Visual and Geospatial Analysis of BFRO Sightings

D3
In the development of the selected five D3.js visualizations, the primary objective was to effectively showcase a variety of data presentation methods that enhance both understanding and engagement through interactive and visually compelling elements. Each visualization was meticulously chosen to address specific analytical requirements and to demonstrate advanced capabilities developed during earlier assignments.

The scatter plot of Bigfoot sightings over the years serves as a robust tool for highlighting trends over time, offering viewers an intuitive grasp of how Bigfoot sighting frequencies have changed annually. This visualization is enhanced with features such as tooltips, color gradients based on data density, and interactive axes, demonstrating proficient handling of temporal data in a user-friendly format.

The pie chart categorizing Bigfoot sightings by class efficiently segments the sightings into distinct categories (Class A, B, C), providing a clear visual summary of the data distribution. Integrating dynamic color scales and interactive tooltips in the pie chart ensures that users can easily understand each category's proportional significance and specifics. This method effectively communicates categorical data distribution, which is essential for quick comprehension of diverse data segments.

Lastly, the bar chart breaking down Bigfoot sightings by season illustrates how different times of the year might influence the frequency of sightings. By employing color gradients and hover effects, the bar chart not only makes the data more accessible but also visually emphasizes the seasonal variation in Bigfoot sightings. This visualization is particularly valuable for demonstrating seasonal patterns, an important aspect of the dataset.

Overall, these visualizations were selected for their ability to succinctly and effectively communicate complex datasets, showcasing the ability to enhance user interaction and data interpretation through sophisticated visualization techniques. The choice of visualizations not only fulfills the basic requirements of the assignments but also highlights advanced data presentation skills in D3.js, catering to both general and specialized data analysis needs. This strategic selection and enhancement of visualizations demonstrate a comprehensive approach to data presentation, crucial for insightful data analysis and engagement in any analytical endeavor.

Apache Solr
Apache Solr was chosen over ElasticSearch for the ingestion of the data from the BFRO sightings JSON file because of greater familiarity with Apache Solr and its ease of use compared to ElasticSearch. The ingestion of the BFRO sightings data into Apache Solr allows for greater exploration and analysis of the dataset. This is because the indexing capabilities of Solr allow for the creation of a structured index of the dataset, making it easier to search for specific information within the data. This is particularly beneficial when dealing with large datasets like the BFRO sightings data, as it allows us to quickly locate relevant records based on various criteria such as year, state, season, or other attributes.

Using Apache Solr for the data ingestion was straightforward, and the ability to easily create a structured index of the dataset is a notable advantage of the platform. The only slight issue that was encountered when using Apache Solr was having to explicitly include the parameter ‘rows=5112’ in the Solr query. At first, the ‘rows’ parameter was left empty, but the Solr results JSON file only returned 10 rows when the parameter was left blank. This is because Solr defaults to only returning 10 rows if the ‘rows’ parameter is left unspecified. After examining the Solr results JSON file and noticing that under the ‘params’ section, the ‘rows’ parameter was set to 5112, it was deduced that the rows parameter needed to be specified to ensure that the full dataset was obtained in the JSON output. However, other than this minor issue, it was an overall uncomplicated experience working with Apache Solr.

