UFO Sightings Project

Ryan Whalen, John Bartlett, Becky Mathews

IST 707: Data Analytics

7/16/2020

# Introduction

Since 2017, the New York Times has been reporting on Pentagon studies of Unidentified Flying Objects(UFOs). This has raised the bar significantly for serious consideration of the UFO phenomena. Before this, and still, UFO sightings were largely written off as hoaxes and delusions by a large part of society. The highly recognizable “flying saucer” can be traced back to 1947, when popular culture was taken by waves of UFO sightings. However, the sightings go back even further to the 1890s air ship sightings. These were also vehicles crossing our skies but were often described as “cigar shaped”. Fantastical explanations have been attributed to the phenomena since the beginning and varied over time with the cultural zeitgeist. In the 1890s they were piloted by eccentric human inventors, in the 1940s by benevolent space brothers here to save us from nuclear annihilation and the 1980s sinister experimentations.

# Motivation

Although long considered a fringe topic, UFOs have now been rebranded as UAPs (Unidentified Ariel Phenomena) and ae now being discussed in mainstream news sources and being taken seriously by the U.S. military. Looking at the data may reveal some insight into what this long-standing phenomenon represents, outside of purely cultural beliefs. With even U.S. senators such as Marco Rubio pushing for investigation of the subject, it’s time to look at the publicly available data.

# Problem Statement

With occurrences UFO sightings reaching new heights during the global pandemic, we have some questions regarding the historical data.

What areas of the country are most likely to have UFO sightings?

Are there any trends in UFO sightings over time? Do they tend to be clustered or seasonal?

What are the most common UFO descriptions, do any other factors correlate to the witnessed shape?

# Analysis

The first step in this project was getting a clear visualization of where sightings were occurring using mapping in R Studio. First, the data was read in from a csv file. This ending up being 80,332 observations of 11 variables: datetime, city, state, country, shape, duration in seconds, duration in hours and minutes, comments, date posted, latitude and longitude. The authors decided to work primarily in a continental United States context for the mapping visualizations. The data was filtered by country and then the states Hawaii and Alaska were filtered out. The result was a data set of 64,533 observations. Next, a function named “statesFromLower” was created to add the full state name to every sighting in the data set. This enabled the usage of ggmap from the ggplot2 package. The US map was created with a dot for every sighting in the data. The alpha parameter, which control the opacity of the dots, was set very low to 0.03. This enabled the highest concentrations of sightings to clearly present themselves, after initial attempt at higher alpha values resulted in a map completely covered in dots. Analysis of the map is its high number of sightings near population dense areas. This is interesting as culturally; UFOs are generally associated with sightings in more remote locations. The results are below:

A picture containing screenshot

Description automatically generated

Next, a second map was creating which showed sighting density by county in the continental US. This map’s data was created by first creating a data frame which contained just the city and state variable and then merging with a US data set on city and state to create a “county sightings” data set contained city, state and county. The data was then aggregated by county to get a “county count” data frame. This frequency of sightings was then rescaled from one to ten. After that, a ten color vector using heat colors was created and matched to the county count’s frequency variable to create a color variable in the data frame. Last, a map was created using the maps package, calling the map function and passing in a county map based on the county count data frame, with the color variable assigning as the map color for that county. Analysis of the map is the notable evenness of most of the country when broken out by county. Many less populous counties are over a greater landmass, which may account for this. The results are below:

A picture containing map, text

Description automatically generated

Next, attention was turned to getting a general sense of what individuals were saying about their experiences. This was visualized in a simple initial word cloud created from the comments variable. The main message that jumps out is that individuals are seeing *something* in the sky:

A screenshot of a cell phone

Description automatically generated

At this point, a great deal of exploration was done on the individual variables within the data. The data was reloaded to its original state. The datetime data was broken down into several numeric variables: year, month, day and hour. The months were converted to their English names and changed to factor data type. This was used to generate four barplots which showed the frequency of sightings by year, month, day and hour. This confirmed a few things. The first is that most sightings were in the evening, after dark. The reporting was somewhat imprecise as an improbably high number were recorded at the first or fifteenth of the month, likely being reported simply happening in the beginning or middle of the month. The data also showed a moderate skew towards the summer months. Finally, the data is heavily trending towards more recent sightings up until 2014.

A screenshot of a cell phone

Description automatically generated

Next came a piechart by country, filtering out any null values for country. This showed the data as heavily US skewed.

A close up of a logo

Description automatically generated

Next was a barchart showing state filtering down to only US sightings

A screenshot of a cell phone

Description automatically generated

Two initial charts were created on duration. The first was a barchart looking at the frequency of various sighting durations. The second was a scatterplot which viewed how durations changed by year.

A screenshot of a cell phone

Description automatically generated

From the previous graphics some things became clear. The first was that there were some extreme outliers in duration. The second was that these extreme outliers were likely throwing off the scale and hiding the trend in duration over time. As a result, the 95 quartiles was then used for a second scatterplot which eliminates the extreme outliers and allows the upward trend in sighting duration over time to make itself clear.

A screenshot of a cell phone

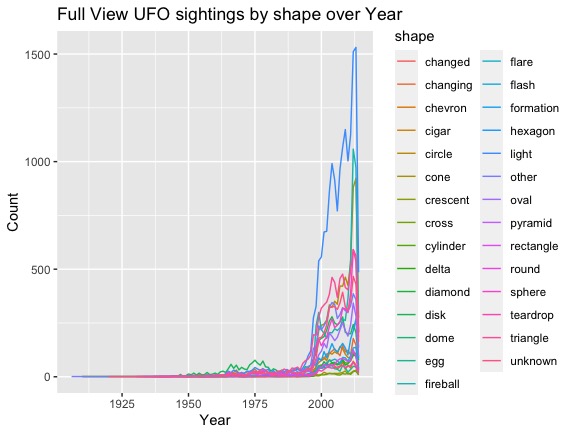
Description automatically generated

Finally, sighting shape was investigated. A simple barplot on shape showing a surprising result: most sightings were quite vague. Although expected cultural ideas of UFO shapes such as triangle and circle had moderately high frequency, the leader by a large margin was light

