1. **The Data Set**
2. **Overview of the Data Set**

The data set used in this analysis is called "fire\_archive\_M-C61\_234859.csv". It was acquired from NASA Fire Information for Resource Management Systems(FIRMS) at the URL: <https://firms.modaps.eosdis.nasa.gov/download/>. It contains fire data from November 2000 to April 2021. The data was originally collected by NASA Moderate Resolution Imaging Spectroradiometer (MODIS) and is one of Collection 6.1 products. A user guide of this product can be found at the URL:

<https://cdn.earthdata.nasa.gov/conduit/upload/3865/MODIS_C6_Fire_User_Guide_A.pdf>

1. **Data Attribute Fields**

This data set have 13 attribute fields; the descriptions of attribute fields are shown in the table below, and can also be found at the following URL: <https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/mcd14dl>

|  |  |
| --- | --- |
| Attribute | Long Description |
| latitude | Center of 1km fire pixel but not necessarily the actual location of the fire as one or more fires can be detected within the 1km pixel. |
| longitude | Center of 1km fire pixel but not necessarily the actual location of the fire as one or more fires can be detected within the 1km pixel. |
| brightness | Channel 21/22 brightness temperature of the fire pixel measured in Kelvin. |
| scan | The algorithm produces 1km fire pixels but MODIS pixels get bigger toward the edge of scan. Scan and track reflect actual pixel size. |
| track | The algorithm produces 1km fire pixels but MODIS pixels get bigger toward the edge of scan. Scan and track reflect actual pixel size. |
| acq\_date | Data of MODIS acquisition. |
| acq\_time | Time of acquisition/overpass of the satellite (in UTC). |
| satellite | A = Aqua and T = Terra. |
| confidence | This value is based on a collection of intermediate algorithm quantities used in the detection process. It is intended to help users gauge the quality of individual hotspot/fire pixels. Confidence estimates range between 0 and 100% and are assigned one of the three fire classes (low-confidence fire, nominal-confidence fire, or high-confidence fire). |
| version | Version identifies the collection (e.g. MODIS Collection 6) and source of data processing: Near Real-Time (NRT suffix added to collection) or Standard Processing (collection only). "6.1NRT" - Collection 61 NRT processing  "6.1" - Collection 61 Standard processing Find out more on collections and on the differences between FIRMS data sourced from LANCE FIRMS and University of Maryland. |
| bright\_T31 | Channel 31 brightness temperature of the fire pixel measured in Kelvin. |
| frp | Depicts the pixel-integrated fire radiative power in MW (megawatts). |
| type\* | 0 = presumed vegetation fire 1 = active volcano 2 = other static land source 3 = offshore |

1. **Techniques Used and Analysis Focuses**

We used PySpark to analyze the data set to see how fire frequency and fire intensity have changed over the years. The techniques used include aggregation, pivoting, window functions etc. The analysis results are visualized with graphs.

1. **Preliminary Data Preparation**

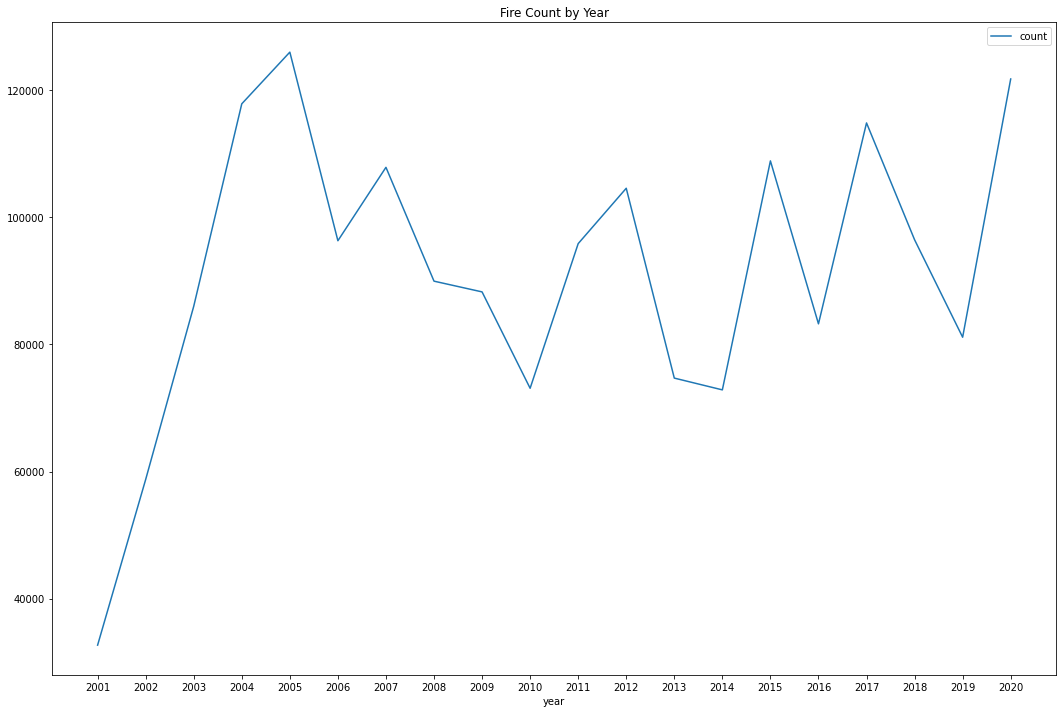
According to the descriptions of the attribute fields, “type” filed has four different values: 0, 1, 2 and 3, representing “presumed vegetation fire”, “active volcano”, “other static land source”, and “offshore detection (includes all detections over water)” respectively. We focus on vegetation fire in our analysis, so we filtered the data set with “where(col("type") == 0)”.

