



UFOs? STA199 Final Presentation

Jordan, Elysia, Abigail, and Gus

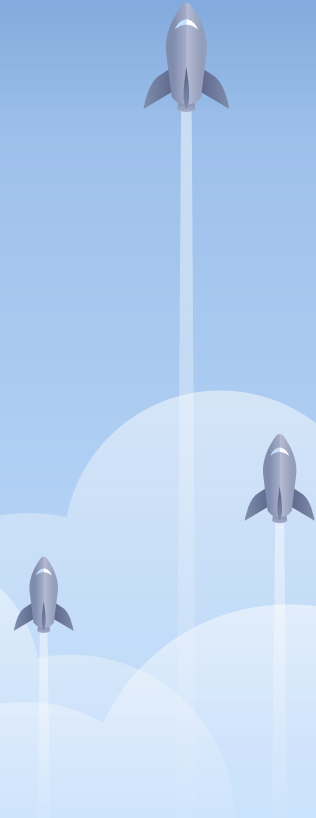
Topic and Motivation

Analysis of UFO Data

- UFO sightings around the world
- What can we learn from data collected worldwide?
- Looking at categorical and quantitative variables

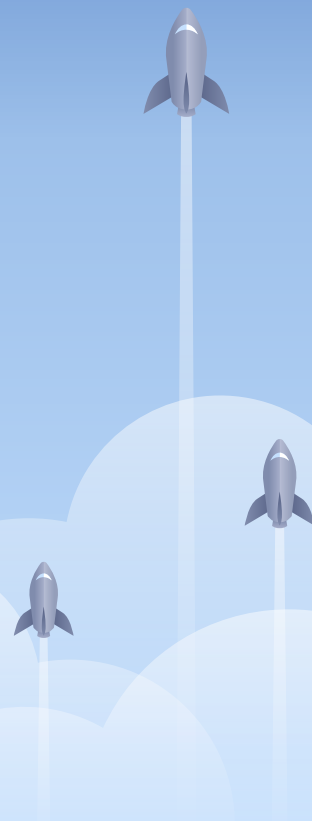
Motivation

- Unique and interesting topic
- Recent Pentagon information
- Space-based current events



THE DATA

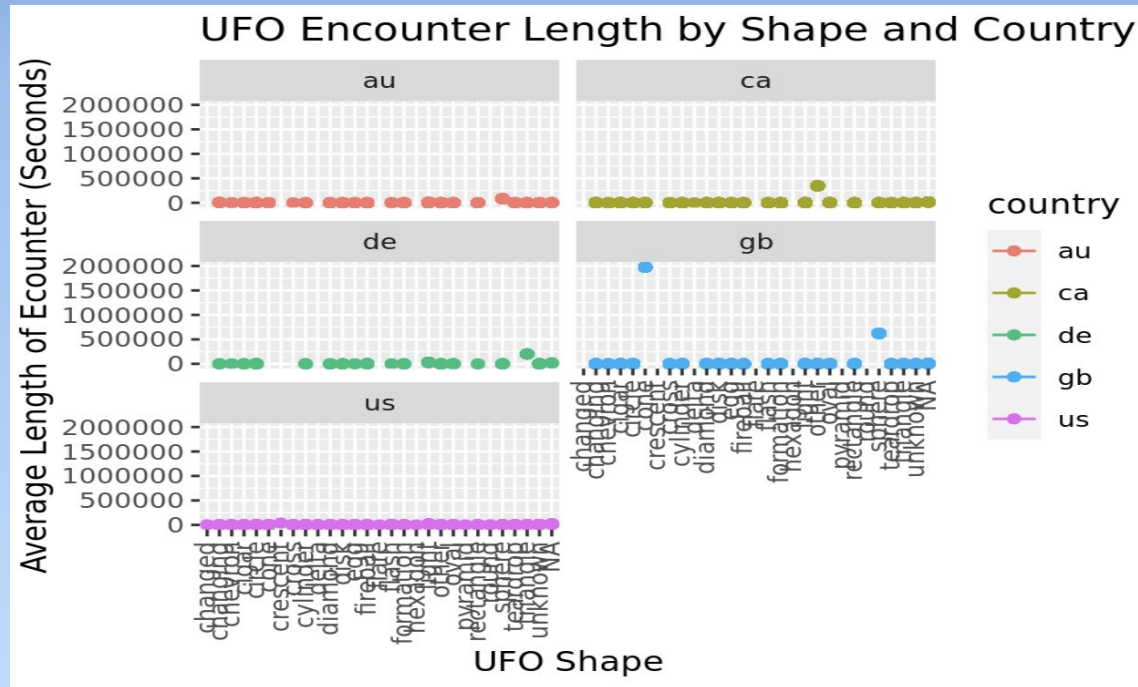
- Data upload inspired by photos and videos of UFOs released by the Pentagon in 2020 and a follow-up article about it from NYT
- 80,000+ observations; 11 variables
 - ▷ state, country, length of the encounter, & qualitative descriptions
 - Ex: “Triangle shaped UFO acting like a lost puppy”, “There were lights on and it was weird”, “Satellite?”
 - data records described as “frightening” and “very disturbing”



