UNIVERSITY OF CALIFORNIA

Los Angeles

Unveiling Hidden Patterns in UFO Sightings:
A Text Mining and Geostatistical Approach

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Applied Statistics and Data Science

by

Yuxin Chen

2023

ABSTRACT OF THE THESIS

Unveiling Hidden Patterns in UFO Sightings:
A Text Mining and Geostatistical Approach

by

Yuxin Chen
Master of Applied Statistics and Data Science
University of California, Los Angeles, 2023
Professor Frederic R. Paik Schoenberg, Chair

This thesis delves into the enigmatic world of unidentified flying objects (UFOs), investigating the diverse characteristics, societal perceptions, and potential associations with military bases. Focusing on UFO sightings reported within the United States and documented by the National UFO Reporting Center (NUFORC) from 1969 to 2022, this study aims to shed light on the complexities surrounding these intriguing phenomena using statistical methods including Exploratory Data Analysis, Text Mining, Sentiment Analysis, and Geostatistical techniques.

The thesis of Yuxin Chen is approved.

David Anthony Zes

Hongquan Xu

Yingnian Wu

Frederic R. Paik Schoenberg, Committee Chair

University of California, Los Angeles

2023

To Mom and Dad

For their endless love and support.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

Introduction

Unidentified flying objects (UFOs) are objects or lights observed mostly in the sky that cannot be immediately explained or identified as conventional aircraft, balloons, drones, or other known objects [Wik23h]. There are several factors that can affect the observation or detection of UFOs, including weather, illumination, atmospheric effects, or the accurate interpretation of sensor data. While most UFOs are later identified as known objects or atmospheric phenomena, such as weather balloons, birds, airborne plastic bags, and reflections of light, the others with unremarkable characteristics remain unexplained and continue to intrigue and fascinate people around the world.

Numerous UFO sightings have been reported over the years, with descriptions ranging from simple lights to more elaborate descriptions of craft with distinct shapes and movements. The study of UFO sightings, also known as ufology, has garnered considerable interest from the public, scientists, and researchers who strive to comprehend the phenomenon and its potential implications.

However, studying UFO sightings presents challenges due to the difficulty in obtaining reliable and verifiable data. Many reports come from witnesses who may lack the expertise or resources to accurately identify or document their observations. Additionally, the absence of standardized reporting procedures or criteria for evaluating and classifying UFO sightings hampers the comparison and analysis of data from different sources. Fortunately, the National UFO Reporting Center (NUFORC), established in 1974, serves as an invaluable resource for collecting and sharing data on UFO sightings. As a non-governmental, non-profit organization, NUFORC operates a publicly-accessible database, enabling individuals to document and share their sightings with the public. With its extensive collection of reports, NUFORC becomes an essential tool for researchers and enthusiasts seeking to understand and analyze patterns and trends in UFO sightings.

It is important to note that studying and tracking UFOs is not necessarily about finding evidence of extraterrestrial life or advanced technology from another planet. Rather, it represents an effort to investigate and analyze phenomena and events that cannot be easily explained or identified through conventional means. By exploring various hypotheses and conducting research, scientists and researchers can develop a deeper understanding of the nature and potential causes of these unexplained phenomena. Similarly, this study does not seek to prove or disprove the existence of extraterrestrial life, but rather to gain a deeper understanding of the world around us and the possibilities that lie beyond our current knowledge.

Furthermore, some theories propose that UFO sightings are more likely to occur near military bases due to the possibility of experimental military technology or secret military operations that may be mistaken for UFOs. Others propose that UFOs near U.S. military bases are sometimes high-tech aircraft assigned by adversaries to conduct espionage activities and collect secret information [Bar23]. However, there is no concrete evidence to support these theories, and it is important to approach such claims with a critical and evidence-based perspective. Therefore, this study also aims to explore the potential relationship between UFO sightings and military bases in the United States, utilizing empirical evidence and statistical analysis to gain insights into this phenomenon.

Specifically, the research questions addressed in this study are as follows:

- Variation in Reported Properties: How do the reported properties of UFOs vary across different locations and over time?
- Variation in Attitudes: How do people's attitudes towards UFO sightings vary across

different locations and over time?

- Impact of Proximity to Military Bases: What is the significant difference between UFO sightings that occurred near U.S. military bases and those that occurred away from U.S. military bases?

Furthermore, the final part of this study incorporates case studies of three well-known UFO incidents to seek supporting evidence within the dataset.

CHAPTER 2

Data

The "UFO Sightings" dataset, provided by Tim Renner and obtained from data.world, is the primary dataset used for this study [Ren23]. The observations in this dataset were initially scraped from the National UFO Research Center (NUFORC), which maintains a publicly-accessible database of reports on a wide variety of UFO sightings as mentioned in the introduction. Therefore, this dataset provides an excellent opportunity to study the patterns and trends of UFO sightings in the United States over the years.

The latest version of the dataset available for this study is dated December 2022, which contains over 140,000 entries, each representing a reported sighting of an unidentified flying object, and 14 variables as follows:

- Summary: A summary of the report, usually the first few sentences.
- Country: The country of the sighting. This study will only focus on sightings in the United States of America.
- City: The city of the sighting.
- State: The state of the sighting.
- Date_time: The date and local time of the sighting in ISO 8601 format.
- Shape: The shape of the sighting.
- Duration: The duration of the sighting in text form.

- Stats: Summary stats of the sighting including when it occurred, when it was reported, etc.

- Report_link: A link to the original report on the NUFORC site.

- Text: The text of the original report.

- Posted: The time when the sighting was reported to the NUFORC site.

- City_latitude: The latitude of the nearest city in which the sighting occurred.

- City_longitude: The longitude of the nearest city in which the sighting occurred.

- City_location: The geocoded location of the nearest city in which the sighting occurred.

As the study focuses solely on the sightings reported in the United States of America, the dataset was filtered down to a subset of about 100,000 complete reports that meet the criteria. Due to the scope of the study and specific research interests, it is not feasible or necessary to analyze all of the variables available in the dataset. As such, the study chooses to focus on a subset of these variables that are deemed most relevant to the research questions at hand, which will be addressed in detail later.

Additionally, the study also makes extensions to obtain additional information that could provide further insights into the data. This may involve creating new variables, combining variables, or using external data sources to enrich the analysis. By focusing on a select set of variables while also seeking to expand upon them, the study is expected to strike a balance between depth and breadth of analysis, and ultimately provide a more comprehensive understanding of the phenomena.

CHAPTER 3

Methodology

3.1 Exploratory Data Analysis

This study only focuses on a selected set of variables in the “UFO sightings” dataset, including “date_time,” “state,” “shape,” “text,” “city_latitude,” and “city_longitude.”

First of all, the “date_time” variable in ISO 8601 format is broken down into two separate variables: “date” and “hour.” The “date” variable represents the day the sighting occurred and is represented as “YYYY-MM-DD” format, while the “hour” variable represents the hour of the day the sighting occurred and is represented as an integer from 0 to 24.

Additionally, “date” is further rearranged into “year” and “decade” in numerical form. Similarly, “hour” is further split into four time groups with 5am to 9am labeled as “morning,” 10am to 1pm labeled as “late morning,” 2pm to 7pm labeled as “afternoon,” and the remaining period of 8pm to 4am labeled as “night.” This categorization allows for a more focused examination of sighting occurrences during different parts of the day and helps to identify potential temporal patterns.

The “UFO sightings” dataset includes sightings reported from 1969 to 2022, with the majority of sightings occurring in the 2010s, accounting for 48.73% of the total reports, followed by 30.41% in the 2000s, and 9.99% in the 2020s, as shown in Figure 3.1. However, only a limited number of reports were documented before the 2000s due to several factors. One of the possible factors is the limitation of technology. Before the widespread use of smartphones and other digital devices, it was difficult for people to capture images or videos of UFO sightings. Additionally, the lack of public awareness and interest in the field of UFOs might have discouraged individuals from reporting their sightings or discussing them openly. Furthermore, there may have been more sightings in the past that were not properly documented due to the lack of record-keeping systems or the destruction of records over time. For instance, the U.S. Air Force's Project Blue Book, which investigated UFO sightings from 1947 to 1969, was criticized for its lack of rigorous investigation and record-keeping practices [Wik23g]. Last but not least, the social stigma surrounding UFO sightings may have deterred people from reporting their experiences. In the past, people might have been reluctant to report sightings for fear of being ridiculed or disrespected.

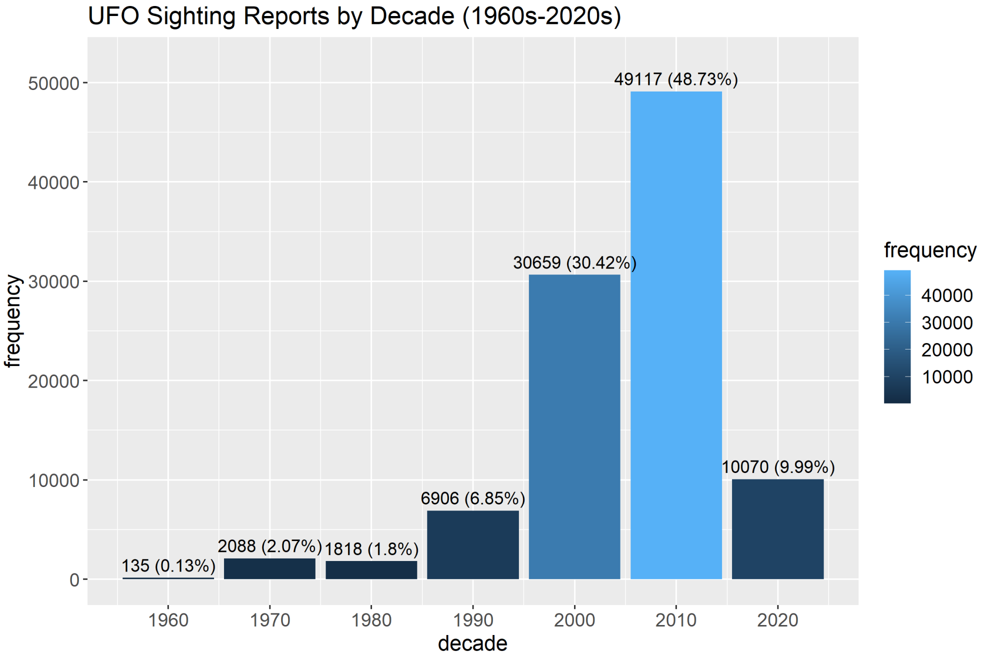


Figure 3.1: UFO Sighting Reports by Decade

After the 2000s, with the U.S. government and media releasing information regarding UFO sightings, along with more openness and acceptance, the topic has been brought to the forefront of public discourse, leading to an increase in reports. Specifically, according to Figure 3.2, the number of sighting reports peaked in the years 2012 and 2014, gradually decreased until hitting a low in 2018, then increased again and reached another peak in 2020. Although the number of reports does not follow a continuous growth in recent years, more records are expected to be added in the future as the study of UFOs continues to develop.

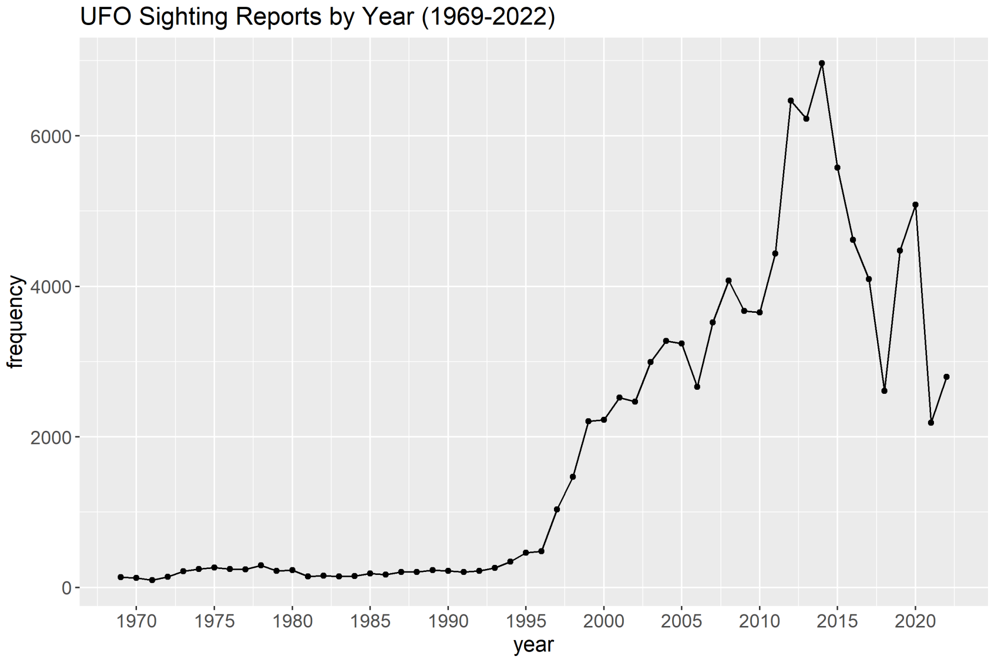


Figure 3.2: UFO Sighting Reports by Year

Regarding the time of day, it is observed in Figure 3.3 that the majority of the UFO sightings occurred at night, with 64.75% of reports taking place during 8pm to 4am. In contrast, only 21.39% occurred in the afternoon, and 7.88% and 6% occurred in the morning and late morning, respectively.

Upon examining the relationship between time and year, Figure 3.4 shows that the number of UFO sightings occurring at night peaked in the years 2012, 2014, and 2020. This result matches with the general pattern, where the overall number of sightings also peaked in those years. Specifically, 70.08% of sightings occurred at night in 2012, while 71.02% of sightings occurred at night in 2014, both of which represent an increase of more than 5% compared to the average proportion of sightings occurring at night.

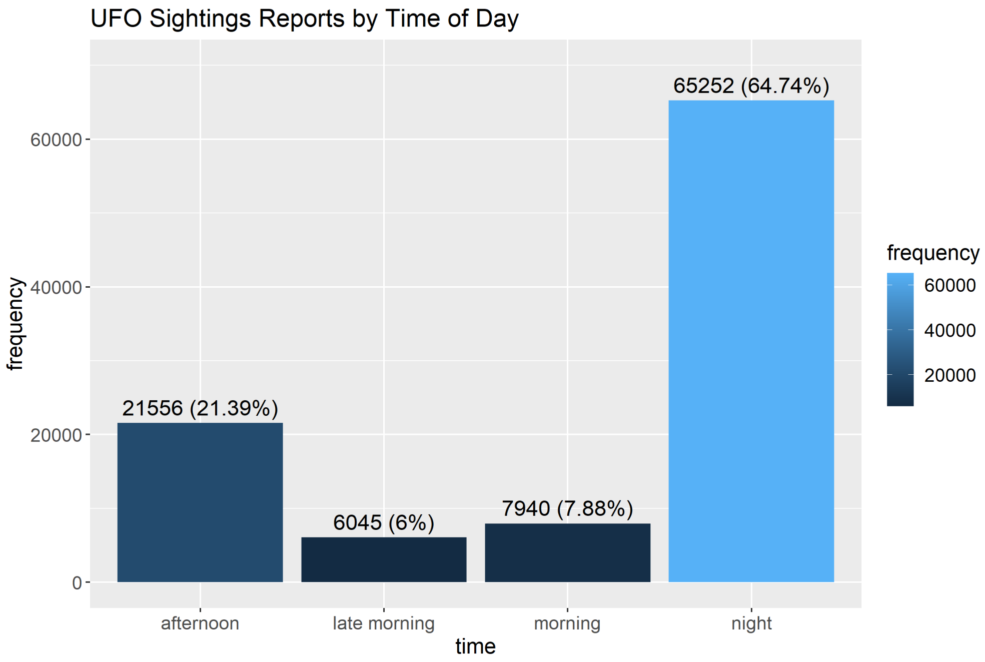


Figure 3.3: UFO Sighting Reports by Time of Day

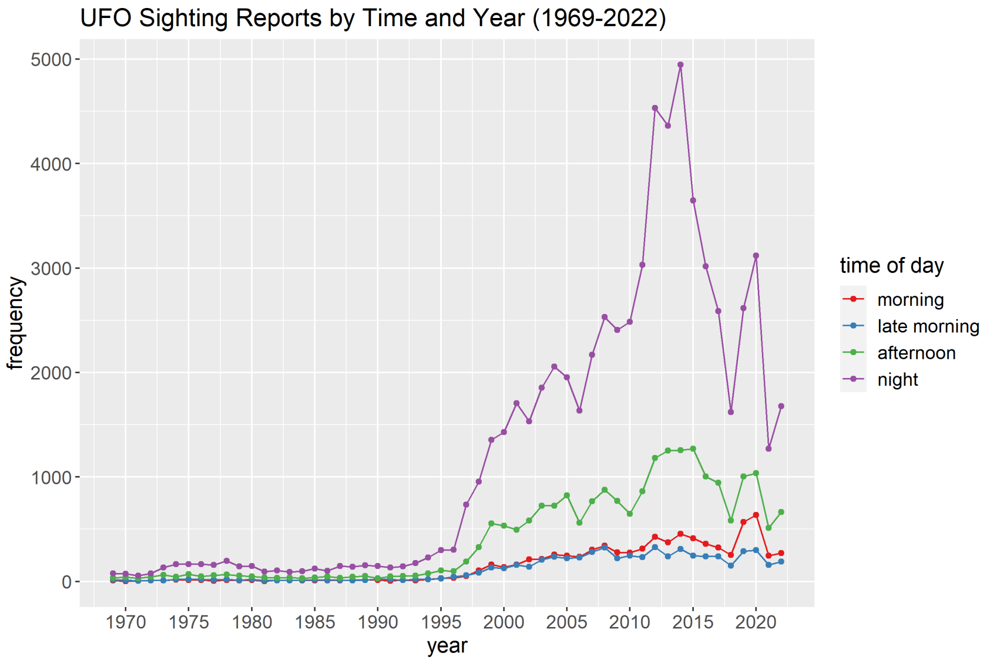


Figure 3.4: UFO Sighting Reports by Time and Year

Second, when examining the “state” variable, it is discovered that few provinces of Canada in the dataset were mistakenly classified as states of the United States. After correcting this error, Figure 3.5 indicates that California has the dominantly highest number of UFO sightings at 12.86% of the total reports, followed by Florida at 6.32%, and Washington at 5.4%.

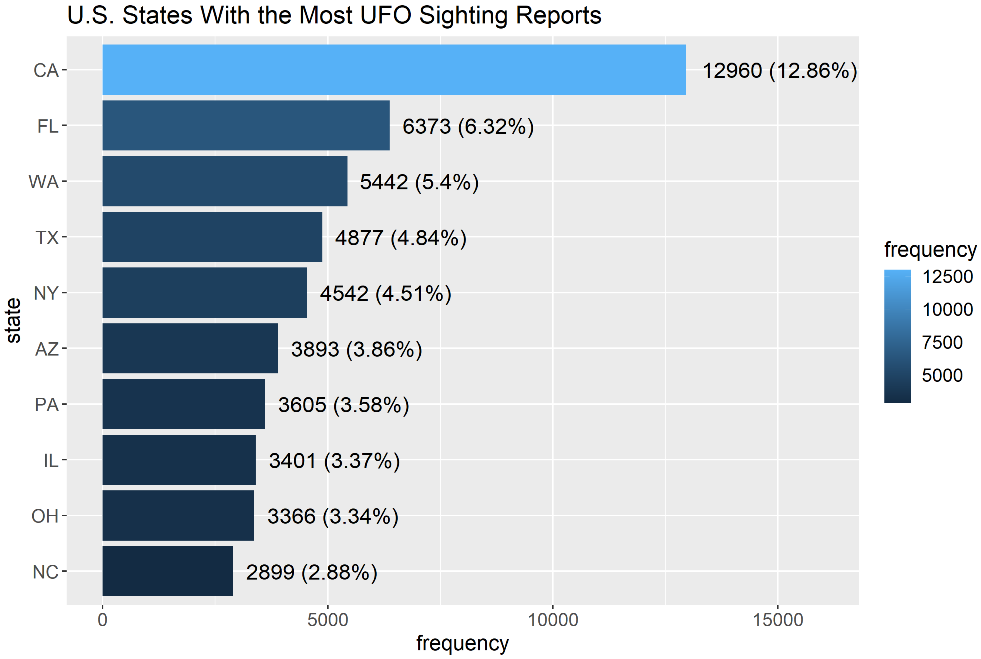


Figure 3.5: U.S. States With the Most UFO Sighting Reports

Additionally, to provide a more comprehensive picture of UFO sightings across the United States, the 50 states (along with Washington, D.C.) are categorized into four statistical regions according to the United States Census Bureau: Northeast, Midwest, South, and West [Wik23c]. This categorization promotes a more equitable distribution of sightings, therefore allows for a balanced comparison between different regions.

- Northeast (9 states): Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont, New Jersey, New York, and Pennsylvania.

- Midwest (12 states): Illinois, Indiana, Michigan, Ohio, Wisconsin, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota.
- South (17 states and DC): Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, Washington, D.C., West Virginia, Alabama, Kentucky, Mississippi, Tennessee, Arkansas, Louisiana, Oklahoma, and Texas.
- West (13 states): Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, Alaska, California, Hawaii, Oregon, and Washington.

As a result, Figure 3.6 reveals that the West and South regions have the highest number of sightings, accounting for 33.89% and 29.94%, respectively, while the Midwest and Northeast regions have a lower number of sightings, accounting for 19.98% and 16.19%, respectively.

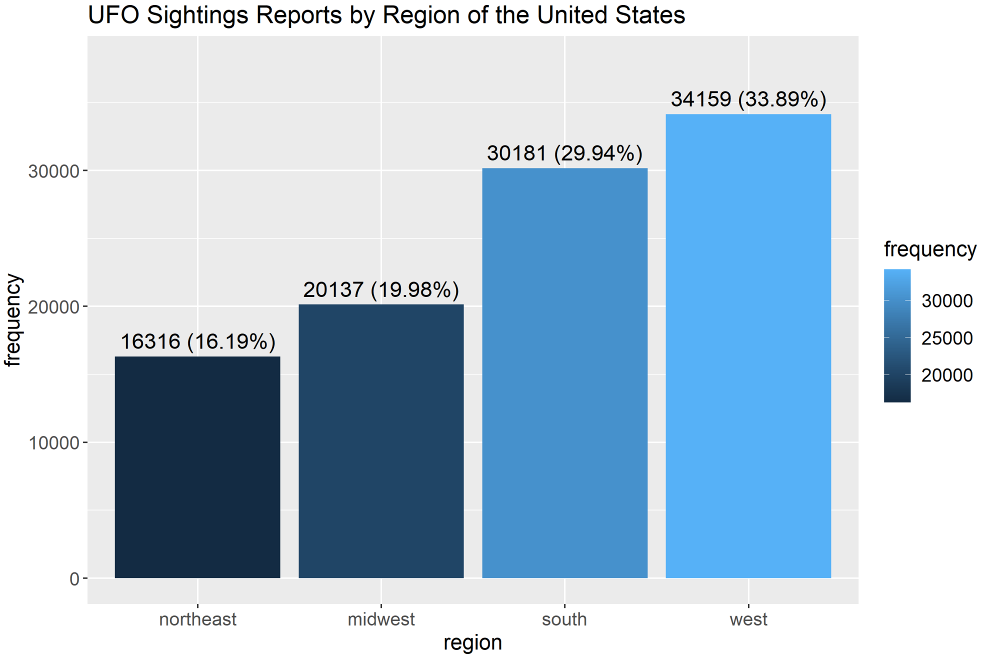


Figure 3.6: UFO Sightings Reports by Region of the United States

Third, the original "shape" variable contains 23 different varieties, some of which are duplicate and not suitable for data analysis. Therefore, to make the data more manageable, the shapes are classified into 10 groups, referred to the common shapes of UFOs [Dee21], as follows:

- Light: includes light, flash, and star (shooting star).
- Circle: includes circle, disk, sphere, and fireball.
- Oval: includes oval and egg.
- Triangle: includes triangle, cone, and delta.
- Chevron: refers to chevron.
- Cigar: includes cigar and cylinder.
- Quadrilateral: includes rectangle and diamond.
- Teardrop: refers to teardrop.
- Cross: refers to cross.
- Other: includes other, unknown, changing, and formation.

