#2 Group Project Report

IMAAN AGHA 1982051
RADHIKA SONDE 1964766
STEPHANY LOPEZ 2142613
ROBERTO JACKSON BAEZA 1861128

Summarization, Mapping, Hotspot Discovery and Change Analysis of High-Intensity Solar Flare Events

1 TASK 1

1.1 Method 1 - Flare intensity estimation based on the total.counts attribute

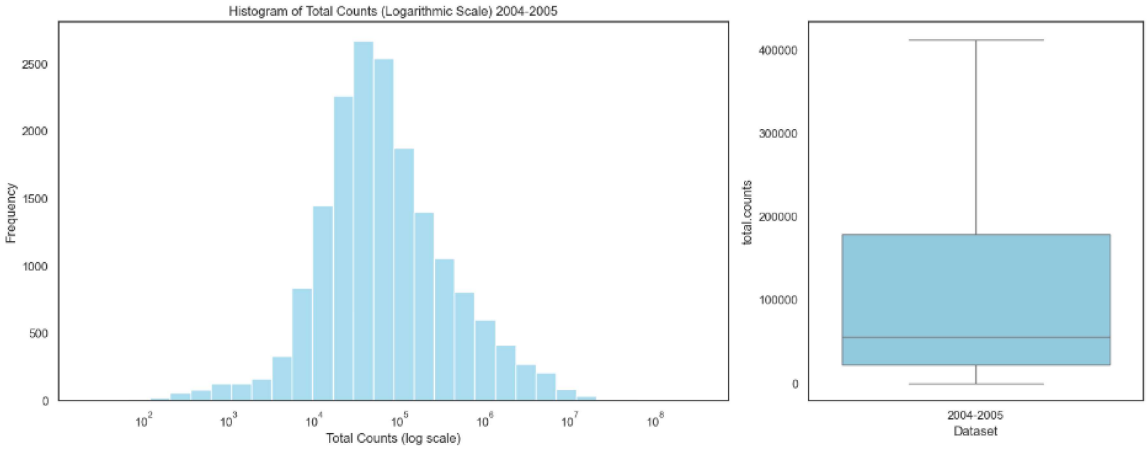


Fig. 1. Method 1 histogram and boxplot depicting total.counts

For Method 1, we struggled to visualize the total.counts data with so much noise from outliers. For our histogram, we found the best method to best understand the distribution of data was to use a logarithmic scale. When reading the histogram, however, it is important to realize that the increments on the x-axis are not uniform and increase exponentially. The Box plot was created by removing the outliers. We can clearly see the range of data without considering outliers. From these visualizations, we can tell that highest frequency of values is when the total.counts are in the range between $10^4 and 10^5$.

1.2 Method 2 - Flare intensity estimation based on the duration.s and energy.kev attributes

For Method 2, we created a bar graph to identify the number of occurrence of each energy band group. With the boxplot and the density plot, we were able to hone in on the relationship between energy bands and duration.

In 2004-2005, it seems that '6-12' energy bands have the highest frequency and the '300-800' energy bands have the lowest indicating that low energy flares occur more frequently than high energy flares. Conversely '300-800' energy bands have the longest duration while '6-12' have the shortest hinting at a relationship between both variables that as the energy levels increase, so does the duration of the flare.

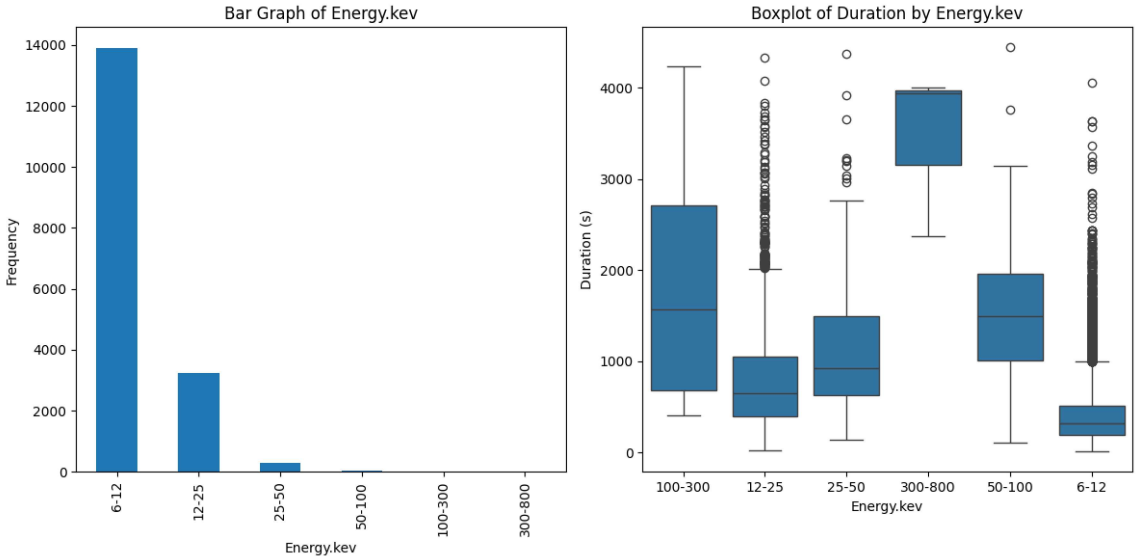


Fig. 2. Method 2 bar graph depicting the energy.kev frequency and boxplot illustrating the relationship between duration.s and energy.kev

