

UFO

Shealagh Brown & Sam Zimpfer

2025-03-20

Introduction

Cleaning the data

The code chunk below reads in our data about UFO sightings. We had to convert the datetime variable to year, month, date format. Then we created a new data frame containing only necessary variables for our purposes. Latitude had to then be converted into a numeric variable.

The variables used are: **datetime**: the date and time of sighting in year, month, day, hours, minutes, seconds format.

city: city where sighting occurred

state: state where sighting occurred

country: country where sighting occurred

seconds: duration of sightings in seconds

latitude: latitude of sighting

longitude: longitude of sighting

year: year when sighting occurred

month: month when sighting occurred

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

## Warning: NAs introduced by coercion
```

Data Summary

The first question we wanted to answer is where do most UFO sightings occur.

We created graphs of sightings based on longitude and latitude

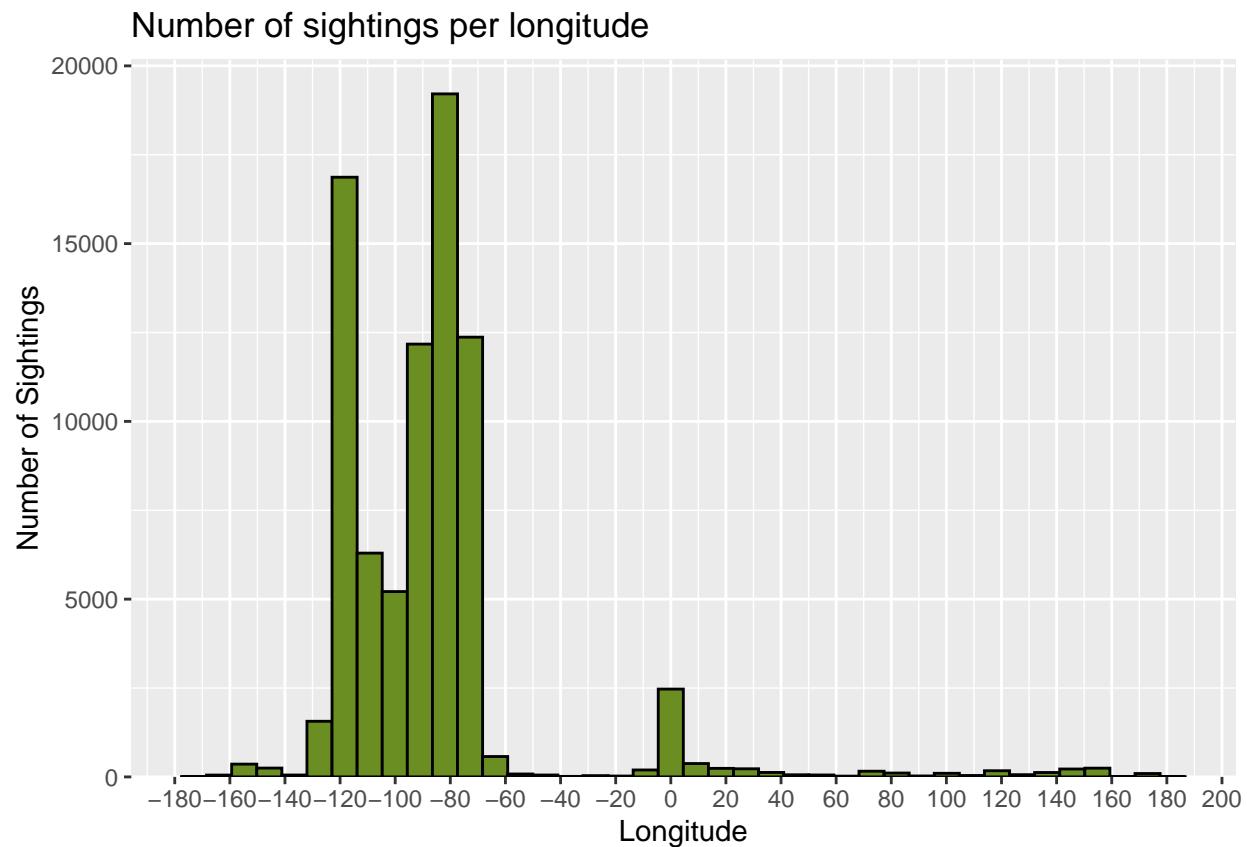
```

