

# UFO Sightings and Presidential Elections

## Project Proposal

**Shelby Bearrows**  
Applied Computer Science  
University of Colorado  
Boulder, CO  
shbe4654@colorado.edu

**Gurhar Khalsa**  
Applied Computer Science  
University of Colorado  
Boulder, CO  
gukh8079@colorado.edu

**Sam Cramer**  
Applied Computer Science  
University of Colorado  
Boulder, CO  
sacr2234@colorado.edu

**Kara Wolley**  
Applied Computer Science  
University of Colorado  
Boulder, CO  
wolley@colorado.edu

### PURPOSE

UFO sightings across the Globe has been a highly controversial topic for over a century. Mainstream media rarely publishes possible sightings for fear of ridicule while individuals that report a sighting are often cast out from popular society. This investigation will focus on trends in UFO sightings in the United States and increase understanding patterns of sightings, occurrences and sighting geographical location. This information will allow the authors to identify trends in UFO shape to better understand advances in alien technology, or at least how human perception of aliens have changed over time. Additionally, the authors will portray differences in sighting frequency as a function of geographical location. A secondary goal of this investigation is to explore potential relationships between presidential elections and reported UFO sightings. Presidential election data will shed light on the relationship between geographical region, UFO sighting frequency and UFO sighting description as a function of political party. Finally, we aim to predict where and when UFO sightings occur based on the outcomes of US presidential and general elections.

### PRIOR WORK

To the author's knowledge, little work has been published on the relationship between

UFO sightings and election information. Regarding UFO data analysis, prior work from the data source (National UFO Reporting Center, 2019) also includes 166 references to additional analyses performed by individuals. Of these resources, most explored frequency of sightings by location (including per capita), word cloud analysis on the description attribute, shape of UFOs by time of year, sightings by shape and season, and shape frequencies. Perhaps the most explored question is whether sightings have increased over time, which some reported a possible correlation. An additional correlation between sightings and geographical region was reported (Adhokshaja, 2016) using a chi-square analysis. While shape over time has been explored, the time variable was usually by week or month, rather than by decade for example (this could be interesting to explore).

Additional work has been done by a group of masters students from USC Viterbi school of engineering, using data scraped primarily from ufostalker.com. They were able to conclude that "sci-fi movie releases correlated with an increase in sightings" and "that events like thunderstorms caused an uptick in reports" (Dawson, 2018). In addition, the data suggested higher frequencies of sightings during summer months. Word cloud created from their data also showed similarity in words used to describe these sightings.

While others have explored some of the main questions regarding ufo sightings, merging our dataset with election data will likely add a new level of novelty to our work not present in the prior work.

## **PROPOSED WORK**

The primary purpose of this research will be a novel, intuitive and visual representation of UFO sightings that provide a new perspective to the authors. A secondary goal of this project is to integrate election data with the UFO sightings data. With the integration of these data, our research will aim to answer the following:

1. Are UFO sightings increasing over time?
2. How has the UFO shape changed over time?
3. If UFO sightings do increase over time, how is that correlated with the turn out rate for elections?
4. Do states that vote Republican have significantly more UFO sightings?
5. How does term number affect the number of UFO sightings? Is this also dependent on the political party?

## **DATA COLLECTION**

Two data sets were collected and compiled for use in this research. The primary data set consists of UFO sighting data published by the National UFO Reporting Center and spans almost a century (National UFO Reporting Center, 2019). Each entry includes date and location of the sighting, along with a description of the sighting that include the shape of the UFO and the duration of the sighting. This data set is approximately 28MB in size and contains 11 different attributes of UFO sightings. The secondary data set includes election data collected by the MIT Election Data and

Science Lab (Tunguz, 2020). The dataset covers the complete voter breakdown in each county for every presidential election since 1976. It also includes information about each candidate that was on the ballot in each county like party affiliation and the number of votes they received in that county. The table is 4.5 MB and contains 11 different attributes for each entry. Preliminary work includes cleaning, preprocessing and integration of data. The UFO dataset is especially 'dirty'. Reformatting of dates, state name and removal of random symbols (&, !, #) from the comments attribute is necessary. The US Elections dataset does not appear to need much cleaning. Preprocessing includes correcting missing values, null handling and identifying erroneous data or outliers. Once these preliminary tasks were completed, the two data sets were integrated into a single table. In total, this research will include 17 attributes for analyses.