According to Figure 3.7, the new shape classifications show that “circle” and “light” are the most frequently reported shapes, accounting for 38.72% and 29.47% of sightings, respectively. “Triangle” is the third most reported shape, accounting for 12.66%, followed by “oval” at 6.66%, and “cigar” at 5.22%. The remaining shapes each account for less than 5% of sightings.

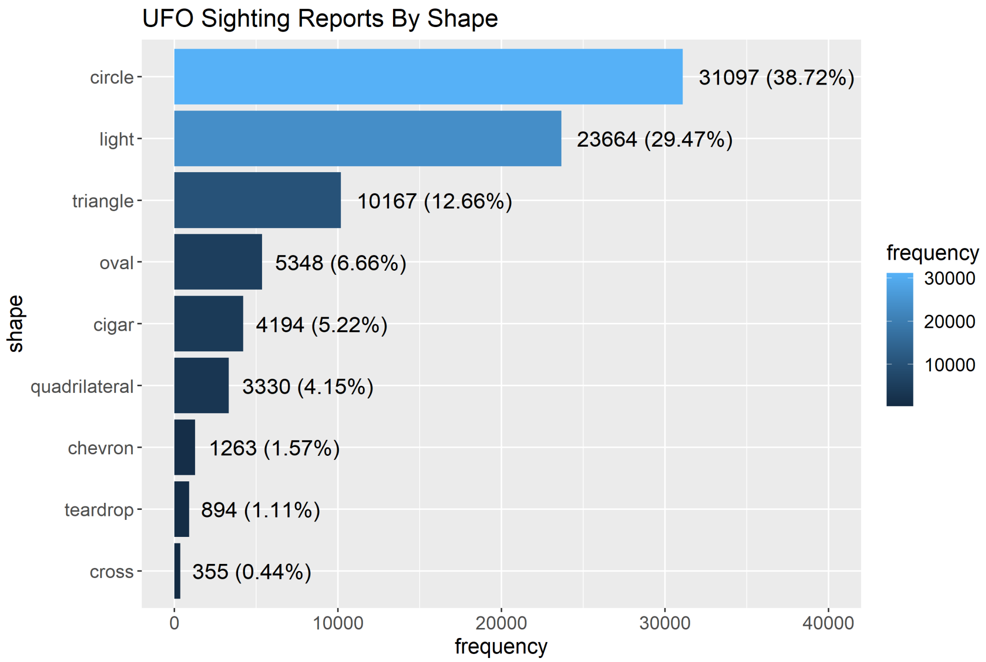


Figure 3.7: UFO Sightings Reports by Shape

Upon examining the relationship between shape and year, Figure 3.8 shows that UFO sighting reports mentioning “circle” and “light” reached peaks in 2012, 2014, and 2020, which again corresponds with the general pattern of the number of reports. However, it is noteworthy that in 2020, the number of reports mentioning “light” exceeded those mentioning “circle,” which is surprising given that, in most cases, “circle” is more frequently mentioned than “light.”

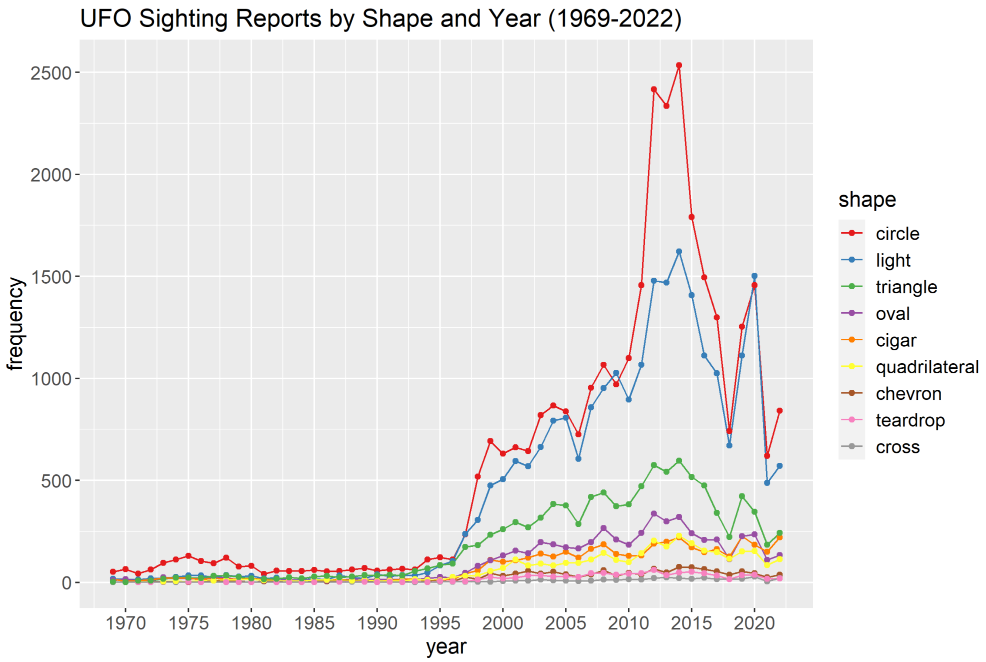


Figure 3.8: UFO Sightings Reports by Shape and Year

So far, the exploratory data analysis has revealed that the majority of UFO sightings in this dataset occurred in the 2010s, with a significant increase in sightings in 2012 and 2014. As to the time when these sightings occurred, most of them happened at night between 8pm and 4am. California is the state with the highest number of sighting reports, while the West is the region with the highest number of reports. Regarding the features of UFOs, the most commonly reported shapes are "circle" and "light."

3.2 Text Mining

3.2.1 Tokenization and Data Cleaning

The information obtained solely from the "date_time," "state," and "shape" variables provides limited explanatory value for a comprehensive analysis of UFO sightings. Therefore, it is necessary to continue exploring the "text" variable, which contains the original UFO sighting reports filed to NUFORC through its online form. These reports can be lengthy and may contain casual language, hence a data cleaning process is performed with the following procedures:

1. Word tokenization: Tokenization breaks down the text into individual words, also known as tokens. This process is important as it creates a consistent format for the text that can be easily analyzed. Additionally, all upper case are converted into lower case to ensure consistency. Punctuations and Unicode characters are also removed to reduce the complexity.

2. Remove numbers: Numbers are removed as they are not relevant in most text analysis tasks and can be a source of noise.

3. Remove English stopwords: Stopwords are commonly used words in a language that do not add significant meaning to the text, such as "the," "an," "a," etc. in English. Removing them helps to reduce the dimensionality of the data and improve the efficiency of the analysis.

4. Text normalization: Stemming is a process of reducing words to their base or root form, such as converting "change," "changing," "changes," and "changed" to their root "chang." This is an essential step in text analysis as it reduces the variations of words with similar meanings, allowing for more accurate grouping and analysis.

After performing the data cleaning process, the resulting data frame of tokens contains over 6.66 million rows and 3 columns, with each row representing a word in its original form, its stemmed form, and its corresponding index in the original text.

3.2.2 Top Frequent Words

Analyzing the most frequent words in the dataset is a crucial step in text mining analysis, as it provides a preliminary understanding of the content of the text data and helps identify any biases or inconsistencies. As listed in Figure 3.9, the most commonly mentioned words in the text include “light,” “object,” “sky,” “move,” and “bright,” which are all closely associated with UFO sightings.

Moreover, the word cloud presented in Figure 3.10 displays the frequent words in reports from different regions of the United States, colored and sized by their frequency. It indicates that the top frequent words in each of the four regions are quite similar. This consistency across regions implies that the data is reliable and representative of the overall sightings in the country.

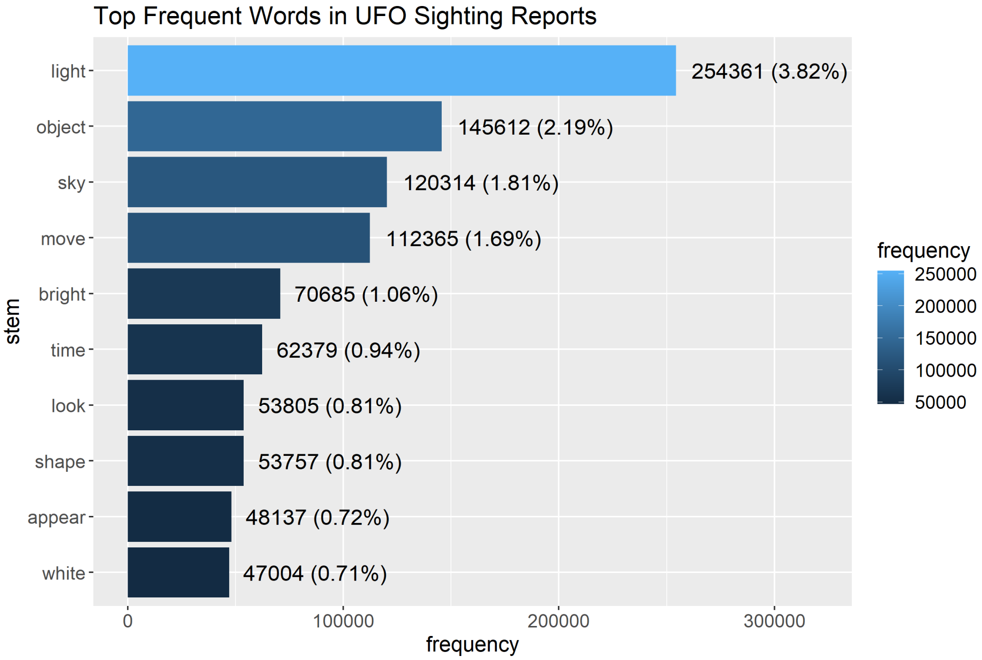


Figure 3.9: Top Frequent Words in UFO Sightings Reports

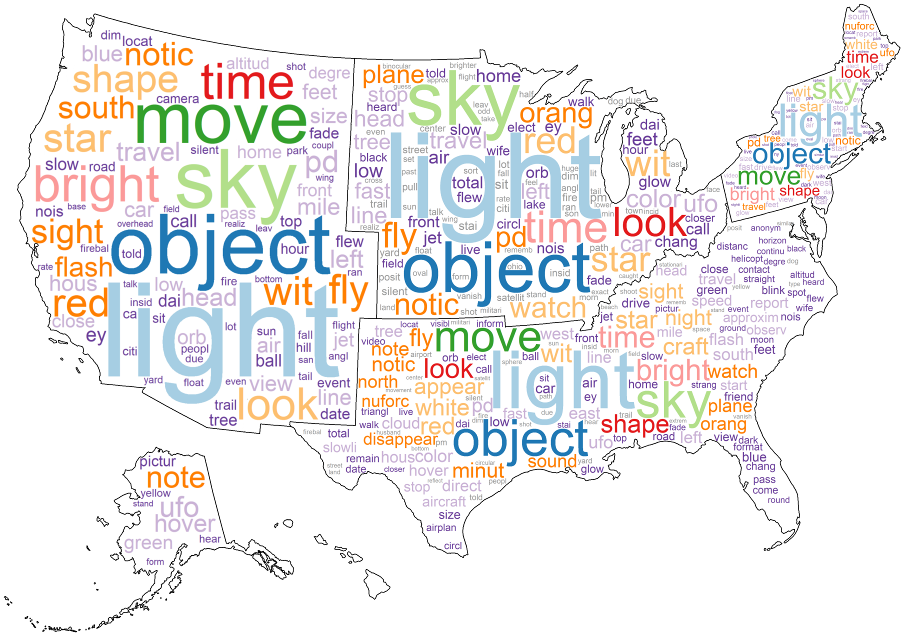


Figure 3.10: Word Cloud of UFO Sightings Reports by Region

3.2.3 Creating New Variables

As the original "UFO sightings" dataset only has a "shape" variable describing the feature of the UFO, a new variable called "color" is created by matching each word (or token) to a list of 28 most commonly used colors [Wik23b]. To simplify the analysis, these 28 colors are further grouped into 11 categories as follows:

- Red: includes red and maroon.
- Orange: includes orange and peach.
- Yellow: includes yellow and gold.
- Green: includes green, lime, olive, turquoise, and teal.
- Blue: includes blue, cyan, navy, and indigo.

- Purple: includes purple and violet.

- Pink: includes pink and magenta.

- Brown: includes brown and tan.

- Gray: includes gray (American English), grey (Commonwealth English), and silver.

- Black: refers to black.

- White: includes white, beige, and ivory.

It is important to note that stemming can occasionally result in non-meaningful words, such as the stem "orang" from the word "orange." Thus, the original forms of words are used for matching to ensure accuracy. Throughout the iterative process, the words (or tokens) from each report are compared against a list of 28 colors to identify intersections. This comparison results in a subset that captures these shared words, in this case, the colors. Each color in the subset is then classified into one of the 11 predefined color categories. The classification outcomes are subsequently recorded in the corresponding row of the "color" columns by its index. For example, if a report mentions the words "gold," "object," "lime," "home," and "navy," the colors "yellow," "green," and "blue" would be recorded for that report. Typically, one or two colors are mentioned in a report, but sometimes more colors are mentioned when describing UFOs with changing lights.

By looking at the frequency of the 11 color categories listed in Figure 3.11, it is observed that "white" and "red" are the most frequently mentioned colors in the UFO sightings reports, accounting for 24.16% and 20.35%, respectively. "Orange" and "blue" followed closely, accounting for 15.43% and 11.17%, respectively. Among the remaining colors, "green," "yellow," "black," and "gray" each account for less than 10%, while "purple," "pink," and "brown" are rarely mentioned, each accounting for less than 1%.

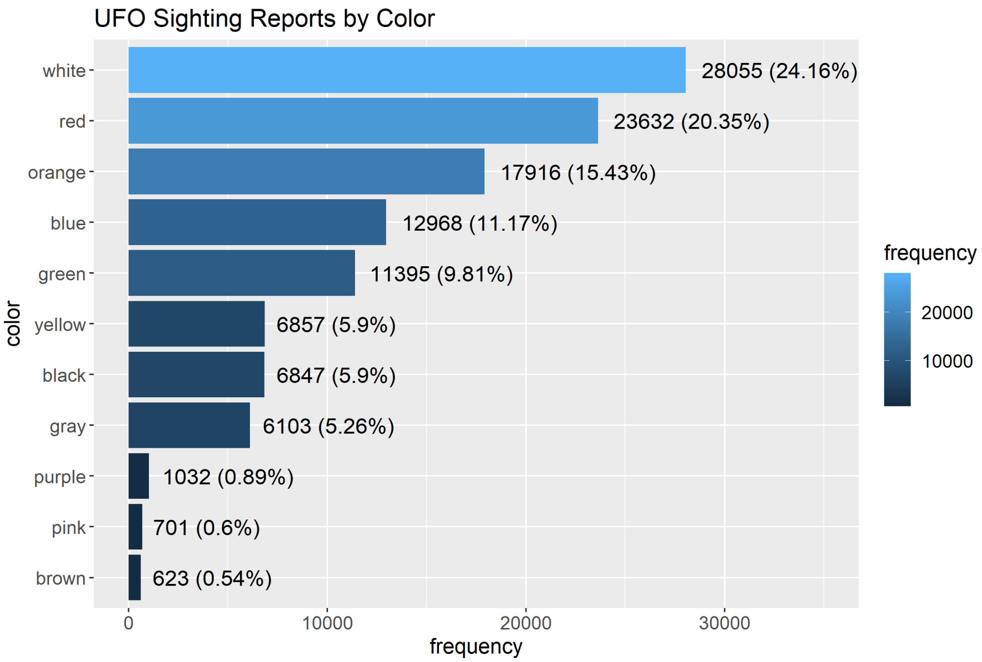


Figure 3.11: UFO Sighting Reports by Color

The following Figure 3.12 illustrates the relationship between the frequency of different color categories and year in UFO sightings reports. “White,” “red,” and “orange” are the most frequently reported colors throughout the years, with “white” being the most common color in most years. However, there were fluctuations across the years. For instance, there was a sudden increase in the frequency of “orange” mentions in 2012, making it the most common color during 2012 to 2013. Notably, in 2014, “red” exceeded “orange,” making it the only year in the 21th century when “red” is the most frequently mentioned color among all 11 color categories.

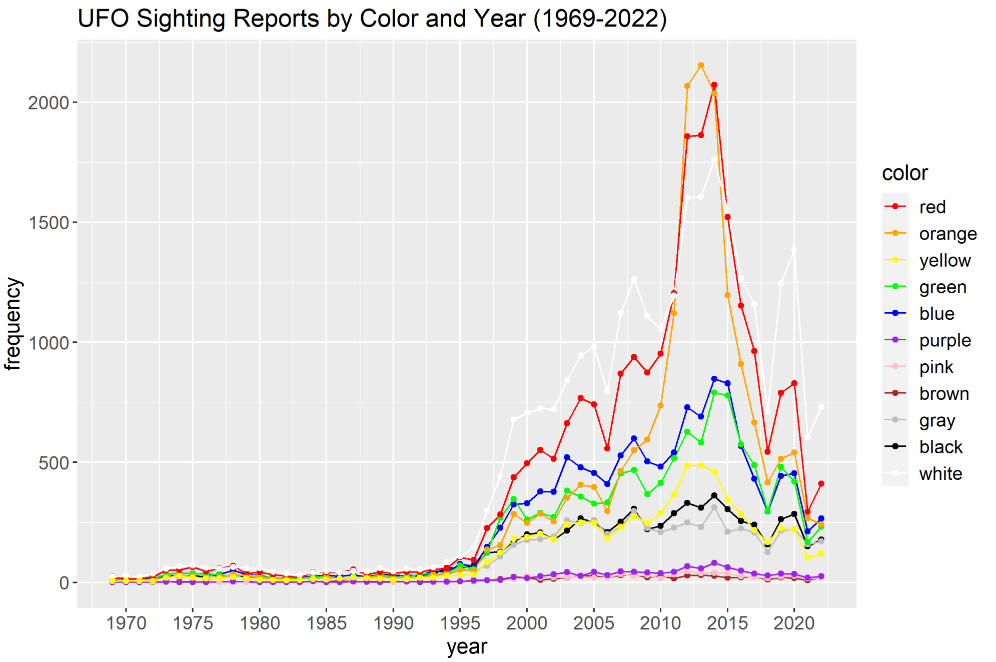


Figure 3.12: UFO Sighting Reports by Color and Year

Furthermore, the “shape” variable with 9 categories (excluding the category of “other”) and “color” variable with 11 categories are merged into a new variable that records the combination of both attributes. For example, if a UFO sighting report describes the shape of UFO as “light” and colors as “green” and “white,” then its “combination” is recorded as both “green light” and “white light.” Instead of further analyzing the “shape” and “color” variables separately, combining them allows identifying possible patterns and relationships between different combinations and other variables.

Figure 3.13 illustrates the top frequent color and shape combinations reported in the four regions of the United States. The most frequently mentioned combination in the Northeast, Midwest, and South regions is “orange circle,” whereas in the West region, it is “white circle.” Other commonly mentioned combinations include “white light,” “red circle,” and “red light.” Overall, the top 10 combinations in each region have relatively similar frequencies, ranging from the highest percentage at around 9% to the lowest at around 3%. One interesting finding is the higher frequency of sightings of UFOs resembling “blue light” in the Western United States than in the others.

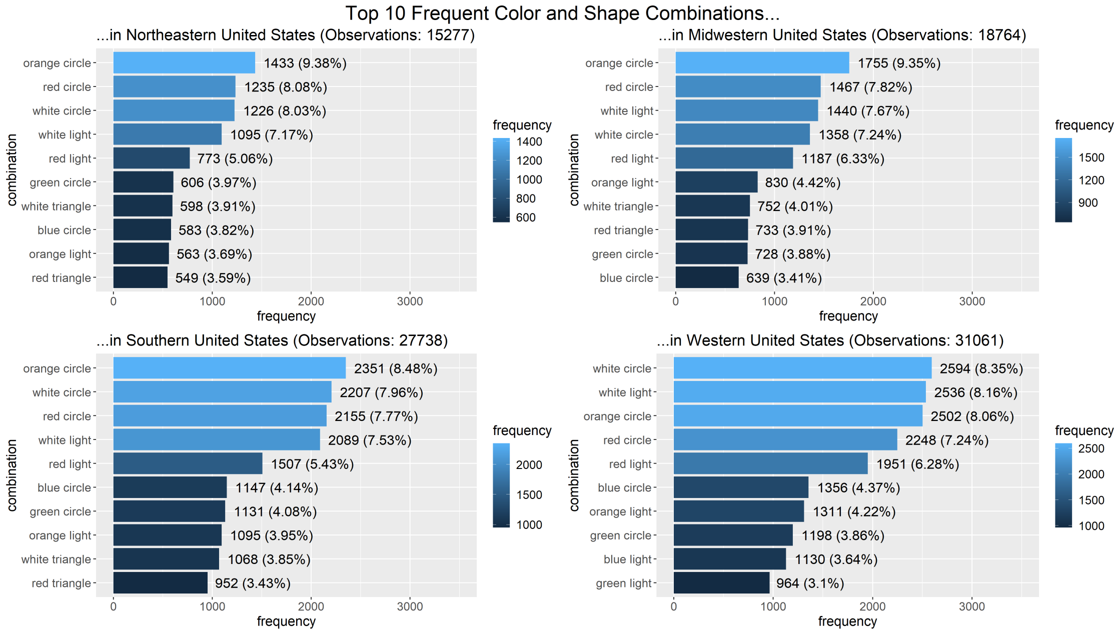


Figure 3.13: Top Ten Color and Shape Combinations in UFO Sighting Reports by Region

In order to further explore patterns and trends over time, it would be valuable to analyze the top frequent combinations by decade. According to Figure 3.13, while there were limited reports before the 2000s, it is noteworthy that “gray circle” was a common combination among reports in the 1970s and 1980s, but became less prevalent from the 1990s onwards. Similarly, sightings of UFOs resembling “black triangle” were ranked among the top 10 frequent combinations in the 1980s and 1990s, which coincidentally corresponds to several documented incidents of “black triangle” UFOs witnessed by large groups of people in New York and Arizona during that period of time [Wik23a]. However, the frequency of such sightings has decreased in the 2020s. Since then, combinations including “white light,” “white circle,” “orange circle,” and “red circle” have remained dominant among the most commonly mentioned combinations.

