Alice Saparov, Alesia Michel, Solomon Freilich, Carlos Cardona

When, What, and Where to be if you want to see a UFO

Overview & Initial Questions

Initially, we had come across a news article about how the Pentagon had formed a task force to investigate UFO sightings. So, after coming across the UFO data set from the National UFO Reporting Center, we were inspired to scientifically analyze the various patterns affiliated with different types of sightings. At the onset of the project, we were interested in analyzing temporal, geographical, and characteristic trends affiliated with each sighting. This included analyzing the location of the sighting, when it was sighted, how long it was seen for, and the shape/characteristics of the sighting. As time progressed, we were also interested in the relationship between real world phenomena and the resulting impact on the number of sightings. Two key phenomena include the end of the Mayan Calendar (2012) which was believed to be the end of the world and Roswell, NM, famed as the site of an alleged UFO crash in 1947.

Data Methodology

The UFO sighting data set is from the National UFO Reporting Center. It is a csv file with over 80,000 reports of UFO sightings from the last century. For the purposes of this analysis, we used a scrubbed version provided by Kaggle, a platform with resources for data scientists.

UFO Sighting Characteristics

In order to analyze the duration of the sightings, the string values in the csv were converted to integers, and stray characters like apostrophes had to be stripped. Next, we selected the duration column from the dataset. We decided that the data would be best represented with a pie chart, and by grouping the data into buckets, we were able to visualize the breakdown of duration times in a very intuitive manner (Fig. 1).

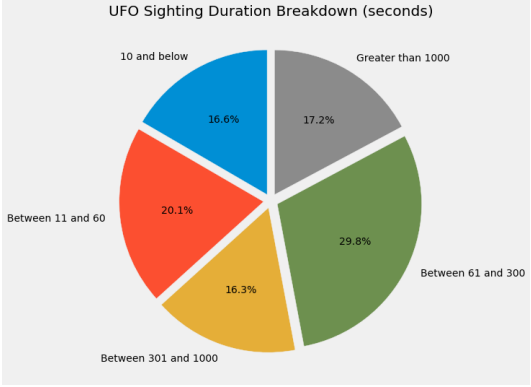

Figure 1. Duration of UFO sightings in seconds.

The results of breaking down UFO sightings by duration were unexpected. We expected the shortest bucket, 10 seconds and below, to be the largest category as logically, the less time one has to view a flying object, the less likely they will identify the object. However, the most common sighting duration was between one to five minutes. One possible explanation is that this time period is short enough for viewers to not have a great chance of identifying the flying object and long enough to give significant attention to the object and not dismiss it as insignificant.

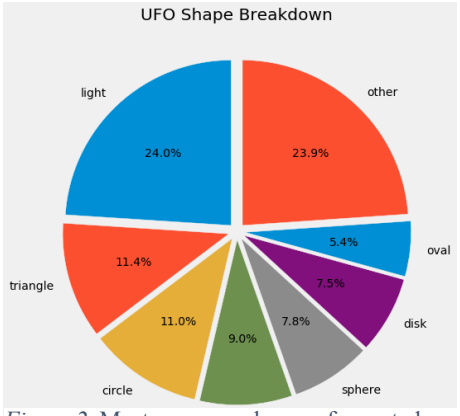

Figure 2. Most common shapes of reported UFO sightings.

We investigated the different shapes of the UFOs by rearranging the data and selecting out relevant column series, grouping the data, and grouping insignificant values into “other categories” (Fig. 2). For visualizing the breakdown of the UFO sightings, we also believed a pie chart would serve as the best representation of the data. Since the data was categorical, the shape data was extracted from the overall table and then grouped by the counts of each shape. Then, all shapes that had insignificant counts were all grouped together in an “other” category in order to gain a comprehensive view of the most common UFO shapes. The results of the UFO sighting shape breakdown analysis were relatively expected. The most common shape was “light” which has been interpreted to mean moving flashes or radiances of light. This is unsurprising as flashes of light are likely to attract more attention than non-illuminated shapes such as dark circles or ovals. Similarly, “fireball” was also among the most common shape descriptions among the UFO sightings. Also, many of the other top categories, “circle”, “sphere”, and “disk”, are very similar, and difference between their frequency in the UFO reports may have been subject to how they were reported by the viewer.