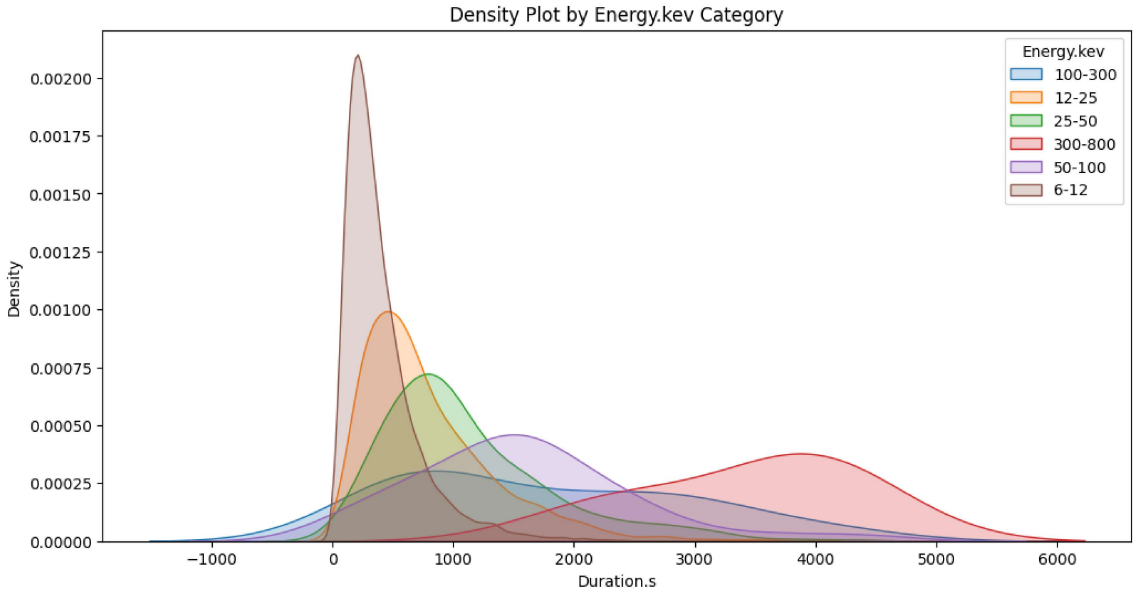


Fig. 3. Method 2 density depicting the relationship between duration and energy

1.3 Method 1 - Intensity heatmaps based on total.counts attribute

For Method 1, we generated two intensity heatmaps corresponding to months 1+2+3+4 and 21+22+23+24. In these heatmaps, the color of each pixel represents the intensity of solar flares at that position. The intensity is directly proportional to the values recorded in the 'total.counts' attribute.

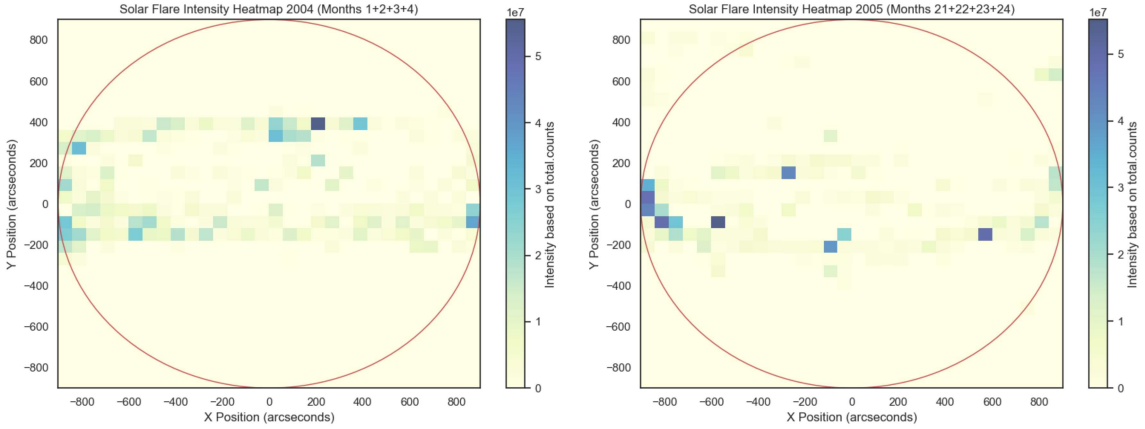


Fig. 4. Method 1 intensity heatmap based on total.counts

1.4 Method 2 - Intensity heatmaps based on duration.s and energy.kev attributes

For Method 2, the intensity estimation is computed using the formula:

$$\text{Intensity} = \text{duration.s} \times \frac{\text{energy.kev.f} + \text{energy.kev.i}}{2}$$

where the duration.s attribute represents the duration of the solar flare, and the energy.kev.i and energy.kev.f attributes contain the initial and final values of the energy range of the solar flare, respectively. The average energy of the solar flare is then multiplied by the duration to obtain the corresponding intensity. A logarithmic color scale had to be applied (LogNorm) to enhance the visualization of intensity difference. This was done because the intensity values mostly spanned in the low-intensity regions. The minimum and maximum values for the logarithmic scale (vmin and vmax) are calculated based on the minimum and maximum values of the intensity computed from the entire dataset. By incorporating a logarithmic scale, the resulting heatmaps effectively display the intensity variations.

