IST687, Ted Tinker, Final Project
Due June 20, Submitted:

June 18, 2019

Introduction

On the website data.world, user timothyrenner has uploaded a csv file of reported UFO sightings[1] from the National UFO Research Center, and also a csv file of reported bigfoot sightings[2] from the Bigfoot Field Researchers Organization. There are 112,000 UFO reports, starting in the 1960's, and 4,800 Bigfoot reports, dating back to 1872. Both kinds of reports span North America and include information like the date, city, state, and latitude/longitude of each sighting.

Unidentified Flying Objects exist—if it flies, and you don't know what it is, it's a UFO; it's just probably not an alien spaceship. Bigfoot, in contrast, doesn't exist at all (I think). The fiction of space-aliens and Bigfeet nevertheless impacts real peoples' real beliefs. Therefore, I would like to explore similarities and differences between the two datasets.

- Where are UFOs usually sighted? Where is Bigfoot usually sighted? Let's compare sightings by US state, absolutely and per capita.

- Did different areas get different proportions of UFO-shapes ("round" vs "pointy" vs "vague")?

- What time of day/week/year are UFOs and bigfoot usually sighted?

- What were people doing when they saw Bigfoot? How many people who saw UFOs also said they saw aliens? Use text-mining to find interesting details and themes among the reports.

Data Acquisition, Cleansing, and Dictionary

First I downloaded the files BFRO_reports_geocoded (of Bigfoot sightings) and NUFORC_reports (of UFO sightings), converted them from CSVs to Excel workbooks, and decided which columns would be worth investigating. One UFO had to be expunged: the 'Moon of Memphis' was a UFO with a description so long it flowed into following row, producing erroneous values.

---

[1]data.world/timothyrenner/ufo-sightings
[2]data.world/timothyrenner/bfro-sightings-data

In the UFO database, the columns I kept were:

| Attribute	Type	Comments
City	Text	20,000 place-names and placeholders. Less than 1,000 missing values.
State	Text	About 10,000 missing values.
City_Latitude	Decimal	Precise up to eight decimal places. About 20,000 missing.
City_Longitude	Decimal	Precise up to eight decimal places. About 20,000 missing.
Date_Time	yyyy-mm-dd T hh:mm:ss	About 3,000 missing.
Date	yyyy-mm-dd	Derived from Date_Time.
Time	hh:mm:ss	Derived from Date_Time.
Season	Text	Spring, Summer, Winter, Fall, and Unknown. Derived from Date.
Day	Text (Abb.)	"Mon," "Tue," etcetera. Derived from Date.
Hour	Integer	Between 00 and 23. Derived from Time.
Shape	Text	Like "circle" or "cigar." About 3,000 missing and 8,000 "unknown."
Duration	Text	Highly irregular. About 3,000 missing values.
Text	Text	Event description. Very few missing.

In the Bigfoot database, the columns I kept were:

Attribute	Type	Comments
County	Text	1,019 counties like "York County" or "Macon County." No missing values.
State	Text (Full)	Interestingly, no bigfoot sightings in Hawaii. No missing values.
Latitude	Decimal	Precise to five decimal places. About 1,000 missing.
Longitude	Decimal	Precise to five decimal places. About 1,000 missing.
Date	mm/dd/yyyy	About 1,000 missing.
Season	Text	Spring, Summer, Winter, Fall, and Unknown.
Day	Text (Abb.)	"Mon," "Tue," etcetera. Derived from Date.
Moon_Phase	Decimal	Precise to two decimal places. About 1,500 missing.
Temperature_High	Decimal	$°F$. Precise to two decimal places. About 1,500 missing.
Temperature_Mid	Decimal	$°F$. Precise to two decimal places. About 1,500 missing.
Temperature_Low	Decimal	$°F$. Precise to two decimal places. About 1,500 missing.
Humidity	Decimal	From 0 to 1. Precise to two decimal places. About 1,500 missing.
Cloud_Cover	Decimal	From 0 to 1. Precise to two decimal places. About 2,000 missing.
Precip_Probability	Decimal	From 0 to 1. Precise to two decimal places. About 2,500 missing.
Precip_Intensity	Decimal	Inches per hour. Precise to three decimal places. About 2,500 missing.
Precip_Type	Text	"Rain," "Snow," or blank.
Pressure	Decimal	Precise to two decimal places. About 2,500 missing.
Visibility	Decimal	Precise to two decimal places. About 2,000 missing.
Wind_Bearing	Integer	0 to 359. About 1,500 missing.
Wind_Speed	Decmial	Precise to two decimal places. About 1,500 missing.
Observed	Text	Event description. Less than 100 missing.

