UFO Activity

Anusarika Prasai, Olisa Omekam

I.Introduction

Unidentified flying objects (UFO) is an object observed in the sky that is not readily identified. The term “UFO” was coined in 1953 by the United States Air Force to serve as a catch-all for such reports. In its initial definition, the United States Air Force stated that a UFO was any “airborne object which by performance, aerodynamic characteristics or unusual features, does not conform to any presently known aircraft or missile type or which cannot be positively identified as a familiar object”. However, in recent times, the term UFO began to used to refer to claims to alien spacecraft. In this project, we just wanted to find out simple answers such as what areas of the country are most likely to have UFO sighting? Are there any trends in UFO Sighting over time? Do cluster of UFO sighting correlate with landmarks, such as airports or government research? What are the most common UFO descriptions? Is there really alien activity in Roswell, New Mexico?

II. Dataset

The dataset that was used for the project comes from the National UFO Reporting Center. The data has over 80,000 data which starts at 1949 and ends at 2013. For the sake of simplicity, we did include the comments, but we made sure all of the extra symbols were deleted. The database we used can be found at http://www.nuforc.org/ which gave us the UFO statistics in CSV format.

III. Conceptual Database Design

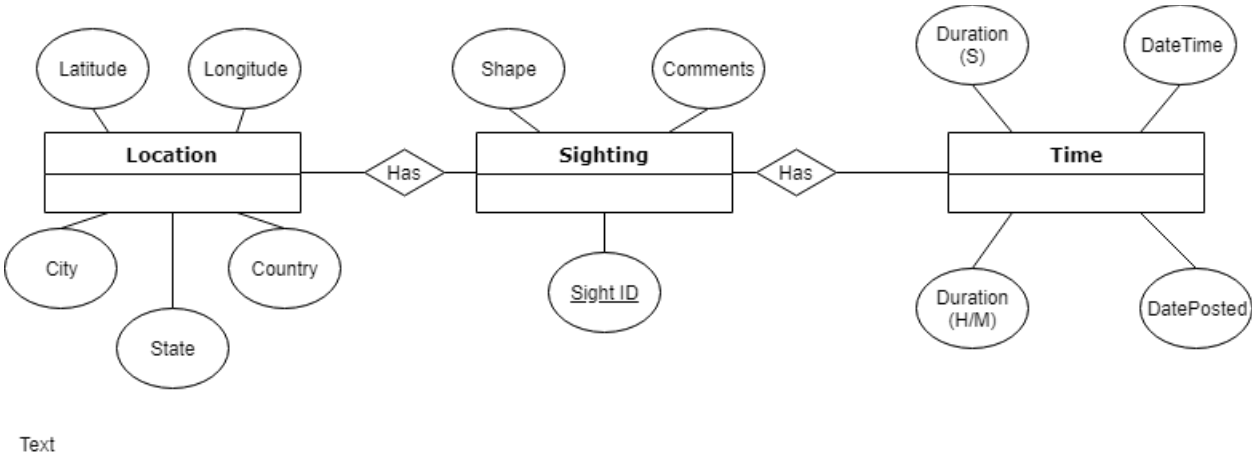


Figure 1. Conceptual Database Model

Above is the conceptual model was used for this project. It was very important that entries, attributes, and relationship, and key constraints were all identified. There are three tables in the database Location, Sighting, and Time. The location table has five attributes. In the location table, longitude is the primary key. The sighting table has three attributes. In the sighting table, sightID is the primary key, which is auto-incremented into the database. Lastly, the time table has four attributes. In the time table, data time has is the primary key and has the sightID on sighting as the foreign key.

