Analysis of Weather on California Wildfires

Daniel Kim

MOTIVATION

Every year, California wildfires burn hundreds of thousands of acres of land, taking lives, homes, and businesses with them. Over the past few years, these wildfires have become more frequent, impacting my own life and my friends' lives on a much greater scale. For this project, I set out to analyze historical weather and precipitation data in the context of these fires, looking for connections and ultimately a way to predict fire frequency and size from these indicators.

STEPS

1. Data Collection, Cleaning, and Merging:
    a. Fire Data https://gis.data.ca.gov/datasets/CALFIRE-Forestry::california-fire-perimeters-all/explore?location=37.895261%2C-119.826701%2C6.84&showTable=true
    b. Temperature and Precipitation Data https://www.ncdc.noaa.gov/cag/county/time-series/CA-001/tavg/ann/1/1900-2022?base_prd=true&begbaseyear=1901&endbaseyear=2000
    c. Unit Mapping Data https://frap.fire.ca.gov/media/2135/admin_units_13.pdf
2. Data Exploration/Visualization
3. Machine Learning

DATA COLLECTION

The fire data came from the CAL FIRE website. The temperature and precipitation data came from the National Oceanic and Atmospheric Association, and was organized by county (since California has such a diverse range of climates). Since the CAL FIRE data was organized by response units rather than counties, I mapped each unit to the corresponding counties, then used those mappings to find the climate data. Some counties were split by units, so I manually searched major fires within those counties and gave that county to the unit that responded to them. I created a function that requests climate data from the counties within a unit, cleaning it

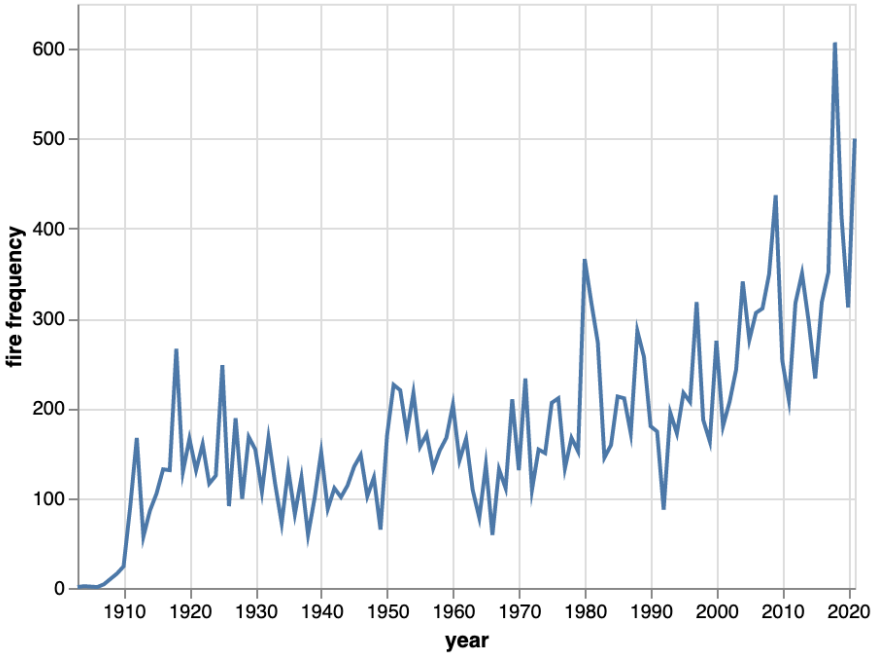
and merging it with those counties' fire data. An example result from one of those calls looks like this.

	temp_value	temp_anomaly	precip_value	precip_anomaly	frequency	acres_sum	acres_mean
1900	51.70	-0.20	67.750	-3.940	0.0	0.000000	0.000000
1901	50.50	-1.40	66.690	-5.000	0.0	0.000000	0.000000
1902	50.90	-1.00	103.350	31.660	0.0	0.000000	0.000000
1903	51.10	-0.80	82.015	10.325	0.0	0.000000	0.000000
1904	51.80	-0.10	107.270	35.580	0.0	0.000000	0.000000
...	...	...	...	...	...	...	...
2017	53.60	1.70	84.640	12.950	6.0	154.809919	25.801653
2018	53.85	1.95	51.540	-20.150	3.0	36.224920	12.074973
2019	53.30	1.40	75.565	3.875	0.0	0.000000	0.000000
2020	54.55	2.65	46.815	-24.875	3.0	122.948170	40.982723
2021	54.40	2.50	64.005	-7.685	0.0	0.000000	0.000000

Notice that the index is on year, so we're dealing with annual data here.

