UFO Sightings Around The World

Project 3 Group 2 with Kerry Oostdyk, Misha Mambully, Alison McCondichie, and William Berry

Our Motivation:
UFOs have captured the fascination of human minds from time immemorial. As a group, we too felt intrigued when we came across this dataset on Kaggle. We thought this could be a good opportunity for us to explore the UFO sightings through data science and visualizations.

Introduction:
Our project explores a dataset of UFO sightings reported from around the world. We are aiming to identify patterns and trends from these encounters through a web application. We are using SQLite as the database, Flask for the back-end, JavaScript with Plotly for data visualization, and we wanted to provide an interactive platform to explore UFO sighting data.

Key Research Questions:
When looking through the data, our team came up with 3 driving questions to guide us through our analysis. We first wanted to determine if there are any trends in the reports over time. In addition, we wanted to see what areas of the world reported the most sightings. Finally, we decided to analyze the duration of the encounters in relation to the time of day the incidents occurred.

Dataset Overview:

Source of the dataset : Kaggle

https://www.kaggle.com/datasets/jonwright13/ufo-sightings-around-the-world-better

Data Cleaning and Feature Engineering:

Renamed columns: For clarity some of the column names were renamed.

```python
ufo_df = df.rename(columns = {'latitude' : 'Latitude', 'longitude' : 'Longitude'})
```

Removed unnecessary columns: Some column names which were not required for the data analysis and visualization were removed from the dataset.

```python
ufo_df = ufo_df.drop(['date_documented', 'Country_Code', 'length_of_encounter_seconds', 'Encounter_Duration', 'Description'], axis=1)
```

Dropped null value rows: Some rows with null values were dropped from the dataset. Since the dataset was large and the dropped rows were very small, it did not affect the dataset as a whole.

```python
ufo_df = ufo_df.dropna(subset=['Region', 'Locale', 'UFO_shape'])
```

Changed data types: The date related columns were changed to appropriate data types.

```python
df['Date_time'] = pd.to_datetime(df['Date_time'])
```

Here is a dataset overview after data cleaning and feature engineering.

```
<class 'pandas.core.frame.DataFrame'>
Index: 77676 entries, 0 to 80327
Data columns (total 12 columns):
 #   Column                               Non-Null Count  Dtype
---  ------                               --------------  -----
 0   Date_time                            77676 non-null  datetime64[ns]
 1   Year                                 77676 non-null  int64
 2   Month                                77676 non-null  int64
 3   Hour                                 77676 non-null  int64
 4   Season                               77676 non-null  object
 5   Country                              77676 non-null  object
 6   Region                               77676 non-null  object
 7   Locale                               77676 non-null  object
 8   Latitude                             77676 non-null  float64
 9   Longitude                            77676 non-null  float64
 10  UFO_shape                            77676 non-null  object
 11  Estimated_Encounter_Duration_Minutes 77676 non-null  float64
dtypes: datetime64[ns](1), float64(3), int64(3), object(5)
memory usage: 7.7+ MB
```

SQL Queries:

We used sql queries within our flask to interact with the UFO sighting data stored in the database.Below is listed some of the example sql queries that were used.

```
SELECT Country as country,count (*) as num_ufosighting

                FROM UFO_data

                WHERE Year >= {min_year}

                GROUP BY Country

                ORDER BY num_ufosighting desc

                LIMIT 10;
```

The above SQL query retrieves the top 10 countries with the most UFO sightings, filtering the data based on a user-defined minimum year. The result is grouped by the country and ordered in descending order by the count of sightings.

```
SELECT Year as year,Season as season,Hour as hour,Region as region,Country
as country, UFO_shape as UFO_shape, Latitude as latitude, Longitude as
longitude

    FROM UFO_data

    WHERE Year >= {min_year}

    ORDER BY Year asc;
```

The above query selects multiple fields from the UF0 sightings including the year,season,hour ,region, country ,ufo_shape,latitude,longitude and filters the results to include only those sightings that occurred in or after the

3

user-specified minimum year.This query is used to plot UFO sightings on the map.

Data Visualizations:

Design:

We created our dashboard so that it answered each of our key questions. It includes a graph showing the number of sightings by year, our sightings by location, the top 10 locations for sightings, a bubble chart showing the duration of sightings based on time of day, the number of sightings by season, and finally our leaderboard. Each of these graphs changes with the two filters; one for the year and one for the UFO shape. In addition we made our dashboard to fit the same Minty color theme we have throughout the web app.