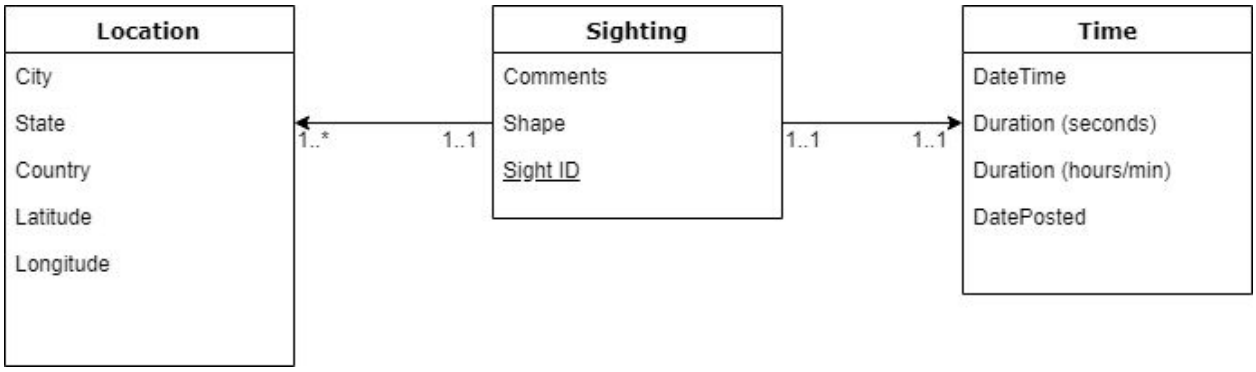


Figure 2: Relation Data Model

Location Table

Location table has five attributes: city, state, country, latitude, and longitude. In which, city, state, and country are text in the database. Latitude and Longitude are double in the database. The longitude is the primary key in the table. The location has a 1.1 relationship with sighting and 1.1 the relationship table with Time.

Sighting Table

The sighting table as shown in the relational model has three attributes. This table has one to many relationships with the location table which it has one to one relationship with time. The sighting table has attributes: comments, shape, and sightID. SightID is the primary key. We will auto increment the sightID so it will unique.

Time Table

The Time Table as four attributes: dateTime, Duration (seconds), Duration(hours/min) and Date posted. This has 1.1 relationships with sighting and location. The primary key in this table is the dateTime and foreign key in this table is going to be SightID, linking sighting to the time table

Normalization

Fortunately, our data was already in the 3 normalization form, we did not have to normalize the data.

	A	B	C	D	E	F	G	H	I	J	K
1	datetime	city	state	country	shape	duration (second	duration (hours/min)	date posted	latitude	longitude	SightID
2	10/10/1949 20:30	san marcos	tx	us	cylinder	2700	45 minutes	4/27/2004	29.8830556	-97.9411111	1
3	10/10/1949 21:00	lackland afb	tx		light	7200	1-2 hrs	12/16/2005	29.38421	-98.581082	2
4	10/10/1955 17:00	chester (uk/england)		gb	circle	20	20 seconds	1/21/2008	53.2	-2.916667	3
5	10/10/1956 21:00	edna	tx	us	circle	20	1/2 hour	1/17/2004	28.9783333	-96.6458333	4
6	10/10/1960 20:00	kaneohe	hi	us	light	900	15 minutes	1/22/2004	21.4180556	-157.8036111	5
7	10/10/1961 19:00	bristol	tn	us	sphere	300	5 minutes	4/27/2007	36.595	-82.1888889	6
8	10/10/1965 21:00	penarth (uk/wales)		gb	circle	180	about 3 mins	2/14/2006	51.434722	-3.18	7
9	10/10/1965 23:45	norwalk	ct	us	disk	1200	20 minutes	10/2/1999	41.1175	-73.4083333	8
10	10/10/1966 20:00	pell city	al	us	disk	180	3 minutes	3/19/2009	33.5861111	-86.2861111	9
11	10/10/1966 21:00	live oak	fl	us	disk	120	several minutes	5/11/2005	30.2947222	-82.9841667	10
12	10/10/1968 13:00	hawthorne	ca	us	circle	300	5 min.	10/31/2003	33.9163889	-118.3516667	11
13	10/10/1968 19:00	brevard	nc	us	fireball	180	3 minutes	6/12/2008	35.2333333	-82.7344444	12
14	10/10/1970 16:00	bellmore	ny	us	disk	1800	30 min.	5/11/2000	40.6686111	-73.5275	13
15	10/10/1970 19:00	manchester	ky	us	unknown	180	3 minutes	2/14/2008	37.1536111	-83.7619444	14
16	10/10/1971 21:00	lexington	nc	us	oval	30	30 seconds	2/14/2010	35.8238889	-80.2536111	15
17	10/10/1972 19:00	harlan county	ky	us	circle	1200	20minutes	9/15/2005	36.8430556	-83.3219444	16
18	10/10/1972 22:30	west bloomfield	mi	us	disk	120	2 minutes	8/14/2007	42.5377778	-83.2330556	17
19	10/10/1973 19:00	niantic	ct	us	disk	1800	20-30 min	9/24/2003	41.3252778	-72.1936111	18
20	10/10/1973 23:00	bermuda nas			light	20	20 sec.	1/11/2002	32.364167	-64.678611	19
21	10/10/1974 19:30	hudson	ma	us	other	2700	45 minutes	8/10/1999	42.3916667	-71.5666667	20
22	10/10/1974 21:30	cardiff (uk/wales)		gb	disk	1200	20 minutes	2/1/2007	51.5	-3.2	21
23	10/10/1974 23:00	hudson	ks	us	light	1200	one hour?	7/25/2004	38.1055556	-98.6597222	22
24	10/10/1975 17:00	north charleston	sc	us	light	360	5-6 minutes	2/14/2008	32.8544444	-79.975	23
25	10/10/1976 20:30	washougal	wa	us	oval	60	1 minute	2/7/2014	45.5827778	-122.3522222	24
26	10/10/1976 22:00	stoke mandeville (uk/england)		gb	cigar	3	3 seconds	12/12/2009	51.783333	-0.783333	25
27	10/10/1977 12:00	san antonio	tx	us	other	30	30 seconds	2/24/2005	29.4238889	-98.4933333	26
28	10/10/1977 22:00	louisville	ky	us	light	30	approx: 30 seconds	3/17/2004	38.2541667	-85.7594444	27
29	10/10/1978 2:00	elmont	ny	us	rectangle	300	5min	2/1/2007	40.7008333	-73.7133333	28
30	10/10/1979 0:00	poughkeepsie	ny	us	chevron	900	15 minutes	4/16/2005	41.7002778	-73.9213889	29