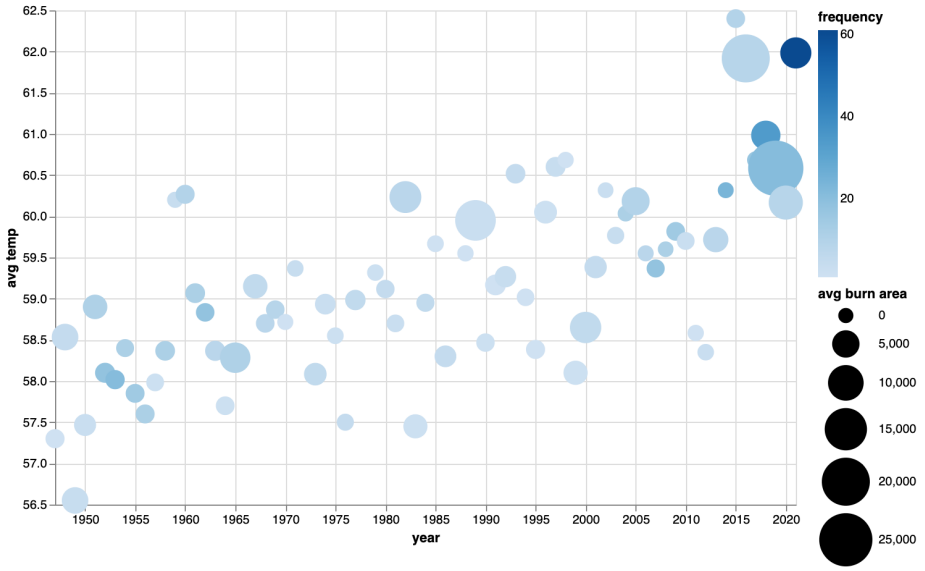
DATA EXPLORATION

Is it just me, or are fires becoming way more frequent?



Turns out they are. Over the last hundred years, fire frequency has been on an upward trend. The spikes in the late 2010's/early 2020's is no doubt from the massive Camp fire that burned down Paradise and the LNU Lightning Complex Fires that nearly burned mine and my friends' houses down.

Temperature and fire frequency/size

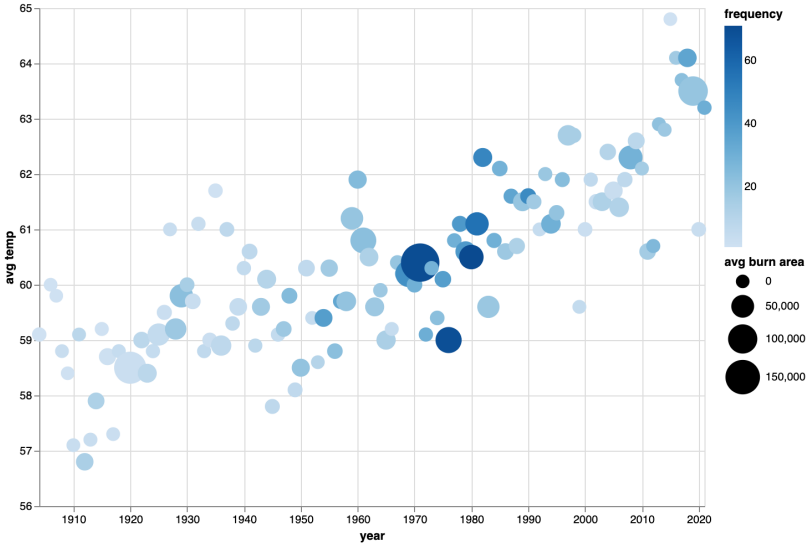


Looking at the LNU group of counties, temperature has definitely increased over time. It also appears that fire frequency and size has increased as well

(which is what I've observed at home).

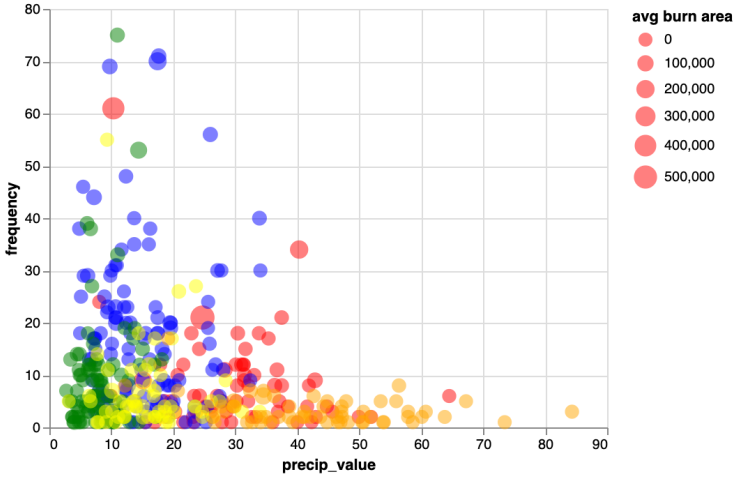
What about counties near LA? Similarly, temperature has increased but frequency and size seems to have peaked and subsided. It's definitely not the primary driver for fires.

