Capstone Project1: Identifying counties that are the most fire-prone and predicting the cause of a fire wildfire.

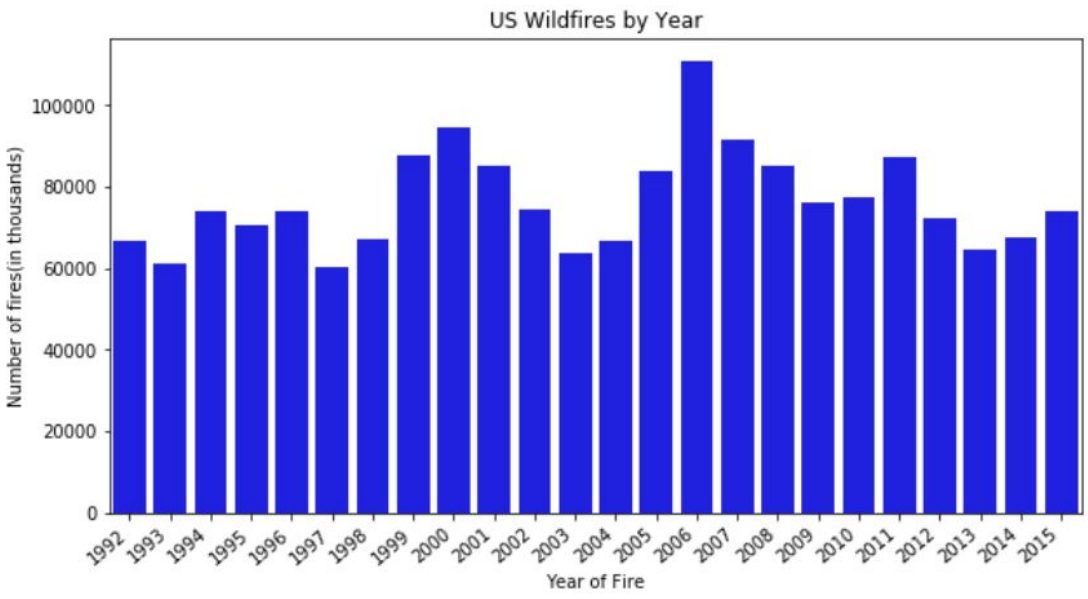
Capstone 1 Data Story Data source :https://www.kaggle.com/rtatman/188-million-us-wildfires

In this exploration of the US Wildfires data we would try to answer below questions

1. Is Global warming affecting the number of fires? Has the number of fires increased over the period 1992-2015?
2. Are we able to limit fire spread with help of growing technology? Has the size of fire decreased over years?
3. What causes the most fires? Which causes are associated with larger wildfires?
4. Which State in the USA is most affected by WildFires? Which county is more prone or less prone to WildFire?

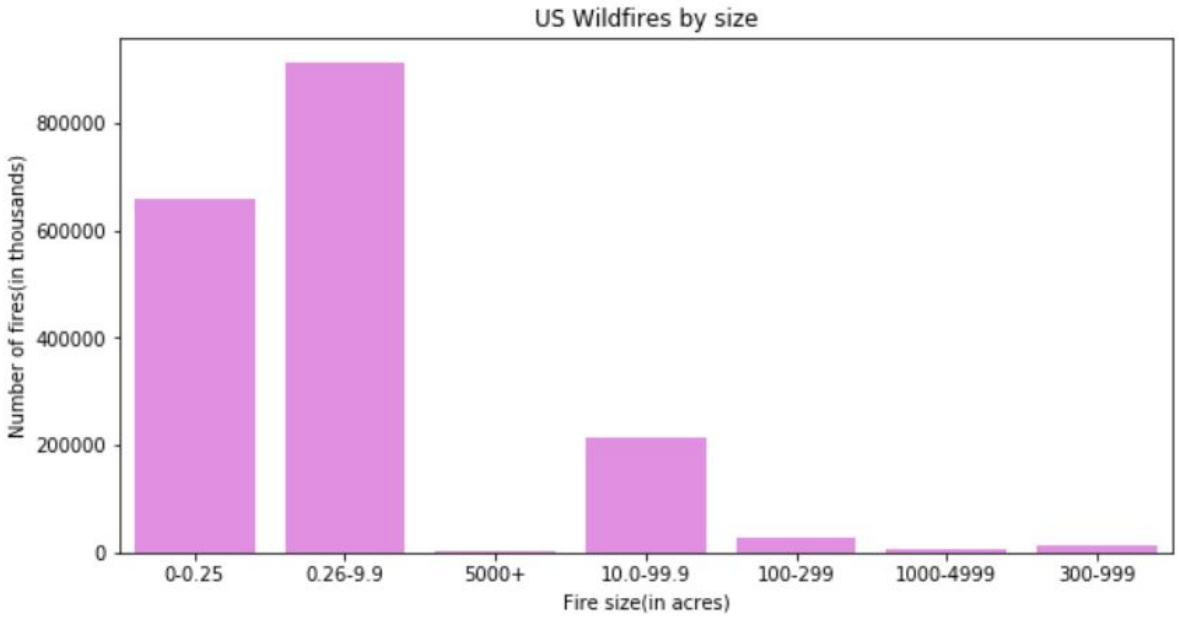
5. What is causing more fires in each state?

To find an answer to question "1. Is Global warming affecting the number of fires? Have the number of fires increased over the period 1992-2015?" Plotting number of fires per year by counting values in 'FIRE_YEAR' column.