## **PRELIMINARY DATA ANALYSIS**

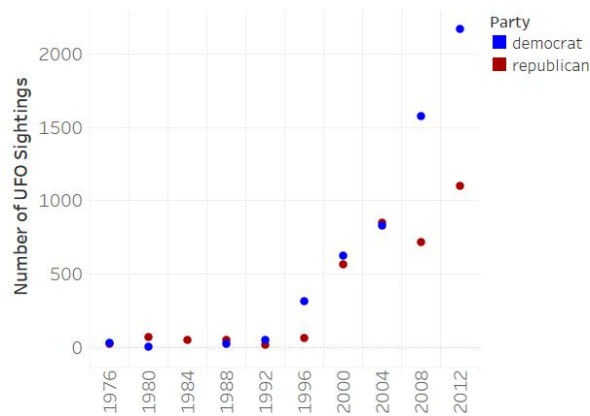
A preliminary data analysis was performed to explore data availability and initial responses to proposed research questions. First, a heatmap was created using the total number of UFO sightings by city in the lower 48 states in the United States (Figure 1). This analysis leads us to believe that more sightings occur in larger cities and more sightings occur in coastal areas. This is in direct contrast to what is typically believed of UFO sightings, which is that they occur primarily in Nevada and the regions surrounding it. Doing more research to understand the political leanings of large cities and coastal cities is required to make further conclusions about the political leanings of these urban areas.



**Figure 1: Heatmap of UFO sightings in the United States**

Second, an initial comparison between UFO sightings and election votes, across all states, was performed (Figure 2).

**Number of UFO Sightings by Election Year**



**Figure 2: Preliminary investigation on the total number of UFO sightings by election year. Sightings are counted as the sum of all sightings by party in that year.**

This preliminary analysis, although insightful, highlighted the need for additional data cleaning techniques. The heat map shown in Figure 1 lacks population density analysis. Additionally, Figure 2 simply removed data that did not report a political party or the political party was not registered as democrat or republican. These elements will be included in future work.

## EVALUATION METHODS

The main evaluation methods we are currently planning to use are histograms, heatmaps, apriori's algorithm, pearson's correlation coefficient, lift, and Student's t-test. We will use histograms to create bins of the different terms used to classify UFO shape. We will try to compare histograms based on different election years to see if the common descriptors change with time, and if they do, are they affected by election results. We'll use heatmaps to visualize geographic location of UFO sightings to better understand where they are happening and compare these to similar heatmaps of how the different regions of the country skew in terms of which party and candidates they vote for in elections. Apriori's algorithm will be used to quantify minimal support for frequent descriptors used in UFO sightings. Depending on the results of this mining, we'll try to group the UFO data by location and see if we can get different results for regional grouping. Pearson's correlation coefficient and Lift will be used to determine correlation between the UFO and election datasets. Student's t-test will be used for a comparison of means between the datasets.

## TOOLS

For our data analysis we will be using a myriad of programming languages and packages. The programming languages we will be using are MATLAB, Python and R. For data cleaning and tidying we will be using numpy and pandas in Python and tidyr and dplyr in R. For data visualization we will use matplotlib and seaborn in Python, and ggplot in R. We will also utilize Tableau for quicker, short-term data visualization.

## MILESTONES

1. Oct. 16: Data cleaning
2. Oct. 21: Final decision on integration of electoral data. Preliminary data selection, transformation and visualization
3. Nov 5: Final decision on integration of other potentially interesting data.
4. Nov 7: Data mining and evaluation with second pass of visualization
5. Nov 21: Preliminary pass of knowledge presentation.

## REFERENCES

[1] Adhokshaja, P. (2016, November 22). *UFO Data Analysis*. EDA and Chi Squared Test. Retrieved October 21, 2020, from: [kaggle.com/adhok93/eda-and-chi-squared-test](https://kaggle.com/adhok93/eda-and-chi-squared-test)

[2] Dawson, C. (2018, July 2). *Data-Crunching Class Looks to the Skies to Explore UFO Sightings*. USC Viterbi News. Retrieved October 20, 2020, from: [viterbischool.usc.edu/news/2018/07/data-crunching-class-looks-to-the-skies-to-explore-ufo-sightings/](https://viterbischool.usc.edu/news/2018/07/data-crunching-class-looks-to-the-skies-to-explore-ufo-sightings/)

[3] National UFO Reporting Center. (2019). *UFO Sightings*. Kaggle. Retrieved September 14, 2020, from: [kaggle.com/NUFORC/ufo-sightings/notebooks](https://kaggle.com/NUFORC/ufo-sightings/notebooks)

[4] Tunguz, B. (2020, October 13). *US Elections Dataset*. Kaggle. [kaggle.com/tunguz/us-elections-dataset](https://kaggle.com/tunguz/us-elections-dataset)