UAP Sightings

Final ENGG 182 - Data Analytics
Master's in Engineering Management - Thayer School of Engineering

Written by Pranav Dharmadhikari

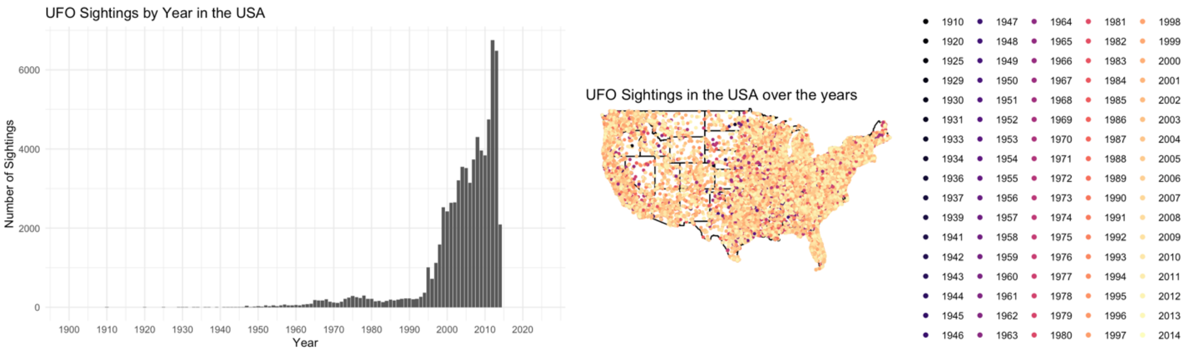
Table of contents

1. Introduction

This analysis investigates Unidentified Aerial Phenomena (UAP) sightings in the United States through a technical lens. The study visualizes the geographic and temporal distribution of sightings, examining variations by shape, time, and duration. Rigorous time series analysis identifies trends or seasonality patterns in UAP occurrences. Forecasting techniques are employed to predict future sightings, along with proposed modifications for improved accuracy. Additionally, the potential for clustering or classifying reported sightings by type is explored, considering the application of machine learning techniques. This multifaceted approach aims to unravel the underlying patterns and contribute valuable insights to the understanding of UAP sightings.

1.1. Visualizing the geographic distribution

To begin, we filtered the dataset to include only sightings within the USA. We then visualized the geographic distribution by plotting sightings on a map, with an interactive R Shiny app allowing users to explore sightings year-by-year. To summarize the temporal patterns, we graphed the number of sightings per year. This revealed peak years with heightened UFO activity, which we could further investigate. This combined geographic and temporal visualization provided a comprehensive yet concise overview of UFO sighting patterns across the USA over time.

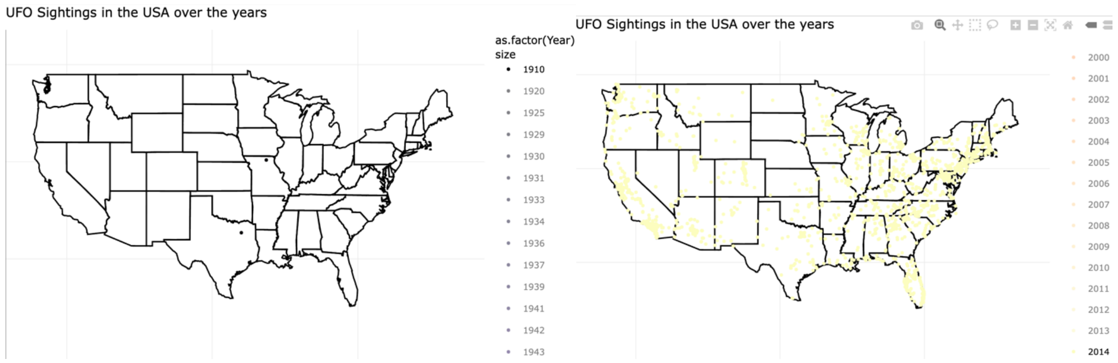


1.1.1 & 1.1.2 UFO sightings by Year in the USA and Geographic distribution

The sightings graph shows a shift from sporadic to consistent UFO reports from 1947 onward, coinciding with the Roswell incident and increased public interest. 2012 emerges as the peak year, potentially influenced by societal, cultural, or technological factors, followed by a declining trend. To understand these fluctuations, key years like 1947 and 2012 warrant analysis to identify contributing events/factors behind sighting spikes. Examining 2014 as a post-peak year and 1910 as the earliest data point provides further context. This targeted analysis of significant years aims to uncover the underlying reasons behind UFO sighting patterns over time and the graph on the right visualizes global UFO sightings, with each point indicating a reported sighting from 1910 – 2014.

