UFO Sightings in the U.S. from 1969 to 2019 - Exploratory Analysis

Introduction:

The goal of our project was to explore a dataset from the National UFO Reporting Center (NUFORC) which covered UFO sighting reports from 1969 - 2019. The objective was to explore this data so we entered this endeavor without a predetermined hypotheses. After the initial exploratory phase we entered trying to answer four key questions:

1. How often do sightings occur at certain times of day and at what times during the year?
2. Where do the majority of sightings occur? Does this directly correlate to population size?
3. Are reported UFO shapes seen more in certain locations?
4. Have sightings increased or decreased over this time period?

Exploratory Phase and Data Cleaning:

During our first pass through of the data we found the summary and text columns, which held the description of the sightings to be very interesting. We initially wanted to do a word cloud report using keywords but after running this we realized the findings were quite meaningless and did not deliver what we expected. Another proposed question we had to set aside was calculating the time the sighting occurred to the time it was reported. As we continued to dig into the data, we found too many inconsistencies with the format in which the time was reported, along with many misspellings. After this, we determined the amount of data cleaning required for a dataset this large (90,000 original entries) would be too much of an undertaking with the skills we had learned thus far.

Once we determined the limitations of our dataset, we proposed the four questions we wanted to answer and dove into cleaning up our dataset to easier answer those questions. The original dataset included 12 columns in the csv:

summary	city	state	date_time	shape	duration	stats	report_link	text	posted	city_latitude	city_longitude

We reduced those columns down to nine, five of which were used to make a converted timestamp column.

city	state	year	month	day	hour	minute	converted_timestamp	shape

The removed columns from the original dataset were, summary, date_time, duration, stats, report_link, text, city_latitude and city_longitude.

Data Breakdown:

From here we broke our code/dataset down into five sections.

Sightings: Covering sightings during time of day and year

- Created bins to break the hours down into day parts creating a new column to answer our question of sightings for time of day
- Created a new column to store the month names
- Created a new dataframe to get the number of sightings per month
- From here charts were made to display the number of sightings per month and per time of the day. We utilized bar, line and pie graphs to show our findings.

Sightings Vs. State: Calculate the total sightings per state

- Created a new dataframe to count the number of times each state shows up
- Graphed the sightings per state totals from 1969-2019

Shapes and States: To find if certain shapes were sighted more often in certain locations

- Calculated the number of times each shape was reported in each state
- Converted our table above using a pivot table to get shape types as column headers
- Created a new dataframe dropping the state column to get our totals. We aimed to find the first, second and third most common shapes per state.
- From here, we created an index column for merging. We merged the total count and most often reported shape.
    - Merged that table with the second max findings
    - Merged table 2 with the third max table and put our findings into one full table
    - Renamed the columns and dropped the extra index column
- Pie charts were then created to show the first, second and third most common occurrence of sighting shapes

Sightings Over Time:

- Started by creating bins to break our years data down into decades
- Created bins to also store our day data down into 4 hour blocks
- We then counted the sightings per year and plotted the data using a bar and line graph
- The sightings per month and the sightings per hour were also plotted using line graphs

Sightings by Population:

- Imported census data for state population from census bureau
- Merged population data for each state with our existing state dataframe
- Calculated the number of sightings per hundred thousand residents
- Charted the new data with populations factored in
- Broke down the census data for each 10 year period in our dataset
- Got population data over time and charted the sightings per year

Findings:

How often do sightings occur at certain times of day and at what times during the year?
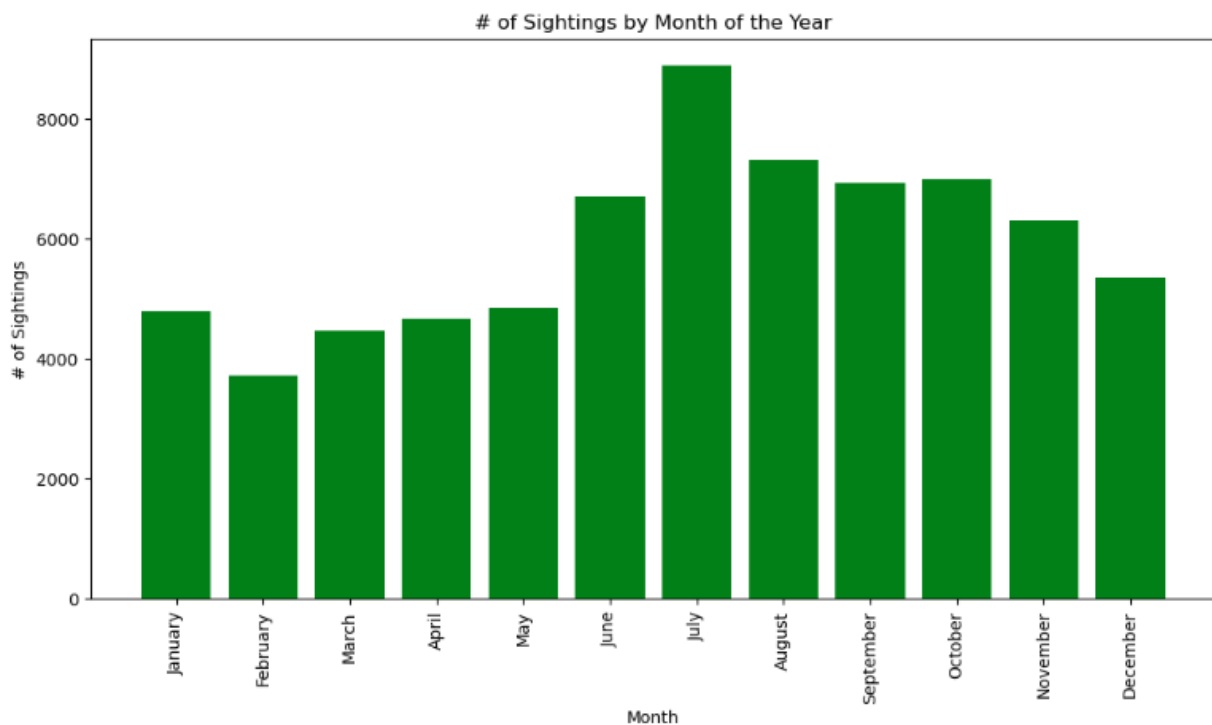
From our findings we can conclude that the majority of sightings occur during the month of July. There is a steady increase in sightings starting in February and have a stark jump up in June before ultimately topping out in July. Then the sightings have a drop off in August then a slow decline as summer wraps up and the colder months begin.

We think we can contribute to the sightings peaking in July to it being the peak time of year across the U.S for warmer weather and more people taking summer vacations and generally being outside more.

# of Sightings by Hour of the Day
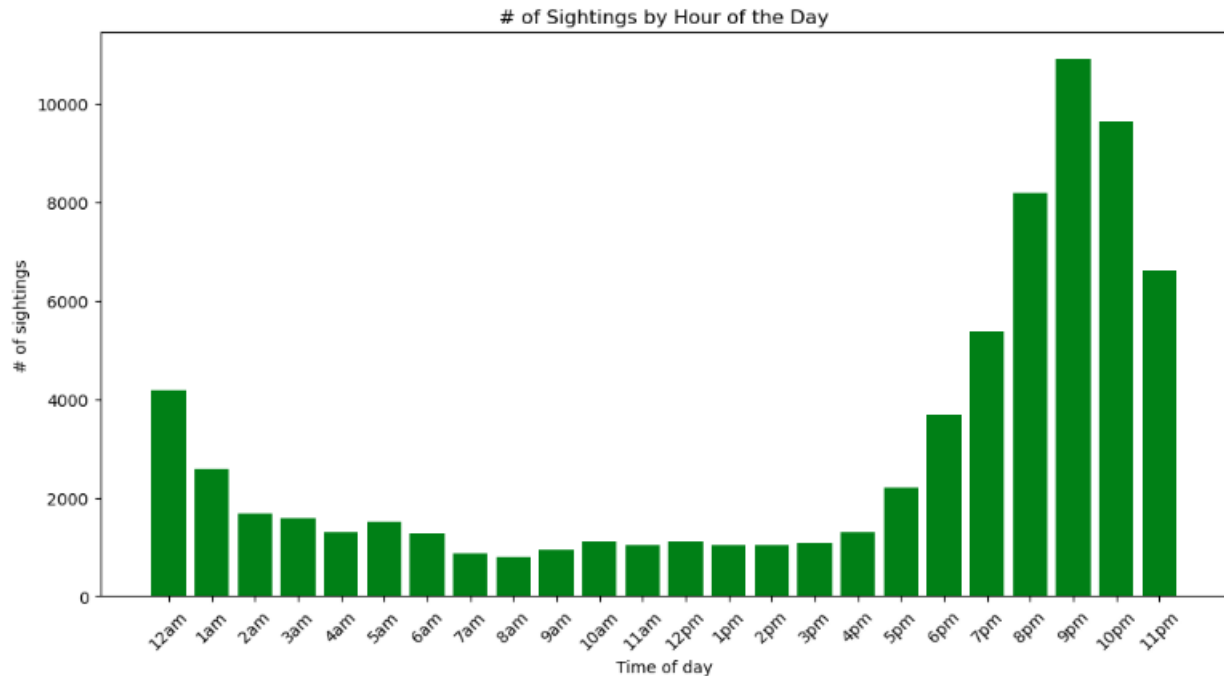
From this chart and data breakdown, we can see the increase in sightings starts to begin the closer you get to sunset and into the night. Sightings peak around 9 - 10pm and then there is a drop off from 11pm - 12am where sightings then level out during the late night hours of 1am through the 4pm hours of the next afternoon.