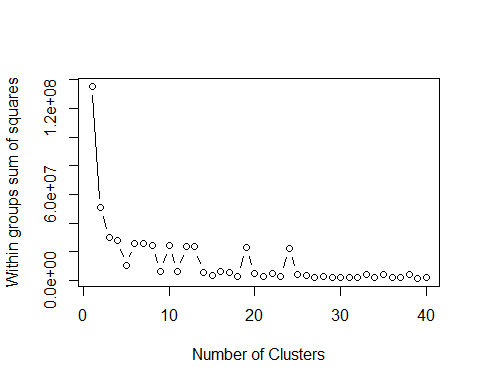
. A screenshot of a cell phone

Description automatically generated

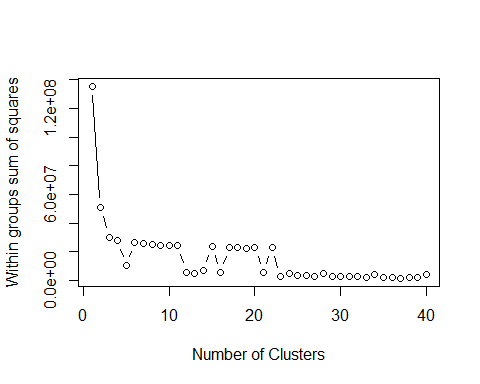
Lastly, sighting was again investigated but this time by year. A complex line chart was created to see if different lines rose to prominence over time. This did end up being the case. There is clear variation in sighting shapes over time.

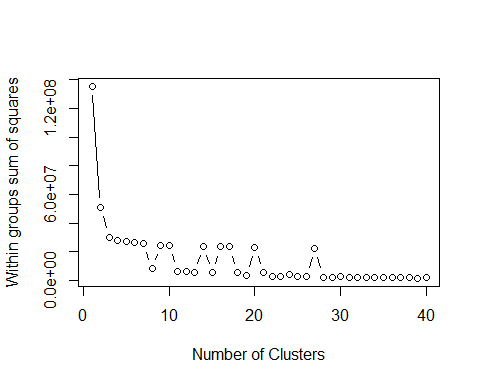


From this point, more comprehensive machine learning analysis began. The first attempt, a decision tree to predict shape, was met with failure. The large number of observations and variables caused the process to run for over 20 hours with no result. From there, variable of interest were identified for use in attempting to discern shape. The hope was that these would be the most likely to determine contributing cultural factors, if any, to how witnesses interpreted the shape of the object seen. Chosen first was latitude and longitude. These would help determine regionality by several factors, with both north/south and coast/interior being common cultural divisions in the US. Other factors chosen were year, state and duration. After twenty hours, this tree also failed to run. Next, a thousand observations were sampled for another attempt which timed out after 6 hours. 100 observations were given three hours to run, also with no results. A tiny sample of ten was attempted to ensure that the code was functioning properly. This returned results in seconds. A larger sample of fifty was attempted which ran within an hour. This resulted in using state at the primary decision factor. With the very tiny sample size and there being almost as many states in the continental US as there were observations, this conclusion was deemed likely spurious.



To definitively support the clustering of the data around the US we decided to attempt a kmeans clustering analysis on the geospatial data. The first step in our kmeans clustering effort is to find the ideal number of clusters for the, to accomplish this we decided to use a technique to visualize the efficiently of the clustering at different cluster numbers. You can see this above, the ideal number of clusters would be at the ‘bend’ of the line.

Here are two more iterations of this technique.

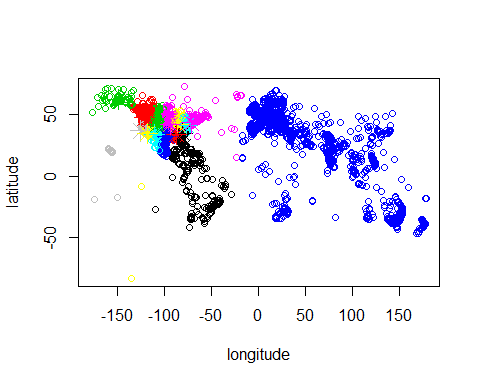


#Takeaway:

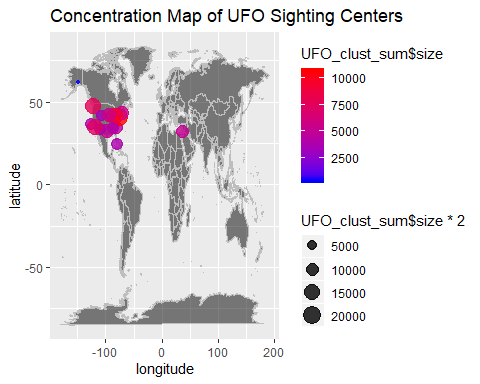
The optimal number of clusters is somewhere inbetween 10 and 30, there seems to be some strange results of a sort of bouncing efficentcy within the that range.

Due this weird kangaroo effect, we chose to go for a optimal number of clusters that sits in the kangaroo range, so we settled on 30 as that is the only real reliable effective number. Any number of clusters before that seems to be a luck of the draw, atleast in terms of the sum of squares of each cluster

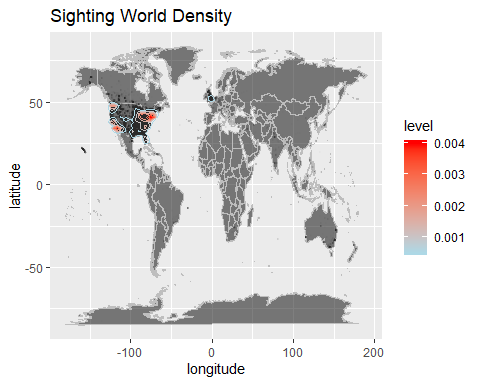
Our assumption is that the kangaroo effect is beacause of how wides spread the data is, and because of over 3 quarters of the data being centered in North America and the remainder being split evenly through the rest of the world.

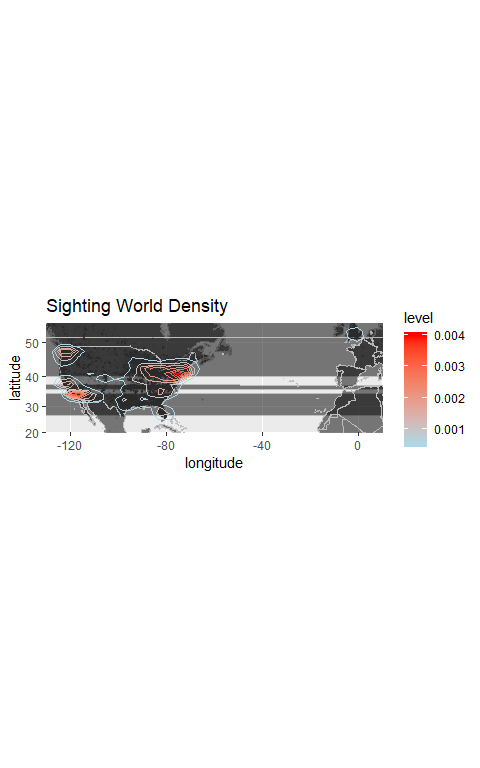


We then wanted to visualize the cluster at our ideal number, 15. You can see above that the size of these clusters is highly varied and the smaller denser clusters are focused on the continental united states.



