UFO Sighting and Statistical Correlations

Amity Ramona Mentis-Cort[1], Minh-Ngoc Huynh[2], Michael Laing[3], Michael Scalisi[1], Tomas Gifford[1]

[1]Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY, United States, [2]Department of Mathematics, Clarkson University, Potsdam, NY, United States, [3]Department of Computer Science, Clarkson University, Potsdam, NY, United States

Abstract

Humanity has always had a fascination with the unknown and that which is difficult to perceive. We as a species are uniquely creative in how we interpret what we can't understand about our reality. Through rigorous study, we have progressively demystified the unknown, which has increased our understanding of our environment. However, as we uncover more about our place in the universe, we always encounter a new boundary that tempts us to defer to the supernatural. With that in mind, across time, one concept we have always discussed and been confused about were UFOs, or unidentified flying objects. Due to the difficulty in observing these phenomena, the populace has space to interpolate and extrapolate information to support their positions on the topic. Though many have speculated about these sightings and why they occur, we decided to apply statistical analysis to the question in order to lead to new insight and guide further statistical inquisition. This paper seeks to apply linear regression to determine if there's any correlation between the number of UFO sightings and other characteristics of a country.

Keywords: UFO, extraterrestrials, statistics, GDP, per capita

1 INTRODUCTION

For decades now, UFO sightings have been making the news. Not only that, but the scientific community has been searching for extraterrestrial life since as long as people could observe stars through a telescope. From reports of other habitable planets to otherworldly radio waves, and even in the civilian reports of UFOs with crop circles and reports of objects in the sky, the interest and fascination with UFOs has been present in today's society for the past century. However, this fascination is not shared in all cultures. In order to understand why that is, this paper explores the potential correlations between GDP per capita, square kilometer of a country, and population of a country (in 2021) to the number of UFO sightings occurring in that country. Square kilometer and population were chosen due to their intuitive link to potential UFO sightings. A larger country with more people will logically have more UFO sightings. As this is an observational study, the data needed to examine these correlations was sourced from Kaggle, Our World In Data, and the US Census Bureau. A positive correlation for both square kilometer and populations to the number of UFO sightings is expected to be found. GDP per capita and its correlation with the number of UFO sightings in a country is the main question of this paper. The most UFO sightings have occurred in the United States, and the second most has occurred in Canada. These are two wealthy nations with high GDP per capita. This paper attempts to find a positive correlation between any country with a high GDP and a high number of UFO sightings. Ultimately, the biggest question is how Americans manage to spot so many more aliens than any other country.

2 METHODOLOGY

2.1 Selection of Data Sources

When collecting data, we initially had to make decisions about the exact points we would use for multivariate analysis. To approach the problem, we sourced Kaggle for datasets related to UFO sightings and found sightings in the past century across the globe. We also used Our World In Data to get the world's gross domestic product per capita and world's area by square kilometer. Additionally, we explored population data from United States Census data for world population.

2.2 Description of Data

The data sets we gathered were from Kaggle, Our World In Data, and the US Census Bureau International Database. Kaggle provided us with the UFO sightings data from NUFORC. Our World In Data provided us with the Land Area in Square Kilometers and GDP per Capita. The data from the United States Census International database was not used in our final calculations, but was explored for further analysis of factors that may influence more or less UFO sightings. When cleaning our data for analysis, we were aiming to have all of our data be around, if not exactly 2021, and evaluate the importance of matching the year on a case by case basis. For instance, land area in square kilometers was based on data from 2017 in our analysis, but due to the nature of the data (land area likely not changing significantly from 2017 to 2021) we elected to continue to use the data for our analysis.

2.3 Analytical Approach

To understand if there is a direct relationship between a country's GDP per capita and UFO sightings, a linear regression model will be used on the UFO data and the GDP per capita. Then, we will observe the coefficient of determination to see if the model is a good fit. If there are any outliers, we will remove them and make a new linear model without those countries.

First, with the GDP per capita dataset and sorted the year to 2021. We did the same thing with the population data to find the world population in 2021. In R Studio, we used the *filter* command to do so.

Then, we merged the population data and GDP per capita data and UFO sightings by country to get each country's total sightings, population, and GDP per capita for 2021. With the dataset, we did a linear model with the command *lm* on sightings by GDP per capita. Then, we plotted a point plot with the linear model on top to see the fit. After that, we would remove large outliers (countries with above 500 sightings) to make another linear model. Additionally, we made another linear regression model for countries with under 150 sightings. Finally, to see how often a country has a certain amount of sightings, we made a histogram with *ggplot* and *geom_hist* with a line to show the mean sightings per country with *geom_vline*. We made 2 histogram graphs for countries with under 500 sightings and a separate histogram for countries with under 150 sightings.