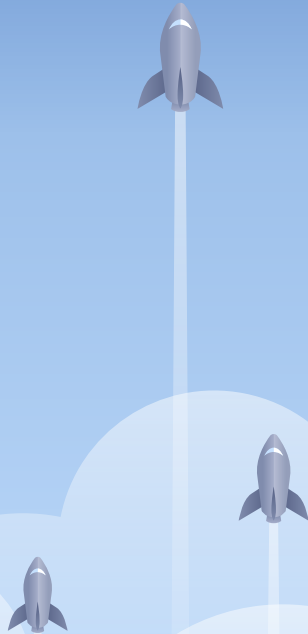
Is the average length of a UFO encounter greater in the decades after 1970 when separated by country?



Highlights from the EDA

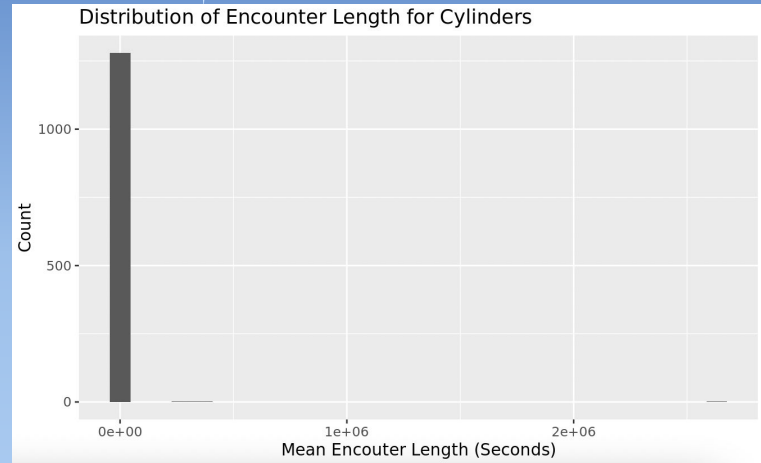
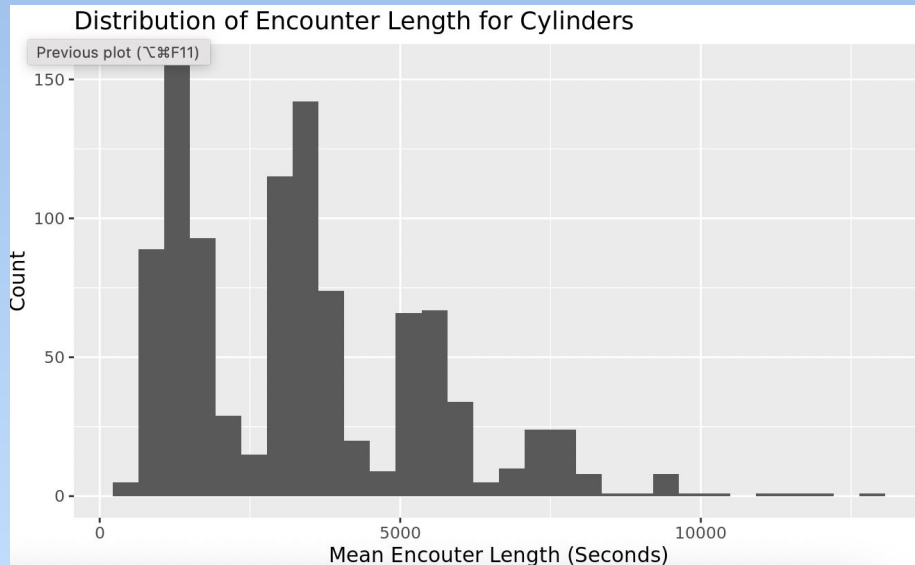


