Exploratory Data Analysis on Reported UFO Sightings

By Alexis Ang, Lily Hallmark, Tim Ilchene
STA 141B Final Project WQ 2019
Professor Nick Ulle

Table of Contents

Part 1: Introduction

The sky has fascinated humankind since the beginning of civilization. The thought of life beyond our planet has found its way into every facet of popular culture. From books to movies to magazines and more, the influence of the *possibility* of alien life and even first contact has captivated generations of people. Since the 1947 Roswell, New Mexico incident, there has been a piqued interest in alien sightings in recent years. Movies such as *Close Encounters of the Third Kind*, *Alien*, *E.T. the Extra-Terrestrial*, and countless more continue to entice viewers and question the probability of life beyond our planet.

The internet has enabled alien believers and conspiracy theorists to connect with each other and share evidence of what they believe proves extraterrestrial beings have visited our planet. One such group, the National UFO Reporting Center (NUFORC), currently based in a former nuclear missile site west of Spokane, Washington has meticulously cataloged Unidentified Flying Objects (UFO) sightings since its founding in 1974. They operate a 24-hour hotline for people to report any UFO activity that they have experienced.

We wanted to use the data the NUFORC has compiled over the years to see if we could find any trends or patterns within the data. We were curious if certain days of the year, times of day, or locations were more likely to have a reported UFO sighting. Additionally, because of its prevalence in the media, we wanted to determine if there was a correlation between sci-fi movie releases and UFO sightings.

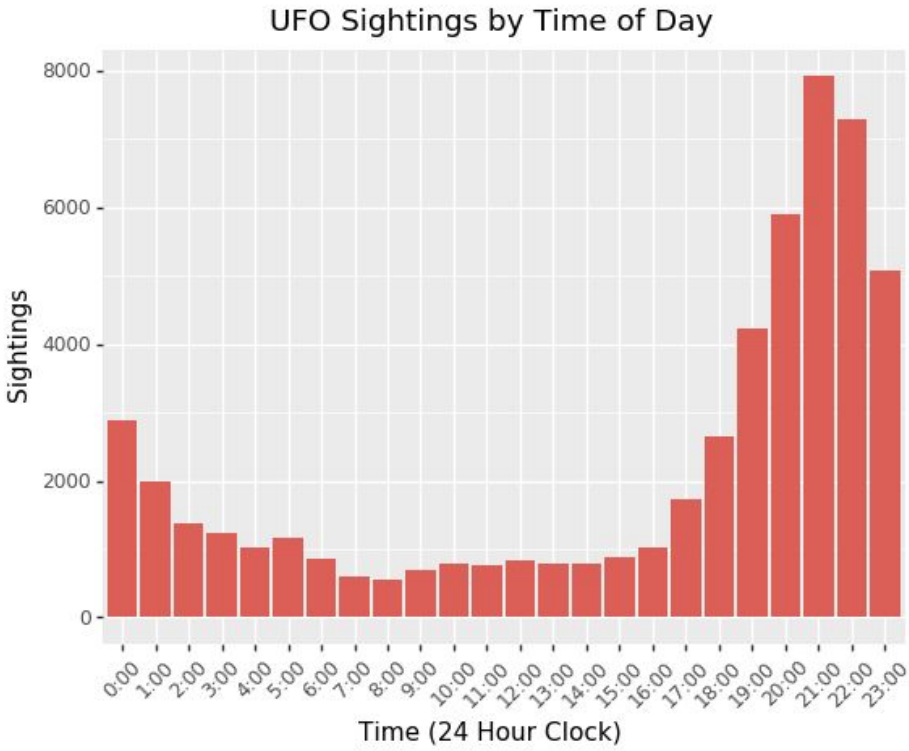
To accomplish our goals, we first contacted the director of NUFORC to gain permission to use their data; at first, he seemed open to allowing us to use their data (with some caveats: namely that we delete the data after), though he failed to respond to our follow up emails. Consequently, as we were still determined to use UFO data, we found a data set on Kaggle that someone had scraped from the NUFORC website which contained data from the early 1900s to 2013. We also collected movie data from The Movie Database on Science Fiction movie releases as well as from various government websites that documented state populations, unemployment rates, binge drinking prevalence, and opioid prescription rates.

To begin, we cleaned the UFO data to pertain only to sightings within the United States within the 14-year time period from 1999 to 2013. Then, we conducted summary statistics to