2.4 Coding Language and Library Tools

All the analysis was run in RStudio with R code in a R script and a R notebook file. The script was used to generate linear models using the lm command and ggplot to analyze the correlation between UFO sightings and gdp per capita of several countries. The notebook file includes an interactive plot with sightings per country.

The notebook file uses the *tidyverse*, *janitor*, *sf*, *tmap*, and *here* packages. *Tidyverse* and *janitor* are for cleaning the data and making new columns for what we wanted to do. The *here* library was used to load the datasets into R Studio. *Sf* and *tmap* were used for geographical data for the interactive map as part of the *plotly* package. We also use *dplyr* functions to *rename*, *count*, *filter*, and *merge*. Additionally, we used *ggplot2* to plot the initial graph of sightings by longitude and latitude.

The R script uses *tidyverse*, *janitor*, and *here*. Otherwise, it also includes *ggplot2* for plotting points and also a linear model on top of the points. We used *geom_point* to plot the sightings and *geom_smooth* to make the linear model on the same plot, additionally, we used *geom_hist* to make a histogram of the quantity of sightings by country and *geom_vline* to plot the mean of the histogram. We also used *gridExtra* and *grid.arrange* to plot many graphs in one image to compare them.

3 RESULTS

Our initial analysis revealed that the United States, Canada, United Kingdom, and Australia had significantly higher sightings compared to other countries. In the interactive map, those countries are highlighted in blue and darker blue, as shown in **Figure 0**. Specifically, the United States had the most sightings at 70878 sightings and the UK had the most sightings per land area with about 0.01 sightings per square kilometer.

There was not a strong correlation with the amount of sightings and GDP per capita, with a coefficient of determination of less than 2% for all linear models. The model for sightings by GDP per capita for countries with under 150 sightings had the highest coefficient of determination, as shown in **Figure 4.2**.

This section includes linear regression models for all countries' UFO sightings by GDP per capita (**Figures 1**), all countries' except the United States' sightings by GDP per capita (**Figures 2**), countries with under 500 sightings' UFO sightings by GDP per capita (**Figures 3**), and countries with under 150 sightings' UFO sightings by GDP per capita (**Figures 4**). Additionally, we have a histogram of how many countries had a certain amount of sightings for countries with under 500 sightings in **Figure 5** with amount of sightings by total countries with that amount of sightings and a similar histogram for countries with under 150 sightings (**Figure 6**). On these graphs is a blue line which is the mean amount of sightings a country has.

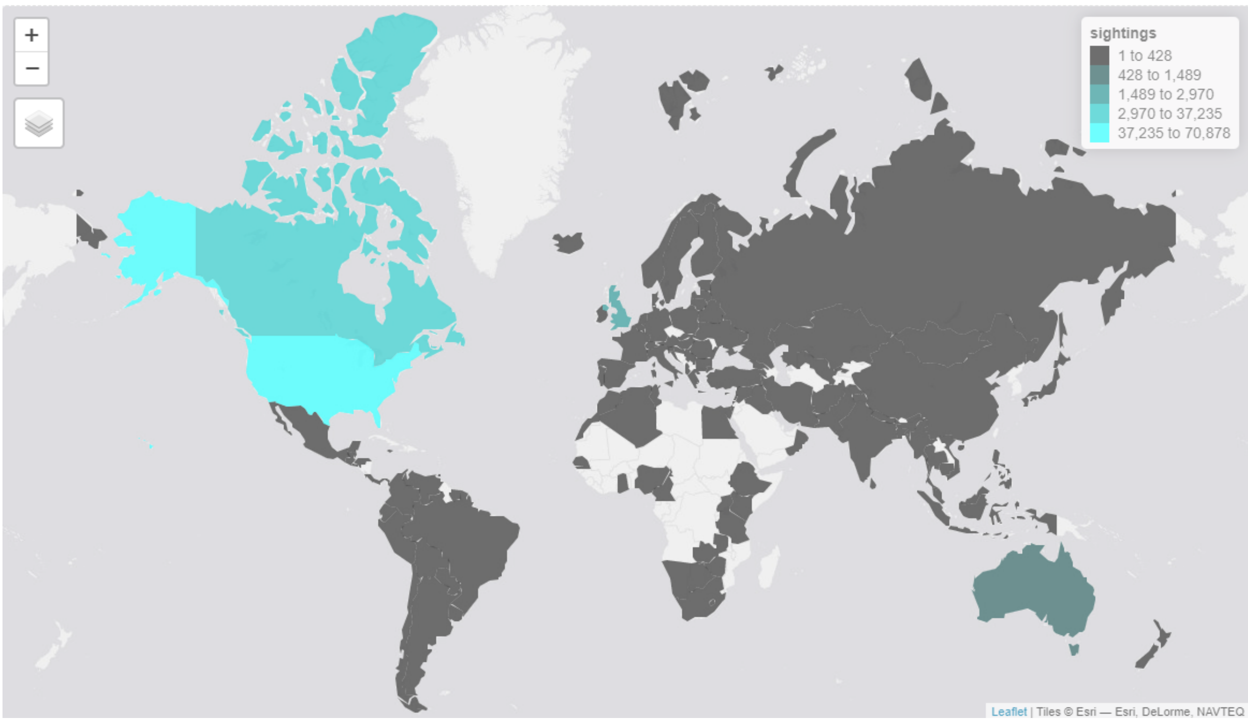

Figure 0: Sightings per country with a global map and scale in blue.

