**Analysis and Prediction of UFO Sightings**

Alexandr Lagornii 1

*1 – Technical University of Moldova*

e-mail: [alexandr.lagornii@iis.utm.md](mailto:alexandr.lagornii@iis.utm.md)

**Abstract**

Unidentified Flying Objects (UFOs) have long captivated the imagination of the public, sparking debates and speculation about extraterrestrial life and advanced aerial phenomena. In this study, 80,000 reports of UFO sightings across the world from 1998 to 2014 were analyzed. In addition to graphically showing a lot of curious facts about those sightings, the linear regression model was built for predicting the number of the UFOs to be seen in the future years.

**Introduction**

The study of Unidentified Flying Objects (UFOs) presents a very fascinating and complex puzzle. The term itself merely denotes an aerial object that can’t be identified, but oftentimes causes us to think about aliens and other extraterrestrial life.

This study is aimed to find out why do people see UFOs in the first place. Do aliens actually exist or is it all just our imagination.

We will show the areas of the world that are most likely to have UFO sightings, find out the most common UFO descriptors, and the time when UFO sightings are extremely common. On top of that, we will see what people say when they report UFOs. In the end, we will try to build a simple model that will show how UFO sightings will never stop, but increase as years go by.

**Materials and Methods**

*The Analysis*

The whole analysis – dataset, plots, this paper, as well as PowerPoint presentation are freely available on the platform named GitHub. The repository is “alexandrlagornii/UFO-Analysis” [1].

*The Dataset*

The chosen dataset is called “UFO Sightings” and is freely available on the popular data science platform – Kaggle [2]. It contains over 80,000 reports of UFO sightings over the last century.

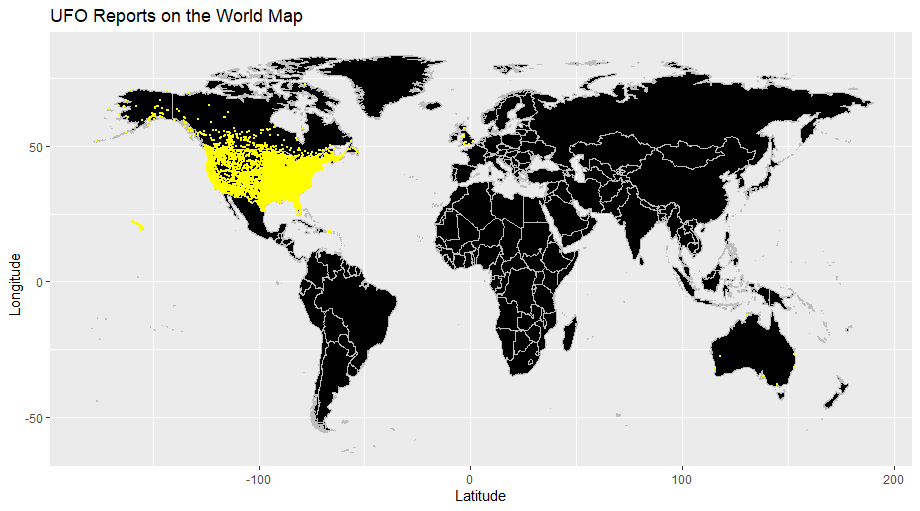
*Tools for Analysis*

The main tool for this analysis is R. This is a language and environment for statistical computing and graphics. It has a lot of useful libraries that allowed us to make good-looking graphs easily, such as “tidyverse”, “ggplot”, “dplyr”. On top of that, a custom library from GitHub was used – “wordcloud2” for building a wordcloud of what people said about UFOs they’ve seen [3]. Libraries for building machine learning models were available and very easy to use in R, which made creating the model a very simple task.

The main analysis was made using technique called EDA (Exploratory Data Analysis), where we show plots that might tell us something about the data.