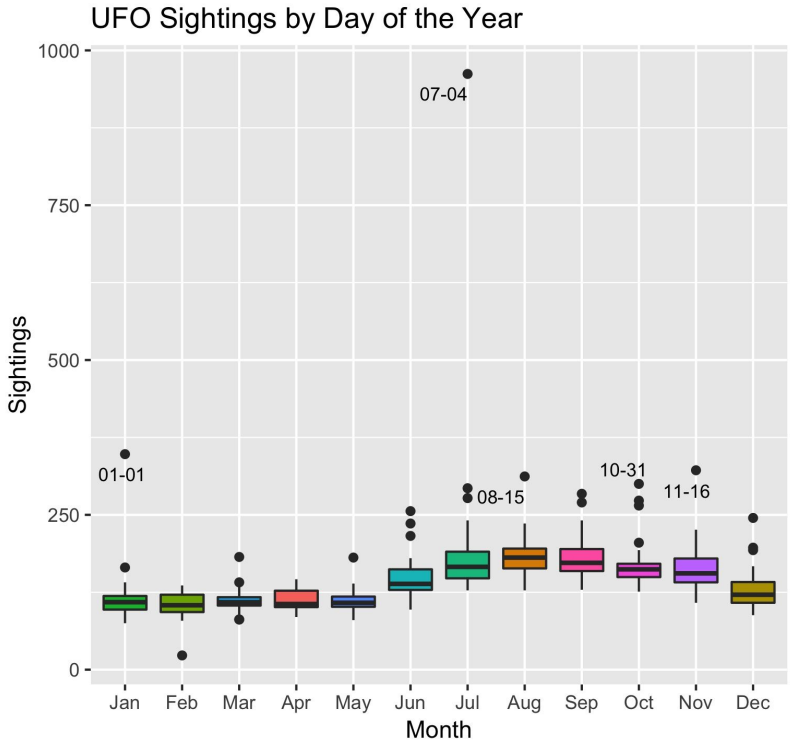
determine basic metrics for the data. We found that the longest reported sightings were around 1.5 years and that the shortest reported sighting occurred for only 0.01 seconds. Afterward, we combined our various datasets in order to examine the data in a more detailed context. We looked at relationships between times of UFO sightings and frequencies, the prevalence of opioid prescriptions and binge drinking, as well as recent science fiction movie releases. While these reported UFO sightings have been meticulously documented by the NUFORC, we take all of our results with a grain of salt and cannot definitively state much beyond the correlations we discover. Our goal is not to determine whether these reported UFO sightings are direct evidence of extraterrestrial life; rather, we would like to point out common factors of reported sightings.

Part 2: Analysis by Date and Time

We wanted to determine if certain months or days in the year saw an increase in UFO sightings when compared to other times of the year. All the graphs below encompass our 14 year period (1999-2013). The bar chart titled 'UFO Sightings by Time of Day' shows that reported sightings typically occur at later times when the night sky allows for greater visibility of illuminated objects, specifically from 7 pm to 1 am, with sightings highly reported between 9 pm and 11 pm. This is a time when most people are home and a time when people are more likely to be inebriated, which is a possible factor for reported sightings.



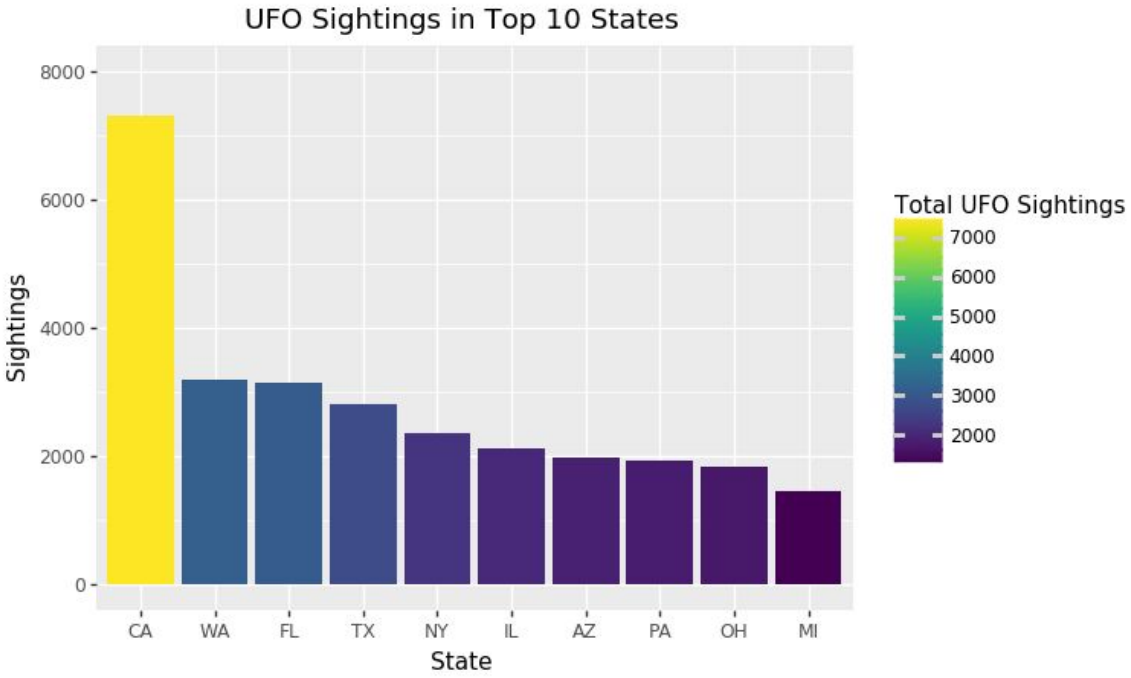
After analyzing the times of day when most sightings occurred, we became interested in what specific days UFO sightings were most likely to occur. The box plot below, 'UFO Sightings by Day of the Year', collates information from 1999-2013 and displays the general range for sightings per month, along with the outliers for the month. An interesting finding from the box plot is that there are specific days in January and July that have a much larger amount of reported sightings. Upon further investigation, the two days that stuck out as the largest outliers were July 4th and January 1st. Over the 14 year period, there were 962 reported UFO sightings on the 4th of July and 348 reported sightings on New Year's day. The box plot also shows that in the latter half of the year there are more reported sightings than in the beginning of the year.



