Final Project: Using Random Forest to Predict Random Forest Fires

Emily Fu, Zhun Yan Chang, Melisa Lee

Due 11:59 PM, December 16, 2019

Contents

1 Executive Summary **2**
1.1 Background . . . 2
1.2 Description of Data . . . 2
1.2.1 Goal . . . 3
1.3 Summary of Findings . . . 4
1.4 Issues and Limitations . . . 4

2 Exploratory Data Analysis **4**
2.1 Data Cleaning . . . 4
2.2 Data Joining . . . 5
2.2.1 National Oceanic and Atmospheric Data (NOAA) . . . 5
2.2.2 American Community Survey (ACS) . . . 5
2.2.3 ArcGIS REST API by United States Department of Agriculture Forest Service . . . 5
2.3 Data Transformations . . . 5
2.3.1 Fire Size . . . 5
2.3.2 Area Burned Over Time . . . 7
2.3.3 Fire Cause . . . 7
2.3.4 Fires by Location . . . 9
2.3.5 What's Sparking the Lightning Fires? . . . 9

3 Model Building **11**
3.1 Testing & Training Data . . . 11
3.2 Linear Regression . . . 11
3.2.1 Forward Regsubsets . . . 11
3.2.2 Backwards Regsubsets . . . 12
3.3 LASSO Model . . . 12
3.4 Random Forest . . . 13

4 Final Model **14**

5 Conclusion **14**
5.1 Results . . . 14
5.2 Recommendation . . . 14

6 Appendix **16**
6.1 Explanantion of Complex Variables . . . 16
6.2 Maps of Complex Variables . . . 16
6.3 Breakdown of Fire Causes . . . 16

1 Executive Summary

1.1 Background

In late October 2019, over a dozen fires raged in California, burning over 100,000 acres and causing nearly 200,000 evacuations and the declaration of a state of emergency. Over the past several decades the amount of area burned in the state has steadily gone up, with scientists attributing the lengthened wildfire season and the increasing likelihood of extreme fires to global warming. The annual cost of fire suppression from these disasters is estimated to cost the state roughly $500 million per year this decade, doubling from the cost in previous decades. As this problem only continues to become more serious, it is important for the Californian government and local fire departments to be able to accurately identify high-risk areas in order to adequately prepare for and respond to wildfire disasters.

In recent years, we have seen an effort to decrease the risk of these fires through targeted action. Most recently, P&G decided to shut down the power grid during high-wind seasons in California in order to decrease the change that a fire would break, resulting in a blackout for over 5 million people. Due to both the high risks and high costs of fires and fire prevention, we want to build a model that helps us predict the size of discovered wildfires in California to inform decision making in the future and improve the safety for state residents.

1.2 Description of Data

Our main data source is from the U.S. Government Forest Research Data Archive. We focused our analysis on 114,558 fires in California from 2001 to 2015. The dataset includes location factors, such as longitude, latitude, and FIPS code, as well as information on the fire discovery like cause and discovery date and time.

To improve our analysis, we also scraped data from the National Oceanic and Atmospheric Administration for the average monthly high temperature, low temperature, and precipitation level and matched it with the month and county of the fire. We used the ArcGIS REST Services Directory hosted by United States Department of Agriculture Forest Service and Python to scrape the vegetation fuel type, mean fire return interval, vegetation departure, wildfire supression index, and drought index by matching with the exact latitude and longitude of the fire.

Our response variable is `FIRE_SIZE` which represents the total number of acres burned for each fire.

Variable	Description
`FOD_ID`	Global unique identifier for each fire
`Fire Name`	Name of the fire
`FIRE_SIZE`	Estimate of acres within the final perimeter of the fire
`FIRE_SIZE_CLASS`	Code for fire size based on the number of acres within the final fire perimeter expenditures
`FIRE_YEAR`	Year that the fire was discovered
`FINAL_DATE`	Date on which the fire was discovered or confirmed to exist
`DISCOVERY DATE`	Date on which the fire was discovered or confirmed to exist in Julian form
`DISCOV_MONTH`	Month that the fire was discovered
`DISCOVERY_DOY`	Day of year the fire was discovered
`DISCOVERY_TIME`	Time of day the fire was discovered
`CONTROLLED_DATE`	Date on which the fire was declared contained or otherwise controlled
`DAYS TO CONTROL`	Number of days it took from discovery to control the fire
`Longitude`	Longitude of fire location
`Latitude`	Latitude of fire location
`AndersonFBFM`	Distinct distributions of fuel loading found among 13 surface fuel components (live and dead), size classes, and fuel types
`ScottAndBurganFBFM`	Distinct distributions of fuel loading found among 40 surface fuel components (live and dead), size classes, and fuel types

Variable	Description
MeanFireReturnInverval	Arithmetic average of all time intervals between fires in a given area over a given time
VegetationDeparture	Indicator of how different current vegetation on a landscape is from estimated historical conditions
ExistingVegType	Representation the current distribution of the terrestrial ecological systems classification developed by NatureServe for the western hemisphere
WildfireSuppressionDifficulty	Quantitative rating of relative difficulty in performing fire control work
pdsi	Palmer Drought Severity Index uses readily available temperature and precipitation data to estimate relative dryness
NWCG_REPORTING_AGENCY	Active National Wildlife Coordinating Group Unit Identifier for the agency preparing the fire report
NWCG_REPORTING_UNIT_ID	Active NWCG Unit Identifier for the unit preparing the fire report
SOURCE_REPORTING_UNIT	Code for the agency unit preparing the fire report, based on code/name in the source dataset
SOURCE_REPORTING_UNIT_NAME	Name of reporting agency unit preparing the fire report, based on code/name in the source dataset
STAT_CAUSE_DESCR	Description of the (statistical) cause of the fire
FIPS_CODE	Three-digit code from the Federal Information Process Standards publication for representation of counties
TotalPop	Total population for the the county the fire is in
IncomePerCap	Income per capita for the county the fire is in
Poverty	Percentage of population under poverty level for county the fire is in
ChildPoverty	Percentage of children living under poverty for the county the fire is in
Unemplyment	Unemployment rate of county the fire is in
precp_winter	Total precipitation for the county in the winter
precp_spring	Total precipitation for the county in the spring
precp_summer	Total precipitation for the county in the summer
precp_fall	Total precipitation for the county in the fall
avg_temp_winter	Average temperature in the county for the winter
avg_temp_spring	Average temperature in the county for the spring
avg_temp_summer	Average temperature in the county for the summer
avg_temp_fall	Average temperature in the county for the fall
precp_month	Average precipitation in the county of the fire for the month of the fire
avg_temp_month	Average temperature in the county of the fire for the month of the fire
high_temp_month	High remperature in the county of the fire for the month of the fire
low_temp_month	Low temperature in the county of the fire for the month of the fire

1.2.1 Goal

Our goal is to use the variables in the dataset to accurately predict the expected fire size. In our analysis and research, we realized that few extremely large fires (less than 10%) accounted for over 80% of the area burned and economic damage. As a result, we believe that being able to predict fire size and, in particular, identify the fires that are likely to be the most destructive can be valuable in determining the amount of resources into the fire suppression effort. Since resources are often limited, determining which fires require the most attention is a difficult an important decision. Our goal is to build a model that predicts fire size and serves as a recommendation for which fires are most likely to grow large. Hopefully, by identifying these fires as high-risk and dedicating more resources to it, we can prevent it from becoming so destructive in the future.

