An Exploration of Extraterrestrial Communication through the Space Industry and Modern Media

Hermione Bossolina, Hrishika Jotwani, Hung Ming Tseng, and Jillian Beck

Table of Contents

Data Description

UFO data - It is a kaggle dataset (https://www.kaggle.com/datasets/rishidamarla/ufo-sightings-approx-100000) of nearly 90,000 reports of UFO sightings with information on where the sightings were, the time, the shape, duration, etc. The dataset comes from https://nuforc.org/databank/, a NUFORC Databank which is the "largest independently collected set of UFO/UAP sighting reports available on the internet" (NUFORC). The data here has data from 1561 (2 sightings), 1864 (2 sightings), 1897 (2 sightings), the number of sightings significantly increased starting 1965. In the dataset from Kaggle it began in 1969 and ended in 2019, it did the initial aggregation of the NUFORC Databank of months to years for us.

Within the UFO Dataset we have a summary and text of the individual description of their sighted sighting, the city and state of where the sighting occurred, the date and time, shape of the UFO, the duration in which the UFO was seen, report_link, and city_latitude and city_longitude of the sightings location.

Space Mission Dataset - Dataset is also from Kaggle (https://www.kaggle.com/datasets/agirlcoding/all-space-missions-from-1957) and it references https://nextspaceflight.com/launches/past/?page=1 while this dataset looks at all space races from 1957 to 2020 due to the UFO dataset being only from 1969 to 2019 we've changed the Space Mission Dataset to be 1969 to 2019 as well. The dataset includes the company of the rocket that launched name, the location of the site where the launch took place, details of the rocket (ex. Falcon Heavy | STP-2), the status of the rocket, how much the Rocket cost, and the status of the mission.

Nasa Space Exploration and Satellites - Dataset is from Kaggle (https://www.kaggle.com/datasets/whenamancodes/space-exploration-and-satellites) and it was used by us to understand the budget of NASA one of the largest space companies in the world. The dataset included three columns (the country, year, and budget). The dataset started in 1959 and the last year that the budget has been recorded was 2022.

Cleaning the Dataset:

- **Combining UFO data and Space Mission Dataset:**

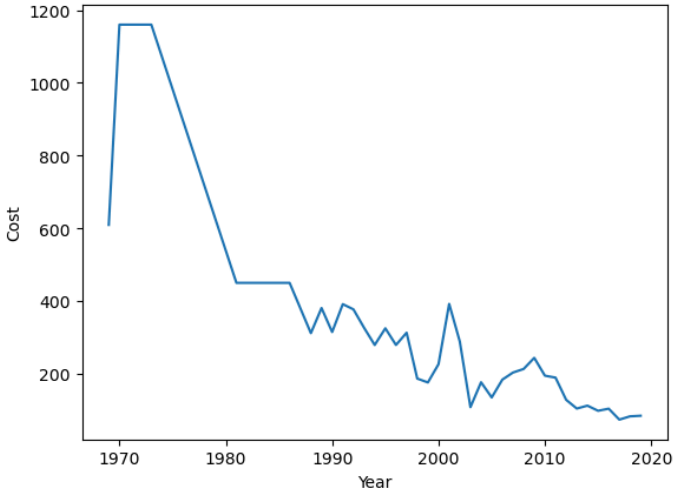
To clean the dataset we began by looking at time frame and time. Notice in Space Missions and UFO have different datum representation. For the UFO dataset they have date and time in the format of Year-Month-DateT[hour:minute:second], while the Space Missions dataset has Day of Week Month Day, Year Hour:Minute UTC. To better work with the data and aggregate it we decided to only look at years rather than looking more zoomed in like month or day. We also as stated before 1969 to 2019 was the years that we use for both Space Mission, Movie Data, and UFO dataset.

