

Reporte de Avistamientos de OVNI's en los Estados Unidos de América

Carlos González Munguía

22 de mayo de 2015

1. Introducción

Los eventos más importantes para aprender más acerca de los avistamientos de OVNI's, son aquellos en los cuales existen múltiples testigos que reportan el mismo evento. Los avistamientos de OVNI's son un fenómeno mundial, con reportes de casi todo el mundo; sin embargo, algunos países tienen más reportes que otros, tal es el caso de los Estados Unidos, del cual se puede llegar a pensar que es debido a su alta densidad poblacional, pero países como México, Alemania e India tienen un número insignificante en proporción al caso de Estados Unidos. Entonces, la idea principal del presente resumen es entender factores culturales involucrados implícitamente derivados de las narrativas escritas por los testigos.

1.1 Objetivo

El principal objetivo de este reporte es realizar un análisis de la base de datos **UFO**¹ principalmente del país de los Estados Unidos de América con el fin de saber cuándo y dónde ocurren los eventos, identificando posibles causas asociadas a la demografía de la gente y las narraciones de los textos escritos de cada uno de los eventos reportados por la gente. Así como también, realizar un modelo de predicción estadístico del número de avistamientos. Dicha base tiene avistamientos desde 1969 hasta el 2015.

1.2 Metodología

Cada archivo fue descargado de la dirección electrónica que aparece al pie de página y concentrado en un sólo archivo que fue cargado a una base de datos *PostgreSQL* con el fin de obtener estadísticas básicas y posteriormente el mismo archivo fue analizado con el paquete estadístico *R* en el que se realizó el modelo y la creación del presente documento.

2. Obtención y Minería de datos

2.1 Obtención de datos

La obtención de datos se obtuvo por medio de la técnica de *scrapeo* el cual consiste en realizar una búsqueda determinada en el buscador y recolectar los datos. La base final tiene un total de 97,148 registros con 6 variables las cuales brindan información de la Fecha, Ciudad, Estado, Forma del Ovni, Duración del avistamiento y Narrativa del evento.

Posteriormente los datos fueron “*tratados*” (limpieza) para tener una correcta lectura de los mismos y obtener estadísticas más confiables; principalmente se realizaron correcciones en cuanto a estandarizar el formato de la fecha y la correcta identificación (estado y/o país) del lugar del evento con el fin de separar los avistamientos de Estados Unidos y el resto del mundo.

El conjunto de datos se observa de esta manera:

¹<http://www.nuforc.org/webreports/ndxevent.html>

date	city	state	shape
2015-04-23	Rock Hill	SC	RECTANGLE
2015-04-23	Mt. Orab	OH	SPHERE
2015-04-23	Columbia City	IN	DISK
2015-04-23	Middletown	RI	UNKNOWN
2015-04-22	Boulder	CO	CIRCLE