Linear model with US Sightings	Estimate	Std. Error
Intercept	-443.73210	787.27102
GDP per capita's coefficients	0.04003	0.02279

Figure 1.1: Linear model of all UFO sightings per GDP per capita with coefficients and standard error

Multiple R-squared	**Adjusted R-squared**
0.02268	**0.01533**

Figure 1.2: The linear model's coefficient of determination

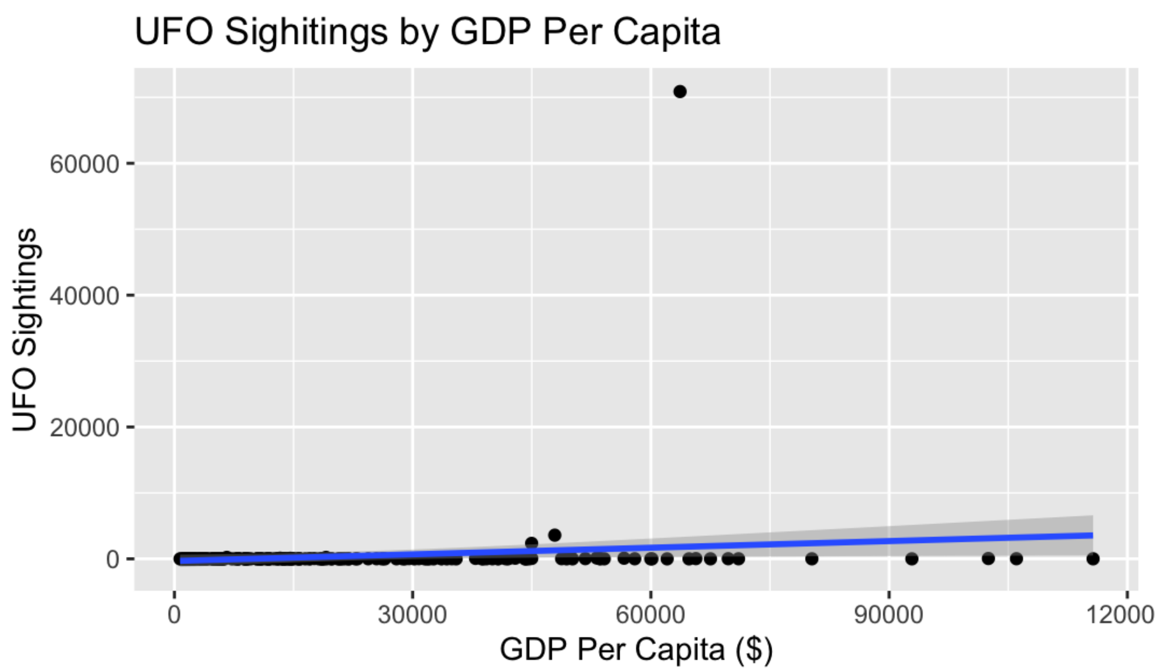

Figure 1.3: Point plot of all UFO sightings per GDP per capita and linear regression

Linear model without US sightings	Estimate	Std. Error
Intercept	10.834975	48.110720
GDP per capita's coefficients	0.002232	0.001405

Figure 2.1: Linear model of all UFO sightings except United States per GDP per capita with coefficients and standard error