Along with this we noticed that the UFO Dataset comes with location, specifically longitude and latitude, but some of these data points are unfortunately missing. More than one thousand are missing, so we decided to remove those datapoints where longitude and latitude

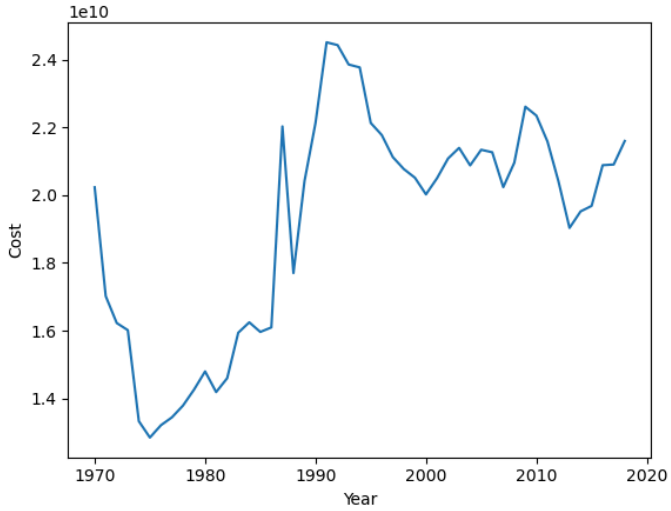
were missing. This cut the UFO Dataset drastically, but there remained more than 6.8K datapoints so we felt that this decision would not affect the data so much. Now Space Mission did not come with longitude and latitude, but it did state the Space Center and State and there were less than 10 so we found each of the space stations' coordinates and entered them using a dictionary! This is how the map visualization was formed!

We decided to use the NASA budget dataset because even though this dataset had 442 space mission's price tags after looking at the data we decided that the data looked incorrect compared to previous trends. Something especially noticeable in the space mission graph was the increase after the 1970s, which seems odd given that we had just landed on the moon, which was the priciest mission.

Using Space Mission Dataset: (y-axis in thousands)



Looking at NASA's budget:



For the NASA dataset the only datacleaning that was done was to make it fit to the same time constraint as the other two datasets: 1969 to 2019.

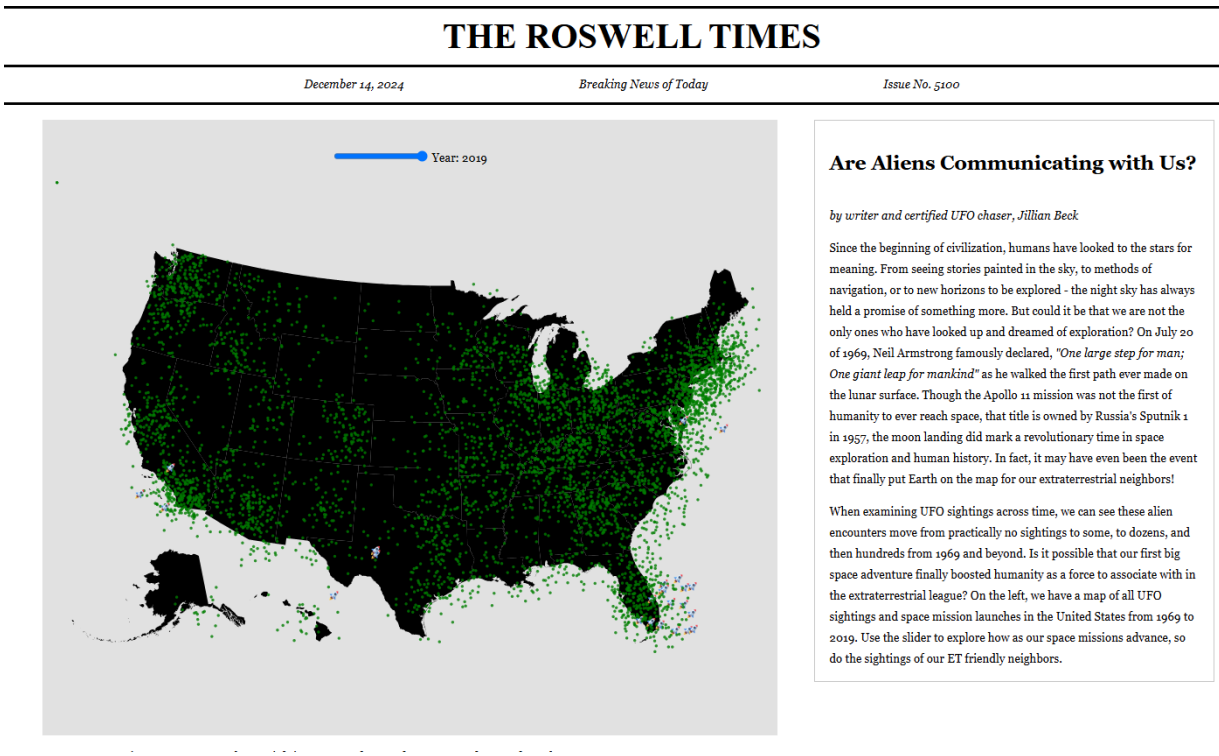
After cleaning the UFO Dataset is 68721 rows while the Space Mission dataset is 981 rows.