I also include state population data from a 2019 census.

# Basic Overview

UFO sightings:
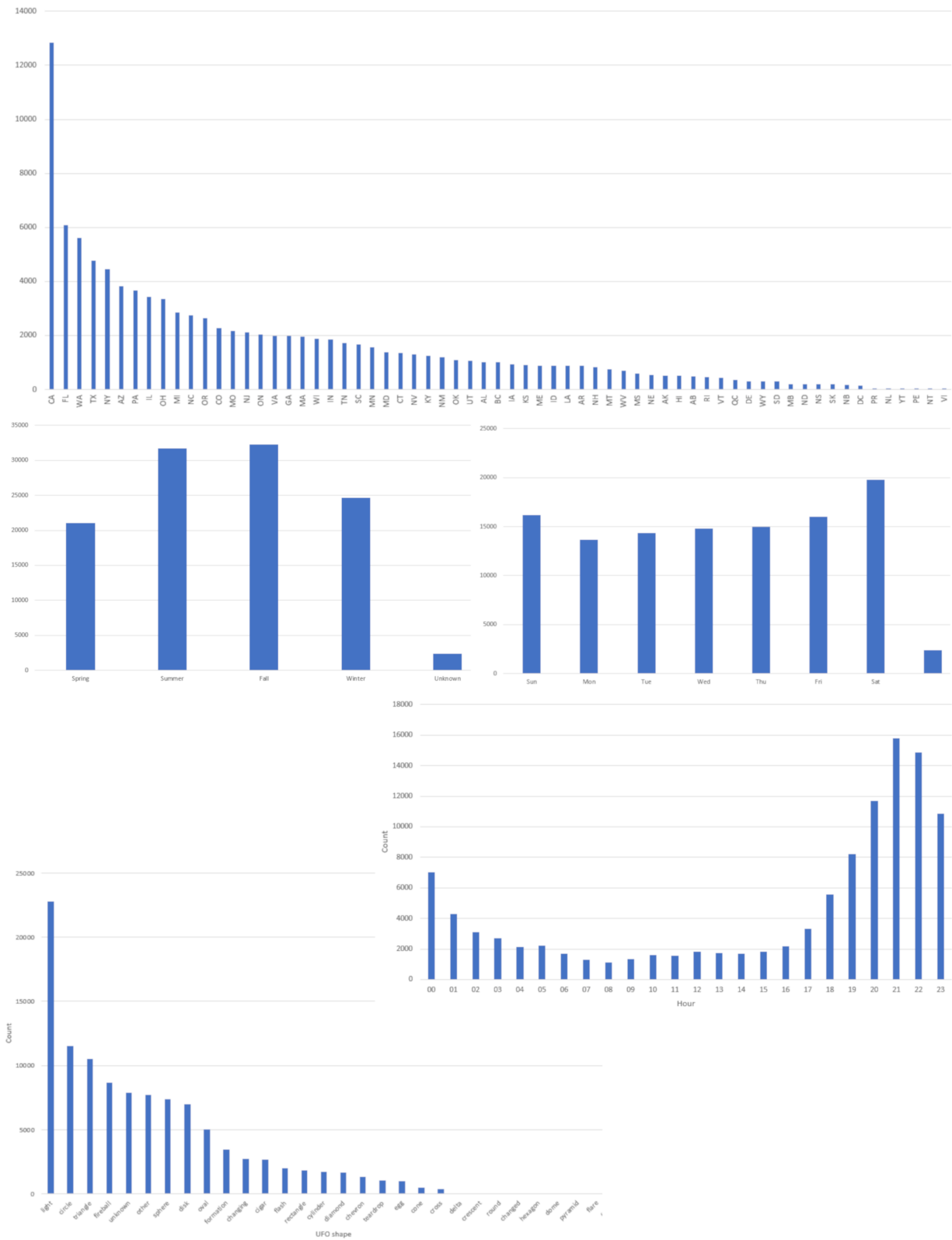
In the UFO data (previous page), we see that UFOs appear most often in California, Florida, and Washington, and most frequently in the Summer and Fall. UFOs are sighted more and more often as the week progresses, hitting their height on Saturday, and typically swing by around 9 PM. The most common UFOs are lights and circles.