Multiple R-squared	**Adjusted R-squared**
0.01875	**0.01132**

Figure 2.2: The linear model's coefficient of determination

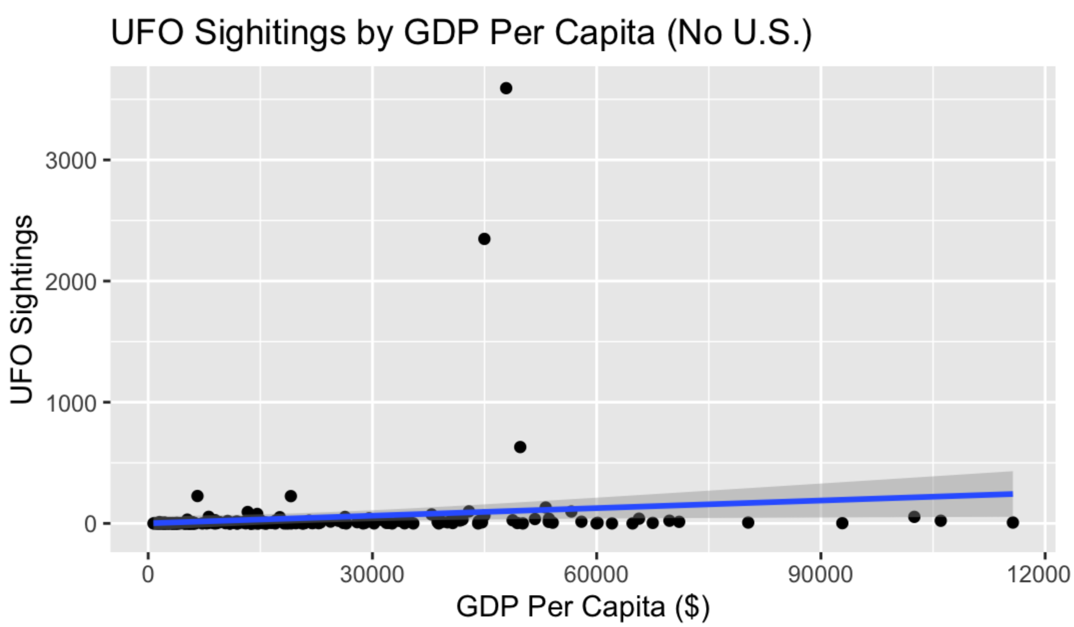

Figure 2.3: Point plot of all UFO sightings (without the United States' sightings) per GDP per capita and linear regression

Linear model with under 500 Sightings	Estimate	Std. Error
Intercept	6.759	2.811
GDP per capita's coefficients	2.863e-04	9.442e-05

Figure 3.1: Linear model of all UFO sightings of countries with under 500 sightings per GDP per capita with coefficients and standard error

Multiple R-squared	Adjusted R-squared
0.04437	**0.03954**

Figure 3.2: The linear model's coefficient of determination

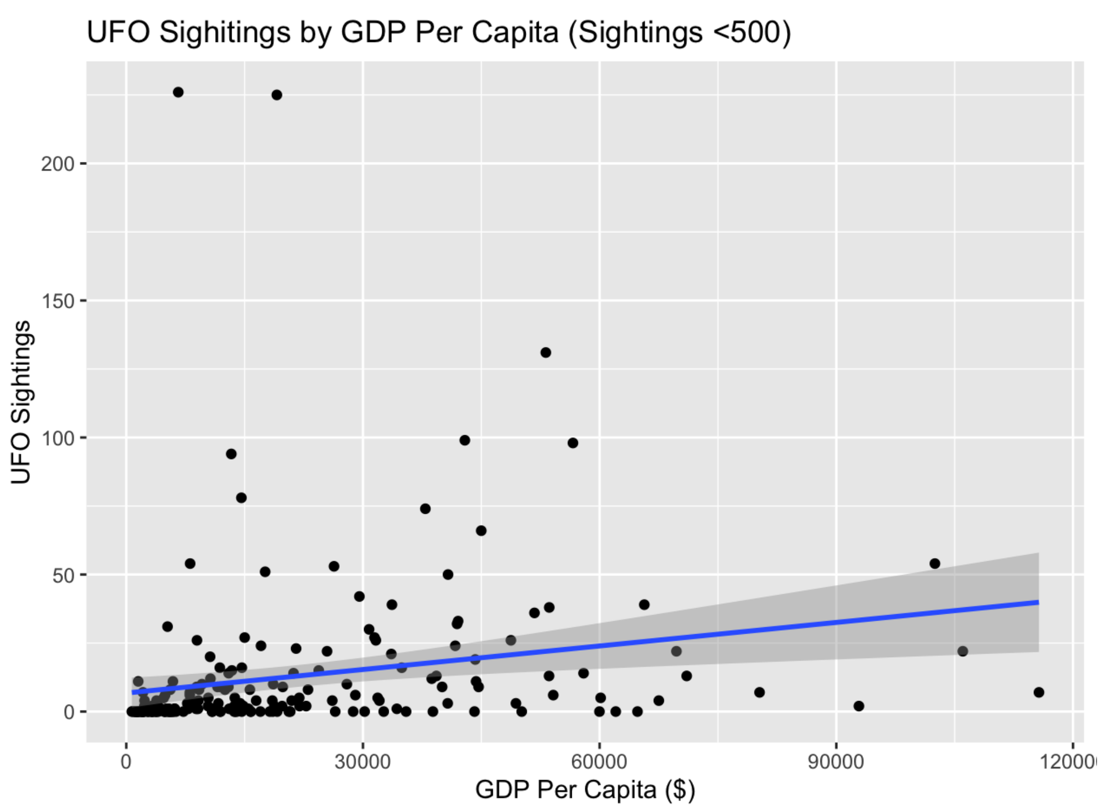

Figure 3.3: Point plot of all UFO sightings (for countries with under 500 sightings) per GDP per capita and linear regression

