

Basic Information

- **Project Title:** Examining the Consistency of UFO Sighting Reports
- **Group Members:**
 - Asa Adomatis, asaa@clermson.edu, C32722636
 - Ellie Painter, rapaint@clermson.edu, C24975483
 - Emily Wilber, emwilbe@clermson.edu, C18656247
- **Repo:** <https://github.com/AsaAdomatis/CPSC-4030-Project>

Overview and Motivation

- **Motivation:** Personal interest in UFO conspiracy and recent news events about UFOs/UAPs like the military released videos and the Mexican aliens.
- **Reasoning:** See if there's any patterns or consistencies within sightings to analyze how valid sightings are.
- **Primary Goal:** Are there consistent patterns in UFO sightings Reports?
- **Secondary Goals:**
 - Is there a consistent location or set of locations UFO sightings are likely to happen at?
 - Is there a consistent time where UFO sightings are likely to happen at?
 - Do similar sightings have a consistent description of the encounter?

Related Work

- **Inspirations:** A lot of inspiration for designs came from other students demoing the work. It gave us ideas and alternatives to the designs we were working on. For example, after seeing the group that analyzed score differential effects in NFL teams, we considered using a heatmap to visualize a combination of durations and shapes.

Questions

- Our original question was just our goal, "Are the consistent patterns in UFO sightings reports?" The spirit of the question was to find out if we could identify something like "tear drops shapes always appear in the South and in a cycle of five years."
- As the project went on we realized a consistent distribution with a lot of the shapes and states, so the question became "Are there any outliers to the distribution?"
- We're still wondering what causes the spike in sightings in Arizona. Other states have distributions mostly consistent to the population, but Arizona in particular, seems to have more sightings than its population.

Data

- **NUFORC Data:**
 - Link: <https://www.kaggle.com/datasets/NUFORC/ufo-sightings/versions/1/data>
- **Shape Files for US Counties and States:**
 - Link: <https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html>

Data Processing

- **Data Clean-Up:**
 - The alternate dataset (2022) needs duration data to be converted from an unformatted string to a discrete number.
- **Derived Quantities:**
 - A county attribute that's derived from long. lat. or coordinates
 - A more generalized shape attribute to group things like oval, and circle together
- **Data Processing Implementation:**
 - We will use Python to convert the lat. long. data into county data

Exploratory Data Analysis

- Other than sightings by month, we included some version of our initial designs into our final dashboard. We excluded sighting by month because the data was consistent across all months.
 - We used a stacked bar chart with the months as entities and number of sightings mapping to the height channel and year mapping to a color channel.
- Initial analysis revealed that the trends without filters maintained themselves with filters but proportionally smaller. Because of this, We wanted more visualization so comparison amongst the individual trends were easier to spot.
 - For example, the "illuminated" shape is always the most sighted and the most sightings are consistently in California.

Design Evolution

- **Sightings by location:**
 - Our final dashboard included 2 choropleths for sightings by area: one for counties and one for states. We went with a choropleth rather than a field with a color channel representing density of sightings because it made for better lookup. We chose to keep both county and state graphs because while the county works better for looking up individual sightings in a specific area, the state can be used to filter other data and still provide a comparison between large areas. In sum, we kept county for lookup and state for summary.
- **Sightings by time:**
 - **By duration:** The initial design for duration was a histogram with each bar being sightings in the duration bracket. However, the dataset was too diverse to work linearly so we switched it to logarithmic scale. Even then, it was hard to parse information from the graph because there were random spikes along the x-axis. We ended up on a boxplot with a logarithmic scale because it showed the wide variation in data but was still legible to understand the general trend (like average).
 - **By month:** Initially we were going to include a stacked bar chart of sightings by month with each stack being a year. The year would use a color channel while the month would use height as a channel. It was scrapped in the end because it was harder to parse than a simple line chart, and the data was consistent across months.
 - **By year:** We went through two different iterations of sightings by year. The first would've have included individual lines for each sighting's shape throughout the years mapped to a color channel. However, we thought it would be easier to understand if we separated shape and year

and instead let you filter and see transitions throughout the years. A separate design would also let you examine the spread of individual shapes in years and vice versa more clearly.