Besides, “confidence” field estimates range between 0 and 100% and are assigned one of the three fire classes (low-confidence fire, nominal-confidence fire, or high-confidence fire). We focus on fires with confidence equals to or more than 50%.Therefore, we filtered the data set with “where(col("confidence") >= 50) ”.

The original data set have a field “acq\_date”, which indicates the date when the hotspot was detected; we added two columns “year” and “month” based on this field.

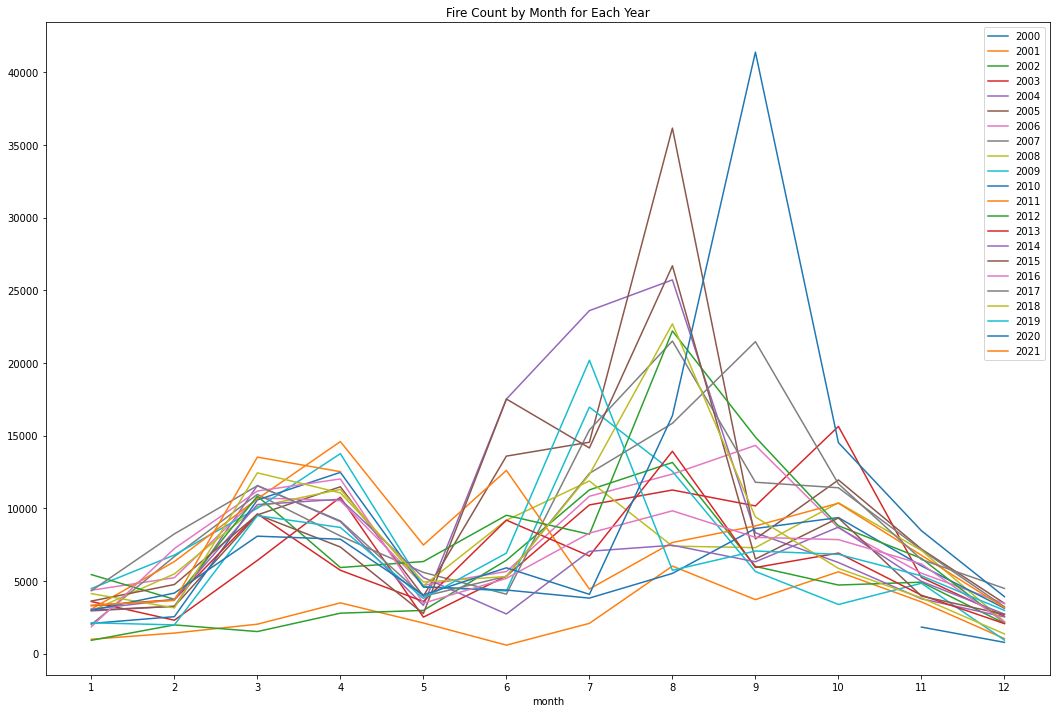
1. **Fire Frequency Change Analysis**
2. **Fire Frequency by Year Across the US**

First, we aggregated the data by “year”, get the counts of fires for each of the years, and produced a graph. As we don’t have the data for certain months in Year 2000 and 2021, we exclude these two years from the graph. We can see from the graph that fire frequencies fluctuated over the years and Year 2020 was one of the peak years.



