Samrawit Basazinew
Eugenia Chien
Drew Garza
Ben Rosensweig

UFO Sightings

Introduction

Our project delves into the intriguing world of Unidentified Flying Objects, or UFO sightings, a topic that has captured the imagination of people worldwide. Interest in analyzing and visualizing this data was driven by our curiosity about the trends and patterns in the sightings over the years. We wondered how often the sightings vary across different regions and at times, and if there are any noticeable spikes which may coincide with some significant events.

Inspiration for our project was drawn from several sources, most notably a 2016 UFO sightings dashboard created by Ken Flerlage. Ken's work stood out to us for its thoughtful design and use of color, which made complex data both accessible and visually appealing. We aimed to incorporate similar design principles into our project to create a dashboard that is not only informative but also engaging for our audience

Through our analysis, we hope to shed light on how often and where UFO sightings are happening, providing users with an easy tool of exploration into this mystery. Whether skeptic or believer, we believe you will find something interesting in our data.

Exploratory Data Analysis – Data Cleaning – SQLite creation

The dataset we chose was UFO Sightings from Kaggle. This was a CSV file of over 80,000 reports of UFO sightings in the last century from all over the world. This data was collected by the National UFO Reporting Center. The dataset from Kaggle contained two files, a complete CSV and a scrubbed CSV. The scrubbed CSV omitted records that had missing or erroneous locations and time fields. Our team decided to use the scrubbed file since it already took care of an essential part of the data cleaning process.

Utilizing the scrubbed file, we began our work of data cleaning and engineering which included trying to narrow the scope of the project from over 80.3k records to something hopefully more manageable. Through some exploratory data analysis, we figured out that most UFO sightings started occurring around the 1990's. Since the latest year included was 2014, we chose to focus on only 20 years of data from 1994 – 2014. To help narrow down the dataset even further, we decided to focus on only sightings from the United States.

For the shapes field, there were 29 unique values. Our intent was to try to use this field as one of the filters on the map. But we felt that 29 ways to filter on the map was too much. Some of these values (hexagon, delta, round, crescent, pyramid, flare, dome, changed, cross, cone) did not

constitute a sizable portion of the data. We dropped those values, confident that it would not impact or skew the data in any significant way. For the remaining shapes, we grouped them by similarity and created a parent category for them: Other, Angular, Rectangular, Circular, Luminous, Elliptical, Formation, and Uncategorized.

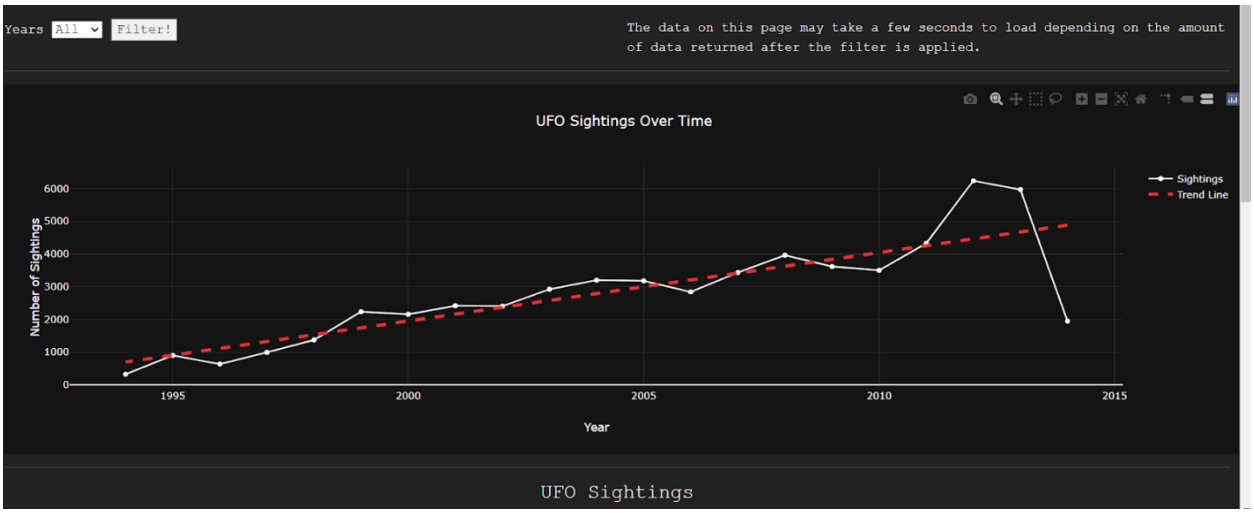
```
def categorize_shape(shape):
    if shape in ['changing', 'other', 'unknown']:
        return 'Other'
    elif shape in ['chevron', 'diamond', 'triangle']:
        return 'Angular'
    elif shape in ['cigar', 'cylinder', 'rectangle']:
        return 'Rectangular'
    elif shape in ['circle', 'disk', 'sphere']:
        return 'Circular'
    elif shape in ['fireball', 'flash', 'light']:
        return 'Luminous'
    elif shape in ['egg', 'oval', 'teardrop']:
        return 'Elliptical'
    elif shape == 'formation':
        return 'Formation'
    else:
        return 'Uncategorized'
# Apply category based on shape function
df['category'] = df['shape'].apply(categorize_shape)
df.head()
```

After our data cleaning and data engineering, we were left with approximately 58.6k records which we decided was the most data filtering and slicing we were willing to do without further skewing the completeness and accuracy of the data. At this point we were ready to send our curated data into an SQLite database.