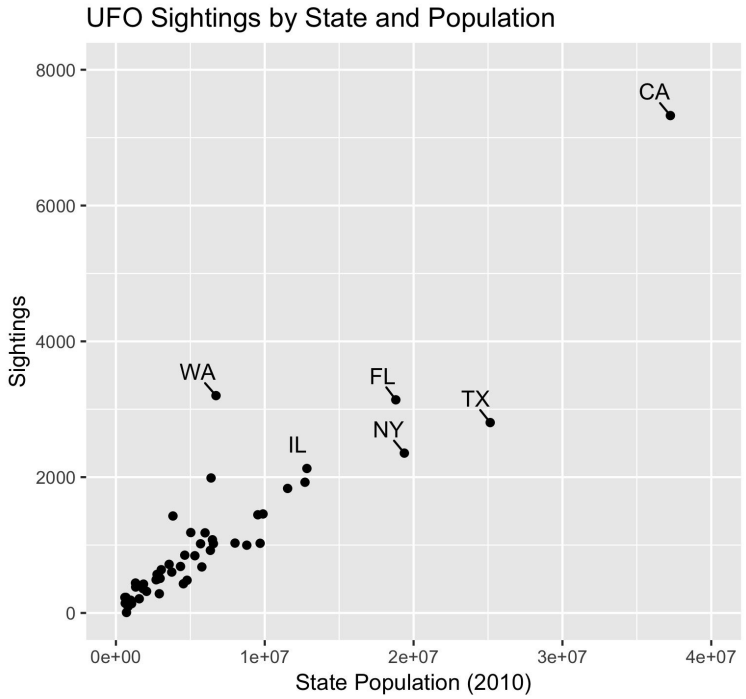
Part 3: Locations of Reported UFO Sightings

We wanted to know which states had the most reported UFO sightings over our 14 year time period. We created the following bar chart which reveals the top 10 states for reported UFO sightings. We immediately noticed that California has significantly more reported UFO sightings

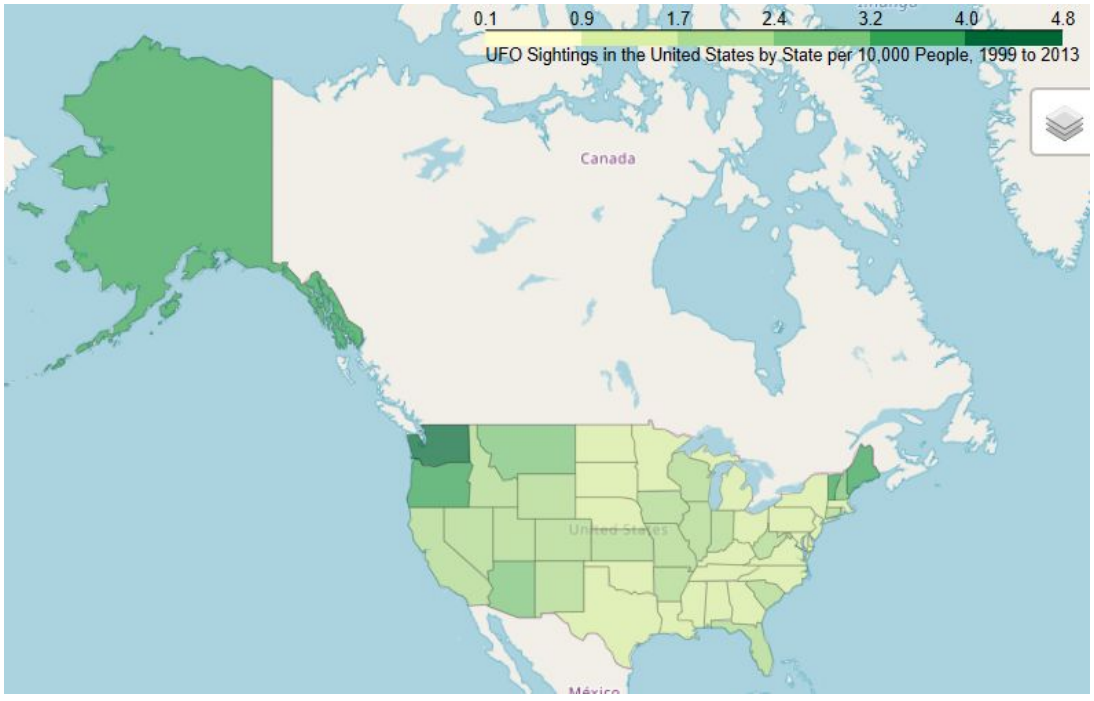
than any other state. It has approximately two times more reported sightings than the next highest state, Washington. The next three states include Florida, Texas, and Washington. Four out of the top five states in terms of reported UFO sightings are also the largest four states in the United States by population. This bar chart confirms to us that a state's population is a significant factor in reported UFO sightings per state. In our later analysis, we directly address this issue by calculating the number of sightings per 10,000 people in each state.