1. **Fire Frequency by Year for Each Month Across the US**

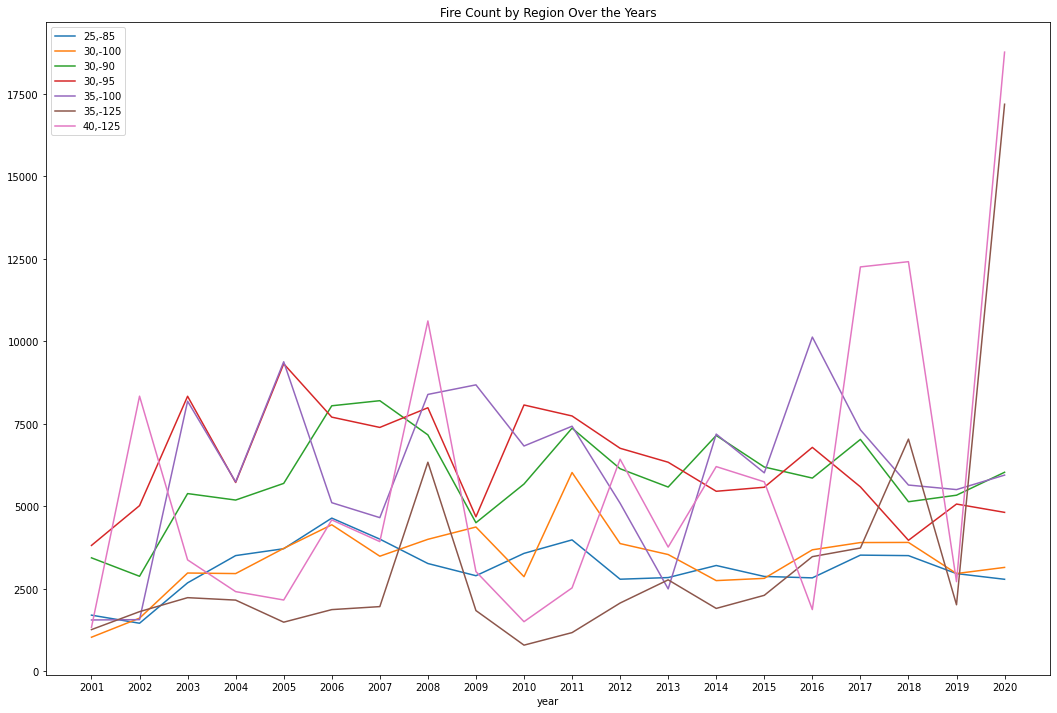
After we had an overview of the fire frequencies over the years, we then looked into more details; we grouped the data by “month” and pivoted column “year”, and produced a graph for monthly fire frequency over the . From the graphs produced, we see a peak of fire frequency in September 2020.



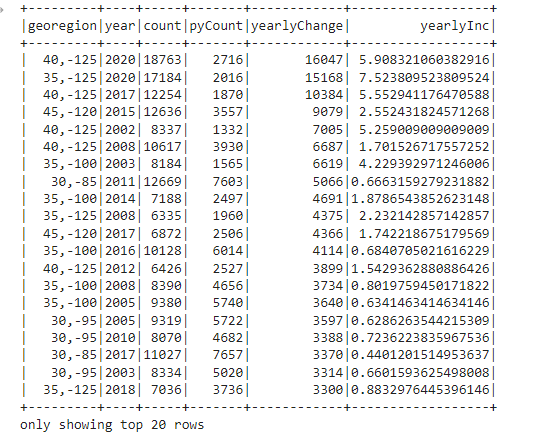
1. **Regional Fire Frequency Over the Years**

The analysis above were based on fire data across the US. We then moved on to analyses on a regional basis. We define a region with a granularity of 5 degrees, and added one column named “georegion”. For example, a “georegion” valued “40,-125” represents a region with latitude from 40 to 45 and longitude from -125 to -120.

We calculated the average early fire counts for each region; extracted data for 10 regions with highest average yearly fire counts and plotted the fire frequencies over the years for certain regions. While we see fluctuations in fire frequencies over the years for each region, year 2020 was an abnormal year for region “35, -125” and “45,-120”; there were abnormal increase in fire frequencies in these two regions. With a bit more effort to find these two regions on a map, we can see that these two regions are in Northern California Area.

