U.C. Berkeley School of Information

W200: Introduction to Data Science Programming

Professor: Gerald Benoit

PROJECT TWO

THE RELATIONSHIP BETWEEN UFO SIGHTINGS, MILITARY BASES AND POPULATION IN THE UNITED STATES

Researchers: Maria DiMedio, Ryan Knorr, Peter Morgan, Kaidong (Kai) Ying

Introduction

Unidentified Flying Objects (UFO) are not new to American history or culture. Humans have described UFOs for millennia, often describing the objects by their various shapes or bright colors in the sky. Unusual flying objects can even be traced back to the art and literature of ancient civilizations like the Greeks, Romans and Egyptians. The term "UFO" was partially devised in 1953[1] by the United States Air Force (USAF) to assist with the labeling of reports on flying objects. Documented sightings of UFOs began to increase in the 20th century around the time that high-tech aircraft began development in the United States.

In 1947, the USAF began a series of studies that investigated approximately 12,618 claimed UFO sightings before it officially ended in 1969[2]. The project operated under the alias "Project Blue Book", which was originally classified at a Top Secret level, which has since been declassified. The project concluded the majority of the 12,618 sightings could be attributed to shooting stars, clouds, ordinary aircraft, or reconnaissance planes, although 700 plus sightings remain unexplained.

In 2017, the New York Times published multiple reports detailing a Pentagon confirmed Intelligence Community project[3] called the Advanced Aviation Threat Identification Program (AATIP) tasked with investigating UFO's, similar in scope to Project Blue Book. The program is said to have reviewed UFO sightings, assessed threats, and collected metal alloys recovered from unidentified aerial phenomena.

While the historical context of UFO sightings in American History is extremely intriguing, our project seeks to investigate the relationship between UFO sightings, military bases and population in the United States throughout the 21st century.

Research Questions

Main Question:

1. Is there a relationship between UFO sightings, population and military bases?

Sub Questions:

1. What effect does proximity to a military base have on reported UFO sightings.
2. What are the trends in UFO sightings in the US between 2000 and 2019?
    a. What are the most common descriptive words used?
3. What are the trends in military bases in the US?
4. Are the most populated states the ones with the highest number of sightings reported?
5. Are there trends in the types of sightings corresponding to population density?
6. What are the highest proportions of sightings per capita?
    a. Do those locations have any military bases nearby?
7. Do larger bases have a higher number of sightings near them?
    a. Do certain military bases have a higher number of sightings near them?
8. Are certain types of UFOs commonly reported near certain types of military bases?

Data Sources

I. Primary dataset:
    A. UFO Sighting Reports from the National UFO Reporting Center (the "UFO Data") - Source: http://www.nuforc.org
II. Secondary datasets:
    A. Military Bases from the US Department of Transportation (the "Military Base Data") - Source: link
    B. Census *[Population Estimates Ranging: 2000-2019] (the "Census Data")* - Source: Census.gov 2000-2010, Census.gov 2010-2019

[1] What Is The Origin Of The Term "UFO"? (2019)
[2] Project BLUE BOOK - Unidentified Flying Objects. (2020)
[3] Cooper, H., Blumenthal, R., & Kean, L. (2021, January 21)

The primary dataset used in this report is data from the National UFO Reporting Center (NUFORC). This dataset will contain the UFO sightings information, location data, time-series data and sighting descriptions. A secondary dataset that will be used in this report is a military base JSON file that provides information on all active and inactive military bases by service and location in 2019. Another secondary dataset that will be used is census data containing population estimates from 2000 - 2019. The Census data will allow the researchers to measure against the other data per-capita. The tables below will show the variable names in each raw dataset.

Dataset Structure

1. The UFO Data:

Date/Time (Datetime - The time of the sighting, ex. “1/15/20 12:30”)	Duration (String - A description of the duration of sighting, ex. “1 minute”)
City (String - The city where the sighting is, ex. “Waterloo”)	Summary (String - A summary of the sighting, ex. “Hovering Black Triangle”)
State (String - Abbreviated State name, ex. “GA”)	Posted (Datetime - The date the sighting was reported, ex. “10/19/12”)
Shape (String - The reported shape of the UFO, ex. “Triangle”)	

The shape of the UFO data is 90335 observations and 7 variables.

2. The Military Base Data: JSON file

datasetid (String - this is “military-bases” for all entries)	recordid (String - A unique string ID for all entries)
fields (Dictionary - Values contain data described by the keys. See table below for more info.)	geometry(Dictionary - type key represents the type of coordinate ex. “Point”, coordinates key represents the lat/long coordinates)
record_timestamp(String - A timestamp of when the data was pulled)	

The shape of the raw military data is 776 observations of 5 variables. Most of the useful data is contained in the “fields” variable which is a dictionary. The table below outlines some of the keys of this dictionary that we will use and what the values mean.

oper_stat (String - Operating status, ex. “Active”)	state_terr (String - State or territory base is located, ex. “Hawaii”)
site_name (String - Name of the military base, ex. “Pohakuloa Training Area”)	component (String - Type of military base, ex. “Army Active”)

3. The Census Data:

For the census data, variables were dropped while reading them in. The variables that were read in are listed below. The shape of this data is 1020 observations and 4 variables.

state (String - ex. “Alabama”)	state_abbrev (String - ex. “AL”)
year (Int - ex. 2000)	population_estimate (String - ex. “4,452,173”)

Data Preparation

The UFO data is originally reported as a series of web tables. The tables are scraped from the website and are then combined into a single dataframe. The original UFO data includes data reported from outside of the U.S. and contains reporting error in some of the observations in the state column. These records are removed for the purpose of this report.

The Census data exported into a time series or wide data format. This was inconsistent with the long format of the other datasets the researchers were using. In order to reformat the data the researchers utilized the Pandas ‘wide_to_long’ to restructure the data.This step enabled the join of this secondary dataset to the primary dataset. A few rows and columns were removed from the original dataset, such as population estimates broken down to broad geographic locations like ‘east’ and ‘west’..

The original format of the Military Base dataset was a json file, which when first imported was formatted in 5 column variables. One of these variables stored a dictionary of data fields and values, which had to be pulled out and assigned to unique columns of their own. Each observation in this dataset represented a military base in the US, and all observations were included in this analysis. A few variables were removed from the original dataset, such as the base record ID, record entry timestamp, and dataset ID because they were not useful to our analysis

The UFO data is used as the primary dataset and needed the least amount of restructuring since the researcher’s looked to join the secondary data to the primary**.** To join the three datasets together, the researchers examined each dataset to find a common data piece to merge on. All three datasets contain a ‘State’ or ‘State Abbreviation’ variable, therefore the researchers decided to use a State Abbreviation variable in each dataset to merge. For the military base and census datasets, this required transforming the State variable to a new ‘State Abbreviation’ variable, instead of the full state name. The State Abbreviation variable was added to datasets by creating a state dictionary that contained the full state name as a key and the abbreviation as a value.