I looked at other counties as well, and they showed a similarly faint connection. Perhaps precipitation will have more of the answers I'm looking for.



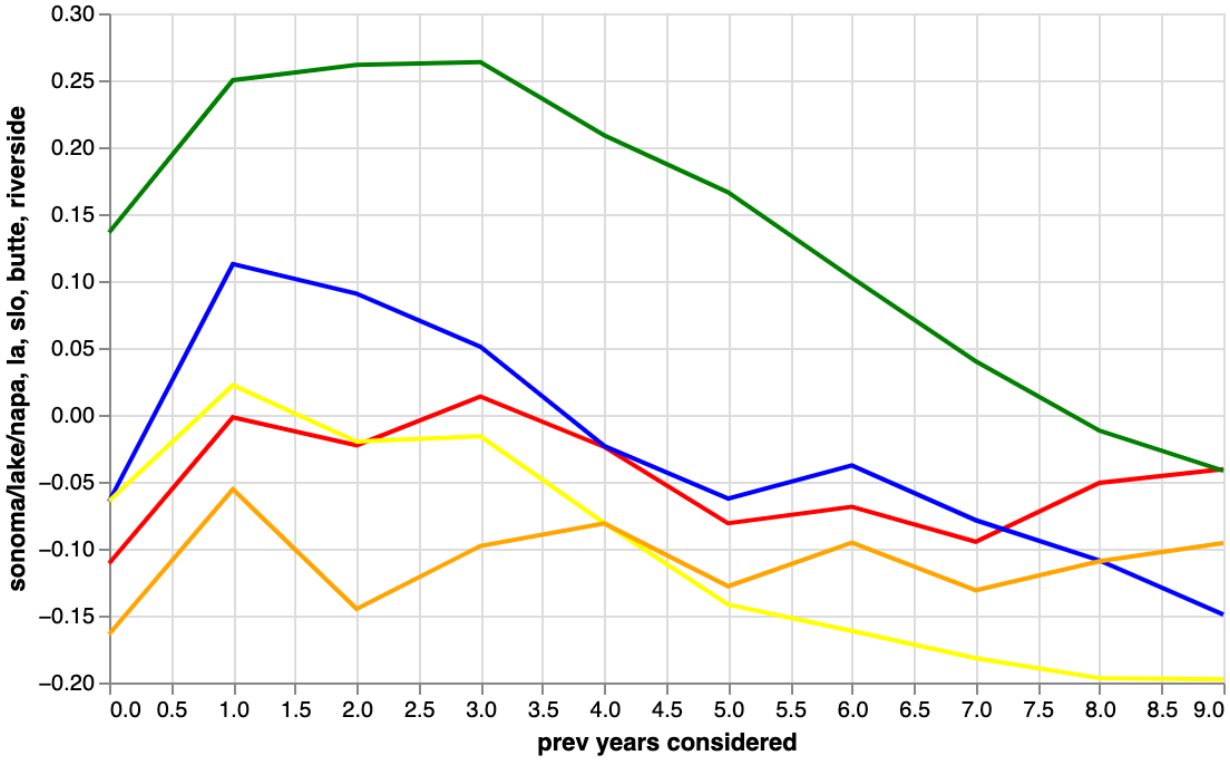
Precipitation and fire frequency/size

Looking at counties scattered around California's various climates, it appears that yearly precipitation might hold some correlation to fire frequency, albeit slight.



Perhaps we need to consider precipitation in the years leading up to a fire.

Now this one is interesting! When we consider average precipitation over the past 1-3 years, the correlation between precipitation and fire frequency increases! Despite the low values, relatively speaking we know that precipitation 1-3 years in the past has stronger correlation.



MACHINE LEARNING

K Nearest Neighbors Regression

Using cross validation to determine the best set of features for the model

```
['temp_value']                                                    79.299974
['temp_anomaly']                                                  79.302103
['precip_value']                                                  79.127282
['precip_anomaly']                                                79.127282
['temp_value', 'temp_anomaly']                                    79.685205
['temp_value', 'precip_value']                                    76.037538
['temp_value', 'precip_anomaly']                                  76.037538
['temp_anomaly', 'precip_value']                                  76.037538
['temp_anomaly', 'precip_anomaly']                                76.037538
['precip_value', 'precip_anomaly']                                79.127282
['temp_value', 'temp_anomaly', 'precip_value']                    74.790487
['temp_value', 'temp_anomaly', 'precip_anomaly']                  74.790487
['temp_value', 'precip_value', 'precip_anomaly']                  72.864359
['temp_anomaly', 'precip_value', 'precip_anomaly']                72.864359
['temp_value', 'temp_anomaly', 'precip_value', 'precip_anomaly']  76.037538
```