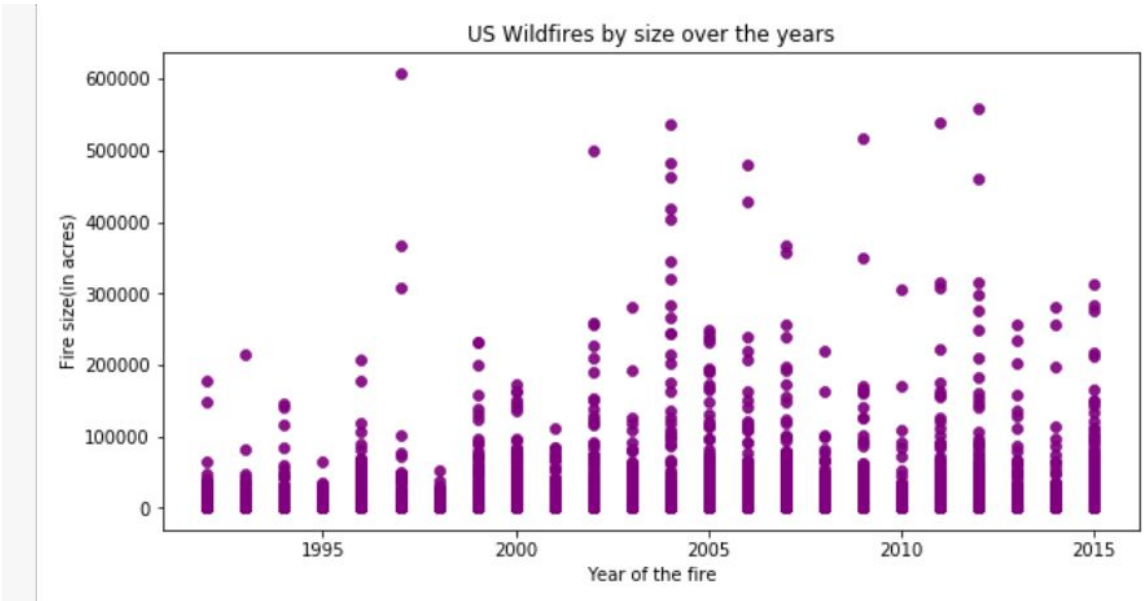


The number of fires per year ran between 60,000 and 100,000 from 1992 to 2015. There was a spike in fires in 2006. Though we can see a small upward trend at certain time periods, there is no continuous upward trend over the period.

To answer question "2. Are we able to limit fire spread with help of growing technology?Has the size of fire decreased over years?" Plotting FIRE_SIZE_CLASS by counting values in 'FIRE_SIZE_CLASS' column.

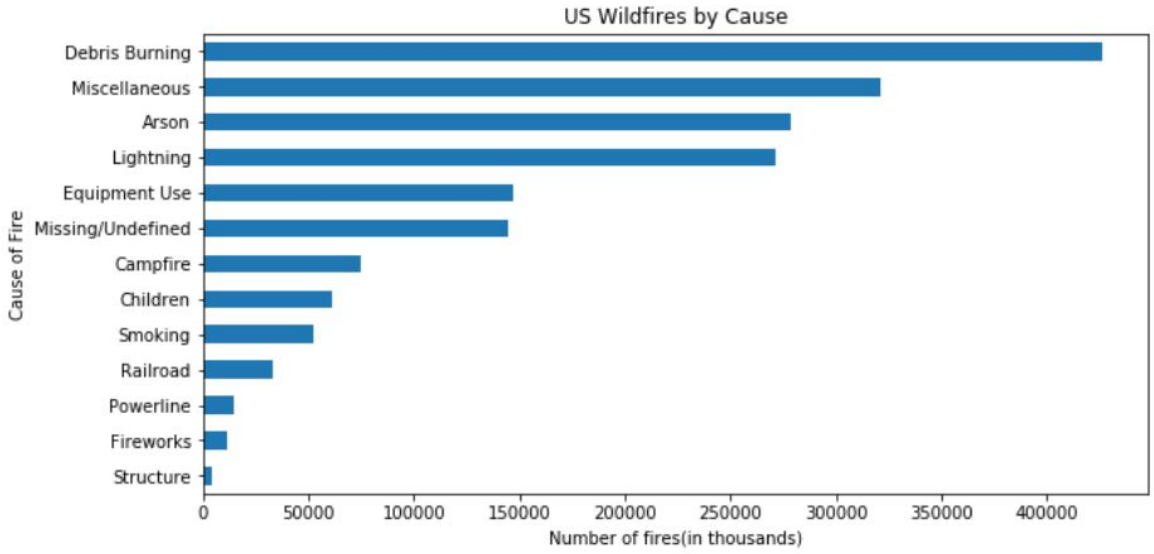


We can see in the above graph that over 800k fire incidents are in between 0.26-0.99 acres within the final fire perimeter expenditures.Now to see if concerned departments are able to restrict the fire size with the help of modern technology over the years, plotting fire year against fire size.

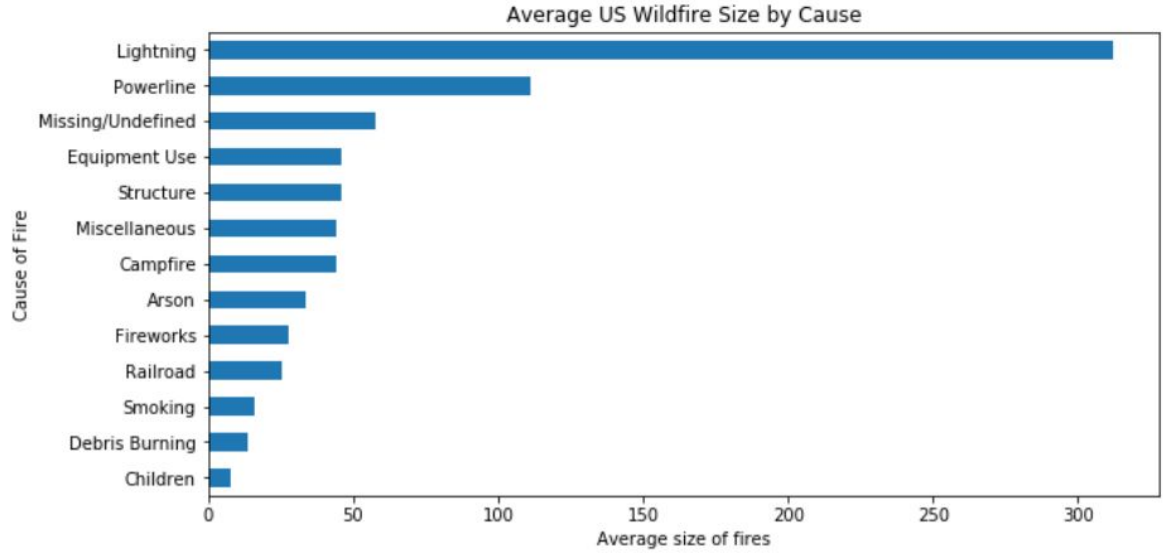


There is no decrease or increase of fire size from 1992-2015.Every year has recorded fire of different sizes.