Linear model with under 150 Sightings	Estimate	Std. Error
Intercept	3.799	1.840
GDP per capita’s coefficients	3.241e-04	6.157e-05

Figure 4.1: Linear model of all UFO sightings of countries with under 150 sightings per GDP per capita with coefficients and standard error

Multiple R-squared	Adjusted R-squared
0.1239	**0.1194**

Figure 4.2: The linear model’s coefficient of determination

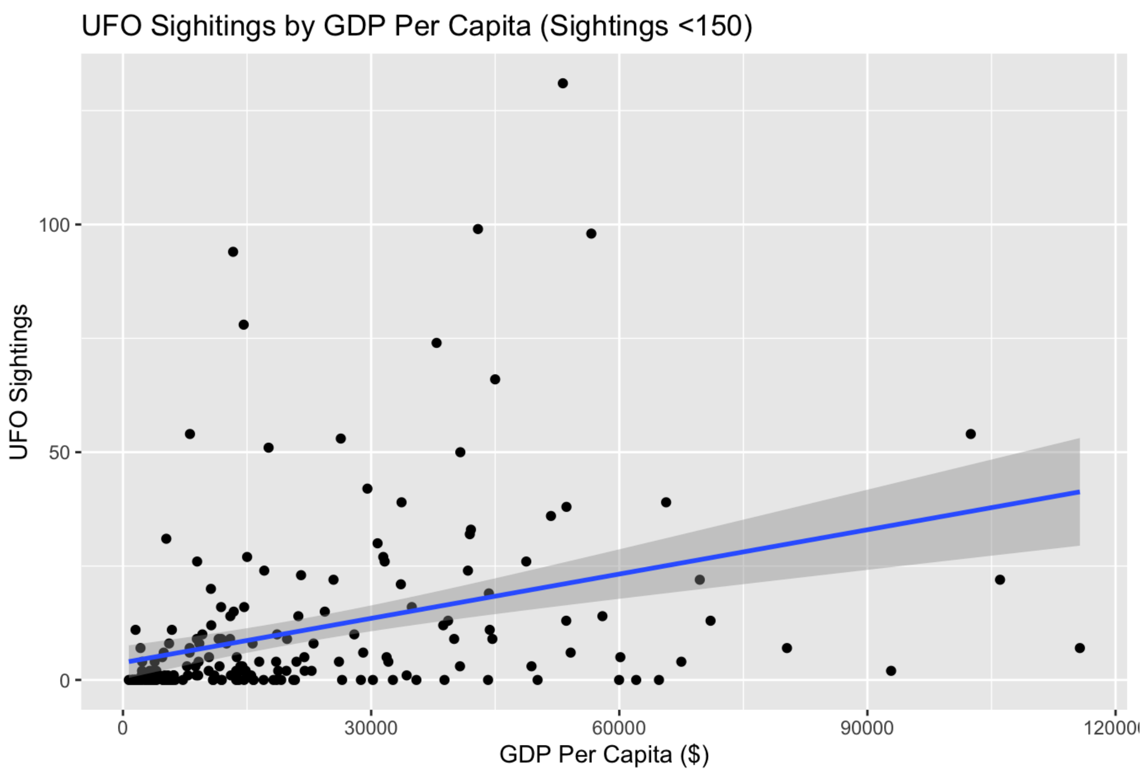

Figure 4.3: Point plot of all UFO sightings (for countries with under 150 sightings) per GDP per capita and linear regression