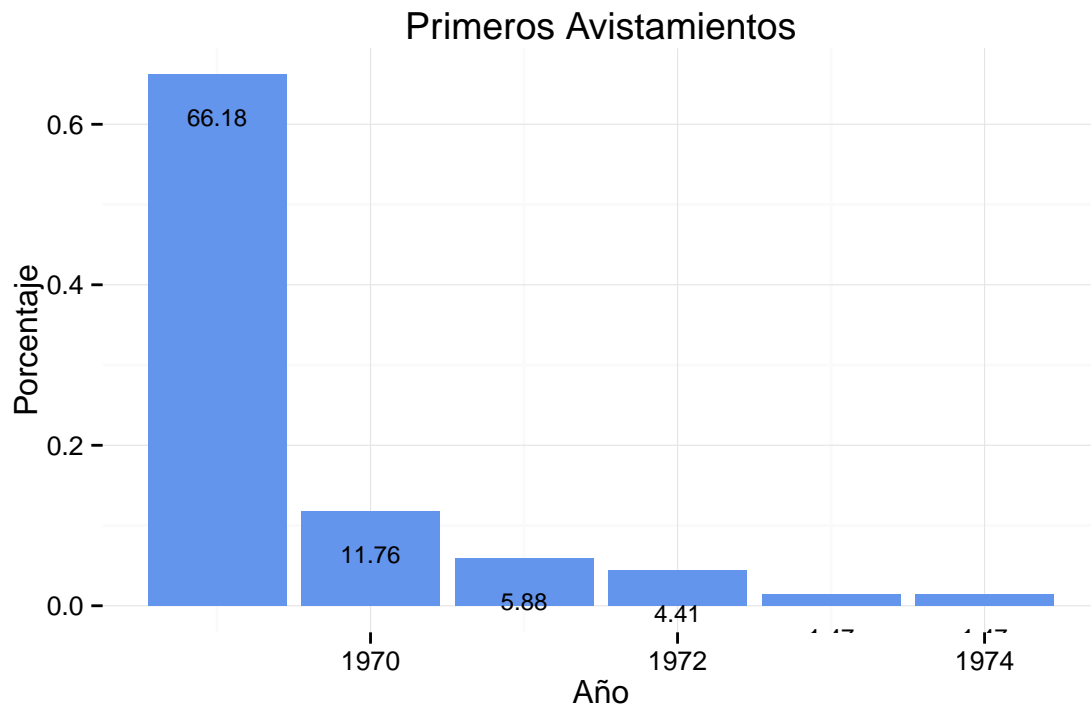
Cada uno de los registros anteriores, viene acompañado con una narrativa de este estilo.

Flying saucer with blinking lights that zigzagged.I was intensely looking all over the sky attempting to see a meteor shower when I spotted what I first thought was an airplane, but I am 99% sure it was not.It was saucer shaped with 3-4 visible horizontal lights, one red, the others white, that flashed one after the other in a row.The saucer stayed almost perfectly in place except for zigzagging every 10 seconds or so.I watched it for 20 minutes before I got too cold and went back inside.

2.2 Minería de datos.

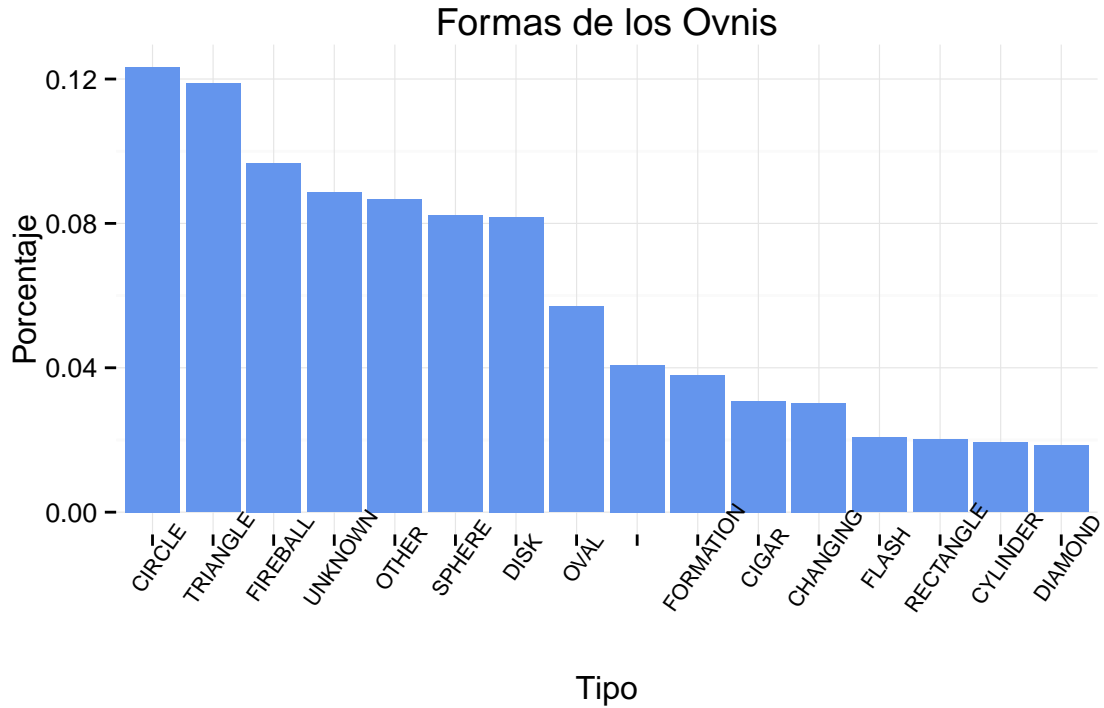
Antes de empezar con la minería de datos, se obtuvieron estadísticas exploratorias y descriptivas que con el fin de entender qué está pasando con el fenómeno *OVNI* y con esto poder realizar el modelo de predicción de avistamientos.