Geo data:
We decided that one of our first visualizations should be a map of the United States to show the spread of UFO data and Space launches. To create this map, we used a geojson dataset sourced from here: https://eric.clst.org/tech/usgeojson/

Best Sci-fi Movies of all Time Dataset
This dataset describes the movie year, title, url, and the score. We use the average of audience score and critic score to determine the significance of the movie.
https://www.kaggle.com/datasets/bwandowando/rotten-tomatoes-best-scifimovies-of-all-time?resource=download

Movie Poster
This consists of the Alien (1979) movie poster in jpg format.
https://www.imdb.com/title/tt0078748/

Map Visualization



Data Handling:

Map Outline and Zoom Feature:
We started making this map by establishing the outlines of each state with our geo data. We used asynchronous and await functions to load in each of the data sets for this visualization. For the map outline, we established a projection with geoAlbersUsa() to create the paths for each state outline. To add a layer of dynamic interactivity, we also initiated a zoom and drag functionality using ".on" for the svg with scales of ".scaleExtent" to set zoom limits and ".translateExtent" to establish the drag limits. This functionality allows users to zoom in on a state when they click on it and allows them to drag the map around.

With the map established, we loaded in the UFO and Space mission data. For the UFO data, we appended circles to mark individual sightings using .data and .join, and we utilized the "city_longitude" and "city_latitude" columns to map the points on the correct map location. We followed the same process for the space mission launch data and we used the "Longitude" and "Latitude" columns of that data to append those points to the correct map locations.

Slider Feature:
After displaying the data points on the map, we realized that due to the large amount of overlapping instances, it would be beneficial to allow users to visualize data across years. With

this in mind, we created a slider to allow users to see the data for each year from 1969 to 2019. Although the UFO and Space mission dataset technically begin in the year 1957, there wasn't enough information on the visualization to justify including those years so we started with 1969. We created a function called UpdateMap() that we linked to the inputs of the slider, and we created filtered variables for both datasets to ensure that only data for the respective year is being shown. Since some of the years (such as 2014) still have copious amounts of overlapping data points, we also implemented a jitter function for the rocket and ufo data so that users can better see the points.

MouseOver:
The final interactive element of our map visualization is a series of hover features on the data points. While markings on a map alone are informative, we wanted to ensure users would be able to gather more information from this visualization than simple location distribution. With this in mind, we implemented mouseover features for the UFO data and Space Launch data. We increased the radius and width of the respective points when users hover over them to offer more clarity to viewers. For the UFO points, we also took the information from the "City", "State", and "Summary" columns to offer users more context on each data entry. The "Summary" column includes a description of the UFO encounter by the witness. For the space mission data, we included information from the "Company Name", "Detail", and "Location" columns to give launch context to viewers. The "Detail" column gives information about the mission and the type of rocket launched.

Design Rationale:

After attending office hours with the professor and discussing potential design styling with him, we decided to create our visualizations in a Newspaper like layout to go with the theme of UFO sightings and alien interactions. We intentionally placed the visualization on the upper left corner of the page, with the text description on the right to mimic standard news convention layouts. Instead of placing a description of the visualization within the SVG, we put a figure disclaimer below the map to mimic the use of reference figures in news articles.

The main goal of this visualization is to demonstrate how over time there may have been a correlation between the number of space missions and ufo sightings showing that maybe the aliens are communicating back to us in a way that we are starting to perceive and understand better. Also this map shows the location on where the ufo sighting and space missions are demonstrating whether there is a correlation between the location of the ufo sighting and where a mission is launched, look for yourself whether there are more ufos in these areas or fewer! In the following visualizations we will cover what forms ufos are starting to present themselves to us, how much do we spend in communicating to aliens and at what cost, and what movies are historically viewed as influential in the alien or extraterrestrial space and whether that has any correlation to ufo's responding to us.