Now that we have seen that number of fires or its size has not changed over years.Let us look at the cause of these Wildfires. What causes the most fires? Which causes are associated with larger and longer-burning wildfires?

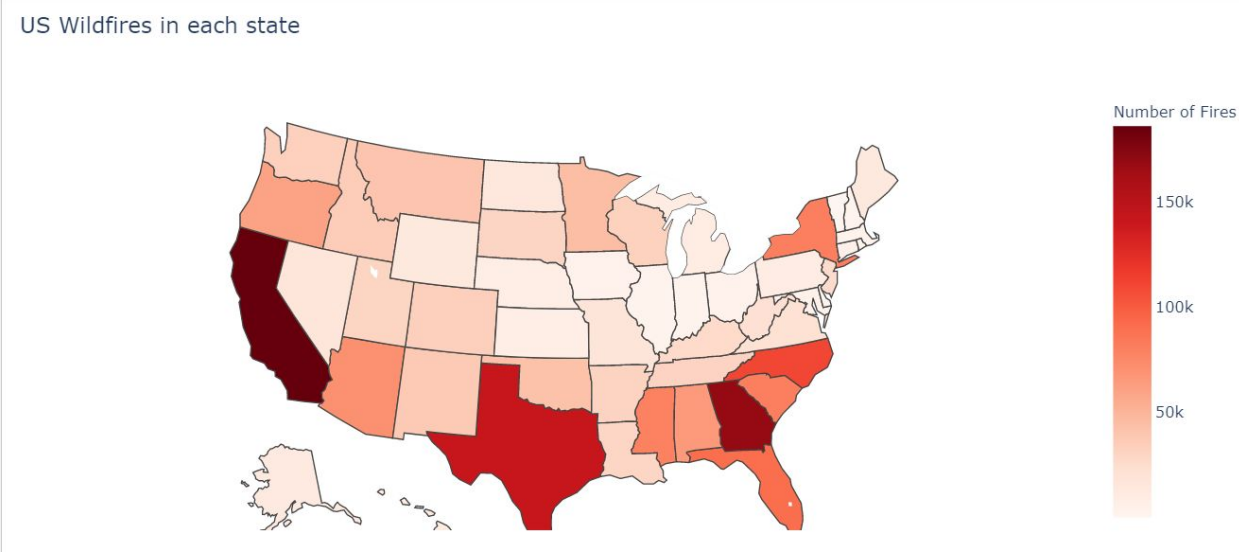


Trash burning was the largest cause of wildfire by a significant margin. Interestingly, slightly more fires were started by arson than by lightning. I am surprised that one of the causes is just ‘children’. This leads to a question. Is there a relationship between cause and fire size?



We can observe that fires started by electricity are the most damaging. Though the number of fires caused by debris burning is large in number, the average size of the fire caused by it is very low.

To answer questions like Which State in the USA is most affected by WildFires? Which county is more prone or less prone to WildFire? Let’s compare the number of wildfires by state.