Additionally, this map shows just the centroid and size of each of the clusters from the previous graph. This visualization just shows how incredibly concentrated the data is on the US. The size of each cluster is so much larger in the US clusters. Also, just the amount of US clusters when compared to international clusters shows a stark difference.

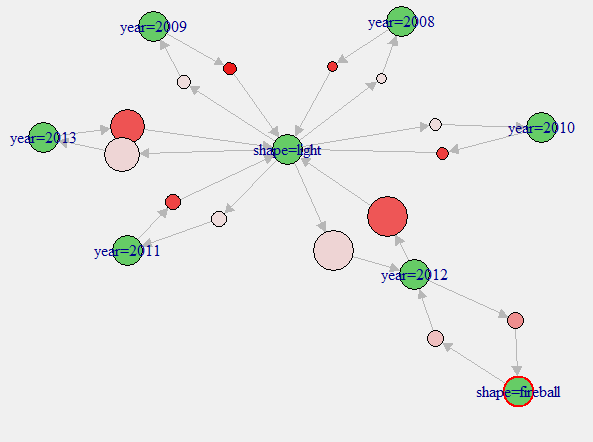
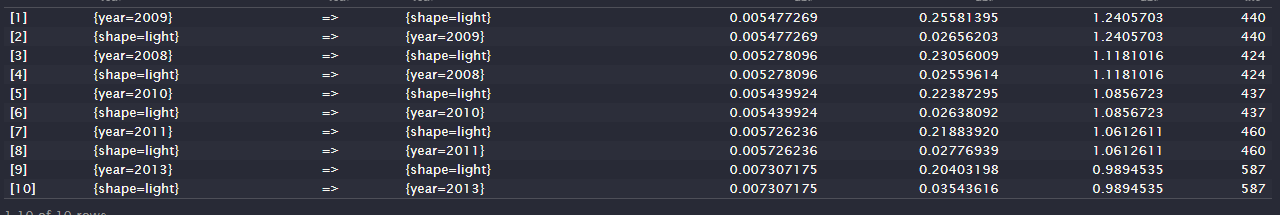




To further reinforce the concentration of these sightings we preformed a 2D stat density plot, essentially a topographical map, drawing polygons circling the areas of highest density. You can see that this shows the same bias towards the united states that was reflected in the kmeans clustering.

# 

Our next attempt at some more advanced data analysis techniques was to mine our data for association rules. More specifically, our first attempt was to find rules associating duration with shape.  We had to set support and confidence parameters fairly low in order to avoid a enormous list of associations, but even with that the existing association rules were not incredibly strong. Although there were many rules that could be pulled out, they represented an arbitrary understanding of the data. Essentially, no valuable inferences could be made from these rules.

We then repeated this process on an additional grouping of shape and year. Unfortunaly we had employed the same techniques to filter out just the strong rules. Alas, these rules appear to be a result of the bias on the data around the early 2010’s. No valuable inference could really be made.

Another technique was to The corpuses are divided up by era, except for the first that covers the reports from 1910-1950 due to the sparsity of reports in relation to subsequent decades. Stopwords were not removed using a standard stopword package, as I wanted to keep certain words such as “I,” “up,” and “near” that would be relevant to the subject. Thus, stopwords were filtered manually. The analysis is done using Python NLTK packages.

**1910-1950**

**Frequency Distribution:**

**Most commonly used words:**

i 35

was 33

object 30

over 20

disk 17

my 17

saw 16

shaped 32

sky 14

ufo 14

this 12

craft 11

above 10

up 10

like 9

our 9

me 8

silver 8

sound 8

around 7

bright 7

early 7

light 7

lights 7

moving 7

objects 7

old 7

round 7

sighting 7

white 7

ball 6

cigar 6

flying 6

large 6

looked 6

near 6

out 6

saucer 6

before 5

house 5

looking 5

minutes 5

observed 5

time 5

us 5

across 4

fire 4

ground 4

high 4

oval 4

roswell 4

seen 4

small 4

summer 4

we 4

appeared 3

away 3

encounter 3

evening 3

event 3

flew 3

flight 3

hovered 3

hovering 3

later 3

military 3

orange 3

playing 3

red 3

saucers 3

slowly 3

straight 3

trees 3

walking 3

witnessed 3

traveling 2

family 2

long 2

experience 2

lighted 2

shape 2

black 2

discs 2

overhead 2

pictures 2

thought 2

approached 2

night 2

cylinder 2

students 2

see 2

fireball 2

late 2

circular 2

atmosphere 2

clear 2

described 2

occurred 2

cylindrical 2

without 2

heading 2

planes 2

far 2

working 2

happened 2

outside 2

down 2

site 2

now 2

check 2

their 2

huge 2

neighbors 2

roaring 2

diameter 2

color 2

dark 2

fast 2

disks 2

behind 2

sometime 2

between 2

many 2

upper 2

unmoving 2

side 2

morning 2

woke 2

front 2

off 2

sure 2

exact 2

been 2

speed 2

where 2

close 2

jet 2

glowing 2

radar 2

screen 2

witness 2

approximately 2

submerging 2

**Notable Bi-grams:**

('cigar', 'shaped'),

('historical', 'sighting')

('moving', 'from')

('object', 'moving'),

('shaped', 'object'),

('early', 'morning'),

('full', 'moon'),

Encounters from 1910-1950 are described as mostly being witnessed alone, shown by the higher use of words such as “I”, “me,” and “my” as opposed to “we,” “us,” and “their.” Words used to describe the sightings are mostly neutral with some positive, such as the objects being described as “playing” in the sky. Many entries were not entered by the first hand witness, so they are very generic and to the point.

A variety of shapes are used to describe these sightings, but are most often cited as being “cigar shaped” and black or silver in color, with lights being white, red, or orange. Objects were cited as being both noisy and noiseless and were seen at night or early morning. Several sightings are reported as specifically occurring during a full moon.