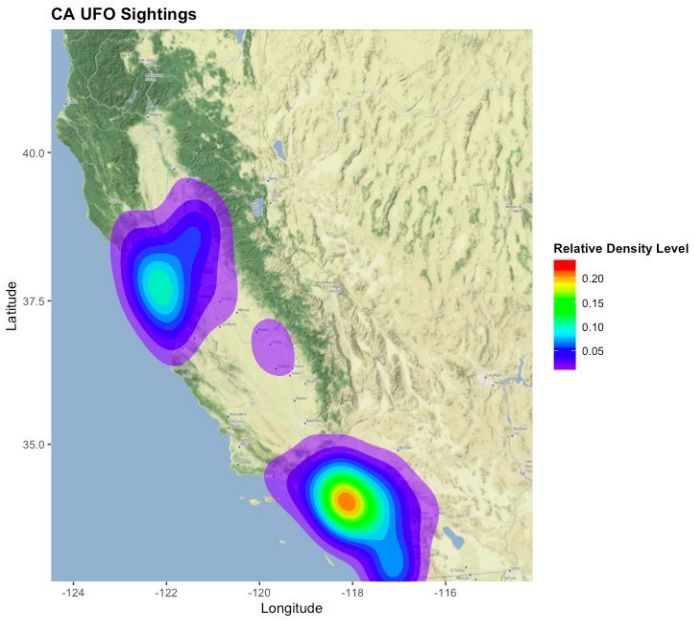


The scatter plot below, 'UFO Sightings by State and Population', shows the correlation between a state's population and reported UFO sightings documented by the NUFORC. We obtained the state population by scraping the table from Wikipedia. We created a scatter plot to represent the states where sightings are reported because it allowed us to simultaneously display sightings and state population. The plot presents a strong positive correlation between state population and reported sightings. Washington is a slight exception where there seems to be more reported UFO sightings than the other states based off of a similar population. This may be due to the fact that the NUFORC is based in Washington.



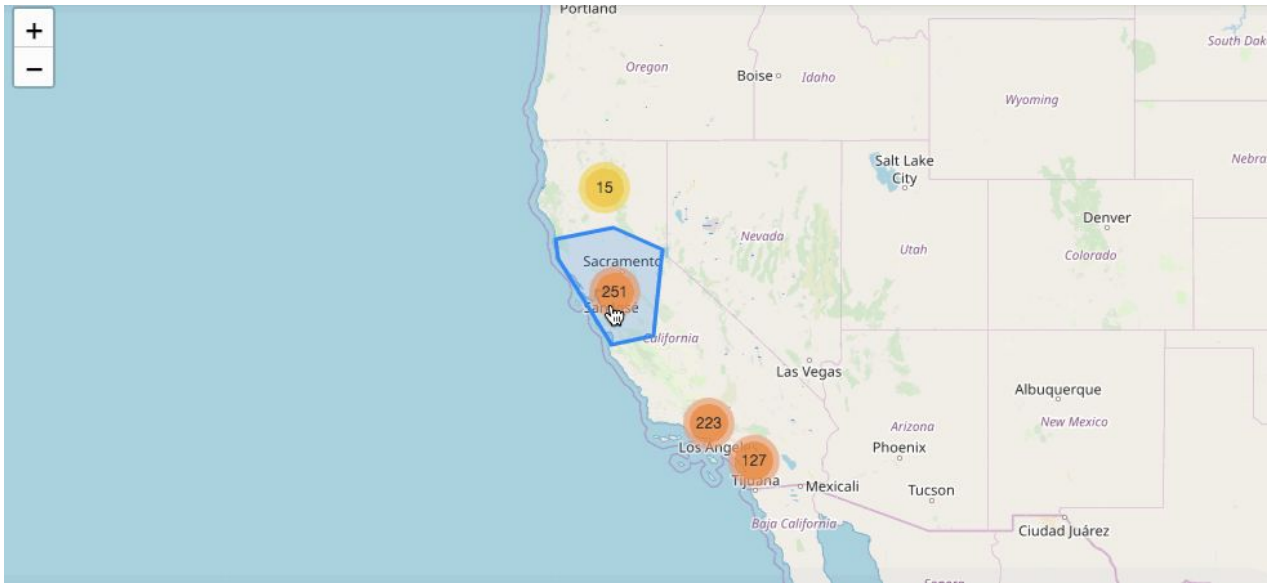
The following map is a screenshot of a choropleth map (which can be found in our GitHub repository, UFOper10000.html) which shows reported UFO sightings in the United States per 10,000 people from 1999-2013. This map agrees with some of our earlier findings in the above scatter plot. The state of Washington has the highest number of reported sightings per people. Oregon, Alaska, and Maine are some of the other states with the highest amount of findings per people. Another interesting find from the map is that states in the middle of the country have a much smaller amount of reported UFO sightings per people when compared to states in the Northeast or in the West.





The above map, 'CA UFO Sightings', is a map of California that shows the location of reported UFO sightings in terms of density. Observing the map, we notice that urban areas of the state with higher populations have more reported UFO sightings. This includes the San Francisco Bay Area, Los Angeles, and the Fresno area of the Central Valley.