1.3 Summary of Findings

Using linear model selection methods such as forward, backward and LASSO, the predictors of fire size that are consistently ranked the strongest include Existing Vegetation and NWCG Reporting Agency. However, non-linear models such as random forests give very different results, ranking more temporal and locational variables as most important such as Discovery Time/Date and Longitude/Latitude.

1.4 Issues and Limitations

One of the major limitations that we ran into during our analysis was that we were dealing with spatial data. The location of the fire played a significant role in multiple predictors such as vegetation type, drought index, climate, and cause of fire. As a result, total fire size also largely depending on the location. This is where our problem came in. Since in this class we focus on linear regression, we are not able to do a spatial analysis. We believe in our analysis we were not able to capture the full relationships across space. This type of geospatial analysis is outside the scope of this class and outside the scope of our data analysis abilities.

Another issue is the accuracy of our weather data. We were only able to find temperature and precipitation data to the precision of a county, rather than the exact location. This is since weather data is tied to a weather station in a particular area, and its density can vary across counties. This makes some of our data inaccurate since counties in California can span vast amounts of area (San Bernardino is 62 square miles). We decided to stick with this less accurate measure since tracking down an individual temperature for 80,000+ longitude and latitude pairs simply would not be productive or feasible given our time frame.

Another issue was that many of our predictors are factors. Significant categorical predictors included vegetation type, fire cause description, and FIPS code. However, in the process of building the model we hit many of the factor and level limits for LASSO and random forest. As a result, the full explanatory power of our variables were not fully captured in some of the models since we had to take them out in order for them to be able to run. One way to combat this would be to do research into each specific vegetation type and bag them in an intelligent way. As for the FIPS code, getting more detailed continuous demographic and geographic data at a greater granularity than the country level could help explain the variability without using factors.

2 Exploratory Data Analysis

2.1 Data Cleaning

Our US Governemnt Forest Research data was quite messy to begin with, our original dataset included many N/A and NULL values for the various factors. In addition, the dataset covered over 1.88 million from 1994 - 2015 across the entire United States. The scope of the entire dataset was simply too large for us to perform analysis on. As a result, we decided to narrow our project down to California, for two reasons. First, California was the state with the largest number of records and the biggest wildfire problem, making the project the most impactful if we could develop an intelligent model to predict the fire size. The second reason ties to the fact that one of our project members grew up in Southern California and had to evacuate due to a wildfire. We also decided to focus soley on fire data after 2001 for our analysis. This is becuase we found that the records from earlier were missing many values and we were concerned about the data quality of the earlier observations.

Some of the intial cleaning we did was dropping the unneccessary columns. For example, we had multiple IDs for each fire (federal, local, etc.) and both the name and code for the cause of the fire. As a result, we removed columns of data that would not be relevant from our dataset. Other variables had too much variability for there to be any meaningful differences between the levels such as `OWNER_CODE`, which assigned a code to the owner of the land the fire burned on. In addition, we also had to convert the date from a Julian date in order to have access to the meaningful relationships between fire year, month, and day or year.

2.2 Data Joining

2.2.1 National Oceanic and Atmospheric Data (NOAA)

While we were doing the research, we came across several climate/vegetation indices that we believed would be valuable for our analysis. As a result, we decided we would pull in external data sets to supplement the information we already had. Since predicting a fire size based only on the fire cause, exact location, and reporting agency was not going to get us very far. The first dataset we joined was weather data from the NOAA. This gave us the average precipitation, high temperature, low temperature, and average temperature by the county based on 10-year averages. We joined this to our current dataset by FIPS code, which corresponded with the county. While we preserved the temperature and precipitation values for the specific month in which the fire broke out, we decided to congregate the rest by season.

2.2.2 American Community Survey (ACS)

We were also curious how socioeconomic and demographic factors play a role in predicting the risk-level of an area to wildfires. As we have seen with other environmental disasters such as flood and pollution, environmental discrimination often appears in the US across ethnic and income lines, with the most at-risk populations having high poverty and minority rates. In order to get data on these parameters, we looked to the ACS which is an ongoing survey done by the US Census Bureau. This granualrity of this imformation also came at the county level, so this was joined to each indivudal fire by county.

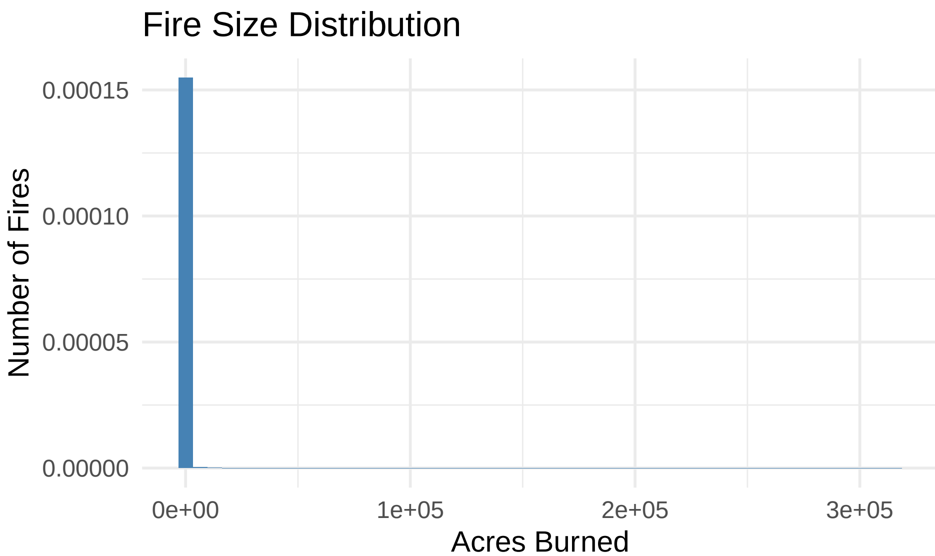
2.2.3 ArcGIS REST API by United States Department of Agriculture Forest Service

The most difficult data to obtain by far was our spatial data. Since we believe key indices such as fire burn fuel models, drought index, mean fire return interval, vegetation type, vegetaion departure, and wildfire difficulty suppression index were going to be the most important factors in predicting the size of the fire, we wanted ot make sure each statistic was tied to the exact location of the fire. This required us to scrape the data one by one from the ArcGIS website. We used python to scrape these 6 indices for each fire. We joined the data on FOD_ID to create our full dataset.