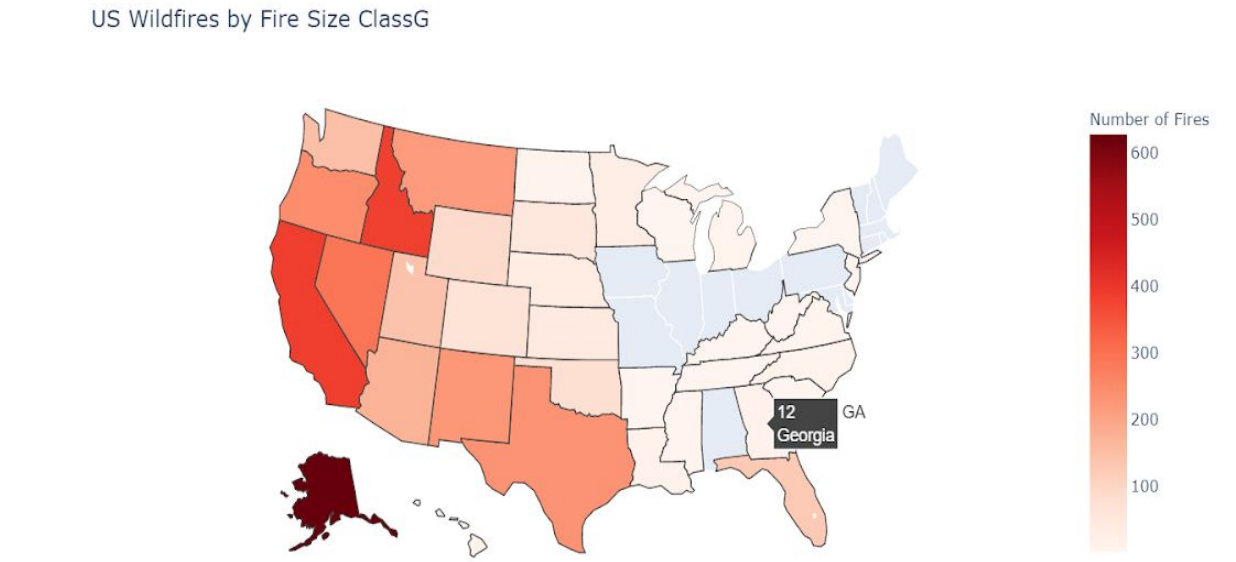


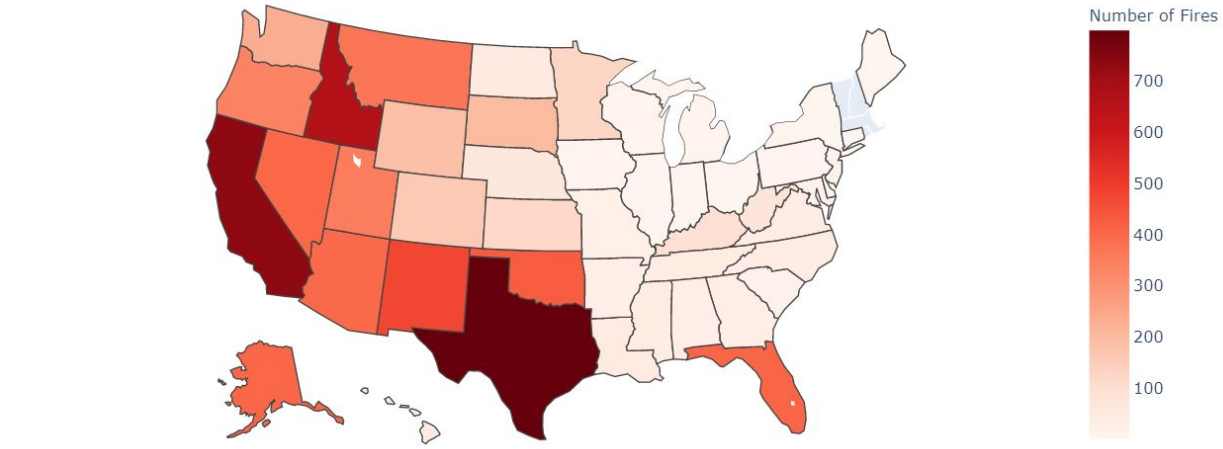
As we might expect California and Texas have the most wildfires due to sheer size and climate. However, it is surprising to see Georgia with so many fires.

Though Georgia reported more number of fires than other bigger states, I am curious about Fire size of fires reported in Georgia . To know more about this let us compare the number of fires per state with respect to fire size reported.

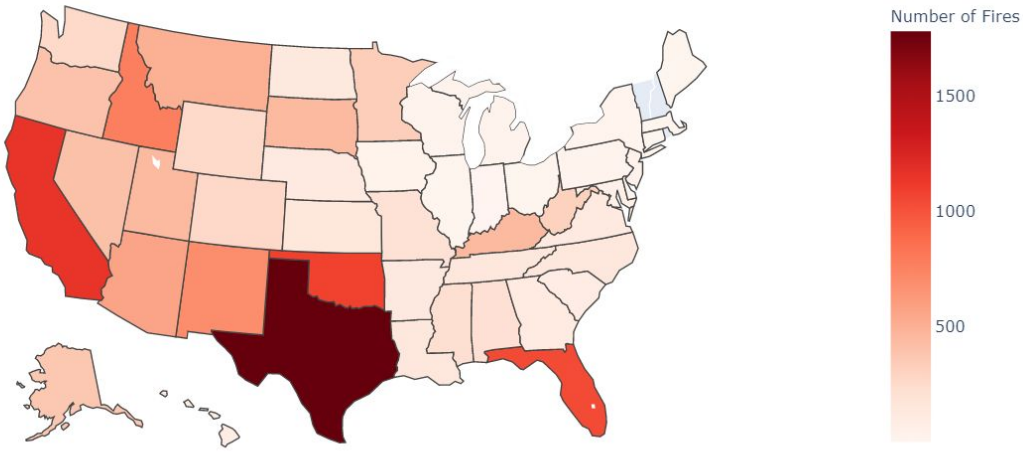
From DataSet we know that **Size 'A' represents '0-0.25' acres,Size 'B' represents '0.26-9.9' acres,Size 'C' represents '10.0-99.9' acres,Size 'D' represents '100-299' acres,Size 'E' represents '300-999' acres,Size 'F' represents '000-4999' acres and Size 'G' represents '5000+' acres.**

We can see that Georgia has 12 fires over the span of two decades which are of size ‘5000”’ acres and Alaska has the most number of fires of largest size.