**What are the trends observed with time in UFO sightings?**

In order to answer this question we plotted a bar graph with the number of sightings and the corresponding year. We observed that the number of UFO sightings increased in the later years. After the year 2000 we can clearly see an upward rise of the number of UFO sightings. See the Figure below.
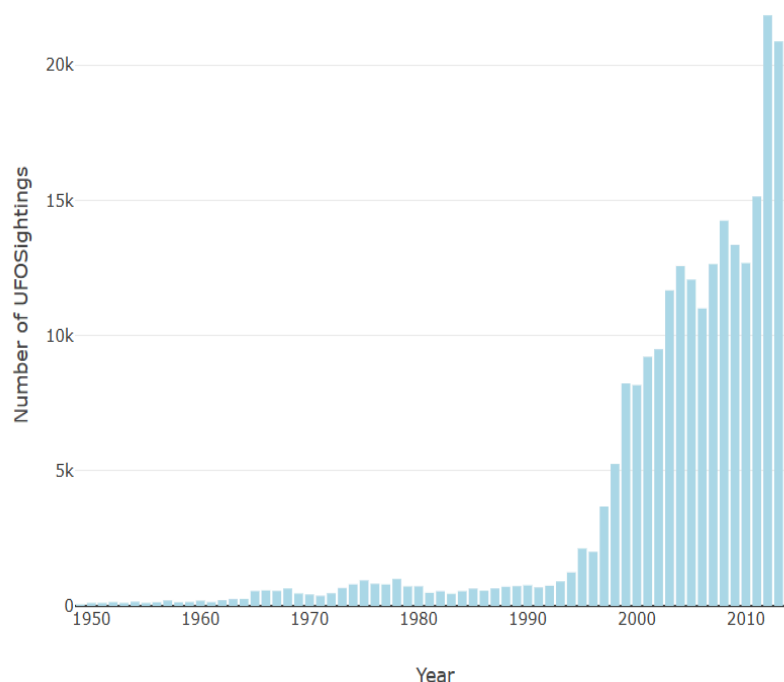
This increase could be because of several factors. The widespread use of phones with cameras and especially with smartphones made it easier for people to capture and report the sightings. Also, the development of online reporting platforms may have encouraged individuals to make reports of these sightings.

**Which regions of the world report the most UFO sightings ?**

We plotted a scatter plot to identify the regions with the most UFO sightings. We found that the USA is the country that has the highest number of sightings. Starting from the year 1949 to 2015, we see that the United States tops the total number of UFO sightings.

UFO Sightings by Location
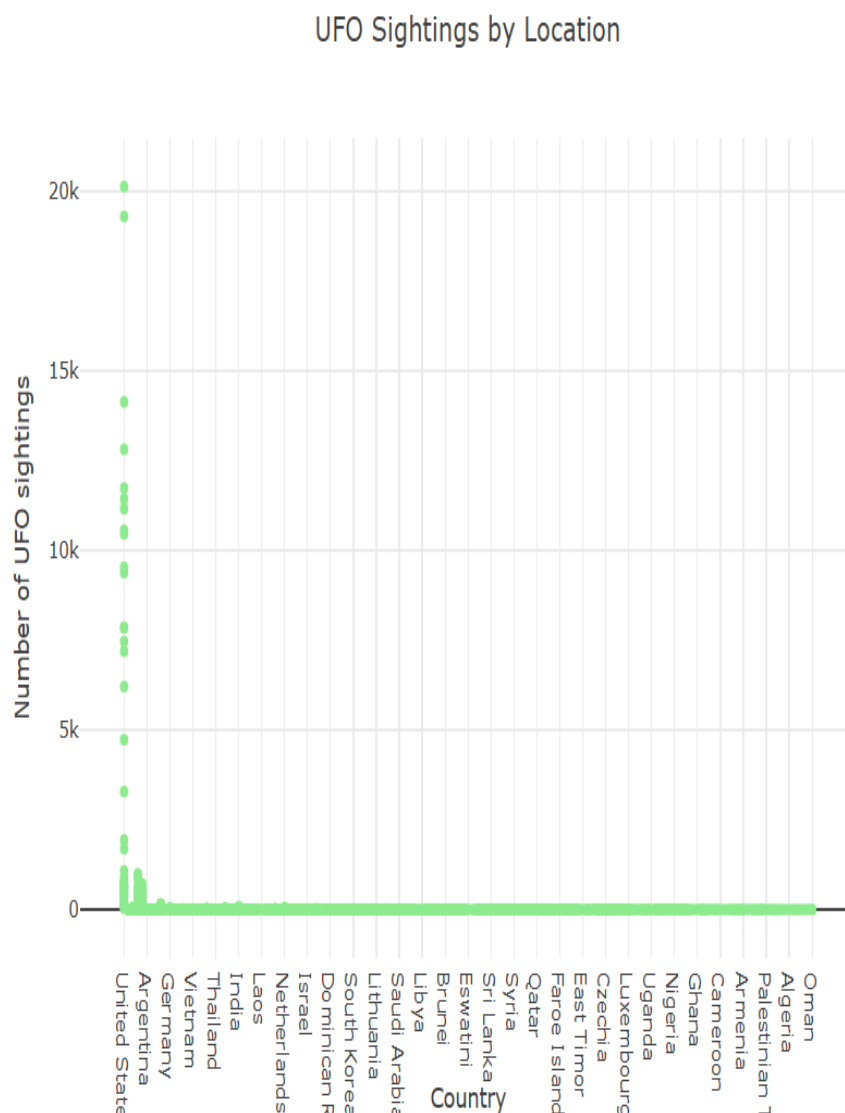
Additionally we also plotted a bar plot for top 10 countries with the most UFO sightings.