Dashboard

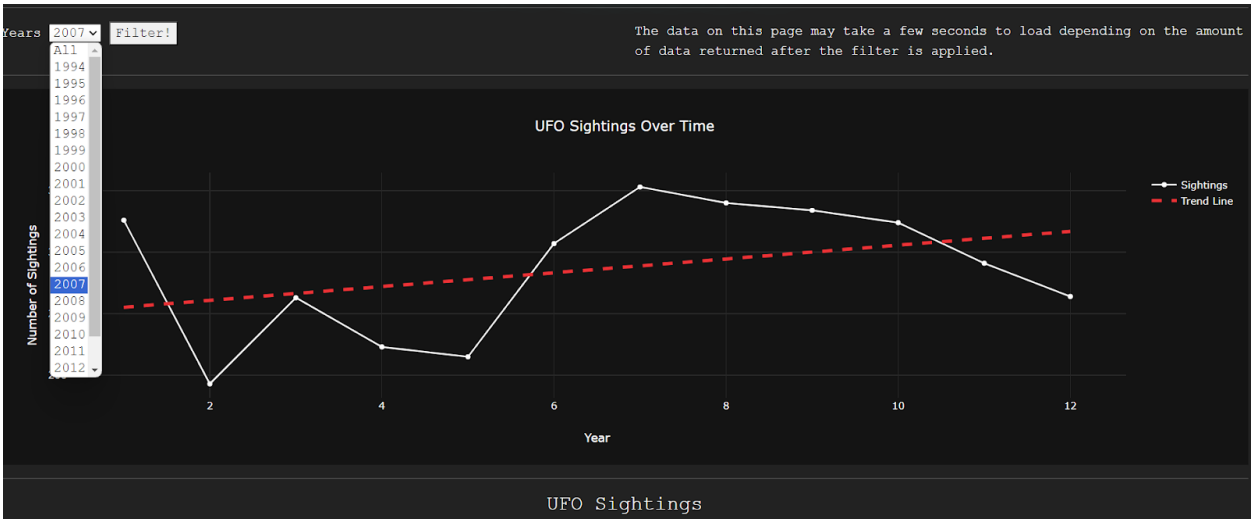
The original dataset we worked with included UFO sightings spanning the entire century. However, we decided it would be more relevant to focus on the most recent sightings and narrow our scope to just those reported in the United States. Our line graph specifically highlights sightings from 1994 to 2014, providing a clearer view of recent trends in UFO activity.

To create the line graph, we first calculated the total number of sightings for each year, allowing us to visualize the trend from 1994 to 2014. Additionally, we calculated the total number of sightings for each month, enabling us to examine the trends on a yearly basis. These calculations helped us to provide a clear and detailed visualization of UFO sighting patterns over time. It's important to mention that the dataset ends in May 2014. As a result, when users filter the data for the year 2014, the graph will only display sightings up until May, showing just the first half of the year.

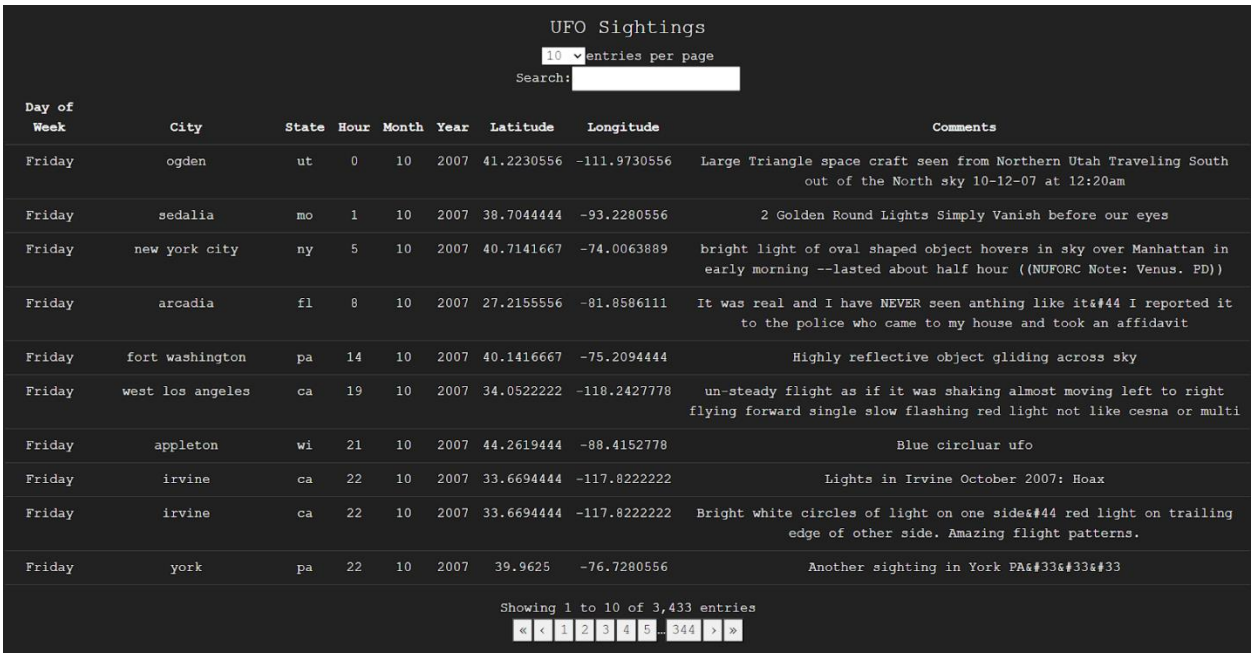


We wanted to give users the flexibility to explore the data by year, allowing them to observe trends over time. Our dashboard enables users to select a specific year and click on 'Filter' to view the number of sightings for that year. The sightings are displayed monthly, making it easy to identify patterns and fluctuations throughout the year.