We can most likely draw some conclusions here that it's easier to see the sky during the night time hours. Also, many Americans are at work during the day and the majority are working inside and not actively looking at the sky. We can also draw some parallels to our later findings in that the most commonly reported shape is a form of light. Which would be easier seen at night and something glowing or moving around would catch your attention more with the backdrop of the night sky opposed to the daytime hours.
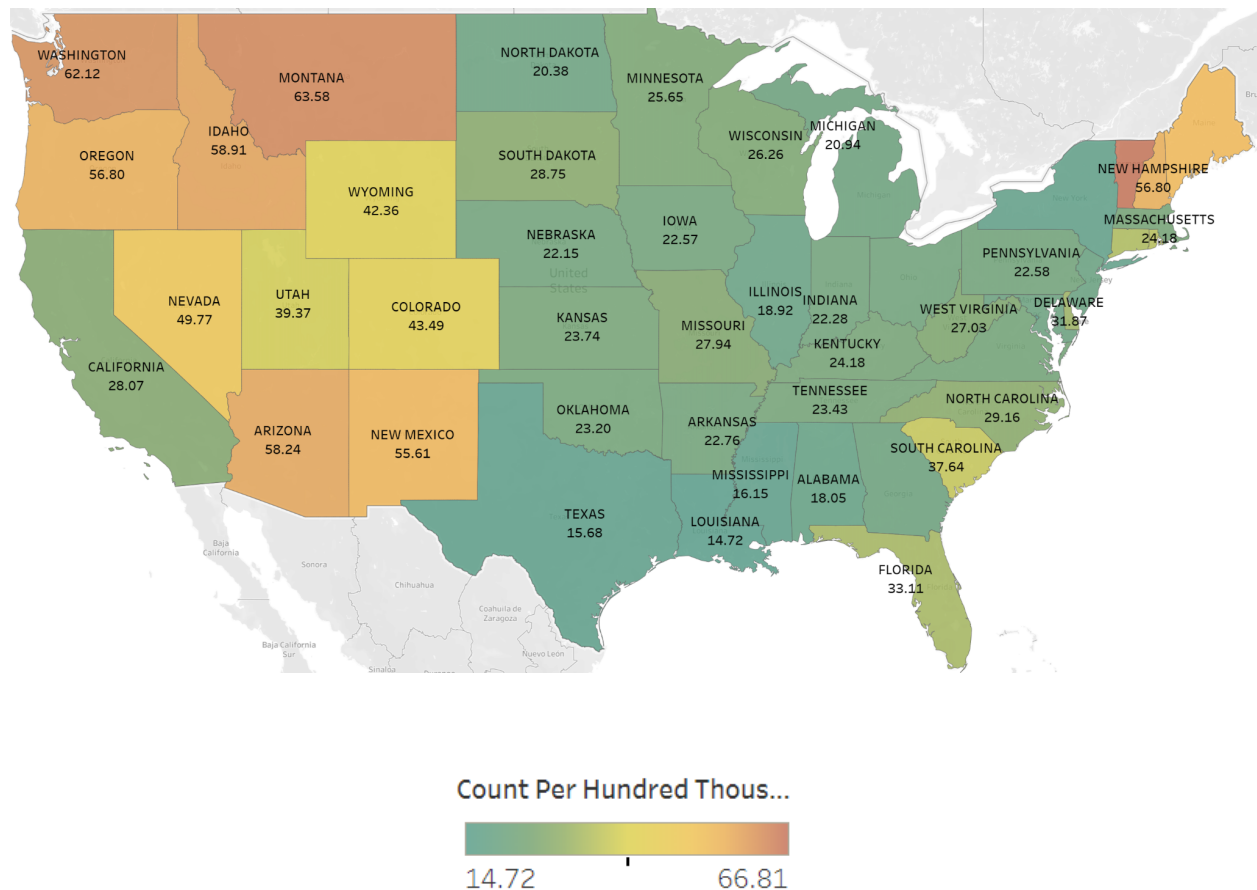