Using grid search and cross validation to determine the optimal hyperparameter k for mean absolute error

```
Pipeline(steps=[('standardscaler', StandardScaler()),
                ('kneighborsregressor', KNeighborsRegressor(n_neighbors=13))])
```

Predicting using these parameters showed that correlation between climate and wildfires is not that strong. As you can see, the amount by which a typical guess is off is about the same if not greater than the actual average number of fires per year.

	name	rmse	mae	r2	avg_frequency
0	la	13.408177	9.769231	0.177368	16.270492
1	lnu	7.177229	4.479823	0.190035	4.778689
2	butte	2.174520	1.511349	0.231203	1.532787
3	river	9.528668	5.421185	0.176052	6.795082
4	slo	5.927219	3.213115	0.235293	3.327869
5	vnc	5.539088	2.511349	0.239295	3.245902

Linear Regression

I used the same parameters on a linear regression, and the predictions were similarly off.

	name	rmse	mae	r2	avg_frequency
0	la	13.408177	9.769231	0.177368	16.270492
1	lnu	7.177229	4.479823	0.190035	4.778689
2	butte	2.174520	1.511349	0.231203	1.532787
3	river	9.528668	5.421185	0.176052	6.795082
4	slo	5.927219	3.213115	0.235293	3.327869
5	vnc	5.539088	2.511349	0.239295	3.245902

CONCLUSION:

In conclusion, while California wildfires have been growing more frequent throughout the years, temperature increases and precipitation only play a small part in it. Neither temperature nor precipitation are accurate predictors for the amount of wildfires that may occur in a year, but looking at precipitation 1-3 years back shows a little bit more promise in terms of correlation.

One thing that I would have done differently in retrospect is focus on different explanatory variables. Certain counties had wildfire frequency spikes in the 70's and 80's, subsiding in the 90's and 00's, and spiking again in the late 2010's. I later came across controlled burn statistics showing that California increased controlled burns in the late 80's and early 90's, and then decreased them again in the 2000's, which seemed to correspond with these frequency spikes. I would have also liked to look at the connection between wind speed and fire size, however I made these realizations deep into the project and could not find easily accessible data on them.

1. Collect fire data from CA.gov, weather data from NOAA (limited to precipitation and temperature)
2. Fire frequency definitely increasing: could it be due to increasing temperatures and drought (lack of precipitation)?
3. Inspecting two areas of interest (LAC and LNU), fire frequency does seem to be increasing, size may be as well. Shown with temp → looking good for the correlation
4. Precipitation correlation maybe a little, but not as much → maybe we need to look at precipitation a few years back to see the correlation
    a. Looks like 1-3 years is a sweet spot for the correlation between precipitation and fire frequency!
5. Tried looking at "prescribed burning" data afterwards, but it seems that the fire dataset was not created with those in mind.
6. Looking at some suburban counties (that have a good amount of both cityscape and nature), it seems there was an increase in frequency during one period, and then another spike now
7. Without a rich dataset, summary statistics from news sources paints a picture that fire frequency increased as California did more prescribed burning – this question makes me so curious, I wish I had the data for it.
8. What I wish I could have done: WIND DATA on fire size, PRESCRIBED BURNING DATA on fire frequency → reason: information not available
9. Since all we have is temp and precip, let's try a nearest neighbor model on it.
    a. First find the categories that minimize MSE with cross validation
    b. Now let's find our optimal k value for those categories
    c. Now let's look at how good these variables really are at predicting fire frequency (and only frequency since size didn't seem to be correlated visually) (later tried with size but the result was that it's terrible at predicting – makes sense because fires spread easily and unless it's actively raining they won't necessarily stop)
10. Yikes, it looks like even with our optimal variables and k value, temperature and precipitation don't play a large role in fire frequency.
11. For the heck of it, let's try a linear regression. Looks like a log regression gives us some pretty good predictions (within 1-2 fires per year), but the low correlation shows that this mae and mse (average amount by which predictions are off) are shallow
12. The results are pretty surprising. I was expecting rainfall and temperature increases to be causing more issues, but apparently not. The words "drought" and "fire" are thrown together quite a bit, but these results show they shouldn't be. I wish I could say for certain, but I have much more suspicions about prescribed burning and wind speeds being the true things to consider here. Guess it's up to the next project, but for now, at least we

don’t have to worry about rain and temp increases (at least when it comes to our homes burning down), although they do play a small part.