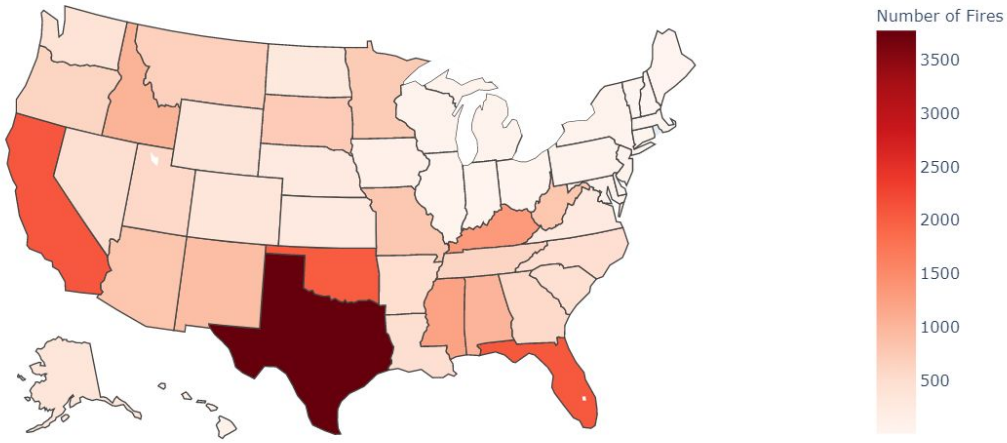




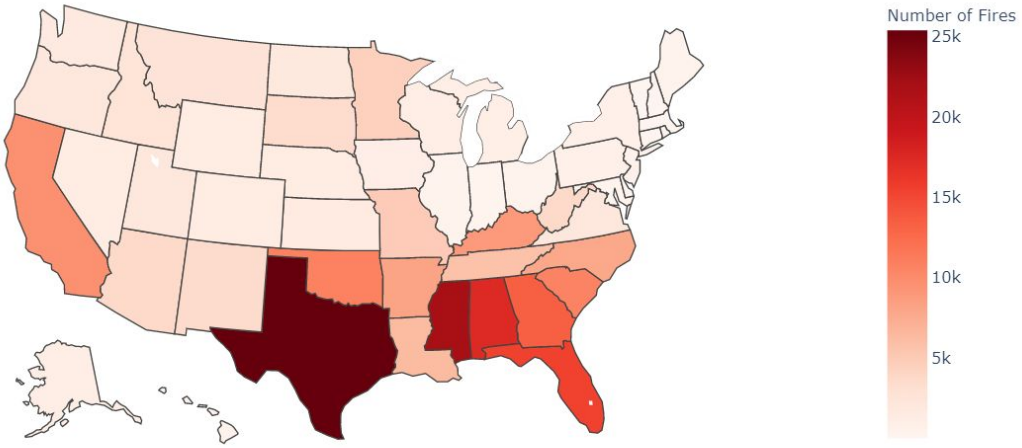
US Wildfires by Fire Size ClassE



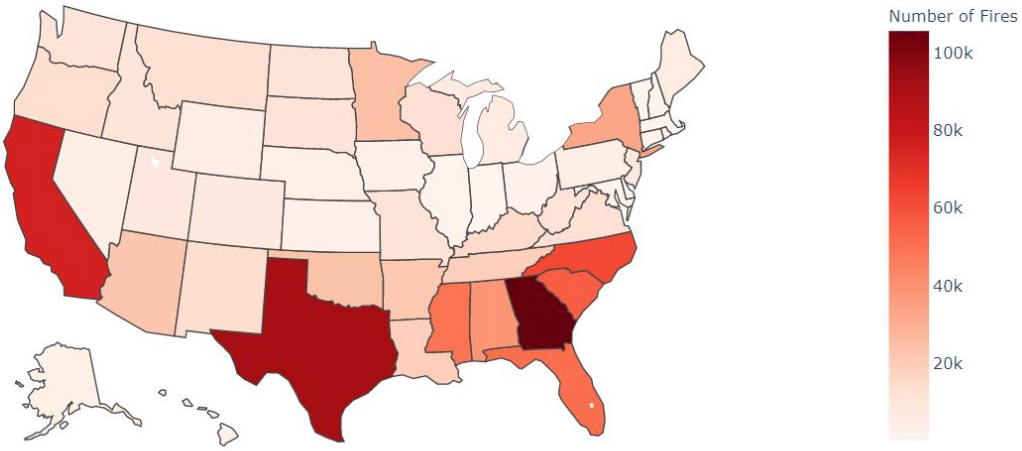
US Wildfires by Fire Size ClassD



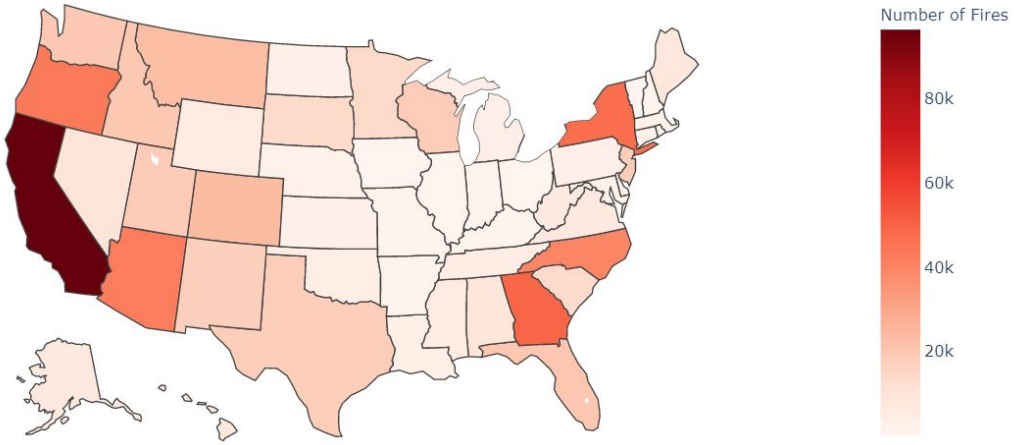
US Wildfires by Fire Size ClassC



US Wildfires by Fire Size ClassB



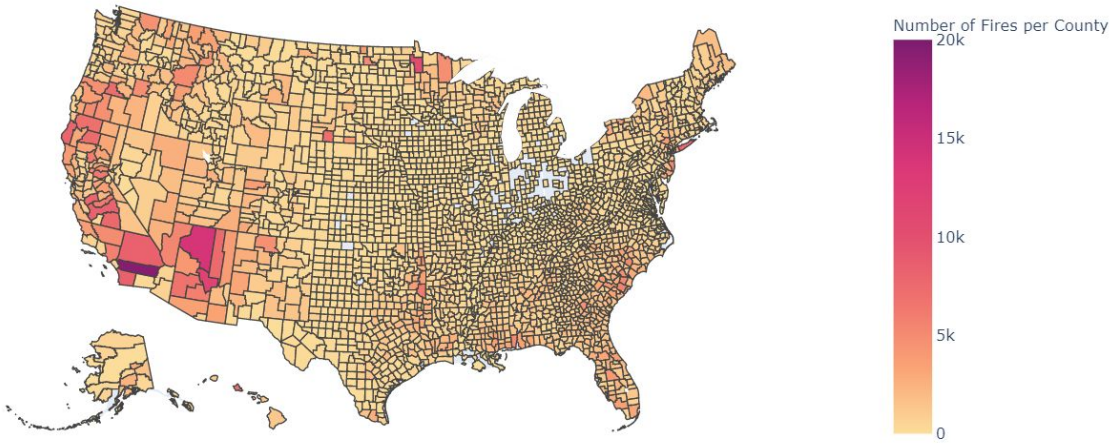
US Wildfires by Fire Size ClassA



After looking at State maps by Fire size class, we can note that Georgia has most fires with size under 100 acres. However, California and Texas have reported fires of all sizes in more numbers over two decades.

Now let's look at fires by county.

US Wildfires by County



	FIPS_CODE	County_Count	COUNTY_NAME	STATE_NAME
0	6065.0	19398	Riverside	California
1	4005.0	13864	Coconino	Arizona
2	4007.0	11662	Gila	Arizona
3	27007.0	10518	Beltrami	Minnesota
4	6071.0	8301	San Bernardino	California
5	6073.0	8271	San Diego	California
6	6019.0	8229	Fresno	California
7	6023.0	7963	Humboldt	California
8	6047.0	7894	Merced	California
9	36103.0	7663	Suffolk	New York

From the above map we can know that “Riverside County” in California has more number of fires. Also we can see that counties like “Hamilton-Indiana” have only 1 case over the period of 1992-2015. We can see that Riverside County of California has the highest number of fire incidents. However, we cannot assume that more fires have caused more damage. Let us see which counties are most affected in terms of acres of land burnt due to wildfire.

