

# UFO Sightings 1914-2010 reported by US Government

Mino Han, Johnny Yu, Sharon Nguyen, Eric Cha

## Abstract

Interest and fascination into UFOs (Unidentified Flying Objects) has only increased since mankind's first experiences with unknown aerial phenomena, though our understanding of the science that explains UFOs have failed to address even recent sightings. The National UFO Reporting Center updates a century-long record of UFO sightings across the globe, and can somewhat represent UFO sightings in the United States. In this project, we have intended to explore some of the spatial, temporal, and shape trends within a sample of the California NUFORC reports and extract meaningful observations from the data that could hint at societal changes in the preconceptions of UFOs. Within a significant increase in UFO reports since the late 1990s, we have found that sightings have mostly been at night and are more often located in well-populated areas in the state. Certain UFO shapes like "light" and "disk" have become more common, and we were unable to find meaningful correlation between variables in the dataset.

## Background

A somewhat hidden inclusion in 2020's last Coronavirus relief-related bill, Congress has asked the Office of the Director of National Intelligence to compile a mostly unclassified report detailing the country's experiences with and knowledge of UFO's, Unidentified Flying Objects (more formally, Unidentified Aerial Phenomena).



UFO's describe mysterious objects or occurrences in the air that are often interpreted as extraterrestrial visits to Earth. These phenomena often can and have been attributed to natural causes, with reasons such as an individual's hallucinations, atmospheric phenomena, or secretive military/science testing and reconnaissance. However, some reported instances, sometimes by trustworthy former military members, have been harder to explain, with speeds and movements that defy physics. From these few unexplainable events, interest in UFO's and extraterrestrial contact is a major interest for people around the world; there still remains a large population that believes in the existence of aliens and that they have visited the Earth. A [2018 poll by Insider](#) had 20% of its respondents report a belief that extraterrestrials have come to Earth, among other theories.

The [National UFO Reporting Center](#) (NUFORC), is a database and reporting center which fields reports from over the world regarding UFO sightings. Their report form seeks the date and time, duration, location, and other information about the instance of UFO sighting. This specific dataset was obtained from the RForDataScience GitHub community's TidyTuesday dataset publication on 2019-06-25, which explained how the original NUFORC data was "cleaned and uploaded to Github by Sigmond Axel, and some exploratory plots were created by Jonathan Bouchet a few years back." Comparing the products from the origin and this dataset, the bulk of this cleaning was compiling variable-indexed tables into one cohesive dataset.

Collection of this data is clearly motivated by the ongoing interest into UFO's and extraterrestrial elements, but this dataset cannot be applied to anything concrete about UFO's, which remain unverified mysteries. Instead, the dataset of UFO sightings describes perception of UFO's. Obtaining effective conclusions about UFO's themselves from these largely unverified reports may be hindered by their veracity, but there is merit to interpreting the changes in the properties of UFO sightings over the past century. Understanding the changes in UFO sightings may not necessarily be indicative of extraterrestrial patterns, but could likely correlate to changes in our society. Trends and patterns can describe spatial distributions, temporal frequency, the public's predisposition to certain UFO concepts, among others.

## Aims

This analysis explores the extent of this change in California by seeing how spatial and temporal patterns of sightings changed over time. The visualizations include examining spatial density of sightings before and after 2000 and identifying diurnal differences in UFO sightings throughout the years. Most reports were made at night and the density increases with increase in reports made throughout the years.

We also explore what UFO shapes have been most prevalent in sightings before and after the year 2000. It is interesting to note the jump in UFO shape, light, after 2000 when the internet and other technological advances started to peak. This may suggest an effect on UFO shape sightings. This leads into the increase of NUFORC reports made, shown in Figure 1 density of reports throughout the years. Furthermore, we used a correlation matrix to investigate the correlation between variables Latitude, Longitude, Encounter Length, Month, Hour, Year, and Year Documented.