Analysis

The main goal of this study, as stated above, was to explore the relationship between military bases and UFO sightings in the United States since 2000. Our approach to this analysis was to first understand the descriptive statistics on UFO sightings and military bases by state. The research team chose to group all data observations and aggregate statistics by state, in order to best understand the trends across the country. Our main variables of interest were related to descriptions of UFO sightings, frequency of sightings, types of military bases, frequency and size of bases, population sizes, time, and state locations. First, we sought to understand the common trends in UFO sightings in the country overall, then by state, and then over time.

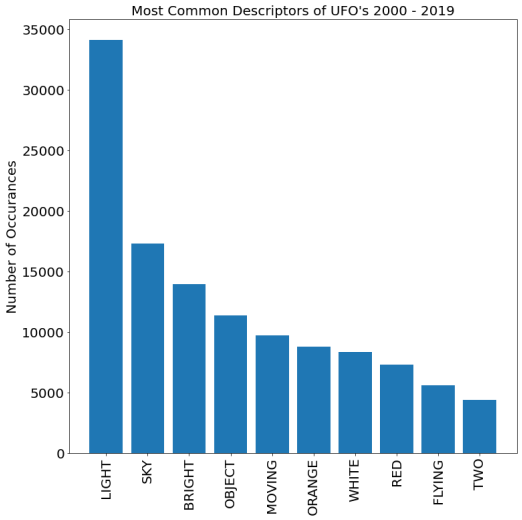
Next, the research team analyzed the trends in military bases across the country, also grouping this data by state for future comparison against UFO sightings. Here the team was primarily interested in uncovering aggregate statistics on each state, including total overall bases, base types, area sizes, and activity. This allowed us to pinpoint where we should concentrate our deeper analysis into relationships between said base statistics and UFO sighting statistics.

After having established a baseline of aggregate statistics across all US states for UFO sightings and military bases, the research team accounted for population disparities across states by normalizing our data against the census data and analyzing the trends in UFOs and military bases per capita in each state.

Finally, the research team analyzed the relationship between UFO sightings and military base statistics in each grouped state.This allowed us to assess if there are any trends in sightings corresponding to aggregate military base statistics, in respect to our research question.

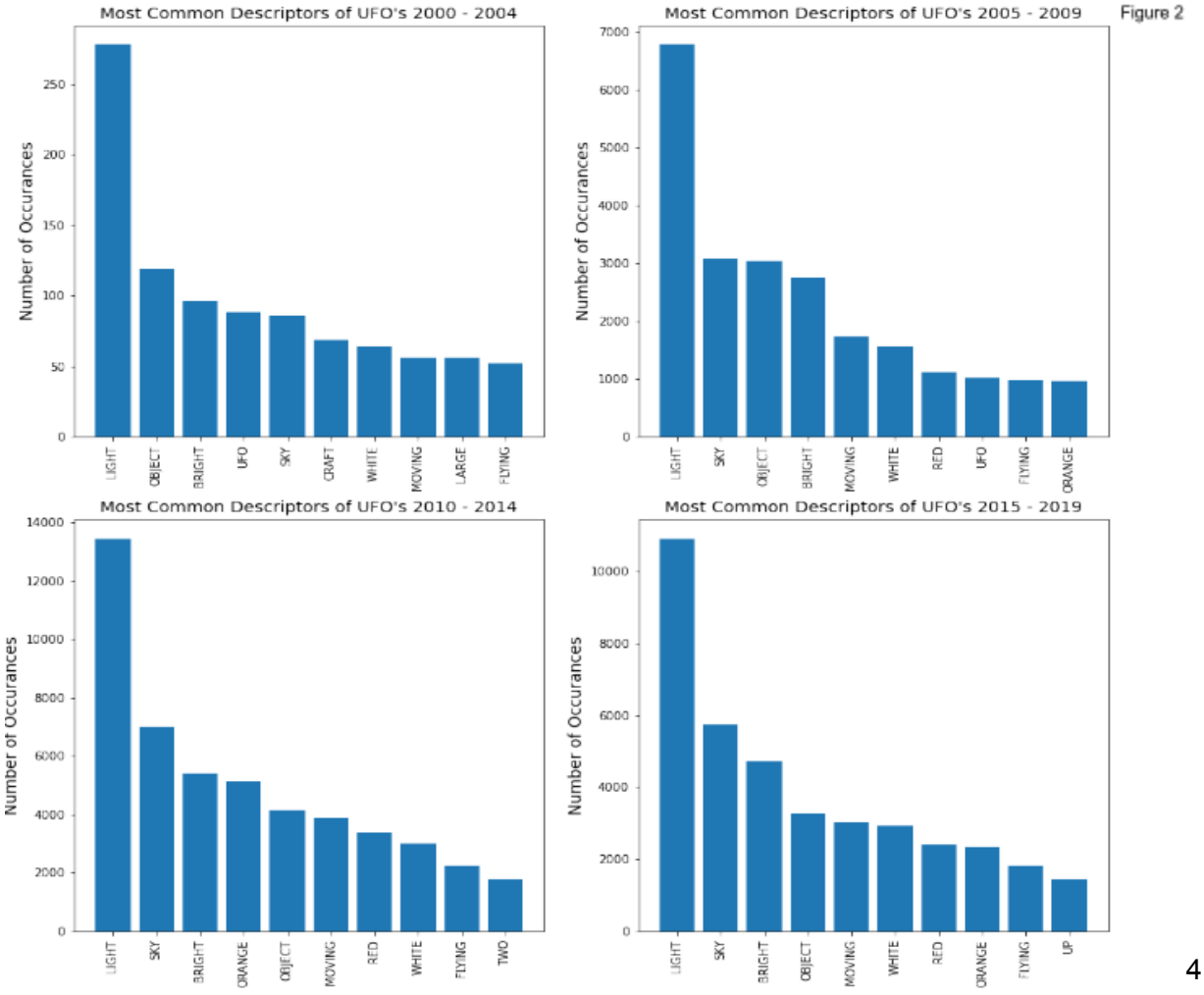
The following report sections include our findings in each stage as mentioned above, with a discussion of each insight. To begin exploring the UFO dataset, we will first examine the most common descriptors of UFO's. This is done by examining the summary variable and making a barplot of the words that appear the most frequently. Some assumptions were made in this. The first assumption was combining common words. For example, "light" and "lights" were both treated the same. The second assumption is to omit common words. These include "a" "the" "is" and so on. The final assumption is to omit words that do not add anything to the description like "saw" and "seen". Figure 1 shows the 10 most common words used in descriptions of UFO sightings from 2000-2019.