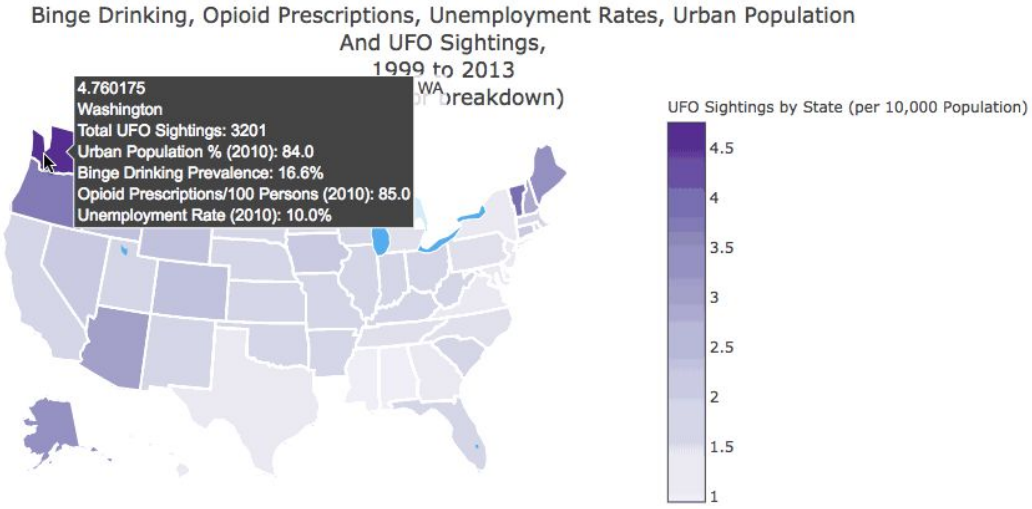
We created two interactive cluster maps that showcase the locations of UFO sightings in California in 2013, as well as reported sightings in the U.S. for every July 4th from 1999 to 2013. These maps can be found in the 'UFO Cleaning.ipynb' file in our project repository.



Part 4: Analysis by Statewide Metrics

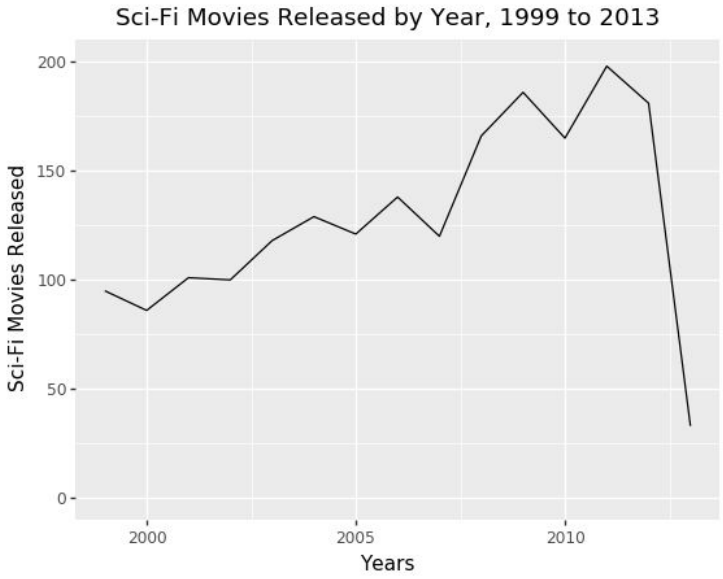
We were interested in what other factors within a state affect UFO sightings, other than population. To determine this, we collected data on binge drinking prevalence, opioid prescriptions, urban population percentage, and unemployment rates in 2010. We found that for every state with over 3 sightings per 10,000 people, the binge drinking prevalence (%) was at least 15%. In 2010, the nationwide unemployment rate was 9.6%; for three of the six states with UFO sightings of more than 3 per 10,000, their unemployment rate was over the national average. While not a definitive statement of causation, it is an interesting metric we noticed in our data. We also found that urban population percentage was not indicative of total UFO sightings. Out of the six states with more than three sightings per 10,000 people, four had more than 50% of the population living in urban areas. We were unable to determine any relationship between opioid prescription rates and UFO sightings, but we found it to be an interesting enough statistic to leave in our plot.

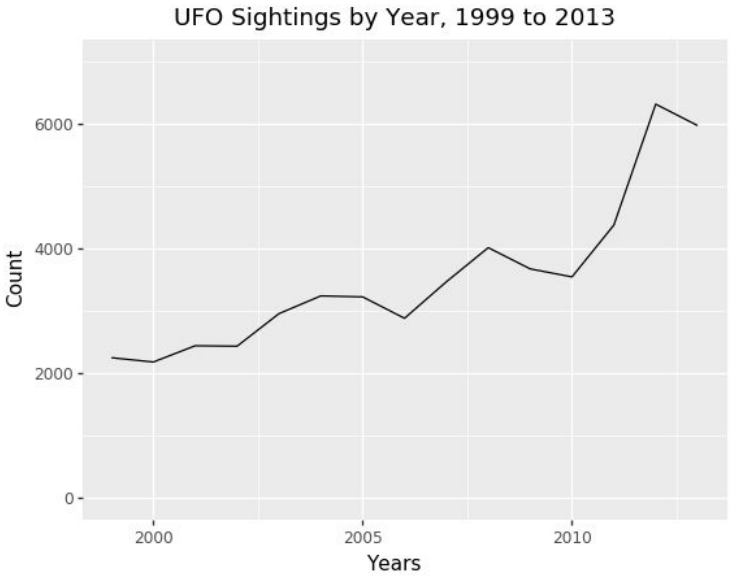
Below is a gif of an interactive plot, colored by the UFO sightings per 10,000 people, which displays the above information for each state. The full interactive plot can be found in our repository, under the 'UFO Cleaning.ipynb'.



