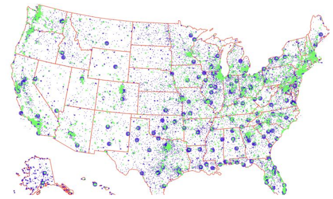
INTRODUCTION

This project consisted of exploring D3 to visually express the UFO data collected in projects 1 and 2 as well as using Image Space and ImageCat to find similarities between UFO sightings images that previously was not easily discernible based on the text captions and class identifications from project 2.

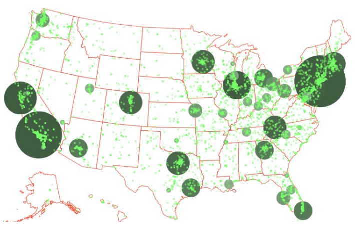
Why did we select our 10 D3 visualizations?



1. Maps

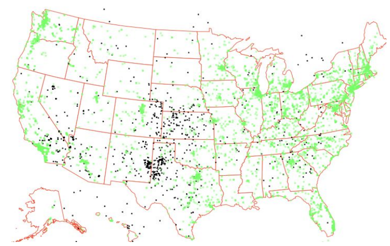
a. Airports

There are 3 classifications of airports: Large, medium, small. Larger airports are larger circles. Vicinity to airports can explain a UFO a sighting as one can mistake an aircraft for a UFO.



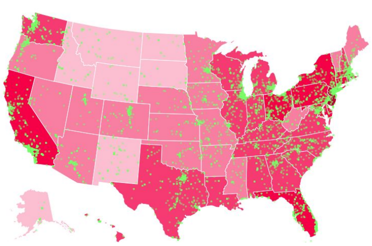
b. Metros / sports teams densities

Professional sports teams are typically located in large, population dense markets. More than 40% of UFO sightings occur in these “metros” and thus further suggest that UFO sightings in population dense areas can be indicative of 1) there is frequent UFO activity there or 2) people in these areas mistake planes, drones, odd light reflections due to skyscrapers, etc. for UFOs. Larger circles represent larger metros while smaller ones represent smaller ones. A metro size is determined by the number of major sports franchises in that region.



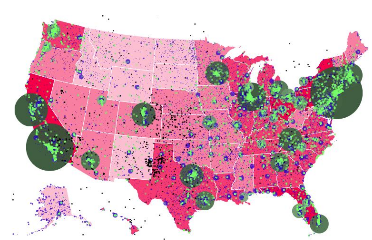
c. Meteorites

Meteorites are occasionally mistaken for UFOs as seen by Wisconsin's meteor madness and even the SpaceX launch in December 2017 that resembles the shape and color of a meteor falling to Earth. Our hypothesis is that meteors can help explain some UFO sightings in rural areas in the U.S and plotting this information makes it easier to see that some rural sightings can be linked to meteorites.



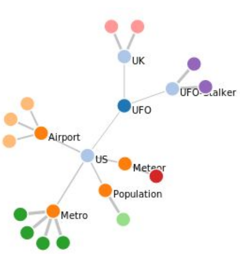
d. Population density (state-wise)

Population dense locations have thus been identified and shown using features that are the result of a location being population dense. Population density can be better visualized with darker states being more population dense than lighter colored states. This can further show users how population density and its resultant features (airport, metros) influences the number of UFO sightings in an area.



e. All map visualizations overlaid/toggle-able with checkboxes:

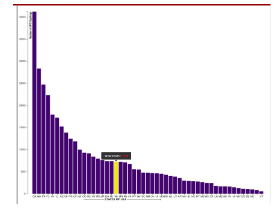
This allows for a great visualization in which the user can play around with the 4 different map features. All 4 overlaid shows that UFO sightings occur commonly in population dense locations, while there are much less sightings in rural areas.



2. Relation Graph

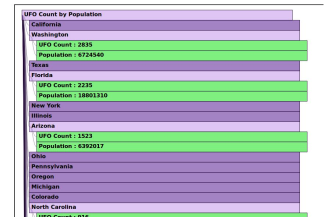
We wanted to show how we have modified the initial UFO data set by adding new features from various others datasets that we decided to choose using this visualization. It starts with the initial UFO dataset as the center and branches out to US and UK datasets which later branch out to show all the other datasets that we used under these category.

3. Bar chart

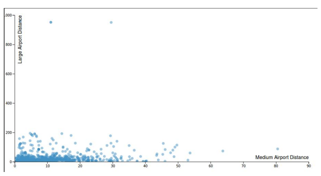


The best way to show frequency in a way that is intuitive and simple is bar chart. So, we chose this representation to show the number of UFO sightings for individual state using bar chart. Also, D3 helps us make it more interactive and attractive.