In the bigfoot data (next page), we see that bigfoot most often frequents Washington, California, and Florida, usually in the Summer and Fall. He visits most frequently on Saturday. Otherwise, it's difficult to tell if the phase of the moon, wind bearing/speed, cloud cover, precipitation, or visibility impacted bigfoot's presence.

These charts were made with Excel Pivot-Tables. Sections after this one utilize R-Studio.

Bigfoot sightings:
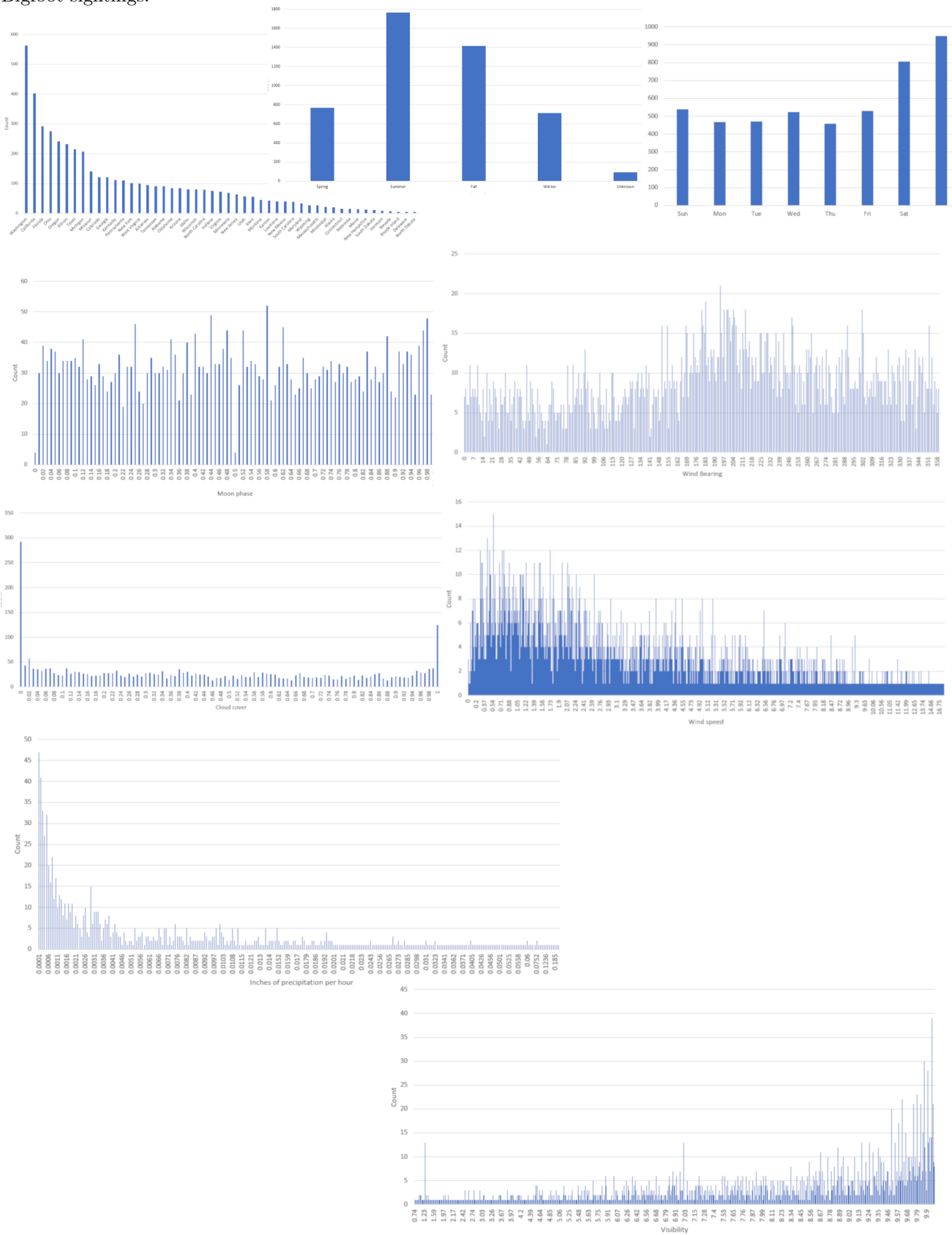
Maps of Sightings

Graphing sightings with ggplot gives a cursory view of the datasets. As expected, the most populous states like California, Florida, and Washington have the most sightings. This version is basically a heat-map of population.