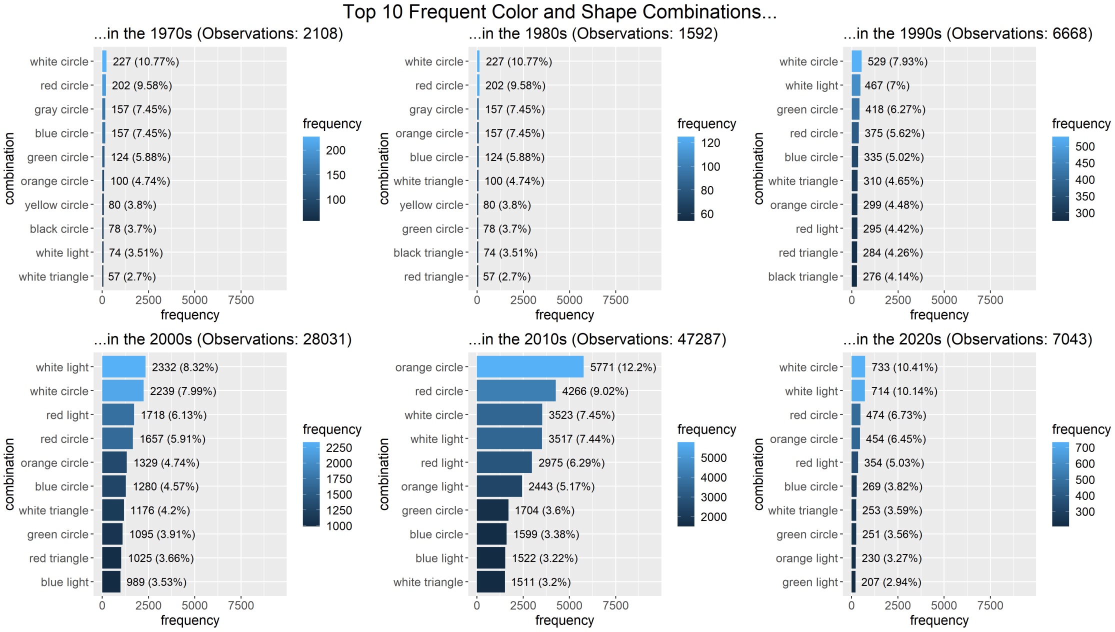


Figure 3.14: Top Ten Color and Shape Combinations in UFO Sighting Reports by Decade

In conclusion, the text mining techniques applied to the UFO sighting reports have yielded several significant findings. The analysis of top frequent words by region of the United States has demonstrated the consistency and reliability of the data. The “color” variable, generated from matching the tokens of text data to a customized list of colors, has enriched the understanding of the features of UFOs and further identifies the most frequently reported color and shape combinations, such as “orange circle” and “white circle.” Moreover, the prevalence of the “black triangle” combination during the 1980s and 1990s is an interesting finding worthy of further research.

3.3 Sentiment and Emotion Analysis

To gain deeper insights into the text of UFO sighting reports, sentiment and emotion analysis are conducted using three general-purpose lexicons: AFINN, bing, and nrc. These lexicons are based on single English words, and each has its own unique approach to sentiment analysis. The AFINN lexicon manually rates words on a scale from -5 to 5, with negative scores indicating negative sentiment and positive scores indicating positive sentiment, while the bing lexicon categorizes words in a binary fashion as positive or negative. The nrc lexicon categorizes words into eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two binary sentiments (negative and positive) [SR22]. It is important to note that not every English word is included in these lexicons, as many words carry a neutral sentiment, and one word may have different interpretations across the three lexicons.

3.3.1 Sentiment Analysis Using AFINN and Bing Lexicons

The AFINN lexicon suggests that the overall sentiment in UFO sighting reports is slightly more positive than negative, with a proportion of 52 to 48%, as shown in Figure 3.15.

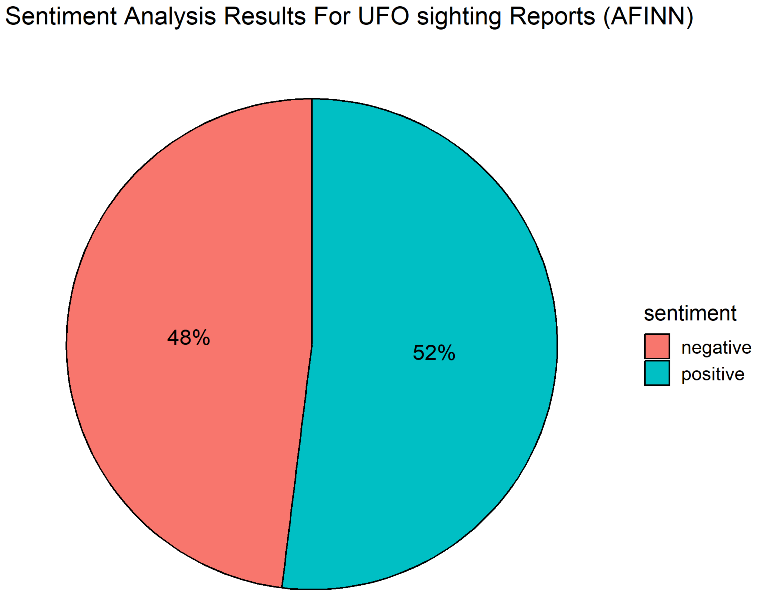


Figure 3.15: Sentiment Analysis Results For UFO sighting Reports (AFINN)

The AFINN lexicon assigns numerical sentiment scores to each word, making it useful for monitoring changes in sentiment over time. The overall sentiment score of a year is calculated by adding up the sentiment scores of individual words included in reports of that specific year. Figure 3.16 reveals a sudden increase in the sentiment score of UFO sighting reports in 1997, indicating a more positive attitude towards UFO sightings during that year. This sudden change in sentiment may be attributed to the widely reported "Phoenix Lights" incident that occurred in March 1997. The event involved multiple sightings of a V-shaped formation of lights in the sky over Arizona and Nevada, witnessed by thousands of people. The U.S. Air Force later released a report concluding that the incident was caused by aircraft participating in Operation Snowbird, a pilot training program of the Air National Guard based in Davis-Monthan Air Force Base in Tucson, Arizona [Wik23f].

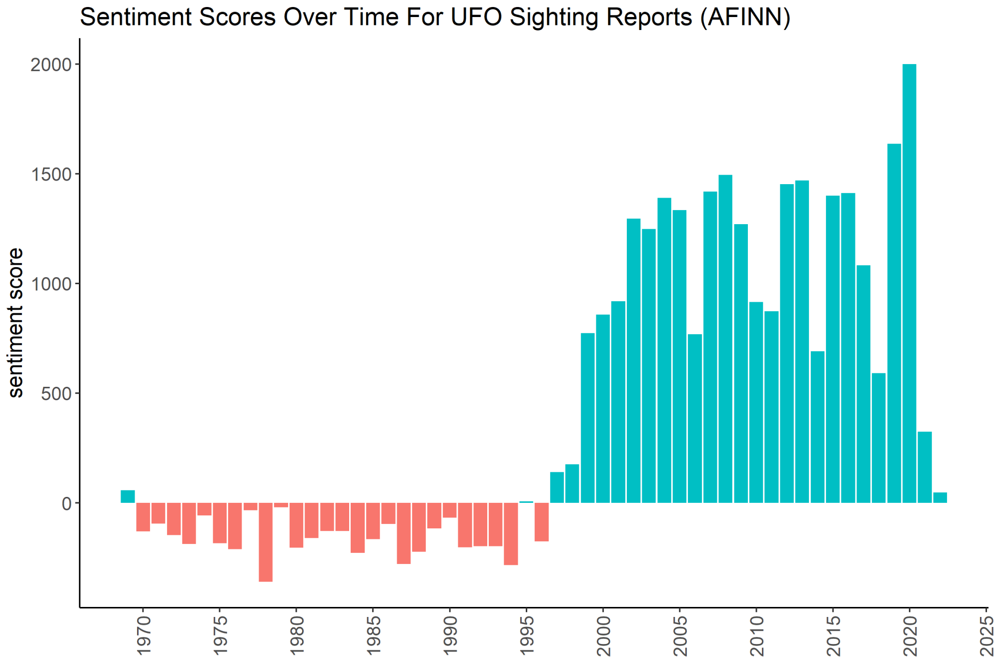


Figure 3.16: Sentiment Scores Over Time For UFO Sighting Reports (AFINN)

However, based on the sentiment analysis using the bing lexicon, Figure 3.17 and 3.18 reveal that the majority of the sentiments expressed in UFO sighting reports are negative, with a proportion of 61% compared to 39% positive sentiments. The most common negative stem words include “object,” “slow,” and “dark,” while the most common positive stem words include “bright,” “fast,” and “glow.”

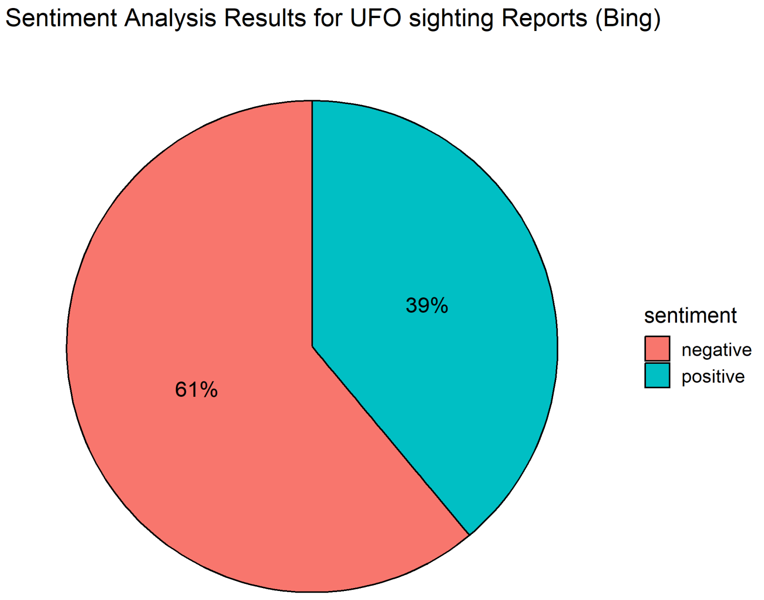


Figure 3.17: Sentiment Analysis Results For UFO sighting Reports (Bing)

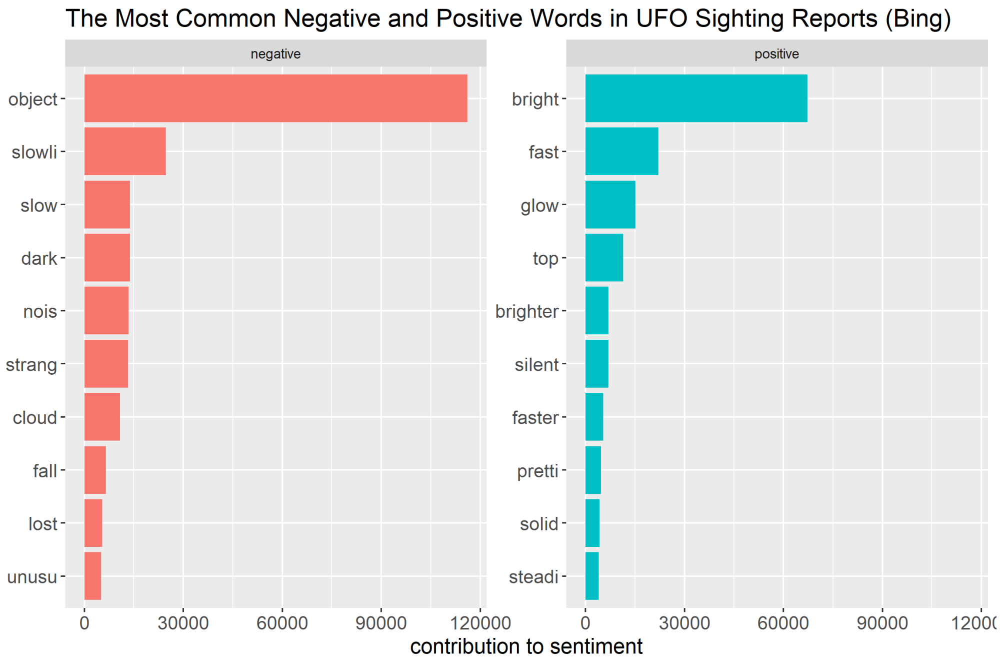


Figure 3.18: The Most Common Negative and Positive Words in UFO Sighting Reports (Bing)

This difference in sentiment results may be attributed to the bing lexicon being based on machine learning algorithms trained on large datasets of text, while the AFINN lexicon is constructed based on a manually rated list of words.

3.3.2 Emotion Analysis Using Nrc Lexicon

Due to conflicting sentiment results obtained from AFINN and Bing lexicons, the analysis of emotions is employed as an additional approach to interpret the sentiments conveyed in UFO sighting reports. The nrc lexicon is used in this analysis, which categorizes words into eight fundamental emotional categories: anger, fear, anticipation, trust, surprise, sadness, joy, and disgust. For example, the word “love” is identified as carrying the emotion of “joy” according to the nrc lexicon.

Figure 3.19 presents a surprising finding that the most commonly expressed emotions in the overall UFO sighting reports are "trust" and "anticipation," accounting for 23.3% and 21.33%, respectively. This may suggest that people who choose to report UFO sightings tend to approach the experience with a sense of curiosity and openness. "Fear" follows closely with 13.03%, indicating a degree of apprehension or uncertainty regarding the sighting. The emotions of "joy," and "sadness" also appear in a significant proportion of reports, accounting for 12.91%, and 10.17%, respectively. In contrast, there is a lower incidence of "surprise," "anger," and "disgust" in reports, suggesting that while UFO sightings may be unexpected, they do not necessarily evoke strong negative emotions in those who report them. This finding raises the possibility that individuals who experience UFO sightings may perceive them as extraordinary, yet not necessarily threatening or harmful events.

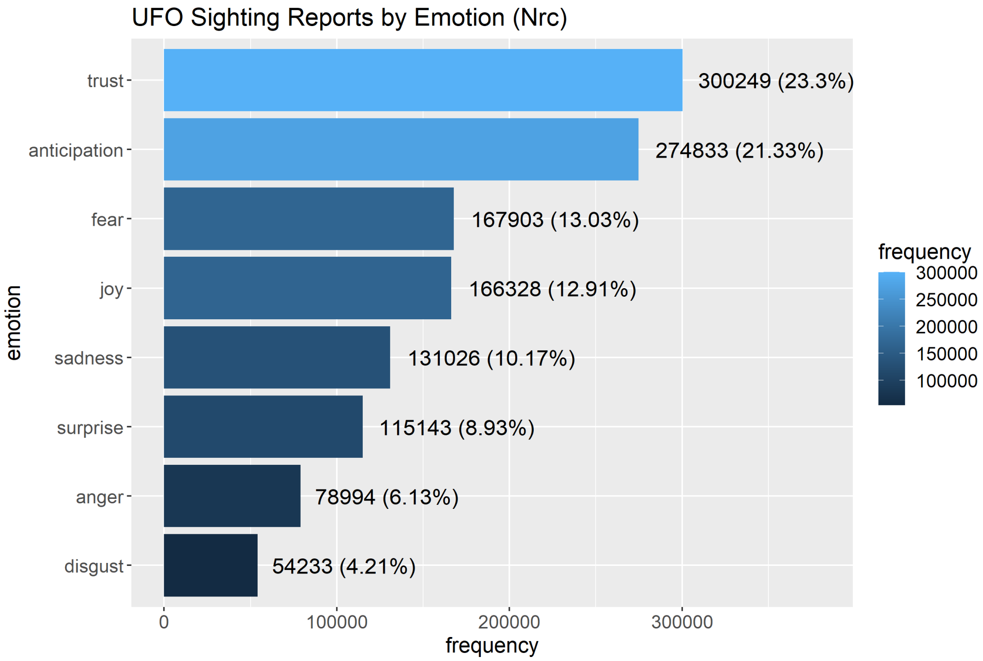


Figure 3.19: UFO Sighting Reports by Emotion (Nrc)

3.3.3 Creating New Variable

A similar method is employed to determine the primary emotion expressed in each of the UFO sighting reports. For example, if a report contains a total of 20 words, and 10 of those words are identified by the nrc lexicon as conveying various emotions, the emotion with the highest frequency is assigned as the primary emotion for that report. In cases where two or more emotions have an equally highest frequency, one of them is randomly selected as the primary emotion to prevent biases. This approach provides a way to categorize the emotional content of individual reports and obtain more detailed insights into the emotional experiences of individuals who have witnessed UFO sightings.

As displayed in Figure 3.20, the frequencies of primary emotions extracted from reports in the four regions of the United States are almost identical. The majority of reports feature "trust" as their primary emotion, accounting for approximately 35%, followed closely by "anticipation" at around 28%. This similarity across regions once again confirms the consistency and reliability of the data. Additionally, this finding provides further support for the notion that individuals who report UFO sightings approach the experience with an open mind, while also anticipating and seeking out the unexplainable.

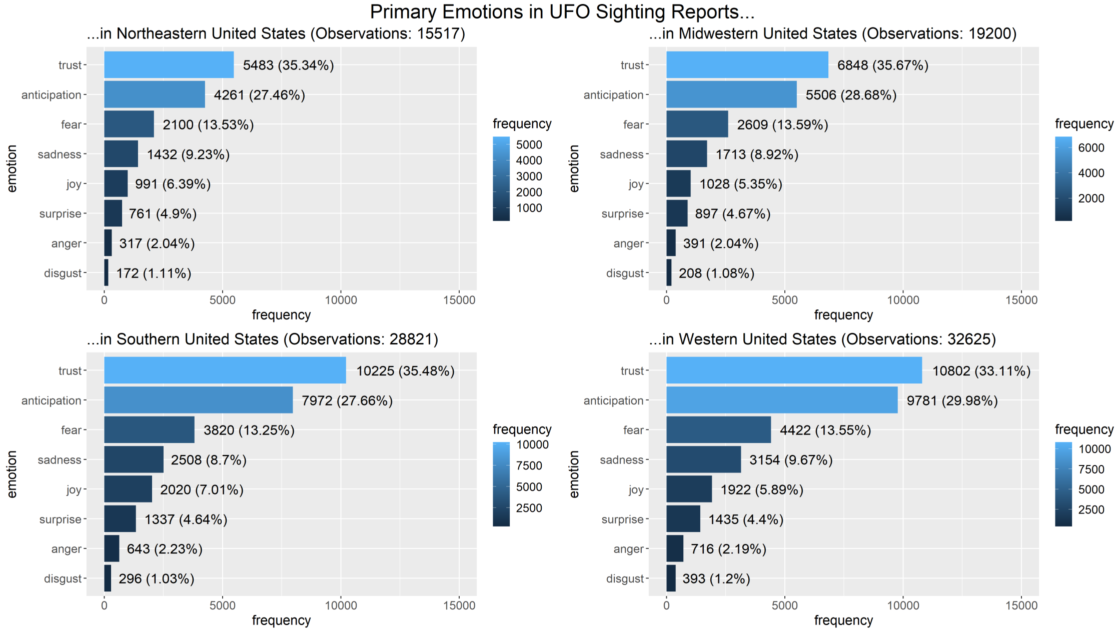


Figure 3.20: Primary Emotions in UFO Sighting Reports by Region

Further examination of the primary emotions in reports across different shapes of UFO sightings and over time, as presented in Figure 3.21 and Figure 3.22 respectively, shows that the frequency of each emotion category remains consistent. This suggests that the emotional response to UFO sightings may not be influenced by the specific feature of the UFO, but rather the overall experience of witnessing an unexplained phenomenon. Moreover, the consistency in emotional responses across regions and over time suggests that the experience of witnessing a UFO may have a universal emotional impact on individuals, regardless of temporal or geographical differences.

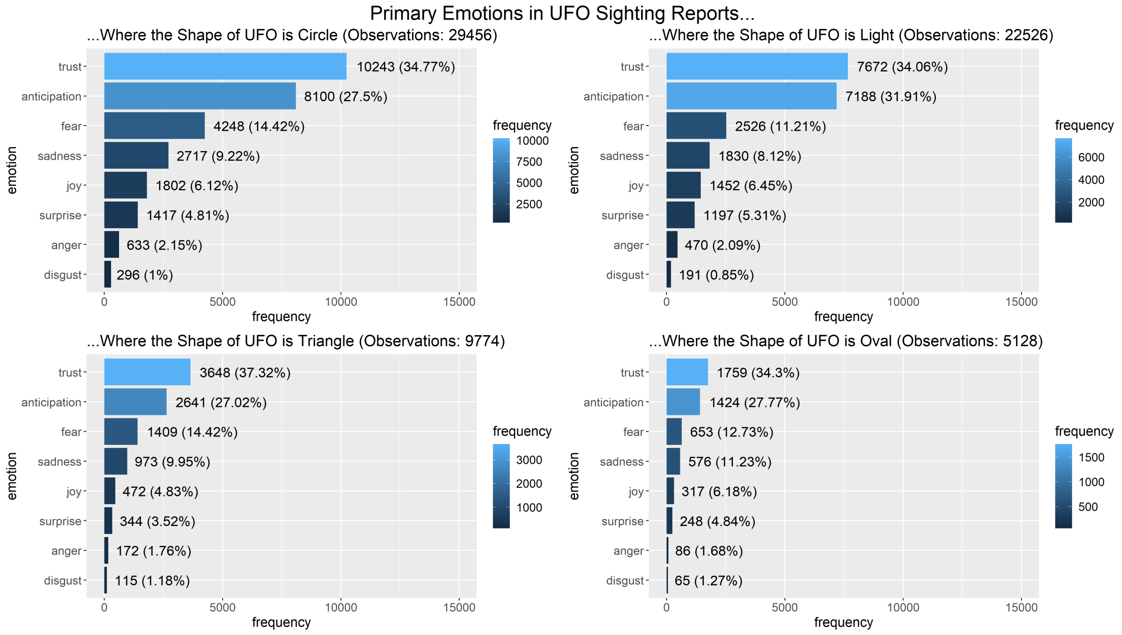


Figure 3.21: Primary Emotions in UFO Sighting Reports by Shape

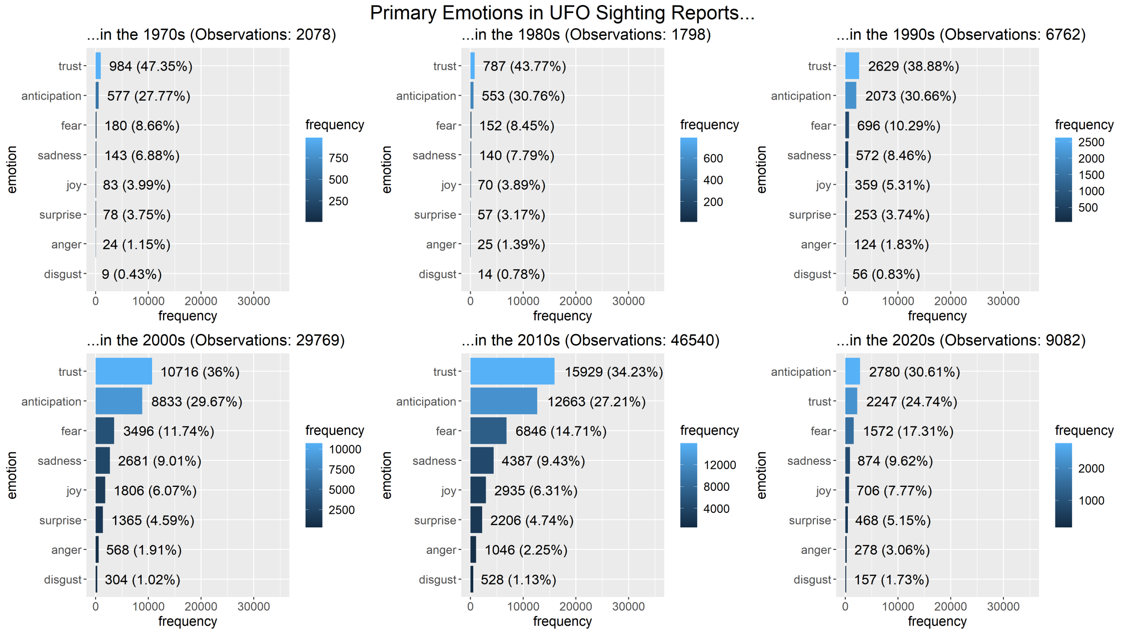


Figure 3.22: Primary Emotions in UFO Sighting Reports by Decade

Overall, the results of sentiment and emotional analysis suggest that, although the experience of witnessing a UFO can elicit a wide range of emotions, most of the individuals who choose to report their sightings typically approach the experience with a positive attitude and a desire for explanations.