- **Sightings by shape:**

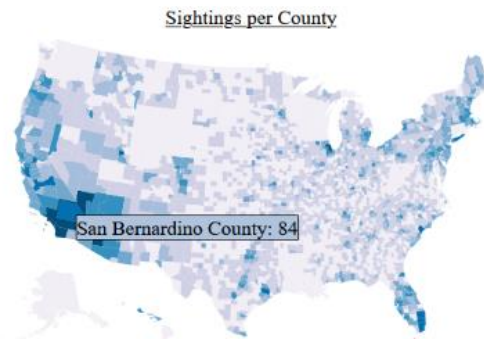
- In many early iterations we considered moving the sightings by shape into other graphs like mentioned with sightings by year. We decided to make it owns graph, a simple bar chart, so users con could compare sightings by shape on an individual level easier

- **Final Philosophy:**

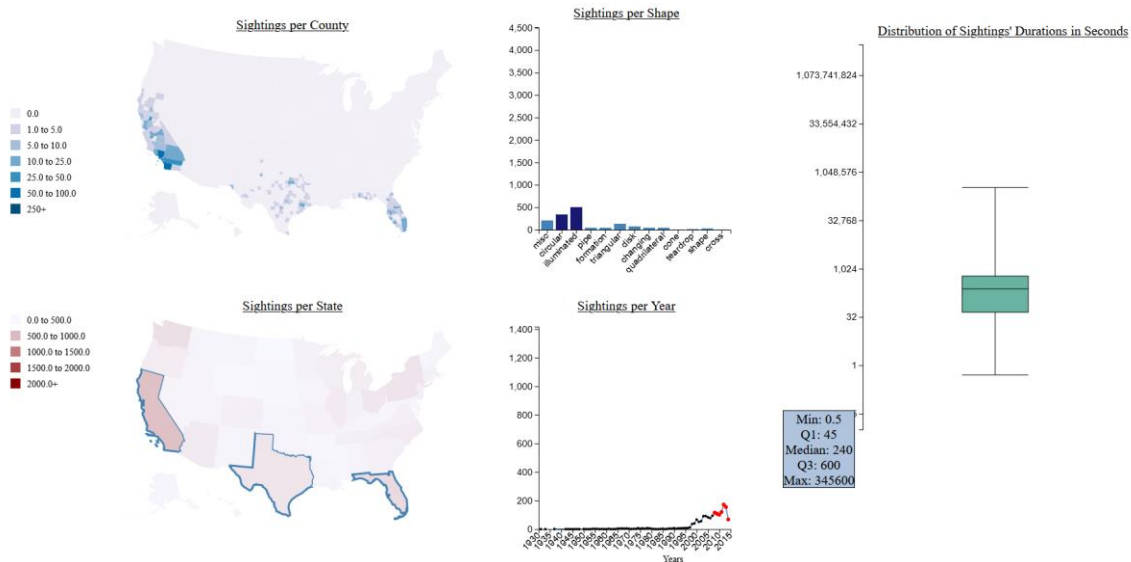
-

Implementation and Interaction

- Hovering for details:



- Filtering by states, shapes, and years in any combination with 1 or more for each



- **Intent:** Allow users to compare information about sightings by state, shape, and year in any combination. By allowing users to filter by almost in any parameter in the use the visualizations to narrow down consistent events. For example, the user could investigate all the sightings of the “illuminated” shape, then see that there’s a spike in 2012 and 2013, then could see that most of the sighting are in Arizona and Southern California.

Evaluation

- The data does not reveal any consistency in the appearances and sightings of UFOs. The largest category of UFO was illumination, this combined with some sightings being over the duration of several months leads us to believe that stars and other astronomical phenomena could be a large portion of sightings. Beyond this the data mostly shows that areas with large numbers of sightings are usually just population centers, with only Arizona showing a higher number of sightings than expected for its population.