UFO & Sci-Fi
Data Viz

Adrien Villemin – Tianyuan Zhang

Link to our work: https://github.com/HalfAChance/UFO_sighting_dataset

1) Choose your Dataset

For the dataset we searched on Kaggle if we could find one interesting. And We found one about UFO's. We started thinking about what we could do with it, and it appears that we found it funny and interesting to work with.

So, we took a dataset on UFO census by people in the world. Here is the link if you want to check for more information's:

https://www.kaggle.com/NUFORC/ufo-sightings

On the website, we see some questions given by the provider of the dataset.

Inspiration

- What areas of the country are most likely to have UFO sightings?
- Are there any trends in UFO sightings over time? Do they tend to be clustered or seasonal?
- Do clusters of UFO sightings correlate with landmarks, such as airports or government research centers?
- What are the most common UFO descriptions?

They all seem to be interesting questions, and they are therefore very much the starting point for our exploration of the dataset.

2) Context and data assessment

There are a lot of variables in this dataset, but we took some that we found more interesting than the others.

For example, the shapes of the census. We can see that there are many shapes possible, and we wanted to know which one was the most seen by people.

Another very interesting information can be the location. And check where there are the most census made.

At the same time, we found some wrong or unusable variables in the dataset, for example, the latitude and longitude coordinates of the UFO observation events do not match the country information. This problem could not be solved by conventional data cleaning process (it would be too time-consuming to recalculate the country information based on the coordinates for all 90,000 data sets), so we had to find some other ways to use this dataset.

After some preliminary research, we found that there seems to have been some interesting changes in the shape perception of UFOs over the last hundred years, and that sightings of UFOs seem to be concentrated in some developed countries. Therefore, we propose some new hypotheses for this dataset:

"Do science-fiction and especially space exploration impact the vision of people on UFO and has an impact on the census of people in this dataset."

So, in this work we tried showing in which case the science fiction especially related to movies impacted the people's minds.

3) Data exploration with visualization

The first step of our work was looking at our dataset. Check all the variables to see which ones could be the more interesting to answer our problematic.

For that here is a list of all the different variables in the dataset:

- Time_clean
- City
- State
- Country
- Shape
- Duration (seconds)
- Latitude
- Longitude
- Year
- Hour
- Month
- Quarter

For our work we took various of these variables. Which we found more interesting. So, the time and especially the Year. To check this census compared to the same period with SF.

The position of the census so we can see a world map with all our census to see where it is more impactful.

And the most important one the shape. Which will probably directly be the impact of SF on people's minds.

To see a visualization of those variables you can check on the introduction parts our graphs. One is literally representing a world map with all the locations of all the census.

And the other one is a bar chart year by year showing the evolution of the shapes by the number of censuses made. So, we can see their evolution by time.

This graph will be the most important one for our work as we will use it to see if we can correlate certain periods of time between the SF and then censuses made. To see if the shapes shown especially in movies can impact the people's minds.

4)Summary

Message Part:

Directly we could see some more important variables in our dataset.

About the location's, we can directly see that the most censuses are made in developed countries, which have much more access to entertainments as movies or books. The mains ones are North America, Europe and a bit Australia. We also have some in India and China as they are the most populated countries, but the percentages are not that high as both previous ones.

Another thing we can see this time in the other graph about the shapes. Is that in the majority there are only 4 or 5 shapes that come out of the box. That people seem to see clearly much more times than the others. The main ones are Circle/disk, oval, cigar, light and also triangle.

Structure Part:

We created a sketch design like this:
(We cover the structural design in more detail later in the visual design section)

Introduction Part

Title 1 → the main problem

Trends of shapes

Title 2 → link to Sci-Fi
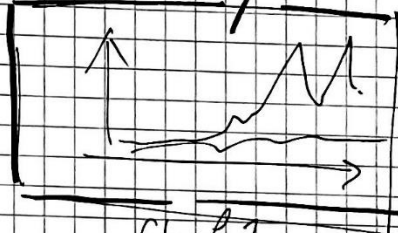
Chart 1

map → distribute shapes

Time Line → Period 1

PIE chart

Image → detail info.

1906

explain of Pie

1933

Image

Image 1945

→ detail info

PIE chart

Image 1996

explain of Pie

5) Story setting: structuring visualization messages

As we already talked a bit earlier. Our main message is to check if SF especially in movies can impact people's minds. And so, regarding the shapes of UFO the census. And compare them to the shapes movies in same period of time are showing.

With that we could probably have a first approach in the way that movies, books, etc. Have an impact in the unconscious of people.

If it results that yes, we could think that at the end we could be manipulated by some sort of propaganda films.

For that we built a specific graph. It is a timeline Going through different periods of time until now. And for each specific period. We compare the most censused shapes from people and the more impactful movies of this time with their homemade shapes for UFO's.

There is not really a main message at first for our graph. It is more an exploratory one. But depending on the results. We could have one message or another. It can be that movies have no impact but in the other hand if it's yes. We could push the reflection further as we mention just before.

