Module Project - PSY6422

ID: 230170519

2024-05-06

Investigating the Relationship Between Major Airport Locations and UFO Sightings in Mainland USA

Background & Research Questions

The phenomenon of unidentified flying objects (UFOs) has captivated the imagination of people around the world for decades. While I don't particularly admire Donald Trump, his decision to express interest in releasing information regarding UFOs and extraterrestrial intelligence, particularly evidenced by the Pentagon's release of UAP videos in 2020, was intriguing. While many sightings remain unexplained, steps have been taken to allow individuals to document their sightings, including factors such as location, duration, shape, and date. For example, The National UFO Reporting Center NUFORC is a widely recognized organization dedicated to collecting and analyzing reports of UFOs from across the United States and beyond. However, the reliability of these sightings are often called into question due to various factors such as misidentifications, hoaxes, and the subjective nature of eyewitness testimony. As such, I wanted to invesigate whether there were any correlations between the locations of these sightings and the loacations of major airports in the USA.

Research Aims:

Aim 1

To Determine Spatial Patterns: Investigate whether there are spatial correlations between the locations of major airports and reported UFO sightings in mainland USA.

Aim 2

To Assess the Duration of Sightings: Investigate the duration of UFO sightings in regard to the shape recorded.

Data Origins

The raw UFO data was sourced from the NUFORC website, and processed by Sigmond Axel to include geolocated and time standardized reports, which is publicly accessible on github.

The locations of major US airports was acquired from data.world and sourced from OurAirports. This dataset includes all North American airports, and is publicly accessible.

Additionally, the R package maps was used to draw a geographical map of mainland USA, which provides access to a diverse collection of geographical data, particularly maps of the world, countries, and states/provinces of various countries.

Project Organization

The project directory includes two main folders: "/data" and "/images". The "/data" folder contains essential datasets for the project, while the "/images" folder houses both images necessary for the project and visual outputs generated during visualization processes.

Additionally, a comprehensive codebook detailing all labels and abbreviations utilized within the project for data, variables, functions, and other relevant entities is available at the location "/codebook.xlsx".

Data Preparation

Loading Packages

The project used the 'renv' package to preserve package versions of the libraries used. The package versions used in this project are listed in the file /renv.lock

```r
#load packages with renv

if (!require('renv'))
{
  install.packages('renv');
}
library(renv)
renv::restore()

#import packages
library(tidyverse)
library(ggplot2)
```

```r
library(dplyr)
library(tidyr)
library(sf)
library(maps)
```

Importing data

```r
#load UFO data
ufo_data <- read.csv("https://raw.githubusercontent.com/tomhbird/PSY6422_UFO_AIRPORTS/main/Raw/complete

#load US airport data
airport_data <- read.csv("https://raw.githubusercontent.com/tomhbird/PSY6422_UFO_AIRPORTS/main/Raw/list-
```

Cleaning the data

```r
#trimming UFO data to only include mainland USA, removing Hawaii, Alaska, and Puerto Rico
ufo_data_usa <- ufo_data %>%
  filter(country == "us" & !(state %in% c("hi", "ak", "pr")))

#remove columns "comments," "duration..hours.min.," and "date.posted"
ufo_data_usa <- subset(ufo_data_usa,
                        select = -c(comments, duration..hours.min., date.posted))

#detangle date.time column, spitting them into individual variables "date" and "time"
ufo_data_usa <- separate(ufo_data_usa, datetime, into = c("date", "time"), sep = " ")

#checking if latitude and longitude columns are in the correct data type
str(ufo_data_usa)

#as the latitude was a character column, it was converted into a numeric column
ufo_data_usa$latitude <- as.numeric(ufo_data_usa$latitude)

#checking to see if it was successfully converted
str(ufo_data_usa$latitude)

# Convert duration column to numerical
ufo_data_usa$duration..seconds. <- as.numeric(ufo_data_usa$duration..seconds.)

# Verify the change
str(ufo_data_usa$duration..seconds.)

# Replace blank recordings of shape with "unknown"
ufo_data_usa <- ufo_data_usa %>%
  mutate(shape = ifelse(shape == "", "unknown", shape))
```

```r
#display table of first 6 UFO sightings data
knitr::kable(head(ufo_data_usa))
```