Number of UFOSightings for Top 10 Countries
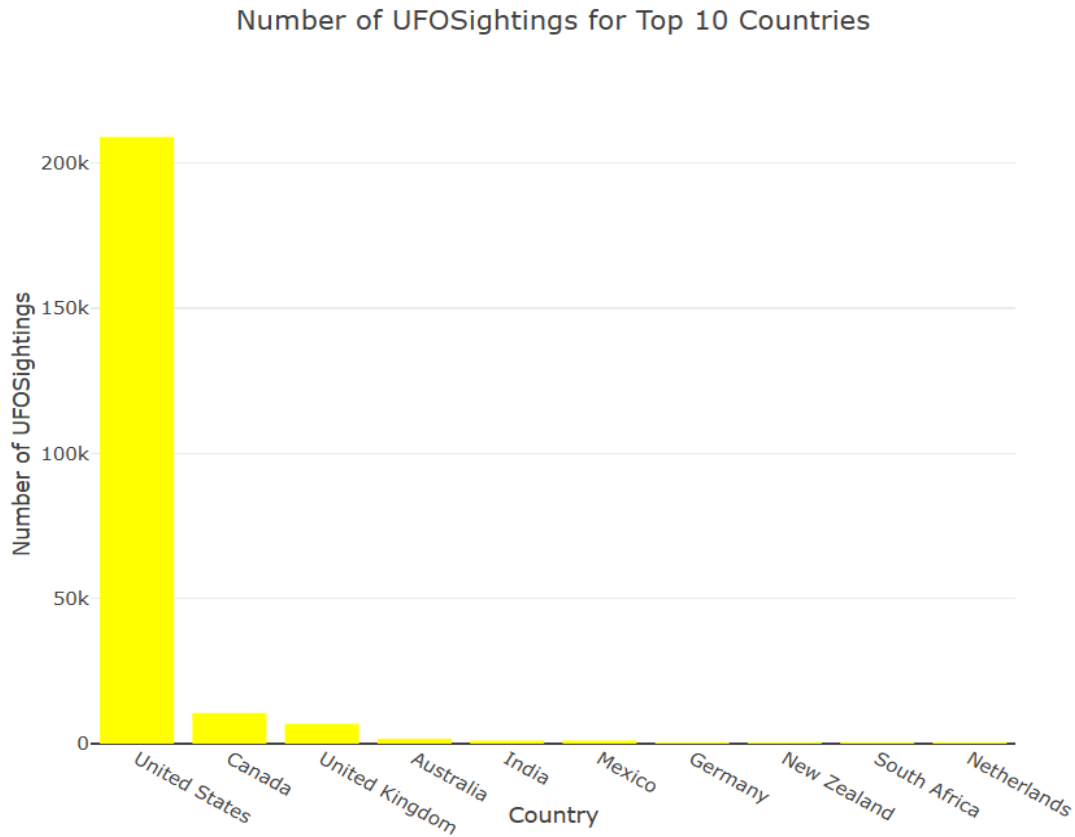
Our heat map also shows a strong concentration of reported UFO sightings in the United States. Compared to other countries, the US is significantly more red, meaning our dataset has a lot more sightings from the US. Or at least, there are more *reported* UFO sightings in the United States.
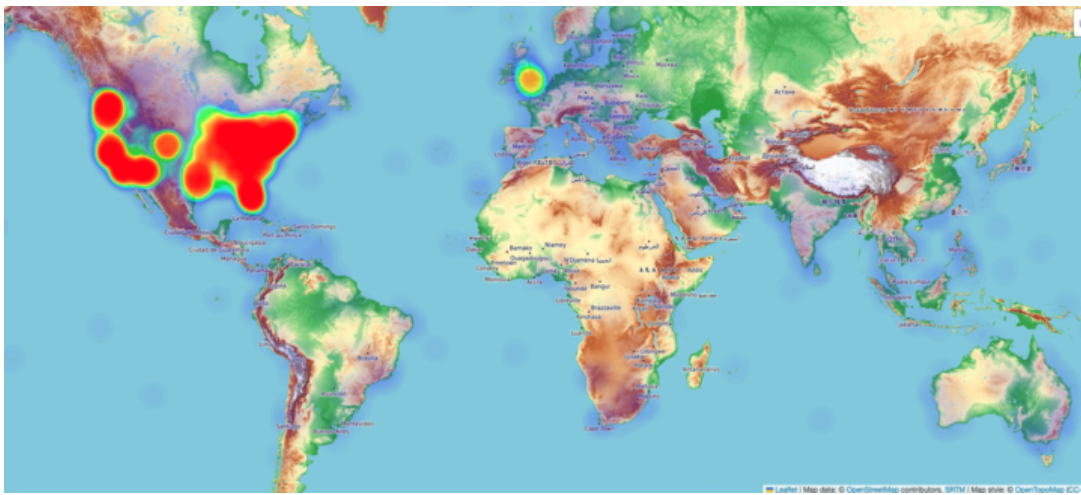
As we see that the United States reports the highest number of UFO sightings. This could be due to a number of reasons. There is a lot of fascination among the people because of the sci-fi movies, television shows, and a wider media coverage. The US military also conducts lots of test flights. These military activities could be mistaken by the public for UFOs. The US has also established a lot of platforms for the reporting of UFO sightings. Plus, our heat map shows increased sightings over heavily populated coastal areas. This too could be contributing to an increased number of UFO sighting documentation.