	FIPS_CODE	FIRE_SIZE_mean	FIRE_SIZE_min	FIRE_SIZE_max	COUNTY_NAME	STATE_NAME
91	2290.0	34400.542941	0.10	312918.3	Yukon-Koyukuk	Alaska
910	20187.0	13467.000000	1.00	40000.0	Stanton	Nebraska
2081	40153.0	9360.666667	604.00	23488.0	Woodward	Oklahoma
823	20007.0	9083.589286	0.25	70000.0	Barber	Kansas
1540	30073.0	8983.950000	0.10	53640.0	Pondera	Montana
71	2068.0	7770.115425	0.10	517078.0	Denali Borough	Alaska
72	2070.0	7342.829783	0.10	606945.0	Dillingham Census Area	Alaska
70	2050.0	6050.744675	0.10	308120.0	Bethel Census Area	Alaska
2527	48263.0	6007.200000	1.00	162625.0	Kent	Rhode Island
2028	40045.0	5000.000000	5000.00	5000.0	Ellis	Texas

Now the picture of counties affected by wildfire has completely changed. Yukon-Koyukuk is most affected county over the period of time. Also it is interesting to note that though the number of fires were more in counties of California,Arizona,Minnesota and New York, more destructive fires are reported in counties of Alaska,Nebraska and Oklahoma.

```
grouped_countyFireData[grouped_countyFireData['COUNTY_NAME'] == 'Riverside']
```

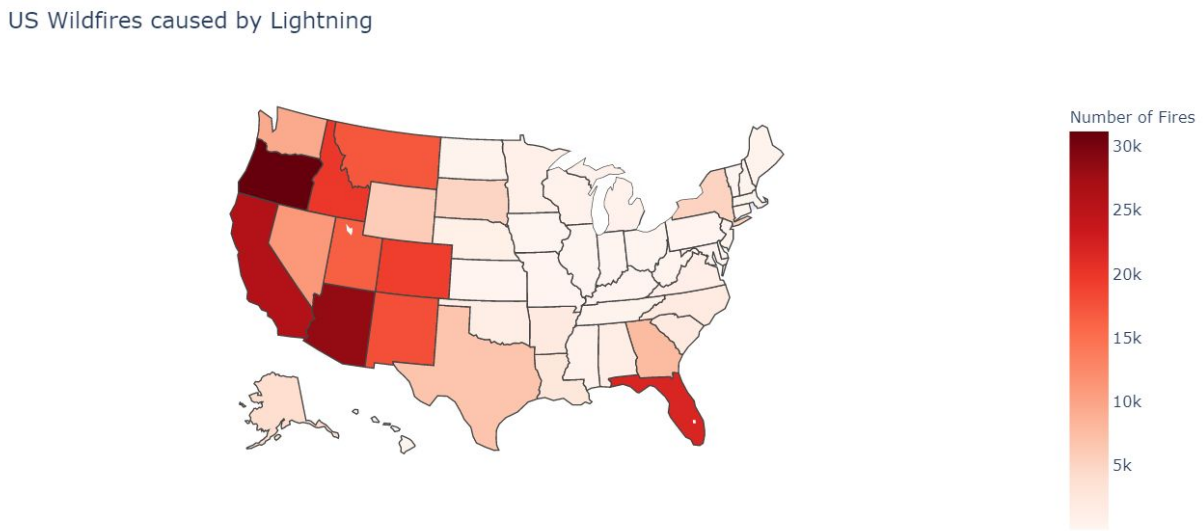
	FIPS_CODE	FIRE_SIZE_mean	FIRE_SIZE_min	FIRE_SIZE_max	COUNTY_NAME	STATE_NAME
214	6065.0	30.995669	0.01	40200.0	Riverside	California

```
countyFireData[countyFireData['COUNTY_NAME']=='Yukon-Koyukuk']
```

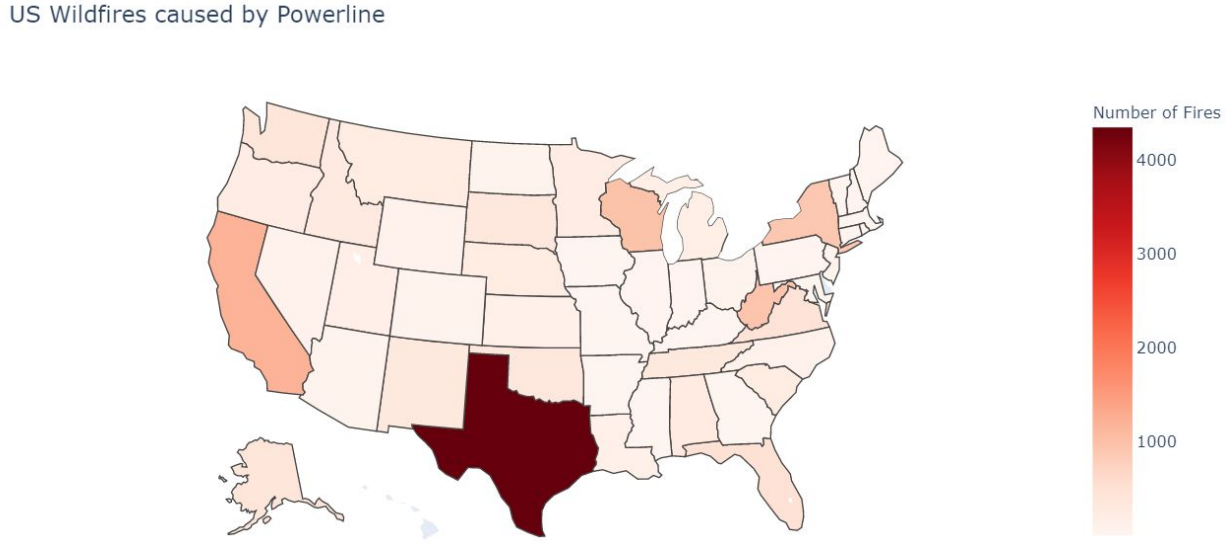
	FIPS_CODE	County_Count	COUNTY_NAME	STATE_NAME
2228	2290.0	51	Yukon-Koyukuk	Alaska

Riverside County of California reported 19398 fire incidents over the period of 13 years with average fire size being approximately 31 acres. However, Yukon-Koyukuk county of Alaska reported only 51 fire incidents over the same period of time but with an average fire size of approximately 35000 acres.

