Stat 196K Final Project

An Analysis of Wildfires in California

By Alec Resha

Overview of project

Goal of project

The goal of this project was to examine trends in wild fires in California over time, based on a data set spanning from 1950 to 2019. The data was retrieved directly from the CA GIS portal which sourced their information from their own surveys with collaboration from Cal Fire. The data is available in multiple formats on the CA State Geoportal website. This website also offers useful visualization tools for viewing the data in non-number based formats. The cleaned and formatted CSV file which was used for this project is available via this google drive link.

Summary of Results

There has been a large increase in the number of fires from 1950-1999 compared to 2000-2019 where the data ends. Breaking these into two sets and retrieving averages, there was a 38% increase in the average number of acres burned per fire. Along with this, the number of fires per year also grew by 45% to an average of 344 fires per year for 2000-2019.

However, while the number and severity of fires grew, the efficiency of firefightes has increased to combat it. From 1950 to 1999, it took an average of 8 days to contain a fire. In 2000 to 2019, it took an

average of 10 days. While the number of days increased by 25%, if fire containment was operating at the same rate as in 1950-1999, we would expect to see an increase of at least 40% from the large increase in the number and severity of fires.

These plots will help to visualize the increase in the number of acres burned per year, the number of fires per year, and the average number of acres that are burned per fire. The average number of acres burned in each fire grows the slowest of the 3 graphed variables, but it is still important to note the upward trend that is visible in all three.

Dataset

The dataset used for this project is available here. This was obtained from the Cal-Fire website. It contains the following columns of information: OBJECTID, YEAR, STATE, AGENCY, UNIT_ID, FIRE_NAME, INC_NUM, ALARM_DATE, CONT_DATE, CAUSE, COMMENTS, REPORT_AC, GIS_ACRES, C_METHOD, OBJECTIVE, FIRE_NUM, GlobalID, SHAPE_LENGTH, SHAPE_Area on 15,565 unique fires since 1950. Not every entry orginated in California, but the included ones did reach California or it's borders. Only 0.8% of the entries are from out of state fires. The direct website shows a visualization of the region for fires as well as more interactive visualization tools.

General Statistics

Yearly Comparison

YEAR_	Number of Fires	Acres Burned	Avg Acres Per Fire
2018	413	1,590,431.15	3,850.92
2017	607	1,424,559.38	2,346.89
2008	438	1,382,462.25	3,156.31
2007	349	1,040,224.30	2,980.59
2003	341	970,479.30	2,845.98
2002	243	963,898.51	3,966.66
1987	286	862,910.91	3,017.17
2012	351	847,714.60	2,415.14
1999	279	801,137.00	2,871.46
2015	318	789,205.99	2,481.78
2006	315	744,764.65	2,364.33

YEAR_	Number of Fires	Acres Burned	Avg Acres Per Fire
1970	233	676,019.52	2,901.37
1996	321	672,216.81	2,094.13
2013	298	569,820.41	1,912.15
1985	211	568,393.59	2,693.81
2016	352	545,984.62	1,551.09
1950	227	538,277.14	2,371.26
1977	167	465,615.28	2,788.12
2009	254	435,839.54	1,715.90
1955	172	386,057.68	2,244.52
1980	319	380,262.59	1,192.05
1979	367	379,128.77	1,033.05
1994	218	377,553.18	1,731.90
2014	232	375,949.59	1,620.47
1990	175	371,804.34	2,124.60

Table generated with:

```
USE finalProject;
SELECT YEAR_,
    count(*) AS "Number of Fires",
    FORMAT (sum(GIS_ACRES), 2) AS "Acres Burned",
    FORMAT (sum(GIS_ACRES)/count(*), 2) AS "Avg Acres Per Fire"
FROM firePerim
GROUP BY YEAR_
ORDER BY sum(GIS_ACRES) desc
LIMIT 25;
```

1950-2000 Vs 2000-2019

Start Year	Number of Fires	Acres Burned	Avg Acres Per Fire
2000	6330	13,293,245.09	2,100.04
1950	9426	14,319,685.50	1,519.17

Analysis of Fire Growth

Based on these two tables, there has been a definite rise in the number and severity of fires. The first list is a majority from 2000 and later. Using the second table, we can see that while 1950-2000 had approximately 50% more fires in this time than 2000-2019, it was also over double the length of time. Adding to the idea that fire severity is worsening, 2000-2019 is already at 92% of the acres burned in the prior 50 years, and there has been a 38% increase in the number of acres burned per fire.
To standardize the numbers, the average number of fires has also increased, from 189 per year for 1950-2000, to 334 per year for 2000-2019.