First UFO data

date	time	city	state	country	shape	duration..seconds.	latitude	longitude
10/10/1949	20:30	san marcos	tx	us	cylinder	2700	29.88306	-97.94111
10/10/1956	21:00	edna	tx	us	circle	20	28.97833	-96.64583
10/10/1961	19:00	bristol	tn	us	sphere	300	36.59500	-82.18889
10/10/1965	23:45	norwalk	ct	us	disk	1200	41.11750	-73.40833
10/10/1966	20:00	pell city	al	us	disk	180	33.58611	-86.28611
10/10/1966	21:00	live oak	fl	us	disk	120	30.29472	-82.98417

```
# Filter the dataset to include only entries where the type is "large_airport" and remove rows where bo
airport_data_filtered <- airport_data %>%
  filter(type == "large_airport") %>%
  filter(longitude_deg != 0 | latitude_deg != 0)

#checking if latitude and longitude columns are in the correct data type
str(airport_data_filtered$longitude_deg)
str(airport_data_filtered$latitude_deg)

#as the longitude and latitude was a character column, they were converted into a numeric column
airport_data_filtered$latitude_deg <- as.numeric(airport_data_filtered$latitude_deg)
airport_data_filtered$longitude_deg <- as.numeric(airport_data_filtered$longitude_deg)

#checking to see if it was successfully converted
str(airport_data_filtered$longitude_deg)
str(airport_data_filtered$latitude_deg)

#removing unecessary columns from airport dataframe
airport_data_filtered <- airport_data_filtered %>%
  select(-id, -ident, -elevation_ft, -continent,
         -iso_country, -iso_region, -scheduled_service,
         -gps_code, -iata_code, -home_link, -wikipedia_link,
         -keywords, -score, -last_updated)
```

```
#display table of first 6 USA major airport data
knitr::kable(head(airport_data_filtered))
```

Second US airport data

type	name	latitude_deg	longitude_deg	municipality	local_code
large_airport	Los Angeles International Airport	33.9425	-118.4080	Los Angeles	LAX
large_airport	John F Kennedy International Airport	40.6398	-73.7789	New York	JFK
large_airport	Chicago O'Hare International Airport	41.9786	-87.9048	Chicago	ORD
large_airport	Hartsfield Jackson Atlanta International Airport	33.6367	-84.4281	Atlanta	ATL

type	name	latitude_deg	longitude_deg	municipality	local_code
large_airport	San Francisco International Airport	37.6190	-122.3750	San Francisco	SFO
large_airport	Newark Liberty International Airport	40.6925	-74.1687	Newark	EWR

```r
# Plot map of USA
usa <- map("state", fill = TRUE, col = "transparent", plot = FALSE)

# Convert map data to data frame
usa_df <- fortify(usa)
```