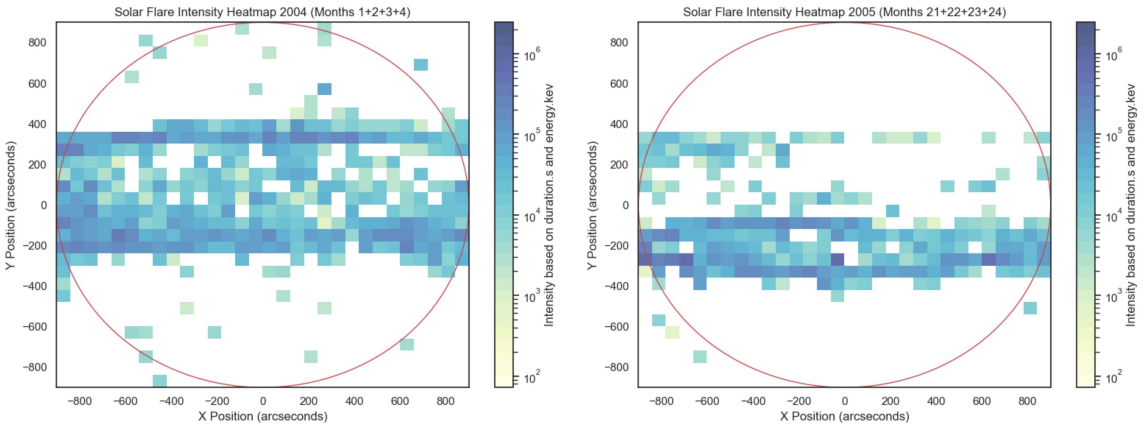


Fig. 5. Method 2 intensity heatmap based on duration.s and energy.kev

1.5 Intensity Map Comparison

Upon comparing the intensity maps based on total counts with those derived from duration and energy, we observe an overall higher solar flare intensity estimation for months 1+2+3+4 than for months 21+22+23+24. This suggests a correlation between the total counts, duration, and energy attributes. Additionally, we note that the intensity estimation of the solar flares is concentrated within position ranges of -400 to 400 on the Y-axis.

2 TASK 2

2.1 Hotspot Discovery Algorithm & Thresholds

```
from scipy.stats import gaussian_kde
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import matplotlib.image as mpimg

# Read data file and set d1
df = pd.read_csv('Solar_flare_RHESSI_2004_05.csv')

# Create kernel density estimation
data = np.vstack([df['x.pos.asec'], df['y.pos.asec']])
kde = gaussian_kde(data)

# Create a gridspace and evaluate KDE
xgrid = np.linspace(-900, 900, 40)
ygrid = np.linspace(-900, 900, 40)
Xgrid, Ygrid = np.meshgrid(xgrid, ygrid)

grid_points = np.vstack([Xgrid.ravel(), Ygrid.ravel()])
Z = kde.evaluate(grid_points).reshape(Xgrid.shape)
thresholdZData = kde.evaluate(np.vstack([Xgrid.ravel(), Ygrid.ravel()]))
d1 = np.percentile(thresholdZData, 75)  # Change this threshold value as needed
d2 = np.percentile(thresholdZData, 95)  # Change this threshold value as needed
```

We developed the hotspot algorithm utilizing kernel density estimation to analyze the spatial distribution and precisely locate the hotspots on the sun using a 2D gridspace. After creating the regions, we filtered the data based on the selected threshold values. Subsequently, we generated a KDE plot and used the "coolwarm" color scheme to effectively pinpoint and enhance the visibility of these hotspots.

To determine our threshold values, d1 and d2, we relied on the 75th and 95th percentiles of the density points. We opted for these percentiles because we recognize that hotspots consist of elevated concentrations of data points, surpassing the mean or 50th percentile. By selecting the 75th percentile, we effectively isolated the top 25% of our data, defining larger, more regionally significant hotspots with intensities exceeding a "medium-high" level. In contrast, the 95th percentile pinpointed the top 5% of our data, accurately representing smaller, highly concentrated hotspots with densities surpassing a "high" intensity threshold.

2.2 Hotspot Visualization

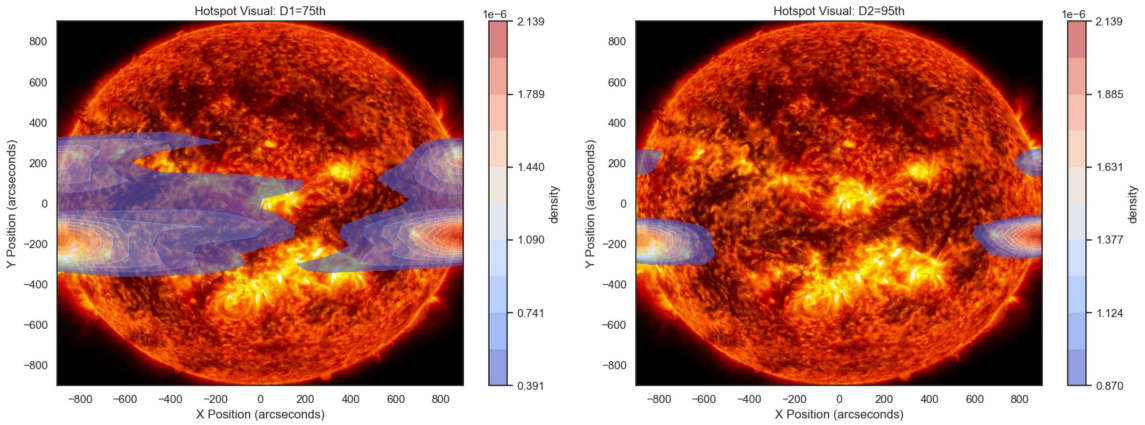
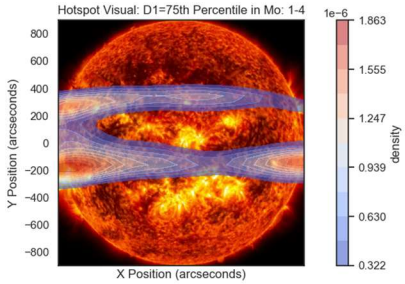
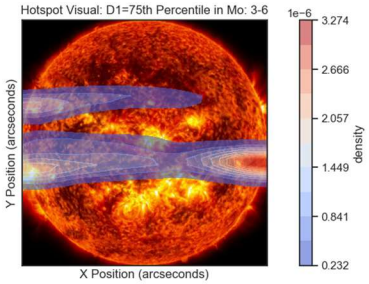
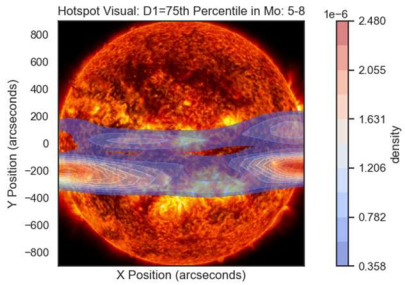