Figure 3: Sample of the dataset

IV. Physical Database

For our physical database design, we used Mysql Workbench and Amazon Web Services to host the database. Since our conceptual design translated well into the logical design we did not have to spend a lot of time setting up the database. Each of the conceptual tables translated well into a physical tables.

```
CREATE TABLE `location` (
  `Longitude` int(11) NOT NULL,
  `Latitude` int(11) DEFAULT NULL,
  `City` varchar(45) DEFAULT NULL,
  `State` varchar(45) DEFAULT NULL,
  `Country` varchar(45) DEFAULT NULL,
  PRIMARY KEY (`Longitude`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8
```

Figure 4: creation of the location table

This is an example of a query used in our database. The key constraints and domain constraints are spelled out clearly in this create table query, where Longitude is the Primary Key and where Longitude has an INT data type and cannot be null.

Result Grid | Filter Rows: | Export: | Wrap Cell Co

state	longitude	latitude	city	country
ct	-73.6288889	41.0263889	greenwich	us
sc	-79.975	32.8544444	north charleston	us
ca	-122.389979	37.615223	san francisco airport	
va	-77.1875	38.7891667	springfield	us
ct	-73.2052778	41.1669444	bridgeport	us
or	-122.5691667	45.4077778	clackamas	us
tx	-100.4366667	31.4636111	san angelo	us
ct	-72.6441667	41.8525	windsor	us
il	-89.6961111	41.7886111	sterling	us
wa	-122.2008333	47.9791667	everett	us
az	-110.6197222	35.8752778	kykotsmovi	us
ny	-73.7566667	42.6525	albany	us
ca	-117.9405556	33.7738889	garden grove	us
fl	-80.1938889	25.7738889	miami	us

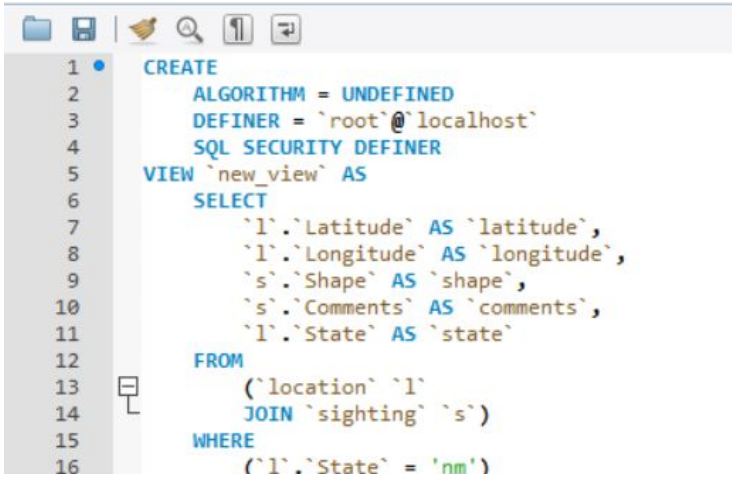
Figure 5: view of the location table

Similar create queries like above were used to create the rest of our database, all according to the logic design.