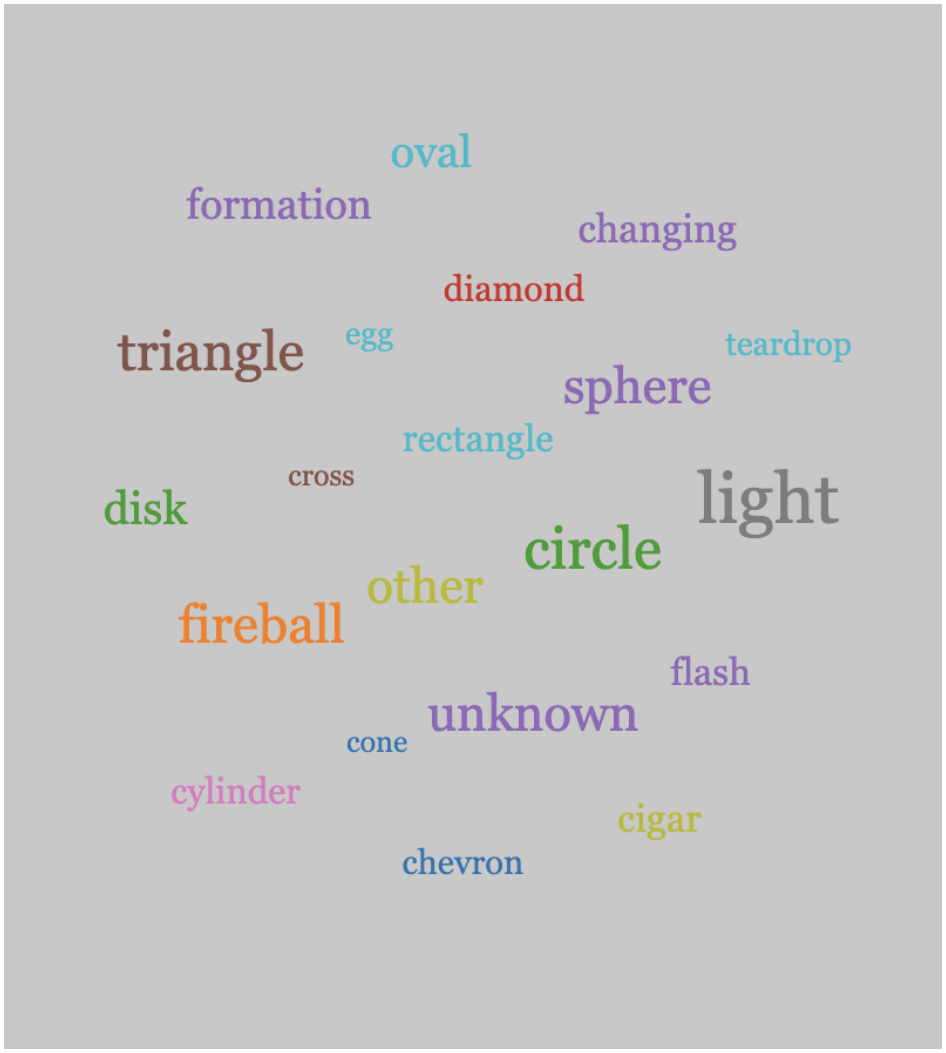
Map Outline and Zoom Feature:

Going with the Newspaper effect, we made the color of the map black with a grey div background against a white overall background for the site. The marks employed in our visualization are circles, which represent instances of UFO sightings, and rocket icons, which represent instances of space launches. The visual channels employed in our visualization are colors and iconography. We made the UFO data points green to symbolize the classic alien assumed color. For the space launches, we knew we needed to differentiate the data from the UFO sightings so we chose to use rocket icons from flaticon.com to establish these positions. The zoom feature allows users to better see the spread of data points within each state, and the drag functionality allows a more intuitive navigation for site visitors as they can move from state to state whilst remaining within a zoomed view.

Slider Feature:
While some may argue that the slider color should have been made black to go with the newspaper theme rather than blue, we decided to keep the default blue color. Our reasoning was that this blue is the standard format across sites so by using a normalized design of this tool, we ensure that users will be inherently familiar with how to use it already. By customizing the slider with any sort of themed motifs might have cluttered the design and made it less familiar for site visitors.

Mousover:
For our data points on the map, we kept a somewhat transparent opacity of 0.8, but when hovering over these points, we amplified the opacity to be fully opaque. This helps users differentiate further between points they're hovering over and others on the map. When users hover over points, we also created a standard pop-up box outline to appear by the points in order to contain the data from the columns discussed earlier. We kept the background of this box white, but gave it a black outline to continue with our newspaper color scheme.