Figure 1



As you can see in Figure 1, light is by far the most common descriptor. This is followed by "sky", "bright", "object" and "moving". Following this are three color descriptions and then flying. Perhaps most interesting to me is the final descriptor of two. This implies that there were multiple UFO's from a single sighting. Next, we will take a look at how these descriptors change over time and any insight we can gain from that. The assumptions are the same as in Figure 1, except the data will be divided into four time periods, 2000-2004, 2005-2009, 2010-2014, and 2015-2019. This is shown in Figure 2 below.

Figure 2



The first thing that stands out about Figure 2 is that light is consistently the most common descriptor among all time periods. Another interesting observation is that there are less descriptions during the earliest time period of 2000-2004. All time periods seem to follow similar patterns. Light is always the most common, followed by a similar 3 or 4 descriptors in different orders, followed by a few colors, and once again followed by a few other descriptors. Next, we will investigate a potential cause of why there were less descriptors in the time period of 2000-2004. This could be because of a difference in total UFO sightings during this time period. To determine this, Figure 3 shows the total number of UFO sightings in the United States by year.

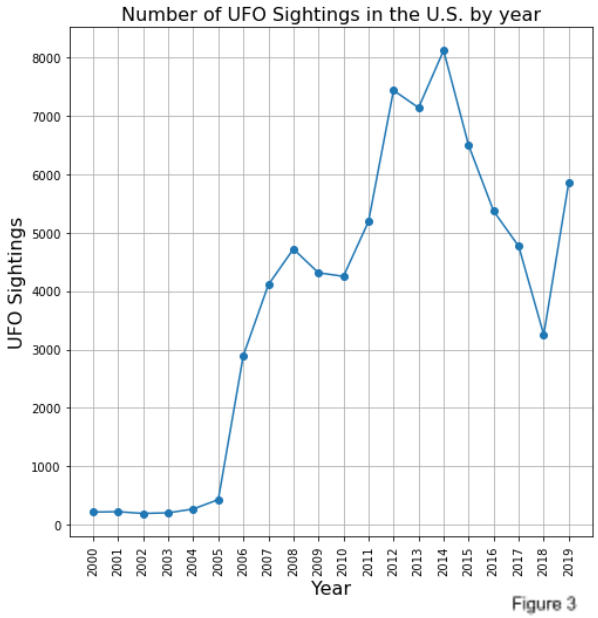


Figure 3

As you can see from Figure 3,there are much fewer UFO sightings in the United States from years 2000-2005 than any of the following years. This could explain why there are less UFO descriptions between the years of 2000-2004. Another interesting finding from Figure 3 is that the peak in UFO sightings in the United States is in 2014 with just over 80,000. There is a consistent drop in sightings from 2014 to 2018 which might be cause for further investigation down the road.

In order to understand which states might have the highest frequency of military base locations for further analysis, Figure 4 shows a map of the United States, colored by a scale of total number of bases.

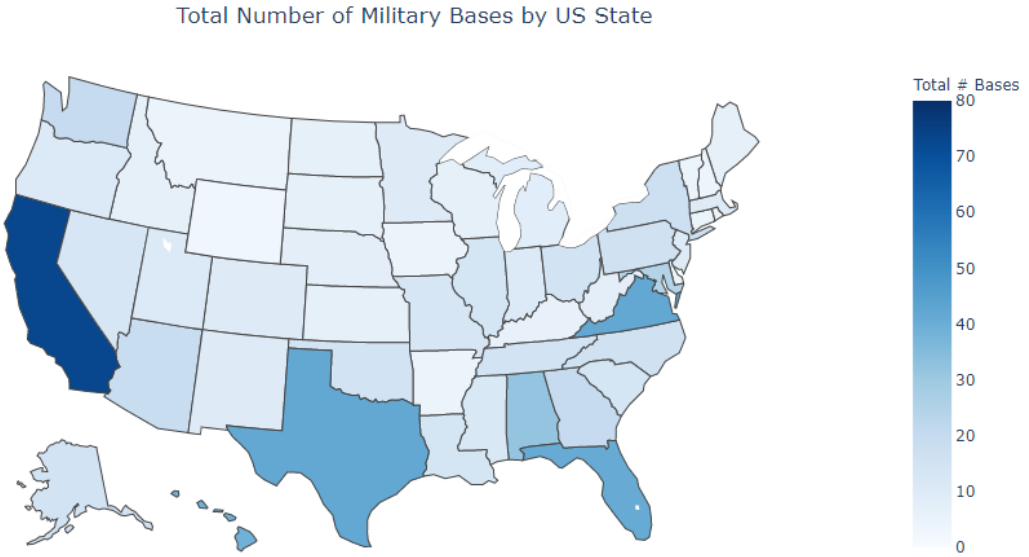


Figure 4

Figure 5 below shows the 20 states with the most total bases. California had the highest total number of bases, with 73. California is somewhat an outlier in our dataset, with almost twice as many total bases than any other state. A second insight which can be drawn from this plot is a breakdown of the most commonly occurring type of base located in each state. Overall, the most common type of base is active Naval Bases, which is followed by active Army and Air Force bases.