V.Queries

Searching For New Mexico

One of the main questions we wanted to answer in the beginnings of the project was: is there any alien sighting in New Mexico. We ran the query below to find out and luckily the output of this query was not an empty table



```
CREATE
    ALGORITHM = UNDEFINED
    DEFINER = `root`@`localhost`
    SQL SECURITY DEFINER
VIEW `new_view` AS
    SELECT
        `l`.`Latitude` AS `latitude`,
        `l`.`Longitude` AS `longitude`,
        `s`.`Shape` AS `shape`,
        `s`.`Comments` AS `comments`,
        `l`.`State` AS `state`
    FROM
        (`location` `l`
        JOIN `sighting` `s`)
    WHERE
        (`l`.`State` = 'nm')
```

Figure 6: query for New Mexico

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

state	latitude	longitude	shape	comments
nm	35.0844444	-106.6505556	light	Three bright lights that were huddled together t...
nm	36.7280556	-10	cle	Single reddish circle in the sky that wasn&#39t ...
nm	32.3122222	-10	ireball	4 fireballs in sky side by side
nm	35.0844444	-106.6505556	sphere	Six Spheres of light move into my vision over th...
nm	35.0844444	-106.6505556	fireball	Vehicle out of southern sky thought to be shoot...
nm	32.3122222	-106.7777778	fireball	4 fireballs side by side
nm	33.1283333	-107.2522222	unknown	Fast moving red ball and then flashing light
nm	0	0	disk	While traveling at night we saw what appeared ...
nm	35.2333333	-106.6638889	fireball	Sighted during the major meteor shower last year.
nm	35.0844444	-106.6505556	triangle	TRIANGLE VERY LARGE THREE LIGHTS AT POIN...
nm	35.6869444	-105.9372222	egg	Oct&#44 2002 &#44 8:30 AM&#44 Santa Fe&...
nm	35.1472222	-107.8508333	fireball	In October 2002&#44 I was driving east on Int...
nm	35.0844444	-106.6505556	oval	lights
nm	35.0844444	-106.6505556	disk	Two blue pinpoints of light merge onto one silve...
nm	35.2333333	-106.6638889	oval	From inside my home I witnessed an oval shape...
nm	35.6869444	-105.9372222	fireball	Several bright lights headed slowly towards hori...
nm	35.5830556	-105.7694444	changing	Large&#44 meteor-type light seen near Santa ...
nm	32.77	-108.2797222	light	UFO over Silver City&#44 New Mexico at very l...
nm	35.6869444	-105.9372222	light	about 23:15 sunday&#44 10/1/06--a slow-mo...
nm	36.4072222	-105.5725	formation	segmented flashing formation w/ "v&quot...
nm	36.9033333	-104.4386111	formation	7 bright lights over northeast new mexico
nm	36.9033333	-104.4386111	formation	Seven Bright Lights in New Mexico Skies
nm	36.7280556	-108.2180556	light	a bright lights traveling very fast with a green t...
nm	33.9177778	-106.8652778	circle	ON SUNDAY 10 01 06 MY WIFE AND I WERE DR...
nm	35.2333333	-106.6638889	disk	The UFO was a sacer shaped&#44 silver object.
nm	35.0844444	-106.6505556	light	Pulsing light of red&#44 green and white seen i...
nm	36.665547	-106.691032	changing	Watched a strange light for about an hour. Duri...
nm	35.6869444	-105.9372222	disk	while driving east something unusuall grabed m...
nm	36.4072222	-105.5725	rectangle	Parallelogram shaped bright&#44 white light lo...
nm	35.6869444	-105.9372222		((HOAX??)) Some kind of aircraft with a HUGE ...
nm	34.955	-107.1841667	light	Bright light emits smaller light
nm	33.1283333	-107.2522222	circle	Shiny object reflected sun in white sands area u...
nm	35.0844444	-106.6505556	flash	3 flashing green lights in formatiom of a triangle...
nm	35.0844444	-106.6505556	disk	Disc close enough to almost touch no sound or ...

-108.2180556

Figure 7: output for New Mexico

We also searched the comments attribute to make sure we did not miss any mentions of New Mexico because sometimes the data does not have the state listed.

```
SELECT * FROM sighting
 where comments like "%new mexico%";
```

Figure 8: searching the comments for New Mexico

Result Grid | Filter Rows: | Edit:

SightID	comments	shape
903	UFO over Silver City&#44 New Mexico at very l...	light
905	October 1&#44 2006 meteor reported across s...	oval
920	7 bright lights over northeast new mexico	formation
921	Seven Bright Lights in New Mexico Skies	formation
927	ON SUNDAY 10 01 06 MY WIFE AND I WERE DR...	circle
2157	Was traveling eastbound on I-40 in New Mexico...	light
NULL	NULL	NULL