Third creating map of the USA

Data Visualisation 1

UFO sightings plotted on a map of the USA in relation to major airport locations

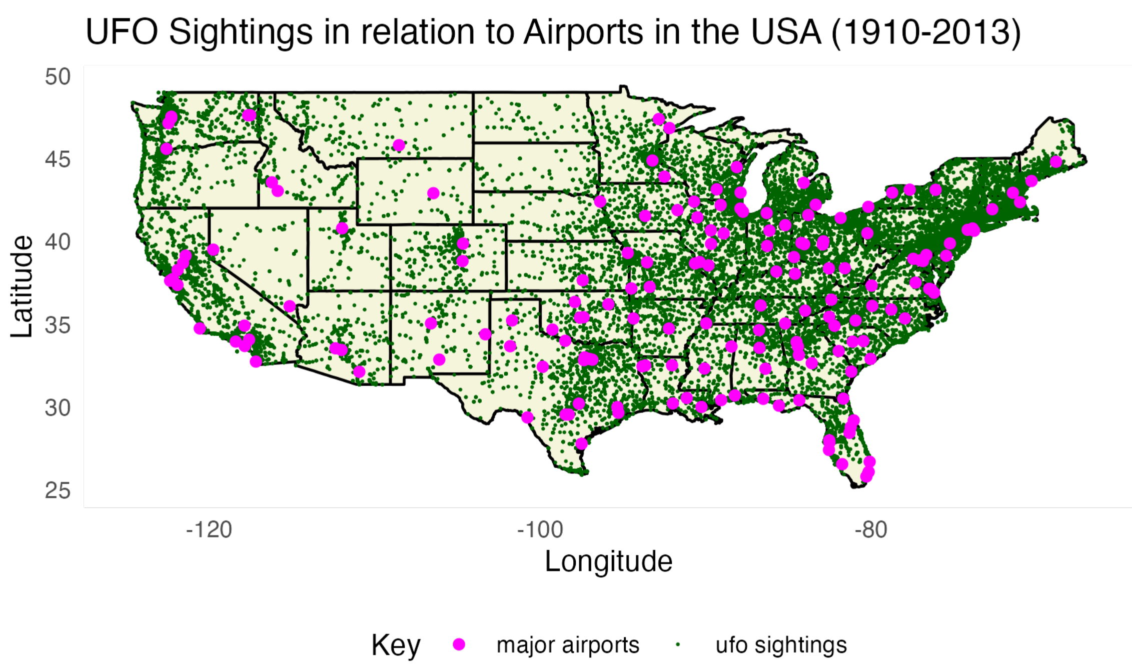
```r
# define data to plot
ufo_map <- ggplot() +

  # Add the USA map polygons
  geom_polygon(data = usa_df, aes(x = long, y = lat, group = group), fill = "beige", color = "black") +

  # Add UFO sightings as points (darkgreen color)
  geom_point(data = ufo_data_usa, aes(x = longitude, y = latitude, fill = "ufo sightings"), color = "da

  # Add airport locations as points (magenta color)
  geom_point(data = airport_data_filtered, aes(x = longitude_deg, y = latitude_deg, fill = "major airpo

  # Set the aspect ratio and limit the plot to the boundaries of the USA map
  coord_fixed(xlim = range(usa_df$long), ylim = range(usa_df$lat)) +

  # Add plot title
  labs(title = "UFO Sightings in relation to Airports in the USA (1910-2013)", x = "Longitude", y = "La
       color = "Data",
       fill = "Key") + # Change legend title

  # Set minimal theme with white background and remove grid lines
  theme_minimal() +
  theme(panel.background = element_rect(fill = "white", color = NA),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank()) +

  # Customize x and y axes
  scale_x_continuous("Longitude", limits = range(usa_df$long)) +
  scale_y_continuous("Latitude", limits = range(usa_df$lat)) +

  # Position the legend at the bottom of the plot
```

```
  theme(legend.position = "bottom")

# Save the graph as an image
ggsave("images/ufo_map.png", plot = ufo_map, width = 6, height = 4, dpi = 300)
```

Aim 1:

To Determine Spatial Patterns: Investigate whether there are spatial correlations between the locations of major airports and reported UFO sightings in mainland USA.

Result



Interpretation

This visualization illustrates the distribution of UFO sightings across the USA in proximity to major airport locations. The concentration of sightings around airports implies a potential correlation between reported "UFO" sightings and the presence of airports. Therefore, this suggests that individuals may mistake aircraft for UFOs and misinterpret their sightings.

Data Visualisation 2

Creating stacked bar graph to show the most common shapes of UFO and their durations in mainland USA

```
#catagorising the top 20 most reported ufo shapes

# Count the frequency of each UFO shape
ufo_shape_counts <- ufo_data_usa %>%
  count(shape) %>%
  arrange(desc(n))

# Select top 20 shapes
ufo_shape_top20 <- head(ufo_shape_counts, 20)
```

```
# Calculate quantiles of duration
quantiles <- quantile(ufo_data_usa$duration..seconds., probs = c(0, 0.25, 0.5, 0.75, 1), na.rm = TRUE)

#check quantiles
head(quantiles)
```

```
##       0%      25%      50%      75%     100%
##        0       15      120      600 66276000
```