Part 5: Movie Analysis

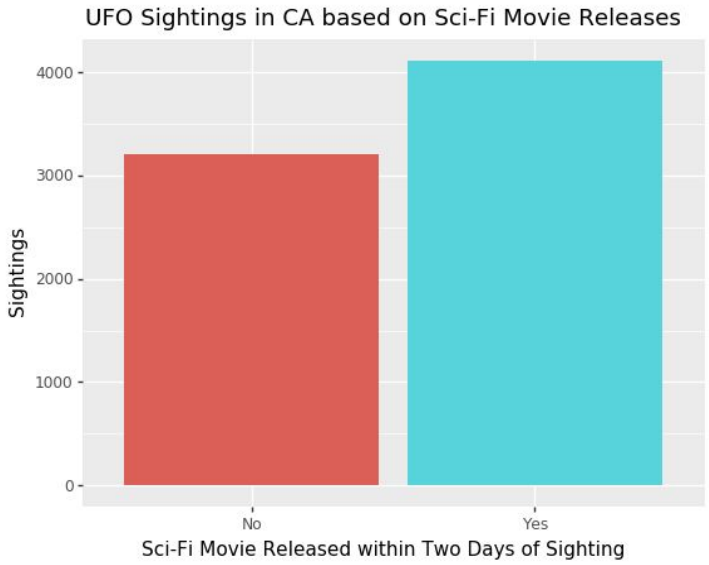
There have been lots of Science Fiction (Sci-Fi) movies over the years solely based on the possibility of extraterrestrial life. As such, we became interested in the possibility of a correlation between the release of Sci-Fi movies and reported UFO sightings. Using an API, we obtained and cleaned data about Sci-Fi movies in our 14 year time period from The Movie Database. The two line charts below detail the number of Sci-Fi movies released between 1999 and 2013 as well as the number of reported UFO sightings per year in that same time range. From the graphs, we can see that over the 14 year time period there was an increase in both the number of Sci-Fi movies released every year and reported UFO sightings every year. There is an anomaly in the Sci-Fi Movie releases in 2013, where there was a huge decrease in the release of movies that year.





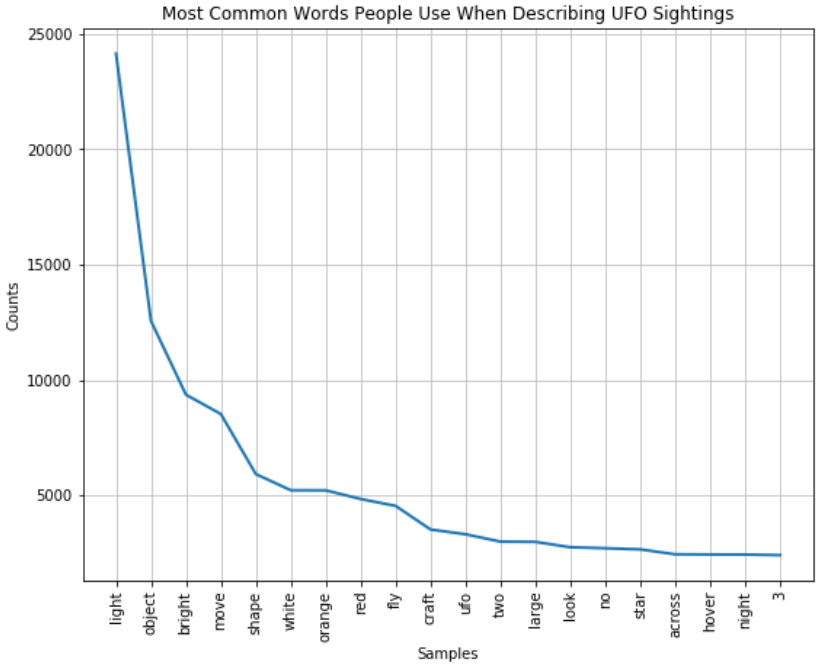
The following bar chart, 'UFO Sightings in CA based on Sci-Fi Movie Releases', shows whether or not a Sci-Fi movie was released within two days of a reported UFO sighting in California. The graph informs us that there is a slightly greater amount of sightings reported within two days of a movie release than not. However, it is important to note that each year there were between 100 and 200 Sci-Fi movie releases. This huge amount of movies means that there is a high probability that a reported UFO sighting happened to coincide with a movie release day. Regardless, we see a correlation between the two (which does not imply causation), and as

CGI/special effects continue to improve perhaps people will have a harder time telling fact from fiction.