Words used to describe the shape: cigar, disk, cylindrical, circular, oval

Colors: Black, red, orange, silver, white

Other attributes: Huge, glowing, speed, bright, lights, hovering

**1951-1960**

**Frequency Distribution:**

**Most commonly used words:**

i 107

was 106

object 95

over 80

my 50

light 47

saw 44

shaped 44

sky 42

ufo 40

lights 38

cigar 32

saucer 31

we 31

craft 30

out 30

about 28

flying 27

seen 26

night 24

observed 21

across 20

large 20

moving 20

speed 20

bright 19

white 19

objects 19

sighting 19

up 19

disk 18

hovered 18

formation 18

old 16

time 16

near 14

off 14

above 14

shape 13

silver 13

glowing 12

sound 12

me 12

hovering 12

moved 12

around 12

saucers 12

car 11

disc 11

daylight 11

fast 11

orange 11

us 11

day 11

ground 10

slowly 10

round 10

before 10

horizon 10

field 10

down 10

our 9

dark 9

landed 9

they 9

during 9

air 9

summer 9

silent 9

after 9

straight 9

looked 9

people 8

overhead 8

sight 8

appeared 8

long 8

oval 8

stars 8

ufos 8

color 8

stopped 8

slow 8

see 7

back 7

metallic 7

aircraft 7

fireball 7

low 7

side 7

red 7

huge 7

star 7

early 7

stationary 7

disappeared 7

traveling 7

altitude 7

circular 7

other 7

morning 6

next 6

made 6

look 6

spheres 6

different 6

force 6

suddenly 6

radar 6

ball 6

silently 6

away 6

several 6

great 6

front 6

small 6

top 5

clear 5

evening 5

watching 5

crash 5

size 5

sped 5

friend 5

sitting 5

turn 5

close 5

them 5

heard 5

sphere 5

direction 5

flew 5

rate 5

standing 5

black 5

family 5

diameter 5

gone 4

spherical4

clouds 4

hover 4

jet 4

hill 4

base 4

believe 4

witness 4

sighted 4

visit 4

fire 4

jets 4

leaving 4

toward 4

watched 4

together 4

shot 4

driving 4

road 4

many 4

football 4

motionless 4

flashing 4

glass 4

approached 4

**Notable Bi-grams:**

('blinking', 'lights'),

('childhood', 'encounter'),

('cigar', 'shaped'),

('different', 'directions'),

('disc', 'darting'),

('disc', 'shaped'),

('disk', 'hovering'),

('fireball', 'headed'),

('flying', 'disks'),

('flying', 'fast'),

('flying', 'saucers)'

('hovered', 'overhead'),

('in', 'formation'),

('incomprehensible', 'speed'),

('large', 'craft'

('light', 'zigzagging'),

('looking', 'towards'),

('major', 'highway'),

('object', 'flying'),

('object', 'zigzagging'),

('objects', 'flying'),

('orange-ish', 'saucer'),

('orange', 'spheres'),

('oval', 'shape'),

('overhead', 'for'),

('Rectangular', 'shape'

('saucer', 'shaped'),

('shaped', 'Disc'),

('shaped', 'object'),

('silver', 'disk'),

('silver-blue', 'disc'),

('spheres', 'observed'),

('ufo', 'crash'),

('v', 'formation'),

('was', 'gone'),),

('yellowish', 'orange'),

While these encounters are also more often witnessed by a singular person, there is a greater number of those witnessed by one or more people, most often while at home or while on the road traveling. .

These sightings expand to those seen visually and on radar, and are reported as being seen during the morning, day, and night. While most are noiseless, some are accompanied by sound. The descriptors of these sights and their actions become both more detailed and varied that those of the previous group. The shapes still include cigar, disk, circle, and oval, but expand to include “football,” “sphere,” “fireball” and “aircraft,” while “cylinder” is absent completely. Colors continue to be black and silver, with the lights being red, orange, and/or white, but colors are described as being “metallic.” Crafts continue to be described as hovering, glowing, and flying at great speeds, but also being able to change directions. Movements are described as “darting,” “zig-zagging” and “in-formation.”

Shapes: Cigar, saucer, disk, spherical, football, fireball, aircraft, circular, oval, spherical,

Colors: black, metallic, orange (and yellow-orange), red, silver (and silver-blue), white.

Other Attributes: bright lights, glowing, hovering, fire, small, long, motionless, flying, speed

**1961-1970**

**Frequency Distribution:**

**Most commonly used words:**

i 298

object 240

over 184

my 165

with 164

light 153

lights 136

sky 130

ufo 121

shaped 119

craft 113

saw 100

bright 98

about 82

hovering 81

saucer 76

sighting 74

we 69

up 67

above 65

seen 64

very 60

night 58

white 56

orange 55

moving 53

large 76

flying 57

disk 56

glowing 46

me 46

objects 43

hovered 43

high 42

our 40

silver 40

across 40

us 39

near 38

around 38

summer 38

back 37

speed 34

red 33

when 33

shape 33

close 32

house 32

slowly 32

home 31

sound 31

appeared 31

small 31

away 31

time 30

sphere 30

ground 30

old 30

observed 29

down 29

clear 29

colored 29

moved 29

disc 29

day 28

circular 28

ball 28

round 28

car 28

cigar 28

looked 27

blue 27

silent 26

air 25

witnesses 24

stopped 23

low 23

star 23

size 20

color 20

field 20

followed 19

daylight 19

straight 19

other 18

traveling 18

witnessed 18

front 18

see 17

altitude 17

ufos 17

moon 17

formation 17

fast 17

overhead 17

several 16

long 16

outside 16

early 16

green 16

flew 16

friend 16

still 15

area 15

directly 15

side 14

family 14

metallic 14

noise 14

new 14

sighted 14

shot 14

looking 14

driving 14

road 14

oval 14

stationary 13

left 13

trees 13

going 13

base 13

friends 13

during 13

dusk 13

many 13

report 13

between 13

huge 13

dark 13

all 13

evening 13

they 13

late 13

chased 12

sight 12

horizon 12

toward 12

watched 12

morning 12

some 12

reported 12

turned 12

tree 12

degree 11

same 11

brother 11

ago 11

before 11

been 11

diameter 11

miles 11

people 11

witness 11

within 11

windows 11

top 11

encounter 11

spheres 11

level 10

military 10

sightings 10

slow 10

behind 10

colors 10

spotted 10

move 10

year 10

only 10

black 10

along 10

first 10

fireball 10

look 10

different 10

directions 10

fire 10

way 10

ship 10

getting 9

larger 9

something 9

hovers 9

young 9

without 9

changed 9

where 9

town 9

turn 9

**Notable Bi-grams:**

('air', 'force'),

('brilliant', 'silver'),

('cigar', 'shaped'),

('colored', 'lights'),

('disk', 'hovered'),

('disk', 'like'),

('doing', 'maneuvers'),

('glowing', 'orb'),

('grey', 'Cylinder')

