
Contra Costa	1,165,896	803	23	3	-0.80	🔥
Del Norte	27,731	1,015	4	72	-0.52	🔥🔥
El Dorado	191,184	1,789	17	42	-0.45	🔥🔥
Fresno	1,008,604	6,010	100	69	2.11	🔥🔥🔥🔥
Glenn	28,914	1,327	26	76	0.16	🔥🔥
Humboldt	136,433	3,610	6	61	-0.55	🔥🔥
Imperial	179,698	4,482	27	93	0.40	🔥🔥🔥
Inyo	19,016	10,229	3	122	-0.21	🔥🔥
Kern	909,039	8,163	198	75	5.24	🔥🔥🔥🔥🔥
Kings	152,480	1,392	39	65	0.37	🔥🔥🔥
Lake	68,163	1,329	16	66	-0.24	🔥🔥
Lassen	32,728	4,721	5	92	-0.35	🔥🔥
Los Angeles	10,006,672	3,951	28	18	-0.55	🔥🔥
Madera	156,224	2,153	38	70	0.42	🔥🔥🔥
Marin	262,271	609	12	20	-0.74	🔥🔥
Mariposa	17,146	1,462	19	65	-0.17	🔥🔥
Mendocino	91,487	3,512	83	68	1.64	🔥🔥🔥🔥
Merced	281,153	1,978	6	59	-0.55	🔥🔥
Modoc	8,697	4,203	2	96	-0.43	🔥🔥
Mono	13,195	3,131	3	92	-0.43	🔥🔥
Monterey	438,876	3,313	59	67	0.94	🔥🔥🔥
Napa	138,014	790	16	45	-0.44	🔥🔥
Nevada	102,234	975	11	43	-0.55	🔥🔥
Orange	3,186,601	799	22	2	-0.81	🔥
Placer	245,274	1,500	14	42	-0.51	🔥🔥
Plumas	19,790	2,614	7	76	-0.40	🔥🔥
Riverside	2,418,121	7,301	42	59	0.35	🔥🔥🔥

Sacramento	1,585,055	994	29	0	-0.77 🔥
San Benito	27,397	1,390	30	96	0.55 🔥🔥🔥
San Bernardino	2,181,698	20,106	55	79	1.09 🔥🔥🔥
San Diego	3,297,473	4,258	36	37	-0.14 🔥🔥
San Francisco	873,818	106	1	-34	-1.22 🔥
San Joaquin	779,231	1,426	42	24	-0.26 🔥🔥
San Luis Obispo	282,590	3,322	69	68	1.24 🔥🔥🔥
San Mateo	764,070	552	10	5	-0.90 🔥
Santa Barbara	447,171	2,551	37	64	0.30 🔥🔥🔥
Santa Clara	1,936,209	1,305	32	27	-0.36 🔥🔥
Santa Cruz	270,645	446	9	31	-0.70 🔥🔥
Shasta	182,143	3,852	24	71	0.05 🔥🔥
Sierra	3,236	961	3	72	-0.54 🔥🔥
Siskiyou	44,080	6,347	49	84	1.00 🔥🔥🔥
Solano	453,931	910	22	35	-0.43 🔥🔥
Sonoma	488,844	1,605	32	27	-0.36 🔥🔥
Stanislaus	552,846	1,516	44	37	0.00 🔥🔥
Sutter	99,630	608	19	52	-0.31 🔥🔥
Tehama	65,835	2,958	74	82	1.73 🔥🔥🔥🔥
Trinity	16,109	3,207	3	100	-0.36 🔥🔥
Tulare	473,100	4,843	67	78	1.44 🔥🔥🔥🔥
Tuolumne	55,622	2,278	4	72	-0.50 🔥🔥
Ventura	843,543	1,832	24	57	-0.12 🔥🔥
Yolo	215,769	1,021	28	43	-0.21 🔥🔥
Yuba	81,572	644	14	44	-0.49 🔥🔥