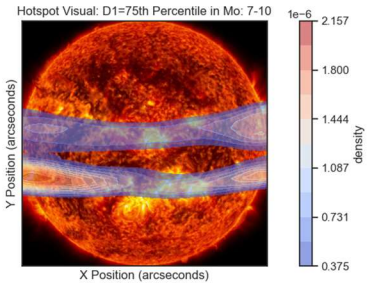
Fig. 6. Hotspot Visualization of the d1 and d2 threshold for all the months

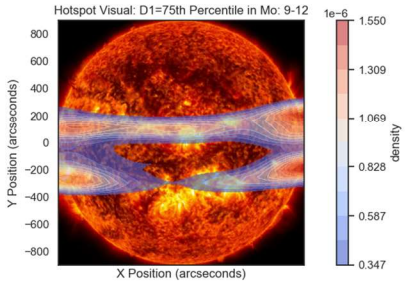
2.3 Times Series Hotspot

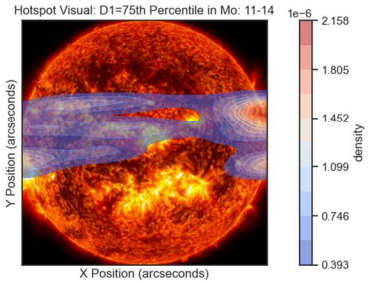

Hotspot Visual: D1=75th Percentile in Mo: 1-4
Y Position (arcseconds)
X Position (arcseconds)
density
1e−6
1.863
1.555
1.247
0.939
0.630
0.322

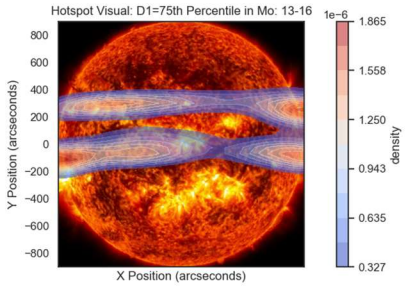

Hotspot Visual: D1=75th Percentile in Mo: 3-6
Y Position (arcseconds)
X Position (arcseconds)
density
1e−6
3.274
2.666
2.057
1.449
0.841
0.232

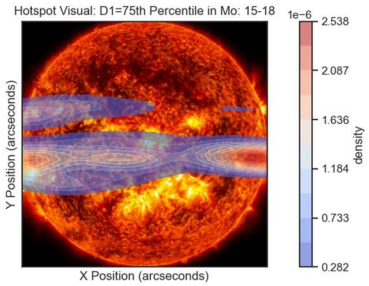

Hotspot Visual: D1=75th Percentile in Mo: 5-8
Y Position (arcseconds)
X Position (arcseconds)
density
1e−6
2.480
2.055
1.631
1.206
0.782
0.358

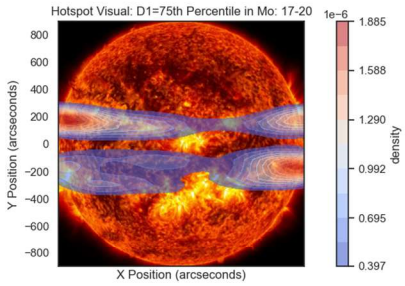

Hotspot Visual: D1=75th Percentile in Mo: 7-10
Y Position (arcseconds)
X Position (arcseconds)
density
1e−6
2.157
1.800
1.444
1.087
0.731
0.375

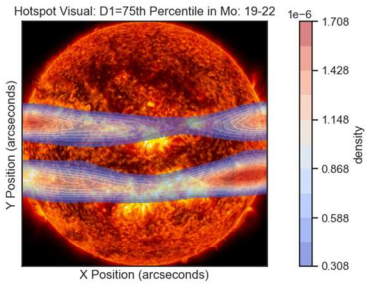

Hotspot Visual: D1=75th Percentile in Mo: 9-12
Y Position (arcseconds)
X Position (arcseconds)
density
1e−6
1.550
1.309
1.069
0.828
0.587
0.347

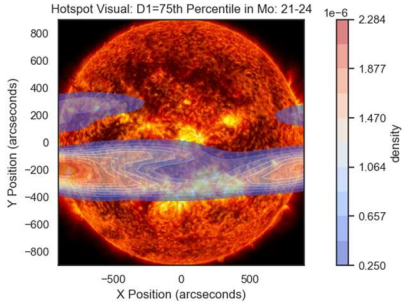

Hotspot Visual: D1=75th Percentile in Mo: 11-14
Y Position (arcseconds)
X Position (arcseconds)
density
1e−6
2.158
1.805
1.452
1.099
0.746
0.393

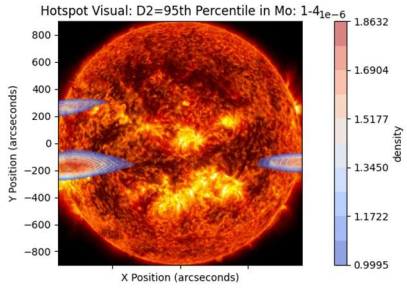

Hotspot Visual: D1=75th Percentile in Mo: 13-16
Y Position (arcseconds)
X Position (arcseconds)
density
1e−6
1.865
1.558
1.250
0.943
0.635
0.327