1.2. Year 1910 and 2014

The visualization on the right starkly contrasts the sparse 1910 sightings with the increase in the sightings in 2014, suggesting a surge in UFO sightings. In 1910, sightings concentrated in Texas and Missouri. By 2014, the landscape transformed dramatically. The East Coast emerged as a hotbed, with Florida, New Jersey, Maryland, and Delaware reporting frequent sightings. This concentration could stem from regional population factors, environmental conditions favoring sightings, or prevalent cultural beliefs about UFOs. Notably, while the East Coast dominated the number of UAP sightings in the year 1910 and 2014, the West Coast, particularly California, also exhibited a significant cluster. This nationwide presence across multiple regions indicates that UFO phenomena manifest broadly, transcending geographic boundaries.

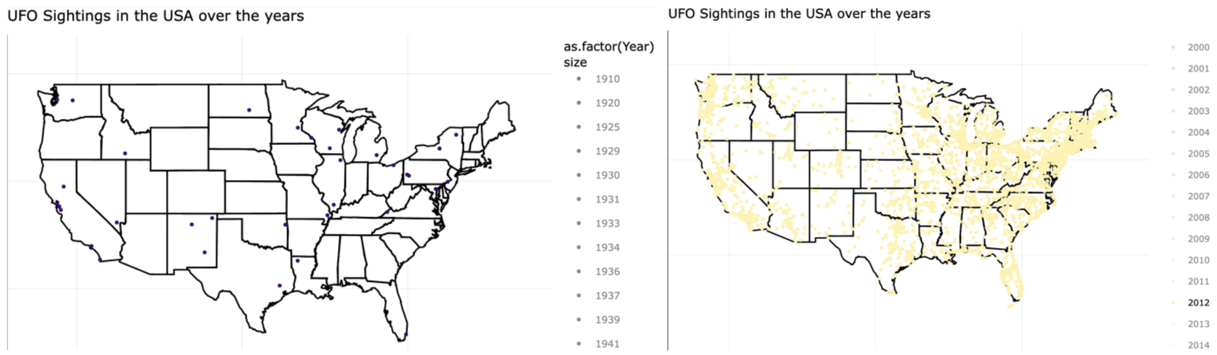


1.2.1 & 1.2.2 Number of sightings during 1910 & 2014.

1.3. Year 1947 and 2012

In 1947, a year that marked increased prominence of UFO sightings, the geographic distribution revealed widespread reports across the United States. While sightings occurred nationwide, certain regions exhibited higher concentrations. A notable northeastern cluster emerged in states like Wisconsin, Michigan, and New York. The southern region, particularly Texas, also witnessed significant activity. On the West Coast, Washington and California stood out with scattered but distinct sightings. This pattern suggests UFO phenomena transcended geographic boundaries in 1947, yet certain areas experienced more frequent occurrences. The nationwide presence, coupled with regional hotspots, underscores the complex nature of UFO sighting dynamics during this pivotal year.

In 2012, the US saw a spike in UFO sightings, mainly clustered in the northeastern states, the southern region, and the west coast (California and Washington).