Figure 9: output table for comments

Most UFO activity

```
CREATE
    ALGORITHM = UNDEFINED
    DEFINER = `root`@`localhost`
    SQL SECURITY DEFINER
VIEW `new_view1` AS
    SELECT
        `location`.`country` AS `country`
    FROM
        `location`
    GROUP BY `location`.`country`
    ORDER BY COUNT(`location`.`country`) DESC
    LIMIT 1
```

Result Grid

country
us

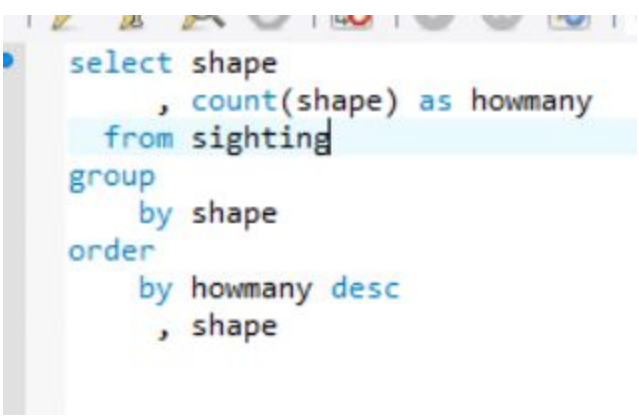
Figure 10: query and output most UFO activity

We ran the query above to see what country in the reported data has the most UFO sighting. The result of the query was the United States. This data might be little biased since the nuforc is

located in the United States and most people who live outside the united states probably do not know that this database exists.

Most Common UFO Shape

We wanted to find out what was the most common reported UFO shape. For this, we took a sample of our dataset because we ran into an error when we tried to do this query with my full dataset. We chose our data points in random and the result is shown below. The most common shape was light which was mentioned 441 times and the second and third common was triangle and circle.



```
select shape
     , count(shape) as howmany
  from sighting
group
    by shape
order
    by howmany desc
     , shape
```

shape	howmany
light	441
triangle	236
circle	214
unknown	182
disk	179
other	153
fireball	144
sphere	144
oval	121
formation	76
cigar	70
changing	61
	60
cylinder	46
chevron	44
rectangle	38
flash	29
diamond	27
egg	25
teardrop	11
cross	7
cone	4
delta	1

Figure 11: query and output of most Common shape

Other sample queries

```
CREATE
    ALGORITHM = UNDEFINED
    DEFINER = `root`@`localhost`
    SQL SECURITY DEFINER
VIEW `view` AS
    SELECT
        `location`.`Country` AS `country`,
        `location`.`City` AS `city`,
        `location`.`State` AS `state`
    FROM
        `location`
    WHERE
        (`location`.`Country` = 'us')
```

country	city	state
us	kaneohe	hi
us	washougal	wa
us	hawthorne	ca
us	san marcos	tx
us	edna	tx
us	pell city	al
us	manchester	ky
us	live oak	fl
us	bristol	tn
us	lexington	nc

Figure 12: Query and Result for UFO Sighting in the US

```
CREATE
    ALGORITHM = UNDEFINED
    DEFINER = `root`@`localhost`
    SQL SECURITY DEFINER
VIEW `above10` AS
    SELECT
        `l`.`State` AS `state`,
        `l`.`City` AS `city`,
        `p`.`Comments` AS `comments`,
        `p`.`Shape` AS `shape`
    FROM
        (`location` `l`
        JOIN `sighting` `p`)
    WHERE
        ((`l`.`SightID` = `p`.`SightID`)
            AND (`p`.`SightID` < 10))
    ORDER BY `p`.`SightID` DESC
```

Figure 13: Query to join two tables and order by sightID Desc

state	city	comments	shape
fl	live oak	This event took place in early fall around 1949-50.	cylinder
ky	manchester	This event took place in early fall around 1949-50.	cylinder
al	pell city	This event took place in early fall around 1949-50.	cylinder
tx	edna	This event took place in early fall around 1949-50.	cylinder
tx	san marcos	This event took place in early fall around 1949-50.	cylinder
tx	lackland afb	This event took place in early fall around 1949-50.	cylinder
ca	hawthorne	This event took place in early fall around 1949-50.	cylinder
wa	washougal	This event took place in early fall around 1949-50.	cylinder
hi	kaneohe	This event took place in early fall around 1949-50.	cylinder