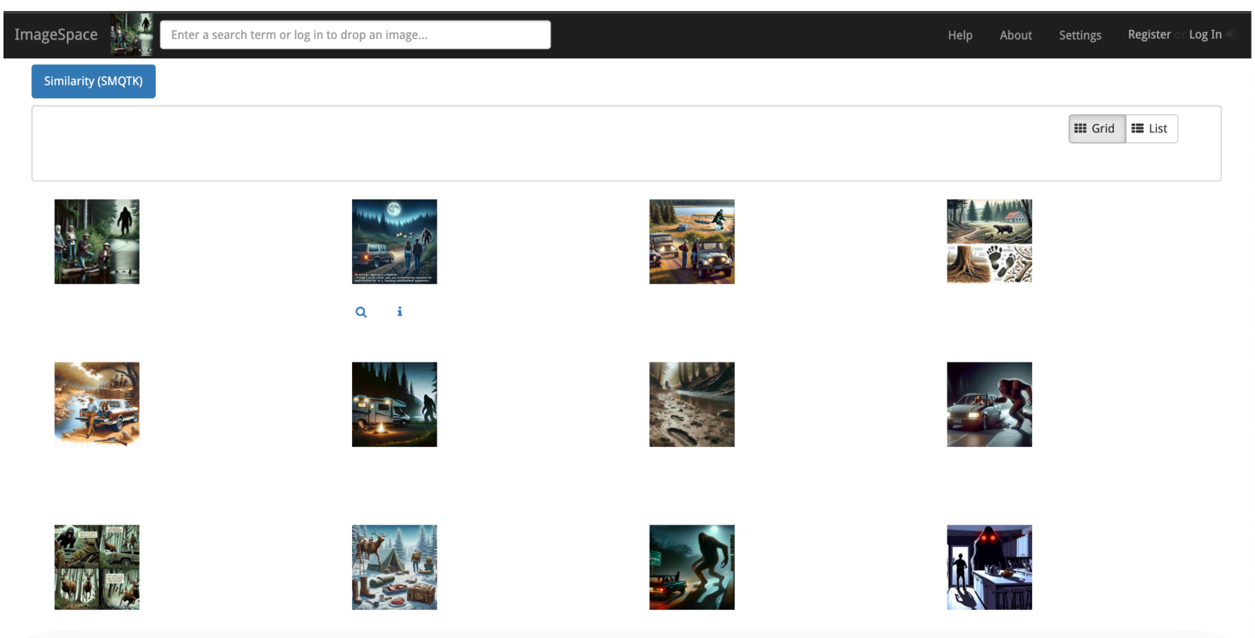
The task of tarring and gzipping the Solr index was also easy to carry out, and this process reduces the size of the file, making it easier to share across networks. Overall, with the use of Apache Solr, it was easy to effectively index, search, and analyze the BFRO sightings data. By ingesting the data into Solr, we can use its advanced features to examine the dataset for any potential valuable insights. Tarring and gzipping the Solr index enhances its portability and accessibility, making the file easier to share with others who are interested in accessing and analyzing the indices.

Image Space

In our exploration with Image Space, we encountered several instances where the tool revealed intriguing similarities between BFRO-generated Bigfoot images that were not immediately apparent through manual observation. One particularly compelling example involved an image depicting people engaged in recreational activities, such as fishing, with a purported Bigfoot figure lurking in the background.

Upon uploading this specific image to Image Space and utilizing the similarity search functionality, we were astounded by the diverse range of scenes that emerged as potential matches. Rather than solely retrieving images of similar woodland settings with Bigfoot-like figures, as expected, the results yielded a fascinating array of scenarios.

Among the similar images returned by Image Space were photographs depicting individuals camping in remote wilderness areas, families embarking on road trips through scenic landscapes, outdoor enthusiasts hiking along rugged trails, and even people enjoying winter sports in snowy mountain regions. Surprisingly, each of these images featured subtle yet unmistakable resemblances to the original image, with enigmatic figures resembling Bigfoot captured in the background or distant foreground.



This unexpected breadth of similarity highlights the versatility and depth of Image Space's image analysis capabilities. The tool facilitated a broader exploration of Bigfoot sightings and their associated contexts by identifying common visual elements and patterns across seemingly disparate images. Moreover, the ability to visualize these similarities side by side enabled us to discern subtle nuances and correlations that might otherwise have gone unnoticed.

This diversity of scenes provided a broader context for understanding Bigfoot sightings, suggesting that encounters with these elusive creatures may occur across a range of environments and geographic locations. From dense forests to open fields, and from summer camping trips to winter expeditions, the presence of Bigfoot-like figures in diverse settings hints at the widespread nature of these encounters and the enigmatic allure of the creature itself.