# Data Description

The data is a sample of recorded UFO "sightings" in California between 1910 and 2014, including the UFO shape, latitude/longitude, specific town/city where the sighting occurred, duration of the "event", and the date and time it occurred. A larger dataset originally comes from the [NUFORC](#) and was [cleaned and uploaded to Github by Sigmond Axel](#).

Reports are collected from the NUFORC online report form, via U.S. mail, or the hotline (if the sighting occurred within the last week) which means it takes on a voluntary response sampling mechanism. There are limitations to voluntary reports because voluntary sampling is highly susceptible to bias. The people who voluntarily make reports may be different from those who do not make a report. Some people will inherently be more likely to volunteer than others. Visitors to the NUFORC website are encouraged to be cautious with regard to the accuracy of data, and none of the posted reports have undergone any degree of investigation. Despite this, the NUFORC believes that the majority of reports are likely truthful and accurate.

The population of the dataset is all NUFORC reports from California, and the sampling frame is the same. This dataset is a 2000 count random sample of reports from California, where each unit is one such sampled report from California, sampled using Pandas' DataFrame.sample() function in Python. As it is a random sample, we can appropriately extend conclusions from the sample to the population, the California NUFORC reports.

Name	Variable description	Type	Units of measurement
Year	Year sighting occurred	integer	year
Month	Month sighting occurred	integer	month (1-12)
Hour	Hour sighting occurred	integer	24 Hour Clock
City	City or area of sighting	character	
State	state/region of sighting	character	
Country	Country of sighting	character	
UFO Shape	UFO Shape	character	
Encounter Length	Encounter length in seconds	double	seconds
Year Documented	Year documented	integer	
Latitude	Latitude	double	degree
Longitude	Longitude	double	degree

Table 1: Provides variable descriptions and units for each column in the dataframe.

	Month	Year	Hour	City	State	Country	UFO Shape	Encounter Length	Year Documented	Latitude	Longitude
0	11	1996	20	stockton	ca	us	NaN	30.00	1999	37.96	-121.29
1	9	2006	21	el cajon	ca	us	formation	0.50	2006	32.79	-116.96
2	8	1999	19	glendale	ca	us	formation	15.00	1999	34.14	-118.25
3	7	1987	2	san juan capistrano	ca	us	light	0.05	2009	33.50	-117.66
4	11	1999	11	half moon bay	ca	us	chevron	5.00	1999	37.46	-122.43

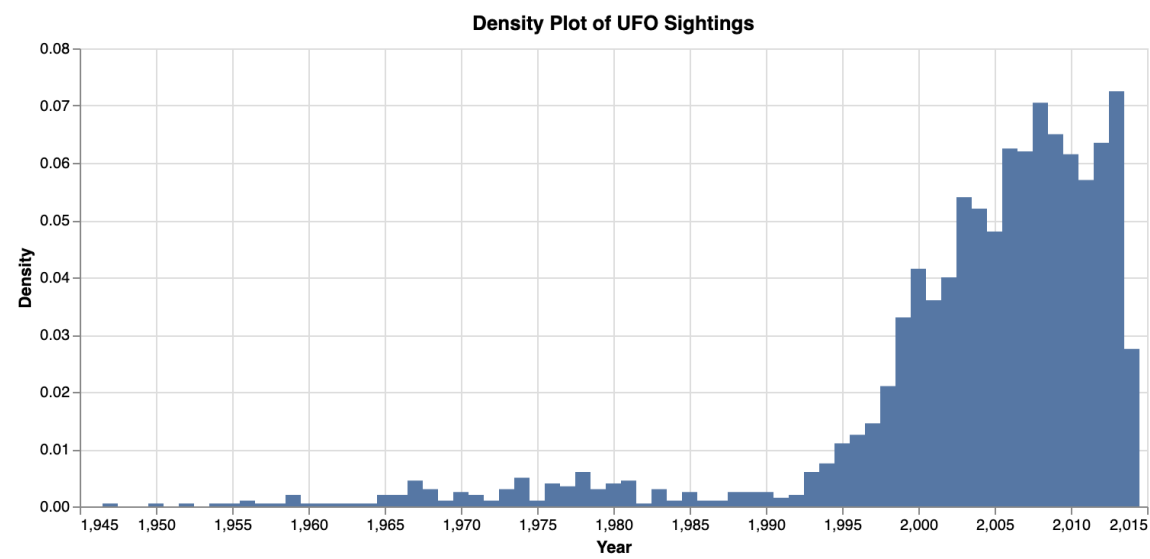
Table 2: Provides the first 5 observations of UFO sightings in the data.