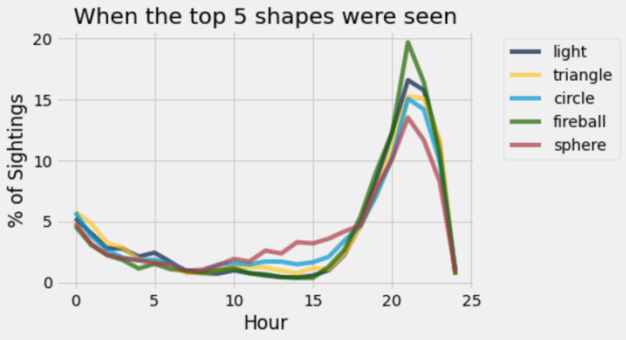

Figure 3. Percent of sightings within each shape category of what time the UFO sighting was reported.

Furthermore, we investigated when the top five UFO shape were seen based on their reported sighting hour (Fig. 3). The top five UFO shapes were scaled within their respective shapes and plotted. The peak percentage of sightings for the top five UFO shapes was at 21 hours or 9 PM. By percent, 'fireball' and 'light' UFO shapes were the most common at night. This makes sense since flashes of light/fire in the sky would be more visible at night compared to during the day. Furthermore, it is much easier to mistake a flash of light or any flying object for a UFO when it's dark out since it is harder to discern details at night. For example, these objects also could have just been planes or comets. In addition, we noticed that during the day, between around 10 AM to 5 PM, sphere, circle, and triangle UFO shapes were reported at a higher percentage than light and fireball shapes. This can be explained by increased visibility during the day, and moreover, flashes of light would be more difficult to notice when it is also light outside.

Temporal Patterns

Next, the data was interpreted temporally. Since each entry within the provided 'datetime' column was written as Month/Day/Year Hour:Minute, we converted the column into a list and used the split function to separate at the "/", ":", and " ". We then created a new table titled "timestable" with individual columns for the hour, month, and year. Each row corresponds to a unique UFO sighting. To understand the temporal patterns of UFO sightings, we looked at the histograms of the frequencies of UFO sightings by hour of day, month, and year.

The year data was binned by every two years, the month data was binned by each month (1-12), and the hourly data was binned by each hour of the day (0-23). By utilizing histograms, we can easily compare differences between bins and pinpoint areas that require further analysis. The sightings by hour histogram revealed that most sightings occur in the evening, particularly after 8PM (Fig. 4). The monthly sightings showed that most sightings occur in July (12%) and the least sightings occur in February (6%) (Fig. 5). However, sightings are generally evenly distributed throughout all months of the year. Finally, the yearly sightings reveal that sightings primarily occurred within recent history, possibly due to improvements in technology and therefore ease of reporting. This increase in sightings could also be attributed to the rapid spread of news and fake news that occurred in recent years, which could make people more aware and 'on the lookout' for UFOs. The most UFO sightings occurred in 2012 (Fig. 6), and we chose to investigate this abnormality.

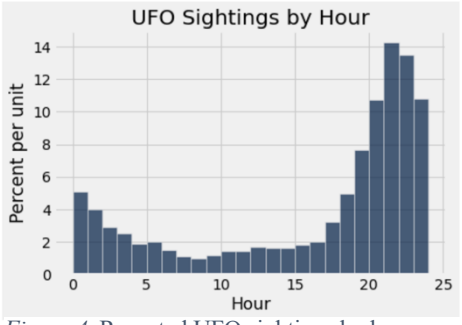

Figure 4. Reported UFO sightings by hour.

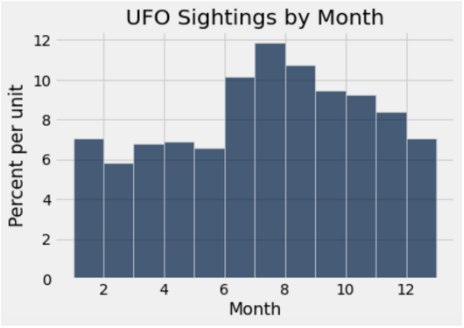

Figure 5. Reported UFO sightings by month.

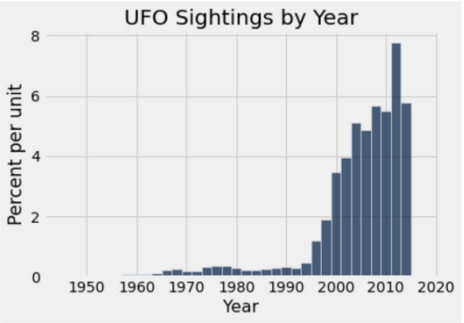

Figure 6. Reported UFO sightings by year.