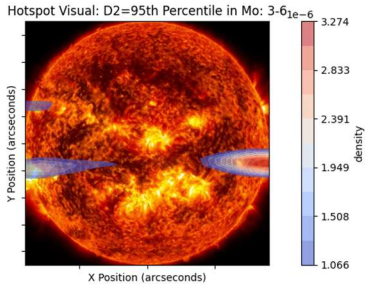

Hotspot Visual: D1=75th Percentile in Mo: 15-18
Y Position (arcseconds)
X Position (arcseconds)
density
1e−6
2.538
2.087
1.636
1.184
0.733
0.282

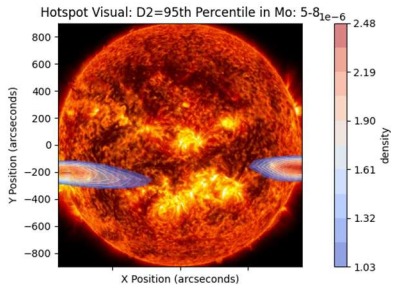

Hotspot Visual: D1=75th Percentile in Mo: 17-20
Y Position (arcseconds)
X Position (arcseconds)
density
1e−6
1.885
1.588
1.290
0.992
0.695
0.397

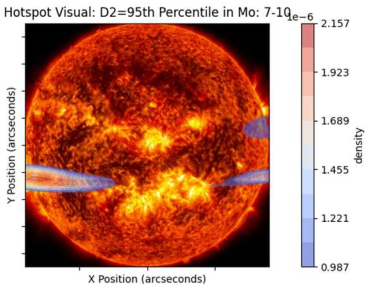

Hotspot Visual: D1=75th Percentile in Mo: 19-22
Y Position (arcseconds)
X Position (arcseconds)
density
1e−6
1.708
1.428
1.148
0.868
0.588
0.308

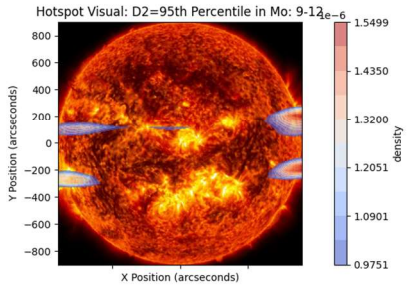

Hotspot Visual: D1=75th Percentile in Mo: 21-24
Y Position (arcseconds)
X Position (arcseconds)
density
1e−6
2.284
1.877
1.470
1.064
0.657
0.250

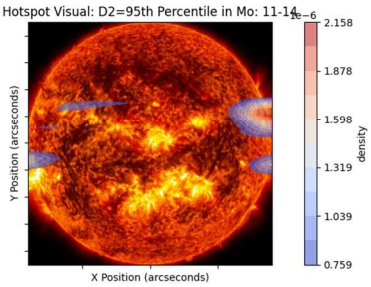

Hotspot Visual: D2=95th Percentile in Mo: 1-4
1e−6
1.8632
1.6904
1.5177
1.3450
1.1722
0.9995
density
Y Position (arcseconds)
X Position (arcseconds)

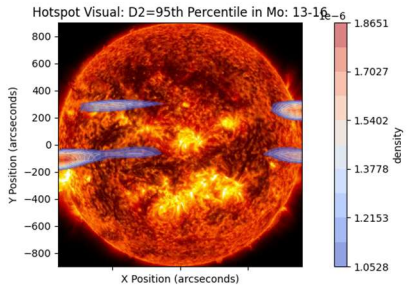

Hotspot Visual: D2=95th Percentile in Mo: 3-6
1e−6
3.274
2.833
2.391
1.949
1.508
1.066
density
Y Position (arcseconds)
X Position (arcseconds)

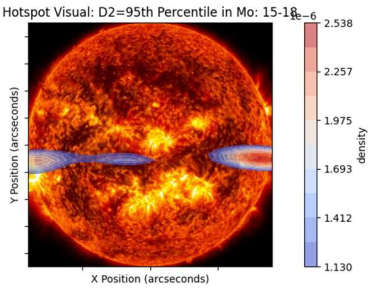

Hotspot Visual: D2=95th Percentile in Mo: 5-8
1e−6
2.48
2.19
1.90
1.61
1.32
1.03
density
Y Position (arcseconds)
X Position (arcseconds)

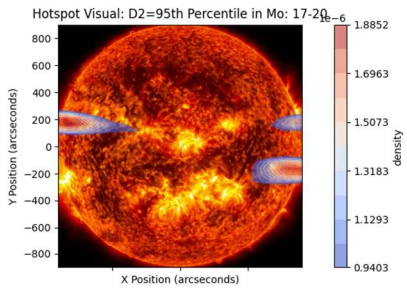

Hotspot Visual: D2=95th Percentile in Mo: 7-10
1e−6
2.157
1.923
1.689
1.455
1.221
0.987
density
Y Position (arcseconds)
X Position (arcseconds)

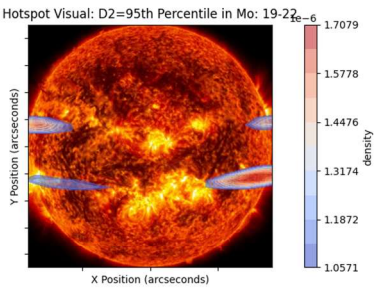

Hotspot Visual: D2=95th Percentile in Mo: 9-12
1e−6
1.5499
1.4350
1.3200
1.2051
1.0901
0.9751
density
Y Position (arcseconds)
X Position (arcseconds)

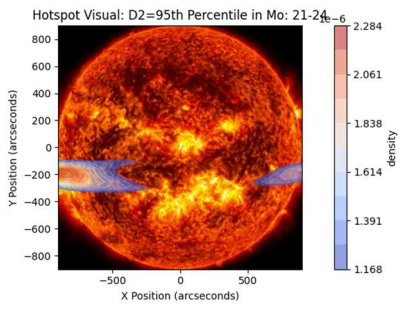

Hotspot Visual: D2=95th Percentile in Mo: 11-14
1e−6
2.158
1.878
1.598
1.319
1.039
0.759
density
Y Position (arcseconds)
X Position (arcseconds)