2.3 Data Transformations

2.3.1 Fire Size

The intial graph of fire sizes shows a very strong skew to the left, which makes sense since around 95% of the fires every year are class A or B fires, the smallest fires. We have included a table before that explains the Fire Size Class Statndards from the National Wildfire Coordination Group. This is also clear in the summary statistics. The IQR runs from 0.1 to 1, however the maximum is over 300,000. This further supports our point that is it only a handful of very large fires causing majority of the damage.

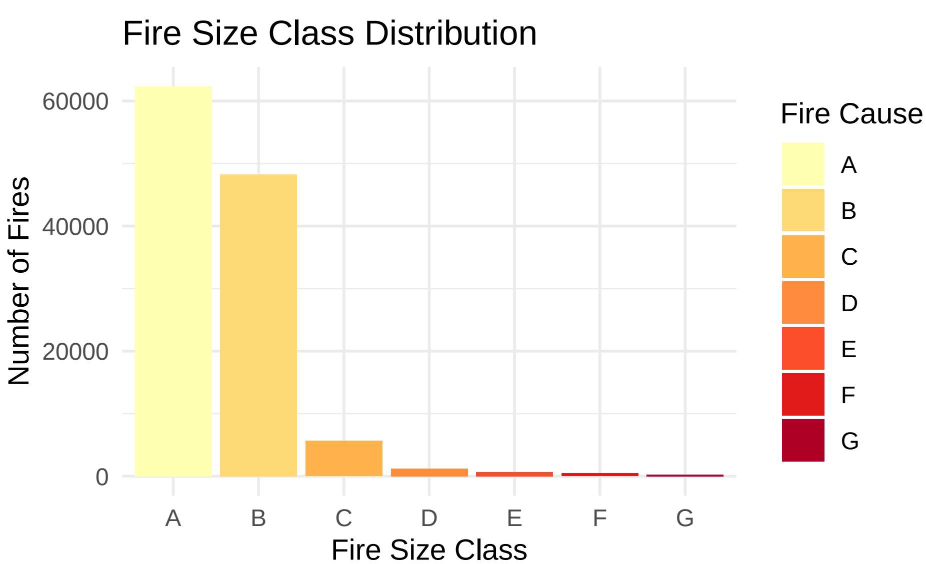


Summary Statistics for Fire Size

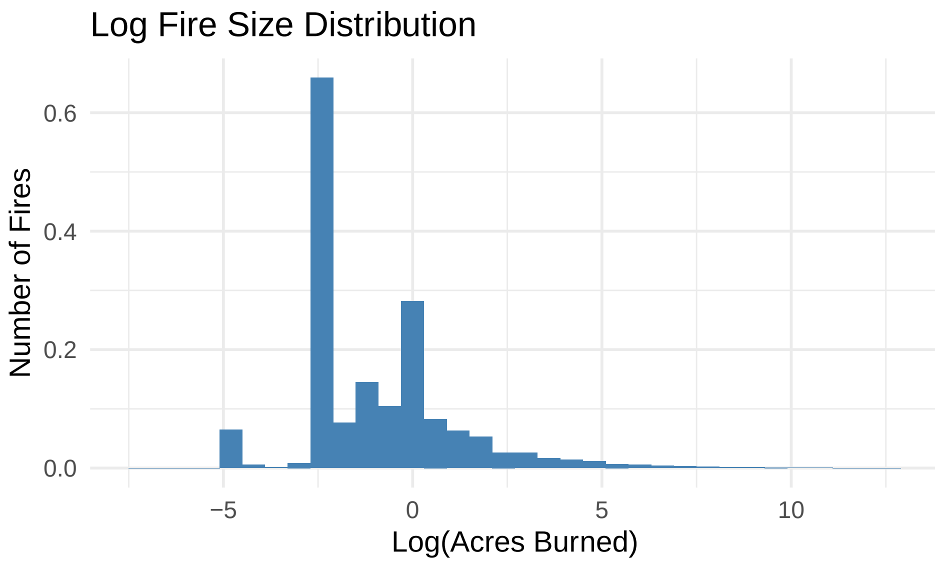
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	0.10	0.20	75.97	1.00	315578.80

Fire Class Size Breakdown

	Fire Size in Acres
Fire Class A	<0.25
Fire Class B	0.25-10
Fire Class C	10-100
Fire Class D	100-300
Fire Class E	300-1000
Fire Class F	1000-5000
Fire Class G	>5000



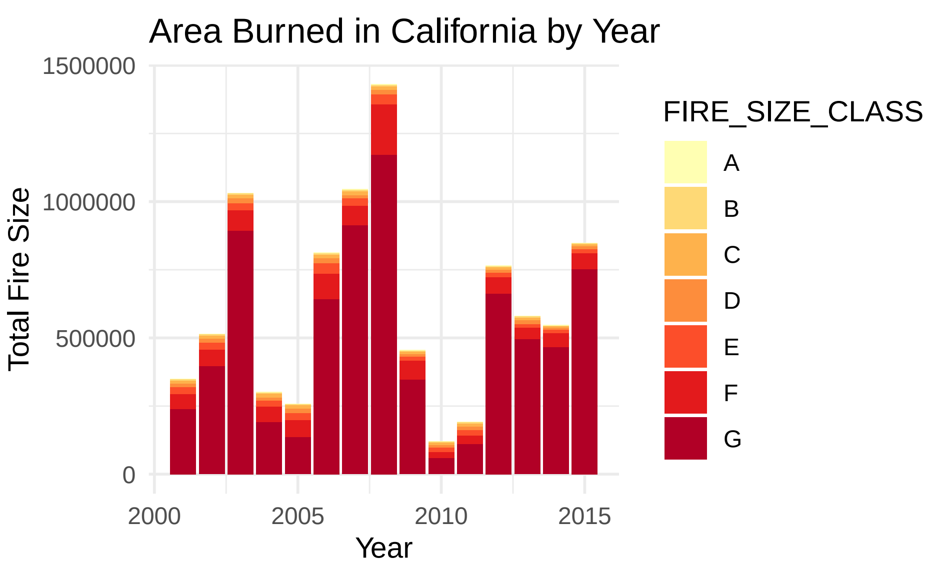
In order to make our response variable more differentiated for our regression model, we decided to take the log of fire size to help normalize the long right tail. This histagram now looks much less skewed than the



earlier one.