Out of curiosity, I plotted sightings near New York. It's interesting that bigfoot never appears on Long Island; he's not a great swimmer. UFOs can fly over bodies of water.

If we divide the number of sightings in each state by the population of the state (according to a 2019 census), we see activity per capita.

Bigfoot sightings by state, per capita

UFO sightings by state, per capita
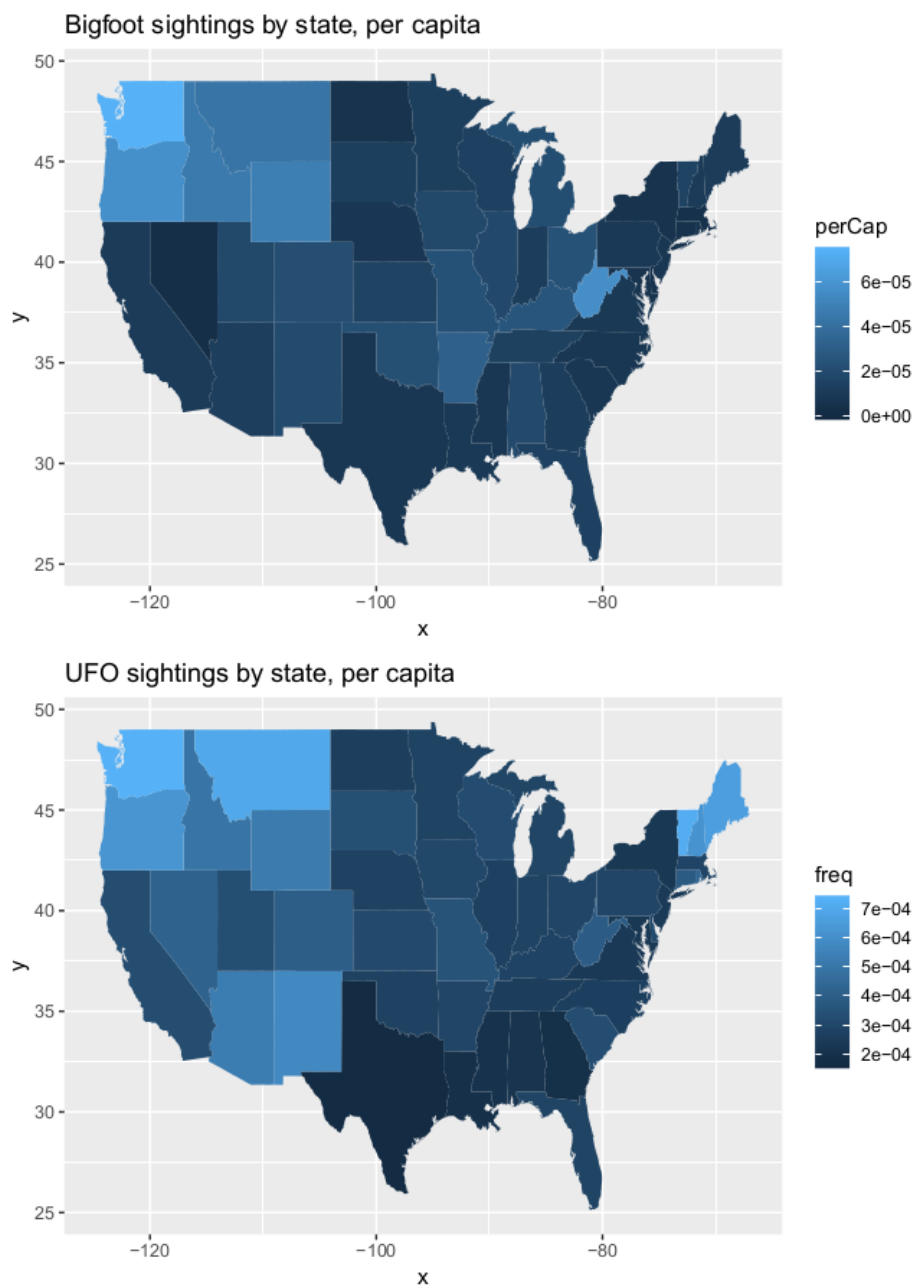
Bigfoot is sighted most often (per capita) in the north-west, and, peculiarly, in West Virginia. UFOs also frequent the north-west, but also visit states near Maine and, of course, Roswell, New Mexico.