6) Visualization design

In this data visualization report, we used three (actually, it should be five, as we explain below) different image tools to present our content, but as we said in the previous section, we wanted to explore and present content that goes beyond the dataset itself. In this report, images are narrative tools and evidence for argumentation, so we can't make a particular image that can present all the content quite perfectly, it's more about letting them each do their job and function as much as their capabilities allow.

1. Map

The report contains two maps, which we have placed at the beginning of the report, one of them is about the time statistics of UFO sightings around the world, and the other one is the number of science fiction movies around the world in the same period.

First, we chose to place these two charts at the beginning because it is their presentation that shifts our attention from UFOs themselves to science fiction films and movies. We hope that readers will use the content of these two charts to ask the same question we did, which is: "Are UFO sightings somehow related to science fiction films?" . For a long report, it is important to capture the reader's attention at the beginning, and that is what we hope to achieve with these two charts. But other than that, we don't expect the reader to get some specific information from the graphs, at least, not yet.

The first map is a density map, which is a compromise. As we mentioned earlier, our dataset lacks country information, only latitude and longitude are available, and the only map that can make good use of this information is the density map, while the observation time statistics are also an indicator of the frequency and the places of UFO sightings.

The second map is a bubble map, created using an IMDB movie dataset, which we only used to create this one map, so we did not mention it in the previous section. Our requirements for this image were not very high: to visually show the countries that produce a lot of science fiction movies. For the sake of aesthetic fatigue, we use color to encode the continents so that readers can

more quickly find the countries in Europe and the Americas that have produced a large number of science fiction works.

2. animated bar chart

This graph was one of the first we completed, and we have seen many cases of using this kind of chart to show the growth of confirmed cases of covid-19 during the pandemic. This type of chart has the advantage that by pausing or selecting a year, it can show the number and order of many variables at that moment in time, and when played, it can show the trend, especially if one variable is growing explosively, or far more than the others. This fits well with the characteristics of the "shape" variable in our dataset: there are almost 30 different shapes in total, with the majority of UFOs having a "disk" shape until 1996, and after which it is quickly overtaken by the rapidly growing "light", "triangle" and other shapes.

This chart has another purpose, and that is to introduce the content of the later section. Because most of the second half of the report is about the detailed explanation of this chart: we have split it into 4 time periods and have given some cases and explanations to help the reader understand it.

3. pie chart and sunburst chart

The charts in this section need to be read together with the text, because the information presented in the charts themselves is very vague and somehow a duplicate of the bar chart above (we have restructured the information for different time periods). Therefore, if we want to present the content more clearly to the readers and not bore them, we have to show something new in how we present the information.

So, the visual design we used in this section is actually a timeline, of which the pie chart is only one component. In this design, the role of the pie chart is to show the distribution of UFO shapes in this period, while the reader can read

some cases on the timeline to see our interpretation of the pie chart, and the scrolling timeline is designed to make it easier for the reader to read.

In this section, we divided the entire dataset into 4 periods based on time, with more "shape" types in the later parts. After some guidance from the teacher, we realized that a pie chart is not a chart for presenting more than 8-10 different types of data: it can show the distribution of each type very well, but not too much, because the reader will have difficulty distinguishing the differences between the smaller categories. To solve this problem, in the last three periods (since they contain many, many different categories), we replaced the pie charts with sunburst charts and grouped their distribution roughly, so that the reader can see a clearer distribution by clicking to see what is in each category.

We then realized that there was a cost to this, in that the inner-circle categories in the sunburst chart were not really meaningful and would interfere with the reader's understanding to some extent, but it also satisfied our need for this part of the chart. We considered dropping the less frequent "shapes" categories, but that would make the chart less credible, so we did not choose to do so. This is a matter of trade-offs, and we have not come up with a better solution...

Overall, we don't think this design approach is the best choice, but it is what we can get by doing our best. During the experimentation phase, we considered a number of different charts and considered introducing additional data sets to make the report better, but due to time constraints and limited capacity, this is all we could come up with.

7) Visualization's implementation

To realize our design, we decided to use dash, which combines text and graphics and the design can be refined through css, the only drawback is that learning to use it is hard task...

We presented this report to several students who gave us some criticisms.

1. It would be nice to upload to a server, otherwise you have to download all the files to see the page
(We tried uploading to heroku, but the free account doesn't have enough space to use...)

2. The page could be more beautiful
(We tried... This is our first work with css, maybe we can do better if we have more time)

3. animated bar chart operation is a bit troublesome
(dragging the progress bar need to double click, otherwise the progress bar will not move when the animation is played again, this seems to be a problem in plotly, we spent a long time studying how to solve this problem, but did not find a solution)

4. the page is too long, the timeline part can make some differences to.
(We need to improve our design ability and web production ability to solve this problem ...)

Overall, we spent a lot of time on creating charts, learning to build web pages and finding information, but the final result, we think, still could be improved: the design of the web pages is not attractive enough, the appearance of the charts could be better suited to the content of the web pages, the articles could be more interesting, and we could import a better dataset to make the report more convincible....