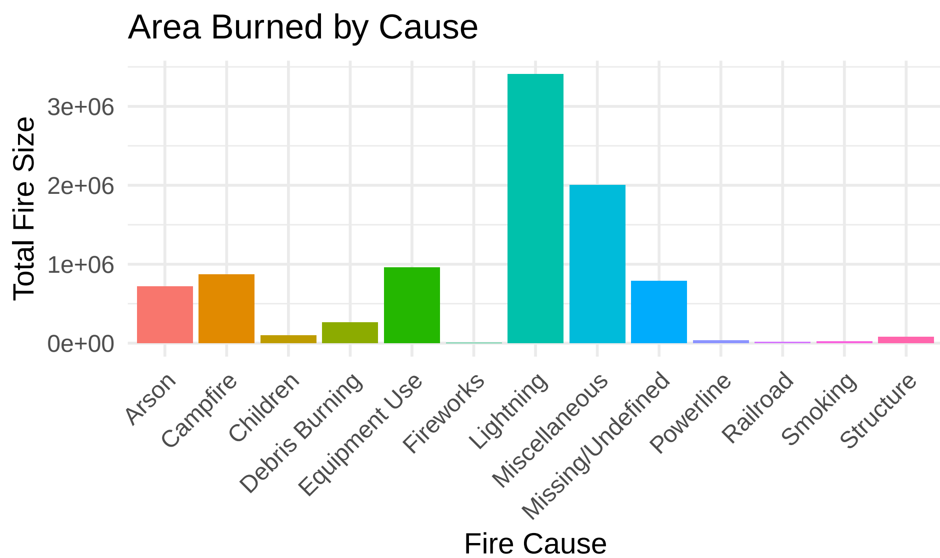
2.3.2 Area Burned Over Time

There are two interesting takeaways from this graph. The first is that despite the large amount of small fires, as seen in our previous graphs, majority of the damage is caused by several very large Class G fires, shown below in the maroon. The second interesting thing to note is the wavy pattern we see in the total area burned from year to year. Rather than being a constant, it seems to cycle up and down, dipping for 2 years before shooting up again. This is due to California's fire years. As we can see in the graph, after a year with a large amount of area burned, there follows a period of very little area burned. This is due to the time it takes for vegetation to grow back and dry out and provide fuel for new fires, since fires cannot catch on already burned land.

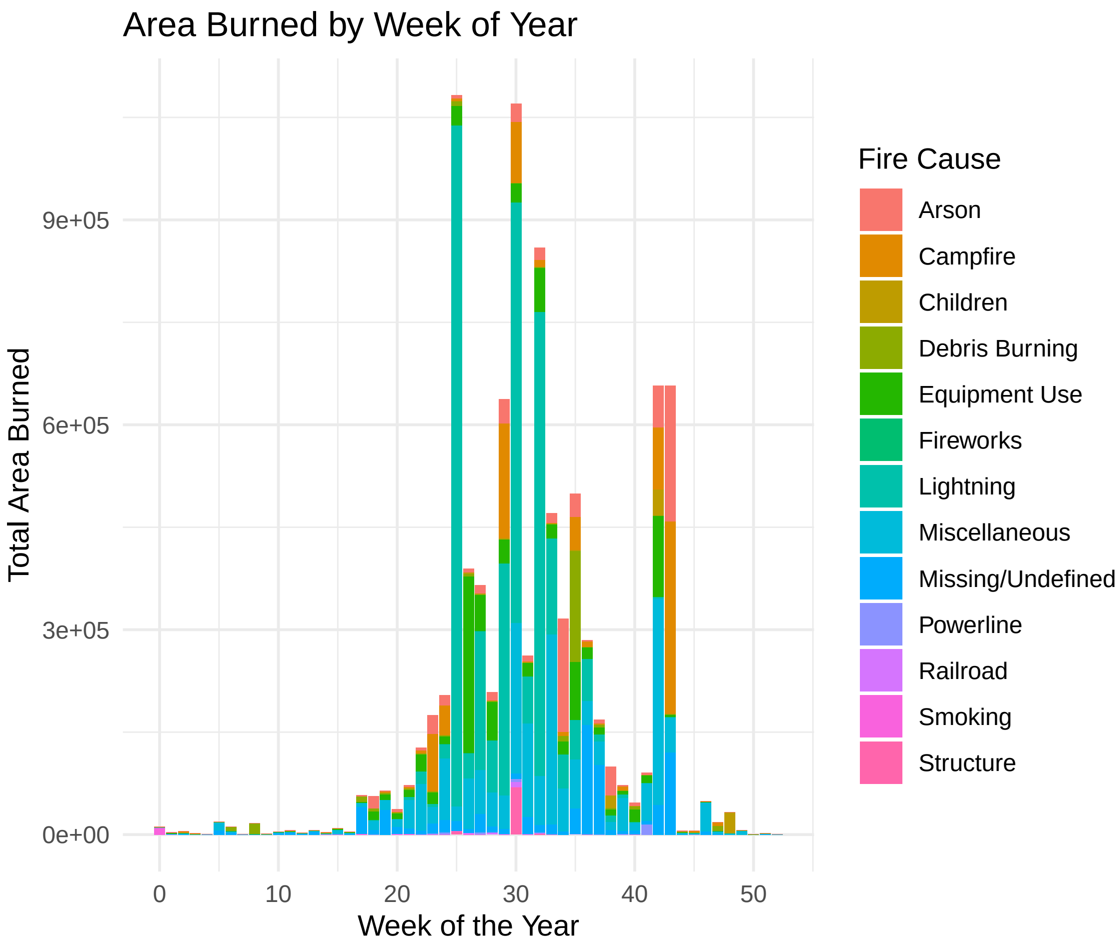


2.3.3 Fire Cause

From this graph, we can see that the largest fires are caused by lightning, followed by miscellaneous and equipment use. Though this seem discouraging since lightning is a natural phenomenon outside of human control, there are still many ways to combat these fires. In recent years, the increase in the destructive power of lightning fires has largely been credited to droughts that weaken and dry out the trees in the Northern Californian forests. In addition, invasive species such as the bark beetle cause more tree deaths and provides more fuel. For a full breakdown of the description for each of these causes, please see our appendix.