3.4 Geostatistical Analysis

Geostatistics is a statistical approach employed to analyze and predict values related to spatial or spatiotemporal phenomena. In this study, several geostatistical techniques are applied to investigate concealed insights regarding the properties of UFO sightings, utilizing the spatial coordinates available in the dataset. These techniques allow for the exploration and extraction of valuable information associated with the spatial distribution and patterns of UFO sightings.

3.4.1 Mapping

To provide a more comprehensive and detailed understanding of the potential geographical patterns of UFOs, this study expands its analysis beyond regional levels and incorporates a mapping of the Contiguous United States at the county level. This visualization allows for a closer examination of the distribution of UFO sightings.

It is important to acknowledge that areas with larger populations tend to report a higher number of UFO sightings. However, it does not necessarily indicate a greater frequency of UFO occurrences in those populated areas. The greater number of reported sightings is primarily attributed to the larger pool of potential observers available to witness and report such incidents. Therefore, when investigating the geographical information of UFO sightings, it is crucial to consider the influence of population density to avoid misinterpreting the data.

The population (from the 2020 census) and area data of 3,243 counties and county equivalents of the United States are derived from Wikipedia to calculate the population density in square miles for each county [Wik23d]. The population density dataset is generated by web scrapping the Wikipedia page with the help of a web tool developed by Gregor Weichbrodt to convert Wikipedia tables to Comma Separated Values (CSV) file [Wei22].

On average, the population density of Contiguous U.S. counties is approximately 244.73 individuals per square mile. Figure 3.23 presents a population density map in each county of the Contiguous U.S., with population density represented by different shades of blue. The most densely populated counties, represented by darker shades of blue, include New York County (the borough of Manhattan in New York City), King County (the borough of Brooklyn in New York City), and Bronx County (the borough of The Bronx in New York City) in New York State, with population densities of 50,170.30, 28,236.06, and 25,642.60 individuals per square mile, respectively. Following closely are the City and County of San Francisco in California and St. Louis County in Missouri, with population densities of 18,595 and 16,221.73 individuals per square mile, respectively. In contrast, the least densely populated county, colored in light blue in Figure 3.23, is Esmeralda County in Nevada, with a density of 0.04 individuals per square mile.

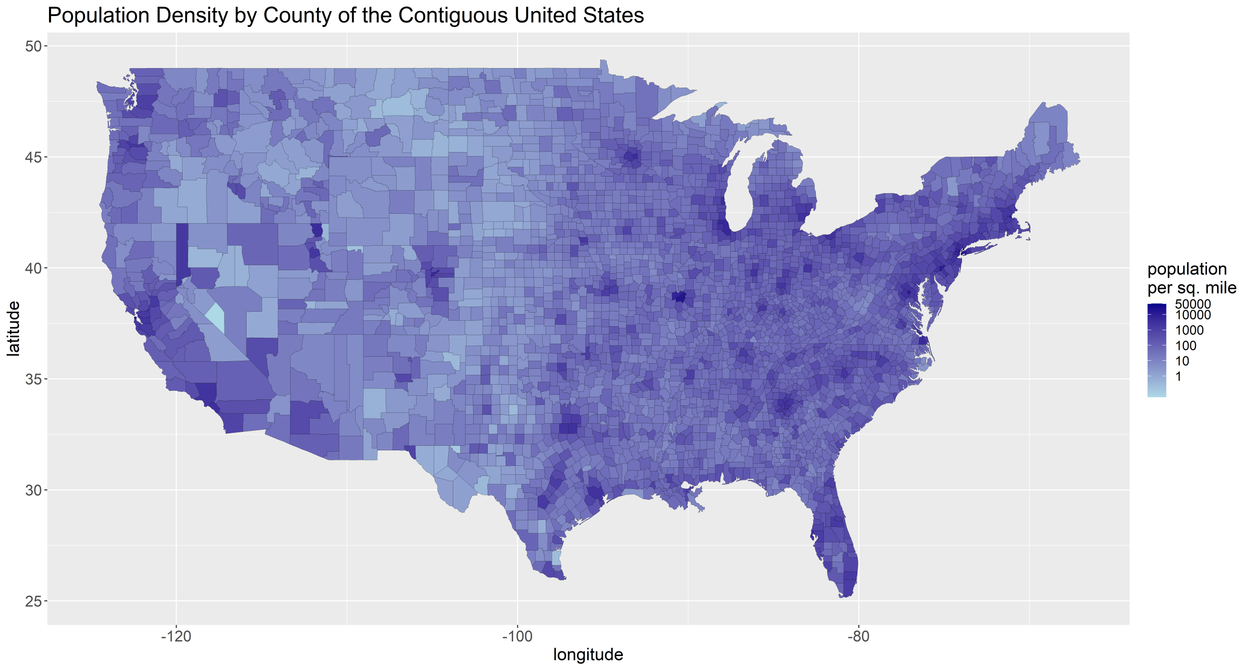


Figure 3.23: Map of Population Density by County in Contiguous United States

Figure 3.24 displays the same map, incorporating UFO sighting reports represented by red points of varying sizes. The size of each point corresponds to the number of reports at a particular coordinate, with larger points indicating a higher frequency of reports at that location. Notably, a significant concentration of sighting reports can be observed along both the west and east coasts, while the central region of the U.S. appears to have a comparatively lower number of reports.

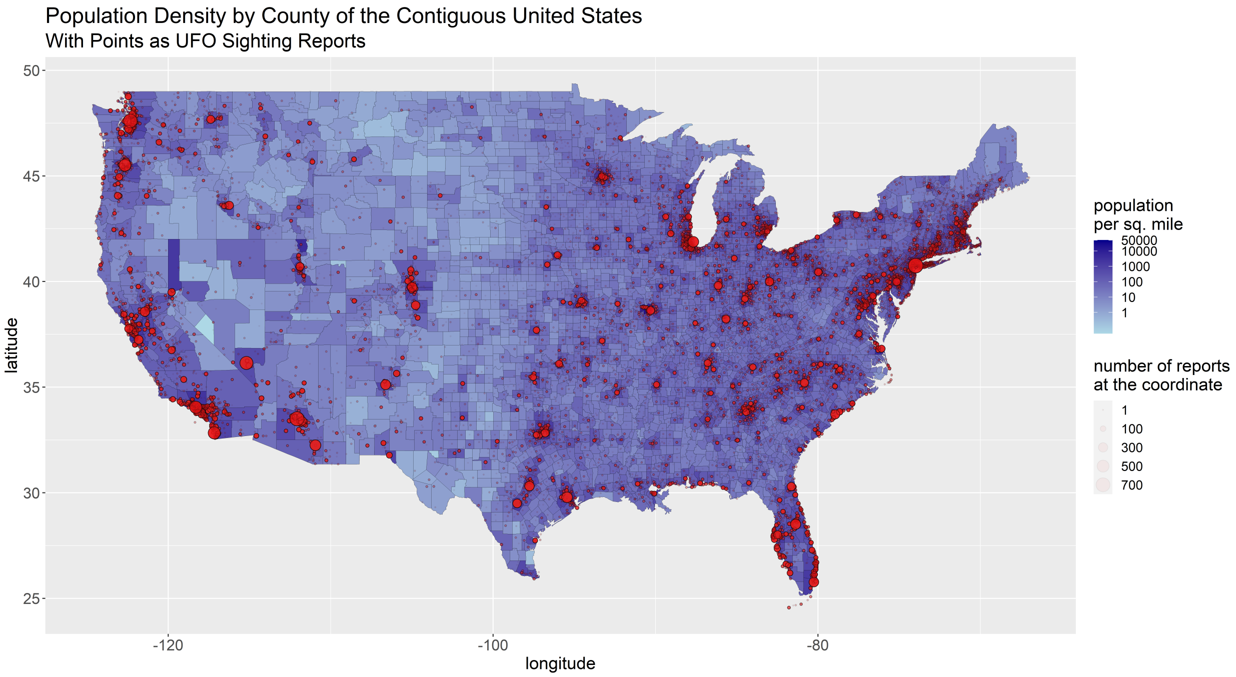


Figure 3.24: Map of Population Density by County in Contiguous United States With Points as UFO Sighting Reports

When a county is represented in a darker shade of blue and also contains larger-sized points, there is a higher likelihood that multiple individuals within that area have observed and reported the same UFO incident, rather than an increased number of distinct individual incidents occurring in that area. For example, a specific location in New York City (NYC) documented a total of 759 sightings over the years, as depicted by the largest-sized red point in Figure 3.24. However, considering the remarkably high population density there, it does not necessarily imply that UFO occurrences are more frequent in NYC compared to other areas. In fact, on November 8, 2003, there were 10 sightings reported to NUFORC at this particular NYC location, all of which were sightings of UFOs during the same lunar eclipse, with most describing the UFO as V-shaped. There are also counterexamples, such as counties surrounding Atlanta in Georgia State, which have relatively lower population densities and more evenly distributed sighting reports.

On the other hand, there are densely populated counties with lower numbers of sightings. For instance, Cache County, located in northeastern Utah, despite having a high population density of 4,252.76 individuals per square mile, has relatively few UFO sighting reports documented in this area.

Taking a closer look at the central region, it becomes evident that the limited number of UFO sighting reports in this area creates a noticeable gap on the map. Intriguingly, this gap coincides with the geographic boundaries of the Ogallala Aquifer, which is one of the largest underground water sources in the United States, spanning across eight states: South Dakota, Nebraska, Wyoming, Colorado, Kansas, Oklahoma, New Mexico, and Texas. Several factors could potentially contribute to the scarcity of sightings in this area, including lower population density, vast rural and agricultural landscapes, and reporting bias related to conservative attitudes towards UFOs. However, it is important to note that despite the relatively lower population density in the Ogallala Aquifer area, there are indeed residents living within this region, as confirmed by Figure 3.25.

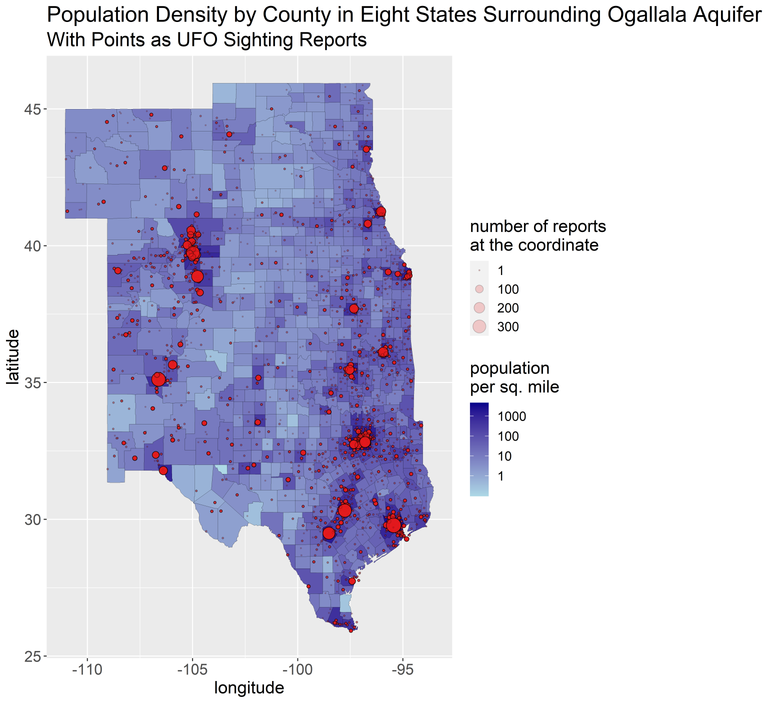


Figure 3.25: Map of Population Density by County in Eight States Surrounding Ogallala Aquifer With Points as UFO Sighting Reports

Figure 3.26 presents an overlay of the map displaying UFO sighting reports in the eight states with the map of the Ogallala Aquifer created by a Wikimedia Commons user “Kbh3rd” [Kbh09]. The population density data is omitted for a clear illustration. A remarkable observation is that UFO occurrences appear to be less frequent in areas where the saturated thickness of the aquifer (i.e., the depth from the water table to the base of the aquifer) is greater. Additionally, there is a noticeable pattern of sightings along the major rivers, suggesting a potential relationship between UFO activity and these waterways.

However, it is important to recognize that no definitive conclusion can be drawn at this point. The observation of a potential connection between UFO activity and water sources presents an intriguing hypothesis that warrants further investigation.

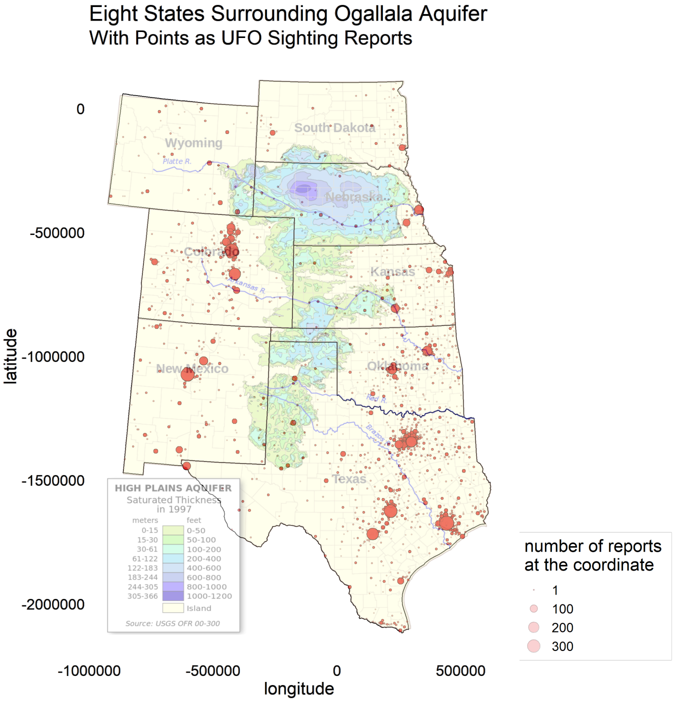


Figure 3.26: Map of Eight States Surrounding Ogallala Aquifer With Points as UFO Sighting Reports Overlaid With the Map of Ogallala Aquifer [Kbh09]

3.4.2 Geodesic Distance to Military Bases

As mentioned in the introduction, some theories suggest that UFO sightings are more likely to occur near military bases due to experiments of advanced military technology or potential espionage activities. To examine these claims from an empirical and statistical perspective, this study intends to investigate whether there is a significant difference between UFO sightings near and away from military bases.

To conduct this analysis, the “Military Bases” dataset, provided by Dominic Menegus in 2019 and obtained from OpenDataSoft, is utilized as an crucial external data source [Tra19]. The dataset was originally derived from the U.S. Department of Transportation / Bureau of Transportation Statistics’s National Transportation Atlas Database. This dataset contains over 700 entries, representing the authoritative boundaries of the most commonly known

Department of Defense sites, installations, ranges, and training areas in the United States and Territories, including their names, locations, and geographic points.

Recall that the “city_latitude” and “city_longitude” variables in “UFO sightings” record the coordinates of each sighting, accurate to the nearest city it belongs to. To determine whether a UFO sighting occurred near a military base, the geodesic distance in kilometers from each UFO sighting to its nearest U.S. military base or installation is calculated [Zes19]. Sightings located within a 20-kilometer radius of a military base are categorized and referred to as “sightings near military bases,” while sightings located more than 20 kilometers away from any military base are categorized and referred to as “sightings away from military bases.” Based on this classification, it is observed that 39.45% of the sightings occurred near military bases, while the remaining 60.54% did not.

Despite there being a greater number of UFO sightings away from military bases, both groups exhibit a remarkably similar trend in reported sightings over the decades. Figure 3.27 illustrates that approximately 49% of sightings occurred in the 2010s, followed by around 30% in the 2000s. Moreover, both groups display an upward trajectory that aligns with the overall trend observed in all reports, with peak years occurring in 2012 and 2014, as demonstrated in Figure 3.28.

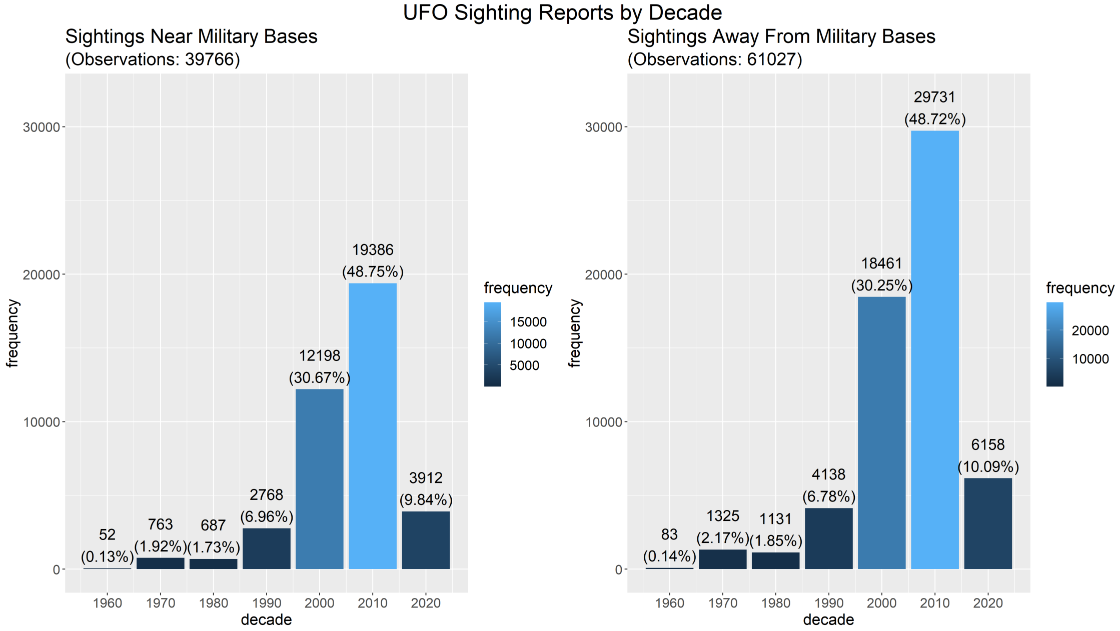


Figure 3.27: Comparison of UFO Sighting Reports Near and Away From Military Bases by Decade

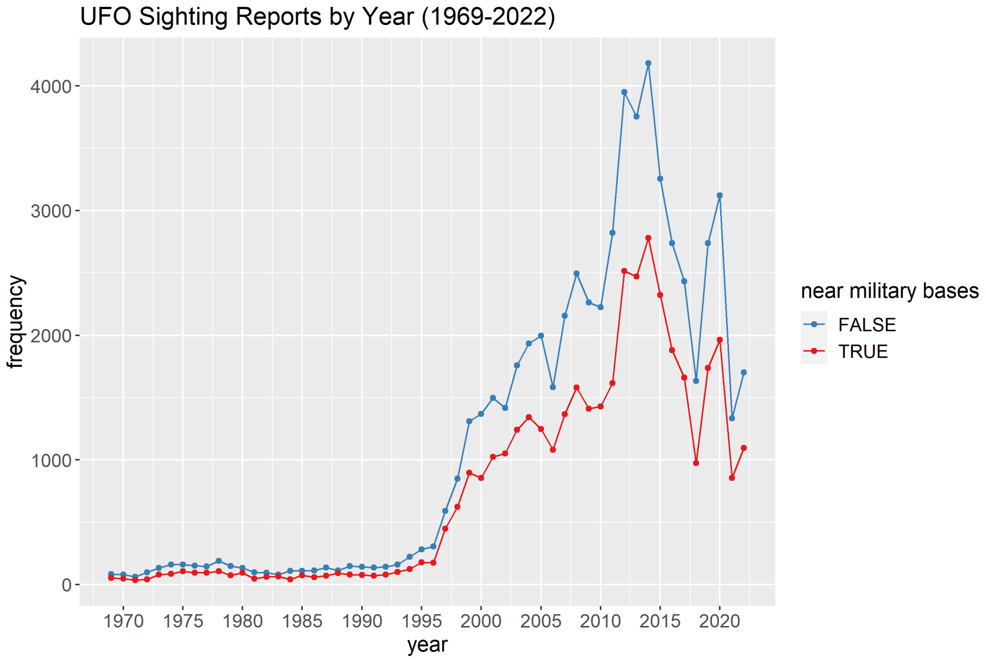
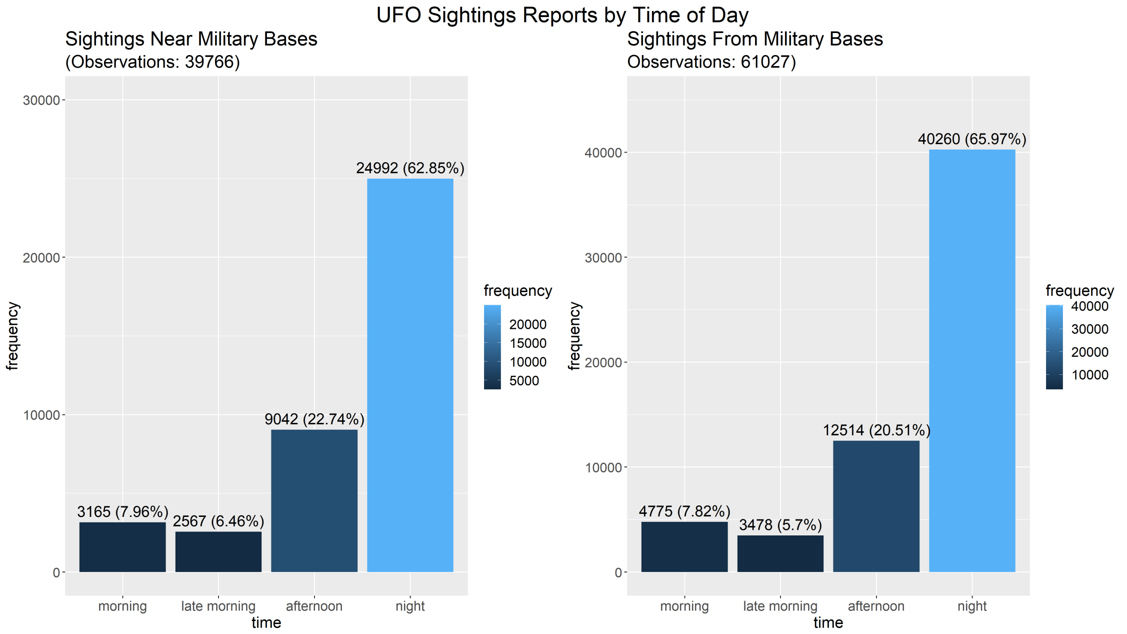


Figure 3.28: Comparison of UFO Sighting Reports Near and Away From Military Bases by Year

Based on Figure 3.29, it appears that UFO sightings near military bases have a slightly higher proportion of sightings occurring earlier in the day. Specifically, approximately 1% more sightings occurred in the late morning (10am to 1pm) and 2% more occurred in the afternoon (2pm to 7pm), resulting in 3% less sightings at night (8pm to 4am). This difference in distribution between UFO sightings near and away from military bases may be attributed to certain peaks observed in sightings near military bases during specific years.