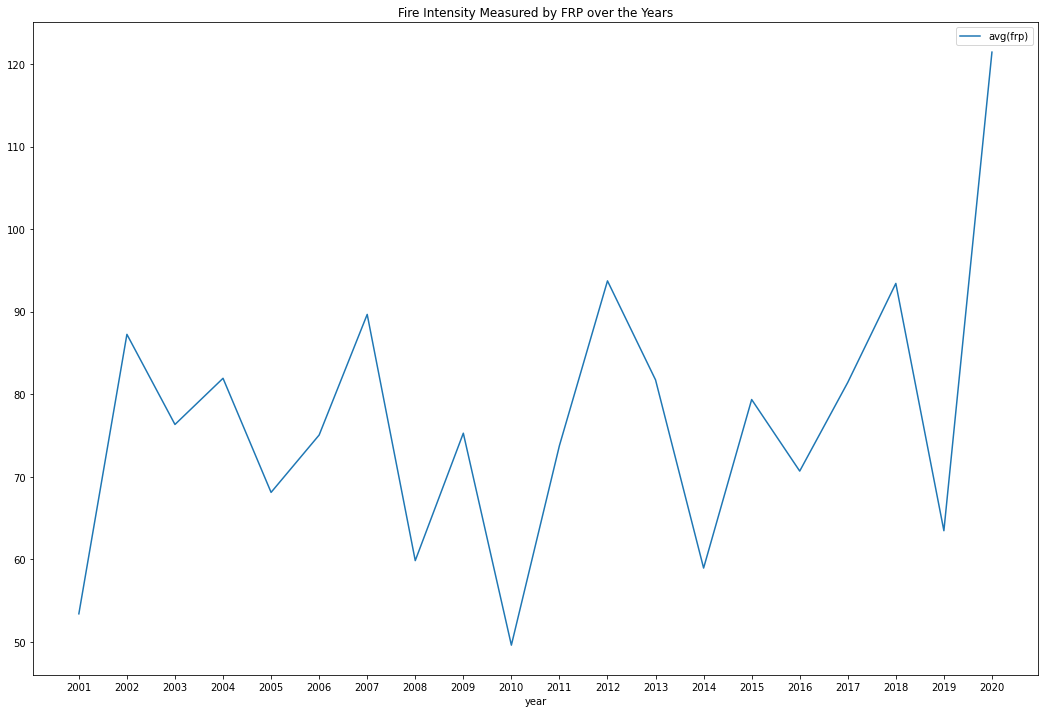


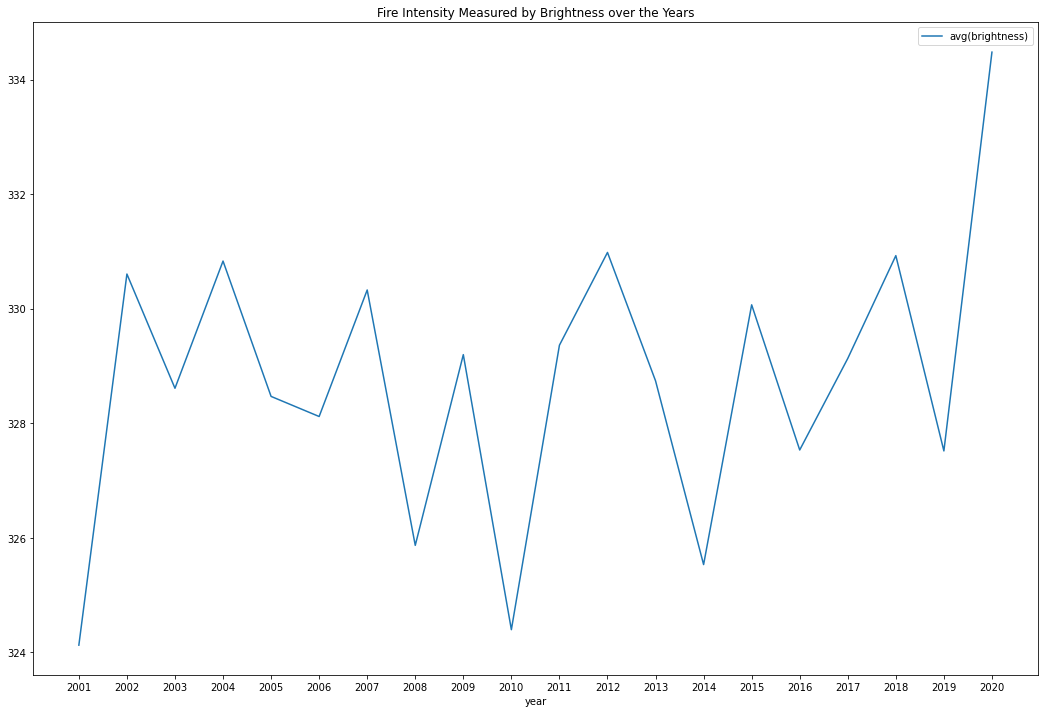
We also used window functions to calculate year-over-year increase in fire frequency for each region and got the following result. We can find from the result that region “35, -125” and “45,-120” saw the greatest yearly changes in fire frequency from 2019 to 2020.