Figure 14: Result of the query above

VI. Visualization

For visualization we used the Python library *matlibplot* to plot on a basemap of all usable 87,184 locations of the supposed sightings to help visualize the geographical locations of those sights. We took the latitudes and longitudes and exported them into a csv file which was then loaded into our program using *pandas*. Plotting all of the points was very cpu intensive and the entire process took almost an hour to complete.

```
import matplotlib.pyplot as plt
from mpl_toolkits.basemap import Basemap
import pandas as pd
import io
import numpy as np


# read in data to use for plotted points
cols_to_use = ['latitude', 'longitude']
buildingdf = pd.read_csv('/home/ollie/Documents/latlon.csv')


lat = buildingdf['latitude'].values.tolist()
lon = buildingdf['longitude'].values.tolist()
combined = np.vstack((lat, lon)).T

# create map using BASEMAP
m = Basemap(projection='merc',llcrnrlat=-80,urcrnrlat=80,\
            llcrnrlon=-180,urcrnrlon=180,lat_ts=20,resolution='c')
m.drawcoastlines()
m.drawcountries()
m.drawstates()
m.bluemarble()
q = len(lat)

for i in range(0,q):
    lats,lons = float(lat[i]),float(lon[i])
    x,y = m(lons,lats)
    m.plot(x,y, 'ro')

plt.show()
```