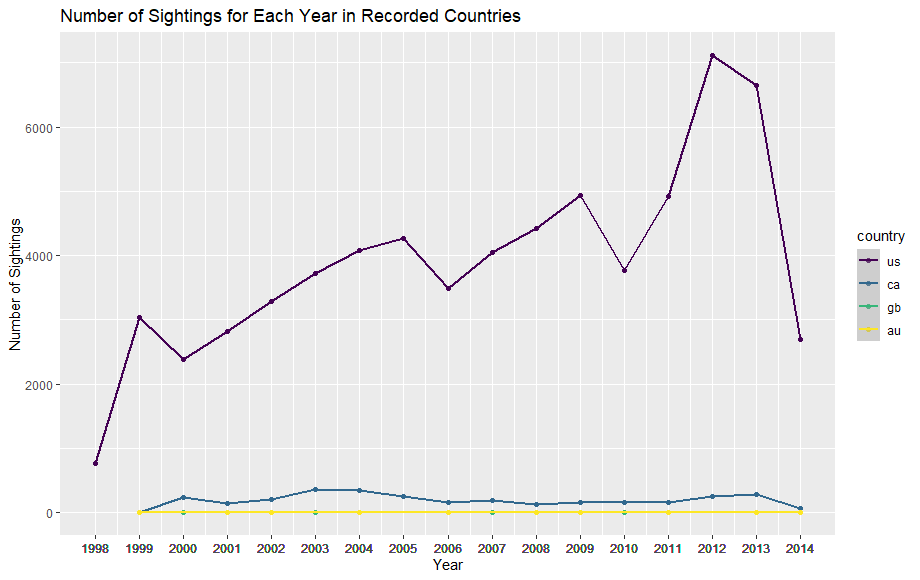
**Results**

Based on 80,000 UFO reports we found out, that the most dominant part of the world that reports UFOs is the Northern part of America continent.



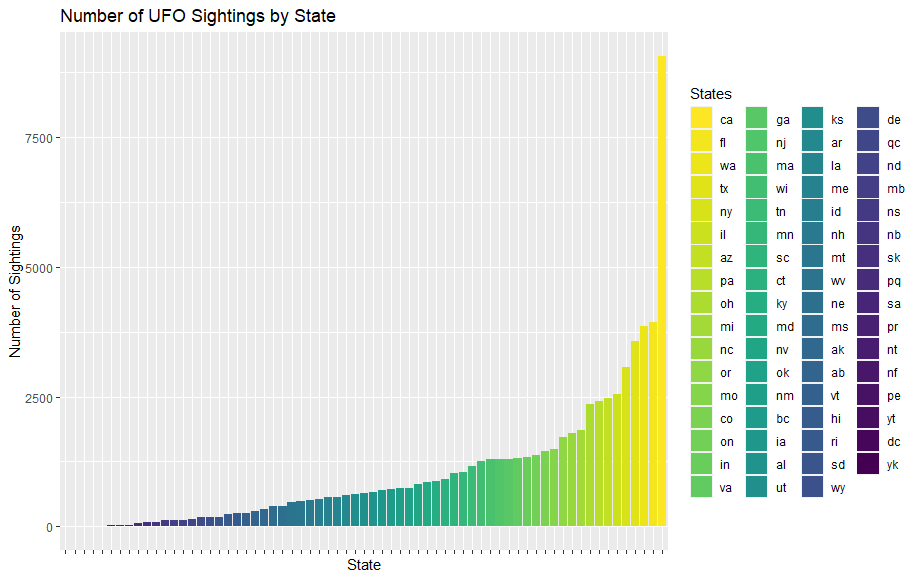
**Figure 1.** UFO Reports on the World Map

The country that reports the largest number of UFOs is the United States of America. Canada places second, Great Britain and Australia are tied for the third place. We can also see that the largest number of reports is in the year of 2012.



**Figure 2.** The number of Sightings in Recorded Countries

The state that has the most reported cases of UFO is California followed by Florida, Washington, and Texas.



**Figure 3.** Number of UFO Sightings by State

The largest number of UFO sightings each time of year is in the summer, followed by winter, fall, and spring.