UFO Shapes Word Cloud Visualization



Data Handling:

We used the UFO shapes dataset for this Visualisation. Since the word cloud visualises the popularity of certain shapes, the size of each word in the word cloud had to be proportionate to the number of occurrences and for that, the actual data used for the word cloud was filtered from the original dataset that consists of various shapes and their corresponding counts. To prepare the data for visualization, the following steps were performed:

1. **Data Transformation:** The raw data is first processed into an array of objects, each containing a shape name (text), frequency (size), and random positions (x and y) to allow for a dynamic, non-overlapping layout.
2. **Scaling:** A square root scale (fontSizeScale) is applied to map the count values to font sizes. This scale ensures that shapes with higher frequencies appear larger, while smaller counts are represented with smaller font sizes. The d3.scaleSqrt() function is

used to map the range of counts to a more visually meaningful font size range of 10 to 30 pixels.

3. **Force Simulation:** To prevent overlap and ensure that words are spaced appropriately, a force simulation (d3.forceSimulation()) is applied. The simulation uses forces such as d3.forceX() and d3.forceY() to center the shapes within the SVG container, while d3.forceCollide() prevents the shapes from overlapping by adjusting their spacing based on their font size.

Although, initially we planned on using a third-party library for word cloud - “d3-cloud/build/d3.layout.cloud.min.js", during Office Hours Prof Jeff mentioned that using forceCollide would be more technical.

Design Rationale:

The design of the word cloud aims to present the frequency of different shapes visually in an engaging and informative way. Instead of doing a classic bar chart visualisation to display the frequency of the different shapes, we felt that a word cloud would be a more interesting approach. The goal of the visualization was to demonstrate how aliens have communicated with us and showcase how we perceive what aliens are. It is meant to make you question whether or not you yourself have ever seen an alien and also wonder, in what form would a ufo showcase itself to you.

Earlier, the words were rotated in random angles that were multiples of 35 degrees to give a more classic word cloud effect, but according to us horizontal alignment looked cleaner.
The word cloud uses a random distribution of words within the SVG, which is then refined through a force simulation. This approach ensures that the words are positioned in a way that is visually appealing, avoiding overlaps while maintaining a sense of organic randomness.
The use of d3.scaleSqrt() to adjust font size based on the frequency count ensures that larger counts are represented with more prominence in the visualization. This is crucial for highlighting the most frequent shapes while still keeping the overall word cloud readable and balanced.
A random color selection from the d3.schemeCategory10 palette is applied to each shape in the word cloud, providing a visually diverse and vibrant design. This enhances the readability of the visualization by distinguishing each shape with a unique color.
On hover, the tooltip displays the exact frequency count for each shape, providing additional context. This allows users to explore the data interactively, offering both an aesthetic experience and detailed insight into the dataset.

NASA Space Budget across the years Line Plot

Data Handling:

For this line plot, the data used represents NASA's annual space budget from 1969 to 2019. The data is sourced from a CSV file (nasa_1969_to_2019.csv), which includes the columns Year and Budget. Each data point in the dataset represents the budget allocated to NASA for that specific year.

The following steps are taken to prepare the data:

1. Sorting: The data is sorted chronologically by the Year field, ensuring that the line plot displays a continuous and accurate representation of the trend over time.
2. Scale Definitions:
    - A linear scale for the x-axis (xScale) is created, with the domain spanning from 1969 to 2019, mapped to the width of the plot.
    - A linear scale for the y-axis (yScale) is defined based on the maximum budget value in the dataset, allowing the y-axis to adjust dynamically depending on the range of values. The numbers in the data were updated

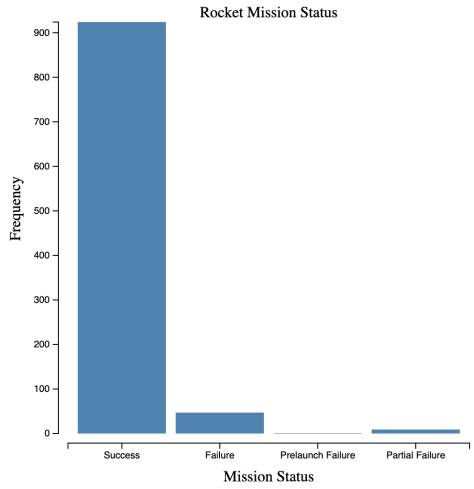
Design Rationale:

In an effort to explore the cost of the missions to communicate to the UFOs we plotted a line chart over the time span to disseminate the cost of our journeys specifically shining light on NASA the landers of the moon's journey. Though the line chart does not emphasize the budget cuts from the 1950s (not pictured) as it was out of scope it does showcase the grueling journey of funding that NASA has gone through over the years.