Realmente se tienen variables categóricas, por lo que se necesita conocer la distribución de cada una de las variables con respecto de las otras; por ejemplo, se sabe que la base de datos tiene registros desde 1969, entonces los primeros avistamientos dado el año serían.



Se tiene que el **78%** de los primeros avistamientos por estado ocurrieron entre 1969 y 1970, sin embargo, existen estados como Las islas vírgenes y Puerto Rico donde su primer avistamiento ocurrió hasta después de 1975; esto puede deberse por el año de incorporación como estado a Estados Unidos.

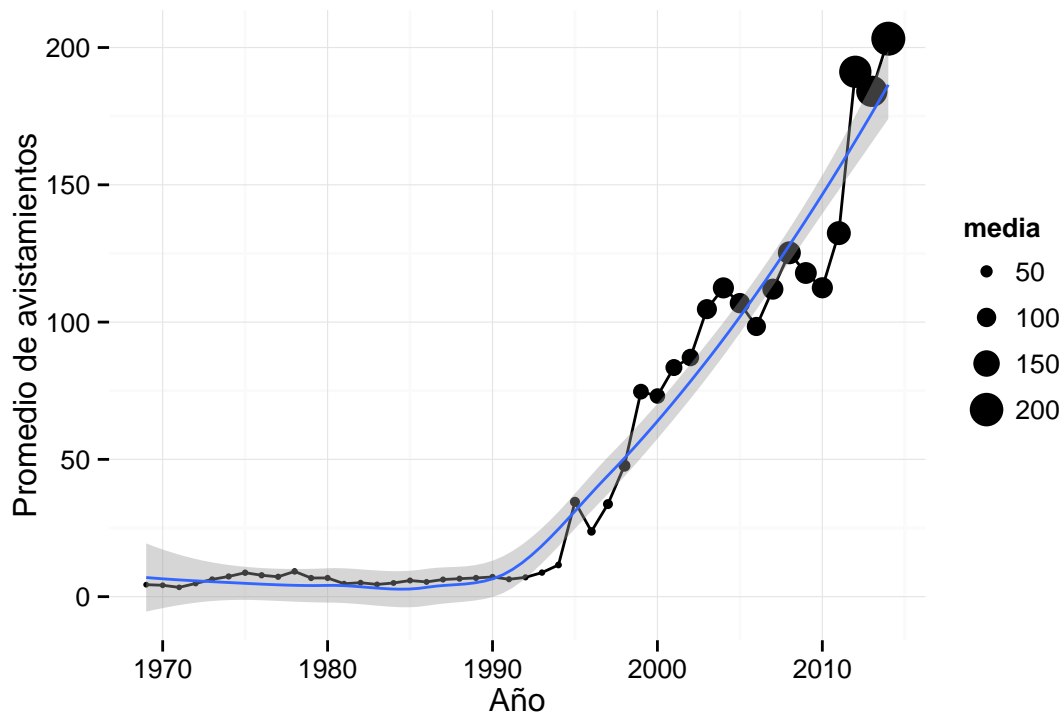
Ahora bien, obteniendo estadísticas del “*tipo*” del Ovni, se tiene que el **34%** de las formas que se han observado son *circulares*, *triangulares* y *bolas de fuego*. Como dato curioso el 8.9% de las formas han sido “*desconocidas*” ocupando la cuarta posición de formas más vistas. Cabe destacar que el 20% de las formas eran *light-luces* pero al no ser un tipo de formas, fue filtrado de las estadísticas anteriores.