Statistical Analysis

In order to further investigate this spike in sightings in 2012, we decided to look at the past ten years of UFO sightings reported in the data (2003-2013) (Fig. 7). As noted previously, the number of sightings has been increasing ever since this UFO sighting data started being tracked. So, to try and make our analysis more consistent, we decided to look closely at the past ten years of UFO sightings since this portion of the 'UFO Sightings by Year' histogram had a more even distribution.

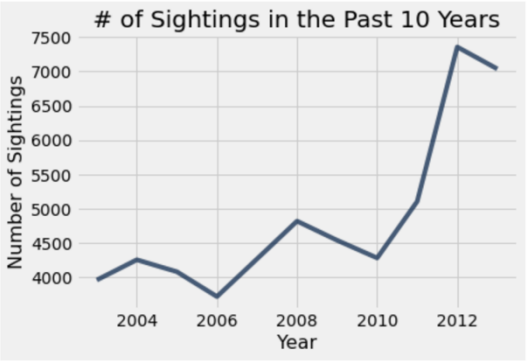

Figure 7. Number of UFO sightings reported from 2003-2013

Mean	SD	2SD	z-score	p-value	95% CI	2012
4858	1164	7185	-2.148	0.016	[2577, 7138]	7357

Table 1. Statistical Analysis regarding the peak in 2012.

We found that in 2012 there were 7357 sightings reported, so we decided to investigate it statistically (Table 1). The average number of UFO sightings per year over the last ten years was 4858 ± 1164 sightings with a 95% confidence interval of (2577, 7138). Furthermore, the number of reported sightings in 2012 was greater than two standard deviations away from the mean (7357>7185). Next, we calculated the z-score which was -2.148 and found its corresponding p-value to be 0.016, and while the distribution was not normal, we decided to investigate this based on the large number of sightings reported. A p-value of 0.016 is very small meaning the probability of having that many sightings in a year is highly unlikely. While the sightings reported in 2012 may require further exploration into their actual significance, we found this peak in sightings very interesting. When coupled with the idea that 2012 was the end of the Mayan calendar, a year when some communities thought that the world would end, it could be possible that people would be hyper alert to flashes of light or random flying objects, and thus report more UFO sightings than normal.

Spatial Patterns

In order to investigate the spatial UFO sighting data we had to remove a few values that were missing location values. We also chose to limit our analysis to the US data.

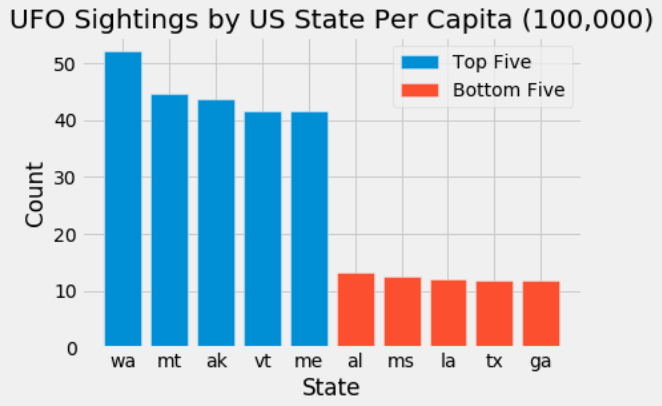

Figure 8. UFO sightings in US states per capita.

Next, because the data was both numerical and categorical, we decided that it would be best represented with bar graphs. Furthermore, in order to scale the sightings reported in each state, we used the latest census population data in order to normalize the data into per capita (Fig. 8). From our results we saw that the top five states had upwards of 40 sightings per 100,000 people. The bottom five states had just over 10 sightings per 100,000 people. Thus, a possible explanation for this could be that certain states see more things fly over them (like planes) based off of their geographical location, which could lead to a higher number of sightings.

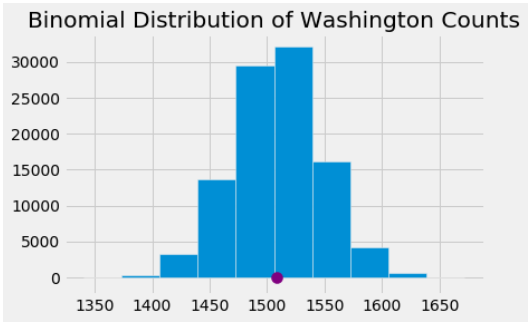

Figure 9. Simulated distribution of sightings in WA.