In this chart we added interactions of hovering allowing you to see the price and years of the NASA space budget. By adding this affordance and pointing out this affordance through the green points along the line we emphasize the user's ability to interact with this.

We hope that this visualization emphasizes the cost it takes us to communicate to UFOs and shines insights on the correlation between our spending, the interest in space, and the UFOs response.

Rocket Success Failure



Data Handling:

To look closely at the Space Mission data we decided to look at the status of each of the space missions between the time range of 1969 to 2019, which totaled 981 total space missions. Luckily the data and this histogram shows a more positive story: 924 successful space missions, and 57 failure missions. These failure missions break down into failure, partial failure, and prelaunch failure, which has 47, 9, and 1, respectively. The rationale behind the histogram of frequency was mainly due to the fact that luckily success is more apparent than other instances in this plot, which meant that a slider of years would result in more only success without failure, prelaunch failure, or partial failure. Thus we decided to combine all the years into one static delivery. Since it is static, there is no additional found elements that an individual needs to look through.

Some successful missions, include the SpaceX Mission in 2018 detailed Falcon 9 Block 5 | Bangabandhu-1 from the location Kennedy Space Center, NASA's Retired Mission called Space Shuttle Atlantis | STS-125 from Kennedy Space Center in 2009.

The Prelaunch Failure was from SpaceX in 2016 called Falcon 9 Block 3 | AMOS-6 launched from Cape Canaveral AFS– "On 1 September 2016, during the run-up to a static fire test, there was an anomaly on the launch pad, resulting in an explosion and the loss of the vehicle and AMOS-6" from Wikipedia showcases why it was called a prelaunch failure.

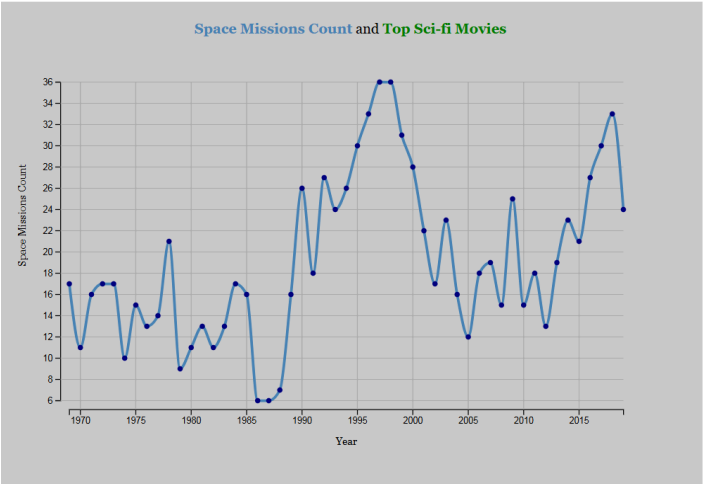
Some Partial Failure was from SpaceX in 2012 called Falcon 9 v1.0 | CRS-1 sent to Cape Canaveral AFS–according to the Wikipedia link "Of its five launches from 2010 to 2013, all successfully delivered their primary payload, though an anomaly led to the loss of one secondary payload", and another example was from Boeing called Delta IV Heavy | DemoSat and 3CS-1 & 2 in 2004 from Cape Canaveral AFS.

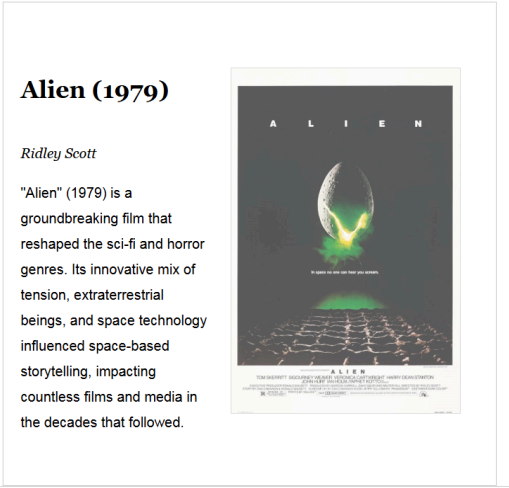
While lastly, some unfortunate failures are in 1999 such as Boeing and Lockheed called Delta III 8930 | Orion 3 and Titan IV(401)B | Milstar DFS-3m, respectively.