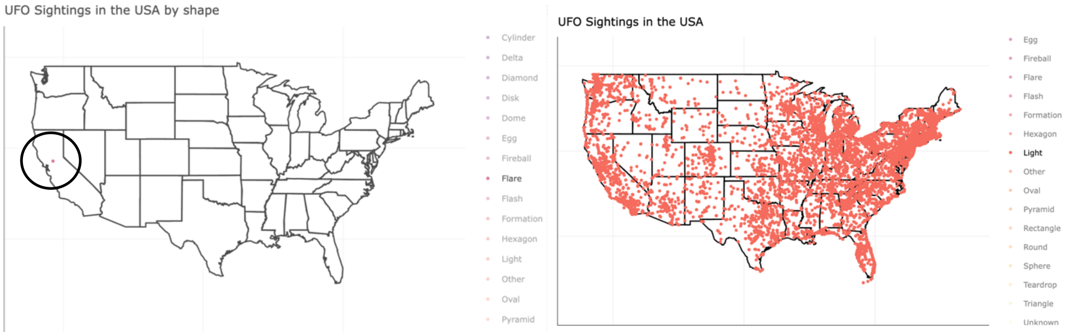


1.3.1 & 1.3.2 Number of sightings in 1947 & 2012

2. Distribution

2.1. Distribution by shape

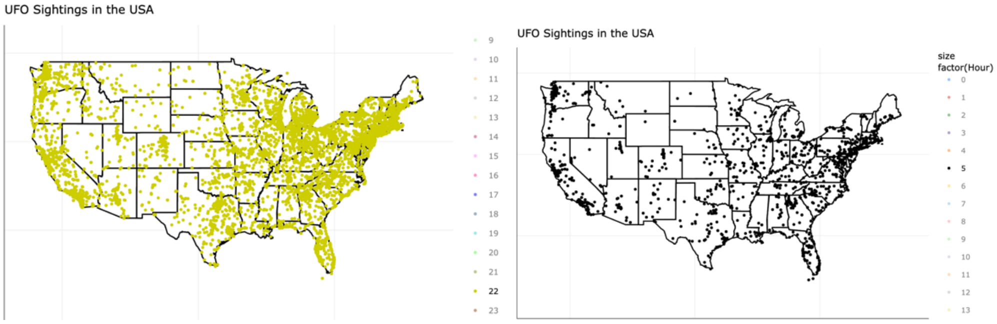
The visualization highlights different UFO shapes recorded over time. At first glance, no clear pattern emerges, but upon closer examination of specific shapes, trends become apparent. Shapes like "light," "fireball," and "circle" stand out as more common, while others like "cone," "flash," and "chevron" are less frequent. Despite the initial lack of pattern, pinpointing specific shapes reveals valuable insights into UFO sightings.



2.1.1 & 2.1.2 Flare shaped sightings & Light shaped sightings

2.2. Distribution by time

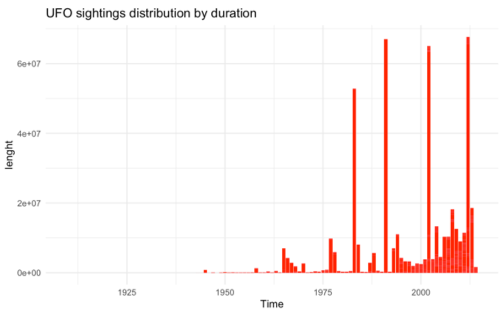
When examining the visualizations, we discover that UFO sighting activity is least between 5pm to 7pm, with a noticeable incline between 10pm and 1am. This temporal pattern suggests that evening hours see the most frequent reports of UFO sightings, while late-night hours experience fewer incidents.