—



PHASE 2: CONSUMER

Visualizations + Ongoing Methodology



	mean_time
1	3426.512

Cylinder



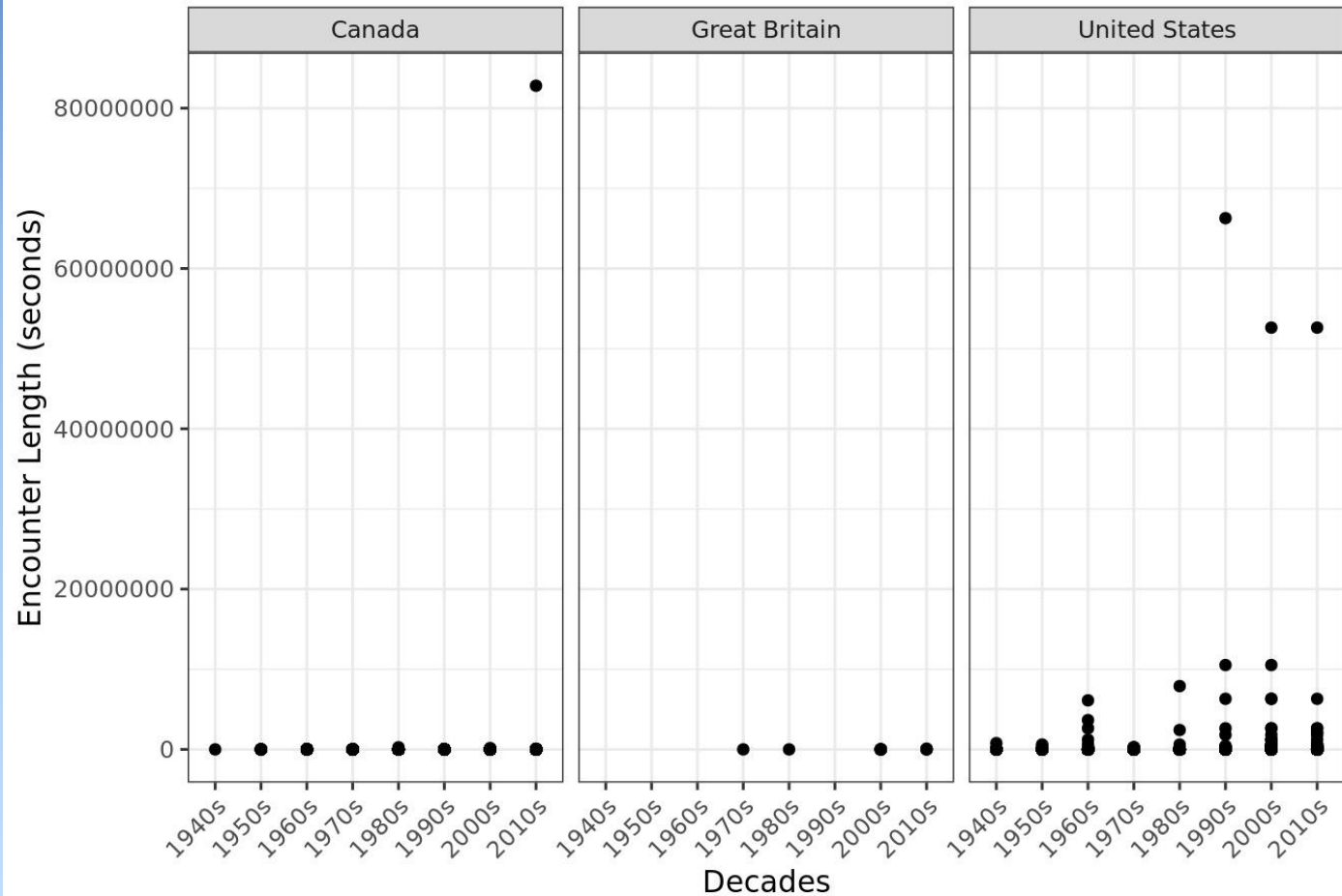
	mean_time
1	4768.093

Circle



The figure consists of three panels, each representing a different country: Canada, Great Britain, and the United States. The y-axis for all panels is 'Encounter Length (seconds)' ranging from 0 to 80,000,000. The x-axis for all panels is 'Decades' with labels from 1940s to 2010s. The data points are black dots.

- Canada:** Shows a single data point in the 2010s at approximately 83,000,000 seconds.
- Great Britain:** Shows data points in the 1970s, 1980s, 2000s, and 2010s, all with encounter lengths near 0 seconds.
- United States:** Shows multiple data points across several decades. Notable points include a cluster in the 1960s (up to ~7,000,000s), a point in the 1980s (~8,000,000s), a peak in the 1990s (~67,000,000s), and points in the 2000s and 2010s (up to ~11,000,000s).



Inference/modeling/other analysis

- Based on the means, there is not a clear difference in length of encounter based on country and shape
- The model highlights the difficulty in discerning particular time, especially with the outliers incorporated

Discussion

- There are never (or rarely) definites in statistics and that's true with our results
- Eliminating outliers would make our models more accurate
- Hypothesis testing worked well because we examined two groups
- Linear regression could work better in future if we wanted to assume time as our x-variable
- Encounter length was not a precise measurement because it was based of estimates (e.g. 1-2 hours)
- We could get more specific about the regions we examine in the future to consider how political affiliation relates to UFO sightings

Conclusions + Future Work

- Calculate the medians, discern diff. for overall trends

More Questions than Answers...

- How do UFOs actually impact lives?
- How do we highlight the narratives surrounding the project? How do these correlate to paranoia and current events?
- Can we collect more specific data?
- Are UFOs even real in the scope of this project?