**Do UFO sightings change in terms of time of day and duration?**

We plotted a bubble chart to answer the question of the relationship between time of day and duration of UFO sighting. When looking at our overall data, there seems to be no correlation between the duration of the sightings and time of day. The duration varies greatly with the time of day. You will notice however, that duration length does increase as time goes on. This is shown in the figure below.

UFO Sightings by Year
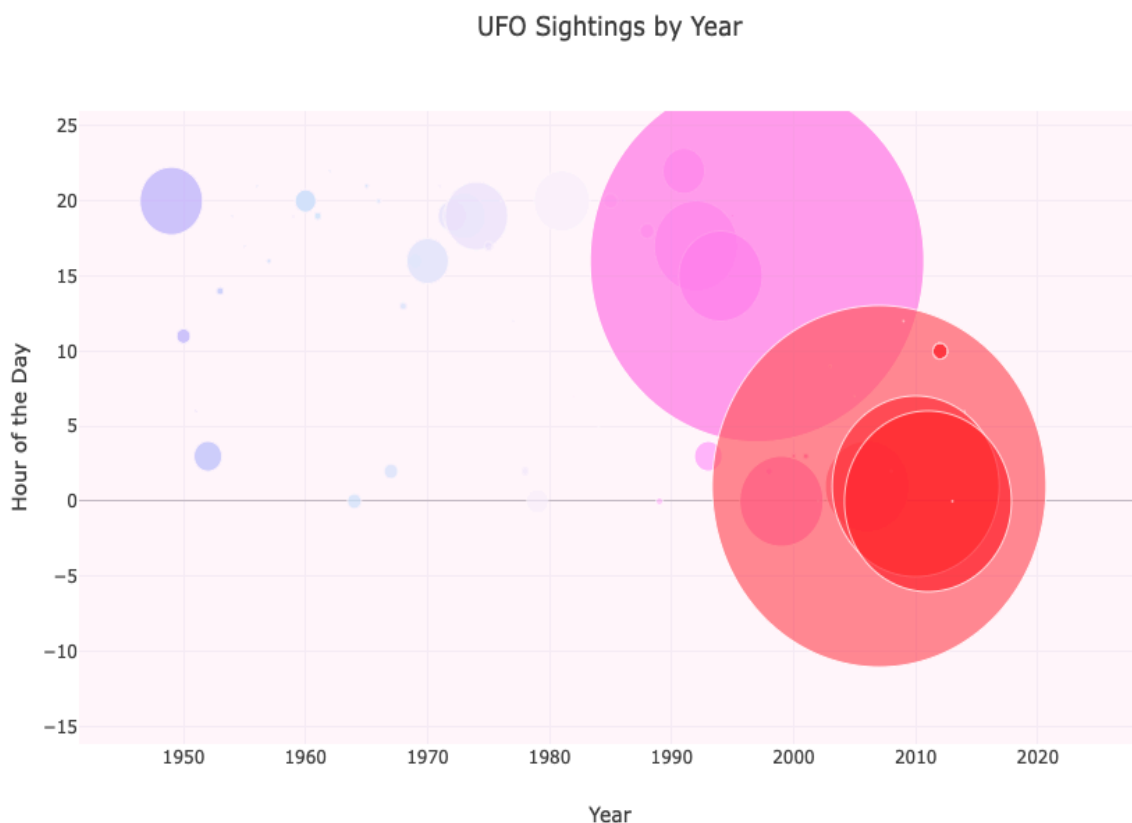
While we don't see a correlation between time of day and duration length, we do see an increase in the encounter time duration after 2000. If we look at the data from 1949 to 1990 we see that the encounter duration is not that much. Once again, this could be due to the increasing amount of video and photography technology.