2.2.1 & 2.2.2 Sightings at 10PM and Sightings at 5PM

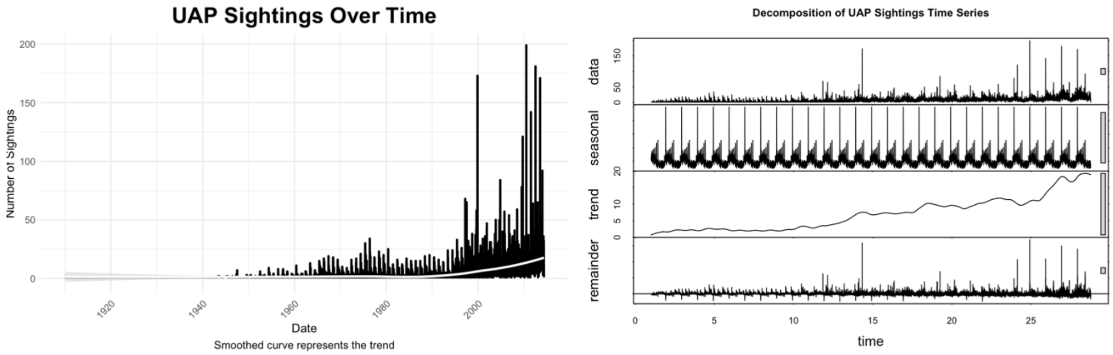
2.3. Distribution by the duration

The graph of UFO sightings in the United States reveals intriguing patterns over decades. A notable surge occurred in 1947, with a significant spike in reported incidents. Another peak emerged around 1990, marked by increased sightings and longer durations. Interestingly, the early 2000s saw UFO sightings peak again, rivaling 1990's intensity. This suggests a shift in sightings' prevalence, possibly influenced by societal changes or technological advancements. The sustained elevation into the new millennium indicates a notable trend in reporting and documenting these phenomena.



2.3.1 Duration of UFO sightings over the years

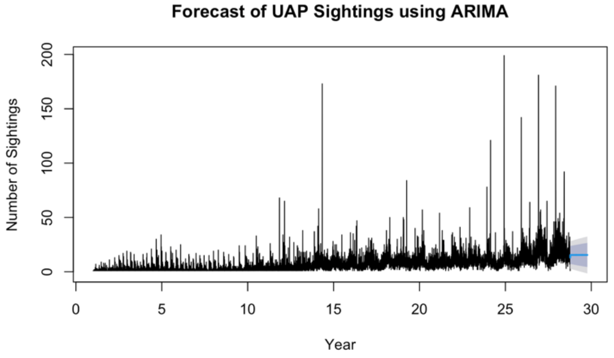
3. Time series analysis



3.1.1 & 3.1.2 UAP Sightings over time & Time Series Analysis

Yes, there are trends according to the time series analysis. The graph shows the time series, seasonal, trend, and irregular components from 1910 to 2014.Note that the seasonal components have been constrained to remain the same across each year. The grey bars on the right are magnitude guides—each bar represents the same magnitude. This is useful because the y-axes are different for each graph. Trend suggests that the UFO sightings have been increasing over the period that we are considering, and it can be attributed to technological advancements enabling better detection through enhanced surveillance capabilities, as well as the widespread availability of recording devices among the public. Improved official monitoring systems coupled with ubiquitous cameras on smartphones have heightened awareness and media coverage of unidentified aerial phenomena, contributing to their rising prevalence.

4. Modelling:



4.1.1 ARIMA Model

4.1. Interpretation

AR (1) = 0.1652: The autoregressive coefficient suggests a weak positive relationship between the current observation and the previous time step. This means that an increase (decrease) in the previous observation is associated with a small increase (decrease) in the current observation.
MA (1) = -0.9597: The large negative moving average coefficient indicates a strong negative correlation between the current error term and the previous error term. This suggests that overestimations (underestimations) are likely to be followed by underestimations (overestimations) in the next time step, allowing the model to correct itself.

Variance: Error variance (sigma^2) = 39.42: The moderate variance value represents the average squared deviation of the residuals from the predicted values. A lower variance would indicate a better overall fit.

Model Fit:
Log Likelihood = -33006.63: This value quantifies the likelihood of observing the data given the model parameters. Higher values indicate a better fit but must be interpreted in the context of the model complexity.
AIC = 66019.25, BIC = 66040.92: These information criteria penalize the log-likelihood based on the number of parameters, preventing overfitting. Lower AIC/BIC values suggest a better trade-off between goodness-of-fit and model parsimony.