One notable pattern we observed when examining the trendlines for each year is that UFO sightings tend to peak during the summer months, particularly in July, August, and September. This trend remains consistent across most years. Additionally, when analyzing data over the past two decades, we found that 2012 stands out with the highest number of sightings



We also wanted to include another visualization of UFO sightings that provides more detailed information. This visualization is a table that displays the day of the week, city, state, hour, month, year, latitude, longitude, and a 'comments' section. The table is a direct reflection of data that is plotted on the line chart which means it is responsive to the Year filter function as well. The comments section contains descriptions from people who reported the sightings, offering insights into what they observed.



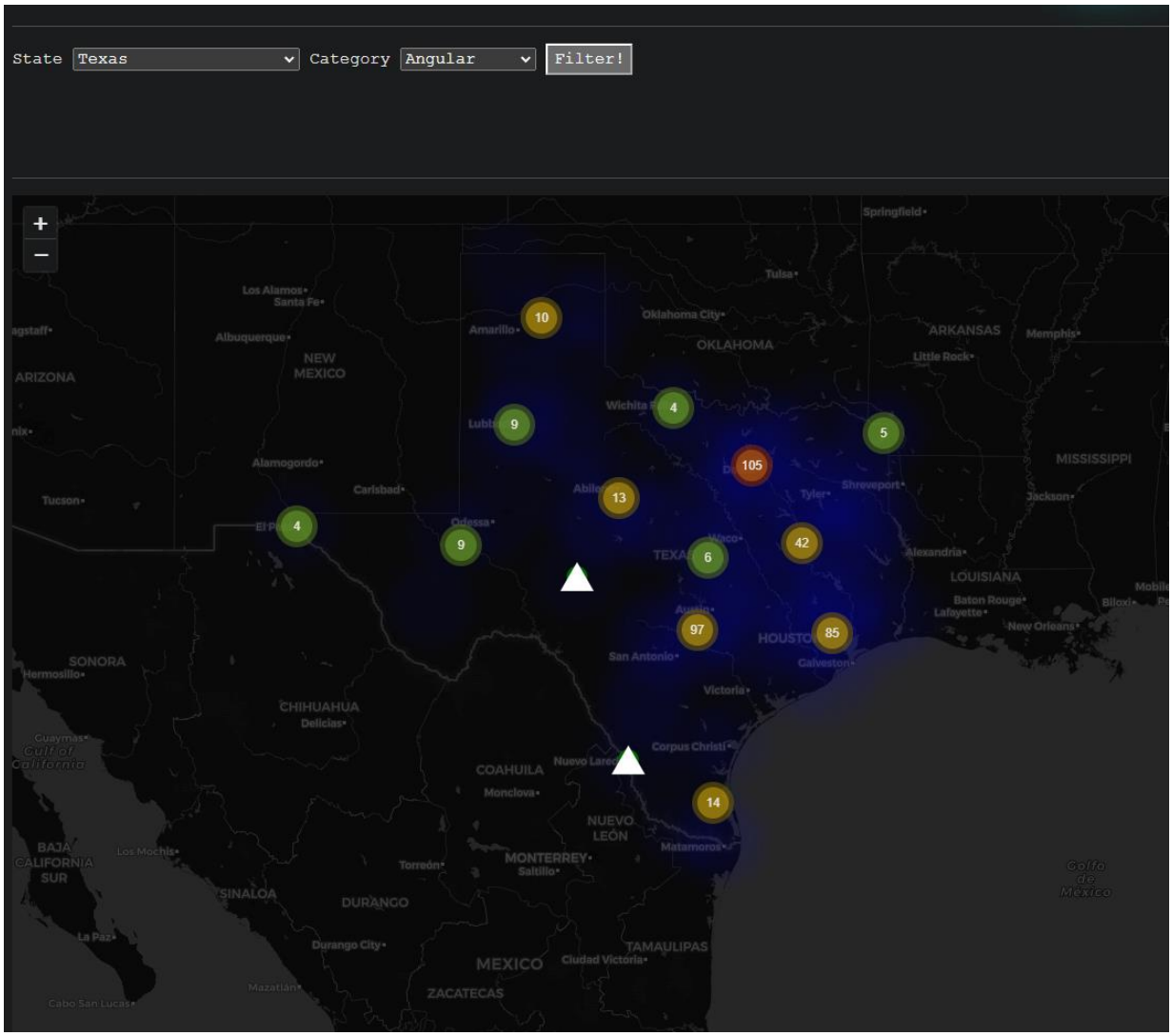
Our dashboard also includes a search feature, allowing users to search by city, state, day of the week, or latitude and longitude. This search capability filters the data to display only the sightings that match the user's criteria, making it easy to find specific information. Additionally, users can select the number of entries displayed per page, though we have set the default to 10 entries per page for optimal viewing. But users also have the option to display 25, 50, or 100 entries per page.