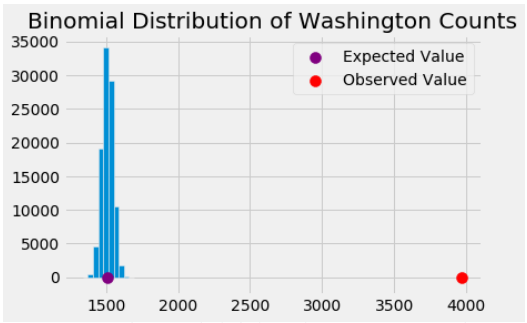

Figure 10. Observed sightings in WA compared to expected.

After exploring state patterns, we wanted to take a closer look at the state of Washington and attempt to conclude how peculiar the results are. Given the population of Washington, about 7.6 million, and the total U.S. population of about 328 million, we would expect the number sightings in the dataset of length 65,114 to be 65,114 * (7.6 / 328) which comes out to about 1,509. This assumes that individuals in all states are equally likely to spot a UFO and are equally likely to report their sighting. We simulated this scenario 100,000 times and is shown as a binomial distribution centered around the expected value (Fig. 9). Then, we compared the observed sightings in Washington, 3,966, to the simulated distribution (Fig.10). This result allowed us to reject the hypothesis that the amount of UFO sightings in Washington follows a binomial distribution, and it implies that there are other mysterious factors contributing to the surprisingly high number of UFO sightings in this state.

We also constructed a heat map using information of sighting reported in states per capita which lets us view the geographical distribution of sightings (Fig. 11). From the heat map, we can see that most of the sightings are from the northwest and northeast of the US, with Alaska also having a high number of sightings.

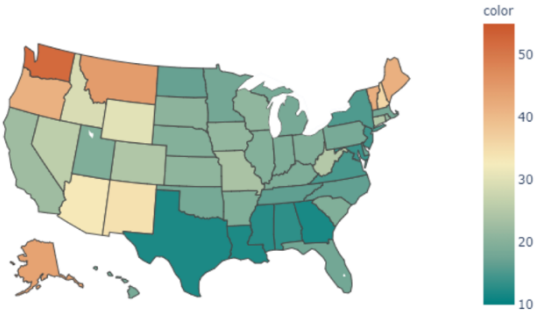

Figure 11. Heat map of UFO sightings in US states.

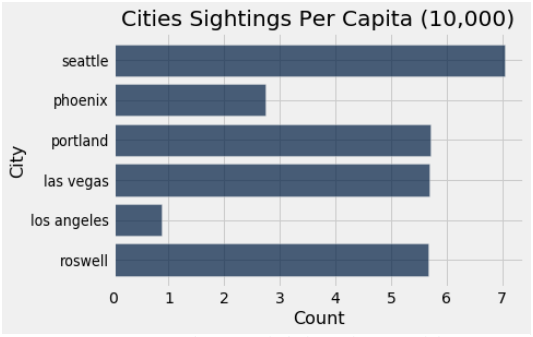

Figure 12. Reported UFO sightings in top cities.

From the city sightings chart (Fig. 12), we included Roswell, NM and compared it to cities with the most sightings per capita, then compared them per 10,000 people. We chose to include Roswell, NM because the city is famous for being the site of an alleged UFO crash. Looking at the chart, we can see that Roswell has a higher or equal number of sightings per 10,000 compared to the other cities. This is surprising because Roswell has a much smaller population at only around 47,600. This means that cities like Seattle have a very large number of sightings considering their large population, and Roswell has a very large number of sightings considering their small population. Lastly, Los Angeles had a very small number of sightings per capita, so their relatively large count is most likely due to sheer population size rather than an actual higher rate of sightings. Ultimately, it seems like historical events regarding UFOs can impact and even potentially increase UFO reporting.

So finally, if you want to 'see' a UFO be on the look out in July around 9PM for a light or fireball in the sky either in WA, or for a more historical feeling, in Roswell, NM!

Group Member Roles:
Alice – Time top 5 shapes were sighted, Statistical Analysis of 2012
Solomon - UFO sighting characteristics, Binomial Distribution
Alesia – Temporal Patterns
Carlos – Spatial Patterns