# Methods

In Figure 1, we used a simple density histogram to examine the frequency of reports over the years. In Figure 2, we focused on analyzing the prevalence of locations regarding UFO sightings within California in two different time frames. To do so, we created a scatterplot of longitude and latitude of reports before and after the year 2000. We also added top and side panels with kernel density estimate (KDE) curves to visualize the density changes specific to latitude or to longitude. In Figure 3, we explored the prevalence of reported UFO shapes in two different time frames. For this, we plotted two sets of histograms from random observations before and after the year 2000 each plotting the densities of the respective UFO shapes. In Figure 4, we focused on examining the diurnal differences in UFO sightings by years. For this, we constructed a scatter plot of hours and years to determine patterns in UFO sightings. In Figure 5, we wanted to further explore the pairwise relationships between all the variables in this dataset. To do so, we created a correlation matrix to visualize the magnitude of correlation between the variables.

# Results

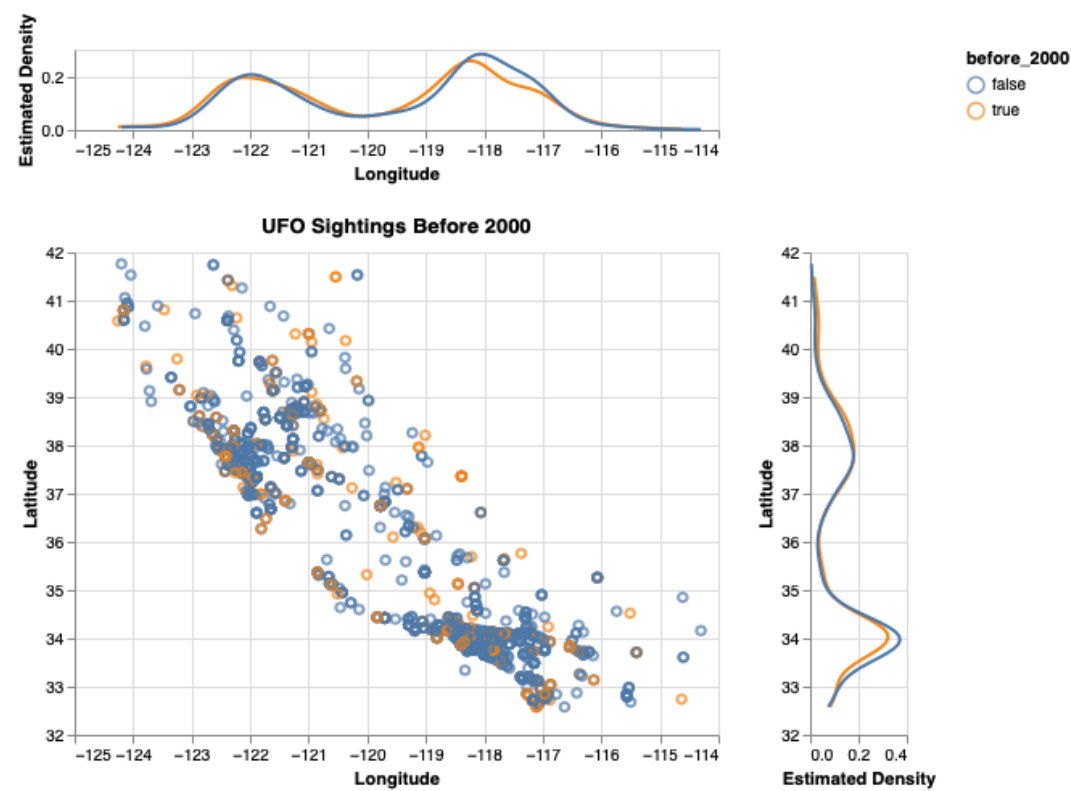
## Examining Density of UFO Sightings Over the Years



**Figure 1:** This graph shows the density of reports made to the NUFORC within each year.

This chart is a simple histogram of density of reports over the years of recording. Clearly identifiable is a significant increase in reports beginning in the late 1990s.