Overall, Image Space's ability to reveal diverse scenes featuring potential Bigfoot encounters transcends traditional image-matching approaches, providing invaluable insights into the multifaceted nature of reported sightings. Its contribution to our understanding of the Bigfoot phenomenon underscores its significance as a powerful investigative tool in the realm of image analysis and forensic investigation.

The experience with Image Space and ImageCat offers intuitive solutions for enhancing the discovery of similarities among photos. Image Space, particularly advantageous for Windows users, compared with other tools I used from previous assignments, facilitates ease of installation through WSL (Windows Subsystem for Linux), allowing similar emulation of a Linux environment. This enables users to navigate and manipulate data without cumbersome processes conveniently. Both platforms empower users to efficiently explore and identify connections within their image collections, simplifying the task of recognizing similarities, because the inclusion of a graphical user interface (GUI) further enhances the user experience, making operations even more convenient and accessible.

MEMEX GeoParser

As shown in Figure 1 below, the locations obtained from the GeoParser are quite dispersed, covering nearly all of the United States. Still, they are primarily concentrated along the West Coast and in the Mid-Eastern and Eastern United States.

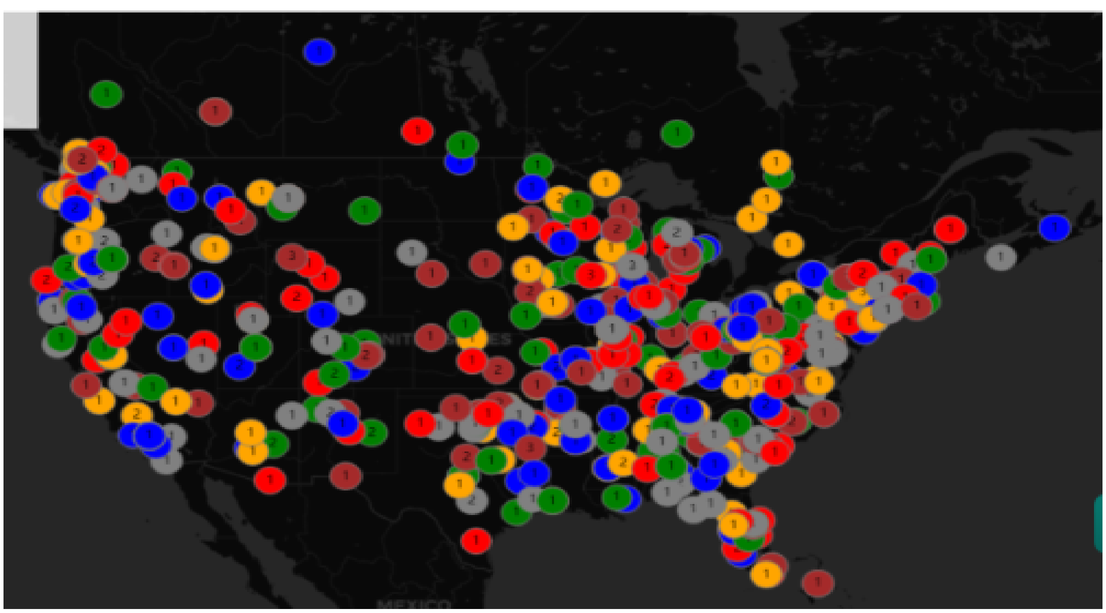


Figure 1: Memex GeoParser Result 1

Also, there are some locations outside of the United States, as shown in Figure 2. The text inputted into the GeoParser has been made as detailed and logical as possible, but it still produces these results. Therefore, these errors may be due to the GeoParser being particularly sensitive to certain terms, resulting in it mistakenly identifying non-location content in the text as locations, and showing them on the map. These locations outside the United States will be ignored since they are definitely errors. The reports only contain sightings in the United States.