Sightings per Day Part
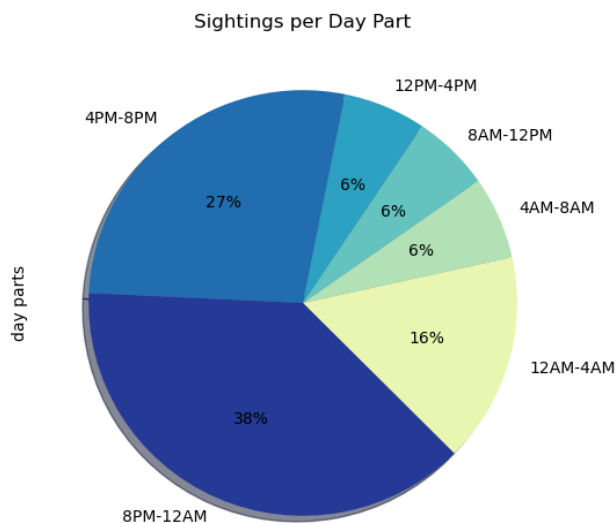
Where do most sightings occur? Is there a correlation to population size?
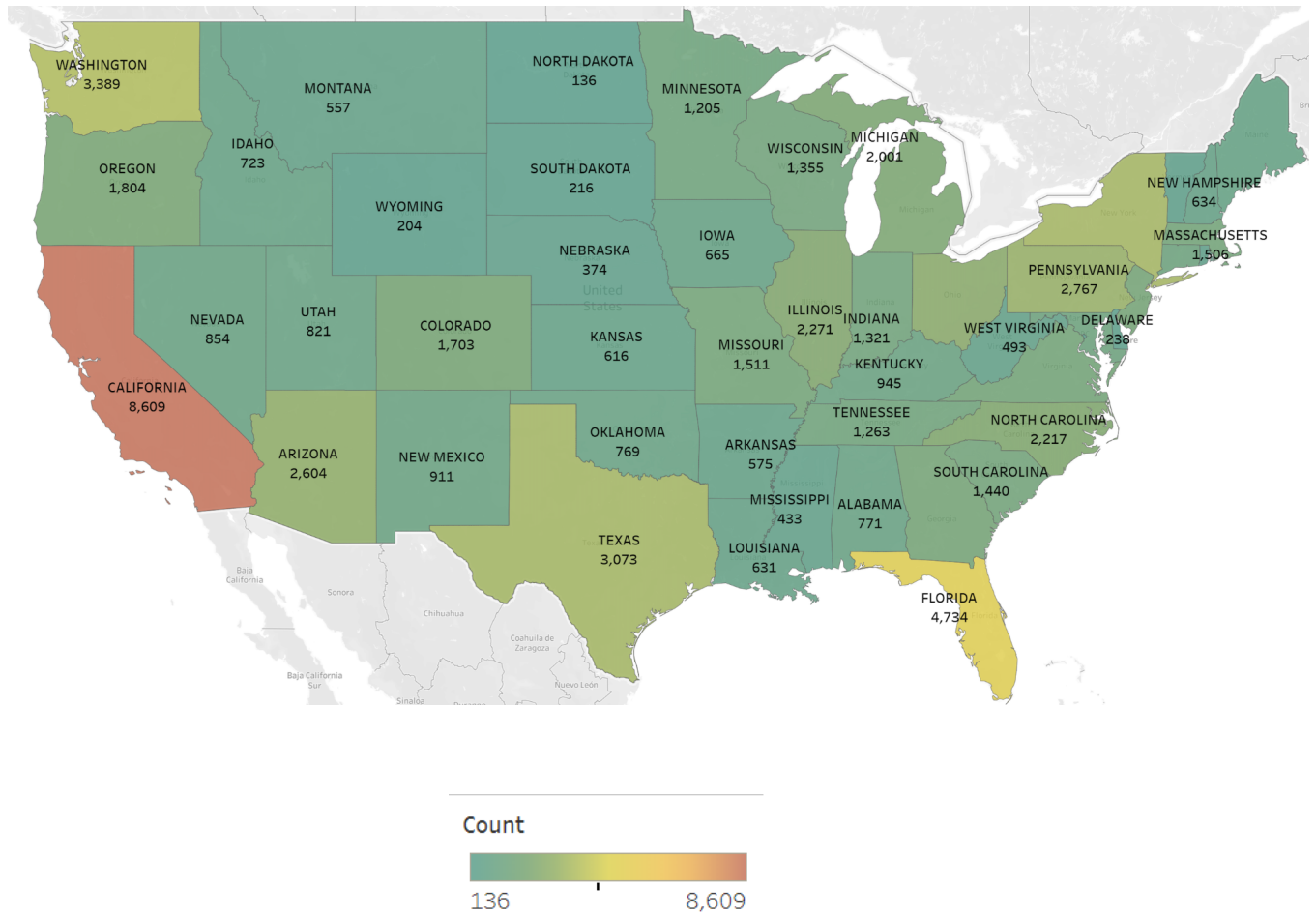
From this data we can clearly see the state of California dwarfing all other states in sightings with nearly 4,000 more reported sightings than the next closest state, Florida. Then there is another steep drop off from Florida to Washington with a drop off of nearly 2,000 sightings to the third most reported state. Some conclusions can be drawn here with California being the most populated state(again need to provide actual numbers). Also California and Florida having the warmest year round weather and outside activity and being heavy vacation destination states. The increase in tourism could directly affect sightings.