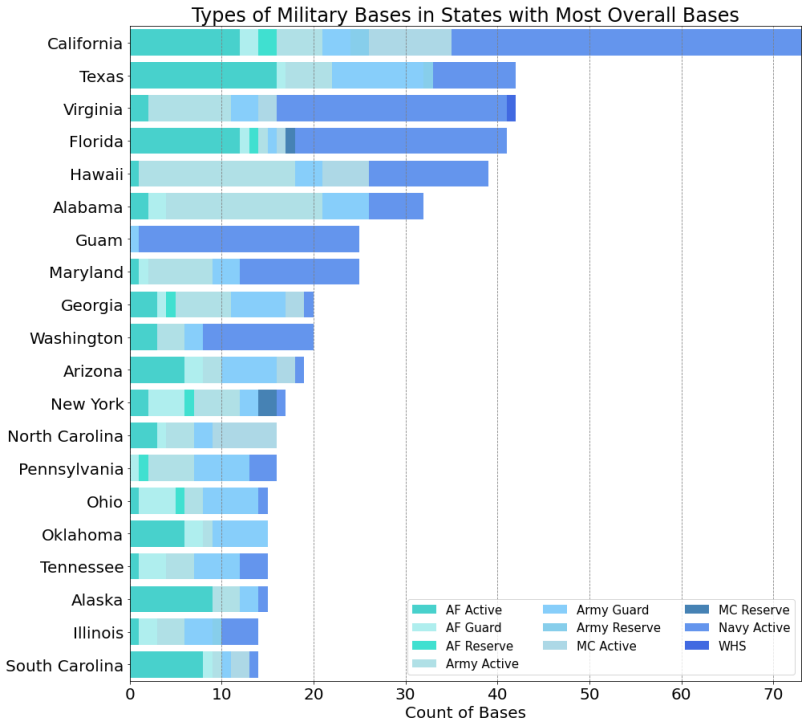


Figure 5

Figure 6

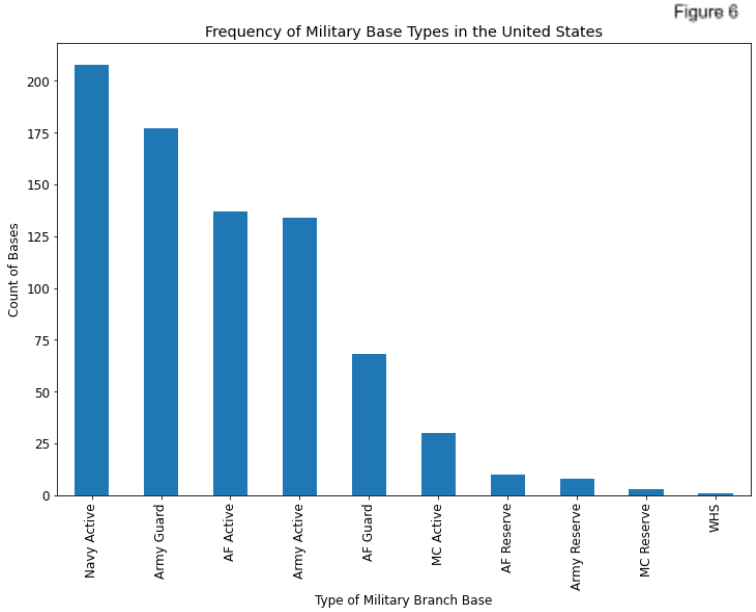
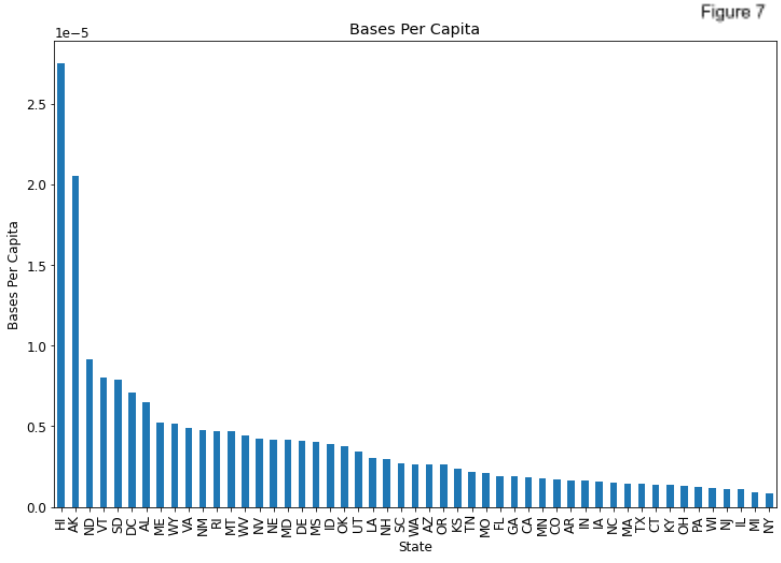


Figure 6 additionally shows the total number of each type of military base in the country overall. As seen in Figure 5, Naval bases and Airforce bases are some of the most common types. However, an interesting insight here is that the second most common type of base is Army Guard, which as seen in the above figure, was not as common in the states with the most bases overall.

In order to understand the size of these bases, the mean area was calculated for each state. Appendix B shows the average size (area, in square miles) of bases in each state. The research team found that Nevada and New Mexico had the largest average base size of any state, which met expectations.

Additionally, the number of bases located in each state was normalized against census population data in 2019 to find the number of bases per capita. This allowed the researchers to adequately assess which states should be further analyzed for trends in UFO sightings as corresponding to

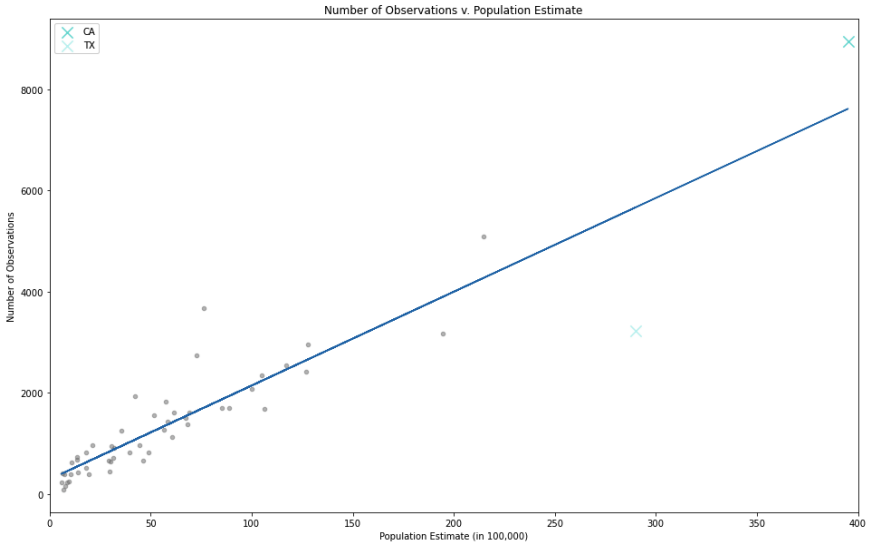
Figure 7



population. Figure 7 shows those results.