Table generated with:

```
USE finalProject;
SELECT MIN(YEAR_) AS "Start Year",
    COUNT(*) AS "Number of Fires",
    FORMAT (SUM(GIS_ACRES), 2) as "Acres Burned",
    FORMAT (SUM(GIS_ACRES)/count(*), 2) as "Avg Acres Per Fire"
FROM firePerim
WHERE YEAR_ >= 2000
UNION
SELECT MIN(YEAR_) AS "Start Year",
    COUNT(*) AS "Number of FIRES",
    FORMAT (SUM(GIS_ACRES), 2) as "Acres Burned",
    FORMAT (SUM(GIS_ACRES)/count(*), 2) as "Avg Acres Per Fire"
FROM firePerim
WHERE YEAR_ <= 2000;
```

Fire Containment

Start Year	AVG days until containment	Longest Burning Fire	YEAR	FIRE_NAME
2000	9.56	377	2007	OCTOBER
1950	7.83	154	1979	NACIMIENTO

Analysis of Fire Containment

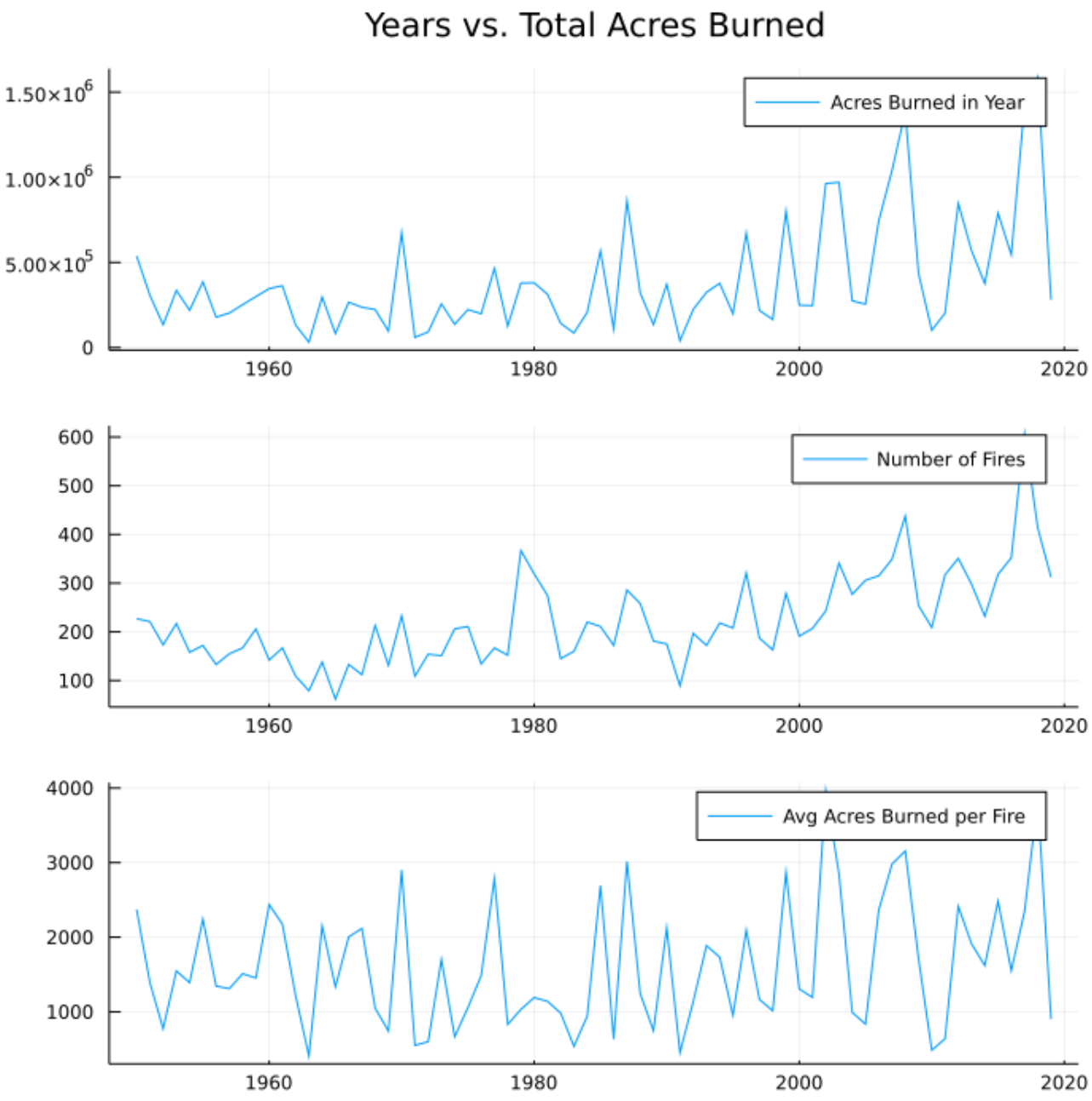
While the number and severity of fires has increase dramatically, the number of days from the fire being reported to the fire being contained has not increased as dramatically. The number of days increased from 8 to 10 days between 1950-2000 to 2000-2019, about a 25% incrase. While this is still a large percentage increase, it is smaller than both the increase percentage in fires, and in fire severity. This increase in the number of days to put out the fires is entirely justified by the large increase in the severity of fires, and means the fires have actually been put out more efficiently.