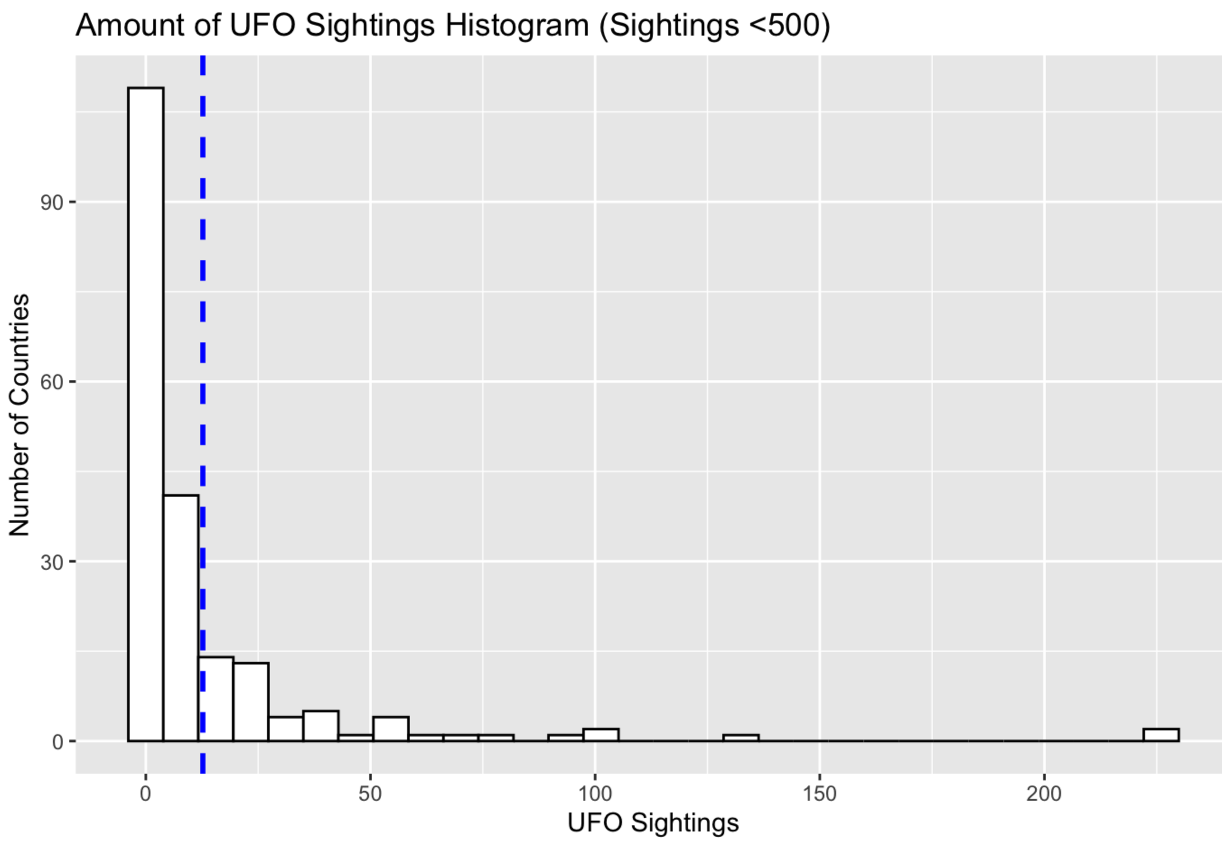

Figure 5: Histogram of how many countries have how many sightings and the mean of all countries' sightings for countries with less than 500 sightings