El mes con más avistamientos es Julio con 963.33 avistamientos promedio por mes, seguido de Agosto con 868.42 avistamientos promedio y el mes con menos avistamientos promedio es Febrero con 463.17. Entonces, no importando el mes, se tienen más de 450 avistamientos promedio al mes.

mes	media
julio	963.3
agosto	868.4
junio	840.3
septiembre	768.9
octubre	752.6
noviembre	669.6
enero	578.0
diciembre	569.4
marzo	545.1
abril	539.5
mayo	537.3
febrero	463.2

Calculando los promedios de avistamientos por año, se obtiene lo siguiente:



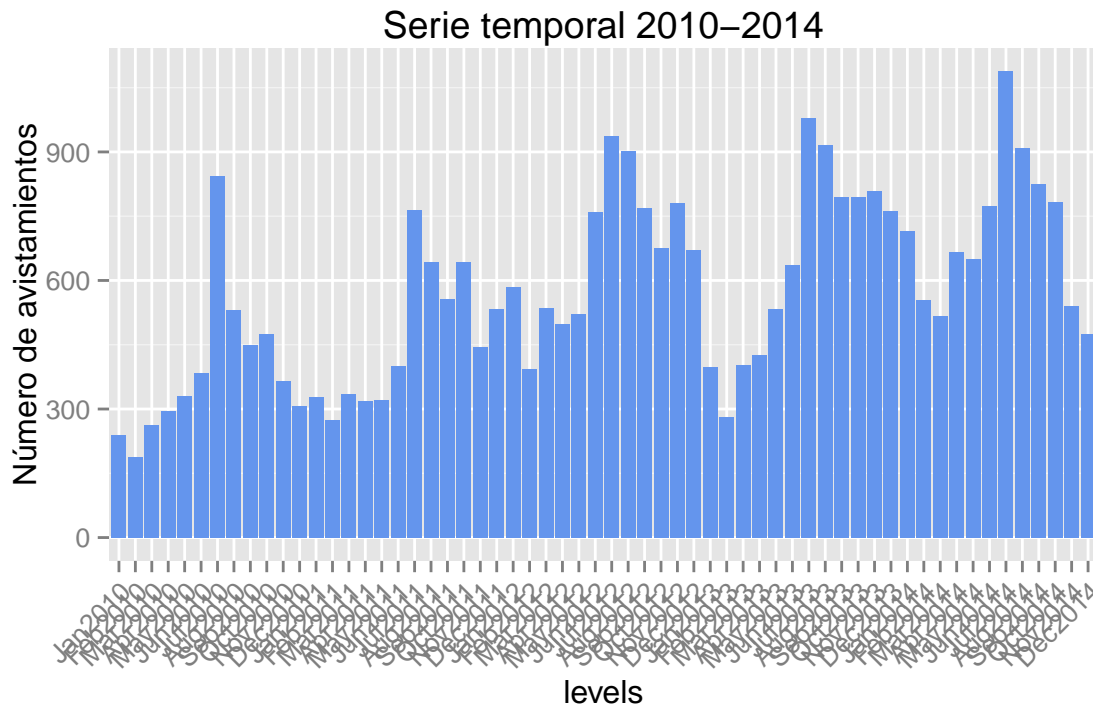
Claramente se tiene una tendencia creciente a principios del año 2000, pasando de 74.66 avistamientos promedio en el año 1999 a 203.21 en el año 2014, es decir, el incremento de avistamientos promedio fue de **274.32%** en los últimos 14 años.

