Final Paper

Shealagh Brown & Sam Zimpfer

2025-05-01

Introduction

The data we were working with came from a data set called "UFO sightings scrubbed" that was found on Kaggle.com from a user named Akhil Goyal. The data was last updated three months ago, making it quite recent data. It contains information on all UFO sightings dating back to 1906. The data could have some bias if ufo sighitngs from certain regions of the world were not recorded or included in this data set, additionally it is observational data collected by differnt people around the globe which can create large amounts of variation in the data. This data is of interest because UFOs have been a topic of public debate for years. With increasing amounts of interest in space travel and extraterrestrials in more recent years, the facinations with UFOs has only grown stronger. For this project we want to explore what influiences sightings as this can be valuable knowledge for those trying to investigate UFOs. In order to work with our data we had to clean it. This includeded converting the datetime column into year, month, day, seconds, minutes, hours format. Then we created a new data set where we added columns for years, seconds, and months and kept the updated datetime, city, state, country, longitude, and latitude columns. We then had to convert both longitude and latitude into numeric values in order to work with them. The code for this data cleaning follows:

```
ufo_raw <- read.csv("ufo_sightings_scrubbed.csv")

#convert date time to ymd_hms format
ufo_raw$datetime <- ymd_hms(ufo_raw$datetime)

#cleaning data
ufo_raw|>
   mutate(seconds = duration..seconds.,
          year = year(datetime), #create year column
          month = month.name[month(datetime)])|> #create month column and convert to name
  select(datetime, city, state, country, seconds,latitude,longitude, year, month) -> ufo

ufo$latitude <- as.numeric(ufo$latitude) #changing lat to numeric
```

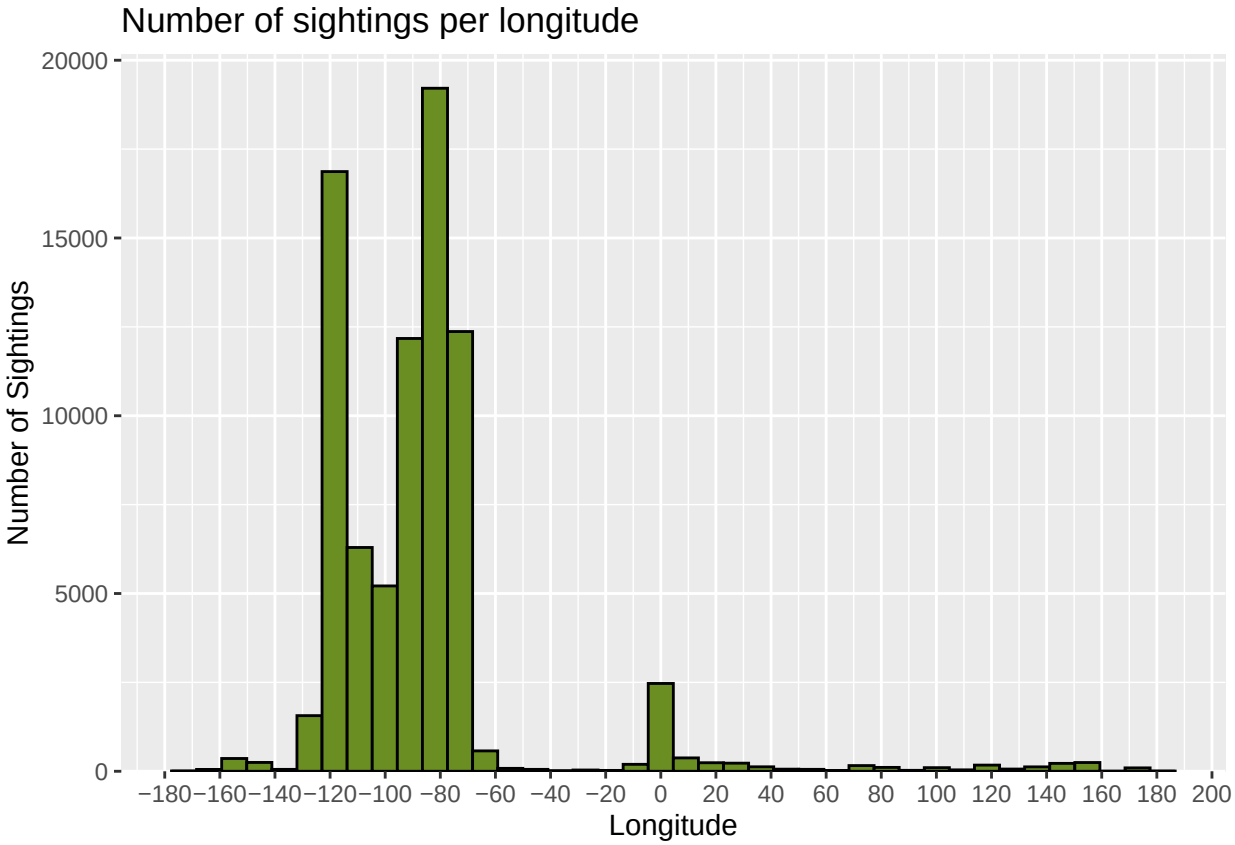
```
## Warning: NAs introduced by coercion
```

```
ufo$seconds <- as.numeric(ufo$seconds) #changing seconds to numeric
```