('hovered', 'motionless'),

('hovering', 'overhead'),

('i', 'witnessed'),

('incredibly', 'fast'),

('large', 'craft')

('large', 'white'),

('military', 'style'),

('multi', 'colored'),

('no', 'noise')

('quiet', 'whirring'),

('reddish-orange', 'sphere')

('saucer', 'shaped'),

('shaped', 'object')

('silver', 'disk'),

('whirring', 'sound'),

('white', 'lights')

With the 1961-1970 reports, we see a continue in the increase of mentions of other witnesses, notably friends and family members. Shapes continue to be described as disk, sphere, circle, oval, and cigar, with cylinder being absent, but grow to include ball, round, star-like, and fireball. Colors continue to be comprised of black and silver, with some being metallic, as well as white, red and orange lights, but expand to include green and blue. Objects continue to regularly being described as being bright and/or having lights or glowing. Sizes are both small and large, speeds as slow and fast, and sound as being silent and noisy.

Types of movements seen expands the most, with objects being reported as flying, hovering, flying straight, in different directions, stopping, being stationary, and flying in formation. Sightings are reported to occur mostly when skies are clear, and during the morning, daytime, evening, dusk, and nighttime. There is an increase in mention of military activity as well.

Overall tone is neutral, with sightings being more mysterious rather than exciting or frightening.

Shape: disk, sphere, circle, ball, round, cigar, star-like, oval, fireball

Colors: black, silver, green, white, orange, red, blue, metallic,

Attributes: bright, lights, large, glowing, fast, slow, low, long

Movements: hovering, flying, straight, stopped, formation, still, stationary, chased, slow, changing directions

**1971-1980**

**Frequency Distribution:**

**Most commonly used words:**

i 643

object 362

over 323

lights 299

light 291

my 268

saw 229

UFO 205

sky 199

shaped 194

craft 183

bright 156

we 150

about 151

large 139

saucer 128

very 127

moving 125

hovering 115

up 111

above 107

night 104

flying 98

seen 93

sighting 93

me 85

white 83

red 97

disk 80

orange 75

sound 74

us 72

house 72

speed 69

time 67

our 66

hovered 65

around 65

glowing 63

shape 62

slowly 62

when 62

objects 60

high 58

across 57

home 56

fast 54

car 53

air 51

round 51

appeared 51

ground 50

looked 49

ball 48

near 48

low 47

disc 47

back 46

oval 45

silent 45

see 44

away 44

green 44

huge 44

moved 42

silver 40

down 40

window 39

summer 39

sphere 39

close 39

friend 38

field 36

blue 35

looking 35

traveling 35

observed 34

flew 34

stopped 34

other 33

driving 33

top 32

star 32

cigar 32

daylight 31

formation 31

noise 30

metallic 30

front 30

road 30

ago 29

went 29

circular 29

long 29

clear 29

small 29

dark 28

area 27

slow 27

late 27

bottom 27

color 26

horizon 26

behind 26

trees 26

tree 26

early 26

same 26

line 26

outside 25

made 24

windows 24

straight 24

followed 23

black 23

altitude 23

Cigar 23

report 23

day 23

colored 23

way 23

left 23

diameter 23

raft 22

disappeared 22

yellow 22

before 22

several 21

friends 21

brother 21

flashing 21

big 20

directly 20

between 20

witnessed 20

colors 20

moon 19

silently 19

size 19

hovers 19

cloud 19

encounter 19

stationary 19

spotted 18

rectangular 18

distance 18

stars 18

middle 18

yards 18

beam 18

stop 18

going 18

another 18

aircraft 18

fire 17

noticed 17

overhead 17

rotating 17

lit 17

witness 17

ship 17

smaller 17

sister 17

happened 17

circle 16

family 16

center 16

sitting 16

watched 16

thought 16

parents 15

crafts 15

direction 15

power 15

level 15

different 15

shot 15

**Notable Bi-grams:**

('aerial', 'maneuvers')

('beamed', 'lights'),

('blinking', 'star')

('bright', 'burning'),

('burning', 'light'),

('changing', 'intensity'),

('chevron', 'shape')

('cloudless', 'sky'),

('colored', 'lights'),

('different', 'colored'),

('disc', 'spinning'),

('floating', 'glowing'),

('flying', 'UFO'),

('incredible', 'aerial'),

('incredibly', 'large'),

('large', 'silent')

('large', 'triangular'),

('lights', 'formation'),

('low', 'flying'),

('metallic', 'saucer'),

('moving', 'erratically')

('objects', 'spotted'),

('red', 'objects'),

('saucer', 'hovers'),

('saucer', 'shaped'),

('spinning', 'silently')

('strange', 'light')

('triangular', 'space'),

('wingless', 'aircraft')

The tone of the descriptions for 1971-1980 sightings has an overall positive tone, with a reaction of wonder. This is noted by the frequent usage of words such as “incredible,” “strange” and in the description of the impressive aerial maneuvers seen during these sightings. These sightings occur mostly at night, with a few still occurring during daytime hours. The presence of sound and silence is noted among the descriptions.

The types of shapes seen are similar to the 1961-1970 reports, with shapes primarily being saucer, disk, round, circle, sphere, star-like and aircraft. Cigar and oval continue to be present but decrease in frequency compared to other shapes, and a continued lack of mention of cylinder shapes, however, some are described as being long.

Colors of crafts and/or lights remain the same as 1961-1970 reports, and continue to be described as bright, large, fast, slow, flashing, glowing, and having lights. We also see new reports describing objects as being dark, and others as blinking or being/having the presence of a beam. Descriptions of movement remain the same as the previous decade, with the addition of spinning and rotating movements. Encounters with alien beings begin to make a more notable appearance as well.