## Examining Spatial Density of Sightings Before and After Year 2000

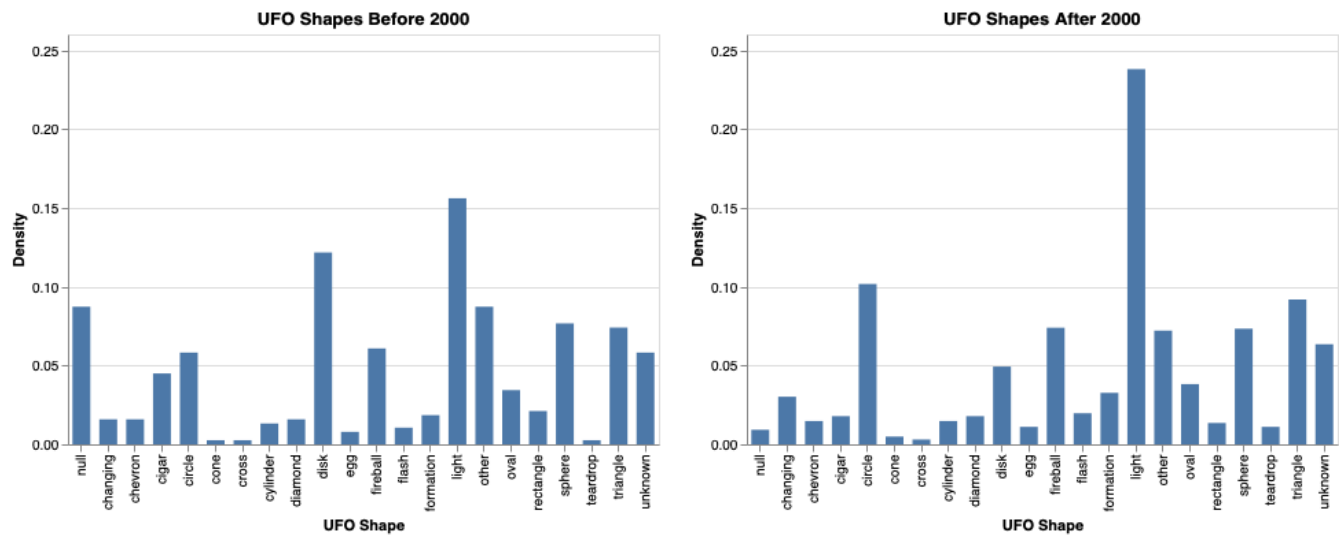


**Figure 2:** This graph shows density of reports related to longitude and latitude in California, for sightings before and after 2000.

We utilized the latitude and longitude attributes of each observation to plot the relative locations of sightings in Figure 2. Evidently, we can loosely see the shape of the state. By separating the data into observations before and after the year 2000, we can assess if there are significant differences in the spatial distribution of observations before and after 2000. Juxtaposed with the main plot, we can use kernel density estimation to visualize the density along both axes independently, as longitude and latitude may have been affected differently over time.

Distribution of the UFO sightings are understandably concentrated in well populated areas, specifically Los Angeles and San Francisco. The kernel density estimation curves reveal a change in the density of observations in the Southern California area, where sightings have become more frequent relative to other areas in the new millenia.