1. **Fire Intensity Change Analysis**
2. **Fire Intensity by Year Across the US**

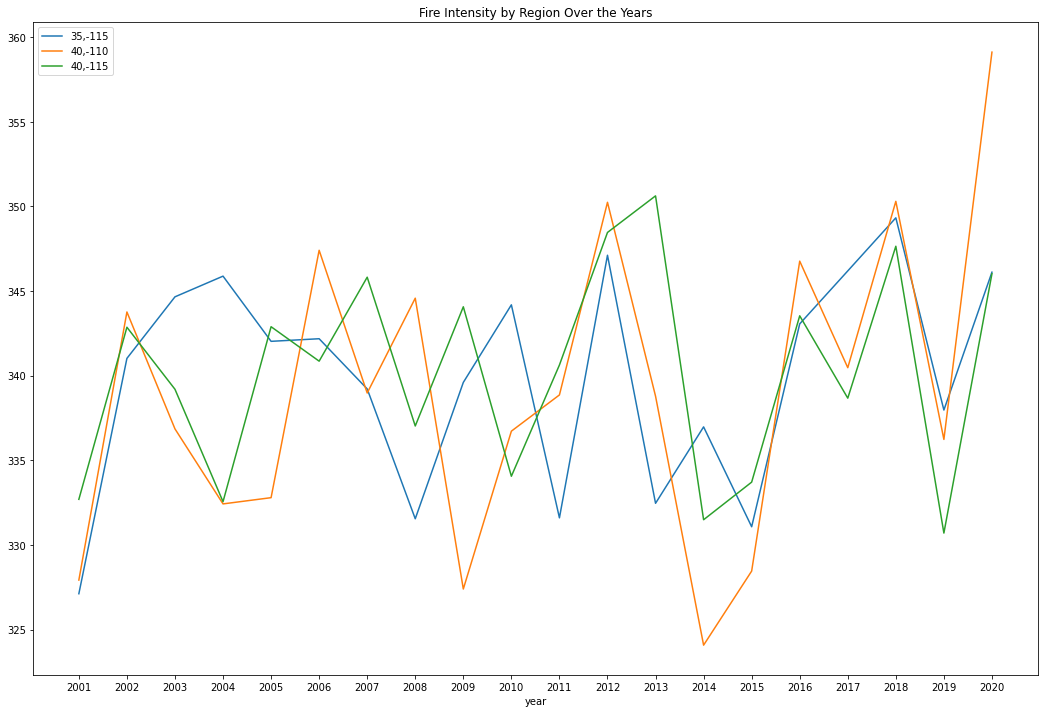
For fire intensity, we used similar techniques. First, we used two different fields “frp” and “brightness” to measure fire intensity, and analyzed how average fire intensity has changed over the years. The two measurements produced consistent results as shown in the two graphs below. We see that the average fire intensity fluctuated over the years and the peaks over the years stayed around 90 for “frp” and 330 for “brightness”; However, in 2020, we see new records of average fire intensity for both of the two measurements.

****

****

1. **Regional Fire Intensity Over the Years**

We also continued our analysis on fire intensity to a regional basis and used the same definition of “region” as we did for fire frequency analysis. We used “brightness” as the measurement of fire intensity. We calculated average fire intensity by region, extracted data for regions with the highest average fire intensity and plotted the fire intensity over the years for certain regions. While we see fluctuations in fire frequencies over the years for each region, year 2020 was an abnormal year for region “40, -110”; there was abnormal increase in average fire intensity in this region. With a bit more effort to find this region on a map, we can see that in the mid west area of the US.



1. **Conclusions**

Our overall findings from analyses on this data set is that Year 2020 was an abnormal year for the US in terms of fire frequency and average fire intensity especially for Northern California Area and the Midwest area.