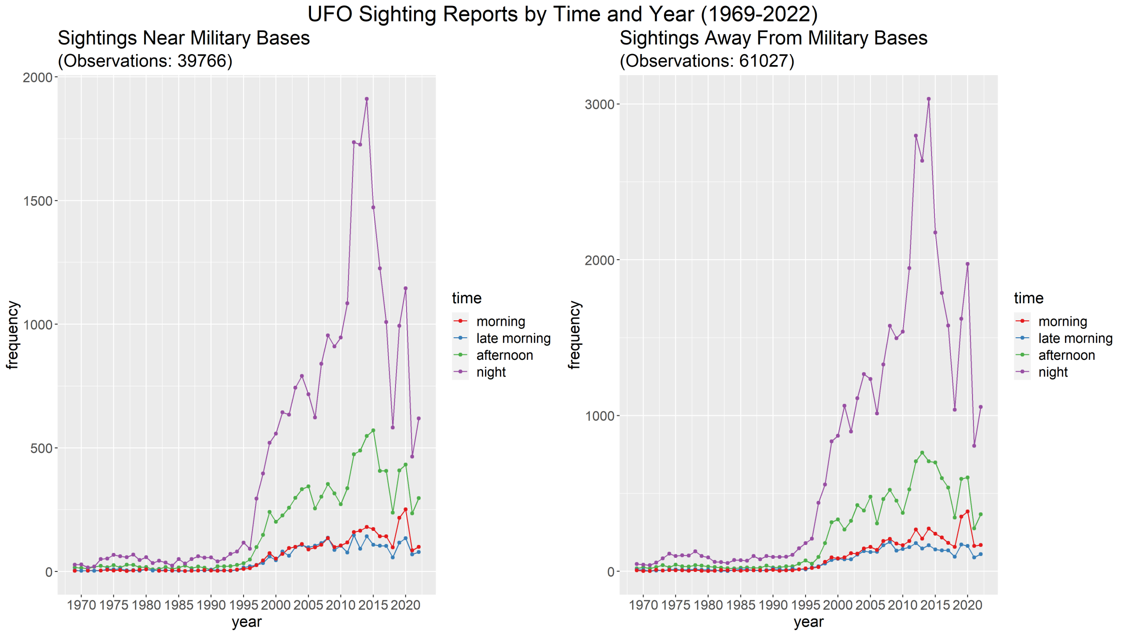
Figure 3.30 provides further insight into these peaks, uncovering notable increases in sightings near military bases during the afternoon in 2014, with a peak in 2015, as well as peaks in the time period of late morning in 2012 and 2014.


UFO Sightings Reports by Time of Day

Sightings Near Military Bases
(Observations: 39766)

Sightings From Military Bases
Observations: 61027)

Figure 3.29: Comparison of UFO Sighting Reports Near and Away From Military Bases by Time


UFO Sighting Reports by Time and Year (1969-2022)

Sightings Near Military Bases
(Observations: 39766)

Sightings Away From Military Bases
(Observations: 61027)

Figure 3.30: Comparison of UFO Sighting Reports Near and Away From Military Bases by Time and Year

Regarding the reported shapes of UFOs, Figure 3.31 demonstrates a comparable distribution in both groups, with approximately 39% of sightings describing the shape as “circle.” While the differences in proportions for each reported shape are minimal, the most notable dissimilarity between the two groups arises in the occurrence of the shape of “light.”

Interestingly, Figure 3.32 uncovers a remarkable finding: sightings occurring away from military bases experienced a sudden increase in the mentions of the shape of “light” in 2009 and 2020. These two particular years stand out as the only instances where the frequency of “light” shape mentions surpassed that of “circle.” This discrepancy is not only observed in sightings away from military bases but also had an impact on the overall trend.

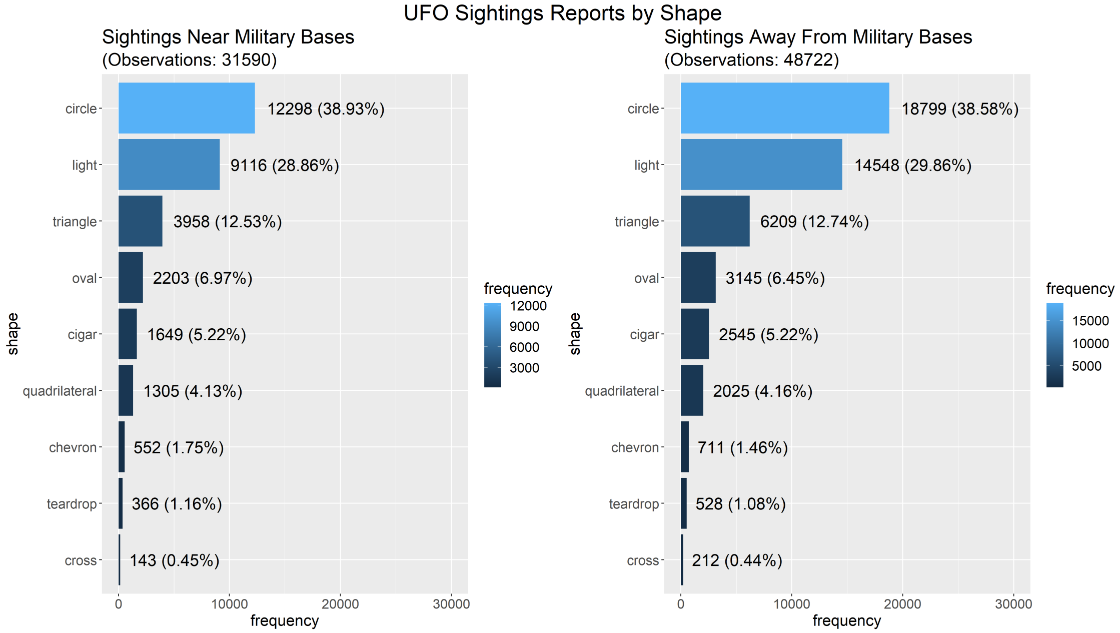


Figure 3.31: Comparison of UFO Sighting Reports Near and Away From Military Bases by Shape

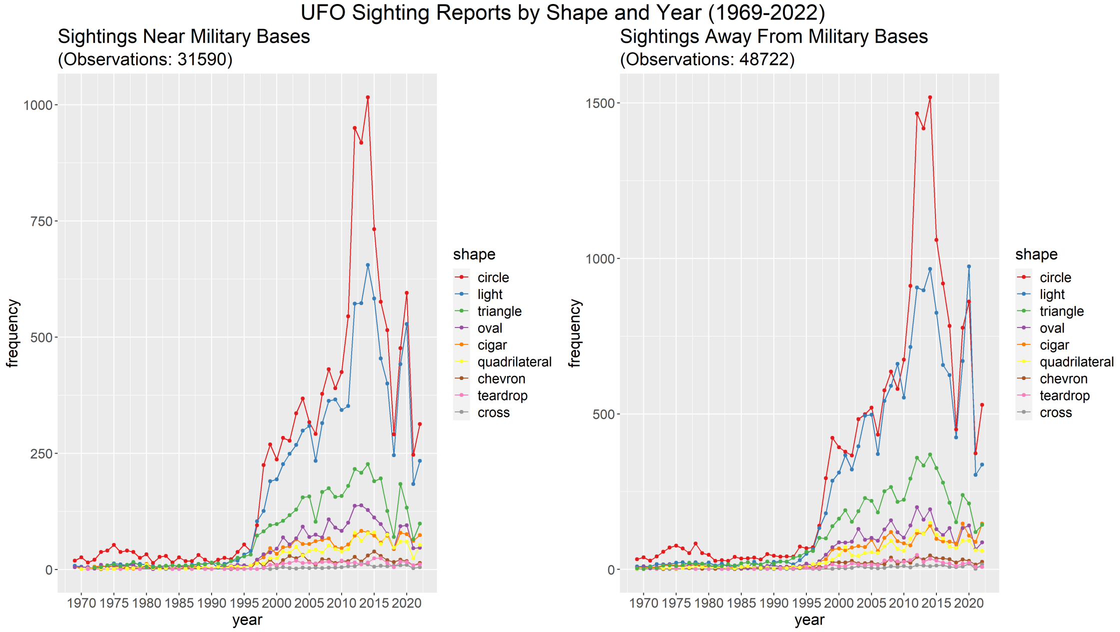


Figure 3.32: Comparison of UFO Sighting Reports Near and Away From Military Bases by Shape and Year

In terms of the reported colors of UFOs, Figure 3.33 demonstrates minimal variations in proportions for each category of reported color, with no significant disparities between the two groups. The most noteworthy dissimilarity between the groups can be observed in the occurrence of the color "orange," albeit with only a slight difference of approximately 1%. It is worth mentioning that a single report may mention multiple colors, which contributes to the larger total number of observations compared to other variables.

Recalling the earlier general explanatory analysis of the "color" variable, it is discovered that there was a sudden increase in the color of "red" in 2014, making it the only year in the 21th century where "red" became the most frequently reported color. By referring to Figure 3.34, it can be explained that this result can be mainly attributed to the increase in mentions of "red" near military bases in 2013 and 2014.

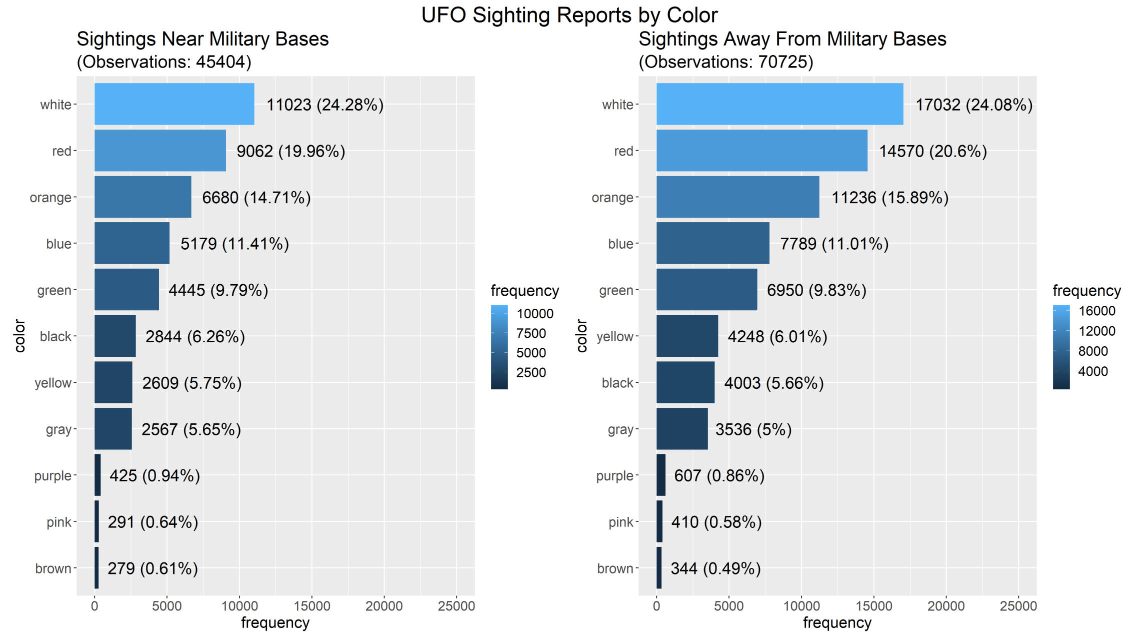


Figure 3.33: Comparison of UFO Sighting Reports Near and Away From Military Bases by Color

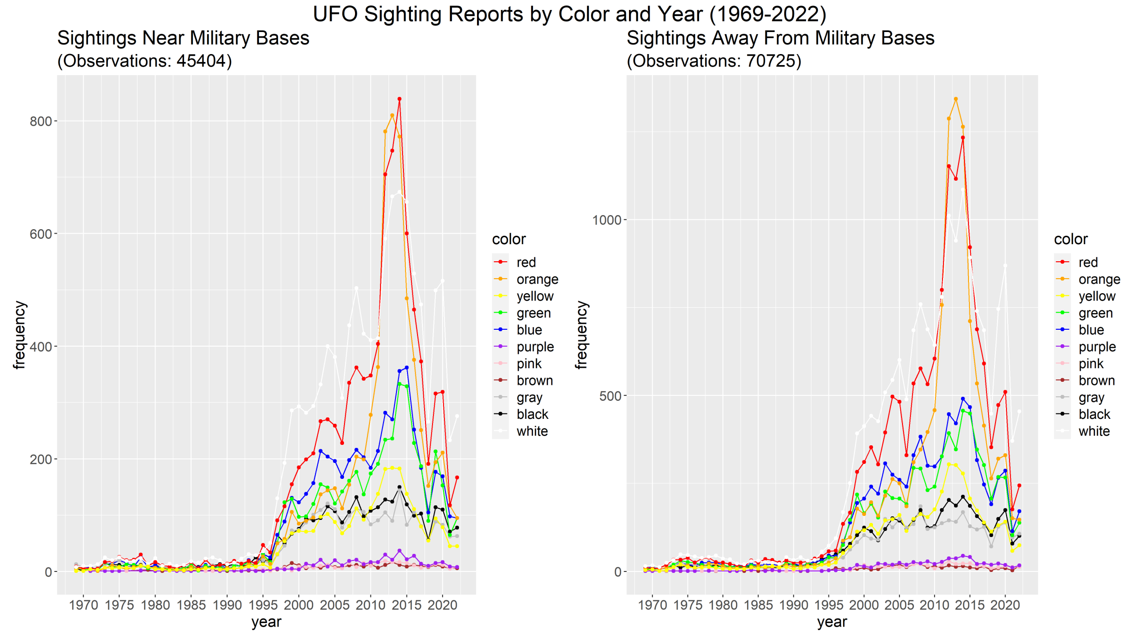


Figure 3.34: Comparison of UFO Sighting Reports Near and Away From Military Bases by Color and Year

Similarly, minimal differences are observed in the combinations of color and shape, as well as the emotions expressed in the text of reports, as shown in Figure 3.35 and 3.36. The disparities in each category are less than 1%, indicating a nearly identical distribution between the two groups.

In conclusion, the explanatory analysis comparing UFO sightings occurred near and away from military bases reveals no significant differences in their properties. The growth pattern of sightings over time, the time of day when the sightings occurred, the reported shapes and colors of the UFOs, and the emotional content of the reports all exhibit remarkably similar trends for both groups. This suggests that the proximity to military bases does not appear to have a substantial impact on these aspects of UFO sightings.

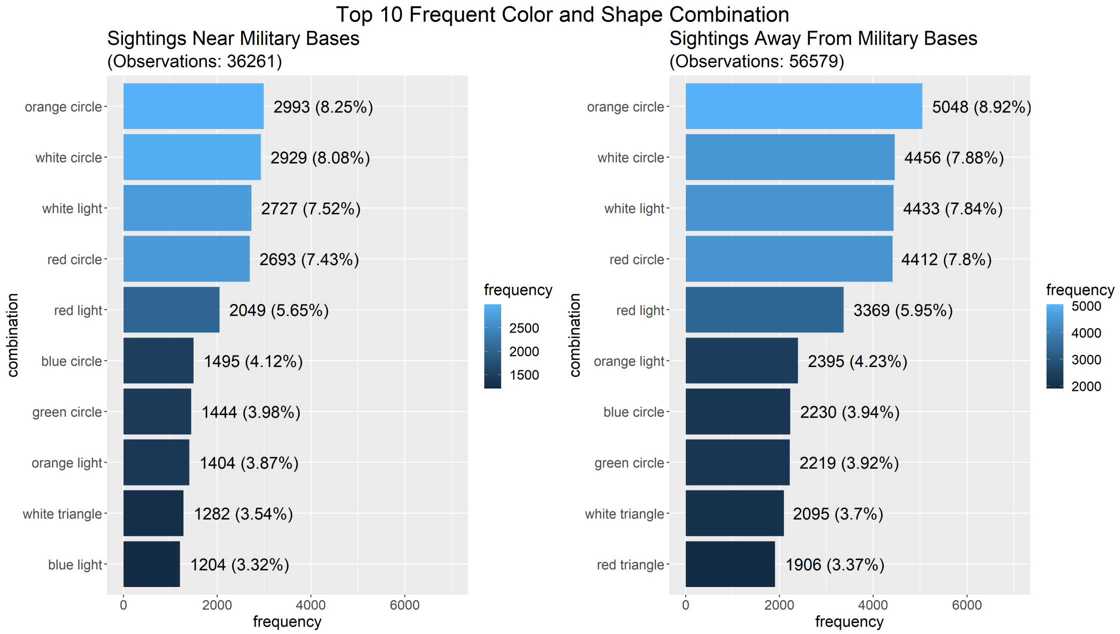


Figure 3.35: Comparison of UFO Sighting Reports Near and Away From Military Bases by Top Frequent Color and Shape Combination

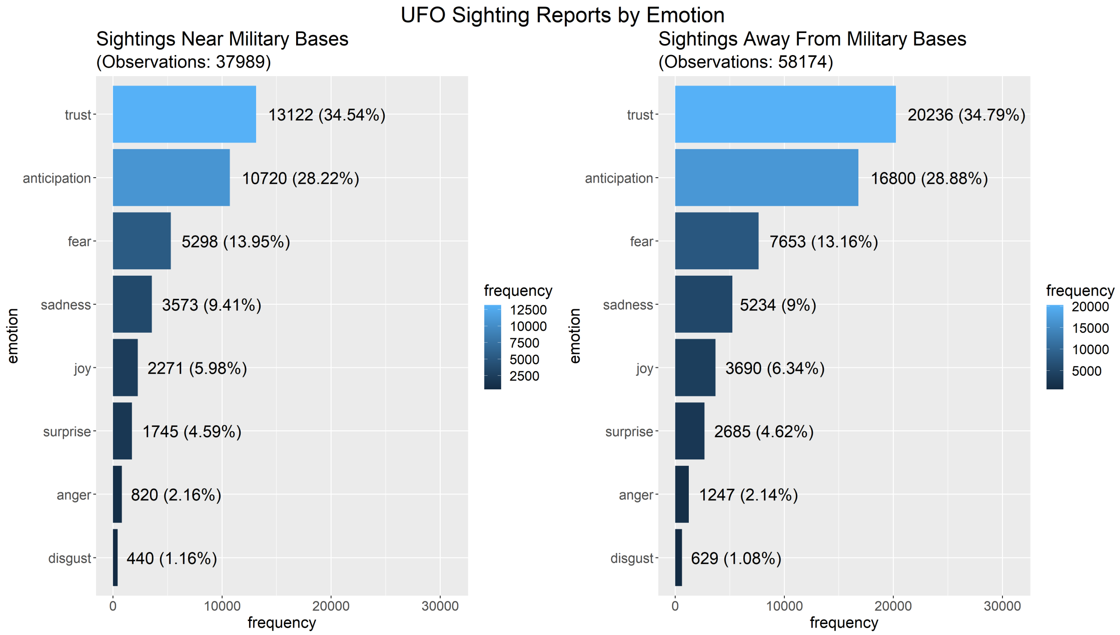


Figure 3.36: Comparison of UFO Sighting Reports Near and Away From Military Bases by Emotion

3.5 Logistic Regression

To examine the potential impact of distance to military bases on UFO sightings, a logistic regression analysis is employed. The dataset is split into a training set, used for model fitting, and a test set, utilized for model evaluation, with a ratio of 70:30. The training dataset consists of 53,742 observations, where 39.3% of them correspond to sightings that occurred near military bases.

The logistic regression is performed to determine the differentiation between UFO sightings that occurred near military bases and those that occurred away from military bases, based on the properties of the sightings. The following variables are taken into account in the logistic regression:

- Near_mb: the sighting occurred near military bases or not. (Binary: True (1) / False

(0))

- Decade: The decade when the sighting occurred. (Categorical with 7 levels: 1960s / 1970s / 1980s / 1990s / 2000s / 2010s / 2020s)
- Time: The time of day when the sighting occurred. (Categorical with 4 levels: Morning / Late Morning / Afternoon / Night)
- Shape: The shape of the UFO described in the sighting report. (Categorical with 9 levels: Circle / Light / Triangle / Oval / Cigar / Quadrilateral / Chevron / Teardrop / Cross)
- Emotion: The emotional content of the UFO sighting report. (Categorical with 8 levels: Trust / Anticipation / Fear / Sadness / Joy / Surprise / Anger / Disgust)

The null hypothesis posits that there is no significant relationship between the response variable, “near_mb”, and the explanatory variables: “decade”, “time”, “shape”, and “emotion”. Rejecting the null hypothesis would suggest that there is a discernible differentiation in the properties of sightings between those occurring near military bases and those occurring away from military bases.

Based on the results of the fitted model, the variables “time” and “shape” are statistically significant, while “decade” and “emotion” are not. These results provide sufficient evidence to reject the null. All levels of the “time” variable yield p-values below 0.05, indicating that the predictor, as a whole, is statistically significant in its relationship with the response variable. Specifically, the level of “night” exhibits an exceptionally low p-value, suggesting a more pronounced association. Holding other variables constant, if a sighting occurs at night compared to the morning, the odds of it being a sighting near military bases decrease by 14.27%, with a 95% confidence interval ranging from 10.46% to 17.92%. Likewise, most levels of the “shape” variable show statistical significance in their relationship with the response variable, with the level of “cigar” demonstrating the most substantial effect. Controlling for other factors, if a sighting involves a cigar-shaped UFO rather than a circular UFO, the odds of it being a sighting near military bases decrease by 22.74%, with a 95% confidence interval ranging from 11.26% to 32.7%. The model achieves an accuracy of 60.41% as assessed by the test set.

While the previous exploratory analysis indicated minimal disparities between the two groups, the hypothesis testing yielded statistically significant results. However, it is important to note that although there are subtle variations in the "time" and "shape" properties of sightings, the practical significance of these results may not align with their statistical significance.

In hypothesis testing, statistical significance is concerned with detecting any difference, regardless of its practical importance. Particularly with a large sample size like the one in this study, even effects that may have little practical relevance can be detected. On the other hand, practical significance refers to whether the observed differences are substantial enough to have meaningful implications in real-life scenarios. Given that the effect sizes of "time" and "shape" variables are relatively small, they may lack practical significance in the real world.

In addition, the problem of multiple comparisons or multiple testing may contribute to the observed statistical significance. When conducting multiple hypothesis tests simultaneously, the probability of obtaining a statistically significant result by chance alone increases. In other words, when performing multiple logistic regressions, the likelihood of obtaining a false positive result (Type I error) rises, potentially leading to variables displaying statistical significance even if no true effect exists.

Moreover, there may exist confounding variables that can affect both the explanatory variable and the response variable. These confounders can create a false association between the variables being studied and consequently lead to spurious results.

Therefore, this study suggests that, based on the NUFORC data, there is no significant practical influence of the distance to military bases on UFO sightings.

CHAPTER 4

Case Study

The last part of this study delves into three case studies of well-known documented UFO incidents in history. The first two incidents draw inspiration from the 2022 documentary "Missing 411: The UFO Connection" directed by David Paulides. In this documentary, Paulides, a former police officer turned conspiracy realist, investigator and writer, investigates the mysterious disappearances of individuals in the wilderness and uncovers a compelling link between UFOs and these vanishings [Pau22].