Hotspot Visual: D2=95th Percentile in Mo: 13-16
1e−6
1.8651
1.7027
1.5402
1.3778
1.2153
1.0528
density
Y Position (arcseconds)
X Position (arcseconds)


Hotspot Visual: D2=95th Percentile in Mo: 15-18
1e−6
2.538
2.257
1.975
1.693
1.412
1.130
density
Y Position (arcseconds)
X Position (arcseconds)


Hotspot Visual: D2=95th Percentile in Mo: 17-20
1e−6
1.8852
1.6963
1.5073
1.3183
1.1293
0.9403
density
Y Position (arcseconds)
X Position (arcseconds)


Hotspot Visual: D2=95th Percentile in Mo: 19-22
1e−6
1.7079
1.5778
1.4476
1.3174
1.1872
1.0571
density
Y Position (arcseconds)
X Position (arcseconds)


Hotspot Visual: D2=95th Percentile in Mo: 21-24
1e−6
2.284
2.061
1.838
1.614
1.391
1.168
density
Y Position (arcseconds)
−500
0
500
X Position (arcseconds)

To generate our time series, we employed our hotspot algorithm and filtered the data based on both the threshold value and batches of months.

In our time series for medium-high level, large hotspots, using a threshold value corresponding to the 75th percentile of the density points, we observe predominant hotspot activity both above and below the Sun's equator. Specifically, in the y-direction, activity spanned from -400 to 200, with a notable gap around the equator (-50 to 50) where no hotspots occurred. These hotspots align parallel to the x-axis, forming two distinct strips. Over time, while medium-high spots are consistently visible between 0 and -400 in the y direction, the activity above fluctuated. For example, in the initial batches, we observe low activity in the positive y-direction, which increases notably in the 9-12, 13-16, and 19-22 month batches.

In the time series for smaller, highly intense hotspots, identified with a threshold value corresponding to the 95th percentile of the density points we observed that these hotspots primarily occurred at the far ends in the x-direction (<-600 and >600). Over time, similar to the medium-high hotspots, these also manifested between -400 to -50 and 50 to 400 in the y-direction and in the southern region of the equator, with an exception in the 17-20 month batch where the hottest hotspot is found in the mid-northern region of the Sun.

Based on this analysis, we can conclude that hotspots on the Sun predominantly occur in the mid-equator region as opposed to the poles, and for the hottest points, they additionally occur at the far corners of the Sun.

3 TASK 3

3.1 Change Analysis for Solar Flares Counts

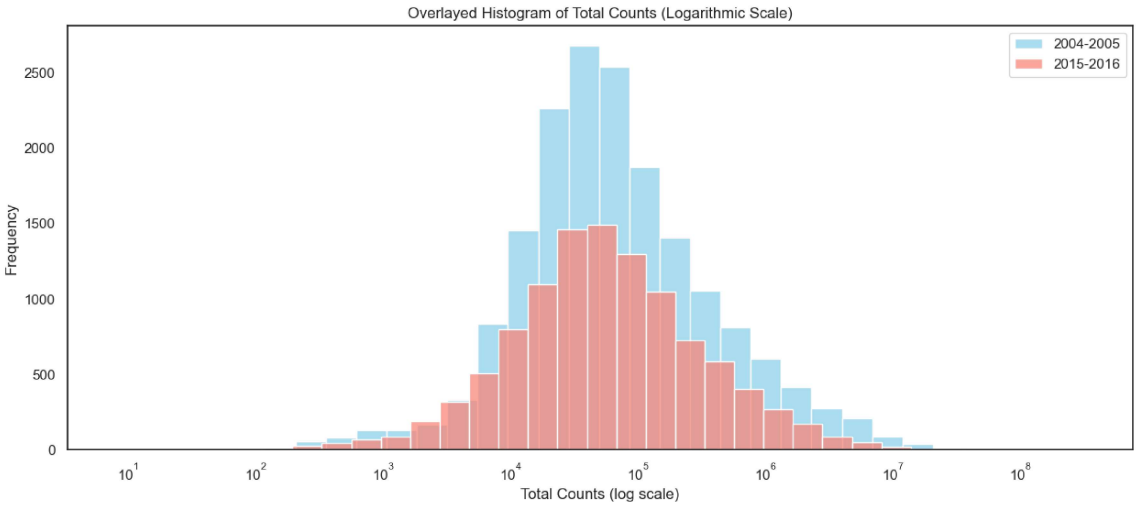


Fig. 7. Histogram comparison of total.counts between 2004-2005 & 2015-2016 dataset

It was hard to visualize the total.counts attribute due to the large number of outliers in the datasets. The total.counts attribute refers to the total number of counts in energy range 6-12keV over duration of a solar flare. The counts are corrected and summed over the entire duration of the flare event and including background noise. Using a logarithmic