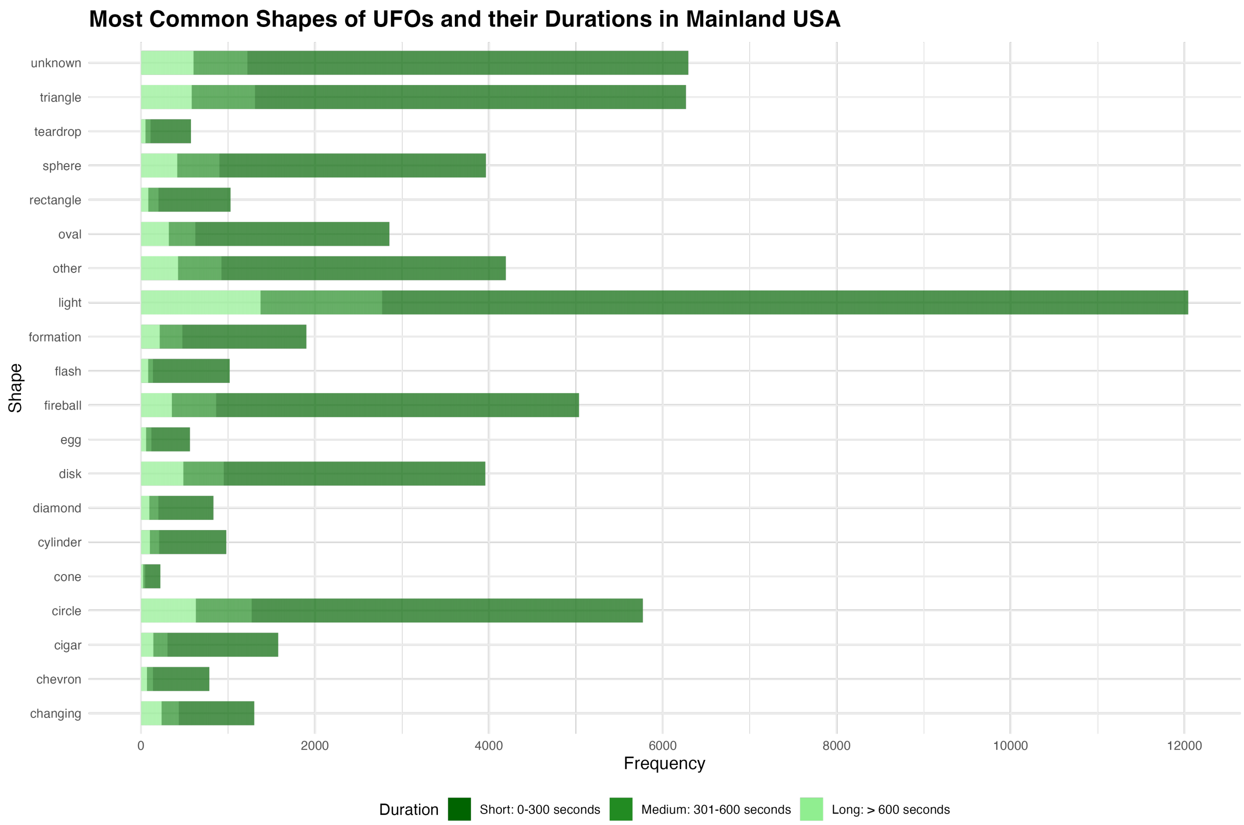
```
#as there is extreme outliers calculate
# Define upper and lower bounds
Q1 <- quantiles["25%"]
Q3 <- quantiles["75%"]
IQR <- Q3 - Q1
lower_bound <- Q1 - 1.5 * IQR
upper_bound <- Q3 + 1.5 * IQR

# Filter data to remove outliers
ufo_data_usa_filtered <- ufo_data_usa[ufo_data_usa$duration..seconds. >= lower_bound & ufo_data_usa$dura

# Create a new variable for duration categories
ufo_data_usa_filtered$duration_category <- cut(ufo_data_usa_filtered$duration..seconds., breaks = c(-In

# Combine both data frames
combined_data <- merge(ufo_shape_top20, ufo_data_usa_filtered, by = "shape")

# Set count to 1 for every data point
combined_data <- combined_data %>%
  mutate(count = 1)

# Create a stacked bar plot to visualize the frequency of top 20 UFO shapes and their durations
# Define data to plot
stacked_barchart <- ggplot(combined_data, aes(x = reorder(shape, count), y = count, fill = duration_cate

  # Add stacked bars representing the frequency of each UFO shape and duration category
  geom_bar(stat = "identity", position = "stack", width = 0.7) +
```

```r
  # Add plot title and axis labels
  labs(title = "Most Common Shapes of UFOs and their Durations in Mainland USA",
       x = "Shape",
       y = "Frequency",
       fill = "Duration") +

  # Manually set fill colors for each duration category
  scale_fill_manual(values = c("Short: 0-300 seconds" = "darkgreen", "Medium: 301-600 seconds" = "fores

  # Flip the x and y coordinates to create a horizontal bar plot
  coord_flip() +

  # Set breaks for the y-axis to ensure proper scale
  scale_y_continuous(breaks = c(0, 2000, 4000, 6000, 8000, 10000, 12000, 14000)) +

  # Apply a minimal theme with white background and bottom legend
  theme_minimal() +
  theme(axis.title = element_text(size = 12),       # Customize axis title font size
        plot.title = element_text(size = 16, face = "bold", vjust = 1.5),  # Customize plot title, font
        legend.position = "bottom")                # Position legend at the bottom of the plot

# Save the graph as an image
ggsave("images/stacked_barchart.png", plot = stacked_barchart, width = 12, height = 8, dpi = 350)
```