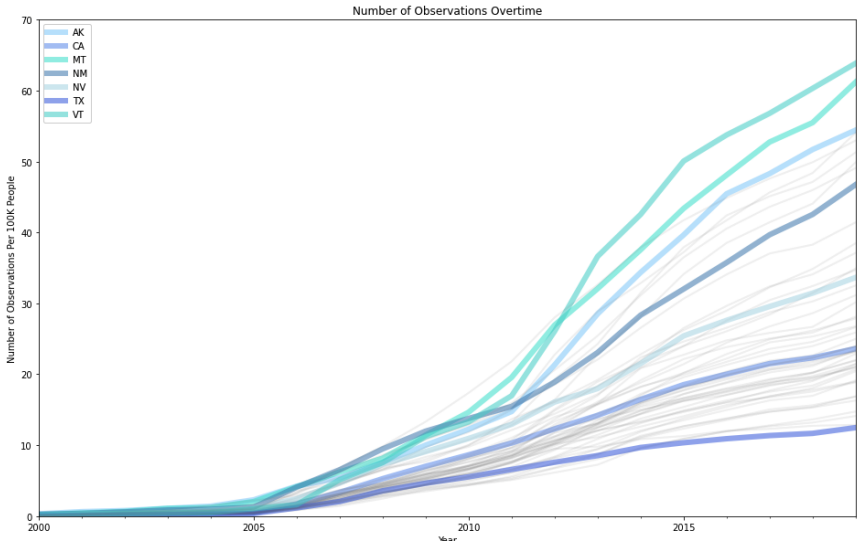
This analysis shows that some unexpected states have a higher proportion of military bases than initially expected, including Alaska and Vermont. This was further analyzed in the following sections of this report.

Figure 8



Similar to military bases, states that have more people staring at the sky and reporting what they see intuitively will have more reported UFO sightings. California has the most total number of sightings in the last 20 years and it is also the most populated state. Yet, as shown in Figure 8, the two most populated states California and Texas appear to have very different relationships with UFO. While they both on aggregate are among the states with the highest number of UFO sightings, California appears to have more sightings than a state its size is expected to have, while Texas on the other hand reported way less sightings than expected. Hence, the number of sightings per capita can more accurately represent the number of UFO sightings in a state.

Figure 9



An UFO sighting in general is a rare event. Each year, out of 100,000 people residing in the US, there are on average only 1.15 UFO sightings. As previously mentioned and shown in Figure 9, the data suggests that residents in California and Texas do not seem to have a particularly higher chance of seeing an UFO in the past 20 years. Figure 9 shows the total number of UFO sightings per 100,000 residents over the last 20 years. In fact, not only do California and Texas do not particularly stand out from the other states, UFO theorists' favorites Nevada, where Area 51 is located and New Mexico, where Roswell is, are not the states with the highest number of sightings as well. Vermont, Montana, and Alaska are actually the top three states with the most UFO sightings per 100,000 people.

Although Vermont, Montana, Alaska, New Hampshire, and Washington are the states with the highest number of sightings per 100,000 residents in the last 20 years, if one would like to see the traditional disk-shaped UFO, the data suggests it is only worth paying a visit to Alaska and not any of the other states. Figure 10 shows the top 5 shapes observed in the states with the highest per-capita sightings.

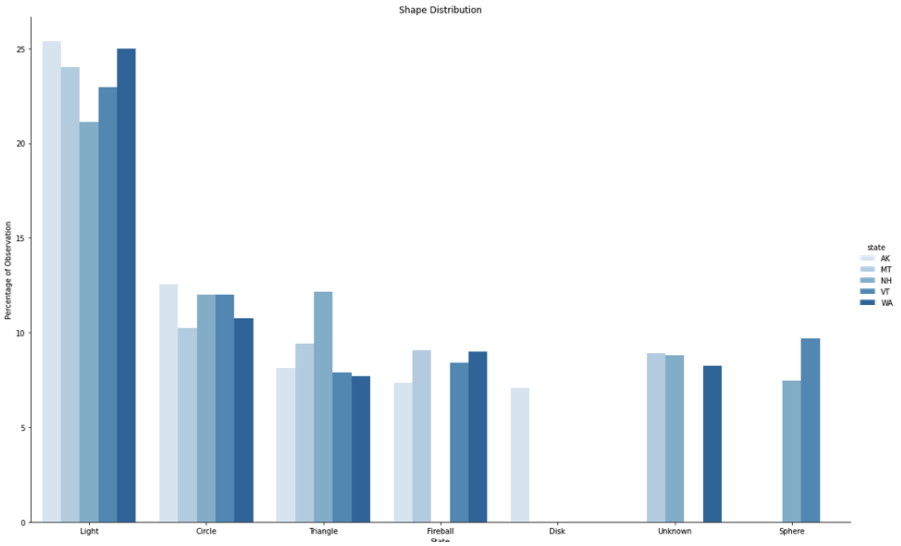


Figure 10

When analyzing the relationship between UFO sightings and military base statistics in each grouped state, Figure 11 shows California, Florida, Pennsylvania, Texas and Washington are states of initial increased interest because of the increased frequency of UFO sightings. These trends in the data are consistent with the census data as was previously shown in Figure 8.

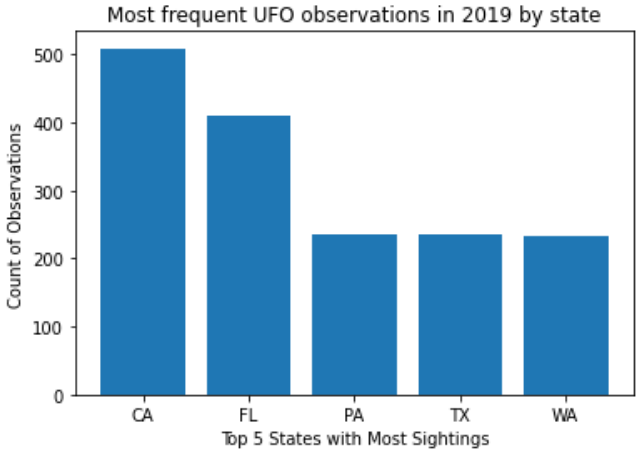
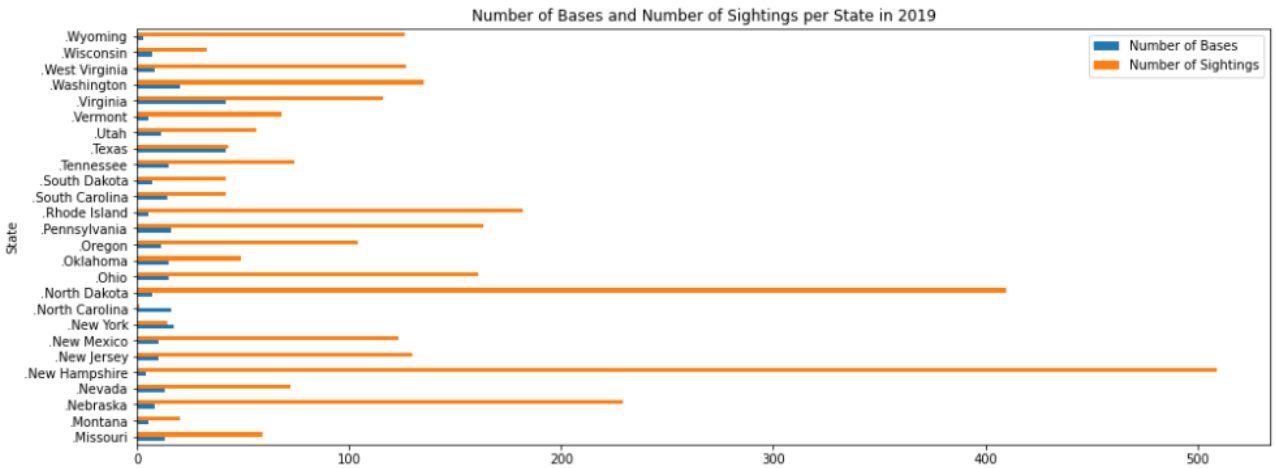