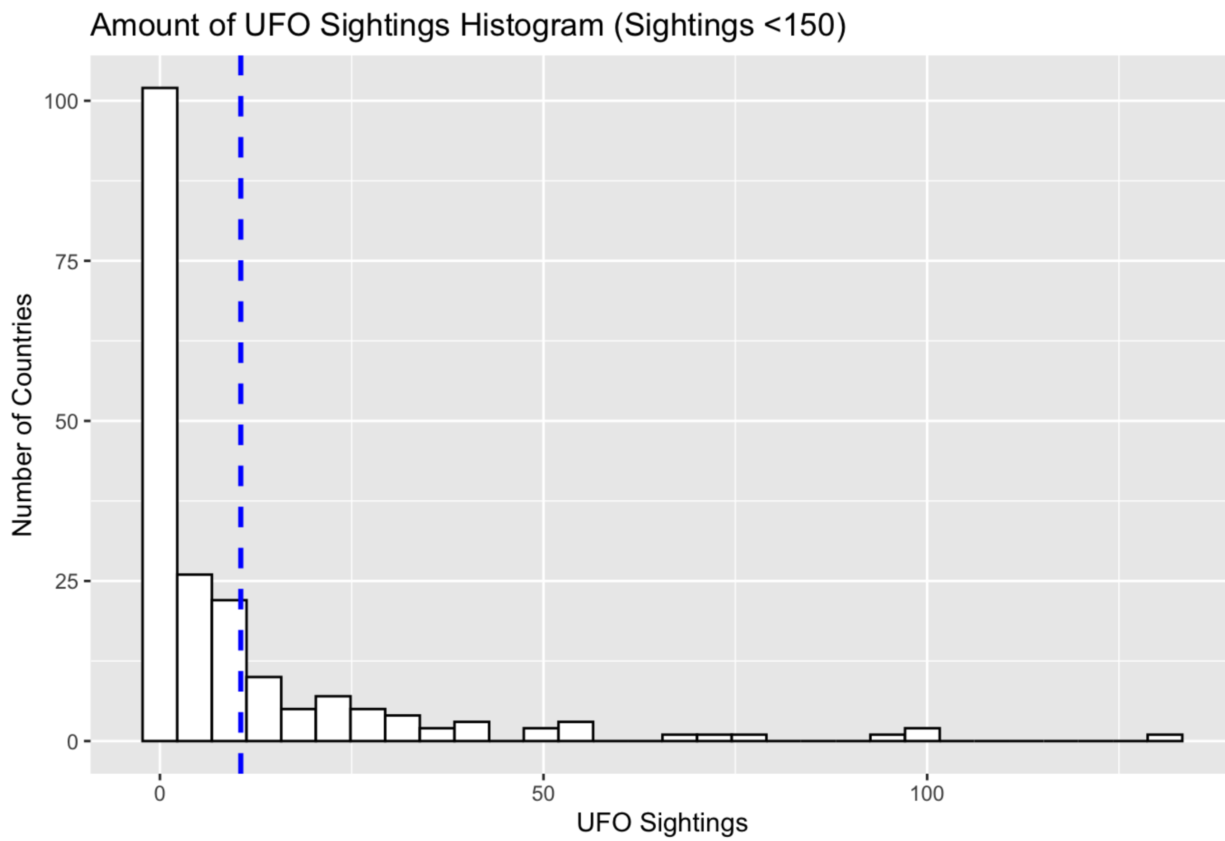

Figure 6: Histogram of how many countries have how many sightings and the mean of all countries' sightings for countries with less than 150 sightings

4 DISCUSSION

R^2 is the value for a linear regression model that determines the accuracy of that model, known as the coefficient of determination. An R^2 value closer to one means a more accurate model. An R^2 value of about 0.9 can be interpreted as an accurate model. The adjusted R^2 values that we obtained for the GDP per capita testing were 0.01533 and 0.01132 for the tests with (**Figure 1.2**) and without the United States (**Figure 2.2**), respectively. Since these R^2 values are much less than 1, we can conclude that there is no statistical correlation between the number of UFO sightings in a country and its GDP per capita. However, it is interesting to make note of how drastic certain countries tend to skew the data to give us our findings.

When interpreting the graphs, we see that when we include the United States in our dataset (**Figure 1.3**), the axis of sightings multiplies from a scale of thousands (**Figure 2.3**) to tens of thousands. The United States alone has around 71000 sightings, which skews our graph to a remarkable degree. We decided that if the U.S. is such a major outlier, it would be of interest to exclude it from our data for a separate examination. This would allow us to see if the exceptional nature of the U.S. was preventing us from detecting any trends or correlations.

In **Figure 0**, we see a heatmap of where sightings are most often reported. In most countries, the number of sightings is under 500 except for four countries that serve to be major values to take into consideration.

Taking this observation into account, the next highest number of sightings after the United States was from Canada, with around 4000 sightings. When looking further at the data, we find very few countries that have a significant amount of UFO sightings.

Country	Sightings
United States	70878
Canada	3592
United Kingdom	2348
Australia	630

It is also of note the specific countries that have noted large numbers of sightings. We were expecting that a higher GDP per capita would imply that there would be more flexibility for citizens of the country to dedicate time to creative endeavors, such as the arts or media. With the proliferation of these new works of art in media, some will naturally trend towards the supernatural and thus could spark interest in the cultural zeitgeist for a certain country and thus, prompt more UFO sightings. However, the data does not suggest that there is a strong correlation between these two data points, at least when not controlling for the outliers. Taking a look at the countries with the highest GDP per capita, we see that there is little to no correlation between a country with a high GDP per capita and the frequency of UFO sightings.

GDP Per Capita Ranking	Country	Sightings
1	Luxembourg	7
2	Singapore	22
3	Ireland	54
4	Qatar	2

We controlled for the major sighting outliers by removing them from the sample data. After this removal (**Figure 3.3**), we find that the number of sightings drops drastically, where the highest sightings are around 200. Out of curiosity, we did the same thing again, but for sightings less than 150.

Looking at the R^2 values for both sightings under 500 (**Figure 3.2**) and sightings under 150 (**Figure 4.2**), we see an R^2 value of 0.03954 and 0.1194 respectively. Compared to the R^2 values we see when considering all the data points with the major outliers, these R^2 values are closer to suggesting a better fit to the data than when the outlier values are included. Especially when noting how the values under 150 sightings have an R^2 value of around 0.1 and originally with all data points included had an R^2 value of about 0.01, we see that the fitness of the data has increased by an order of magnitude. This highlights how drastically the outliers impact the data even though the amount of outliers was so few.

When viewing the histograms for the number of countries and the amount of sightings, for both under 500 sightings (**Figure 5**) and under 150 sightings (**Figure 6**), the mean (indicated by the blue line) is close to zero, suggesting that the majority of the data is closer to zero than what the outliers would suggest. However, without removing these outliers, the mean would have skewed significantly towards the larger sightings and would have made the histogram mean less valuable for interpreting and confirming the information presented regarding the heatmap in (**Figure 0**).