In addition we can see how the size of the fires depends on the cause and the week of the year. To get the rough week that a fire was discovered, we simply took the rounded answers of `DISCOVERY_DOY` divided by 7. Unsurprisingly, most of the land burned happens during California's fire season, which runs from summer through fall (weeks 24 - 43). Again, we see that lightning causes the most destructive fires in the summer, since they are most likely to strike the redwood forests which are often dried out at that time.



2.3.4 Fires by Location

One of the most interesting parts of our data is the spatial relationship. So here we graphed the longitude and latitude of each fire point using Tableau. The color indicates the cause of the fire. Comparing this with our vegetation and precipitation data yields some interesting insights. So we see the top of California has a lot of lighting fires, forests where lightning is likely to strike. We see arson, the dark blue is a little more in the center, and along the coasts, that's where people are concentrated and therefore more likely to commit arson. There's a lot a lot of empty space in the east even though it is the driest. The desert, has nothing to burn.

2.3.5 What's Sparking the Lightning Fires?

As shown by our earlier analysis, the most destructive fires are lightning fires. We can see in the map on the right below (the yellow circle represent lightning fires), that they are mainly congregated in the northern forests. When we did further research into the spread of lightning fires, we found that an underlying cause was the existence of bark beetles. They prey on trees that have been weakened by years of drought and kill them. this leaves millions of dead trees scatterred across the Northern Californian forests and provides perfect kindling for the lightning strikes to turn into large forest fires. In fact, when we compare our map of lightning fires with a map of invasive species in California on the bottom left, we see that the two images line up almost perfectly with the purple spots. The purple spots represent invasive species that are fir engravers, which include the bark beetles that have killed over 129 million fir trees.

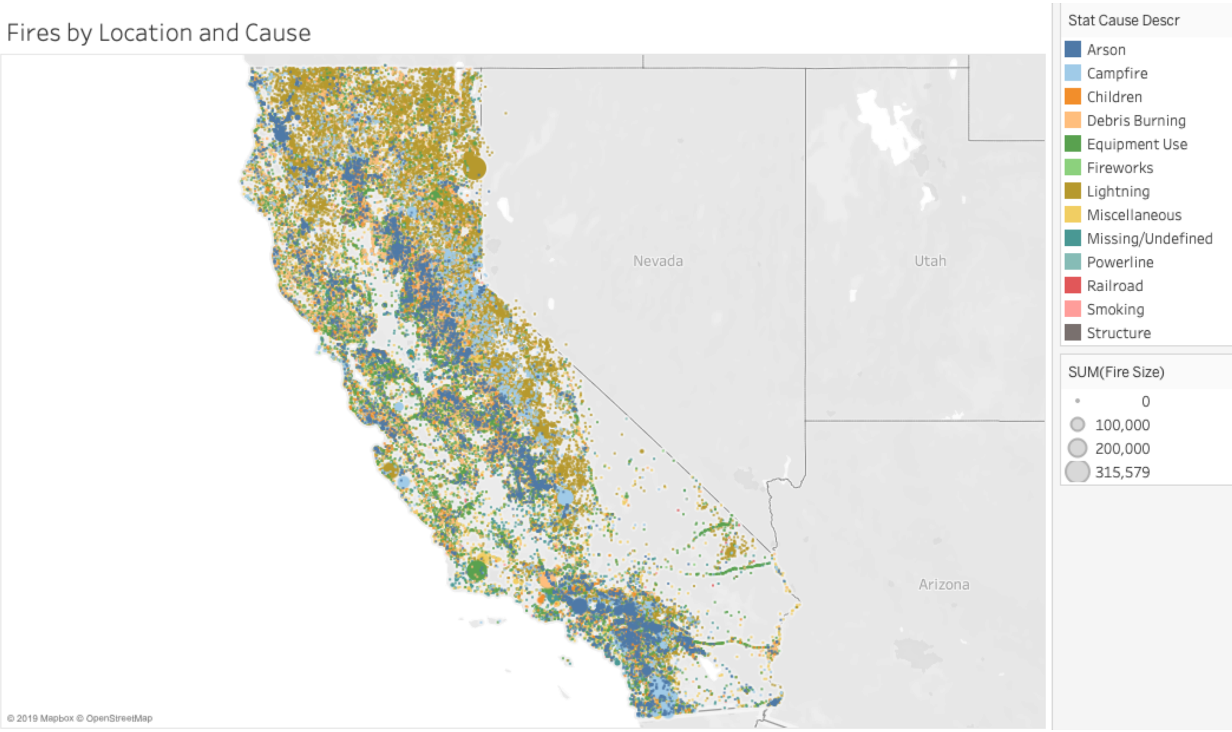


Figure 1:

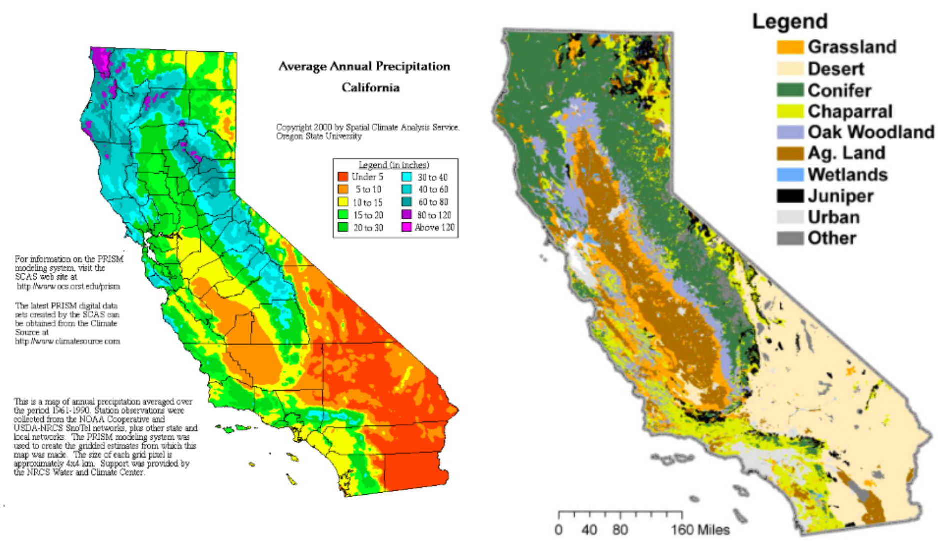


Figure 2: California Precipitation and Vegetation

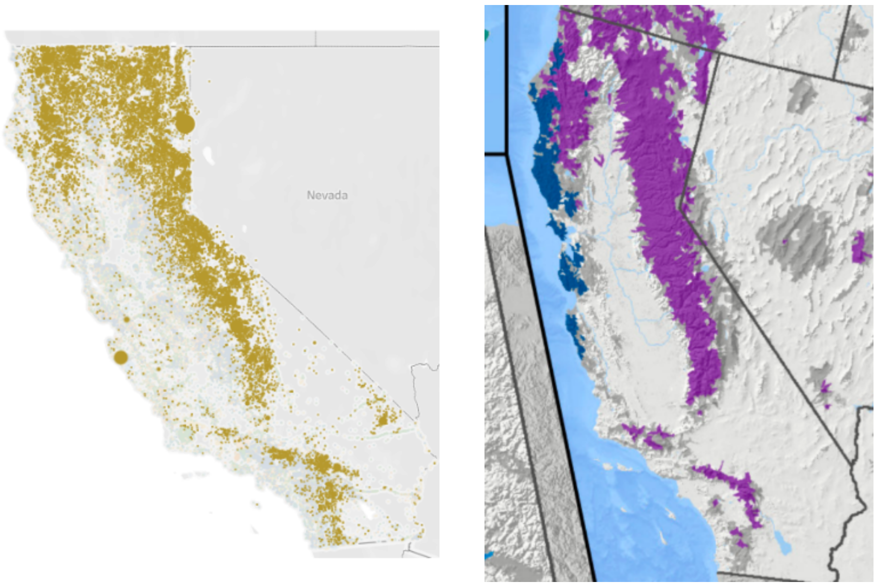


Figure 3: Lightning Fires vs Invasive Species

3 Model Building

3.1 Testing & Training Data

Due to limitations on the computational ability of our student laptops, we decided to focus the model building on a subset of 37,000 data points of fires from 2011-2015. We believe due to an increased climate change risk and other factors, it is most pertinent for us to focus on more recent years. Our training/testing/validation data account for 70/20/10 percent of our full data, respectively. In order to preserve outliers, we made sure to take a proportional sample of the data for east set based on fire size, cause, and year.

3.2 Linear Regression

With a linear model, a linear relationship between our predictors and the size of the fire. The model is fit by estimating coefficients of each independent variable such that, if we were to sum the product of the independent variables and their coefficient, we would get an estimated value of our dependent variable for each observations. The coefficients are tuned to minimize the sum of squared differences between the estimated value and the actual value for each observation. The parametric form of a linear model requires certain assumptions.

3.2.1 Forward Regsubsets

Forward regsubsets requires us to pick the first best variable that minimizes the deviance of the model then moves downwards to find the second best variable that minimizes the deviance of the model. This is less optimal than an exhaustive search but requires less computational effort. We decided to use Cp which represents prediction error.

Our forward regsubsets produced a model that included the factors: `DISCOVERY_DATE + Latitude + Longitude + ScottAndBurganFBFM + MeanFireReturnInterval + ExistingVegType + WildfireSuppressionDifficul + NWCG_REPORTING_AGENCY + STAT_CAUSE_DESCR + FIPS_CODE + Poverty + IncomePerCap + precp_month + pdsi + avg_temp_month + high_temp_month`. We then ran those factors in a linear regression and removed factors until they were all significant at the 0.01 level. The 5 most important predictors and their coefficients are shown below:

These first four factors are all colerated with an increase tin the predicted log size of the fire. Vegetation

	Forward Resubsets Coefficient Estimates
NWCG_REPORTING_AGENCYDOD	8.3521942
ExistingVegType3960	3.715112
ExistingVegType3966	2.766639
ExistingVegType3068	2.546763
ExistingVegType3004	-2.536253

Types 3960, 3966, and 3068 represent Western Cool Temperate Orchard, Western Cool Temperate, and North Pacific Dry and Masic Alpine Shrubland vegetaion, respectively. The DOD reporting agency is Department of Defense, meaning that the areas they reside over tend to have larger fires. Existing Vegetation type 3004 is North American Warm Desert Sparsely Vegetated Systems and negatively correlated with log fire size, makes sense since there are close to no large fires in the dessert.

The model produced a testing error of 3.961597.

3.2.2 Backwards Regsubsets

Backwards regsubset first removes the variable that increases the deviance the most followed by the next variable that increases the deviance of the model the second most.

Our backward regsubsets produced a model that included the factors: `FIRE_YEAR + DISCOVERY_DATE + DISCOVERY_DOY + DISCOVERY_TIME + Longitude + Latitude + MeanFireReturnInterval + VegetationDeparture + AndersonFBFM + ScottAndBurganFBFM + ExistingVegType + NWCG_REPORTING_AGENCY + STAT_CAUSE_DESCR + FIPS_CODE + pdsi + avg_temp_month + high_temp_month + precp_month`. The 5 most important predictors and their coefficients are shown below:

	Backward Resubsets Coefficient Estimates
FIRE_YEAR	-64.54891
NWCG_REPORTING_AGENCYDOD	8.283172
AndersonFBFM11	-3.350822
ExistingVegType3960	3.326509
AndersonFBFM12	-3.310215

It is interesting to note that the backwards selection found `FIRE_YEAR` as an extremely important predictor while forward did not. The backwards selection also included Fire Burn Fuel Model factors. FBFM 11 represents fairly active fire, fuels that consist of slash and herbaceous materials, slash originates from light partial cuts or thinning projects, fire is limited by spacing of fuel load and shade from overstory. FBFM 12 represents Rapid spreading and high intensity fires, dominated by slash resulting from heavy thinning projects and clearcuts, slash is mostly 3 inches or less. These FBFM show that places that are burned often result in vegetation that that are not good fuel models for fires. These three predictors were all negative and therefore correlated with a smaller log fire size, all else held equal. DOD and vegetation type 3960 are overlaps between the two models.

The model produced a testing error of 3.963663.