Realizando un análisis por día de avistamiento, se tiene que los fines de semana son los días donde más avistamientos ocurren con un **46.85** del total.

day	count
Sat	17113
Fri	14257
Sun	14255
Thurs	13339
Wed	13220
Tues	12831
Mon	12133

Todo el análisis exploratorio anterior indica que al parecer existe una tendencia en el incremento del número de avistamientos relacionado con la temporalidad en que la gente tiene más tiempo libre; por una parte en los meses de julio y agosto que son las vacaciones de verano y los fines de semana que son los días en los que la gente comúnmente descansa. Cabe destacar que en tiempos decembrina, la gente tiene vacaciones pero está tan ocupada con todos los eventos de las fechas que se olvida de mirar al cielo.

Entonces, si lo anterior es cierto se esperaría tener un comportamiento de la serie de tiempo estacional donde los máximos se alcancen en los meses de julio y posteriormente decaigan en los meses de diciembre y enero.



Por lo que las olas de avistamiento parecen ser cada año donde se alcanza el máximo es en Julio y el mes donde se alcanza el mínimo es febrero; el comportamiento es similar en todos los años.

Análisis de narrativas

Para realizar el análisis de las narrativas, se creó una matriz de términos documentos con la finalidad de un mejor tratamiento de los datos. Todas las narrativas tuvieron un proceso de depuración en cuanto a quitar palabras función como artículos, signos de puntuación, “*stripeo*” que se refiere a quedarse con la raíz de cada palabra y finalmente se pasaron todas las palabras a minúsculas.

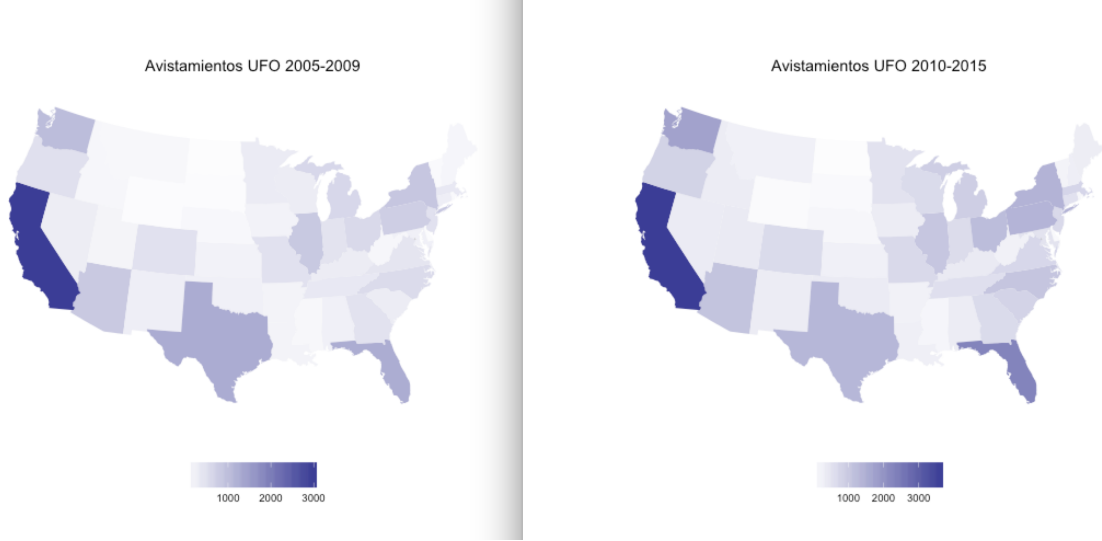
De todo lo anterior, se decidió hacer un dendograma que agrupa las palabras que tienen una relación de más del 85%.