UFO Shapes

I grouped UFO shapes into three categories: Round (for shapes like circles, disks, ovals, and eggs), Pointy (for shapes like triangles, rectangles, and pyramids), and Vague (for shapes like fireball, flare, other, and unknown).

UFO sightings by shape
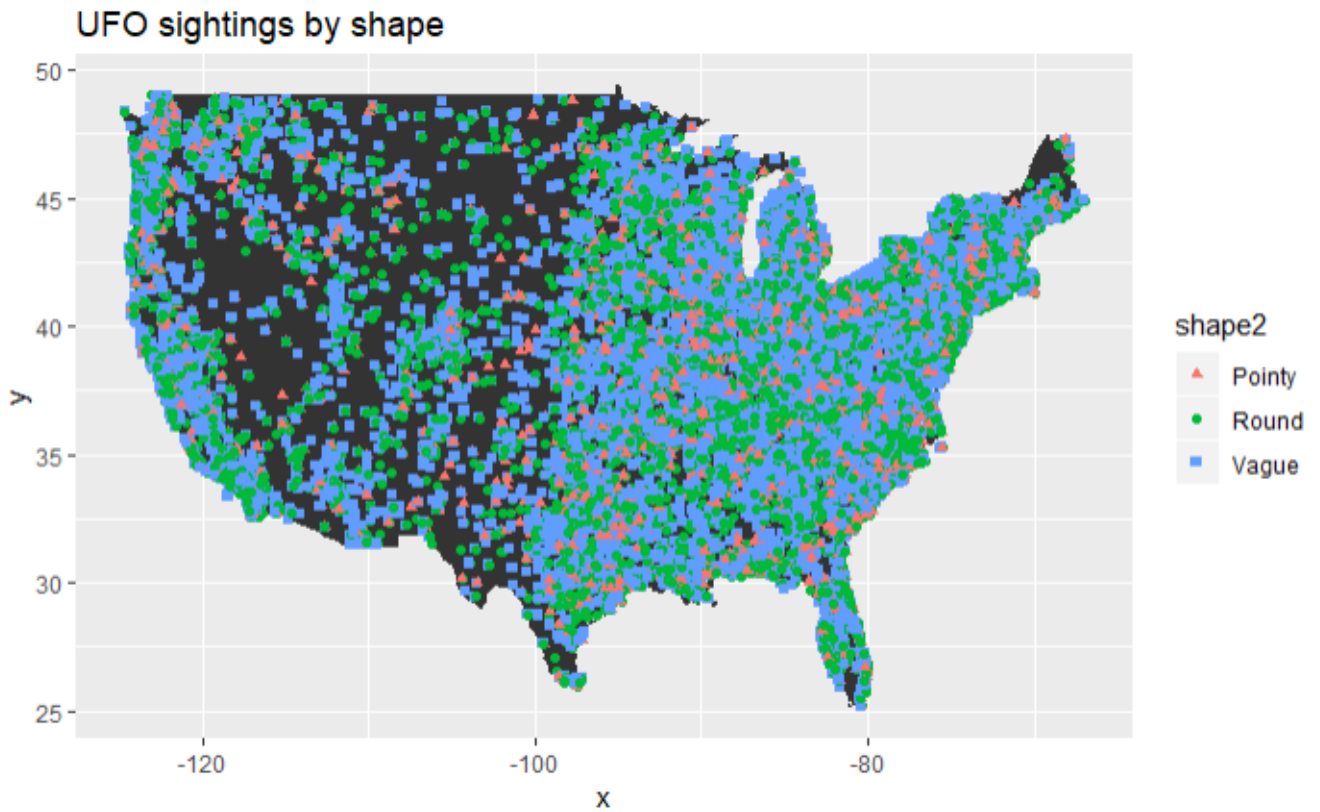
Vague shapes account for about 44,000 UFOs, Round shapes for 40,000, and Pointy shapes for the remaining 14,000. The shapes appear to be uniformly distributed; there are no obvious collections of pointy UFOs, for example.