Figure 11

Although California, Florida, Pennsylvania, Texas and Washington are the states with the highest frequency of sightings, comparing the number of bases to the number of sightings might provide a more comprehensive idea of a relationship between military bases and frequency of UFO sightings. Both Figure 12 and Figure 13 exhibit that three of the five states with the highest frequency of sightings (California, Texas & Florida) also have the highest number of military bases.

Figure 12



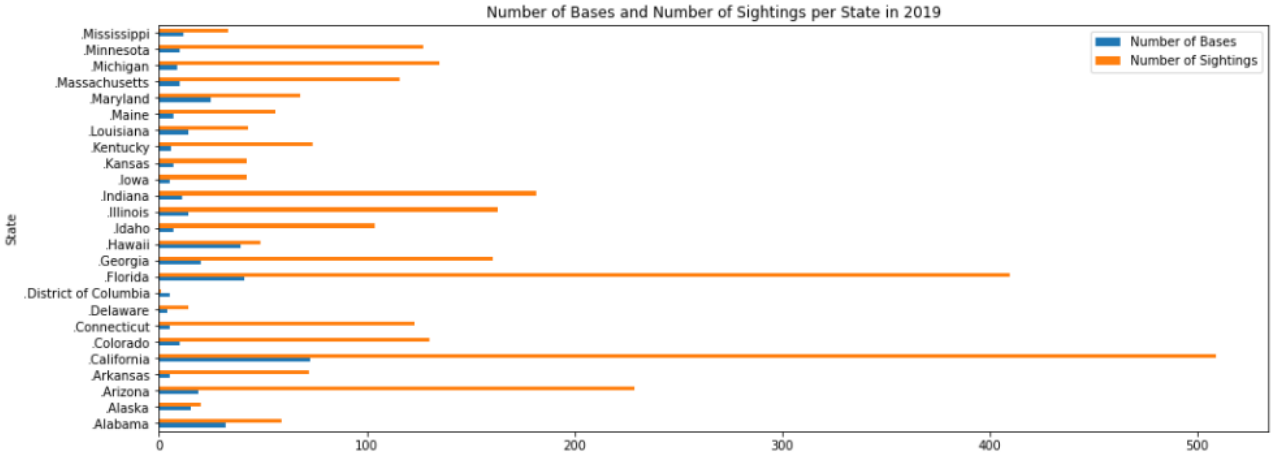


Figure 13

While examining the relationship between UFO sightings, the number of military bases, and type of military base the researchers focused on exploring states that had greater than one hundred sightings and greater than fifteen military bases. By focusing on the outlined subset of data, the researchers hope to control for potential outliers that could skew the trends in data. While some of the larger states may appear less exciting between 2000-2019 when controlling for population, that is not the case when looking at the most recent UFO sighting data from 2019. Data from 2019 was selected because it is the most recent military base, and UFO sighting data that is not affected by COVID.

State	Number of Bases	Sightings
.California	73	509
.Texas	42	233
.Virginia	42	193
.Florida	41	410
.Hawaii	39	49
.Alabama	32	59
.Maryland	25	68
.Washington	20	233
.Georgia	20	161
.Arizona	19	229
.New York	17	188
.Pennsylvania	16	235
.North Carolina	16	162
.Tennessee	15	126
.Alaska	15	20

Figure 14 is a more granular view displaying the service type of each military base compared to sightings. Initial assumptions led the researchers to believe Air Force bases would be predominant in States with the highest frequency of UFO sightings. However, Figure 14 shows that for the states California, Florida, Pennsylvania and Texas this is not the case. California, Florida, and Pennsylvania have more Naval Bases than they do Air Force bases.