Observando el dendograma se tiene aproximadamente 5 clusters, por ejemplo, uno con las palabras *sky*, *bright* y *move* que hace alusión a “movimientos en el cielo brillantes”, otro *object* con *light* y así sucesivamente. Realmente el cluster de color verde con *look* y *like* hace alusión a “luce como” y esta dentro de la rama de las palabras de color naranja, es por eso que se puede ver como uno solo.

Análisis geográfico

Dado los resultados anteriores, se puede observar que el número de datos es representativo en los últimos 10 años en comparación al resto anterior de los años, es por esto que se ha decidido comparar estos últimos 10 años de número de avistamientos por estado. Graficando lo anterior se tiene:



Realmente no hay una diferencia significativa en cuanto al número de avistamientos en los últimos 10 años, simplemente el número de avistamientos incrementa pero en cuanto al comportamiento en cada estado sigue siendo el mismo.

3. Modelo de predicción.

Para realizar las predicciones del número de eventos por estados, uno de los modelos más adecuados para realizar esta tarea es por medio de Métodos Monte Carlo vía Cadenas de Markov (MCMC) ya que son Formas analíticas no cerradas y funcionan bien con situaciones donde se tienen datos longitudinales y espaciales.

El modelo a utilizar es el siguiente :

$$x[i] \sim \text{Poisson}(l[i])$$

$$\log(l[i]) \sim \alpha + \beta * \log(\gamma[i]) + e[i]$$

$$\alpha \sim N(0, \sigma_1)$$

$$\beta \sim N(0, \sigma_1)$$

$$\sigma_1 \sim \text{gamma}(.001, .001)$$

$$\sigma_2 \sim \text{gamma}(.001, .001)$$

Donde

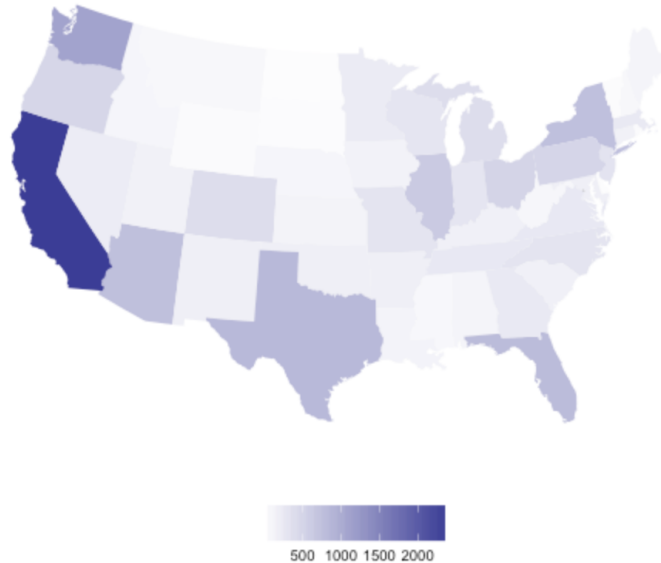
$x[i]$ es el número de avistamientos por estado.

$\gamma[i]$ es la densidad de la población

Como se observa en el modelo, se realizó un suavizamiento *Poisson* de los avistamientos.

Resultados

Realizando las predicciones de avistamientos del año 2015, se observa un comportamiento similar a los “bloques” de historia anteriores. Visualmente, no se tiene algún especie de incongruencia por estado.



Conclusiones

Podemos observar que la mayoría de los avistamientos se dan en los estados que se encuentran en la frontera con México, el Océano Pacífico y Atlántico. Sin embargo en la parte de la frontera con México y el Océano Pacífico hay una densidad de población baja, pero en la frontera con el océano Atlántico hay una alta densidad de población, por tanto podemos concluir que el número de avistamientos no está relacionado con la densidad de población y está relacionado con los periodos vacacionales de la gente ya que en las temporadas de verano, es decir, entre junio y agosto, en todos los años se reporta el máximo de avistamientos; además, los fines de semana son los días que más avistamientos se reportan.