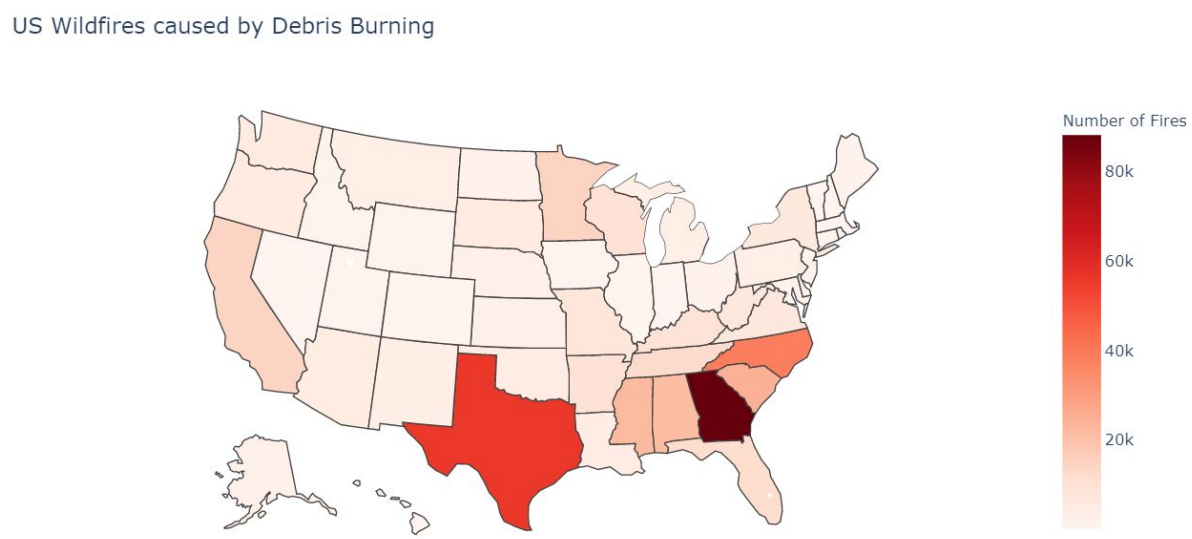
Though we see the number of fires per state and county in above maps, it is not clear what is causing these fires in any given state. If we know the reason behind these wildfires, State/County departments can take appropriate precautionary actions depending on the most likely reason for fire in their state/county.



We can note that many states are affected by lighting.

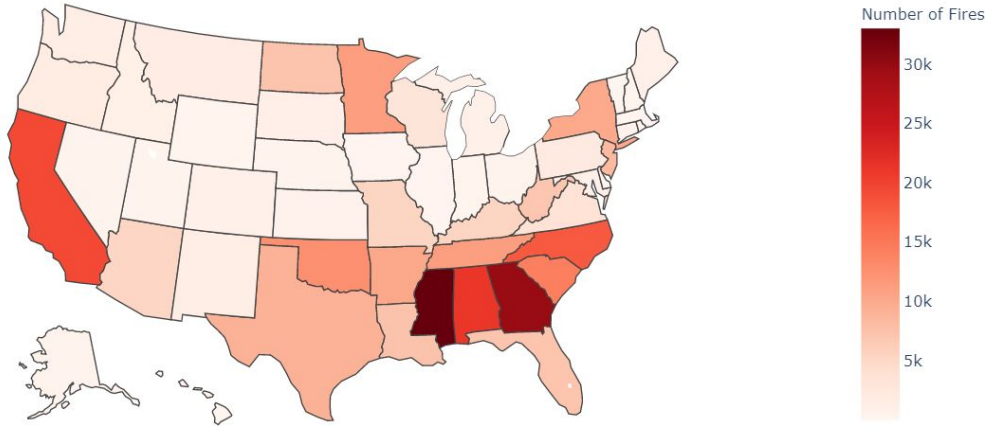


Powerline has caused around 4K wildfires in Texas.



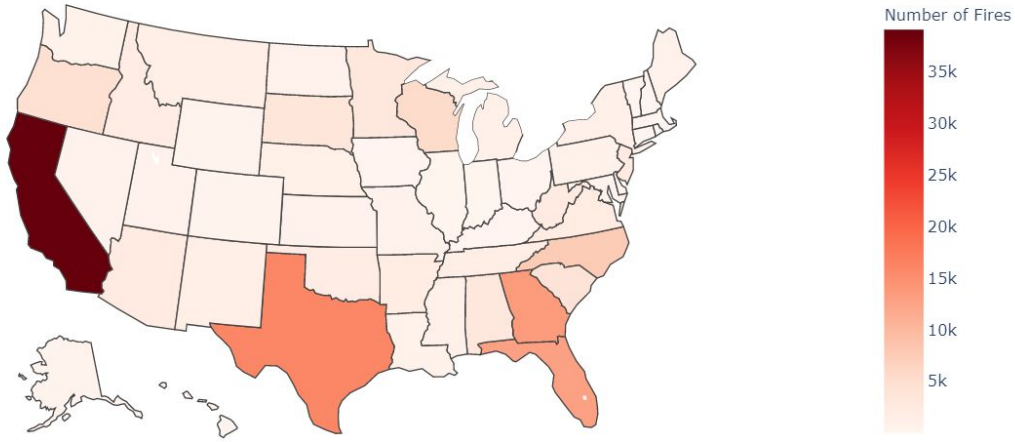
Looks like trash burning is the cause for almost all the cases in Georgia.