4. Indented Tree



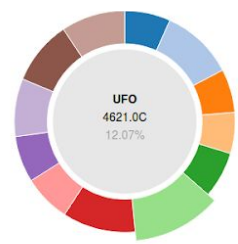
We wanted to see correlation between population and UFO sightings reported. A good way to represent this is show state wise population and UFO sightings. We use indented tree to represent these numbers for each state as it is ideal to visualize the states as a tree structure.



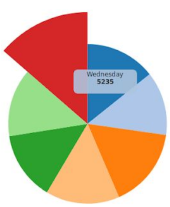
5. Scatter Plot

Sometimes flights can be mistaken for an UFO. In order to depice this, we have a scatter plot of UFO sightings and distance of the sighting from the airport. We plot the UFO sightings with nearest large airport on x-axis and nearest medium-sized airport on y-axis. So, the sightings which are at reasonably equal distances from largest and medium sized airport would be clustered together. **This implementation is done using elasticsearch.**

6. Donut Chart



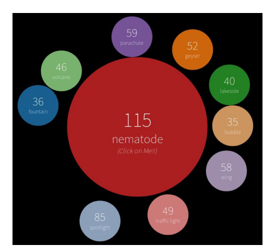
We plot the data of UFO sightings month wise using an interactive donut chart. Here, the number in the center represents the total UFO sightings, whereas we can get UFO sightings for a given month as we hover over the month.



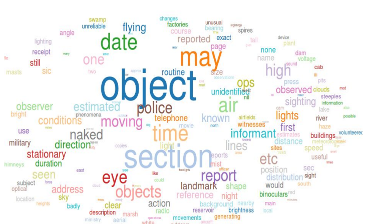
7. Pie chart

We wanted to see correlation between days and sightings reported. We show the number of UFO sightings day wise using an interactive pie chart. And we can see that there are maximum sightings on Saturday. We had an intuition that as people are comparatively free on weekends, there are more observations on weekends, and the data supports that.

8. Bubble Chart



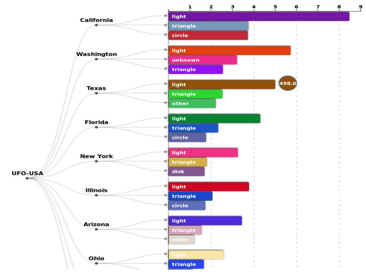
Getting the count of UFO sightings of different shapes gives us an idea of the most commonly seen shapes. We plot this using a clickable Bubble Chart. Every bubble represents a shape and has the number of UFO sightings corresponding to that shape



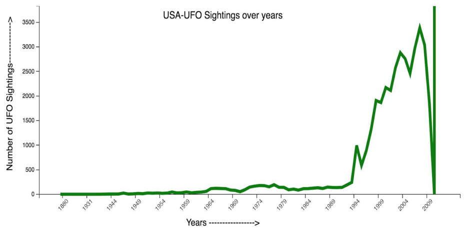
9. Word Cloud

In order to explore what words occur in description of UFO sightings, we draw a word cloud of the most frequently used words in description of UFO sightings in UK

10. Dendrogram



We are using a dendrogram to depict the count of UFO sightings per shape for different shapes. It is a hierarchical representation, depicting the count of few shapes for few states of US. We select only few states so that the visualization looks neat.



11. Trend Line

Trend line is an efficient way of showing how the trends of the number of the UFO sightings in USA vary. We extracted the years of the sightings in USA and calculated the number of sightings seen in those year span.

ImageSpace and ImageCat

We were successfully able to complete the **ImageCa**t and **ImageSpace.** We indexed 3600 images.

File name	Notes	Number of images returned (N)	How many we thought were actually similar (x)	Accuracy (x/N)
83543_submitter_file1__IMG4727	Sky, land, clouds	21	16	.76
83565_submitter_file2__IMG0354	Small light, sky	21	16	.76
84651_submitter_file1__DSC92992.jpg	Dark images	21	11	.52
84735_submitter_file2__1009am2	Black & white cloud	21	2	.09
F89382_submitter_file3__IMG4573	Sky, trees, people	21	13	.619
83543_submitter_file1__IMG4727	moon	21	14	.66
85721_submitter_file1__Ontop1	View from airplane	21	6	.28
86245_submitter_file1__20170814104918	lightning	21	13	.619
86560_submitter_file2__DSC011133	Sunset, text in image, trees	21	2	.09
88305_submitter_file4__16.11.04RRUMarsMoonVenueSatturnIMG644916	Satellite launch	21	4	.19

Note: We've included the file name that we searched for in flann_index. You can find this image in the /images_to_index folder and look for the results flann index outputted in the /flann_screenshots folder

We noticed that there is a similarity calculation in flann_index for each image. The similarity scores for dark images tends to be high while the images are typically unrelated. Histograms are computed from the images and similarities are computed across all combinations of target image and indexed images. Because histogram method is used, same images from different angles are often not included or have low similarity scores -- computing histograms and using a mixture of similarity scores on convolutions of the images may increase the similarity score of same UFO images from different angles.