Map

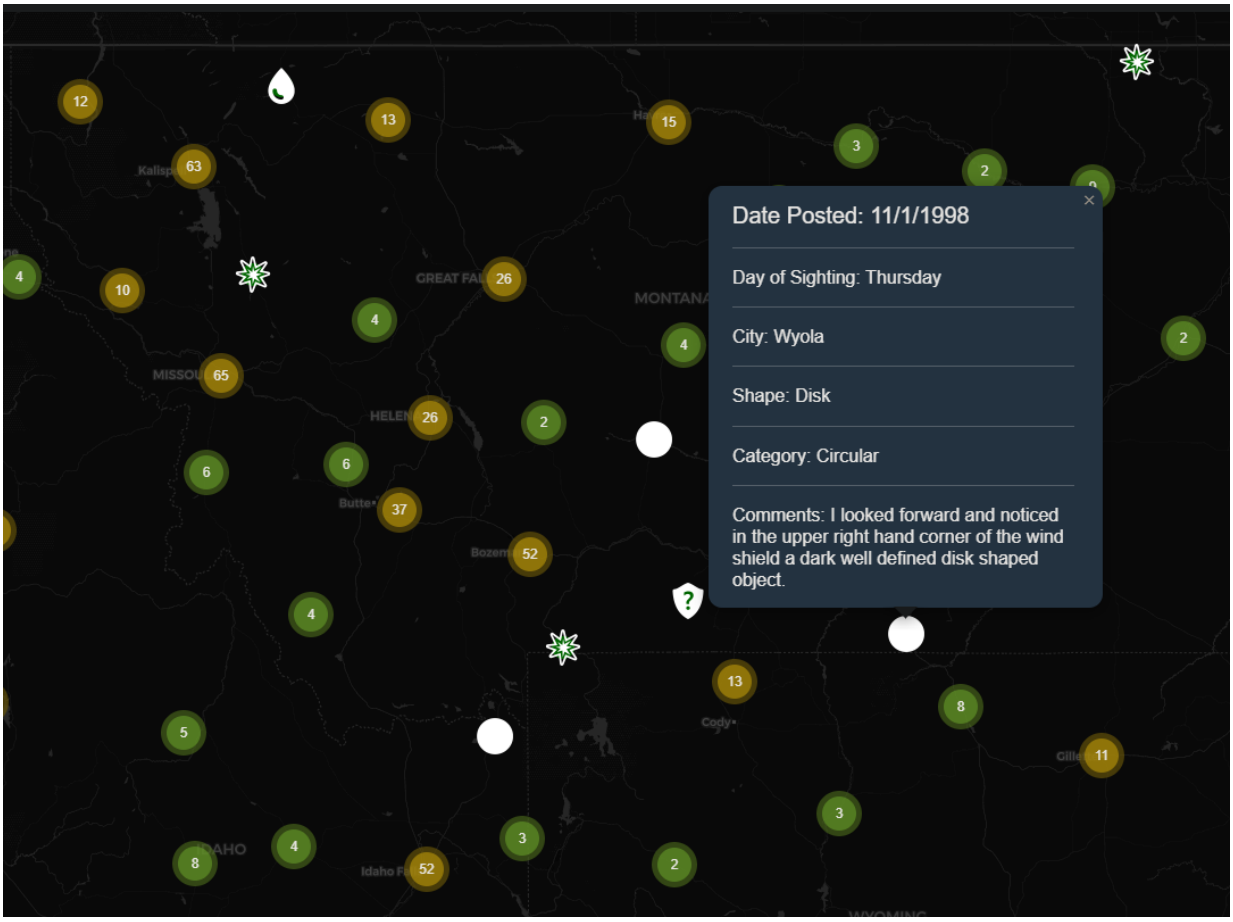
Our initial dataset was very large and included some international sightings, but we wanted to specifically focus on sightings within the USA. Additionally, with almost 20 different shapes across

all sightings, we decided to apply a custom category value to each sighting so that we could group multiple shapes together which narrowed us down to 7 unique category values.

With the location and category datapoints finalized, we applied a separate dropdown filter so that each could be used together to filter the same data. For example, we start with All sightings and can filter it to a specific State and/or a specific shape Category. Once the filter is applied, the markers on the map are replaced with only markers which both filters apply to.

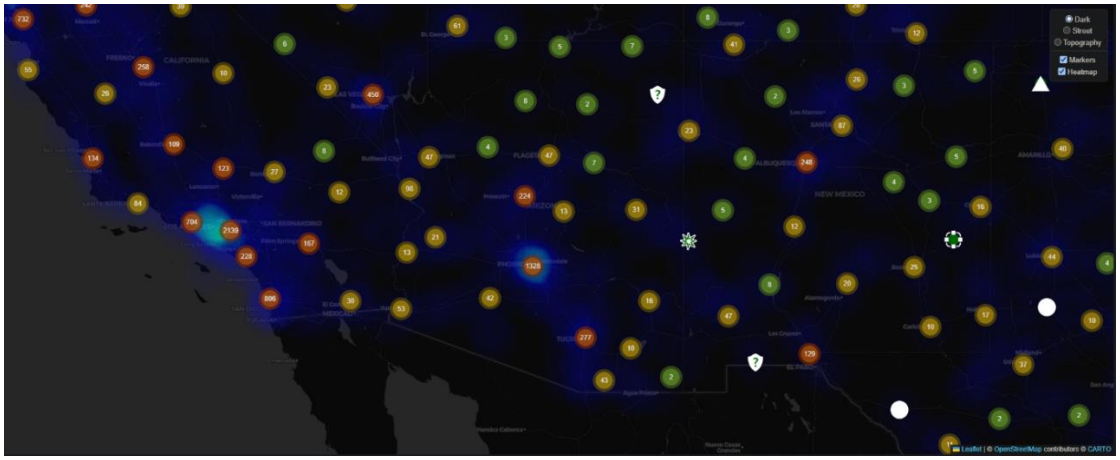


Due to the number of sightings needing to be mapped, most markers will be clustered when the map is set to a zoom wide enough to allow users to seamlessly navigate the map without system resource concerns. When the map is zoomed in enough to see individual markers, you will notice that each marker has been associated with a custom icon to indicate the shape category of that sighting.