US Wildfires caused by Arson

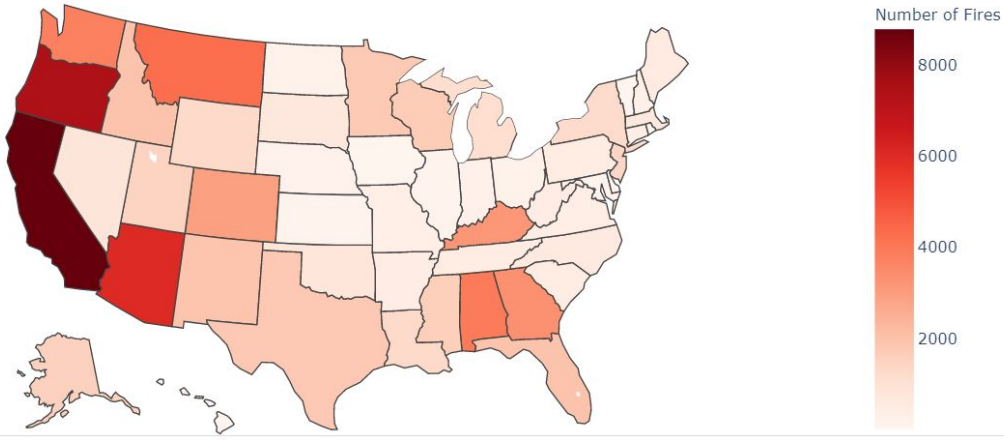


Mississippi has 30k wildfires caused by trash burning and arson alone.

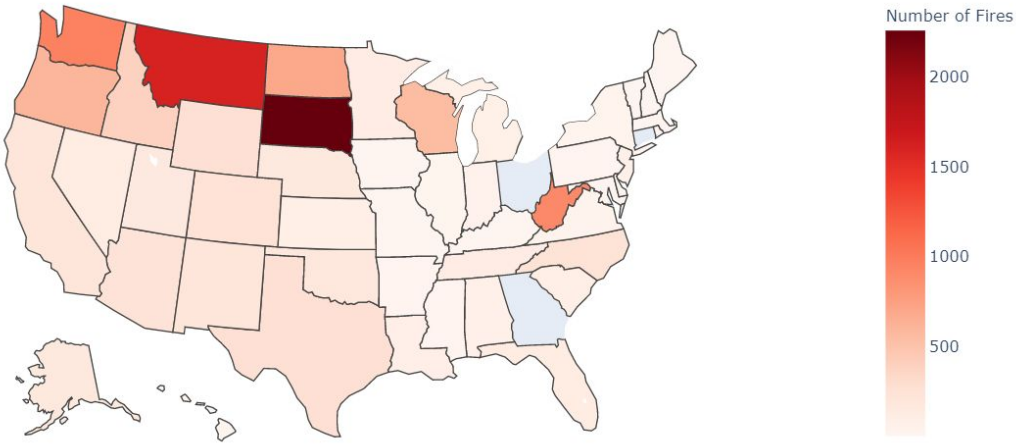
US Wildfires caused by Equipment Use



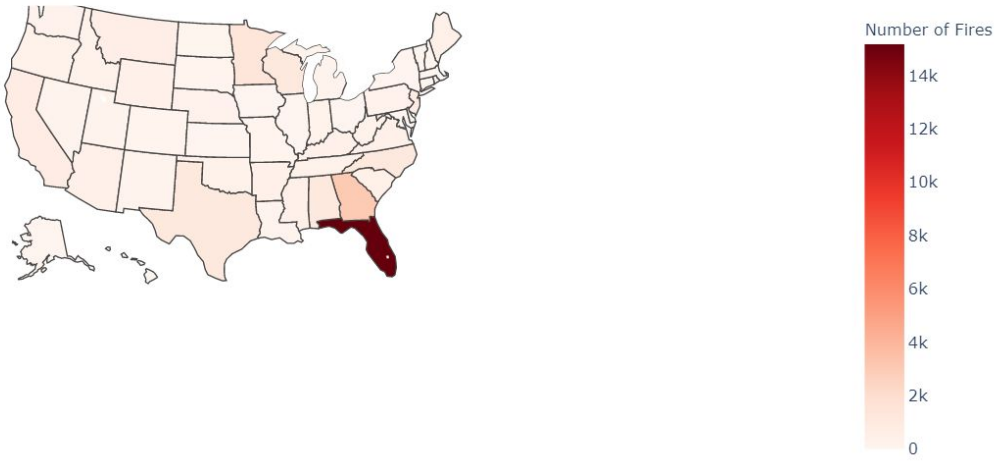
US Wildfires caused by Campfire



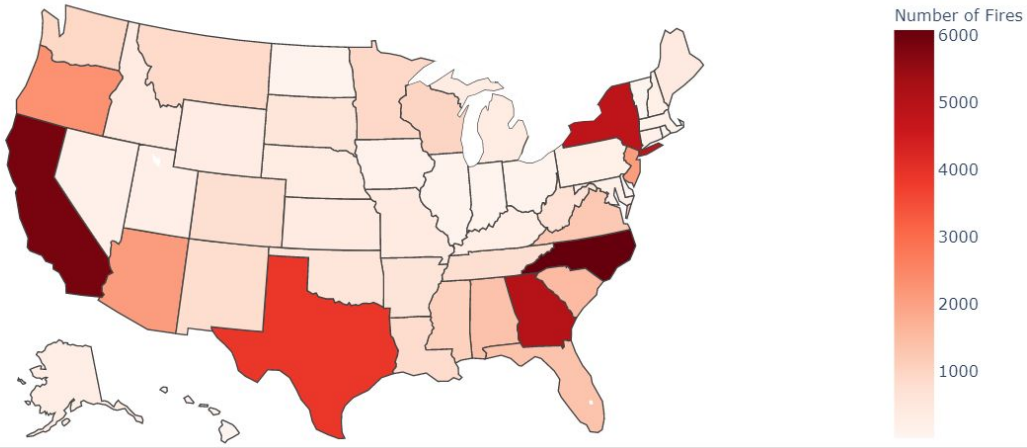
US Wildfires caused by Fireworks



US Wildfires caused by Railroad



US Wildfires caused by Smoking



Below is the link to Github repository of jupyter notebook file with Data wrangling code.
https://github.com/lasyabheemendra/Sprigboard-DatascienceProjects/blob/master/Capstone1_US-Wildfire-Prediction/DataStoryTelling.ipynb