But when we break this data down by sightings per 100,000 people, the findings get interesting.

WASHINGTON 62.12
MONTANA 63.58
NORTH DAKOTA 20.38
MINNESOTA 25.65
IDAHO 58.91
OREGON 56.80
WYOMING 42.36
SOUTH DAKOTA 28.75
WISCONSIN 26.26
MICHIGAN 20.94
NEW HAMPSHIRE 56.80
MASSACHUSETTS 24.18
NEBRASKA 22.15
IOWA 22.57
PENNSYLVANIA 22.58
NEVADA 49.77
UTAH 39.37
COLORADO 43.49
KANSAS 23.74
ILLINOIS 18.92
INDIANA 22.28
MISSOURI 27.94
WEST VIRGINIA 27.03
DELAWARE 31.87
KENTUCKY 24.18
CALIFORNIA 28.07
TENNESSEE 23.43
NORTH CAROLINA 29.16
ARIZONA 58.24
NEW MEXICO 55.61
OKLAHOMA 23.20
ARKANSAS 22.76
SOUTH CAROLINA 37.64
MISSISSIPPI 16.15
ALABAMA 18.05
TEXAS 15.68
LOUISIANA 14.72
FLORIDA 33.11

Count Per Hundred Thous...
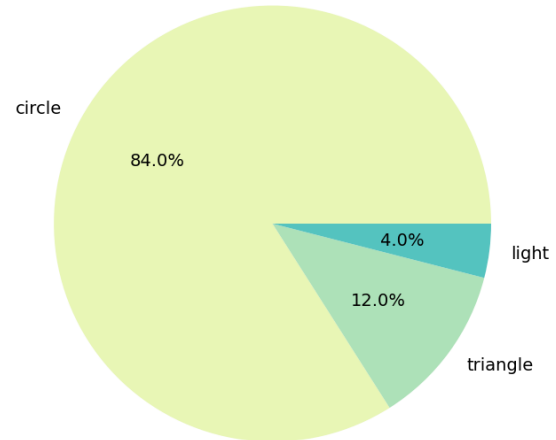
14.72   66.81

Here we can see that Alaska and Vermont lead the way with 67 sightings per 100,000. Followed closely by Montana and Washington State. The fifth through tenth most sightings per 100,000 are Indiana, Arizona, Oregon, New Hampshire, New Mexico and Maine. California, the state with the most total sightings falls to 23rd when broken down by sightings per 100,000 and Florida with the second highest total sightings, is 19th when broken down by sightings per 100,000.
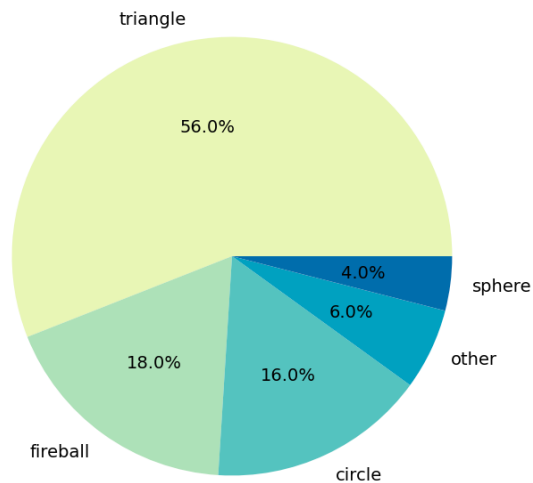
Are certain shapes sighted more in certain locations?
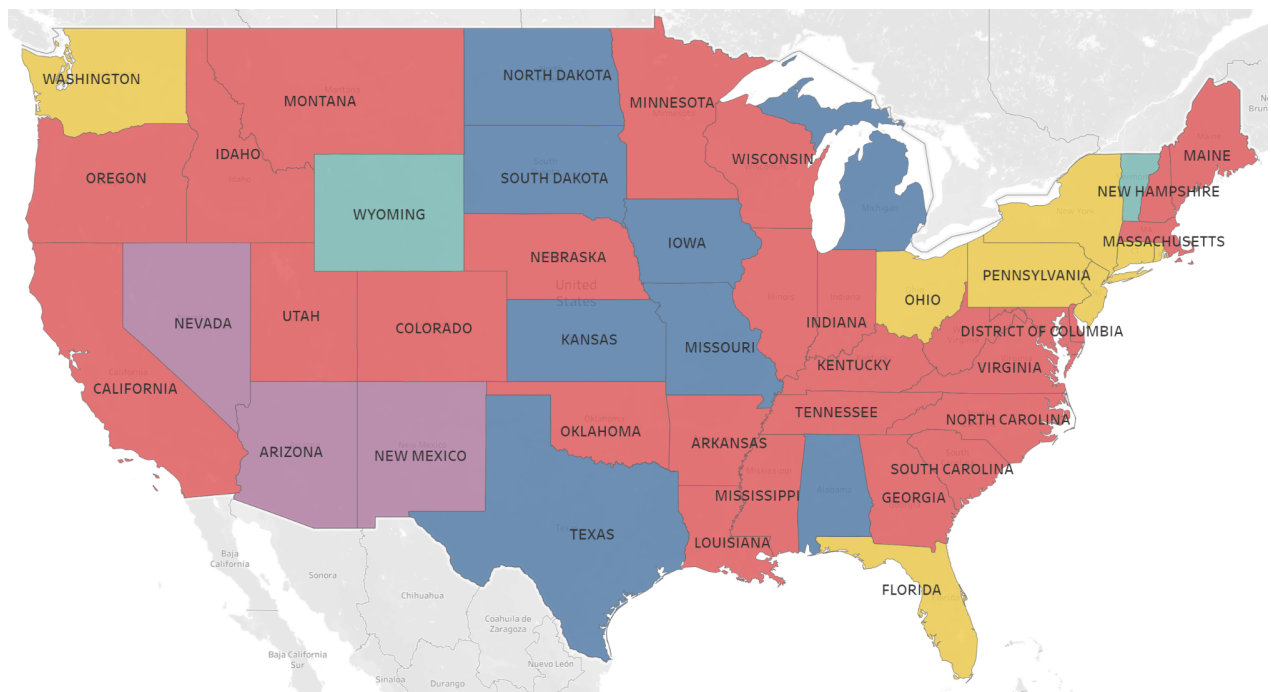
Most Common Shapes Reported in UFO Sightings