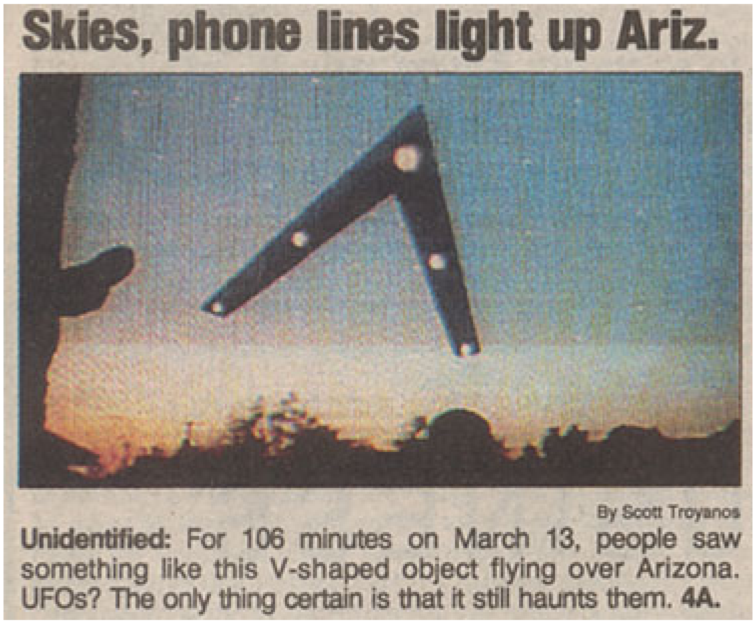
4.1 The "Black Triangle" UFOs

The documentary "Missing 411: The UFO Connection" features a chapter titled "The Idaho Triangle," which recounts the notable UFO encounter of Chris Bales and his brothers. During an elk hunting trip in Challis, Idaho on September 27, 2000, they witnessed a UFO in the mountains at midnight. Bales describes the UFO as a black triangular object with dots emitting red and white glows [Pau22].

In this study, an intriguing observation emerges regarding the prevalence of UFOs described as "black triangle" during the 1980s and 1990s. In fact, numerous documented large-scale triangular UFO sightings occurred during this period. One example is the "Hudson Valley UFOs" sightings, where several hundred people reported witnessing UFOs flying over the Hudson River in New York state in the 1980s. Most of the reports came from reliable witnesses who claimed to spot V-shaped objects rimmed with multiple colored lights

[Sha22].

Additionally, sentiment analysis using the AFINN lexicon reveals a sudden increase in the sentiment score of UFO sighting reports in 1997, indicating a significant shift in attitudes towards UFO sightings. This finding may be attributed to the "Phoenix Lights" incident, which took place on March 13, 1997. The event involved widely sighted triangular formations with spherical lights observed in the skies over the southwestern states of Arizona and Nevada [Wik23f].



Skies, phone lines light up Ariz.

By Scott Troyanos

Unidentified: For 106 minutes on March 13, people saw something like this V-shaped object flying over Arizona. UFOs? The only thing certain is that it still haunts them. **4A.**

Figure 4.1: Image of the Phoenix Lights Newspaper Article From USA Today [Wik20]

Therefore, the first case study attempts to find evidence in the dataset specifically related to "black triangle" UFOs, which are defined in this study as triangular UFOs exhibiting black, red, or white colors. Applying filters to the dataset yields 5,687 reports that meet the criteria of having a shape classified as "triangle" or "chevron" and featuring at least one color from "black," "red" and "white." Figure 4.2 provides an overview of the distribution of sightings with such features. It presents a relatively even distribution that aligns with the overall trend observed in all sighting reports, with the majority of sightings clustered along the west and east coasts, as well as in regions encompassing major cities.

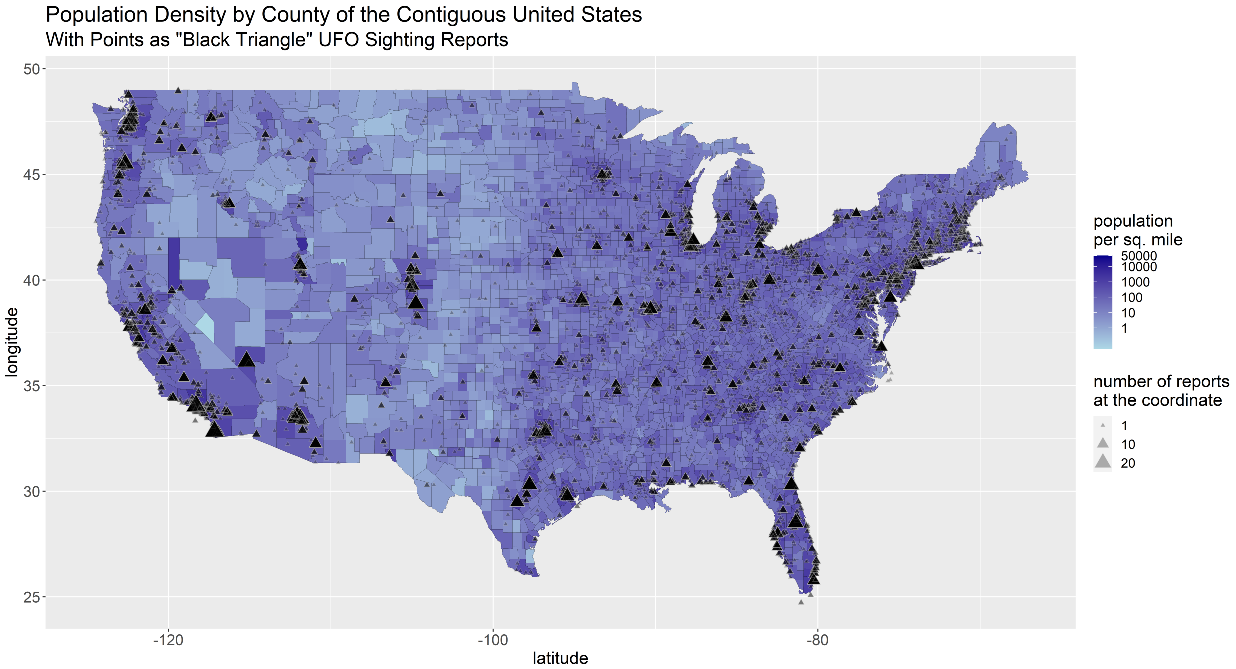


Figure 4.2: Map of Population Density by County in Contiguous United States With Points as “Black Triangle” UFO Sighting Reports

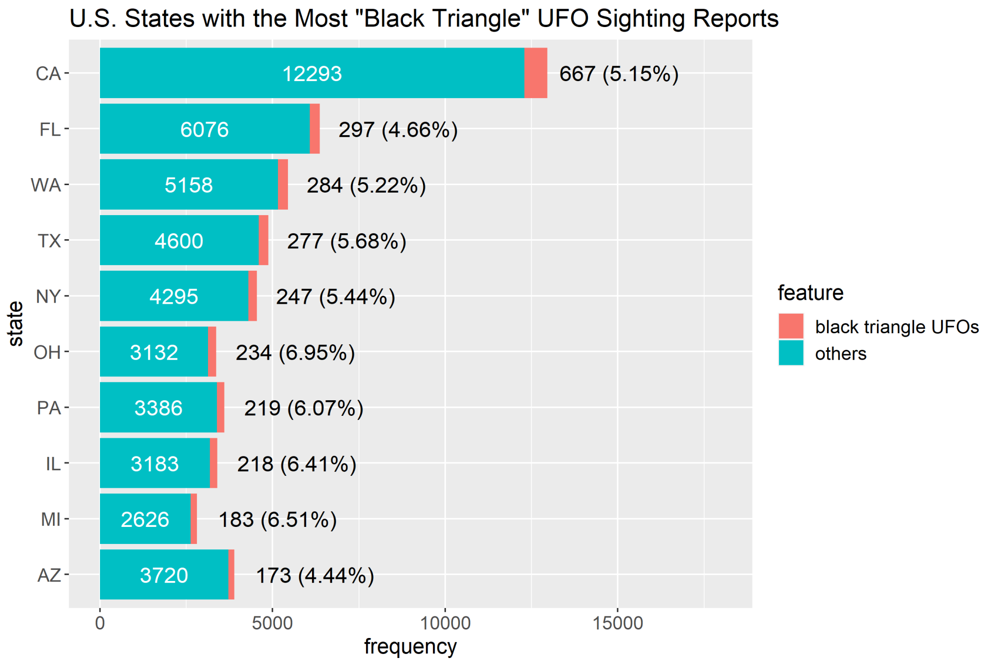


Figure 4.3: U.S. States With the Most “Black Triangle” UFO Sighting Reports and Their Proportion to Total State Reports

According to Figure 4.3, California stands out as the state with the highest number of sighting reports for “black triangle” UFOs, with 667 reports representing 5.15% of all reports within the state. This can be highly attributed to California having the highest number of overall sightings. Other states such as Florida, Washington, Texas, and New York also have a significant number of reports for such sightings. Interestingly, Ohio has a relatively low number of overall sighting reports compared to these states; however, it still has 234 reports of “black triangle” UFOs, accounting for 6.95% of all reports within the state. Similarly, Michigan also has a low number of overall reports, but a notable 6.51% of these reports are for “black triangle” UFO sightings.

Upon examining the states previously mentioned with notable incidents of “black triangle” UFO, it is discovered that the dataset includes 51 reports describing such UFOs in Idaho, accounting for 4.8% of all reports within the state. Figure 4.4 presents that most of these sightings were witnessed along the southern border of Idaho.

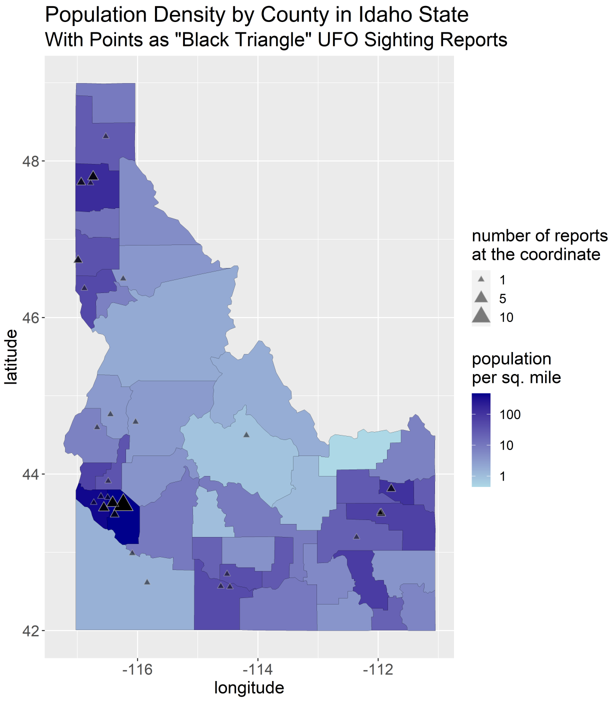


Figure 4.4: Map of Population Density by County in Idaho State With Points as "Black Triangle" UFO Sighting Reports

Taking a closer look at New York State, there are 247 reports documenting qualified triangular UFOs, representing 5.44% of all reports within the state. Figure 4.5 highlights a series of sightings along the Hudson River, which winds up from the southeastern corner of New York State. These sightings provide compelling evidence for the existence of the "Hudson Valley UFOs."

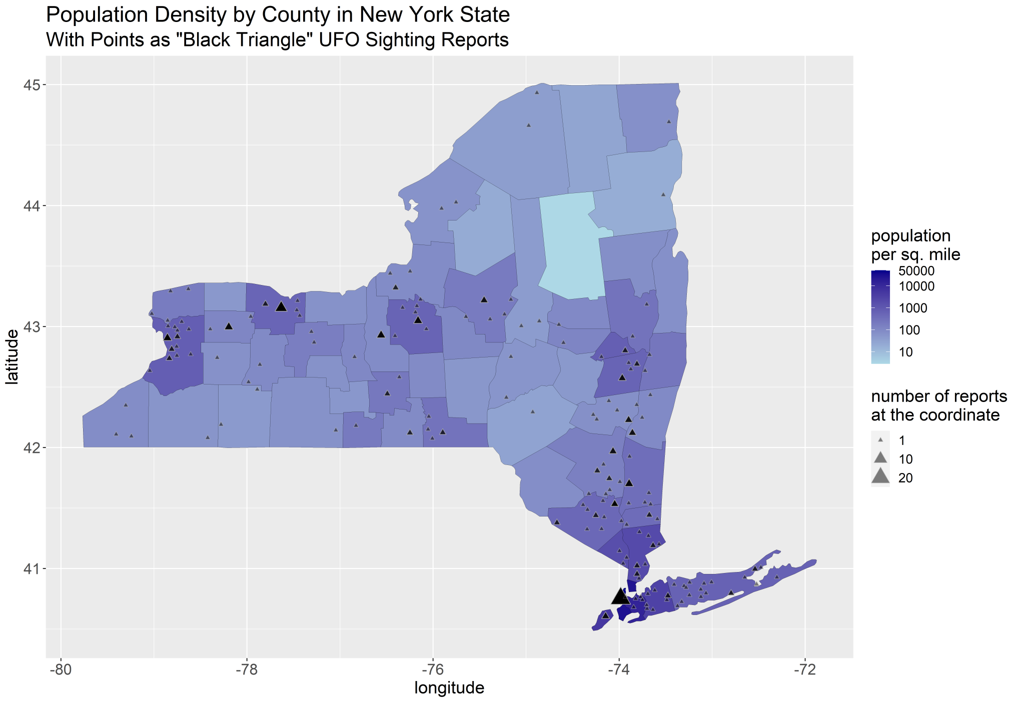


Figure 4.5: Map of Population Density by County in New York State With Points as “Black Triangle” UFO Sighting Reports

Regarding the “Phoenix Lights” incident, Arizona and Nevada had 173 and 48 reports of “black triangle” UFOs, respectively, representing 4.44% and 3.71% of all reports within each state. Specifically, 20 reports were documented in the year of 1997. Figure 4.6 clearly illustrates sightings surrounding Las Vegas, Phoenix, and Tucson, which further support the accounts of the “Phoenix Lights.”

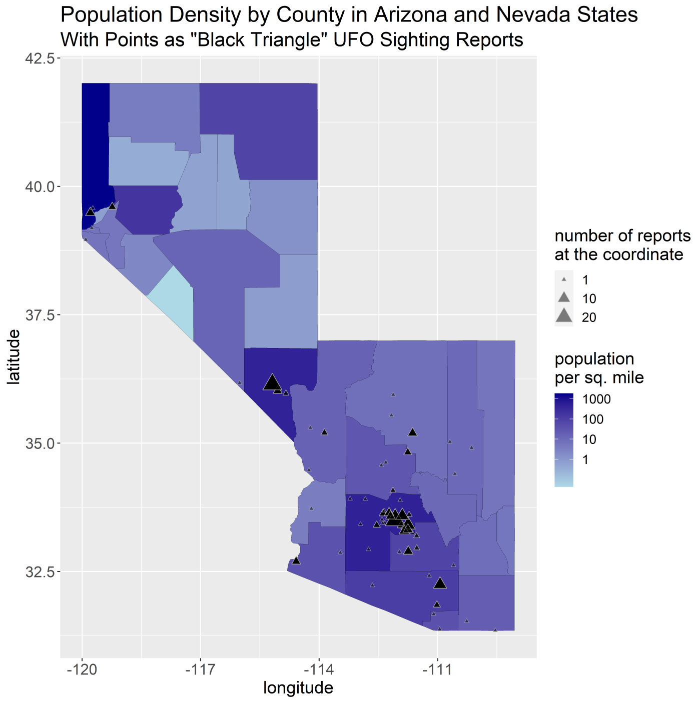

Population Density by County in Arizona and Nevada States
With Points as "Black Triangle" UFO Sighting Reports
42.5
40.0
37.5
35.0
32.5
latitude
-120
-117
-114
-111
longitude
number of reports at the coordinate
1
10
20
population per sq. mile
1000
100
10
1

Figure 4.6: Map of Population Density by County in Arizona and Nevada States With Points as "Black Triangle" UFO Sighting Reports

In conclusion, the dataset from NUFORC provides convincing evidence supporting the existence of these widely witnessed incidents of "Black Triangle" UFOs.

4.2 Animal Abduction

Another chapter titled "The Elk Abduction in Washington" in the documentary recounts the extraordinary event of elk abduction witnessed by a team of approximately 14 forestry workers on the Mountains near Mt. St. Helens in Washington state on February 25, 1999. They claim to have witnessed an elk being lifted off the ground and transported away by a distinctly disc-shaped object. Notably, the elk was suspended beneath the craft without any visible cables. Approximately one week later, a deceased elk was discovered two miles away from the sighting [Evi11]. What makes this event even more astounding is the absence of apparent mutilation, gunshot wounds, or broken bones indicative of a fall [Fil10].

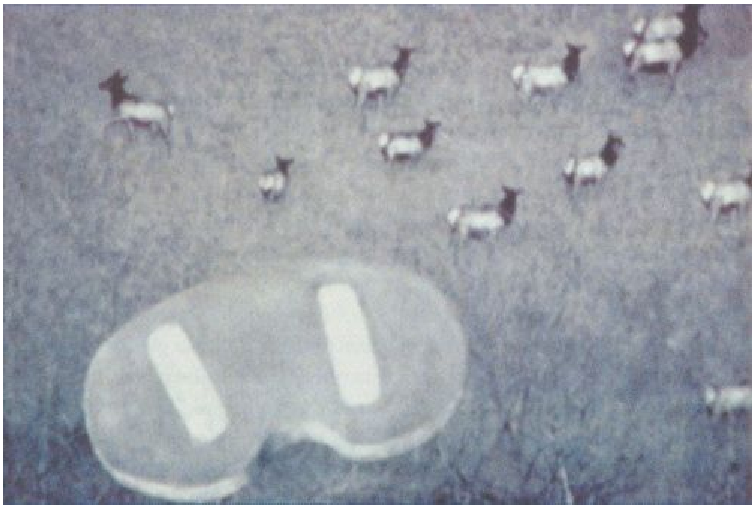


Figure 4.7: Rendition of Elk Abduction by UFO Researcher Robert Fairfax (Credit: FATE Magazine) [Evi11]

The enigma of animal abductions remains mysterious as time goes on. When the topic of UFOs is mentioned, the graphic depiction of an animal being lifted by a UFO, often a cow, tends to come to mind. Thus, the second case study aims to find evidence related to animal abduction within the NUFORC dataset.

Recalling the text mining methodology employed to extract colors from the reports, a similar approach is used to derive the information regarding animal abduction. In this process, the stemmed forms of words (or tokens) from each report are compared against two distinct lists of keywords: one comprising terms related to animals and the other consisting of terms related to abduction. Whenever an intersection is detected, a binary result is recorded for that particular report. The list of animal-related keywords includes the stemmed forms of "elk," "deer," and "moose," as well as livestock including "cow," "cattle," "sheep," "goat," "horse," "donkey," and "mule." The list of abduction-related keywords includes "abduct" and "kidnap". Utilizing stemmed forms helps simplify the process by treating variations as the same. For example, plural and singular forms of animals share the same stemmed form, and terms like “abduction,” “abducted,” and “abduct” are all represented by the stemmed form “abduct.”

As a result, it is determined that animal-related keywords are mentioned in 975 reports, while the abduction-related keywords are mentioned in 526 reports. Applying a strict filter to isolate reports mentioning both keywords related to animals and abduction gives only 26 observations. Nevertheless, it is important to recognize that the presence of these keywords does not guarantee that the reports explicitly discuss witnessing animal abductions by UFOs. A counter example could be individuals claiming to be abducted at a location with an animal reference in its name, such as “Elk Lake.” To obtain a more accurate understanding, a closer examination reveals that only 2 reports genuinely address the topic of animal abductions. One report recounts an incident in Crabapple, Georgia, in 1975, where cattle mutilation was observed without any apparent cut marks after witnessing four gray metallic disc-like objects. The other describes a sighting in Rockdale, Texas, also around 1975, where a cigar-shaped object with multiple rows of stacked red lights was witnessed on the ground. Notably, this sighting coincided with a period when cattle mutilation was reported in the local newspaper.

With a more flexible filter, Figure 4.8 maps out the reports mentioning either animal-related or abduction-related keywords, distinguished by different colors. Several regions exhibit overlapping occurrences of these keywords, such as Maricopa County in Arizona and Bexar County in Texas. However, unlike the mapping results of “Black Triangle” UFOs, the distribution of these reports does not appear to align with the general trend observed in all sighting reports. Rather, it presents a sense of randomness. Without further investigation, it is premature to draw any conclusion at this stage.

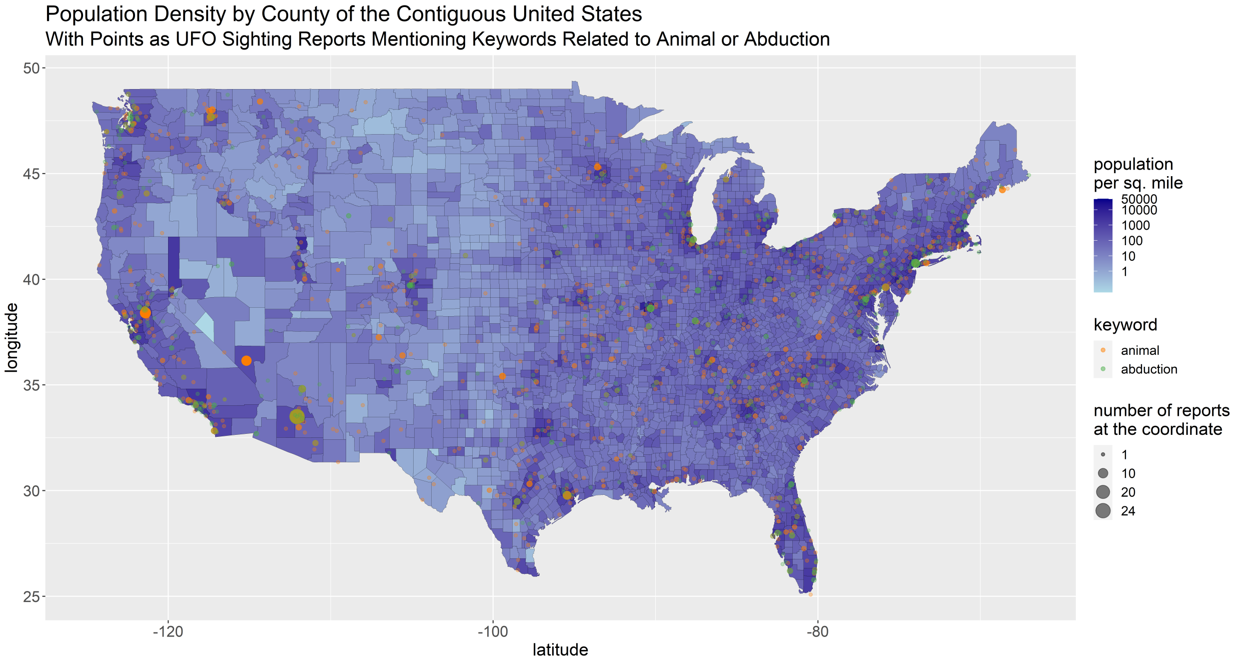

Population Density by County of the Contiguous United States
With Points as UFO Sighting Reports Mentioning Keywords Related to Animal or Abduction
50
45
40
35
30
25
longitude
-120
-100
-80
latitude
population
per sq. mile
50000
10000
1000
100
10
1
keyword
animal
abduction
number of reports
at the coordinate
1
10
20
24

Figure 4.8: Map of Population Density by County in Contiguous United States With Points as UFO Sighting Reports Mentioning Keywords Related to Animal or Abduction

4.3 The Malmstrom Air Force Base UFO Incident

Malmstrom Air Force Base, located in Cascade County, Montana, serves as the headquarters for the United States Air Force's 341st Missile Wing, an intercontinental ballistic missile unit. On the morning of March 16, 1967, an extraordinary UFO incident occurred, profoundly impacting numerous personnel stationed there. The nuclear missiles in Malmstrom Air Force Base experienced rapid and simultaneous disabling, immediately after a saucer-shaped UFO with red glows was observed hovering over the facility. Within seconds, all ten intercontinental ballistic missiles became inoperable, despite the power supply to the sites remaining unaffected. The missile shutdown was attributed to unexplained malfunctions in the guidance and control systems, and this disabling lasted for an entire day [SK00]. In 2014, Robert

Salas, an experienced launch officer at the base who experienced the incident, provided testimony that further adds to the credibility and significance of this inexplicable phenomenon [Cha14].