Donut Chart showing the number of UFO sightings across different seasons.

A donut chart displaying the percentage of UFO sightings across different seasons showed a relatively balanced distribution. However, sightings were slightly higher in summer (32.8%), likely due to increased outdoor activity.
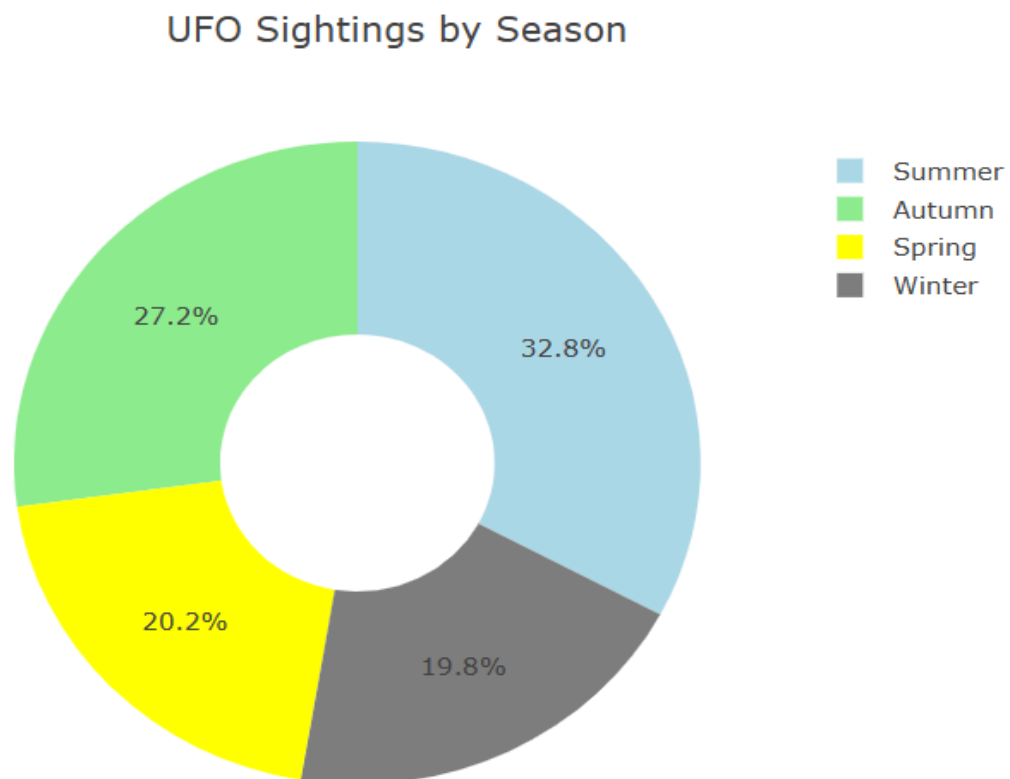
LeaderBoard of UFO sightings across years for season shape country and coordinates.