light — 96.0%
triangle — 2.0%
circle — 2.0%

Second Most Common Shapes Reported in UFO Sightings

circle — 84.0%
light — 4.0%
triangle — 12.0%
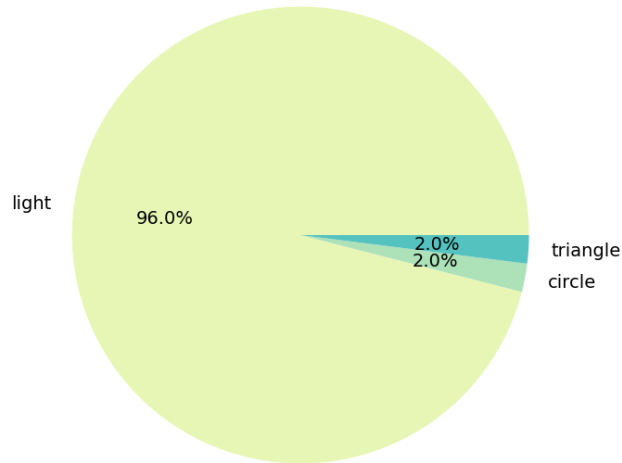
Third Most Common Shapes Reported in UFO Sightings

triangle — 56.0%
sphere — 4.0%
other — 6.0%
circle — 16.0%
fireball — 18.0%

The findings here were pretty cut and dry. The most common "shape" reported was a form of light with a resounding 96%. Which can be understandable when most reports occur at night and depending on the distance of the sighting, it might be hard to make out anything more than an abnormal light. The second most common reported shapes are circle and triangle, which light again showing up as 4% of these findings. It's not until we get to the 3rd most common shape reported that we see the data start to differentiate some and show variety. Triangle makes up 56% of the 3rd most reported sightings, followed by fireball with 18% and circle with 16%.
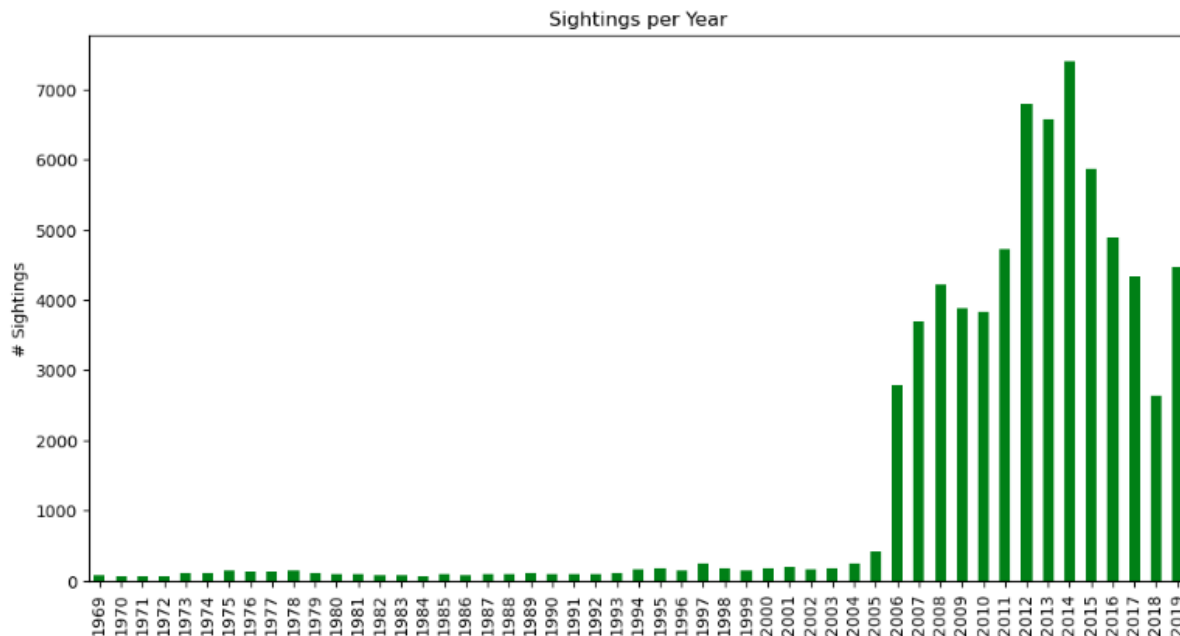
Third Most Common Shape
- circle
- fireball
- other
- sphere
- triangle