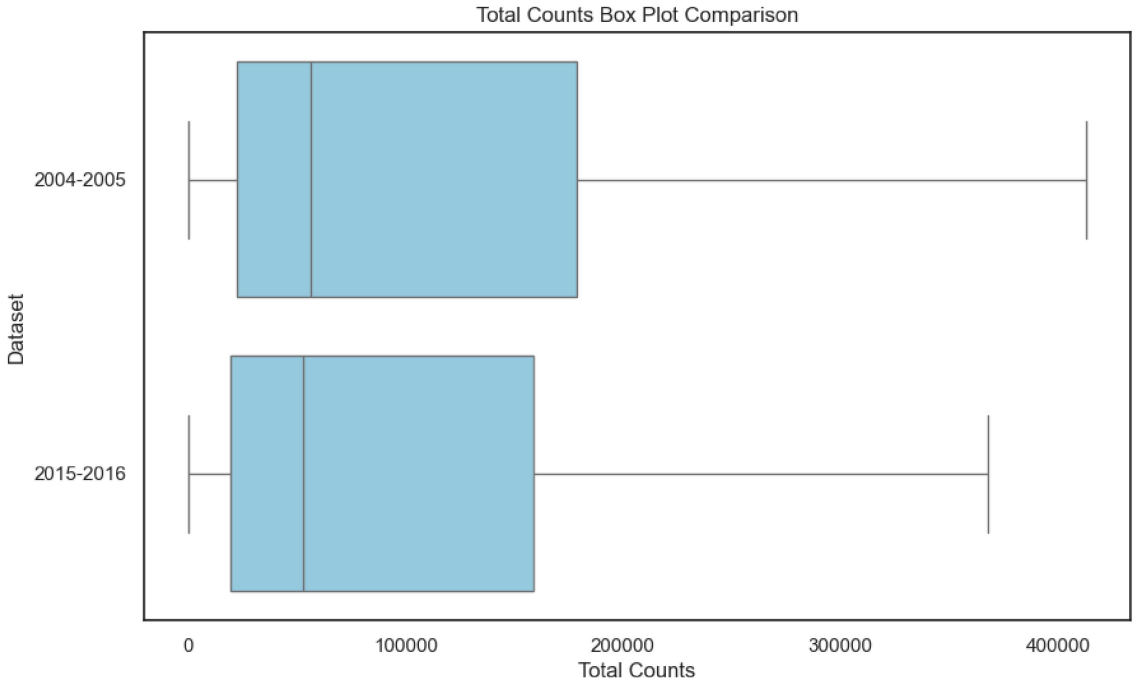


Fig. 8. Boxplot comparison of total.counts between 2004-2005 & 2015-2016 dataset

scale we were able to use a histogram to visualize the data. Using this technique helps to better understand the distribution of data when there are extreme or a wide range of values. The width of the bins increases exponentially and the distance between ticks are not equal intervals but rather ratios.

While the distribution of both histograms are similar the graph from 2004-05 has the higher frequency of events (over 2500). The histogram of 2015-16 has its peak frequency at just above 1400.

Plotting a boxplot was also a challenge with so many outliers, to make it easier to visualize the total.counts data we omitted the outliers to have reduced variability. Both boxplots have similar interquartile ranges but the 2004-05 plot has the bigger range of values that are not outliers.

3.2 Change Analysis for Solar Flares Duration and Energy Levels

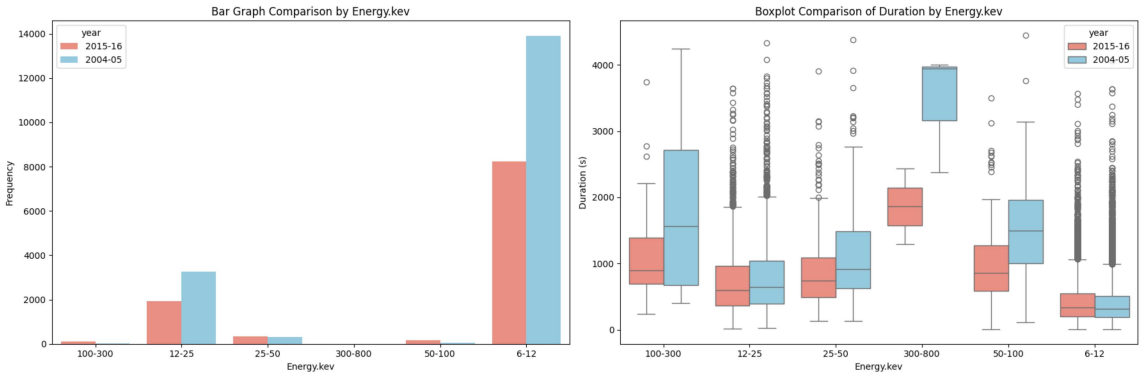


Fig. 9. Bar graph & Boxplot comparison of energy.kev and duration.s between 2004-2005 & 2015-2016 dataset

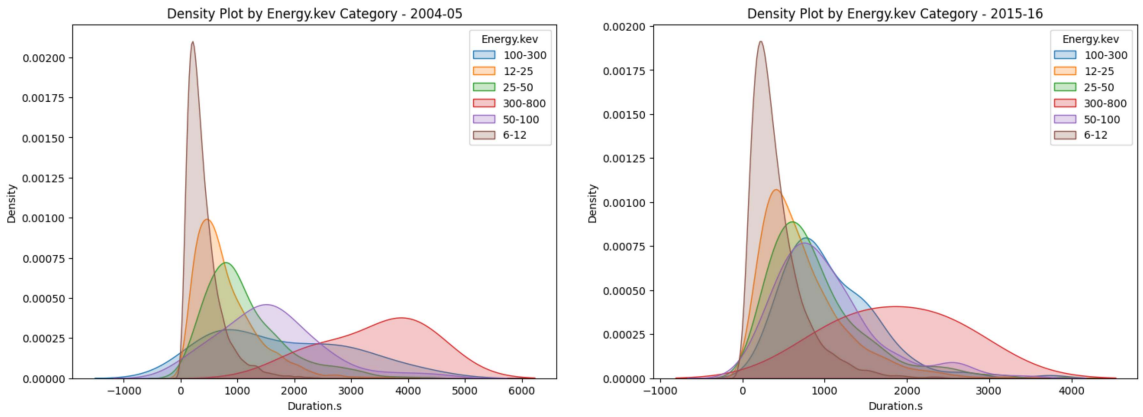


Fig. 10. Density plot comparison of energy.kev and duration.s between 2004-2005 & 2015-2016 dataset

Like the 2004-2005 dataset, the '6-12' energy bands have the highest frequency while the '300-800' energy bands have the lowest solidifying the observation that flares with high energy occur less frequently. Additionally, in the 2015-2016 period, the peaks of each energy category seem to be in order as the duration increases also bolstering the relationship between high energy levels having a longer duration than low ones.