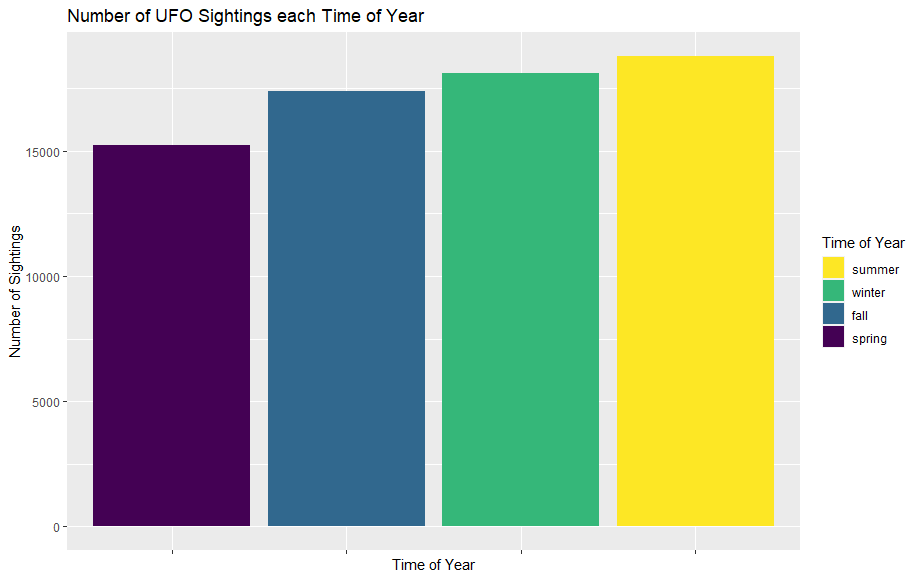
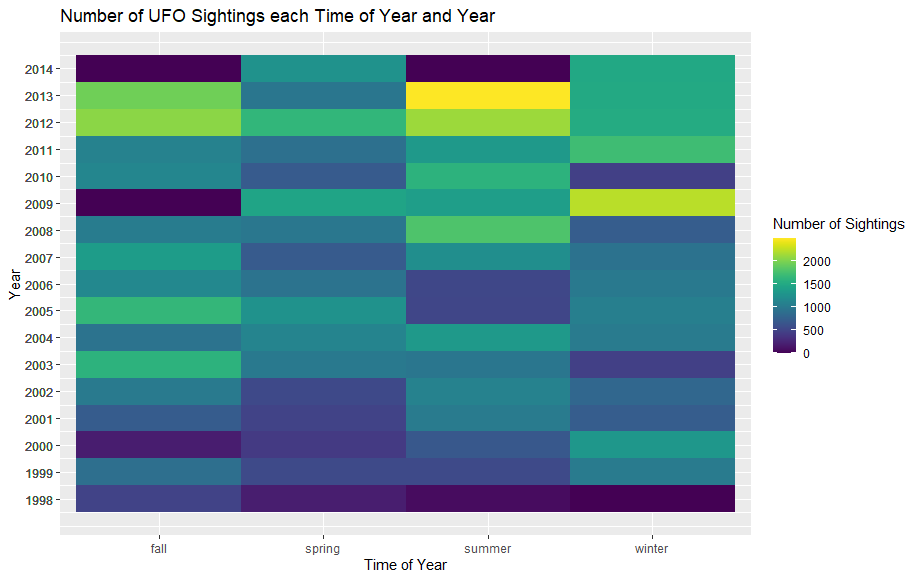


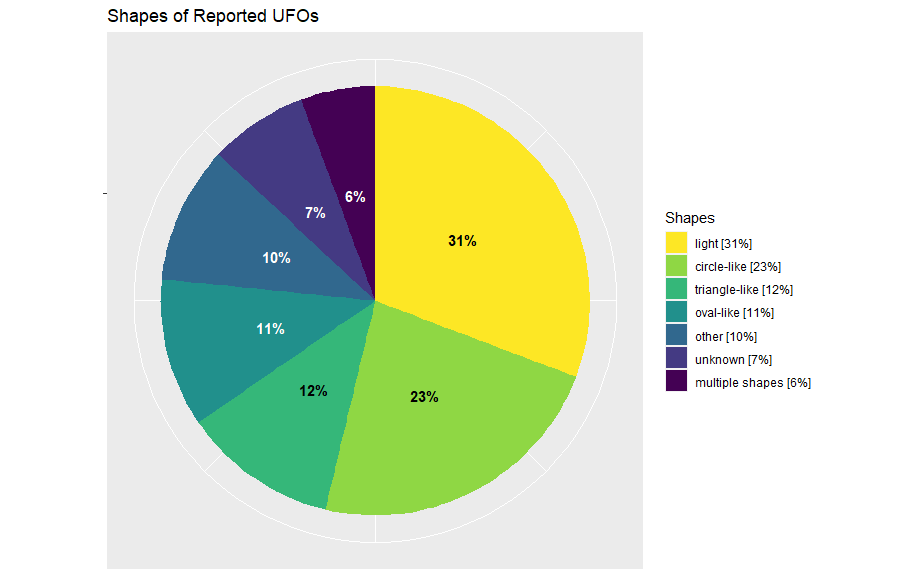
Figure 4

It’s hard to pinpoint the correlation between the time of year and the number of reports each year



**Figure 5.** UFO reports each Time of Year

The most common shape of reported UFOs is either a blink of light or some circle-like object. The rest of the reports show that people see triangles, ovals, and other shapes. Sometimes they see multiple different shapes or can’t even tell what kind of object is in front of them.