Aim 1:

To Assess the Duration of Sightings: Investigate the duration of UFO sightings in regard to the shape recorded.

Result



Interpretation

This visualization highlights a notable trend: the majority of reported sightings describe objects with a "light" shape. Additionally, there are frequent mentions of other light-related shapes such as "flash" and "fireball." This suggests that aircraft may account for these sightings, given their characteristic appearance, especially at night time where they often have blinking lights. This further emphasises the influence of airport locations on recorded UFO sightings, as not only does visualisation 1 show that reports tend to congregate around airports, visualisation 2 shows that light related terms account for the majority of shapes recorded.

Furthermore, the duration of these sightings is particularly relevant. As the majority of sightings tend to fall into the short (0 - 300 second) variable, this suggests individuals may just be observing airplanes flying overhead and are mistaking it for a UFO.

Conclusions

Overall, the visualizations presented shed light on the intriguing relationship between reported UFO sightings, major airport locations, and the characteristics of observed objects. The concentration of sightings around airports suggests a potential correlation, raising the possibility that individuals may misinterpret aircraft as UFOs. Moreover, the prevalence of "light" related shapes in sightings, coupled with the typical duration of these events, reinforces this notion. It appears that many reported sightings align with typical aircraft activity, especially considering the prominence of blinking lights at night. These findings underscore the importance of considering environmental factors, such as airport proximity, when analyzing UFO reports. Ultimately, while the mystery of UFOs continues to captivate public interest, a critical examination of the data suggests that many sightings may have terrestrial explanations rooted in human perception and environmental context.

Limitations & Future Directions

A limitation of this project was that I only focused on mainland USA, it would be interesting to see if this trend persists on a global scale. Especially, in countries that may not have as many large airports. Additionally, this project only looked at the relationship between UFO sightings and major US airports, it could add value to investigate smaller airports as well, and see whether UFO sighting reports also pool around them, even if its on a smaller scale. Furthermore, the project did not investigate the time of day of these sightings. Future research could investigate the correlation between the time of day of sightings and airport locations, to see whether factors such as nighttime may affect the number of sightings reported.

A further limitation comes from the data used. The UFO data set ranged from 1910-2013, which may have had some confounding effects on the data presentation, considering many of the airports investigated were build after 1910. As such, it may have been more beneficial to use a more updated data set that with more recent sightings included and older sightings removed. Despite this, the project still highlights a relationship between airport locations and UFO sightings, suggesting these “unidentified flying objects” may just be identifiable flying objects - planes!

References

The citations provided in the “Data Origins” section of this project are not permanent references as the websites utilized lack a DOI for inclusion in this project. Consequently, these web pages are subject to change or removal, potentially affecting the reproducibility of certain project components.