4.1 Practical Implications and Future Work

The results of our analysis suggest that it is unlikely that there is a correlation between a country's GDP per capita and the frequency of UFO sightings within that country. We made an assumption that with a higher GDP per capita, countries might be able to invest more into culture, the arts, and media (Statista), which could lead to more superstition about what lies beyond the stars. However, we discovered that while GDP per capita is not the selection we needed to measure against to find a correlation, there is still cause for investigation. Future work could entail a sociological study that investigates the top four countries with sightings. Although it cannot be readily identified by the researchers, further research may reveal another marker of analysis that can be used to discover what distinguished the US, Canada, UK, and Australia so drastically from other countries with regard to UFO sightings.

4.2 Limitations

This study is understandably limited as due to data being collected from around the world, language barriers and sheer quantity make it possible for UFO sightings to slip through the cracks. As these are human reports (Artangel), there is a report bias within the UFO data. Furthermore, UFOs can just be other flying objects made from our world that are not extra terrestrial. The interpretation of UFOs can vary depending on the context of the report.
Additionally, GDP and population is an estimate and therefore has some innate variance. Censuses cannot be perfectly accurate due to reporting bias.
Currently, with countries' land area, there are some warring nations that are having area disputes. Therefore, the area data can have errors. Also, the measurement of area can have errors in itself.

5 CONCLUSION

The purpose of this paper is to determine whether the population, square kilometer, and GDP per capita of countries impacted how many UFO sightings were reported from them. For the overall area correlation, we, surprisingly, didn't find enough evidence to determine it had a significant impact on UFO sightings. Whether or not

the United States was included in the data did have a relatively large impact on the R-value for area, but it was ultimately at too small a scale.
Based on the data gathered, and the R-squared values that followed from it, it cannot be concluded that there is a correlation between a country's GDP per capita and the amount of UFO sightings within its borders. This raises questions as far as the United States is concerned.
After conducting this research, it has become clear that the United States is a major outlier compared to the rest of the world. It is also worth noting that the countries immediately following the United States in terms of sightings were Canada, the United Kingdom, and Australia, all of which speak English as a primary language. As such, it could be worth looking into a possible correlation between English-speaking rate and UFO sightings in further studies. Although, even among English-speaking countries, The US has several magnitudes more sightings than its peers.

6 DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article and the code is made available by the authors, without undue reservation, in the repository at https://github.com/JonathanDoety/UFO-Sightings-Analysis-STAT383-.

7 AUTHOR CONTRIBUTIONS

ARMC: Writing—Abstract, 2.2: Description of Data, 4.Discussion (minus 1st paragraph and 4.2: Limitations), coding, data curation, editing, data visualization, formatting, formal analysis, statistical analysis, validation, review.

MNH: Writing— Methodology (minus 2.2: description of data), 3.Results, 4.2: Limitations, coding, statistical analysis, data organization, data visualization, conceptualization of linear models, validation, review.

ML: Writing—1.Introduction (Lit Review), 4.2 Limitations, 5.Conclusion, investigation, data curation, formal analysis, review.

MS: Writing—4.Discussion (first paragraph), Abstract, 1.Introduction, formatting, data curation, investigation, conceptualization, analysis, review.

TG: Coding, data curation, investigation, conceptualization, review.

All authors contributed to the article and approved the submitted version.

8 CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

9 REFERENCES

(2023) National UFO Reporting Center. In: NUFORC. https://nuforc.org/. Accessed 8 Dec 2023.

GDP per capita. In: Our World in Data. https://ourworldindata.org/grapher/gdp-per-capita-worldbank. Accessed 8 Dec 2023.

Gibin WO (2023) UFO sightings. In: Kaggle. https://www.kaggle.com/datasets/willianoliveiragibin/ufo-sightings. Accessed 8 Dec 2023.

In: Artangel. https://www.artangel.org.uk/witness/ufo-sightings-from-around-world/. Accessed 8 Dec 2023.

International database. In: United States Census Bureau. https://www.census.gov/data-tools/demo/idb/#/table?COUNTRY_YEAR=2023&COUNTRY_YR_ANIM=2023&menu=tableViz&quickReports=OVW&CCODE_SINGLE=**&TABLE_YEARS=2021&TABLE_USE_RANGE=N&TABLE_USE_YEARS=Y&TABLE_STEP=1&TABLE_ADD_YEARS=2021. Accessed 8 Dec 2023.

Land area in square kilometers. In: Our World in Data. https://ourworldindata.org/grapher/land-area-km . Accessed 8 Dec 2023.

Visé, D. de. (2023, August 2). *The truth is out there: More Americans believe in ufos*. The Hill. https://thehill.com/homenews/space/4131768-the-truth-is-out-there-more-americans-believe-in-ufos/. Accessed 8 Dec 2023.

Statista Research Department, 6 S (2023) Global art market share by country 2022. In: Statista. https://www.statista.com/statistics/885531/global-art-market-share-by-country/. Accessed 8 Dec 2023