When an individual marker is selected, a pop-up appears with certain specific details for each sighting. For this deployment, we focused on the date and shape while also displaying associated comments to provide further context.

Additionally, a heatmap layer has been added to the map to visualize UFO hotspots in the filtered area. This layer is disabled by default but can be toggled on using the layer controls in the top right section of the map.

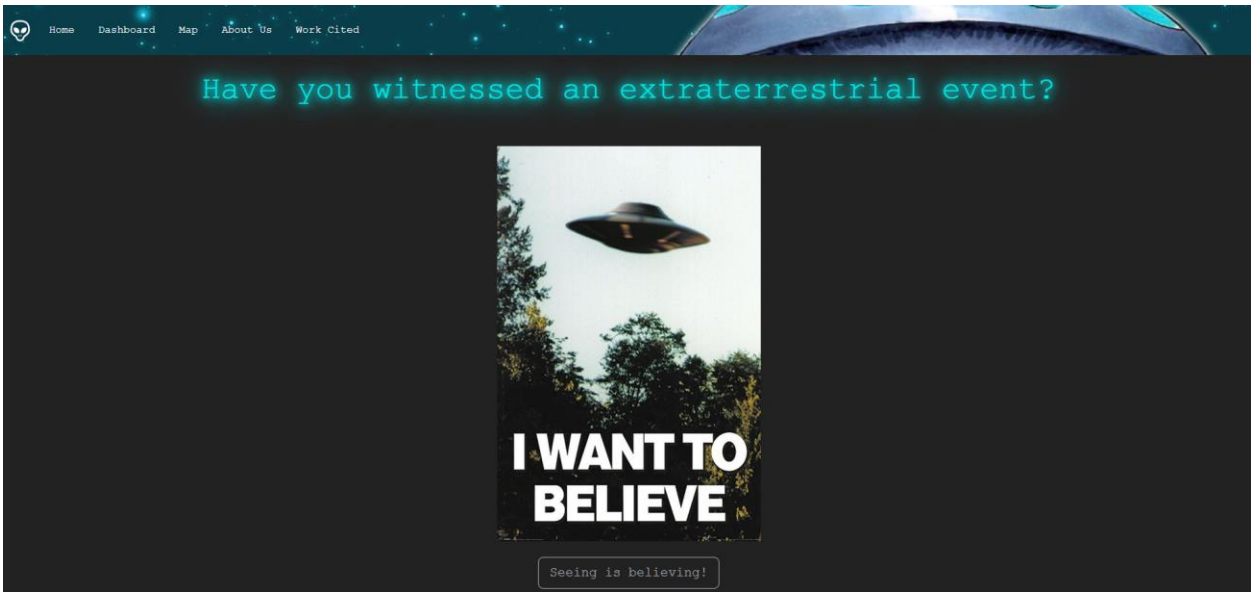


By default, the map layer is using a dark themed map provided by CARTO with attributions displayed in the bottom right of the map. Two additional map tiles have been added for user consideration that may assist with visibility of markers or geographical context.

Styling, Design Concepts, Website Architecture

Originally, we had wanted the color scale to be green. We knew that the background of the website would be dark like the night sky. But after testing some greens against the dark background, we decided to go with something less “loud”. So, we settled on the “Darkly” theme for our website design. The background was not pitch black but rather a dark grey (#2c3e50) which was more pleasant to read text on. The font color for each page was a light blue (#00e6e6) which we added a glow effect to. This closely mimics the blues for the banner image we selected which was a cartoon UFO flying in the night sky.

Even though we were initially inspired by the bold, high contrast imagery from the 2016 UFO sightings dashboard created by Ken Flerlage, we decided to go with softer contrast and muted colors so that it’s more comfortable for the viewer to read and spend time on the website.