Shape: saucer, disk, round, circle, cigar, star-like, sphere, oval, ball, aircraft, rectangular, triangle, wingless,

Colors: white, red, orange, metallic, blue, green, black, yellow

Attributes: lights, bright, large, fast, glowing, slow, high, stopped, rotating,

flashing, dark, hovers, beam, blinking

Movements: hovering, flying, in-formation, in a line, straight, changing, shooting quickly, spinning, stopping,

**1981-1990**

**Frequency Distribution:**

**Most commonly used words:**

i 428

lights 279

light 248

over 228

my 211

object 196

saw 159

craft 155

sky 153

ufo 147

shaped 141

large 126

bright 118

we 101

hovering 100

above 88

seen 85

night 84

out 84

moving 78

red 77

up 77

sighting 72

very 71

white 71

ball 62

orange 59

flying 57

feet 56

me 56

sphere 56

huge 54

sound 54

objects 53

saucer 52

off 51

car 50

shape 50

speed 50

around 49

slowly 47

observed 46

time 45

away 44

house 44

formation 42

moved 41

round 41

silent 41

high 40

looked 39

near 39

across 38

driving 38

green 38

close 37

our 36

tree 36

us 36

ground 35

glowing 33

star 33

low 32

made 32

colored 31

home 31

disk 30

hovered 30

appeared 29

black 29

down 29

side 29

area 28

clear 28

fast 28

flew 28

traveling 28

field 27

noise 27

outside 26

oval 26

see 26

silver 26

friend 25

other 25

circular 24

disc 24

**Notable Bi-grams:**

('air', 'Force'),

('Bell', 'shaped'),

('blueish', 'green'),

('bright', 'star'),

('bright', 'yellow'),

('cigar', 'shaped'),

('circular', 'lights'),

('circular', 'object')

('cylindrical', 'object'),

('disc', 'shaped'),

('disks', 'moving'),

('green', 'disc'),

('illuminated', 'disks'),

('large', 'helicopter'),

('lights', 'growing'),

('moving', 'disks'),

('moving', 'rapidly'),

('object', 'sighted'),

('octagon', 'shape'),

('parallel', 'path'),

('red', 'circular'),

('saucer', 'shaped'),

('shape', 'pods'),

('shaped', 'craft'),

('shaped', 'objects'),

('Shaped', 'ufo'),

('silver', 'cigar'),

('slow', 'moving'),

('space', 'shuttle'),

('sphere', 'hovering'),

('sudden', 'movement'),

('ufo', 'hovering'),

('unmanned', 'craft'),

('winged', 'craft'),

('yellow', 'lights')

The reports for 1981-1990 return to a more neutral/serious nature, while the array of different sightings continues to increase in variety. As with previous decades, reports include both silent and non-silent objects, and having descriptions ranging from bright, dark, large, small, glowing, flashing, long, slow, and fast. Movement descriptors aren’t as varied as the previous decade, but continue to include the standard hovering, flying, as well as stationary, floating, stopping, changing directions, flying parallel, and flying in formation(s). Colors of objects and lights remain the same as previous decades, but there is an increase in objects being a light or light(s) and being associated with stars. The increase in variety comes with the descriptions in shapes. The terms used to describe the shapes become more specific: ball, circle, sphere, round, orb, square, rectangle, as well as more varied shapes such as octagon, hexagon, helicopter, diamond, aircraft, and football,

Shape: ball, sphere, saucer, disc, circle, oval, star-like, round, football, cigar, aircraft, diamond, rectangular, fireball, orb, cylindrical, octagon, helicopter

Colors: red, white, blue, silver, green, metallic, yellow,

Attributes: Lights, large, bright, huge, low, high, glowing, long

slow, small, dark, flashing,

Movements: hovering flying, formation, stationary, floating, stopped, parallel

**1991-2000**

**Frequency Distribution:**

**Most commonly used words:**

i 2951

light 2135

object 1849

sky 1759

bright 1511

lights 1433

saw 1352

my 1208

over 1018

moving 871

white 794

very 765

we 659

shaped 612

red 594

out 566

green 553

up 530

seen 510

craft 508

no 482

across 470

orange 444

large 437

objects 435

above 423

ufo 394

looked 383

appeared 377

night 374

flying 371

traveling 371

speed 351

high 337

fireball 333

disappear 327

observed 308

ball 305

blue 301

hovering 300

star 299

fast 297

slowly 293

shape 283

moved 280

driving 273

horizon 262

formation 247

looking 231

sound 219

me 214

glowing 211

noticed 210

color 197

small 196

direction 193

low 191

size 187

round 185

home 181

going 180

our 180

other 175

flew 174

near 173

away 169

behind 168

sighting 167

sphere 164

tail 162

clear 157

flashing 156

house 155

car 154

thought 154

us 150

witnessed 149

outside 148

friend 147

straight 146

aircraft 145

flash 143

black 142

circular 142

moon 142

altitude 139

stationary 139

see 138

hovered 137

slow 132

strange 132

disk 131

silver 129

towards 129

line 122

oval 122

yellow 120

colored 119

air 117

move 116

changed 115

plane 115

**Notable Bi-grams:**

('bright', 'crescent'),

('bright', 'orange'),

('bright', 'yellow'),

('brilliant', 'gold'),

('brilliant', 'vertical')

('bright', 'light'),

('chased', 'after'),

('chevron', 'luminous'),

('circle', 'shaped'),

('colored', 'cylindrical'),

('crescent', 'shape'),

('cylindrical', 'object'),

('dark', 'vertical'),

('different', 'directions'),

('disk', 'sighted'),

('egg', 'shaped'),

('extremely', 'fast'),

('fast', 'light'),

('figure', 'floating'),

('flashing', 'lights'),

('flight', 'patterns'),

('gold', 'colored'),

('green', 'luminous'),

('light', 'chevron'),

('light', 'hovering'),

('light', 'moving'),

('lights', 'hovering'),

('low', 'flying'),

('luminous', 'orange'),

('luminous', 'round'),

('moving', 'erratically'),

('moving', 'rapidly'),

('noiseless', 'craft'),

('noiseless', 'object'),

('orange', 'light')

('red', 'lights'),

('round', 'object'),

('saucer', 'shaped'),

('shaped', 'disk'),

('shaped', 'figure'),

('shooting', 'across'),

('silver', 'circle'),

('silver', 'disk'),

('sphere', 'moving'),

('soundless', 'object'),

('triangle', 'formation'),

('umbrella', 'shape'),

('vertical', 'line'),

('without', 'wings'),

('yellow', 'tail'),

('yellowish', 'beams')

1991-2000 continues the trend of increasing variety and detail in descriptions. Colors, attributes, and types of movement remain similar to 1981-1990, and sightings are reported to occur during day and night-times, being silent and non-silent. Shapes cited include previously cited shapes plus some. Reports continue of objects shaped such as fireball, star-like, sphere, tailed, aircraft, circle, disk, oval, plane, triangle, cigar, meteor-like, saucer, orb, satellite, diamond, crescent, cylinder, round, egg, winged, wingless, and even more unusual shapes such as chevron and umbrella.