Another noteworthy thing we noticed is that some UFO Stalker images were repeated - meaning multiple UFO sightings submitted exactly the same images. Imagespace with flann index can be used to identify plagiarized images.

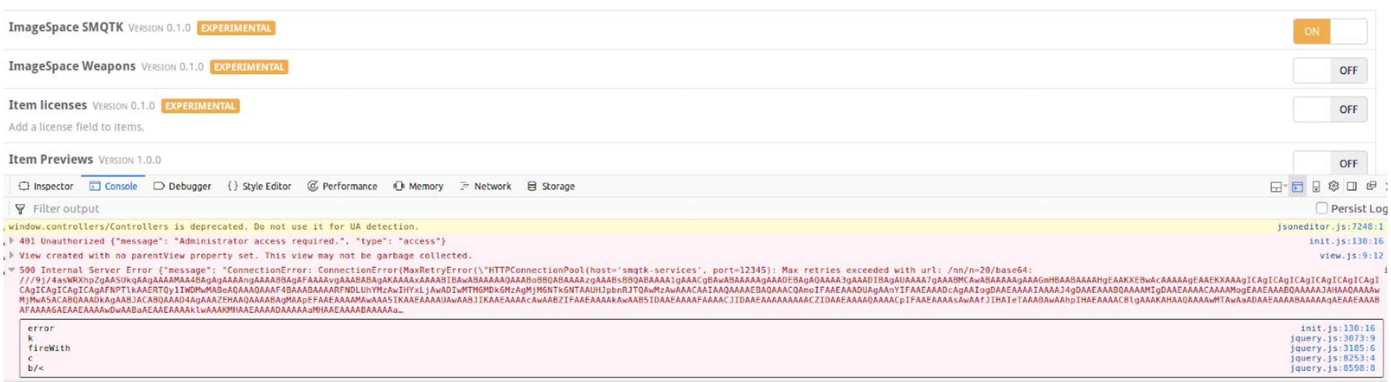
Did Image Space allow you to find any similarity between UFO sightings images that previously was not easily discernible based on the text captions and object identifications you did?

Overall, Imagespace does a great job of identifying similar images. ImageCat performs OCR on images and extracts text from images similar to the way we did in project 2, but also adds rich metadata about the image that Imagespace uses to allow users to search for images. Flann index adds more granularity to image search as it builds histograms from each image and returns images with similar histograms (essentially similar images) when Flann index is queried.

Imagespace allows for much greater image exploration than Tika Vision and Caption due to the fact that similar images don't rely on labels from a classification model that previously has not seen UFO training data. For example, when we inputted an image of the Spacex satellite launch from December 2017, ImageSpace returned similar UFO images that had bright circular/oval shaped lights. To conclude, ImageSpace is better to use to find similarity between UFO sighting images.

Extra Credit:

1) First, we exported data generated by ImageSpace that was put into ImageSpace Solr and batch put this data into our elasticsearch instance. Next, we modified imagesearch_rest.py found in image_space//imagespace/server/imagesearch_rest.py to use the elasticsearch python package to retrieve data. This allows girder to query elasticsearch instead of solr. Please take a look at our imagesearch_rest.py
2) **Trying different plugins for Image Space:** We tried to use SMQTK plugin in Image Space. We were able to successfully complete the installation and had enabled it, but an error when we attempted indexing images with SMQTK enabled. Here is the screenshot of the enabled plugin and the error that we ran into:



NOTES:

1. We have implemented the d3 visualizations in **both ways** ,i.e. Using json/csv files and also using elasticsearch. Both of the implementations have been pushed.
2. To view the Scatter Plot implementation, the elastic search has to be up and running at ufo.usc.edu website. Then please uncomment the <iframe> for scatter plot in "HW3_website/html/team3_html/team3_index.html". Also, change the link to elastic search in "HW3_website/js/team3_js_files/scatter.js".

TEAM MEMBERS:

- Khyati Ganatra
- Pavneet Kaur Mukar
- Prerana T H M
- Sachin Kumar GB
- Sanjay Nadhavajhala