Figure 14

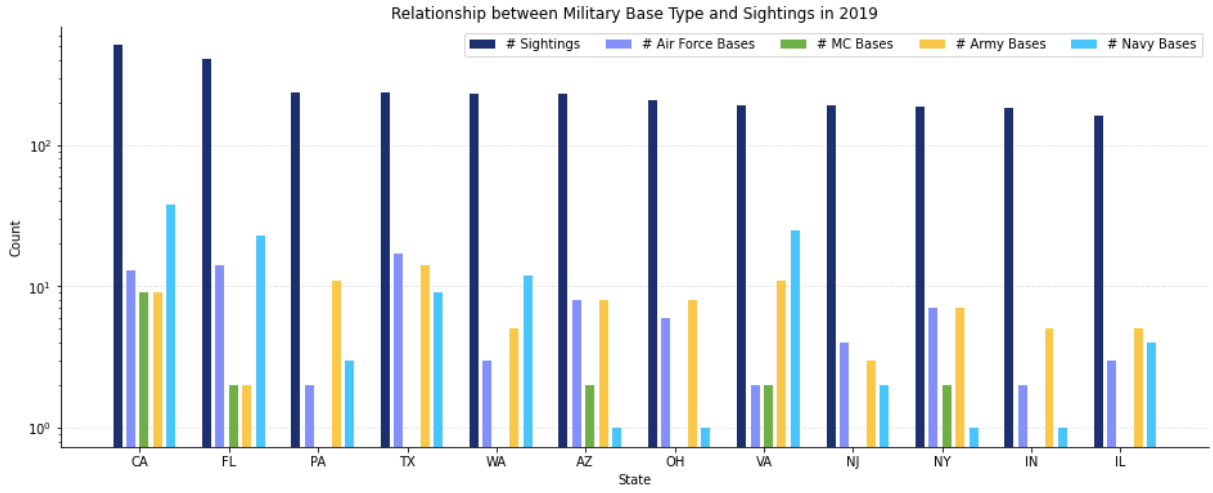
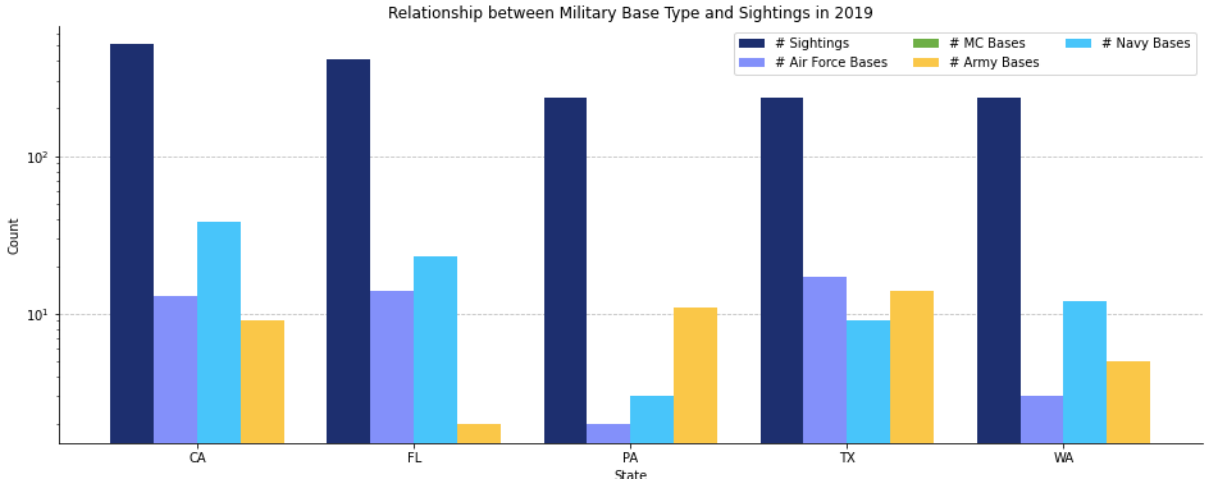


Figure 14 specifically focuses on the top five states with the most sightings to see if the trend observed within Figure 13 continues. Washington follows the trend of previously described by having a higher count of naval bases than Air Force bases. However, Texas does not follow the same trend as the other states listed. Texas has a higher count of Air Force bases than both Navy and Army bases.

Figure 15



Limitations

The first limitation that will be acknowledged is how the proximity to military bases was dealt with. When the research team analyzed proximity to military bases, the distance from a UFO sighting to the nearest military base was not considered. Instead, proximity of the UFO sightings to military bases was determined by the number of bases in the state. This might not provide as effective insight. Given more time, this could be addressed by adding geolocation data for all the cities listed in the UFO sightings and comparing this to the military base geolocation and calculating a distance between the two.

The second limitation that will be acknowledged in this section is a concern about the integrity of the military data. The researchers noticed that in the state of Maryland, there were more military bases listed in the data than there actually were in the state. Upon further investigation, it was found that “Solomons” is listed as a Naval Base. In reality, it is a camping site that navy veterans have access to. The researchers decided to keep the data because it was determined that this would still satisfy the requirements to analyze, explore, clean, and draw conclusions from data. Given a longer timeline, the researchers would remove the entries in the military database that are not truly military bases.

Conclusion

The analyses presented in this research have examined various relationships between UFO sightings, military bases, and population in the United States. Specifically, the data suggests that UFO sightings have been increasing rapidly in the early 2010s and less so in recent years. While states with higher populations tend to have more sightings, after we normalize the data against population, Vermont, Montana, and Alaska appear to have the highest per capita sightings. Interestingly, Vermont and Alaska are among the states with the most military bases per capita between 2000 and 2019 as well. Moreover, states with more sightings tend to have more naval bases. In summary, our data demonstrates a positive relationship between the number of UFO sightings and the number of military bases. Given the limitation of the data and approach this research relies upon, the findings are mostly for research purposes and are not intended to draw any conclusion.

References

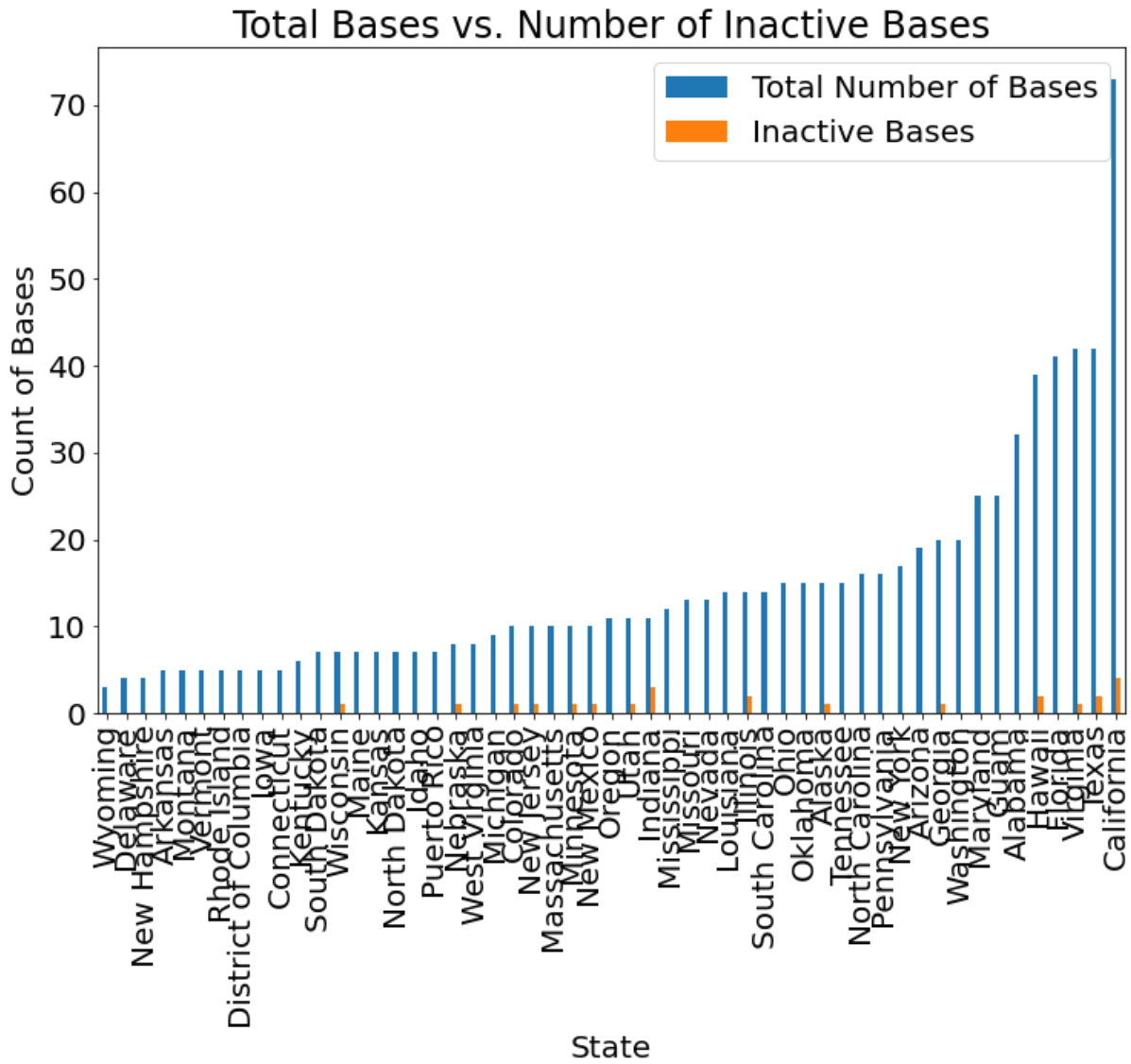
Oxford University Press (OUP). (2019, December 9). What Is The Origin Of The Term “UFO”? Lexico.Com. https://www.lexico.com/explore/what-is-the-origin-of-the-term-ufo