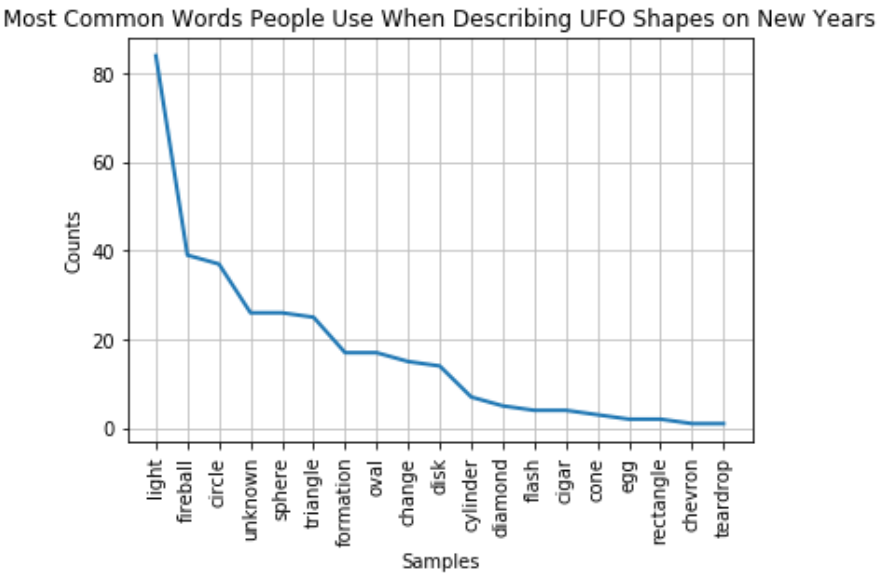
Part 6: Word Frequencies

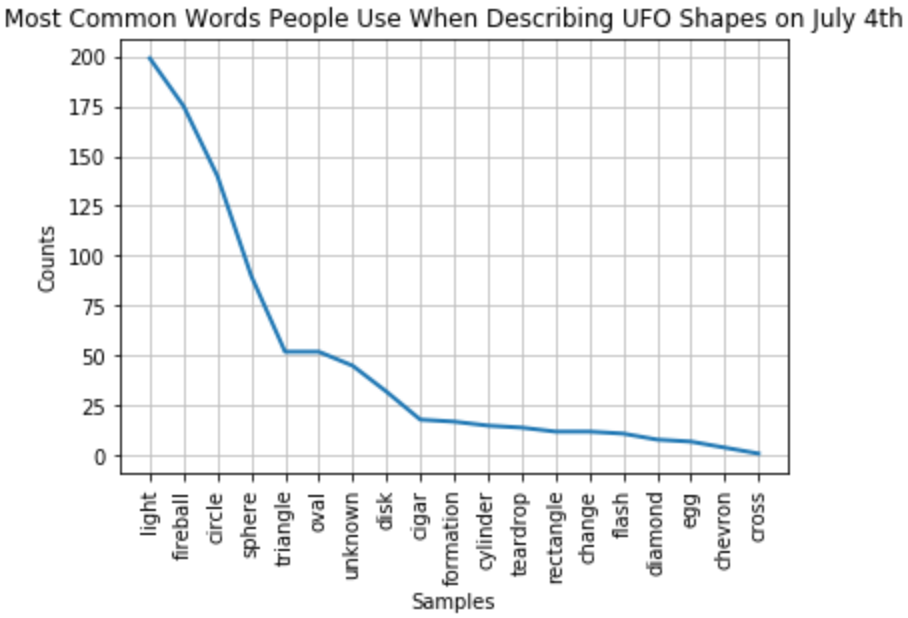
We used the *comments* category from the UFO data set to perform a word frequency analysis on the most common words people use to describe their UFO sighting. The words found in the chart below, specifically “light,” “bright,” “white,” “orange,” and “red,” show that some of the most popular words people use to describe their reported UFO sighting revolve around light

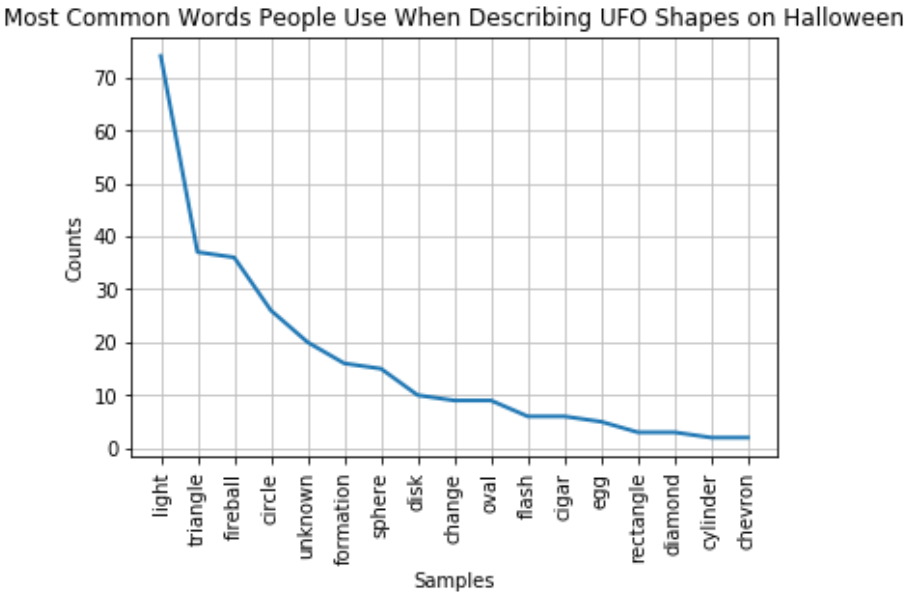


and color. We also see large counts of the words “object” and “craft,” possibly indicating they saw not just a bright light but an actual object in the sky.