Therefore, the last case study aims to investigate the possible relationship between sightings of UFOs classified as “Malmstrom AFB Incident” UFOs and their distance to U.S. military bases. In this study, these UFOs are defined as circular in shape and displaying red or orange colors.

Among the 12,968 qualified reports, 37.66% of sightings occurred near military bases (i.e. within a 20-kilometer radius of any military base), while the remaining 62.34% occurred away from military bases, with their geographical distribution depicted in Figure 4.9. Recall that, when considering the entire dataset, 39.45% of the sightings were reported near military bases. This suggests that the occurrence of “Malmstrom AFB Incident” UFOs does not appear to be significantly more frequent near military bases. Examining the temporal distribution of these UFO sightings, both near and away from military bases, Figure 4.10 indicates consistent results with the general trend observed in all sighting reports. Therefore, based on the NUFORC dataset, it appears that the Malmstrom AFB Incident is more likely to be an anomalous event rather than a part of deliberate UFO activities targeting U.S. military bases.

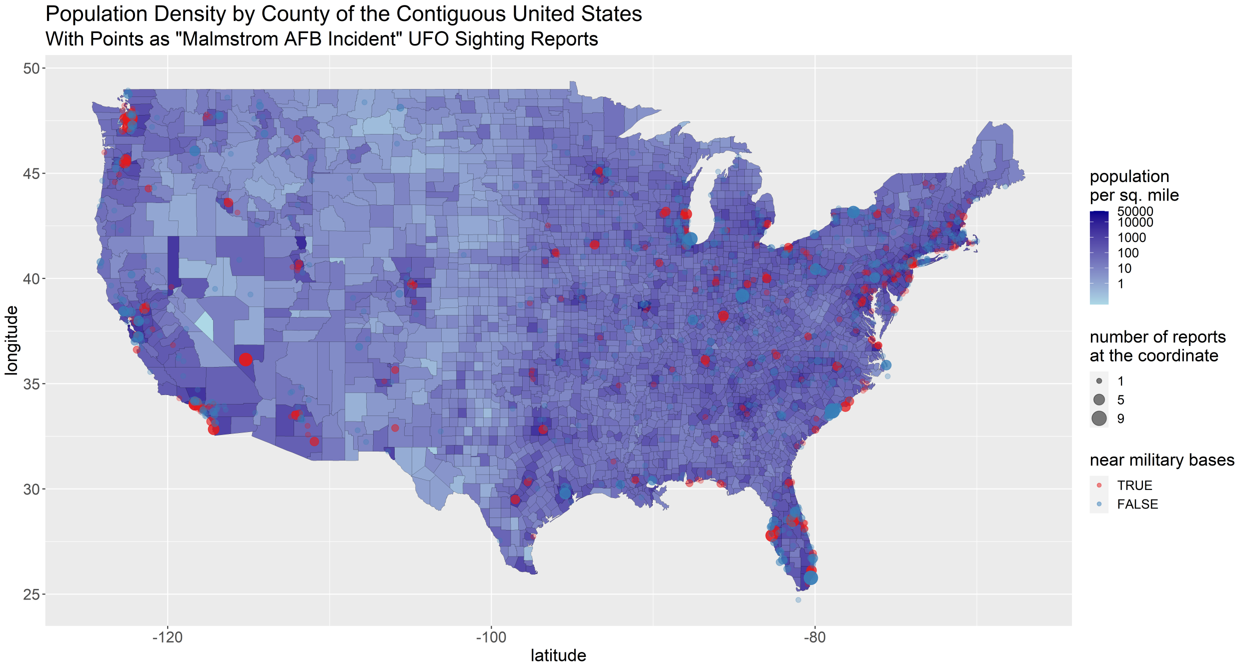


Figure 4.9: Map of Population Density by County in Contiguous United States With Points as “Malmstrom AFB Incident” UFO Sighting Reports

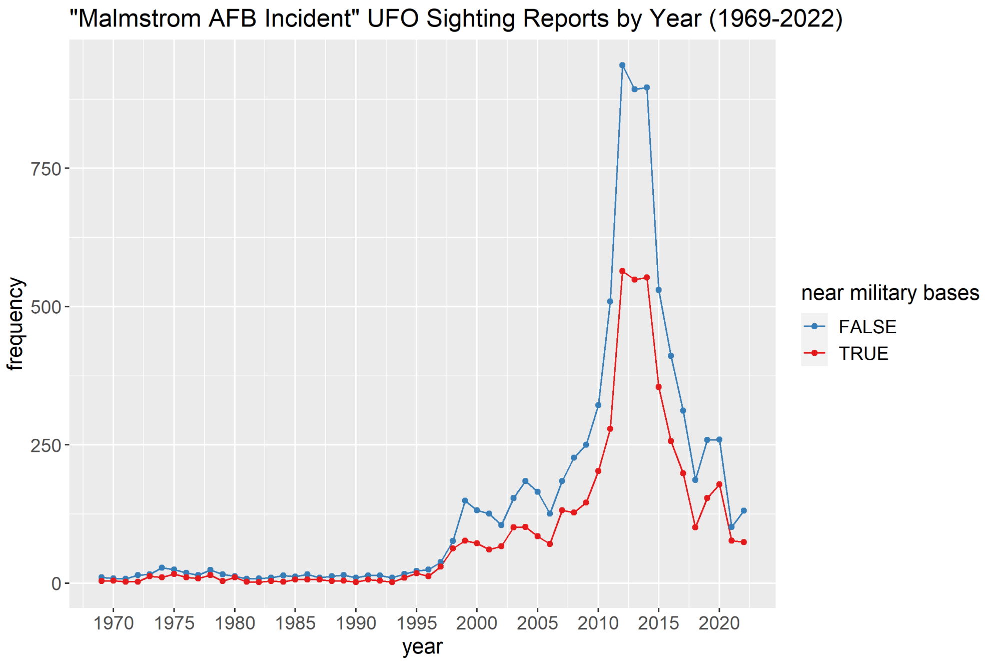


Figure 4.10: “Malmstrom AFB Incident” UFO Sighting Reports by Year

CHAPTER 5

Conclusion and Discussion

In conclusion, this study presents fascinating findings regarding the properties of UFOs reported in the United States, as documented by the National UFO Reporting Center (NUFORC) between 1969 and 2022. The exploratory analysis highlights that the majority of sightings occurred during the 2010s, with notable peaks recorded in 2012, 2014, and 2020. Moreover, the geographical distribution of sightings reveals a significant concentration along the east and west coastlines, with California emerging as the state with the highest number of documented sightings and the Western United States as the region with the most occurrences. In contrast, the central region exhibits a lower frequency of sightings, resulting in a distinct gap on the map that coincides with the geographic boundary of the Ogallala Aquifer.

The most frequently mentioned shapes in the reports were “circle,” “light,” and “triangle,” accounting for 38.72%, 29.47%, and 12.66%, respectively. Notably, in 2020, there was a sudden surge in reports describing the shape of UFOs as “light,” making it the only year in the 21st century where “light” had the highest frequency among the nine shape categories.

Colors mentioned in the text of the reports were extracted using text mining techniques. The most frequently mentioned colors were “white,” “red,” and “orange,” accounting for 24.16%, 20.35%, and 15.43%, respectively. During 2012 and 2013, a sudden increase in reports mentioning the color “orange” made it the only period where ”orange” had the highest frequency among the 11 color categories. Similarly, in 2014, a sharp rise in reports mentioning the color “red” made it the only year in the 21st century where “red” held the highest frequency.

In terms of the combinations of color and shape, combinations involving “white light,” “white circle,” “orange circle,” and “red circle” consistently stand out as the predominantly mentioned combinations throughout the years. However, an intriguing observation arises as UFOs resembling “black triangle” were prominently reported during the 1980s and 1990s. This coincides with documented incidents of “black triangle” UFOs witnessed by substantial groups of people in various locations across the United States during that period. Subsequently, a case study focusing on “black triangle” UFOs provides supporting evidence of these incidents, based on the NUFORC dataset.

Regarding public attitudes towards UFO sightings, the sentiment analysis uncovers a noteworthy transition in attitudes from negative to positive sentiment in 1997. This shift in attitudes could potentially be attributed to the widely recognized “Phoenix Lights” UFO incident in that year. Interestingly, the emotional analysis indicates that the majority of reports conveyed emotions of “trust” and “anticipation,” indicating that individuals who choose to report their sightings generally approach the experience with an open mind and a genuine curiosity, seeking plausible explanations for the phenomena they encountered.

To evaluate the theories suggesting a higher likelihood of UFO sightings near military bases from a statistical perspective, sightings were divided into two groups based on their geodesic distance to the nearest military bases. Minimal differences were observed between the two groups across various factors, including the number of sightings over time, the time of day when the sightings occurred, the reported shapes and colors of the UFOs, and the emotional content behind the reports. As a result, this study does not provide support for the hypothesis proposing a potential impact of proximity to military bases on UFO sightings.

This study also acknowledges several limitations that can be addressed for further improvement. One limitation is the presence of additional factors within the NUFORC dataset that warrant investigation. For instance, the “duration” variable captures the length of time for each UFO sighting. However, the absence of consistent units and documentation formats within this variable poses challenges for data cleaning and accurate data analysis. Additionally, the peaks observed in the number of sightings in 2012, 2014, and 2020 require further investigation to gain a deeper understanding of the underlying factors contributing to these notable increases. Addressing these limitations would enhance the comprehensiveness and accuracy of future research endeavors in the field of UFO sightings.

APPENDIX A

Appendix of R Code

A.1 R Code for Word Cloud

The following R code creates data frames of the top frequent words (in stemmed form) in reports from each region of the United State and generates word clouds by region, using the shape of region as the mask for the word cloud visualization.

```
library(dplyr)
library(tibble)
library(ggplot2)
library(SnowballC)
library(textstem)
library(ggwordcloud)
library(usmap)

######## Load Dataset ########

dir <- "C:/Users/yuki/OneDrive/Thesis/"
setwd(dir)

# dataset from: https://data.world/timothyrenner/ufo-sightings
ufo <- read.csv(file = paste0(dir, "nuforc_reports.csv"),
                encoding = "UTF-8",
                na.strings = c("", "NA"))

ufo <- ufo %>%
  filter(country == "USA") %>%
  na.omit()

######## Region ########

northeast <- c("CT", "ME", "MA", "NH", "RI", "VT",
               "NJ", "NY", "PA")
midwest <- c("IL", "IN", "MI", "OH", "WI",
             "IA", "KS", "MN", "MO", "NE", "ND", "SD")
south <- c("DE", "FL", "GA", "MD", "NC", "SC", "VA", "DC", "WV",
           "AL", "KY", "MS", "TN",
           "AR", "LA", "OK","TX")
west <- c("AZ", "CO", "ID", "MT", "NV", "NM", "UT", "WY",
          "AK", "CA", "HI", "OR", "WA")

ufo <- ufo %>%
  mutate(region = case_when(state %in% northeast ˜ "northeast",
                            state %in% midwest ˜ "midwest",
                            state %in% south ˜ "south",
                            state %in% west ˜ "west"))

ufo <- ufo %>%
  filter(!is.na(region)) %>%
  filter(city_latitude >= 19 & city_longitude <= -67)

######## Top Frequent Words by Region ########

ufo_tidy_northeast <- ufo %>%
  filter(region == "northeast") %>%
  select(text) %>%
  unnest_tokens(output = word, input = text) %>%
  filter(!grepl("[0-9]", word)) %>%
  anti_join(stop_words, by = "word") %>%
  mutate(stem = wordStem(word)) %>%
  count(stem, sort = TRUE)

ufo_tidy_midwest <- ufo %>%
  filter(region == "midwest") %>%
  select(text) %>%
  unnest_tokens(output = word, input = text) %>%
  filter(!grepl("[0-9]", word)) %>%
  anti_join(stop_words, by = "word") %>%
  mutate(stem = wordStem(word)) %>%
  count(stem, sort = TRUE)

ufo_tidy_south <- ufo %>%
  filter(region == "south") %>%
  select(text) %>%
  unnest_tokens(output = word, input = text) %>%
  filter(!grepl(’[0-9]’, word)) %>%
  anti_join(stop_words, by = "word") %>%
  mutate(stem = wordStem(word)) %>%
  count(stem, sort = TRUE)

ufo_tidy_west <- ufo %>%
  filter(region == "west") %>%
  select(text) %>%
  unnest_tokens(output = word, input = text) %>%
  filter(!grepl("[0-9]", word)) %>%
  anti_join(stop_words, by = "word") %>%
  mutate(stem = wordStem(word)) %>%
  count(stem, sort = TRUE)

######## WordCloud ########

# create background/mask images
plot_usmap(regions = "states") +
  theme(panel.background = element_blank())

ggsave(filename = "./Figures/text_mining/us_background.png",
       width = 16, height = 9, units = "in",
       dpi = 300)

plot_usmap(include = .northeast_region, fill = "black", labels = FALSE) +
  theme(panel.background = element_blank())

ggsave(filename = "./Figures/text_mining/northeast_background.png",
       width = 16, height = 9, units = "in",
       dpi = 300)

plot_usmap(include = .midwest_region, fill = "black", labels = FALSE) +
  theme(panel.background = element_blank())

ggsave(filename = "./Figures/text_mining/midwest_background.png",
       width = 16, height = 9, units = "in",
       dpi = 300)

plot_usmap(include = .south_region, fill = "black", labels = FALSE) +
  theme(panel.background = element_blank())

ggsave(filename = "./Figures/text_mining/south_background.png",
       width = 16, height = 9, units = "in",
       dpi = 300)

plot_usmap(include = .west_region, fill = "black", labels = FALSE) +
  theme(panel.background = element_blank())

ggsave(filename = "./Figures/text_mining/west_background.png",
       width = 16, height = 9, units = "in",
       dpi = 300)

# ggwordcloud
set.seed(123)
```

```
my_palette <- c(RColorBrewer::brewer.pal(12, "Paired")[1:10],
                # replace the light color in Paired brewer
                RColorBrewer::brewer.pal(9, "Set1")[9],
                RColorBrewer::brewer.pal(12, "Paired")[12])

wordcloud_northeast <- ufo_tidy_northeast %>%
  filter(n > 1000) %>%
  ggplot(aes(label = stem, size = n, color = factor(-(n - n %% 2000)))) +
  geom_text_wordcloud_area(
    rm_outside = TRUE,
    mask = png::readPNG(
      paste0(dir, "Figures/text_mining/northeast_background.png"))) +
  scale_size_area(max_size = 20) +
  scale_color_manual(values = my_palette) +
  theme_void()

ggsave(filename = "./Figures/text_mining/wc_northeast.png",
       plot = wordcloud_northeast,
       dpi = 600)

wordcloud_midwest <- ufo_tidy_midwest %>%
  filter(n > 1000) %>%
  ggplot(aes(label = stem, size = n, color = factor(-(n - n %% 2000)))) +
  geom_text_wordcloud_area(
    rm_outside = TRUE,
    mask = png::readPNG(
      paste0(dir, "Figures/text_mining/midwest_background.png"))) +
  scale_size_area(max_size = 40) +
  scale_color_manual(values = my_palette) +
  theme_void()

ggsave(filename = "./Figures/text_mining/wc_midwest.png",
       plot = wordcloud_midwest,
       dpi = 600)

wordcloud_south <- ufo_tidy_south %>%
  filter(n > 2000) %>%
  ggplot(aes(label = stem, size = n, color = factor(-(n - n %% 3000)))) +
  geom_text_wordcloud_area(
    rm_outside = TRUE,
    mask = png::readPNG(paste0(
      dir, "Figures/text_mining/south_background.png"))) +
  scale_size_area(max_size = 20) +
  scale_color_manual(values = my_palette) +
  theme_void()

ggsave(filename = "./Figures/text_mining/wc_south.png",
       plot = wordcloud_south,
       dpi = 600)

wordcloud_west <- ufo_tidy_west %>%
  filter(n > 2000) %>%
  ggplot(aes(label = stem, size = n, color = factor(-(n - n %% 5000)))) +
  geom_text_wordcloud_area(
    rm_outside = TRUE,
    mask = png::readPNG(paste0(
      dir,"Figures/text_mining/west_background.png"))) +
  scale_size_area(max_size = 20) +
  scale_color_manual(values = my_palette) +
  theme_void()

ggsave(filename = "./Figures/text_mining/wc_west.png",
       plot = wordcloud_west,
       dpi = 600)
```

A.2 R Code for Creating Color Variable

The following R code extracts the colors mentioned in the original text of reports.

```
######## Create a Color List of 28 Colors ########

color_list <- c(
  "red", "maroon",
  "orange", "peach",
  "yellow", "gold",
  "green", "lime", "olive", "turquoise", "teal",
  "blue", "cyan", "navy", "indigo",
  "purple", "violet",
  "pink", "magenta",
  "brown", "tan",
  "gray", "grey", "silver",
  "black",
  "white", "beige", "ivory")

######## Create a Color Dictionary of 11 Colors ########

color_dictionary <- c(
  "red" = "red", "maroon" = "red",
  "orange" = "orange", "peach" = "orange",
  "yellow" = "yellow", "gold" = "yellow",
  "green" = "green", "lime" = "green", "olive" = "green",
  "turquoise" = "green", "teal" = "green",
  "blue" = "blue", "cyan" = "blue",
  "navy" = "blue", "indigo" = "blue",
  "purple" = "purple", "violet" = "purple",
  "pink" = "pink", "magenta" = "pink",
  "brown" = "brown", "tan" = "brown",
  "gray" = "gray", "grey" = "gray", "silver" = "gray",
  "black" = "black",
  "white" = "white", "beige" = "white", "ivory" = "white")

######## Derive Color From Tokens ########

ufo <- ufo %>%
  mutate(color1 = NA) %>%
  mutate(color2 = NA) %>%
  mutate(color3 = NA) %>%
  mutate(color4 = NA) %>%
  mutate(color5 = NA) %>%
  mutate(color6 = NA) %>%
  mutate(color7 = NA) %>%
  mutate(color8 = NA) %>%
  mutate(color9 = NA) %>%
  mutate(color10 = NA) %>%
  mutate(color11 = NA)

color_id <- c(
  "red" = which(colnames(ufo) == "color1"),
  "orange" = which(colnames(ufo) == "color2"),
  "yellow" = which(colnames(ufo) == "color3"),
  "green" = which(colnames(ufo) == "color4"),
  "blue" = which(colnames(ufo) == "color5"),
  "purple" = which(colnames(ufo) == "color6"),
  "pink" = which(colnames(ufo) == "color7"),
  "brown" = which(colnames(ufo) == "color8"),
  "gray" = which(colnames(ufo) == "color9"),
  "black" = which(colnames(ufo) == "color10"),
  "white" = which(colnames(ufo) == "color11"))

for (i in 1:nrow(ufo)) {

  candidates <- ufo_token %>% filter(index == i) %>% pull(word)
  derived_colors <- intersect(unique(candidates), color_list)
  final_colors <- unique(as.character(color_dictionary[derived_colors]))

  if (length(final_colors) > 0) {
    for (j in 1:length(final_colors)) {
      ufo[i, as.numeric(color_id[final_colors[j]])] <- final_colors[j]
    }
  }

}
```

A.3 R Code for Creating Emotion Variable

The following R code extracts the emotions (defined by the nrc lexicon) mentioned in the original text of reports.

```
library(tidytext)


set.seed(123)


ufo <- ufo %>%
  mutate(emotion = NA)


emotion_id <- which(colnames(ufo) == "emotion")


ufo_emotions <- ufo_token %>%
  inner_join(nrc, multiple = "all") %>%
  filter(sentiment != "positive" &
           sentiment != "negative" &
           !is.na(sentiment))


for (i in 1:nrow(ufo)) {
  target_emotions <- ufo_emotions %>%
    filter(index == i) %>%
    count(sentiment, sort = TRUE)


  # frequency of primary emotion(s)
  primary_freq <- as.integer(target_emotions[1, 2])
  # number of primary emotion(s)
  primary_num <- nrow(target_emotions %>%
                        filter(n == primary_freq))

  # report i has only one primary emotion
  if (primary_num == 1) {
    ufo[i, emotion_id] <- as.character(target_emotions[1, 1])
  }

  # report i has more than one primary emotions
  # randomly choose one as the final result
  if (primary_num > 1) {
    x <- sample(1:primary_num, 1) # generate a random integer
    ufo[i, emotion_id] <- as.character(target_emotions[x, 1])
  }
}
```