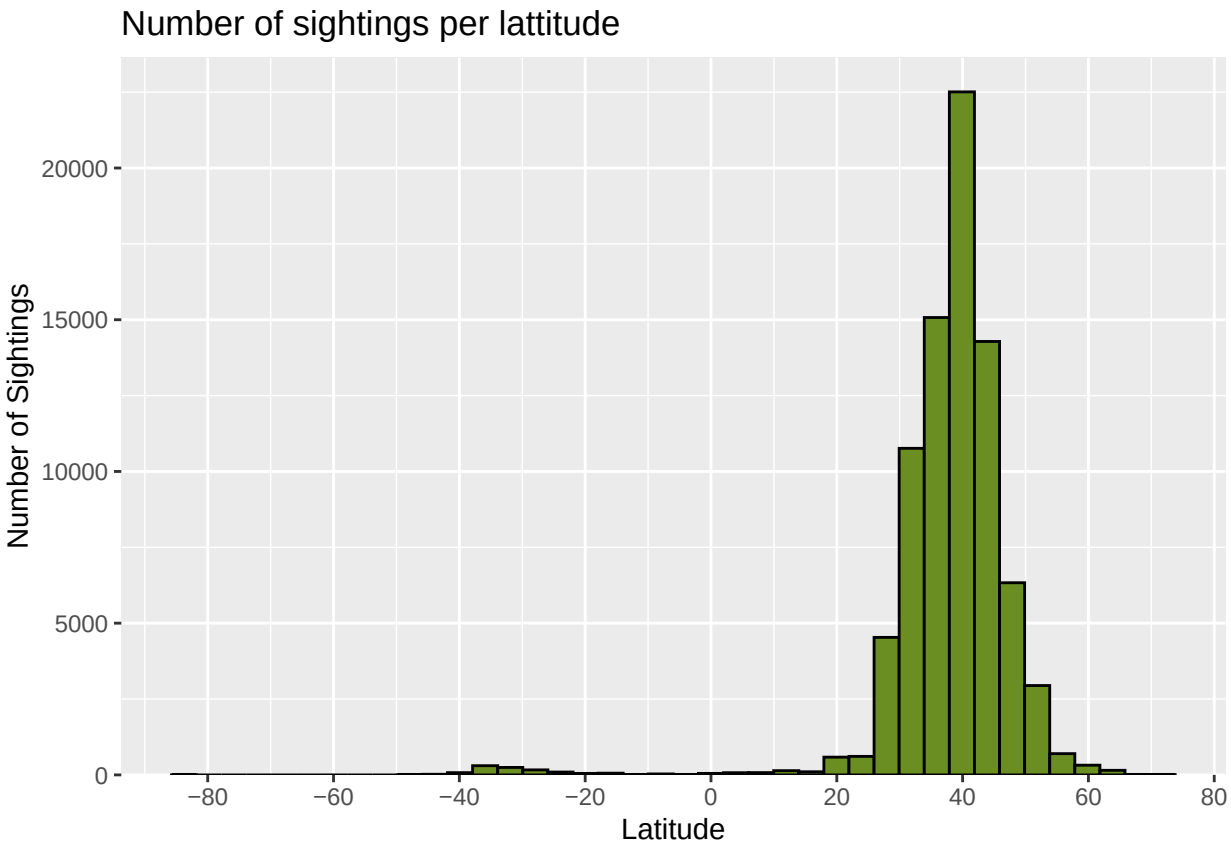
```
## Warning: NAs introduced by coercion
```

Data Analysis

The first question we had about UFO sightings is how the location around the world influienced




```
## Warning: Removed 1 row containing non-finite outside the scale range
## (‘stat_bin()‘).
```


Number of sightings per lattitude
Number of Sightings
20000
15000
10000
5000
0
−80
−60
−40
−20
0
20
40
60
80
Latitude