The UFO data also included a *shape* category, which is the shape used to describe the reported UFO sightings. To determine what the most popular shapes are for reported UFO sightings we looked at three different holidays: New Year’s Day, the 4th of July, and Halloween (New Year’s Day and the 4th of July are also the two days with the highest amount of reported UFO sightings; Halloween is the fifth) and performed individual word frequency analyses. Both New Year’s Day and the 4th of July showed a similar display of words used to describe the shape seen, in fact, the top 20 words are in almost the exact same order for both graphs. The top 20 most common words on Halloween are nearly identical to the words used on New Year’s Day and the 4th of July. However, on Halloween the word ‘triangle’ is the 2nd most common word, compared to the two other days where ‘triangle’ is only the 6th most common word. Interestingly, the word ‘light’ is used to describe the shape of the UFOs. This is interesting because light is not a shape, but we acknowledge that ‘lights in the sky’ and similar variations are a common way to describe UFO sightings. The use of the word ‘fireball’ is also interesting because both the 4th of July and New Year’s Day are common days that fireworks appear in the sky, and the word ‘fireball’ is an accurate description of the shape of a firework. However, ‘fireball’ is the 3rd most common word used on Halloween, and fireworks don’t typically appear in the sky on Halloween. This analysis of the shapes used to describe UFO sightings on New Year’s Day, the 4th of July, and Halloween appears to show that there is little change in how people describe their UFO sightings based off of the day.







Part 7: Conclusions and Considerations

Overall, our analysis of the UFO data from 1999 to 2013 garnered mixed, yet interesting results. We found that reported UFO sightings are most likely to occur at night, specifically around 10 or 11 pm. We also found that there are certain days that have much more reported UFO sightings, such as New Year's Day or the 4th of July. One of our speculations for this is

that on those two particular days people are out late into the night and typically drink a good amount of alcohol which could impair judgment.

We also found a high correlation between state populations and how many reported UFO sightings there are per state. The state of Washington was a slight anomaly; however, we believe this could be related to the fact that the National UFO Reporting Center is located in Washington. When we performed a more in-depth analysis of the locations of reported UFO sightings within California we found that in areas with higher densities of people there were also more reported UFO sightings from there as well.

When we included metrics on binge drinking, opioid prescriptions, urban population, and unemployment rate, we found some slight correlations between drinking, unemployment, and reported sightings. However, we cannot state that there are any direct causal relationships between these metrics without further investigation into confounding variables, possible seasonal trends, and deeper statistical testing.

Our analysis of Science Fiction movie release with reported UFO sightings didn't produce very fascinating results. About 100-200 Science Fiction movies were released every year so it's very common for a reported UFO sighting to be within a two-day range of a movie release date. We did notice that reported durations for UFO sightings increased on average if a sci-fi movie was released recently.

We performed word frequency analyses on the *comments* and *shapes* categories of the UFO data set to see if there were any trends with words people were using to describe their reported UFO sightings. Some of the most popular words people used in *comments* to describe their reported UFO encounter dealt with words surrounding light (bright, light, red, orange). A common word also found in the *comments* section was ufo. Some of the most common words used in *shapes* were triangle, disk, fireball, sphere, cigar, and egg.

Overall, our analysis of the data provided interesting conclusions, but simultaneously raised more questions of the relationships between reported sightings, human, and environmental elements that could be affecting the observations.

Part 8: Works Cited

Age-adjusted prevalence of binge drinking among adults aged 18 years and older, by state, United States, 2015. Center for Disease Control and Prevention. https://www.cdc.gov/alcohol/data-table-text.htm#intensity

Annual Unemployment Rates by State. Iowa State University. https://www.icip.iastate.edu/tables/employment/unemployment-states

List of U.S. States by Population. Wikipedia. Accessed March 2019. https://simple.wikipedia.org/wiki/List_of_U.S._states_by_population

National UFO Reporting Center. Wikipedia. https://en.wikipedia.org/wiki/National_UFO_Reporting_Center

The Movie db. https://www.themoviedb.org/

The National UFO Reporting Center. http://www.nuforc.org/index.html

UFO Sightings Data set. Kaggle. https://www.kaggle.com/NUFORC/ufo-sightings

Urbanization in the United States. Wikipedia https://en.wikipedia.org/wiki/Urbanization_in_the_United_States

U.S. State Prescribing Rates, 2010. Center for Disease Control and Prevention. https://www.cdc.gov/drugoverdose/maps/rxstate2010.html