Time of Day/Week/Year

Both bigfoot and UFOs live for the weekend, appearing most often on Saturday.
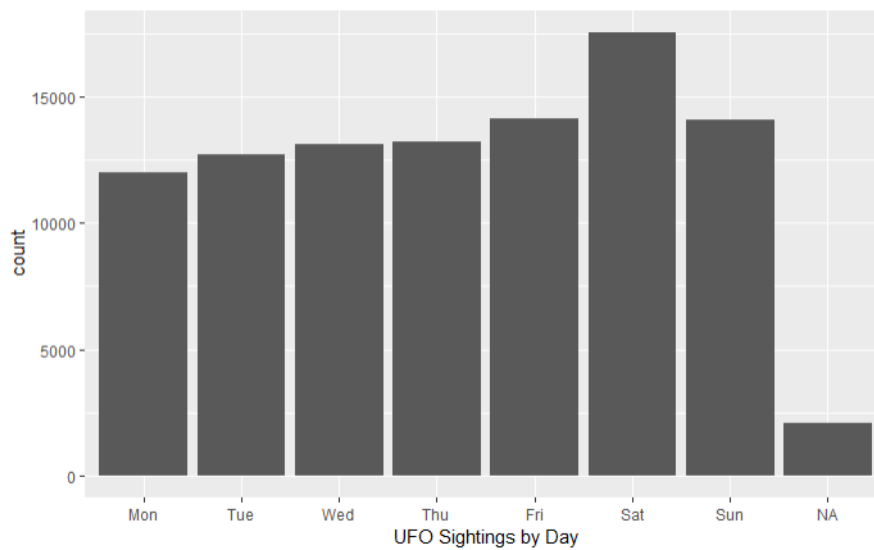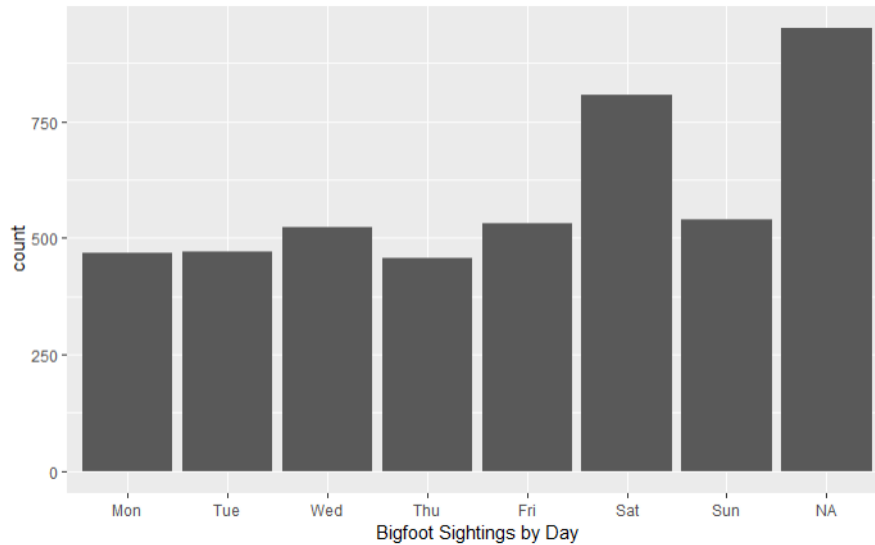
As seen in the basic overview, UFOs and bigfoot appear most often in the Summer and Fall, and UFOs tend to arrive around nine o' clock and leave by midnight.