Figure 15: Code for the program

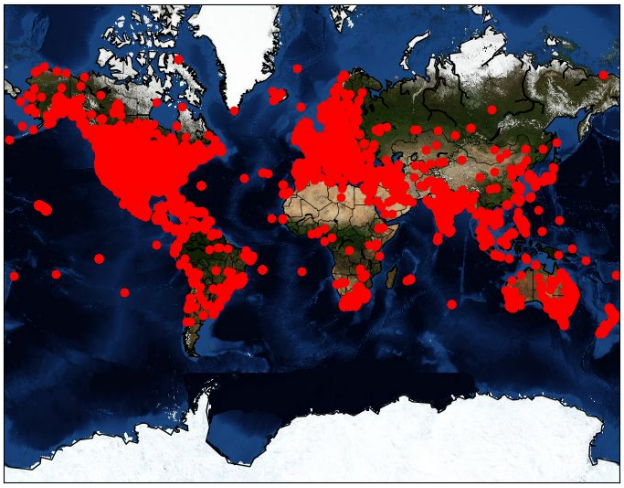


Figure 16: Plots of the world

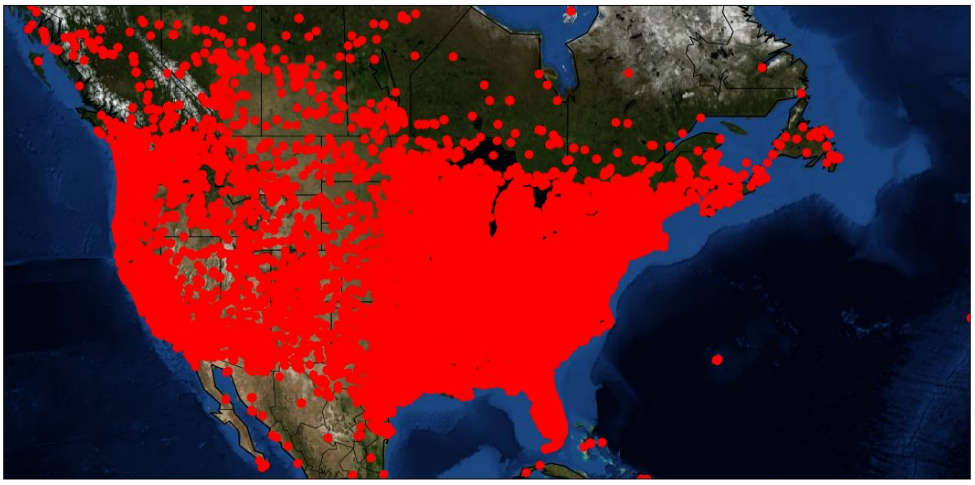


Figure 17: Plots of the United States

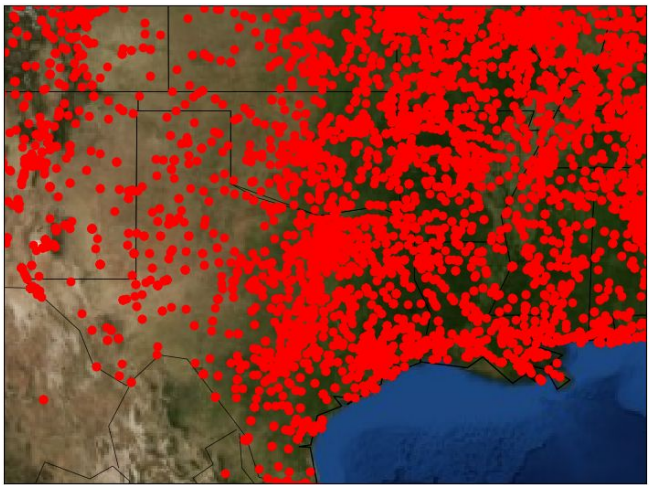


Figure 18: Plots of Texas

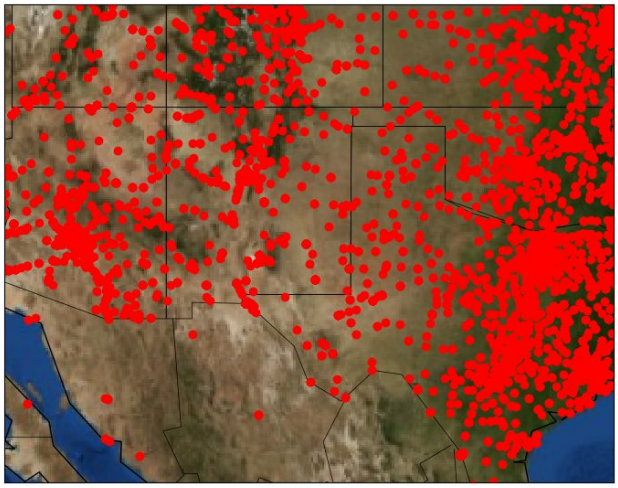


Figure 19: Plot for New Mexico

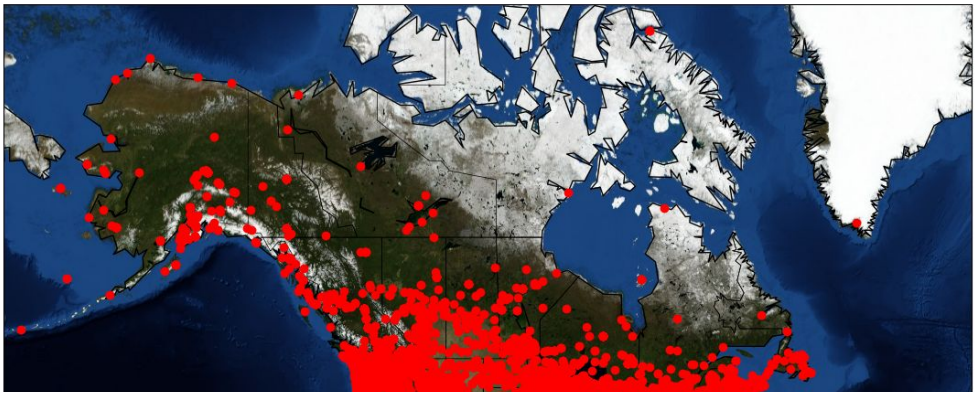


Figure 20: Plot of Canada/Alaska

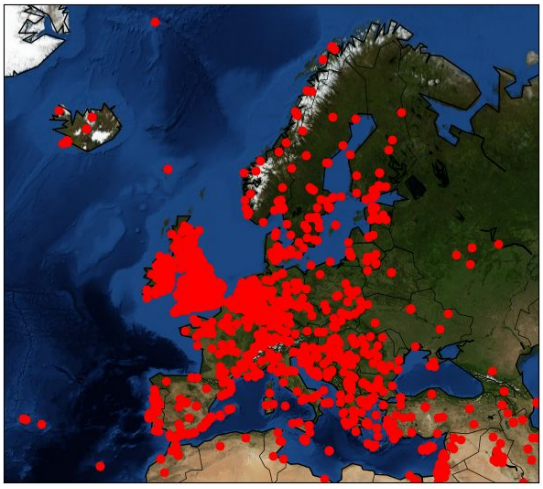


Figure 21: Plot of Europe

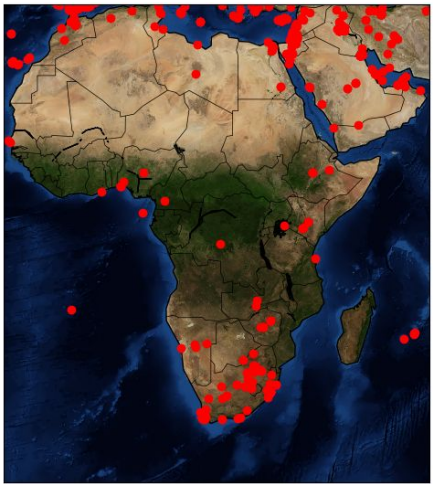


Figure 22: Plot of Africa

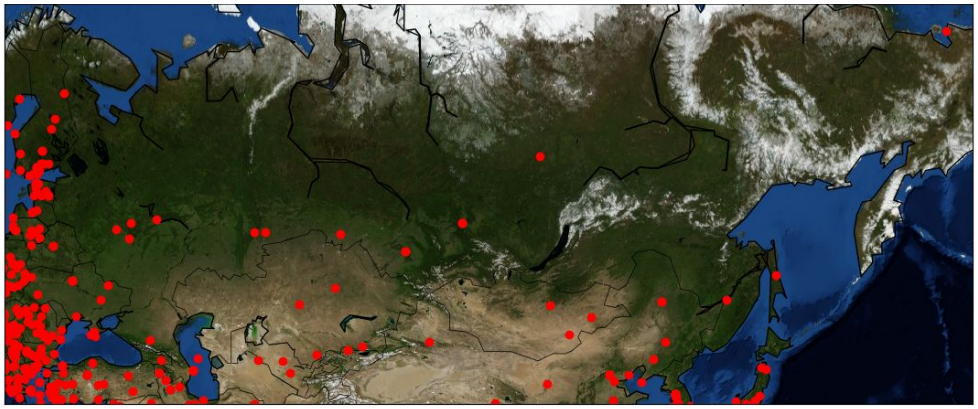


Figure 23: Plot of Russia/Mongolia

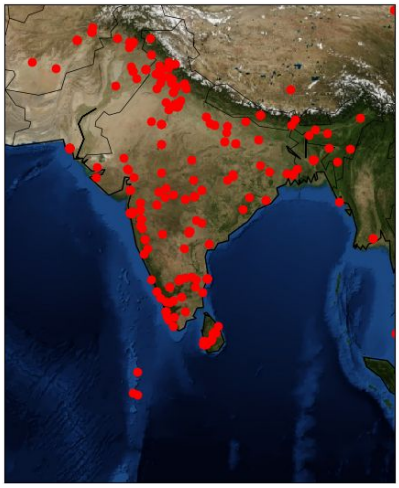


Figure 24: Plot of India

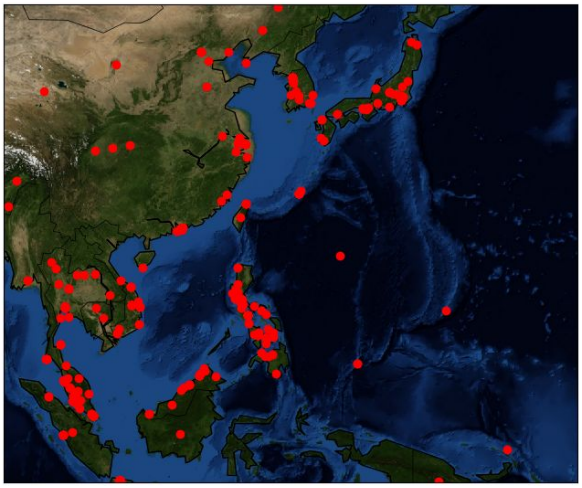