Research Questions

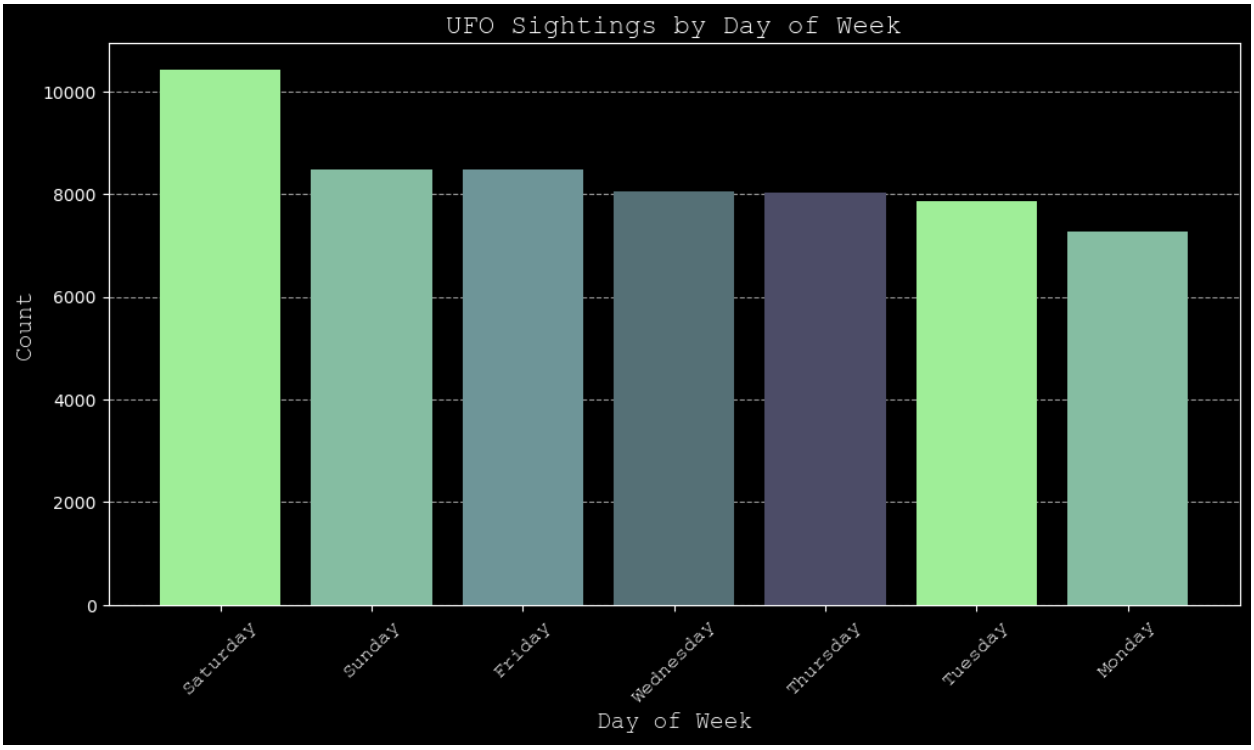
These are the four research questions that we chose to focus on:

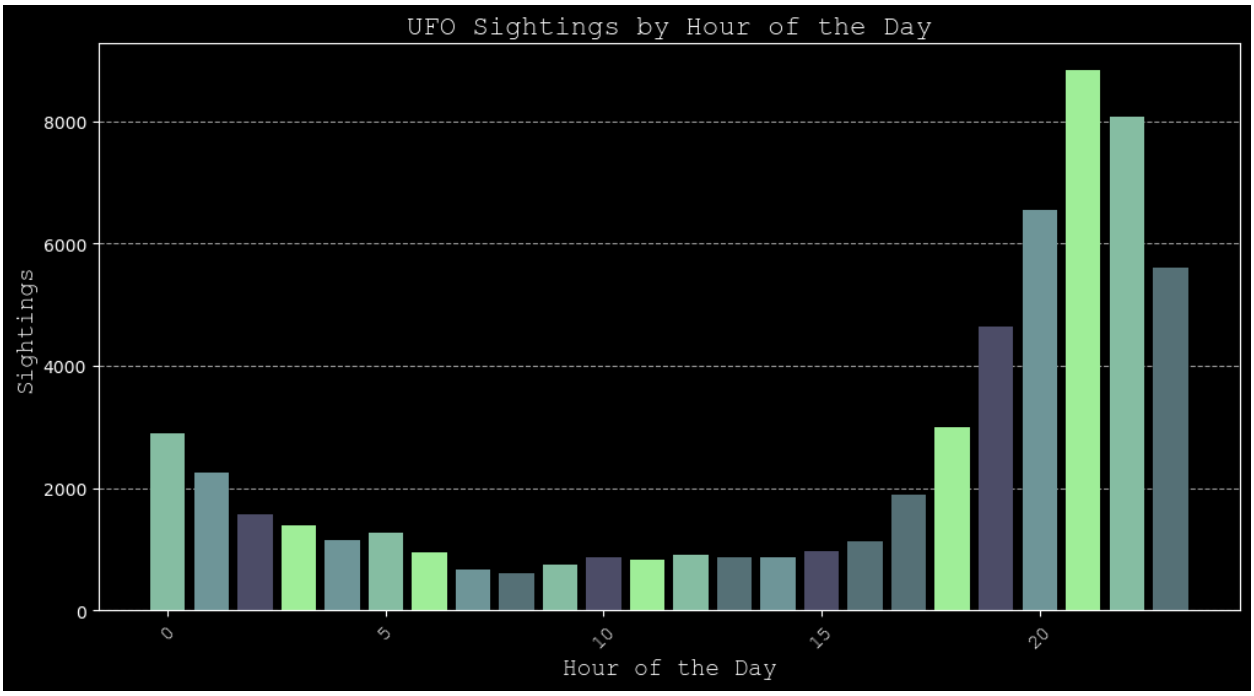
1. What areas of the United States are most likely to have UFO sightings?
2. What days of the week and time of day do UFO sightings tend to occur?
3. What months / years have the highest occurrences of UFO sightings?
4. What is the distribution of shapes observed from UFO sightings?