Text Mining

A wordcloud of terms from the UFO dataset shows a trend of describing attributes of abstract objects, like their colors and shapes, and their actions, and their position in the sky relative to the horizon, clouds, and the viewer. For stability reasons, I only used the first 13,000 entries.
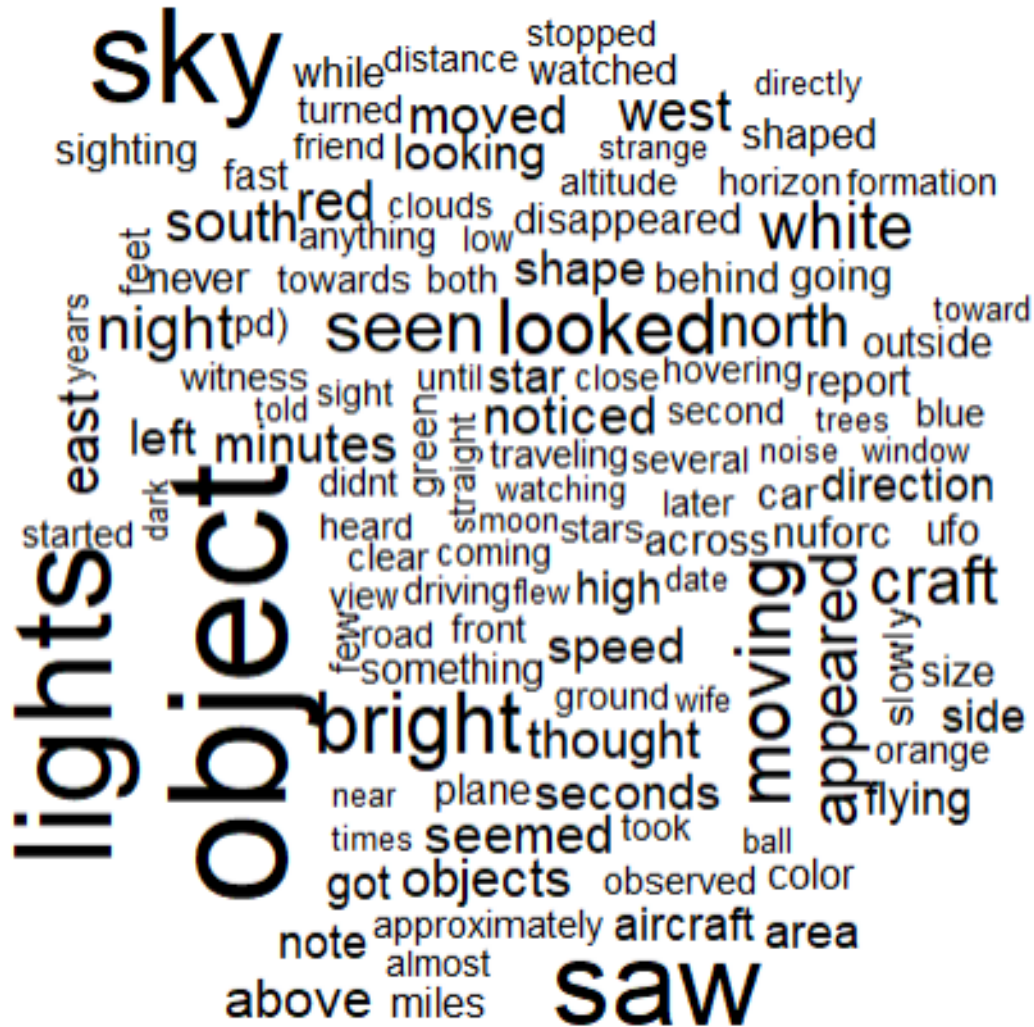
Disappointingly, less than 2% of the reports mention aliens.

In contrast, bigfoot-sighting reports tend to describe everything except bigfoot, including trees, camps, roads, dogs, deer, and the viewer's family-members. People also described why they were in the area (eg fishing, hunting, walking, camping).

About a third of reports call him 'bigfoot,' about a tenth call him 'sasquatch,' and only a third of a percent call him a 'yeti.'

Intriguingly, total AFINN score for the bigfoot text is -19,046—very negative. In three times as many text entries from the UFO dataset, the AFINN score is 11,128. Apparently people are happier to see UFOs than bigfoot.

How to meet bigfoot

Considering what we've learned, if we're eager to witness the paranormal, we should go camping every Summer weekend in Washington state. We should do outdoor activities in the evening, like drive to fishing spots along old country roads, so we'll spot bigfoot if he's around. Meanwhile, we should scan the sky and the clouds.

Thankfully, if enough people take this advice, Washington will fill with like-minded individuals looking for bigfoot in the night. I'm sure if any of those individuals bump into each other in the dark, each might mistake the other for sasquatch. The collective interest in bigfoot would conjure him.