Figure 25: Plot of South/Northeast Asia

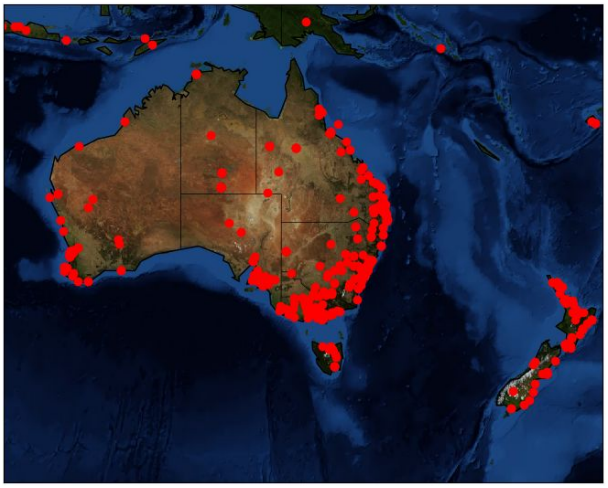


Figure 26: Plot of Australia/New Zealand

VII. Conclusion

Our database is a useful tool for convening a wide array of interesting information regarding related to UFO sighting in the world. With this database, we can effectively visualize the data instead of looking at a long list of data. Our project touched upon all the major topics we learned in class and helped us understand more about the topics.

Future plans for our database would be to improve the UI, add more functionality to the database. We want to create more hands and interactive on visualization graph. We would also want to auto update the database when someone makes a report on the nuforc.