Por otra parte, la mayor variabilidad en el modelo se encuentra en los estados donde existe menor densidad de población y toda la frontera de estados.

Bibliografía

- <http://www.nuforc.org/webreports/ndxevent.html>
- <http://www.jstatsoft.org/v51/i07/paper>
- <http://rpubs.com/gaston/dendrograms>
- http://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States

Preguntas de interés-Anexos.

¿Primer avistamiento en cada estado?

	state	date
1	LA	1969-01-01
2	NV	1969-01-01
3	PA	1969-01-01
4	FL	1969-01-04
6	MT	1969-01-15
7	IN	1969-02-15
8	IL	1969-03-01
9	NE	1969-03-01
10	NJ	1969-03-01
11	OK	1969-03-15
12	WA	1969-03-15
13	GA	1969-04-01
14	SC	1969-04-04
15	CA	1969-04-05
16	HI	1969-04-15
17	AR	1969-04-21
18	CT	1969-05-01
19	MD	1969-05-01
20	OH	1969-05-01
21	MO	1969-05-15
22	TX	1969-05-15
23	BC	1969-06-01
24	CO	1969-06-01
25	KY	1969-06-01
26	MI	1969-06-01
27	MS	1969-06-01
28	NY	1969-06-01
29	VA	1969-06-01
30	OR	1969-06-10
31	ON	1969-06-11
32	MA	1969-06-30
33	AL	1969-07-01
34	AZ	1969-07-01
35	MN	1969-07-01
36	AK	1969-07-05
37	IA	1969-07-10
38	UT	1969-07-10
39	NC	1969-07-27
40	MB	1969-08-07
41	KS	1969-08-15
42	VT	1969-08-15
43	ME	1969-08-27
44	TN	1969-10-13
45	DE	1969-10-15
46	NH	1969-10-15
47	AB	1970-04-15
48	RI	1970-06-15
49	WY	1970-06-20
50	PR	1970-07-13
51	ND	1970-07-20
52	NS	1970-08-14
53	NB	1970-08-15

	state	date
54	NM	1970-12-05
55	SA	1971-01-10
56	WV	1971-05-01
57	WI	1971-06-10
58	PQ	1971-08-15
59	SK	1972-06-15
60	QC	1972-07-01
61	SD	1972-07-15
62	ID	1973-07-24
63	DC	1974-06-06
64	NF	1975-01-01
65	PE	1975-07-15
66	NT	1978-06-01
67	YK	1985-05-15
68	YT	1993-09-01
69	VI	1997-10-25

¿Primer avistamiento de cada forma?

shape	date
OTHER	1969-01-01
TRIANGLE	1969-01-01
UNKNOWN	1969-01-01
OVAL	1969-01-04
LIGHT	1969-01-15
RECTANGLE	1969-01-15
SPHERE	1969-01-15
DISK	1969-01-20
CIGAR	1969-02-12
CROSS	1969-03-15
CYLINDER	1969-04-21
CIRCLE	1969-06-01
FORMATION	1969-06-01
	1969-06-30
CHANGING	1969-06-30
CONE	1969-06-30
EGG	1969-06-30
FIREBALL	1969-07-11
FLASH	1969-07-15
TEARDROP	1970-04-17
CHEVRON	1970-09-15
DIAMOND	1970-09-15
DELTA	1974-07-18
ROUND	1995-12-14
DOME	1996-03-15
CHANGED	1996-06-24
PYRAMID	1997-01-11
CRESCENT	1997-03-22
FLARE	1997-11-30
HEXAGON	1997-12-16
TRIANGULAR	2015-03-09

¿Promedio de entre avistamientos, por mes, por año? ¿Por estado?