Analysis and Conclusion

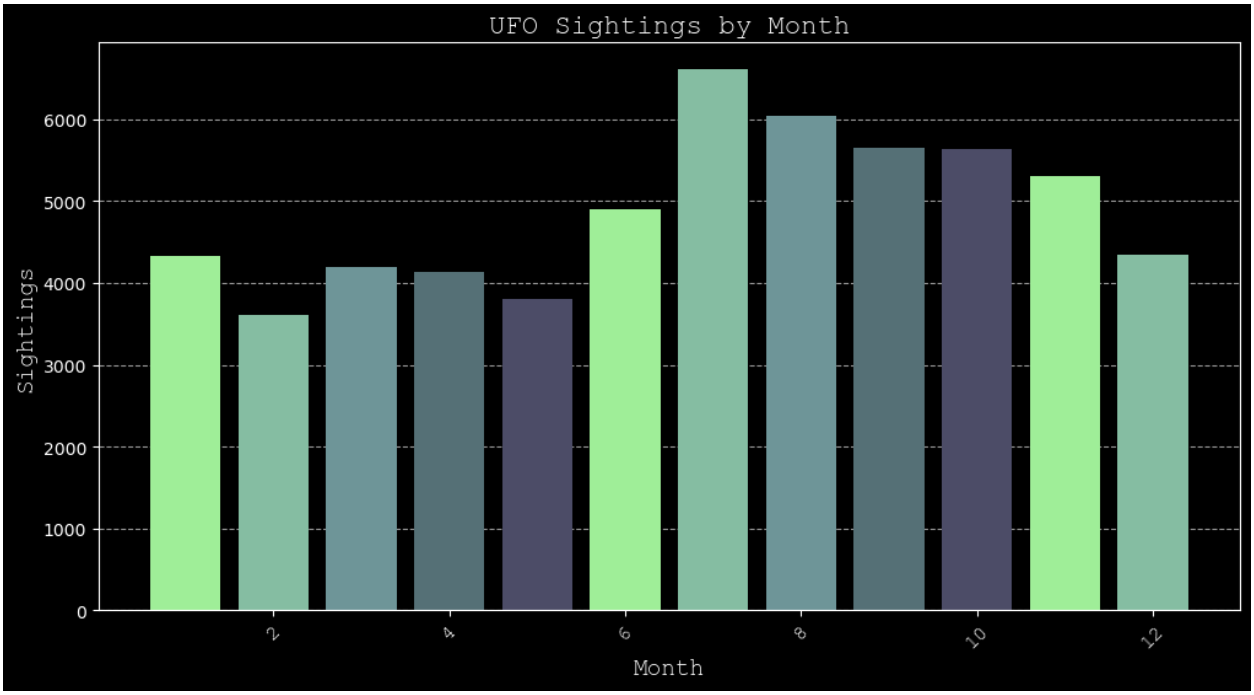
Between 1994 and 2014, California had the most sightings of UFOs with 8,081 sightings, or roughly 13.8% out of all 50 states and Puerto Rico and DC. These sightings were concentrated particularly in the Los Angeles area, and particularly in the year 2012. Since the dataset ended partly through 2014, it is difficult to know whether 2012 has particular significance or whether it was simply part of a trend.

Saturday had the most sightings (10,419 or 17.8% of the weekly sightings), and Monday had the least (7,275 at 12.4%). Sightings peaked at 9 PM, with most sightings occurring after 6 PM and before 1 AM. One hypothesis is that more people are out looking at the sky late Saturday night, and that more people are indoors or going to sleep earlier Sunday night into Monday to start the typical work week. The 9 PM peak tracks as not only does it begin to get dark after 6pm, 9 PM is also when most people are probably still up, out, and able to see the sky.