Training Error:
ME = 0.0295: The small positive value indicates a slight tendency to overestimate the actual values on average.
RMSE = 6.2774: As a measure of the typical magnitude of prediction errors, the RMSE suggests deviations from actual values within a reasonable range.
MAE = 3.1169, MAPE = 70.5173%: These metrics quantify the average absolute errors, with MAPE expressing it as a percentage. The high MAPE could indicate potential for improved accuracy.
MPE = -48.1156%: The large negative value suggests a systematic tendency to underestimate the actual values by a considerable margin on average.
MASE = 0.7357: As the MASE is less than 1, the model performs better than a naïve forecast that uses the last observation as the prediction.
ACF1 = 0.0007: The nearly zero autocorrelation at lag 1 suggests little remaining serial correlation in the residuals, indicating the model captured most of the time series dynamics.

Overall, while the ARIMA (1,1,1) model exhibits reasonable fit and error measures on the training data, the high MAPE and tendency to underestimate point to potential areas for improvement, such as exploring alternative ARIMA configurations or exogenous variables to enhance forecasting accuracy.

4.2. Is the model accurate?

The provided forecast, while offering insights into potential high-level trends, may lack precision in predicting accurate future UFO sighting counts. This limitation could arise from the following factors:

Data Quality Concerns: The forecast's accuracy could be affected by incomplete or inconsistent data in the dataset. Variations in reporting practices, recording methodologies, or biases in data collection over time may obscure or distort the underlying patterns, leading to biased forecasts.
Model Fitting Inadequacies: The chosen forecasting model, such as ARIMA, may not fully capture the intricate dynamics and complexities associated with UFO sighting occurrences. Factors like seasonality, long-term trends, or external influences might not be adequately accounted for, compromising the model's ability to generate highly accurate predictions.

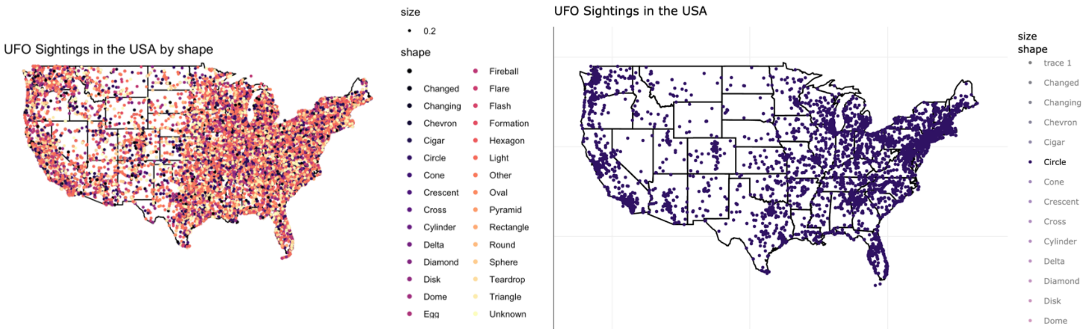
5. Potential improvements

Weather Data: Knowing weather conditions helps discern if sightings could be atmospheric reflections, weather balloons, or other natural events like storms. Factors like wind speed, visibility, and atmospheric pressure are crucial.
Satellite Imagery: High-resolution satellite images can confirm sightings and rule out misinterpretations like satellites, weather balloons, or atmospheric disturbances.
Sensor Data: Data from surveillance cameras, aircraft instruments, or other sensors offer additional evidence. This includes infrared imaging, electromagnetic measurements, and acoustic recordings.

We will use the following Machine Learning Techniques.

Random Forest: Handling of categorical and numerical data is very efficient when it comes to Random Forest methodology. It can handle both categorical and numerical data types, making it suitable for analyzing diverse features such as shape, duration, geographic location, and time of day of sightings, as well as any additional data that could be collected.
Neural Networks: Neural Networks can scale effectively to large datasets, accommodating the potentially vast amount of historical and future data on UFO sightings. As the dataset grows over time with additional reports, Neural Networks can continue to learn from the expanding data to improve classification performance.
K-nearest Neighbors (KNN): KNN is effective at handling multi-modal data distributions, where different types of UFO sightings may exhibit distinct clusters in feature space. By considering the nearest neighbors of each data point, KNN can accurately classify sightings even when they belong to different modes or clusters.

6. Appendix



6.1.1 & 6.1.2 Sightings by Shape & Circle Shaped sightings