3.3 LASSO Model

LASSO is useful in regularization of predictors, helping us detemine what is actually significant by pushing coefficients towards zero. We decided to use lamdba min as our tuning parameter which gave us 10 non-zero coefficients. Since LASSO results tend to be biased, we use relaxed LASSO to get our model which included: `DISCOVERY_DATE + Longitude + ScottAndBurganFBFM + ExistingVegType + WildfireSuppressionDifficultyIndex + NWCG_REPORTING_AGENCY + STAT_CAUSE_DESCR + FIPS_CODE + pdsi + low_temp_month`

When we ran a linear regression, we found that the top 5 most important predictors:

	LASSO Coefficient Estimates
NWCG_REPORTING_AGENCYDOD	8.485139642
ExistingVegType3960	3.76045773
ExistingVegType3966	2.87839369
ExistingVegType3068	2.627957954
NWCG_REPORTING_AGENCYBORG	2.5931130360

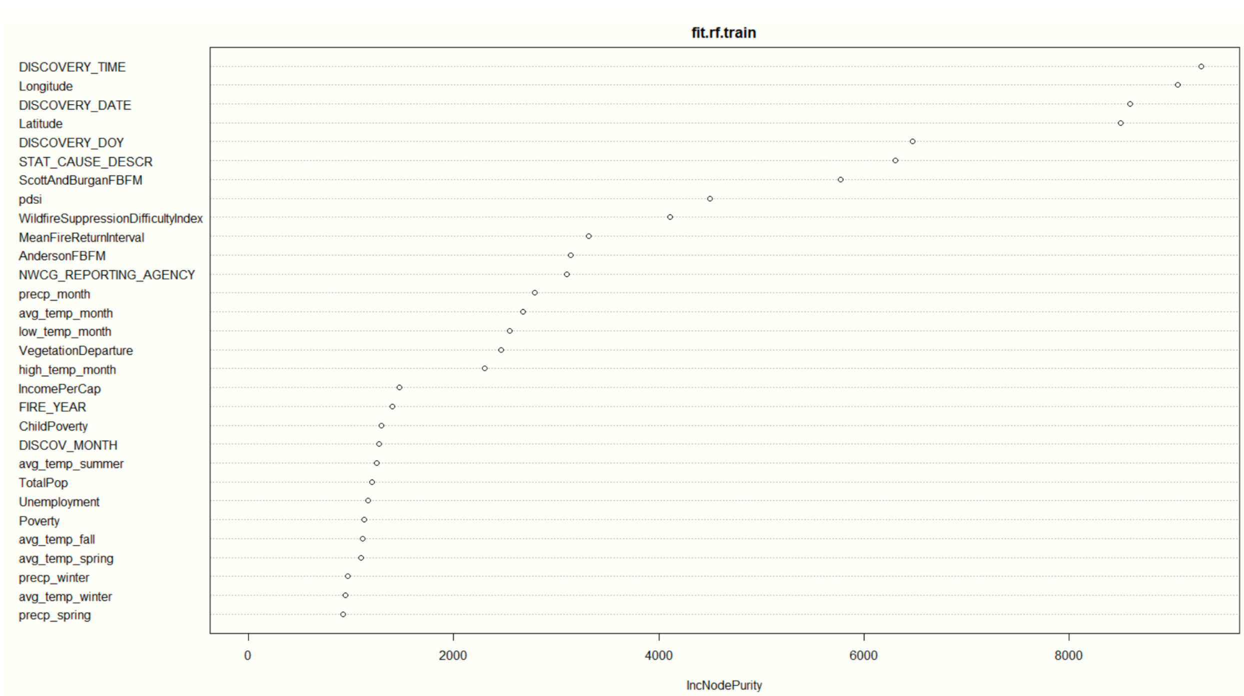
The LASSO output was very similar to the forwards regsubset output in both the factors chosen and their coefficients. The main difference is that LASSO also included Reporting Agency BOR which were in areas overseen by the Bureau of Reclaimation.

The LASSO model had a testing error of 3.965501.

3.4 Random Forest

Random forest relies on the idea of a large number of relatively uncorrelated trees operating as a committee will outperform any of the individual constituent trees. Although we must sacrifice the ability interpret how each of these selected variables directly affect fire size, we may still analyze the variable importance measures.

The important predictors from our Random Forest are shown in the image below:



Node purity for regression is used to measure how closely a randomly chosen data point from the set would be correctly predicted if it was randomly grouped according to the distribution of leaves in the splitted sub-branch. It is, therefore, used to rank which are the most important predictors (read: most accurate) included in random forest. The top 5 predictors are Discovery Time, Longitude, Discovery Date, Latitude, and Discovery Day of Year. Surprisingly, many of these were not highlighted among the top 5 variables in other models, although they were included as selected predictors. Unfortunately, two variables, FIPS Code and Existing Vegetation Type, were removed due to their high number of factor levels and our limitations in computing power.

Still, random forest still performs well relative to its competitor models with an out-of-bag error of 3.620212 and testing error of 3.679664.

4 Final Model

We evaluated our model based on the testing errors of the different models.

	Testing Error
Fwd Regsubsets	3.961597
Bck Regsubsets	3.963663.
LASSO	3.965501
Random Forest	3.679664

Random forest makes sense since many of the relationships are not linear or single directional, think back to year and how it ebbs and flows. In addition, since we have many factor variables that are significant such as Fire Burn Fuel Model, random forest can help model these interactions. Linear regression cannot show interaction and only models different factor levels as stacked lines wihtout accounting for the fact that slope could be different. Therefore, it was unable to capture the full value of the temporal and locational variables included the way random forest can.

One caveat to note about using random forest and node purity to rank the importance of variables, however, is that it may introduce bias into the model, preferring variables with many factor levels (which we have largely taken care of by removing predictors with ~100 levels). It also does not handle collinearities well, and it may discount a predictor because its collinear counterpart has already been ranked highly. This may explain why some variables that were ranked highly in the other models such as NWCG Reporting Agency and Fire Year rank lower in importance according to node purity.

Our final validation error for the random forest is 3.48764.

5 Conclusion

5.1 Results

We conclude that random forest is the best model to predict the size of the wildfires evidenced by the low testing error and validation error. Across all the linear models, we consistently saw that these predictors were significant: Department of Defense who reports the fire, and the Existing Vegetation Type are Western Cool Temperate Orchard, Western Cool Temperate Fallow/Idle Cropland and the North Pacific Dry and Mesic Alpine Dwarf-Shrubland or Fell-field or Meadow. Most of the agricultural properties are located either immediately adjacent to natural vegetation or within a mile of it, an area referred to as the Ember Zone. These properties are all at risk from the flaming front of an advancing wildfire or the embers it produces.

Severe fires burning in recent years have destroyed large areas of forest and created ideal conditions for shrubland to spread to areas it didn't historically cover. Because forest conditions have changed, where they burn now are switching to large shrublands that can persist. Forest can creep in on shrubland from the edges, but it is a slow process. More than a century of putting out all fires have left unnatural conditions leaving forests prone to more destructive wildfires.