It’s the adjectives used in describing the sight of these objects that give the corpus an overall positive tone, with the feeling of excitement and wonder. Colors and objects are described beyond “bright” with words like luminous, brilliant, strange and fluorescent.

Shape: fireball, ball, star-like, sphere, tailed, aircraft, circle,

disk, oval, plane, triangle, cigar, meteor-like, saucer, orb, satellite, diamond, crescent, cylinder, round, egg, chevron, umbrella, wingless, winged

Colors: white, red, green, orange, blue, silver, yellow, metallic, gold

Attributes: light, bright, large, speed, high, fast, slow, small, low,

flashing, glowing, fire, dark

Movements: flying, disappearing, hovering, in-formation, changing directions, straight, stationary, changing, floating, beams

**2001-2010**

**Frequency Distribution:**

**Most commonly used words:**

light(ing)(ed)(s) 13077

objectless 7762

sky/s 7183

object(s) 6124

bright(ness)(est)(er)(ly) 5833

star(s)/starlike 5644

I 5529

over/head 5282

shape/d/s 3693

white/ish 3128

me/my/self 2962

red/ish 2684

fly/ies/ing/flew 2448

saw/seen 2366

ufo/s 2249

orange(ish) 2159

very 2147

disc/s 2122

like 1878

hover(ing)(ed)(s) 1800

disappear(ing)(ed) 1779

craft/s 1668

sight(ings)(ing)(ed) 1653

look(ing)(ed)(s) 1598

large/r 1591

move/ing/ed/s 1567

slow(ing)(ly)(er)(ed) 1500

flash(ing)(ed)(es) 1429

circle(ing)(ular)(es)(ed) 1415

green/ish 1408

fast(er) 1407

night/s 1406

we 1342

across 1330

out 1302

stationary 1287

appear(ance)(ing)(ed)(s) 1286

up 1251

about 1235

sped/speed(ing)(s) 1229

blue 1190

above 1165

sphere(s)/spherical 1153

color/s/ered/ful 1128

travel(ing)(ed)(s) 1084

formation 1083

glow(ing)(ed) 1041

change(ing)(ed)(es) 992

fireball/s 905

around 900

small/er 900

Strange/ly 896

round(ed) 814

near/by 803

low/er 797

minute/s 792

ball 750

orb/s 725

off 724

sound(s) 720

second/s 706

oval/s 697

observe(ed)(er)(ing)(ion) 693

day-light/time 672

time/s 663

spot/ted 649

possible/ly 647

cloud/y/s 633

silver/ry 625

other/s 596

triangle(ular)(s) 587

black 580

direction 580

shot/shoot(ing)(s) 575

silent/ly 550

area 546

yellow(ish) 531

report/ed/s 523

turn(ing)(ed)(s) 505

clear/ly 493

toward/s 493

make(ing)(s)/made 490

witness(ed)(es) 486

linear/line/s 482

blink(ing)(ed) 476

down 474

after 469

notice/d 460

photo(graph)(ed)(s)/picture(s)456

drive/ing 455

head/(ing)(ed)(s) 452

side(s)(ways) 447

cigar 445

our 445

place 438

stop(ing)(ed)(s) 438

goes/going 430

see/ing 428

take(ing)(en)(s) 420

big/ger 416

quick/ly 410

follow(ing)(ed)(s) 407

straight 404

vanish(ing)(ed)(s) 404

altitude 393

saucer/s 390

fall(ing)(s) 389

float(ing)(ed)(s) 386

horizon 386

ground 376

jet/s 374

dark/ness 373

air 361

close/ly/er 360

fade(ing)(ed)(es) 358

outside 351

seem(ingly)(ed)(s) 347

extreme/ly 341

trail(ing)(s) 337

us 334

highway 332

moon 332

cylinder/ic 328

pulsing/ating 328

several 327

dim(ing)(ed)(er)(ly)(s) 324

size/d 319

rectangle/ular 316

watch(ing)(ed) 316

huge 314

pattern/s 314

friend/s 313

long/er 313

metallic 313

fire 312

sudden/ly 302

last(ing)(ed) 293

shiny/ing 293

streak(ing)(ed)(s) 291

space 289

rapid/ly 287

satellite 285

Venus 285

diamond 284

different 284

mile/s 278

meteor/ite 270

thought 269

behind 268

aircraft 265

during 256

noise 254

airplane(s)/airliner 253

morning 248

rate 246

front 243

something 243

pass(ing)(es)(ed) 241

point(ing)(ed)(s) 235

another 233

balls 233

erratic/ly 223

amber 221

multiple 219

tail 217

missile 216

car 214

window/s 214

movement(s)/moving 211

illuminate(ing)(tion)(ent) 210

many 210

brilliant 208

more 203

cross(es)(ed)(ing) 202

top/s 199

view(ing)(ed) 199

Due to the size of the document, different tenses of the same word were combined in order to get the full range of frequently used terms. This did not change the order of frequency of these terms in any significant way. Bigrams could not be extracted due to the length of the corpus.

2001-2010 is the first time period where "I" is not the more frequently used word (even before stemming/combining). This grouping has more than twice as many different shapes described and including shapes previously sighted. Many are compared to actual objects: rock-like, dome, blimp, boomerang, balloon, basket, missile, shuttle, jet, cigar, egg, rocket, bell as well as general shapes such as circle, sphere, round, ball, orb, oval, triangle, elongated, and oblong. The range of colors expand to include amber, gold, pink, and brown.

Movements described are similar to previous decades, but with a wider variety of terminology used. Objects are described as moving in unison, parallel to one another, disappearing and re-appearing. Movements are further described using more colorful terms such as zooming, darting, bounding, chasing, dancing, fading, shooting, rotating, and zig-zagging. Sightings occur all times of the day and during sunny, clear, and cloudy skies.

One noticeable shift is the increase in mention of technology and capturing visuals of the event. While previous decades had occasional mentions of telescopes and binoculars, we see frequently used words such as camera, photos, phone, video, etc.

Another noticeable shift is the move towards use of military-type terms in descriptions. While previous decades had the occasional use of “military,” “shuttle,” and “jets,” 2001-2010 has plentiful use of terms such as missile, military, stealth, jet, aircraft, police, drone, and shuttle.

The tone of the text is largely negative and one of anxiety. In addition to the heavy use of military aircraft related terms, these sightings are also described as inciting fear and terror. Other negative terms such as “scary,” “strange,” “force,” and “suspect” are joined with terms of something out of the ordinary such as irregular, unknown, unusual, odd, erratic, impossible, and mysterious.