Design Rationale:

This visualization depicts our space mission dataset more closely, breaking down that not all space missions are successful and put next to the cost of space budget for NASA highlights how sometimes truly costly these mistakes and missions truly are. It showcases that while UFOs are giving us more and more signals, we've also poured more and more money every additional year, more and more time, to reach these UFOs.

Space Missions & Sci-fi Movie Visualization





Data Handling:

Space Missions:
We used the space mission dataset and summed up the total space missions for each year to understand how the amount changed from 1969 to 2019. In order to create a line chart to represent the changes and trends in space missions, we set the y axis as space missions count and the x axis as year.

Sci-fi Movies & Poster:
We cleaned up the Best SciFi Movies Of All Time dataset and chose the top 150 most significant movies. First, we calculate the average of the audience score and critic score. Then we sort them in a descending order to get the top movies. Next, we use the movieYear to match the year of space missions For the movie poster, we sourced the Alien (1979) poster image online and import in a text panel.

Design Rationale:

Our main goal was to understand the relationships between aliens, space missions, and the movie culture. Specifically, how does the perceptions of aliens and space explorations influence public perception and the mass media.

Space Missions & Sci-fi Movies Chart:
We overlaid the sci-fi movies dataset for each year onto the space missions count. A line chart was used to show a smooth progression of mission counts. The steel-blue line color was selected to represent the technological nature of space exploration. The green color was

selected to represent the alien and made consistent with the map visualization above. When users hover over each year, they can see the most important movies in that particular year. For mouseovers, a dedicated interactive region (activeRegion) ensures responsive and accurate mouse events. Marker lines, circles, and labels provide real-time feedback on data points, enhancing user engagement and allowing for precise exploration. We also integrated a default marker highlighting 1979, a significant year in the dataset of the Alien movie, to immediately engage users when they start the webpage.

Movie Poster:

We picked Alien (1979) as a representative movie for the visualization and introduced the audience to the dominant perception of aliens at that time. Ridley Scott's vision in this film left a lasting impact on the science fiction and horror genres. Its ability to create tension within the environment of the spaceship set a new standard for atmospheric storytelling. The use of unique designs the alien creature brought a haunting realism for the audience. Decades later, Alien remains a masterpiece, inspiring generations of creators in film, games, and visual media. We chose to dedicate a section of our newspaper layout to explain its significance. We used a flexbox layout to structure the section. By dividing the section into two parts—text on the left and an image on the right—we created a clear and intuitive flow for viewers. The content hierarchy is as follows: the title is presented in a bold <h4> tag, making it immediately stand out; the director's name is styled in italics to highlight it subtly; the main paragraph uses a modern, sans-serif font to ensure readability and a clean design aesthetic.

Team Contribution

Hung Ming Tseng

- 10-14 hours (includes team meetings, data cleaning, movie & space missions line chart, and movie poster visualization)
- Data cleaning – Cleaned Sci-fi movies Dataset and combined with Space Missions Dataset.
- Helped merged code including newspaper layout, charts, word cloud, and histograms
- Aliens, Space Missions, and the Movie Culture journal article

Jillian Beck

- 10-14 hours (includes team meetings, CSS newspaper styling, Map visualization, and journal write-up)
- Planned and implemented site CSS styling
- Created dynamic map visualizations

Hrishika Jotwani

- 10-14 hours (includes team meetings, word cloud, chart visualization, and journal write-up)
- Planned and implemented the NASA budget line chart
- Created word cloud that visualizes UFO shape

Hermione Bossolina

- 10-14 hours (includes team meetings, data cleaning, histogram, and other responsibilities)
- Data cleaning – I cleaned the UFO Dataset, NASA Dataset, and Space Mission Dataset.
- Organized Meetings and Helping to Set Deadlines + Agendas for Meetings
- Rocket Success and Failure + Journal Article