Examining Density of UFO Shapes in Reports Before and After Year 2000

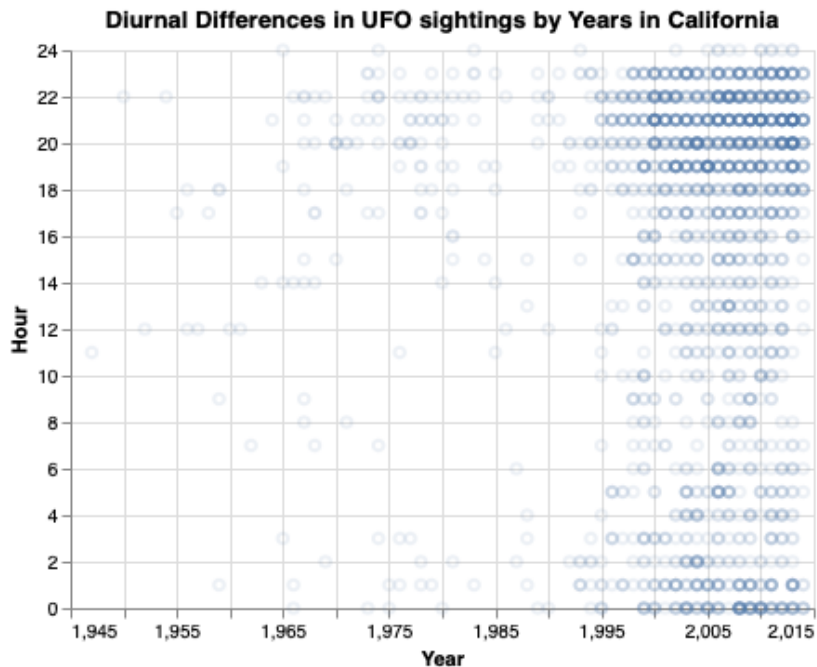


**Figure 3:** This graph shows the density of UFO Shapes found by NUFORC reporting before and after the year 2000.

We also wanted to assess the differences in UFO shapes in these two time frames. The two charts reflect the densities of the 21 UFO shapes that appeared in our sample. We can examine whether the shapes changed between the two time periods and how they changed.

Within minor changes in densities for most UFO shape types, the most notable changes were in the null, circle, disk, and light categories, where null sightings go considerably lower, circle sightings density doubles, disk sightings density nearly triples, and the light density nearly becomes a quarter of all sightings.

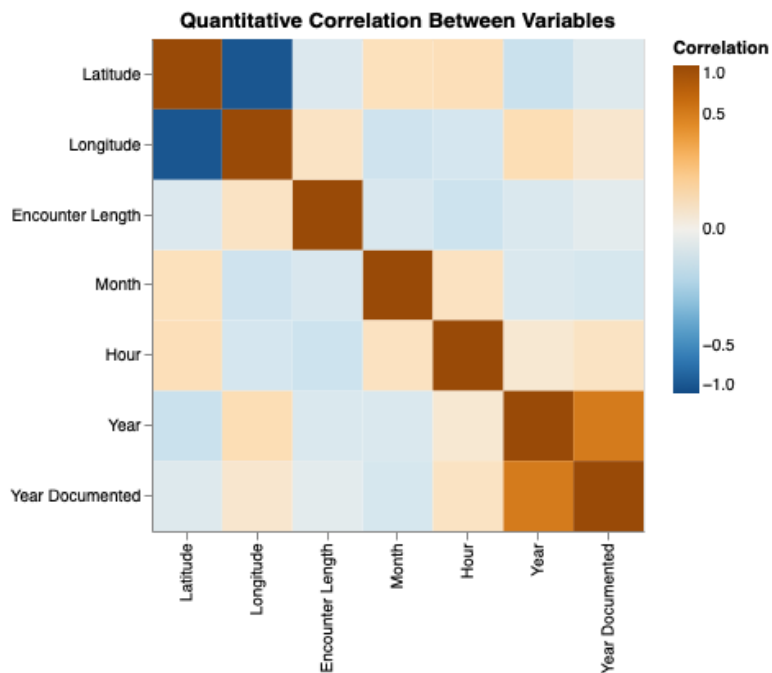
Identifying Diurnal Differences in UFO Sightings Throughout the Years



**Figure 4:** This graph shows diurnal differences in UFO sightings reported in California and its change over the years.