From 2004-2005 to 2015-2016, there seems to be a major change in the energy levels and duration although the initial observations stay the same. The frequency of the energy bands decreased with the '6-12' band experiencing the greatest decline. There also seems to be a trend of decrease between the energy bands and the duration with '300-800', the highest energy level, experiencing the sharpest. Therefore, from 2004-2005 to 2015-2016, we can conclude the activity of the solar flares have decreased.

3.3 Change Analysis for Solar Flares Intensity

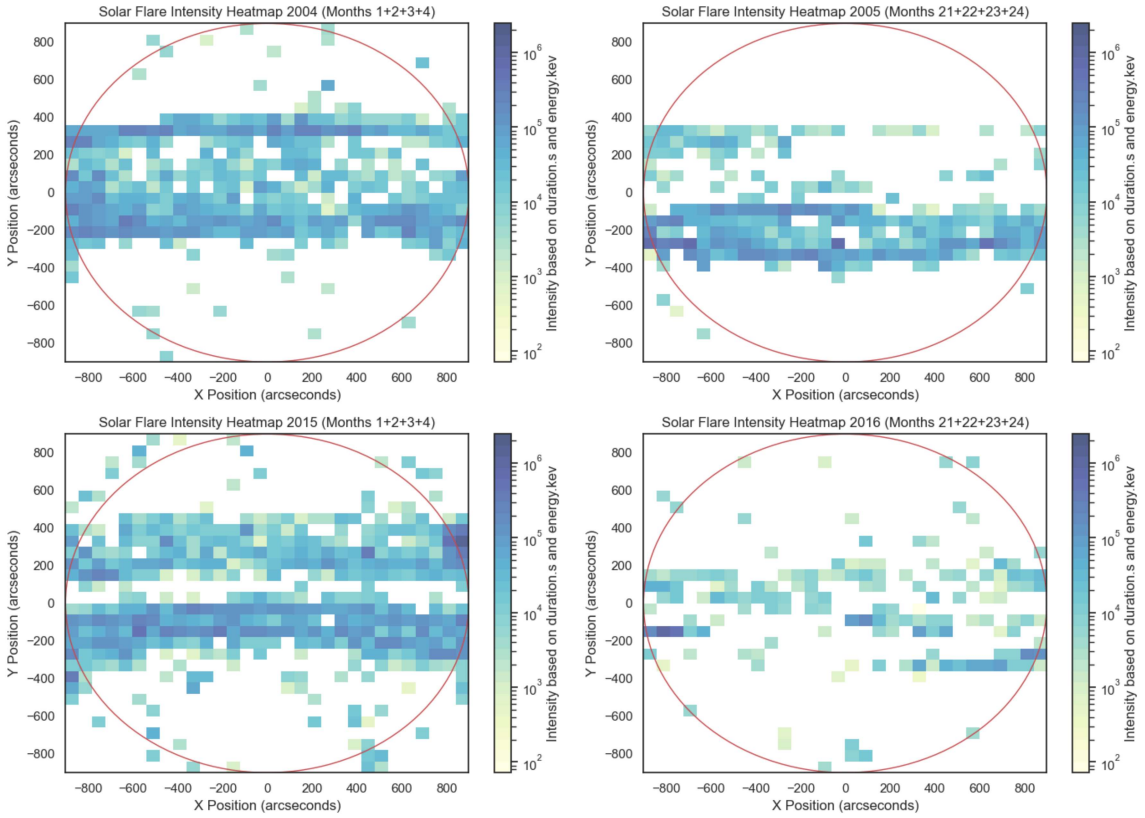


Fig. 11. Intensity heatmap comparison between 2004-2005 & 2015-2016 dataset

The solar flare intensity heatmaps from both datasets reveal higher intensities during the initial four months of 2004 and 2015 compared to the concluding four months of 2005 and 2016. Additionally, the distribution of solar flares is concentrated within the -400 to 400 range across all four plots. However, in the intensity maps for months 1+2+3+4 for 2004 and 2015, there are additional positions scattered outside the previously mentioned concentration range. The same phenomenon occurs for months 21+22+23+24 for the years 2005 and 2016 which could be an indicative of specific spatial patterns or anomalies in solar flare activity during those periods.

4 CONTRIBUTIONS

- **Imaan Agha** = Method 1 graphs for Task 1 & 3, Report
- **Roberto Jackson Baeza** = Method 1 intensity heatmaps, Times Series hotspots visualization & algorithm
- **Stephany Lopez** = Method 2 intensity heatmaps, intensity heatmaps comparison writeup, formatting time series graphs, Report
- **Radhika Sonde** = Method 2 graphs for Task 1 & 3, explanation summarization and interpretation of Task 2, Report

5 REFERENCES

- Intro to Spatial Analysis and HotSpot Discovery PowerPoint
- R Tutorial: Hotspot Analysis using Getis Ord Gi: https://rpubs.com/heatherleeleary/hotspot_getisOrd_tut
- COSC 3337 Group Project Introduction PowerPoint
- ChatGPT
- Exploratory Spatial Data Analysis (ESDA): https://darribas.org/gds_scipy16/ipynb_md/04_esda.html
- Exploratory Spatial and Temporal Data Analysis (ESTDA): https://darribas.org/gds_scipy16/ipynb_md/05_spatial_dynamics.html