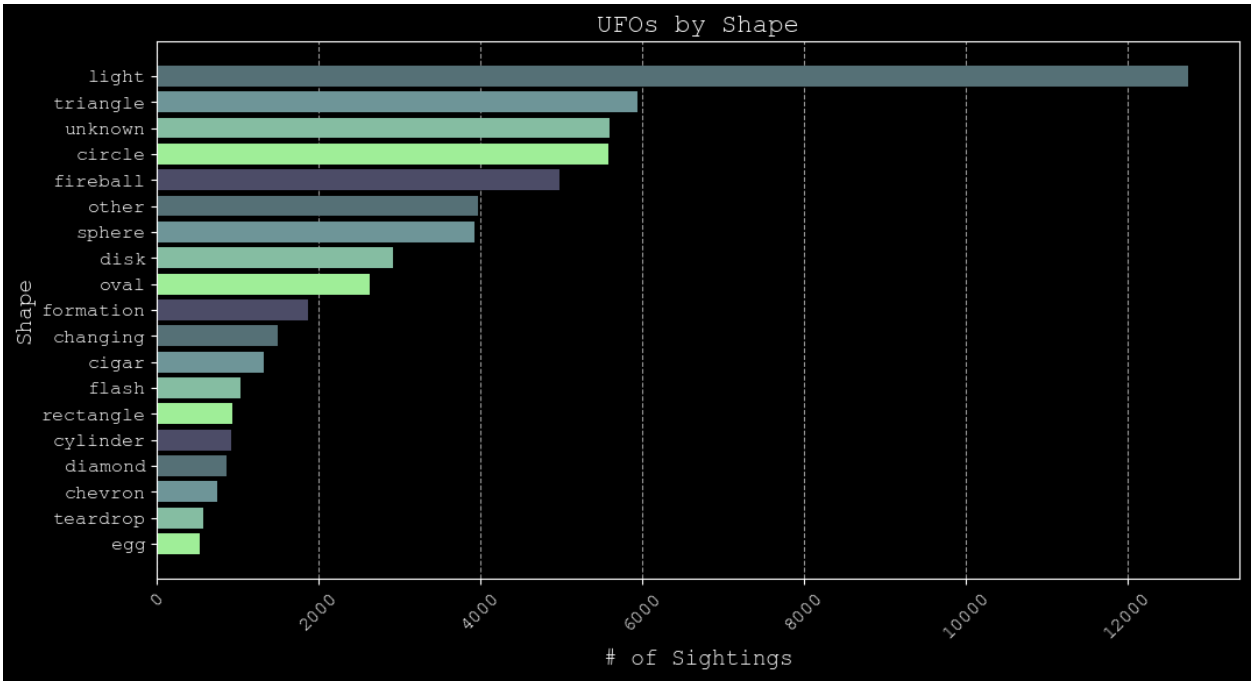


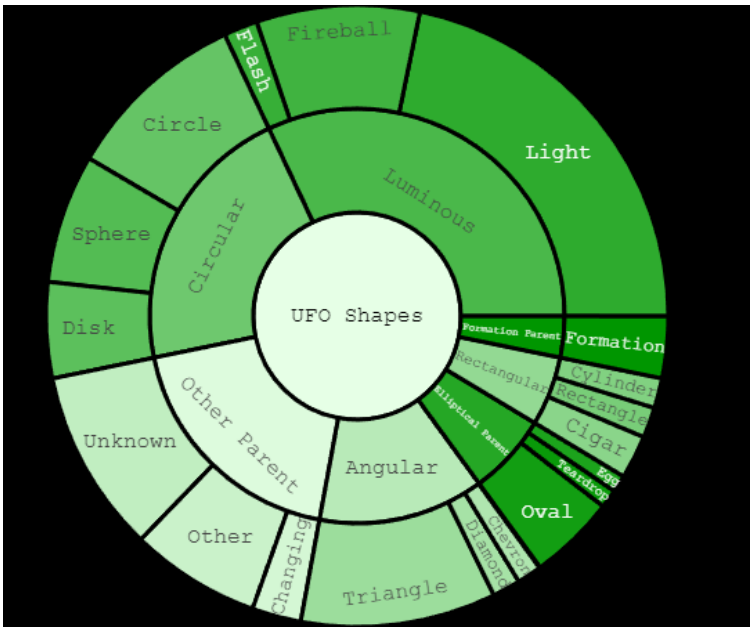
Sightings have been steadily increasing since 1994, peaking in 2012 (2014 had only about half a year's data, which is why it appears to crash after 2013 in the line graph). July had the highest number of sightings (6,605 or 11.3% out of all months for the year). This was followed by June with 6,034 sightings or 8.4%. February, with 3,612 sightings or 6.2%, had the least number of sightings. This could be because February has the fewest days, and/or because it's one of the coldest months. However, disregarding February, it is interesting to note that sightings steadily decrease from January, with the lowest amount in May (3,808 sightings or 6.5%) before spiking in June. One hypothesis is that even though one could assume sightings would increase in March as it begins to warm up, the continued amount decrease could correlate to more people at work or in school, staying inside more and more before finishing the school year before going on summer vacations in June.



Out of the seven categories of shapes, Luminous was by far the most seen shape with 18,754 sightings or 32% of all shapes, encompassing the specific shapes of 'light', 'fireball', and 'flash'. Formations had the lowest shape sightings at 1,874 or 3.2%. Outside of the Luminous category, of the specific shapes, the "triangle" was most seen with 5,947 sightings or 10.2%, followed closely by a virtual tie between "unknown" shapes at 5,591 sightings and the "circle" at 5,587, both 9.5%. After filtering 'light' and 'unknown' shapes, when observing the map, there appears to be a loose correlation of circles seen more often in the northeast and along some coastlines, while triangles

are predominantly seen inland and in the parts of the south. Notable exceptions are DC and Puerto Rico, the only places where 'cylinder' and 'disk', respectively, were seen the most.





Limitations and Bias

NUROC sightings are self-reported and not corroborated by an impartial organization. While the site currently encourages people not to file a report before cross-referencing aerial and celestial events

- satellites, rocket launches, camera lens anomalies, and stars - it is unknown whether this cross-referencing was recommended between 1994 to 2014.

A lack of current data after 2014, contextual information, demographics, personal beliefs are a few limitations of the dataset. The lacking contextual information could include factors such as potential nearby Air Force bases conducting exercises, weather phenomena such as ball lightning, and celestial objects and events, such as satellites, comets, and stars. Demographics, and knowing whether the person reporting a sighting already believes in aliens, could further shape analysis. It would be interesting to be able to compare the nature of sightings from those who were already convinced the sighting to be an alien spacecraft to those who simply did not know what they were seeing. Finally, having data after 2014 could further inform whether peak sightings 2012, and concentrations in Los Angeles were an anomaly, or part of a greater trend.

Future Work and Conclusion

UFO sightings continue to occupy a small subset of the population of the United States. While not the most pressing consideration for the well-being of society, it offers insight into certain observational habits of those who report sightings. One can surmise that the average UFO report will be filed by someone in the greater Los Angeles area on a warm summer night in July right around 9 o’clock. Future work would need to include data up to the present day to update findings such as sighting density and seasonality. Lingering questions include why sightings are increasing. Are there simply more objects in the sky, such as planes, weather balloons, and satellites? Is it simply because the national population increased between 1994 and 2014? Have there been more movies on aliens influencing the imaginations of the public? Were more people going outside and looking up? Or maybe we are just being steadily more visited by aliens since 1994.

Work Cited

Dataset

https://www.kaggle.com/datasets/NUFORC/ufo-sightings

Inspiration

https://towardsdatascience.com/data-analysis-everything-youve-ever-wanted-to-know-about-ufo-sightings-e16f2ed34151

https://towardsdatascience.com/data-analysis-everything-youve-ever-wanted-to-know-about-ufo-sightings-e16f2ed34151

https://public.tableau.com/app/profile/ken.flerlage/viz/UFOSightings2016/Final

Banner image source

https://seller.tcgplayer.com/welcome-to-our-media-library/tcgs/metazoo/ufo-1st-edition/

Icons image source

https://www.iconsdb.com

Code

https://github.com/Leaflet/Leaflet.heat

https://github.com/mourner/simpleheat