A.4 R Code for Mapping Population Density by County

The following R code creates a data frame of population density by county in Contiguous United States and maps the results.

```
library(dplyr)
library(tidyverse)
library(ggplot2)
library(maps)
library(maptools)

######## Load Population Density Dataset ########

dir <- "C:/Users/yuki/OneDrive/Thesis/"
setwd(dir)

# dataset from: Wikipedia
#               List of United States counties and county equivalents
# generated by: https://wikitable2csv.ggor.de/
pop_density <- read.csv(
  file = paste0(
    dir, "List_of_United_States_counties_and_county_equivalents.csv"),
  encoding = "UTF-8")

pop_density <- pop_density %>%
  select(1:4) %>%
  rename(county = 1,
         state = 2,
         population = 3,
         area = 4) # area in square mile

# data cleaning
pop_density_cleaned <- pop_density %>%
  mutate(county =
           replace(county,
                   county == "District of Columbia", "Washington")) %>%
  # xx Parish in Louisiana
  mutate(county = gsub("Parish", "", county)) %>%
  # San Francisco, City and County of -> San Francisco
  mutate(county = gsub("(.*),.*", "\\1", county)) %>%
  # Rose Atoll (Rose Island) -> Rose Atoll
  mutate(county = gsub(" \\s*\\([^\\)]+\\)", "", county)) %>%
  # St. Clair -> St Clair
  mutate(county = gsub("[^0-9A-Za-z///’ ]", "", county)) %>%
  # O’Brien -> OBrien
  mutate(county = gsub("’", "", county)) %>%
  # remove white space
  mutate(county = gsub(" ", "", county)) %>%
  # Dona Ana in New Mexico
  mutate(county =
           replace(county,
                   county == "DoaAna", "DonaAna")) %>%
  mutate(state = gsub("[^0-9A-Za-z///’ ]", "", state)) %>%
  mutate(state = gsub("’", "", state)) %>%
  mutate(state = gsub(" ", "", state))
```

```
options(scipen = 999)

pop_density_cleaned <- pop_density_cleaned %>%
  mutate(county_state =
           paste0(tolower(county), "_", tolower(state))) %>%
  mutate(population = as.numeric(gsub(",", "", population))) %>%
  # density_mi2 = population per square mile
  mutate(density_mi2 = population / area) %>%
  # density_km2 = population per square kilometer
  mutate(density_km2 = density_mi2 * 0.3861) %>%
  select(county_state, density_mi2)

######## Load Mapping Dataset ########

us_county <- map_data("county")

# data cleaning
us_county_cleaned <- us_county %>%
  rename("county" = "subregion",
         "state" = "region") %>%
  mutate(county = gsub("[ˆ0-9A-Za-z///’ ]", "", county)) %>%
  mutate(county = gsub(" ", "", county)) %>%
  mutate(county =
           replace(county,
                   county == "stlouiscity", "stlouis")) %>%
  mutate(county =
           replace(county,
                   county == "yellowstonenational", "yellowstone")) %>%
  mutate(county =
           replace(county,
                   county == "oglaladakota", "oglalalakota")) %>%
  mutate(state = gsub("[^0-9A-Za-z///’ ]", "", state)) %>%
  mutate(state = gsub(" ", "", state))

us_county_cleaned <- us_county_cleaned %>%
  mutate(county_state =
           paste0(tolower(county), "_", tolower(state))) %>%
  select(county_state, state, long, lat, group)

######## Create Population Density by County Dataset ########

density_by_county <- left_join(x = us_county_cleaned,
                               y = pop_density_cleaned,
                               by = "county_state")

summary(density_by_county$density_mi2)

######## Mapping Population Density by County ########

# customized font size
my_theme2 <- theme(plot.title = element_text(size = 20),
                   plot.subtitle = element_text(size = 16),
                   axis.title.x = element_text(size = 14),
                   axis.text.x = element_text(size = 12),
                   axis.title.y = element_text(size = 14),
                   axis.text.y = element_text(size = 12),
                   legend.title = element_text(size = 14),
                   legend.text = element_text(size = 12))

map_density_by_county <- ggplot() +
  geom_polygon(data = density_by_county,
               aes(x = long, y = lat,
                   fill = density_mi2, group = group),
               color = "black", linewidth = 0.1) +
  scale_fill_gradient(name = "population \nper sq. mile",
                      low = "lightblue",
                      high = "darkblue",
                      breaks = c(1, 10, 100, 1000, 10000, 50000),
                      trans = "log10") +
  coord_fixed(1.3) +
  labs(title =
         "Population Density by County of the Contiguous United States",
       x = "longitude",
       y = "latitude") +
  my_theme2
map_density_by_county

ggsave(filename = "./Figures/mapping/map_density.png",
       plot = map_density_by_county,
       width = 16, height = 9, units = "in",
       dpi = 600)

######## Add Sighting Points ########

# load cleaned ufo dataset
ufo <- read.csv(file = paste0(dir, "ufo_v3_emotion.csv"),
                encoding = "UTF-8")

# customize the size of points by frequency
ufo <- ufo %>%
  mutate(coords = paste0(city_latitude, ", ", city_longitude))

freq_df <- ufo %>% count(coords, sort = TRUE)
freq_dictionary <- by(freq_df, freq_df$coords, FUN = function(x) x$n)

ufo <- ufo %>%
  mutate(freq = as.numeric(freq_dictionary[coords]))

# add sighting as points
map_add_sightings <-
  map_density_by_county +
  geom_point(data = ufo %>%
               filter(state != "AK" & state != "HI"), # contiguous US
             aes(x = city_longitude, y = city_latitude, size = freq),
             shape = 21,
             color = "black",
             fill = "#E41A1C", # set1 red
             alpha = 0.05) +
  scale_size_area(name = "number of reports \nat the coordinate",
                  breaks = c(1, 10, 100, 500, 700)) +
  labs(subtitle = "With Points as UFO Sighting Reports") +
  my_theme2
map_add_sightings

ggsave(filename = "./Figures/mapping/map_sightings.png",
       plot = map_add_sightings,
       width = 16, height = 9, units = "in",
       dpi = 600)
```

APPENDIX B

Appendix of Supplemental Figures

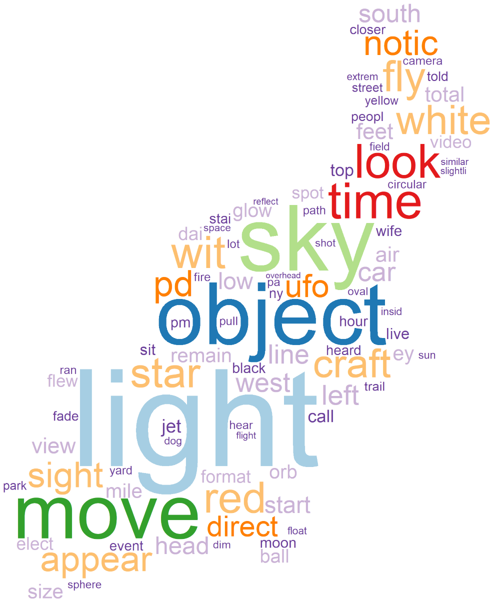


Figure B.1: Word Cloud of UFO Sightings Reports in Northeastern United States

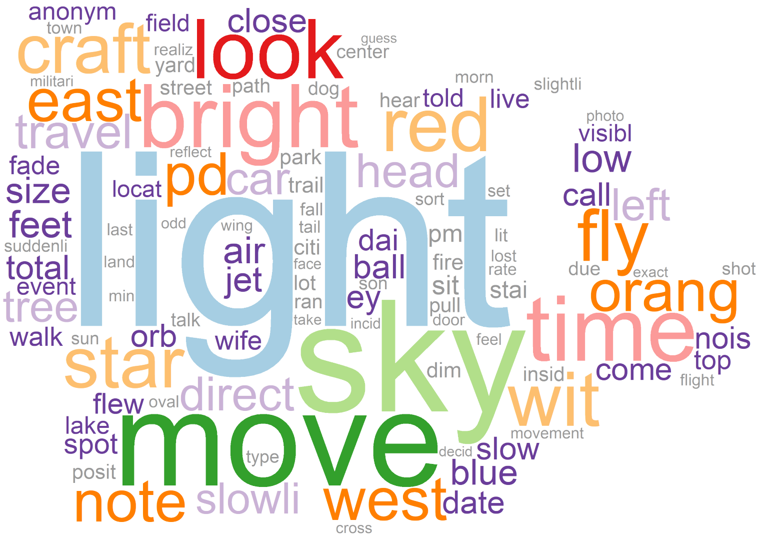


Figure B.2: Word Cloud of UFO Sightings Reports in Midwestern United States

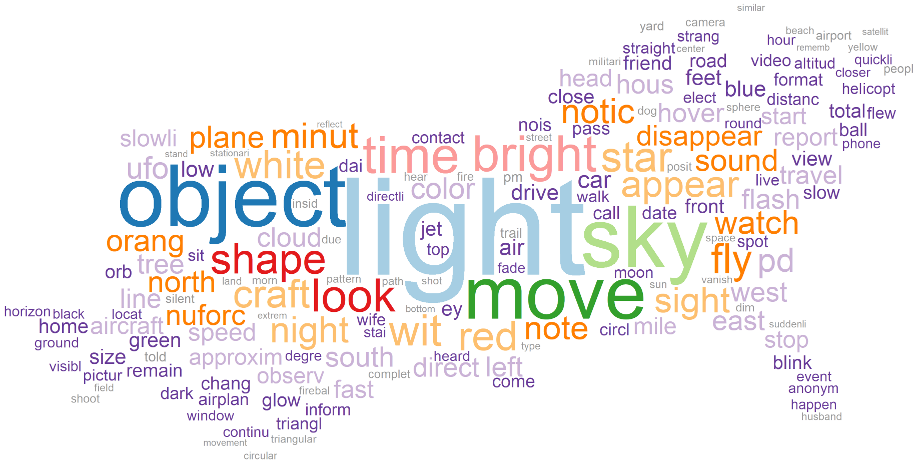


Figure B.3: Word Cloud of UFO Sightings Reports in Southern United States

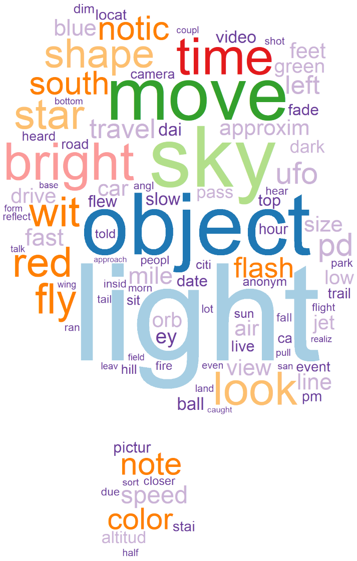
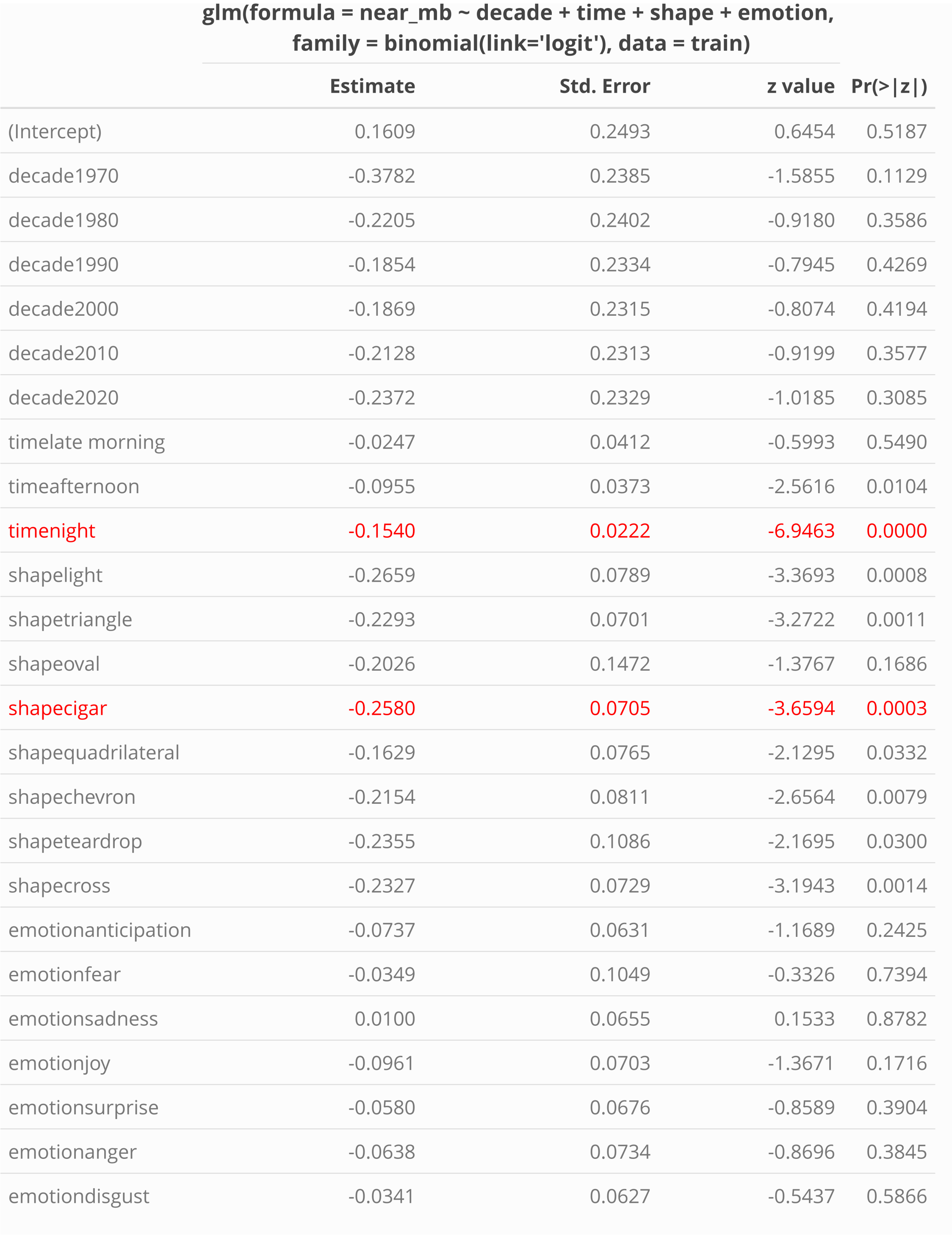
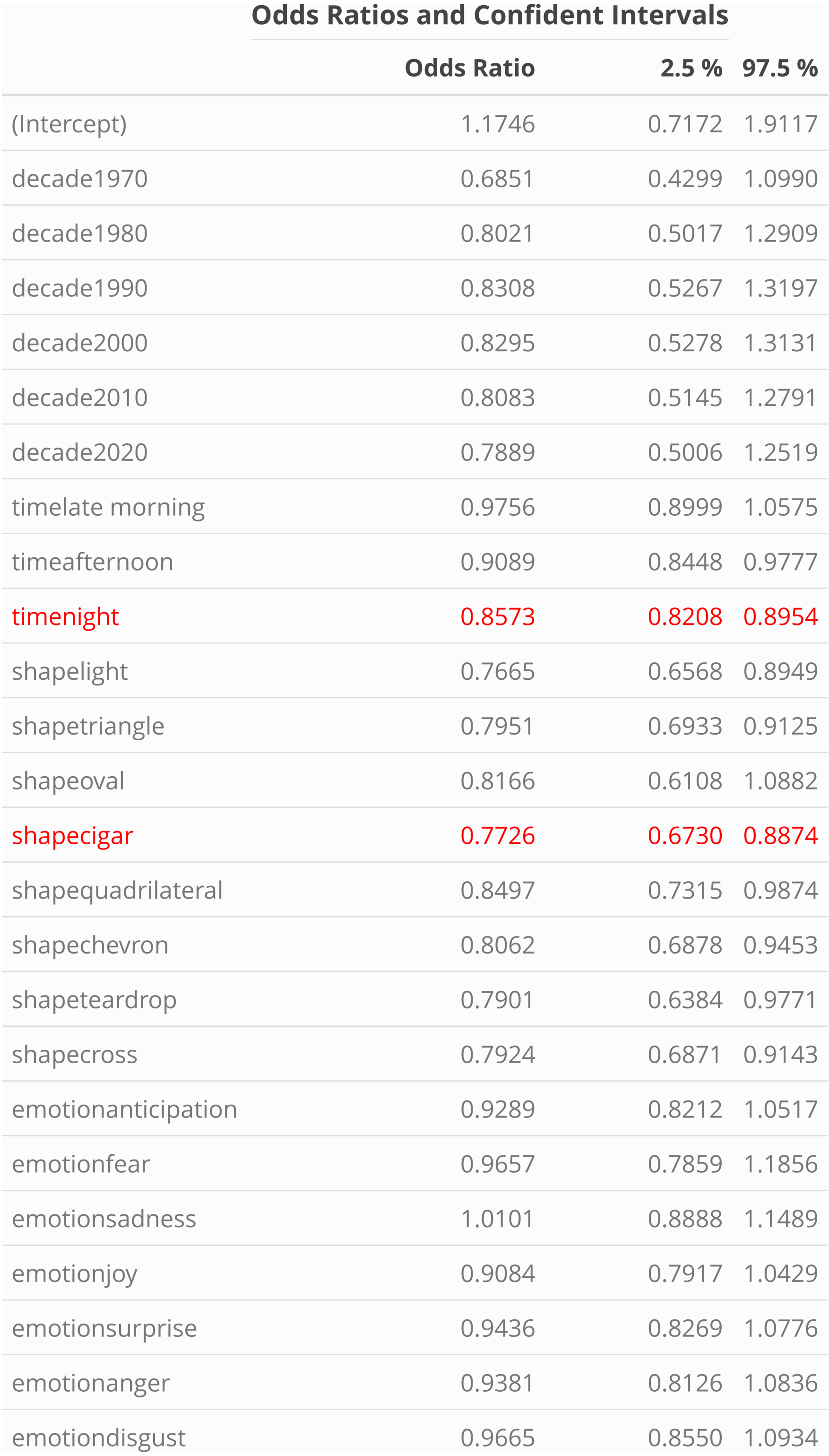


Figure B.4: Word Cloud of UFO Sightings Reports in Western United States



	glm(formula = near_mb ~ decade + time + shape + emotion, family = binomial(link='logit'), data = train)			
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.1609	0.2493	0.6454	0.5187
decade1970	-0.3782	0.2385	-1.5855	0.1129
decade1980	-0.2205	0.2402	-0.9180	0.3586
decade1990	-0.1854	0.2334	-0.7945	0.4269
decade2000	-0.1869	0.2315	-0.8074	0.4194
decade2010	-0.2128	0.2313	-0.9199	0.3577
decade2020	-0.2372	0.2329	-1.0185	0.3085
timelate morning	-0.0247	0.0412	-0.5993	0.5490
timeafternoon	-0.0955	0.0373	-2.5616	0.0104
timenight	-0.1540	0.0222	-6.9463	0.0000
shapelight	-0.2659	0.0789	-3.3693	0.0008
shapetriangle	-0.2293	0.0701	-3.2722	0.0011
shapeoval	-0.2026	0.1472	-1.3767	0.1686
shapecigar	-0.2580	0.0705	-3.6594	0.0003
shapequadrilateral	-0.1629	0.0765	-2.1295	0.0332
shapechevron	-0.2154	0.0811	-2.6564	0.0079
shapeteardrop	-0.2355	0.1086	-2.1695	0.0300
shapecross	-0.2327	0.0729	-3.1943	0.0014
emotionanticipation	-0.0737	0.0631	-1.1689	0.2425
emotionfear	-0.0349	0.1049	-0.3326	0.7394
emotionsadness	0.0100	0.0655	0.1533	0.8782
emotionjoy	-0.0961	0.0703	-1.3671	0.1716
emotionsurprise	-0.0580	0.0676	-0.8589	0.3904
emotionanger	-0.0638	0.0734	-0.8696	0.3845
emotiondisgust	-0.0341	0.0627	-0.5437	0.5866

Figure B.5: Logistic Regression Coefficients with Standard Error, T-Statistic, and P-values



Odds Ratios and Confident Intervals

	Odds Ratio	2.5 %	97.5 %
(Intercept)	1.1746	0.7172	1.9117
decade1970	0.6851	0.4299	1.0990
decade1980	0.8021	0.5017	1.2909
decade1990	0.8308	0.5267	1.3197
decade2000	0.8295	0.5278	1.3131
decade2010	0.8083	0.5145	1.2791
decade2020	0.7889	0.5006	1.2519
timelate morning	0.9756	0.8999	1.0575
timeafternoon	0.9089	0.8448	0.9777
timenight	0.8573	0.8208	0.8954
shapelight	0.7665	0.6568	0.8949
shapetriangle	0.7951	0.6933	0.9125
shapeoval	0.8166	0.6108	1.0882
shapecigar	0.7726	0.6730	0.8874
shapequadrilateral	0.8497	0.7315	0.9874
shapechevron	0.8062	0.6878	0.9453
shapeteardrop	0.7901	0.6384	0.9771
shapecross	0.7924	0.6871	0.9143
emotionanticipation	0.9289	0.8212	1.0517
emotionfear	0.9657	0.7859	1.1856
emotionsadness	1.0101	0.8888	1.1489
emotionjoy	0.9084	0.7917	1.0429
emotionsurprise	0.9436	0.8269	1.0776
emotionanger	0.9381	0.8126	1.0836
emotiondisgust	0.9665	0.8550	1.0934

Figure B.6: Odds Ratios and Confident Intervals

REFERENCES

[Bar23] Julian E. Barnes. “Many Military U.F.O. Reports Are Just Foreign Spying or Airborne Trash.” https://www.nytimes.com/2022/10/28/us/politics/ufo-military-reports.html, 2023.

[Cha14] Science Channel. “What Disabled Missiles at Malmstrom Air Force Base?” https://youtu.be/VgziDyPSUog, 2014.

[Dee21] James M. Deem. “The What’s What of UFOs: The Five Most Common Shapes.” https://jamesmdeem.com/stories.ufo.shapes.html, 2021.

[Evi11] UFO Evidence. “The Washington State Elk Abduction.” http://www.ufoevidence.org/cases/case14.html, 2011.

[Fil10] George Filer. “Elk abducted by UFO.” https://www.mufonohio.com/mufono/Washington%20Elk%20Abducted%20by%20UFO.html, 2010.

[Kbh09] Kbh3rd. “Ogallala saturated thickness 1997-sattk97-v2.svg.” https://commons.wikimedia.org/w/index.php?title=File:Ogallala_saturated_thickness_1997-sattk97-v2.svg&oldid=708987839, 2009. CC BY-SA 3.0.

[Kea17] Leslie Kean. “Glowing Auras and ‘Black Money’: The Pentagon’s Mysterious U.F.O. Program.” https://www.nytimes.com/2017/12/16/us/politics/pentagon-program-ufo-harry-reid.html, 2017.

[Pau22] David Paulides. “Missing 411: The UFO Connection.” https://www.amazon.com/gp/video/detail/B0B8JGPHML, 2022.

[Pri21] Hannajane Prichett. *“Understanding Patterns of Extraterrestrial Phenomena: An Exploratory Spatial Analysis of UFO Sightings Throughout the Contiguous United States from 1910-2014.”*. Master’s thesis, The University of Arizona, 2021.

[Ren23] Tim Renner. “UFO sightings.” https://data.world/timothyrenner/ufo-sightings, 2023.

[Sha22] Jazz Shaw. “The Hudson Valley UFOs: How the Media Reacted to a 1980s UFO Flap.” https://thedebrief.org/the-hudson-valley-ufos-how-the-media-reacted-to-a-1980s-ufo-flap/, 2022.

[SK00] Robert Salas and Jim Klotz. “The Malmstrom AFB UFO/Missile Incident.” https://www.cufon.org/cufon/malmstrom/malm1.htm, 2000.

[SR22] Julia Silge and David Robinson. *Text Mining with R: A Tidy Approach.* O’Reilly, 2022. Chapter 2 Sentiment analysis with tidy data.

[Tra19] U.S. Department of Transportation. "Military Bases." https://public.opendatasoft.com/explore/dataset/military-bases/information/, 2019.

[Wei22] Gregor Weichbrodt. "Convert Wiki Tables to CSV." https://wikitable2csv.ggor.de, 2022.

[Wik20] Wikipedia. "Image of the Phoenix Lights newspaper article from USA Today." https://en.wikipedia.org/w/index.php?title=File:PhoenixLights1997model.jpg&oldid=951876080, 2020.

[Wik23a] Wikipedia. "Black triangle (UFO)." https://en.wikipedia.org/w/index.php?title=Black_triangle_(UFO)&oldid=1156592520, 2023.

[Wik23b] Wikipedia. "Colour." https://simple.wikipedia.org/w/index.php?title=Colour&oldid=8784805, 2023.

[Wik23c] Wikipedia. "List of regions of the United States." https://en.wikipedia.org/w/index.php?title=List_of_regions_of_the_United_States&oldid=1156309693, 2023.

[Wik23d] Wikipedia. "List of United States counties and county equivalents." https://en.wikipedia.org/w/index.php?title=List_of_United_States_counties_and_county_equivalents&oldid=1154738119, 2023.

[Wik23e] Wikipedia. "Malmstrom Air Force Base." https://en.wikipedia.org/w/index.php?title=Malmstrom_Air_Force_Base&oldid=1151043382, 2023.

[Wik23f] Wikipedia. "Phoenix Lights." https://en.wikipedia.org/w/index.php?title=Phoenix_Lights&oldid=1156931545, 2023.

[Wik23g] Wikipedia. "Project Blue Book." https://en.wikipedia.org/w/index.php?title=Project_Blue_Book&oldid=1146419552, 2023.

[Wik23h] Wikipedia. "Unidentified flying object." https://en.wikipedia.org/w/index.php?title=Unidentified_flying_object&oldid=1156263789, 2023.

[Zes19] Dave Zes. *widals: Weighting by Inverse Distance with Adaptive Least Squares*, 2019. R package version 0.6.1.