Table generated with:

(Length of fires limited to 0 to 1000 days to filter out mis-entered data that severly skews results)

```sql
USE finalProject;
SELECT MIN(YEAR_) AS "Start Year",
    FORMAT (AVG(DATEDIFF(CONT_DATE, ALARM_DATE)), 2) as "AVG days until containment
    MAX(DATEDIFF(CONT_DATE, ALARM_DATE)) as "Longest Burning Fire",
    YEAR_,
    FIRE_NAME
FROM firePerim
WHERE YEAR_ >= 2000
    AND DATEDIFF(CONT_DATE, ALARM_DATE) >= 0
    AND DATEDIFF(CONT_DATE, ALARM_DATE) <= 1000
UNION
SELECT MIN(YEAR_) AS "Start Year",
    FORMAT (AVG(DATEDIFF(CONT_DATE, ALARM_DATE)), 2) as "AVG days until containment
    MAX(DATEDIFF(CONT_DATE, ALARM_DATE)) as "Longest Burning Fire",
    YEAR_,
    FIRE_NAME
FROM firePerim
WHERE YEAR_ <= 2000
    AND DATEDIFF(CONT_DATE, ALARM_DATE) >= 0
    AND DATEDIFF(CONT_DATE, ALARM_DATE) <= 1000;
```

Plotting Annual Data



Analysis of plots

All three plots above show a clear increase in the issue of fires in California. The number of fires shows the most dramatic rise, with the end of the graph spiking much higher than previous years, but also the general upward trend of all three variables. The average number of acres per fire has the slowest growth, but in 2000 to 2019 there is a clear increase in the number of annual fires. There are two years in the 2000-2019 range that had more acres burned than any previous year. These two years alone likely skew the average acres burned for this time period, but the intermediate years are all still highed than the lows in the previous years.

Plots generated with:

```
using Plots;
```

```
f = open("fireData.txt");                                    # Loading data
year = Array{Integer}(undef, 0)                              # Initializing arrays
fireCount = Array{Integer}(undef, 0)
acresBurned = Array{Float64}(undef, 0)
acresPerFire = Array{Float64}(undef, 0)
for line in eachline(f)
    temp = replace.(split(line, "\t", limit = 4), "," => "") # Formatting data
    append!(year, parse(Int64, temp[1]))                     # Loading data into va
    append!(fireCount, parse(Int64, temp[2]))
    append!(acresBurned, parse(Float64, temp[3]))
    append!(acresPerFire, parse(Float64, temp[4]))
end
p1 = plot(year, acresBurned, title="Years vs. Total Acres Burned",
         label="Acres Burned in Year");                      # Plotting total acres
p2 = plot(year, fireCount, label="Number of Fires");         # Plotting number of f
p3 = plot(year, acresPerFire, label="Avg Acres Burned per Fire"); # Plotting avg ac
plot(p1, p2, p3, layout = (3,1), size=(700,700))             # Creating stacked gra
```

Based on a list of annual amounts for number of fires, fire size, and average annual fire size, stored as "fireData.txt" from the following SQL query:

```
USE finalProject;
SELECT YEAR_,
    COUNT(*) AS "Number of Fires",
    FORMAT (SUM(GIS_ACRES), 2) as "Acres Burned",
    FORMAT (AVG(GIS_ACRES), 2) as "Avg Acres Per Fire"
FROM firePerim
GROUP BY YEAR_
ORDER BY YEAR_ ASC;
```