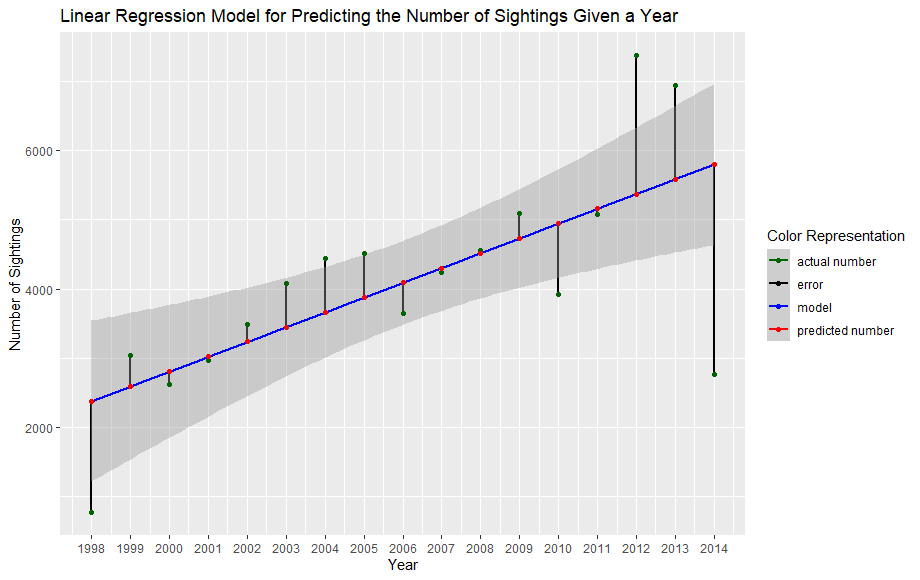
**Figure 6.** Shapes of Reported UFOs

When people describe what they saw they usually use such words as – “lights”, “moving”, “green”, “night”, “white”, “flying”, “saw”, “seen”, “red”, “large”, “ufo”.



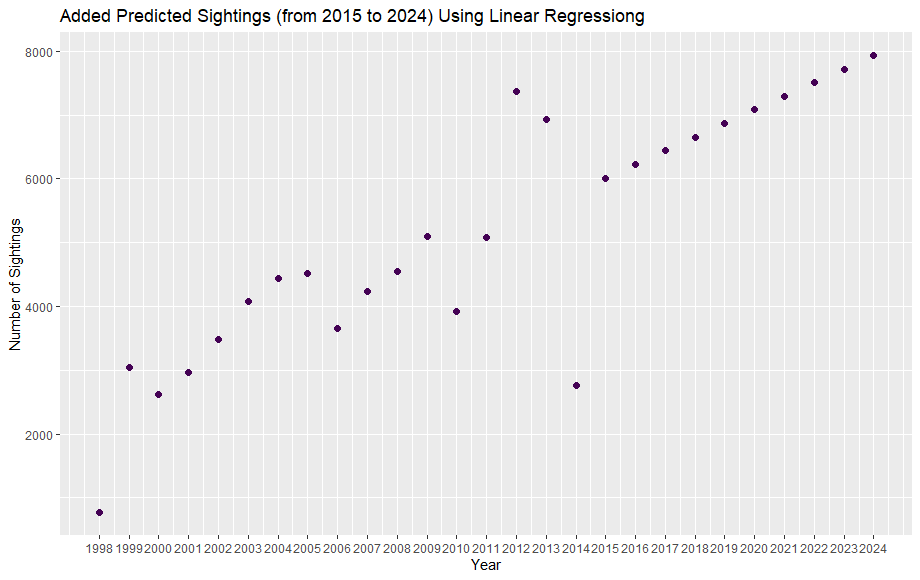
Figure

Utilizing the available data, we built a linear regression model for predicting the number of reports in the future years. The model is y = wx + b, where x is the year and y is the number of UFO reports.



Figure

After that we can try and see how the number of reports will increase each subsequent year.



Figure

**Discussion**

The part of the world that has by far the most reports of UFO sightings is The United States of America, which might be influenced by the huge territories that are still not explored, as well as by media and superstitious nature of a lot of Americans and their belief in supernatural [4].

The State of California has the biggest number of UFO reports, which can be connected to the fact that there are a lot of airports, as well as military bases [5]. On top of that California has a warm climate which allows a lot of folks to spend a lot of their time outside, even at night.

The year that has seen the most UFO reports, coincidentally, is the year of 2012 phenomenon [6]. It can be argued that the more human mind is disturbed by thoughts about the supernatural, the more chance it is to see the UFO for them.

**References**

1. https://www.kaggle.com/datasets/NUFORC/ufo-sightings
2. https://github.com/Lchiffon/wordcloud2
3. https://github.com/alexandrlagornii/UFO-Analysis
4. https://phaneuf.net/blog/americans-believe-in-ghosts#:~:text=Forty%2Done%20percent%2C%20according%20to,2009%20Pew%20Research%20Center%20survey.
5. https://www.security.org/resources/states-highest-military-burden/
6. https://www.history.com/topics/religion/december-21-2012