Year	Season	Country	UFO_shape	Latitude	Longitude
1952	Summer	United States	Circle	41.85	-87.65
1952	Spring	United States	Disk	47.3858333	-122.8263889
1952	Spring	United States	Triangle	33.1383333	-96.1105556
1952	Spring	United States	Disk	41.85	-87.65
1952	Winter	United States	Disk	36.241416	-115.050807
1952	Winter	United States	Disk	42.0986111	-75.9183333
1952	Winter	United States	Other	41.85	-87.65
1952	Winter	United States	Cylinder	44.009855	-90.683212
1952	Autumn	Canada	Cylinder	53.55	-113.5
1952	Autumn	United States	Circle	46.2858333	-119.2833333

Marker Clusters on the Sightings Map

Our map is organized with marker clusters. The user can click on a cluster to zoom in and further break them down to specific sightings. As you continue to break down the clusters to their smallest point, we have included some extra information regarding the sighting. Each marker has a pop up menu with the region of the sighting, the year, the hour of the day, and the shape of the UFO. Some sightings on the map are indicated with multiple markers, as shown in the picture below, while other individual sightings have just one marker. This is because, occasionally, UFO encounters are seen and reported by multiple people. You'll find that in less populated areas of the world, there are less clusters and more individual markers.
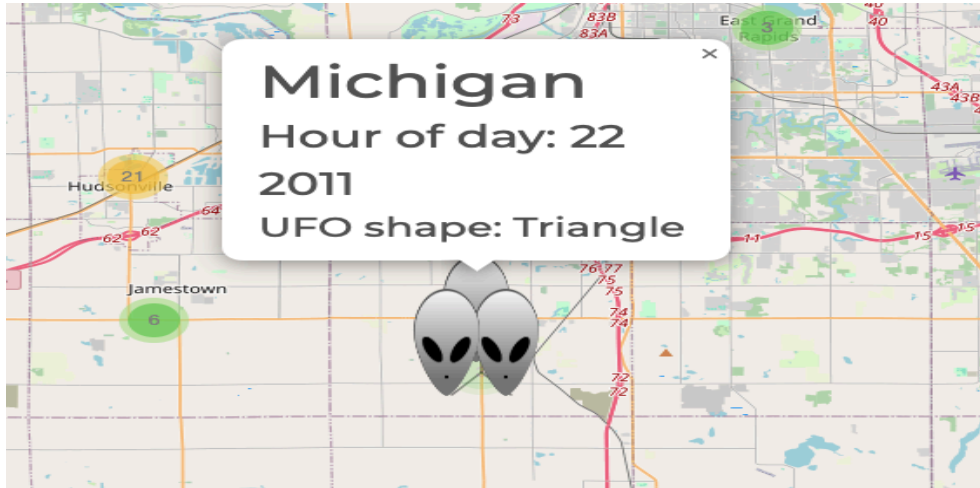
Bias and Limitations

One of the large biases from our data set is the lack of consistency in data collection. For the better understanding of the UFO sightings it is important that we have more data accuracy. This can be achieved only if we collect data through standardized ways.

The data set is only up through the year 2014 which doesn't give us a full picture of UFO sightings current day. With the changes that have arisen since the turn of the century, more up to date data is definitely needed.

The large collection of data from the United States, is itself a bias. Our data might be skewed because of the location of the reporters. Since US citizens have access to reporting sites, it automatically changes the bias of our data.

Call To Action/Future Work

Future research should investigate why UFO sightings are disproportionately high in the U.S. Is it due to cultural factors, military activity, or another unknown reason? Additionally, further study is needed to understand why encounter durations have increased over time. Finally, many reported sightings lack verification, raising the possibility that some are fabricated.

Conclusions

In conclusion, the United States has by far the most UFO sightings. The data show that sightings are most common by the coastlines of America. Lastly, the year in time seems to correlate with the number of UFO reports. The closer to present day, the greater the number of reported sightings and the longer in duration the sightings become.

Works Cited:

Roswell Incident

https://en.wikipedia.org/wiki/Roswell_incident

The History of UFO Sightings Around the World

https://newspaceeconomy.ca/2025/02/04/the-history-of-ufo-sightings-around-the-world

The Black Vault

https://www.theblackvault.com/documentarchive/category/the-fringe/ufo-phenomena/

BBC News

https://www.bbc.com/news/uk-68515515

National UFO Reporting Center

https://nuforc.org/

Chat GPT

https://openai.com/chatgpt/overview/