Promedio de avistamientos por mes

mes	media
julio	963.3
agosto	868.4
junio	840.3
septiembre	768.9
octubre	752.6
noviembre	669.6
enero	578.0
diciembre	569.4
marzo	545.1
abril	539.5
mayo	537.3
febrero	463.2

Promedio de avistamientos por año

year	media
2014	203.214
2012	191.167
2013	184.048
2011	132.405
2008	125.238
2009	117.881
2010	112.500
2004	112.452
2007	112.000
2005	106.905
2003	104.714
2006	98.405
2002	87.095
2001	83.452
1999	74.667
2000	73.071
1998	47.691
1995	34.524
1997	33.714
1996	23.762
1994	11.548
1978	9.191
1993	8.762
1975	8.714
1976	7.786
1974	7.357
1977	7.238
1990	7.167
1992	7.024
1980	6.809
1989	6.809

year	media
1979	6.786
1988	6.548
1991	6.333
1973	6.286
1987	6.238
1985	5.881
1986	5.357
1982	5.071
1984	4.976
1972	4.809
1981	4.714
1983	4.452
1969	4.357
1970	4.143
1971	3.381

Promedio de avistamientos por estado

state	media
CA	160.1857
FL	73.4143
WA	71.6571
TX	62.4143
NY	54.9286
AZ	46.2000
PA	45.0000
IL	43.7143
OH	41.6143
MI	35.0571
NC	32.5571
OR	32.0429
MO	26.8429
CO	26.4571
ON	26.4000
NJ	25.7714
VA	23.9143
IN	23.7857
GA	23.7714
MA	23.6857
WI	22.5571
TN	20.6857
SC	20.1286
MN	18.3571
CT	16.7571
MD	16.0286
NV	15.5571
KY	15.4143
NM	14.2429
OK	13.3857
BC	13.2286
UT	12.3571

state	media
AL	12.2857
IA	11.9286
AR	11.0429
KS	10.9571
LA	10.8000
ME	10.6571
ID	10.0857
NH	9.5714
MT	9.0143
WV	8.5429
MS	7.1714
NE	6.8143
HI	6.3286
AK	6.2714
AB	5.8714
VT	5.4714
RI	5.0143
WY	3.5286
SD	3.4714
DE	3.3143
QC	3.0429
MB	2.6714
NS	2.5000
ND	2.3286
NB	2.1000
DC	1.7857
SK	1.7429
PQ	1.4571
SA	0.5571
PR	0.5143
NF	0.4714
NT	0.3286
PE	0.3286
YT	0.2000
YK	0.1000
VI	0.0143

¿Narrativas parecidas?

Para responder esta pregunta, con la matriz de términos documentos se crearon 10 clúster de narrativas con un largo de 30 palabras. Estos parámetros pueden variar y obviamente las narraciones parecidas no tienen sentido en cuanto a la escrituras, pero serían las 10 narrativas más parecidas de 30 palabras.

cluster 1: light move sky bright saw white red one look craft appear see like two object just flash west disappear east second watch time around seen back minut shape fli south

cluster 2: report jet sight see texa day houston one mile post occur sound air area just light nois perhap regard santa disk fighter heard saw way west wife afterburn alert approx

cluster 3: look light like see sky move plane saw just side front right left shape craft friend object almost back bright night one time red seem get toward two hour know

cluster 4: object sky light bright move look shape appear seen time saw wit like fli west minut notic see one just first white red travel cloud hover second around north two

cluster 5: light move see look one sky red white saw bright like seem craft object back minut appear just seen two north time first south line fli around second plane turn

cluster 6: flash craft south move train appear second away like see degre look right river time back shape ufo white bright color one side sound west drive first light sky turn

cluster 7: object sky light bright observ look saw appear like move second sound time toward east hous jet wit seem aircraft first fli west group seen south speed travel view visibl

cluster 8: light sky move bright like look note wit saw object provid inform contact white seen red total disappear one remain shape elect anonym orang fli appear see second west just

cluster 9: look like light saw sky see star move one object just time seen went bright outsid plane white notic back toward appear seem come shape first someth orang two night

cluster 10: light look see craft night one sky seen hous move star tri back bedroom like room time bed door fli around close walk come notic right seem start tree wave