Project BLUE BOOK - Unidentified Flying Objects. (2020, September 29). National Archives. https://www.archives.gov/research/military/air-force/ufos

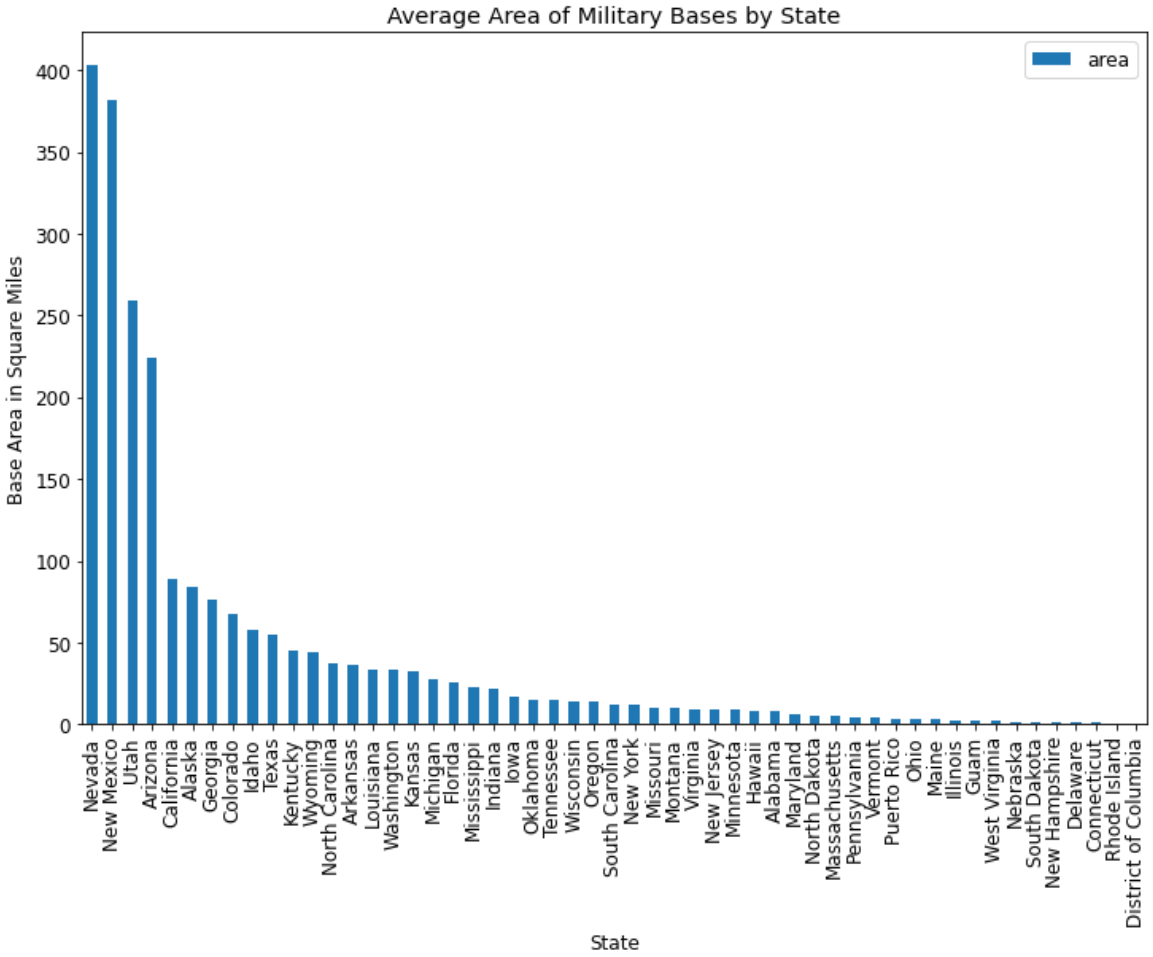
Cooper, H., Blumenthal, R., & Kean, L. (2021, January 21). Glowing Auras and ‘Black Money’: The Pentagon’s Mysterious U.F.O. Program. The New York Times. https://www.nytimes.com/2017/12/16/us/politics/pentagon-program-ufo-harry-reid.html

Appendix

Appendix A.



Appendix B.


Average Area of Military Bases by State
area
Base Area in Square Miles
0
50
100
150
200
250
300
350
400
Nevada
New Mexico
Utah
Arizona
California
Alaska
Georgia
Colorado
Idaho
Texas
Kentucky
Wyoming
North Carolina
Arkansas
Louisiana
Washington
Kansas
Michigan
Florida
Mississippi
Indiana
Iowa
Oklahoma
Tennessee
Wisconsin
Oregon
South Carolina
New York
Missouri
Montana
Virginia
New Jersey
Minnesota
Hawaii
Alabama
Maryland
North Dakota
Massachusetts
Pennsylvania
Vermont
Puerto Rico
Ohio
Maine
Illinois
Guam
West Virginia
Nebraska
South Dakota
New Hampshire
Delaware
Connecticut
Rhode Island
District of Columbia
State