But while the overall tone of this decade is negative, there is still a significant presence of positive descriptors. The objects themselves are sometimes described as beautiful, brilliant, incredible, and amazing. There is also a heavier emphasis on attempts to explain sightings as known celestial objects, such as Jupiter, Mars, Venus, or comets.

Shapes: star-like, disc, circle, sphere, fireball, round, ball, orb, oval, triangle, cigar, saucer, jet, cylinder, rectangle, satellite, diamond, meteor-like, aircraft, airplane, pointed, ball, missile, dots, helicopter, chevron, tear-drop, egg, boomerang, winged, balloons, clusters, oblong, cone, shuttle, comet-like, fireworks, square, crescent, football, ring, globe, rocket, blimp, dome, bell, cell, chemtrails, rock, elongated, halo, arc,

Colors: white, red, orange, green, blue, silver, black, yellow, metallic, fiery, amber, gray, gold, blue, pink, brown.

Attributes: lights, bright, large, slow, flashing, fast, glowing, small, changing, low, big, pulsating, dim, huge, shiny, streaking, beams, flaming, twinkling, solid, cloud-like, strobing, reflecting, broad, hazy, thin, giant, burning, dull, sparks, exploding, metal, vapor, translucent, humming, connected, fuzzy.

Movements: hovering, disappearing, flying, stationary, reappearing, in formation, changing directions, shooting, turning, in a line, blinking, sideways, stopping, quick, straight, floating, fading, trailing, patterned, zig-zagging, rotating, dropping, motionless, chasing, dancing, steady, splitting, darting, parallel, accelerating, gliding, crashing, ascending, unison, zooming, hanging, shifting, bouncing,

**2011-2015**

**Frequency Distribution:**

**Most commonly used words:**

light(ed)(s) 8605

bright(er)(ly) 4975

sky/ies 4695

orange(ish) 4159

over(head) 3681

object(s) 3455

move(ments)(ment)(ing)(ed)(s) 3256

red(ish) 2796

seen/saw 2490

i 2337

white 1709

fly(ies)(ing)/flew 1598

shape(s)(ed) 1551

look(ing)(ed) 1465

orb(s) 1356

disappear(ing)(ed)(s) 1272

fireball/s 1262

hover(ing)(ed)(s) 1160

ufo/s 1117

my 1086

slow(ly) 1037

ball 1035

night(s) 1018

across 977

star(like)(s) 929

circle(ing)(s)(ular) 928

glow(ing) 893

flash(ing)(ed)(s) 885

sphere(s)(ical) 867

formation 866

large(er) 811

sight(ing)(ed)(s) 796

fast(er) 764

craft(s) 755

travel(ing)(ed) 726

green 722

appear(ing)(ed)(s) 704

blue 689

above 668

we 628

color(ed)(s) 625

speed(s) 574

low(er) 568

fier/y 553

round 524

change(ing)(ed)(s) 495

high 476

strange 432

sound 431

small(er) 427

line/s 421

near 420

witness(ed)(es) 416

yellow(ish) 412

direction 372

come(ing)/came 364

head(ing)(ed) 352

oval 349

time(s) 329

down 321

other 318

spot(ted) 315

cloud/s 314

floating 312

aircraft/plane 305

go(ing) 303

toward/s 301

observed 295

fade(ing)(ed)(s) 286

possible/ly 283

vanish(ing)(ed)(es) 281

before 272

stationary 271

silent/ly 270

turn(ing)(ed)(s) 270

shot/shoots/shooting 264

straight 261

followed/ing 259

same 252

driving 251

pulsing/ating 249

black 245

see(ing) 240

me 239

multiple 238

quick/ly 235

several 235

blinking 232

outside 225

triangle(ular) 224

plane(s) 220

our 217

big(ger) 216

noticed 214

close/r 211

stop(ing)(ed) 208

horizon 206

altitude 205

disc 204

each 195

ground 195

silver 191

house 189

dark 182

pattern(s) 180

amber 179

different 177

home 166

air 165

noise 163

long 162

diamond 161

behind 157

county 155

huge 153

photo(s)/Picture(s) 150

taking/took 149

extremely 148

cigar 147

last(ing)(ed) 140

clear 139

falling 139

few 139

As with the 2001-2010 corpus, combining of stemmed words is utilized in order to get the full range of frequently used terms. This did not change the order of frequency of these terms in any significant way. Bigrams also could not be extracted due to the length of the corpus.

The reports since 2010 have moved away from the military-attack-anxiety, though it is still present, and continues with efforts to explain sightings with known celestial objects such as Venus, Jupiter, and Sirius, and the presence of technological references, such as phones, camera, videos, while not as strong as the previous decade, is also still present.

The tone of the text is less negative than 2001-2010, and more neutral with emphasis on the unknown, with common terms such as strange, unidentified, suspect, odd, weird, unknown, and unusual. There is also the presence of a number of cryptic and ominous entries that don’t go into much detail beyond being cryptic such as “They came back...” and “THERE'S SHIPS UP THERE.” Shapes, colors, and other attributes are less numerous and varied than the previous decade, but remain more diverse and specific that other previous decades.

Shapes: orb, fireball, star-like, circle, sphere, round, oval, aircraft, triangle, plane, disc, cigar, jet, cluster, rectangle, cylinder, helicopter, meteor, saucer, dots, fireworks, satellite, cigarette, tear drop

Colors: orange, red, white, green, blue, yellow, black, silver, amber, metallic, gray, gold, purple,

Attributes: lights, bright, slow, glowing, flashing, large, fast, low, fiery, high, pulsating, blinking, big, dark, flaming, trail, dim, erratic, flickering, solid, trail, beams, flare, twinkling, shiny, thin, illuminating, strobe,

Movements: flying, disappearing, hovering, formation, changing, stationary, turning, straight, stopping, spinning, streaking, dancing, dropping, rotating, descending, chasing, zoomed,

# Conclusions

After thorough investigation of the data set, the answers to what the UFO sighting phenomena truly represents, if anything, remains as elusive as ever. Relatively few observations conformed to cultural expectations of UFO shape compared to vague sightings of some sort of light in the sky. With rapidly increasing drone technology and military interest in the subject, perhaps these sightings should be paid attention to on a case by case base and always not written off as delusions of little green men. The authors were reminded of the closing of the official government UFO study, Project Bluebook, which stated that there was nothing to be gained from scientifically studying the phenomena at this point. Something way be actually witnessed but the vague descriptions and imprecise reporting making coalescing a single explanation extremely difficult, if not likely impossible. The only unifying phenomena may be a mental one, which desires everything seen in the sky to fit neatly into a single box.