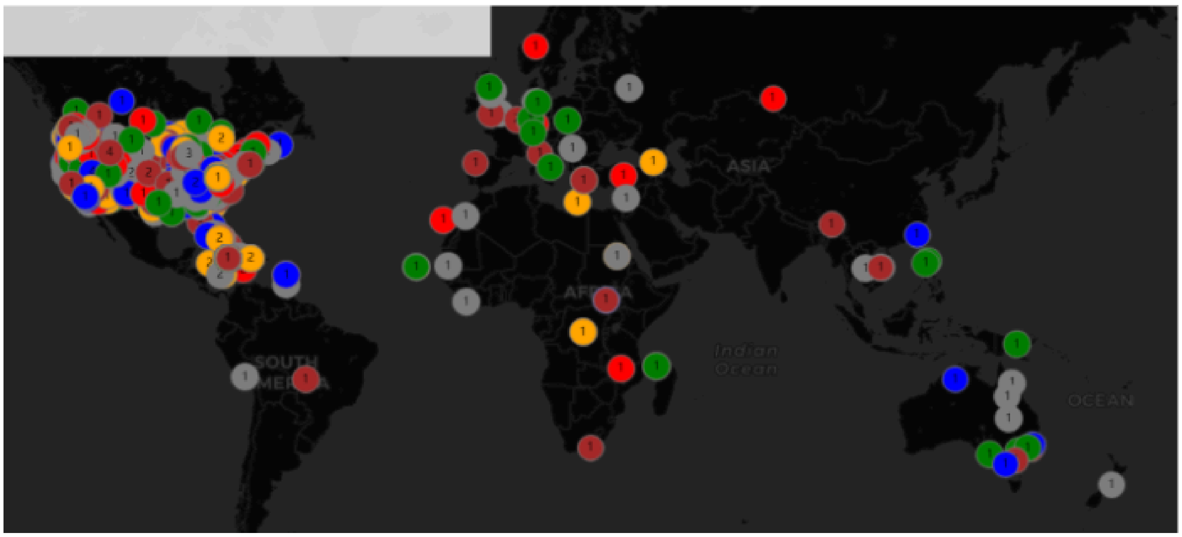


Figure 2: Memex GeoParser Result 2

The locations we get are quite dispersed, but we can still see some patterns and make some deductions. The sighting locations are primarily in places where there are forests, or the vegetation coverage is relatively high. As we can see, there aren't many locations in Nevada, Utah, or Wyoming. These areas are quite arid and have sparse tree coverage, whereas Bigfoot may generally be found or sighted in wooded or forested areas. In assignment 1, we added the fields "Forest cover rate" and "Number of lakes." By filtering these fields, we can see that 60% of the states where Bigfoot sightings occur have forest coverage of over 30%. More than half of these states have over 1,500 lakes, indicating that most Bigfoot sighting locations are in areas with high forest coverage or are vegetated and not very dry.

However, it is important to note that the Memex GeoParser is extracting multiple locations from the same text or JSON, while we only expect it to extract and show the one that is most detailed and most frequently mentioned by the text we inputted. The text we inputted contains features including the headline, the state, the county, the location details, the nearest town, and the nearest road, and these features are combined into a complete sentence, like "the headline of the report is…, and the state of the location is…". When we just started to test the GeoParser, we also included in the text that these locations are in the United States, but this resulted in the GeoParser identifying the United States as a parsed location and showed it on the map, which was represented as a point in the center of the United States. If we didn't exclude this information, we may see this point as some sightings of BigFoots in some Middle States in the US. But even if we have excluded the United States from the text in our final practice, it is difficult to create text inputs that make the GeoParser able to identify only one location per json file. In this case, this map created by the GeoParser might be biased in some ways, because we don't know which of the obtained locations are categorized by state, which is by town, or county, it makes it difficult for us to make accurate judgments based on the map. Still, we can make some inferences by looking at the possible "clusters" or areas that have more locations and areas that have fewer, just as we mentioned above regarding the possible relationship of the density of locations with the forest cover rates in different regions.