Have sightings increased or decreased over time?

Sightings per Year

From this dataset we can see that reports were steady in the hundreds from 1969 - 2005. Starting in 2006 the number of reported sightings jumped to nearly 3,000 and then steadily climbed to its peak in 2014 with around 7,500 sightings being reported. There was then a drop off per year from 2015 - 2018 where it then spiked to a 3 year high in 2019.

Year Over Year Sightings Percent Increase/Decrease
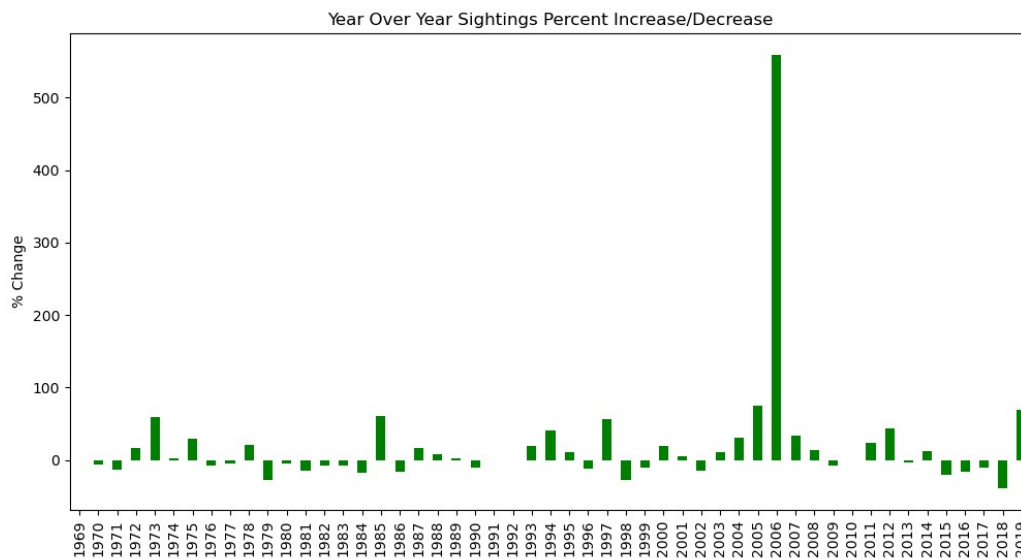
It's hard to really determine the cause of the sudden spike in 2006 but we could draw correlations to the popularity of the internet and forum discussions bringing the topic more into the social forefront. Also the ability to quickly report sightings via the internet could explain the jump. Also, as the years have gone on, access to high speed internet has spread more across the country to also include rural areas bringing the ability to report to more and more people.

So it's with safe certainty that we can conclude that yes, reports have increased over time, but specifically between the time from 2006 - 2019 showing the highest growth rates.

Limitations and Next Steps:

We acknowledge that our dataset is only from one source and we could do some cross comparisons from other sources. The National Archives could be an interesting place to start to also catalog official government/military sightings that have been reported and made public over the last few years.