This plot identifies a concentration of observations during the night time hours. While the more recent sightings are more plentiful, sightings more in the past also follow this nighttime trend. Also, compared to pre-midnight, observations in the later hours of night towards morning are less common.

Investigating Correlation Between Quantitative Variables in the UFO Report Dataset



**Figure 5:** This graph shows correlation between all variables in the UFO sightings dataset.

In our last analysis, we developed a correlation matrix to assess the relationship between the quantitative variables in the dataset. This analysis did not reveal any significant correlation between most of the variables, only in ones we expected. Of course, the latitude and longitude variables were strongly inversely correlated with each other, and there was moderate positive correlation between the year of the sighting and the year it was documented. Besides these two exceptions, all the other variables had weak pairwise correlations.

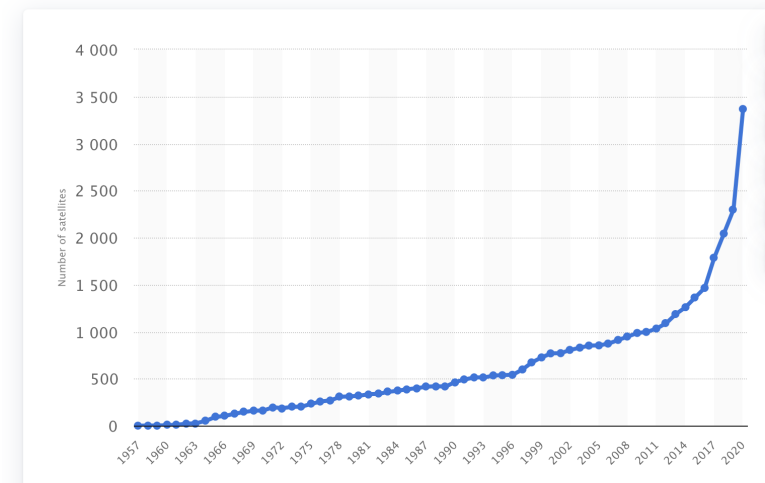
## Discussion

By the year 2000, almost half of the population in the U.S. was accessing information through the internet. This progression towards the internet age and technological advancements may affect the reports made to NUFORC. When exploring density in reports made throughout the years (Figure 1), it reveals a drastic increase in reports made after the early 90s. More widespread internet use has made it easier and quicker to submit reports to the NUFORC, which had previously been an obstacle for those who were not strongly committed to their UFO sightings.

Analyzing the spatial distribution of the UFO sightings revealed an interesting change between the before and after 2000 time periods, where Southern California sightings became more frequent (Figure 2). However, this change could mostly be explained by the increases in population that Southern California has experienced in the past century, and so like we have seen for other areas that have not changed, the more populous areas report UFO sightings more frequently.

The fact that most reports are made at night (Figure 4), and the major increases in the density of UFO shapes, light, after the year 2000 (Figure 3) could be explained by global technological advances, wherein increased use of satellites, planes, and advanced aerial technology could explain increases in sightings of these types. Thus, these increases may very well be explained by human error/misjudgement. ([www.statista.com/statistics/897719/number-of-active-satellites-by-year/](http://www.statista.com/statistics/897719/number-of-active-satellites-by-year/)).

Number of active satellites from 1957 to 2020



Further topics to be investigated may include exploring specific technological advances in satellites and planes that could explain rises in UFO reports. Location may also have an effect on reports made, the findings show that most reports were made from the highly populated areas of Los Angeles and San Francisco (Figure 2). Normalizing these spatial densities by local populations may reveal areas with atypical UFO reports.

It is important to note that the NUFORC reports are not representative of UFO activity or even of UFO sightings in general. Any conclusions we make can only be applied to the reports received by the NUFORC. Yet, we can still make interesting conclusions about all of the >16,000 reports from California over the past century and identify the changes in their attributes over time. As these reports indicate changes in societal experiences of UFOs, it may be interesting to correlate changes in UFO types with popular UFO depictions in US media. For instance, there may have been an interesting increase in sightings of “disk” shaped UFOs after a popular disk-shaped movie UFO.