5.2 Recommendation

With that, we have a few recommendations. If we proactively ramp up the use of controlled fire as a restoration tool, we can remove the dense vegetation that fuels fast-moving blazes, giving fire crews better opportunities to stop future wildfires. Hence, fire management should focus on strategic prescription burns to both insure the most efficient fire hazard reduction. There is a need to increase the pace and scale of forest treatments, including controlled burns, and focus on removing the most problematic vegetation - small diameter trees and brush - and not the large, most fire-resistant trees.

There is also increasing community responsibility in locating and constructing new homes. More attention to fire prevention are likely to be avenues for the greatest decreases in community vulnerability to wildfires. In future planning, cities should encourage smarter development, discouraging sprawling new subdivisions in fire-prone areas and instead favoring higher-density construction in cities and established neighborhoods.

6 Appendix

6.1 Explanantion of Complex Variables

Variable	Description
`AndersonFBFM`	13 Anderson Fire Behavior Fuel Model (FBFM13) describes the 13 fuel models listed by Albini and provided aids to
selecting a fuel model. Anderson listed as model parameters only fuel load by size class, fuelbed depth, and dead fuel extinction moisture.	
`ScottAndBurganFBFM`	40 Scott and Burgan Fire Behavior Fuel Model (FBFM40) represents distinct distributions of fuel loading found among surface fuel components (live and dead), size classes, and fuel types. This set contains fuel models in every fuel type (grass, shrub, timber, slash). The number of fuel models representing relatively high dead fuel moisture content increased, and fuel models with an herbaceous component are now dynamic, meaning that loads shift between live and dead (to simulate curing of the herbaceous component) rather than remaining constant.
`MeanFireReturnInverval`	The Mean Fire Return Interval (MFRI) quantifies the average period between fires under the presumed historical fire regime. MFRI is intended to describe one component of historical fire regime characteristics in the context of the broader historical time period represented by the LANDFIRE (LF) Biophysical Settings (BPS) layer and BPS Model documentation.
`VegetationDeparture`	Vegetation Departure (VDep) indicates how different current vegetation on a landscape is from estimated historical conditions. VDep is based on changes to species composition, structural stage, and canopy closure. The VDep metric ranges from 0 - 100.
`ExistingVegType`	RExisting Vegetation Type (EVT) represents the current distribution of the terrestrial ecological systems classification, developed by NatureServe for the western hemisphere, through 2016. A terrestrial ecological system is defined as a group of plant community types that tend to co-occur within landscapes with similar ecological processes, substrates, and/or environmental gradients.
`WildfireSuppressionDifficulty`	Quantitative rating of relative difficulty in performing fire control work
`pdsi`	Palmer Drought Severity Index (PDSI) uses readily available temperature and precipitation data to estimate relative dryness. It is a standardized index that generally spans -10 (dry) to +10 (wet).

6.2 Maps of Complex Variables

6.3 Breakdown of Fire Causes

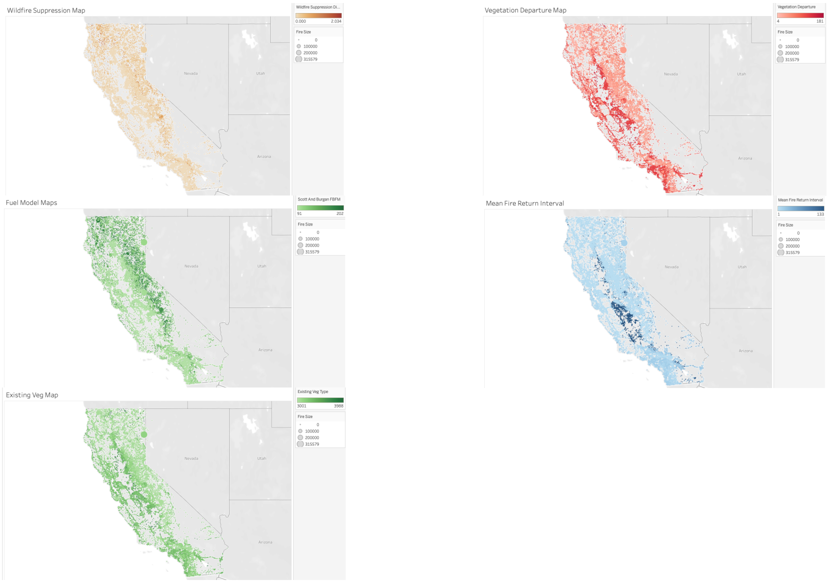


Figure 4: Maps of Complex Variables

Cause	Description
`Arson`	Arson is the crime of willfully and maliciously setting fire to or charring property. Though the act typically involves buildings, the term arson can also refer to the intentional burning of other things, such as motor vehicles, watercraft, or forests. The crime is typically classified as a felony, with instances involving a greater degree of risk to human life or property carrying a stricter penalty. A common motive for arson is to commit insurance fraud. In such cases, a person destroys their own property by burning it and then lies about the cause in order to collect against their insurance policy.
`Campfire`	Fires that originated from human-made campfires unintentionally getting out of hand.
`Children`	Fires started by children
`Debris Burning`	Fire caused by burning debris which often come from open burning on residential property. Open burning is the burning of a bonfire, vegetation debris fire or other fire in an outdoor location where fuel burned is not contained in an incinerator, outdoor fireplace, barbecue grill or barbecue pit.
`Equipment Use`	Most common causes of fires in commercial buildingscome from equipment use, such as microwaves
`Fireworks`	Fireworks can spark wildfires, especially in very dry areas of the country. As a result, the state has banned fireworks in many counties and has a zero tolerance policy for the salend use of illegal fireworks.
`Lightning`	Lightning strikes cause many fires exspecially in the northern Californian forests. Dry thunderstorms, drought, and the abundance of dead trees makes these fires particularly destructive.
`Miscellaneous`	The miscellaneous category includes explosives, glass refraction, shootings, car accidents and oddball things that don't completely fit in another category
`Missing/Unidentified`	The cause of the fire was never identified

Cause	Description
`Powerline`	When powerlines collapse, they often can cause fires. It can be combatted through quality installations as well as shutting off the power grod in times of high speed winds
`Railroad`	Though not very common now, railroads were one of the common causes of wildfires. Smokestacks and brake shoes failing, along with bad vegetation clearance practices for new tracks, started catastrophic fires.
`Smoking`	Smoking materials, including cigarettes, pipes, and cigars, started majority of home structure fires
`Structure`	A structure fire is a fire involving the structural components of various types of residential, commercial or industrial buildings, such as barn fires. Residential buildings range from single-family detached homes and townhouses to apartments and tower blocks, or various commercial buildings ranging from offices to shopping malls.