long_hist <- ggplot(
  data = ufo,
  mapping = aes( x= longitude
  )
) +
  geom_histogram(fill = "olivedrab",
                 color = "black",
                 bins = 40) +
  labs(
    title = "Number of sightings per longitude",
    x = "Longitude",
    y = "Number of Sightings"
  ) +
  scale_x_continuous(breaks = seq(-200, 200, 20)) +
  scale_y_continuous(expand = c(0, 0, 0.05, 0))

lat_hist <- ggplot(
  data = ufo,
  mapping = aes( x= latitude
  )
) +
  geom_histogram(fill = "olivedrab",
                 color = "black",
                 bins = 40) +
  labs(
    title = "Number of sightings per lattitude",
    x = "Latitude",
    y = "Number of Sightings"
  ) +
  scale_x_continuous(breaks = seq(-100, 100, 20)) +
  scale_y_continuous(expand = c(0, 0, 0.05, 0))

long_hist

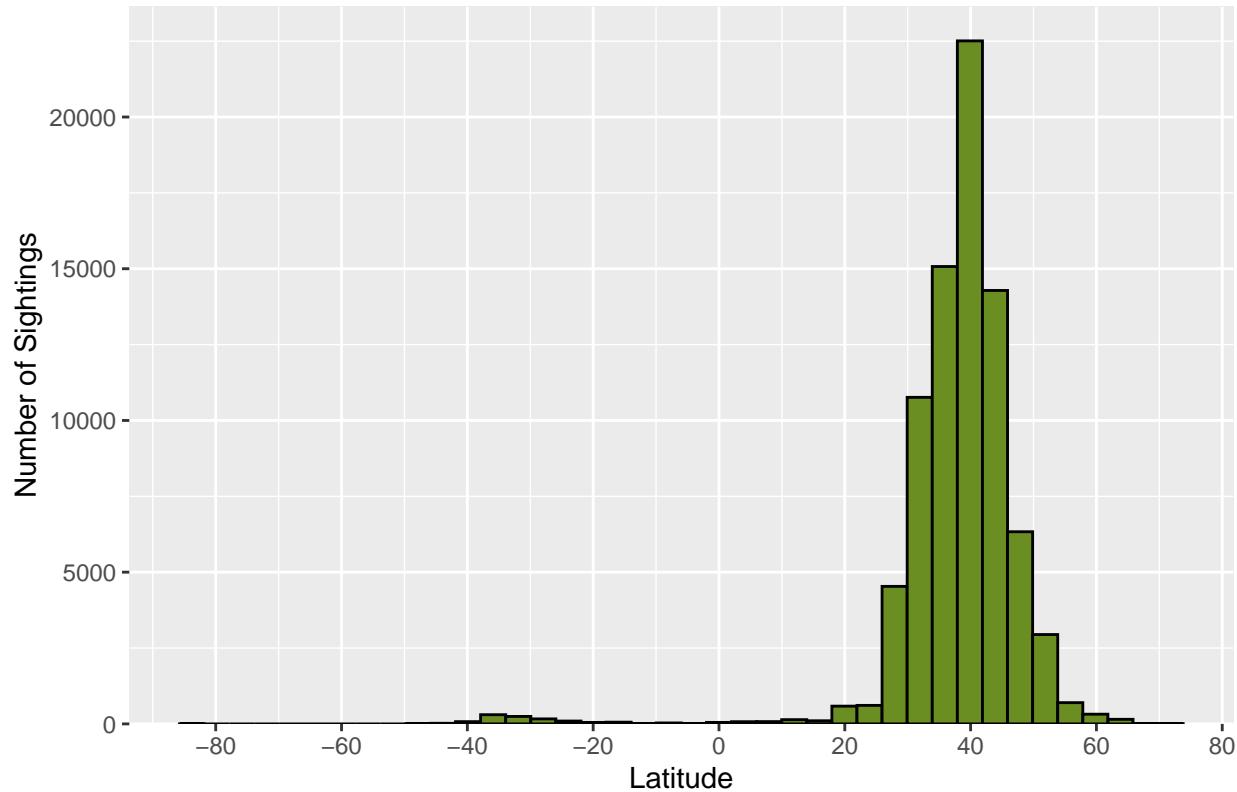
```



```
lat_hist
```

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

Number of sightings per latitude



Observations * Longitudes of sightings appear to be concentrated around -120 through -70 (Number of sightings per longitude) * Latitude of sightings appear to be concentrated around 30 through 50 (Number of sightings per latitude)

Scatter plot of sighting locations

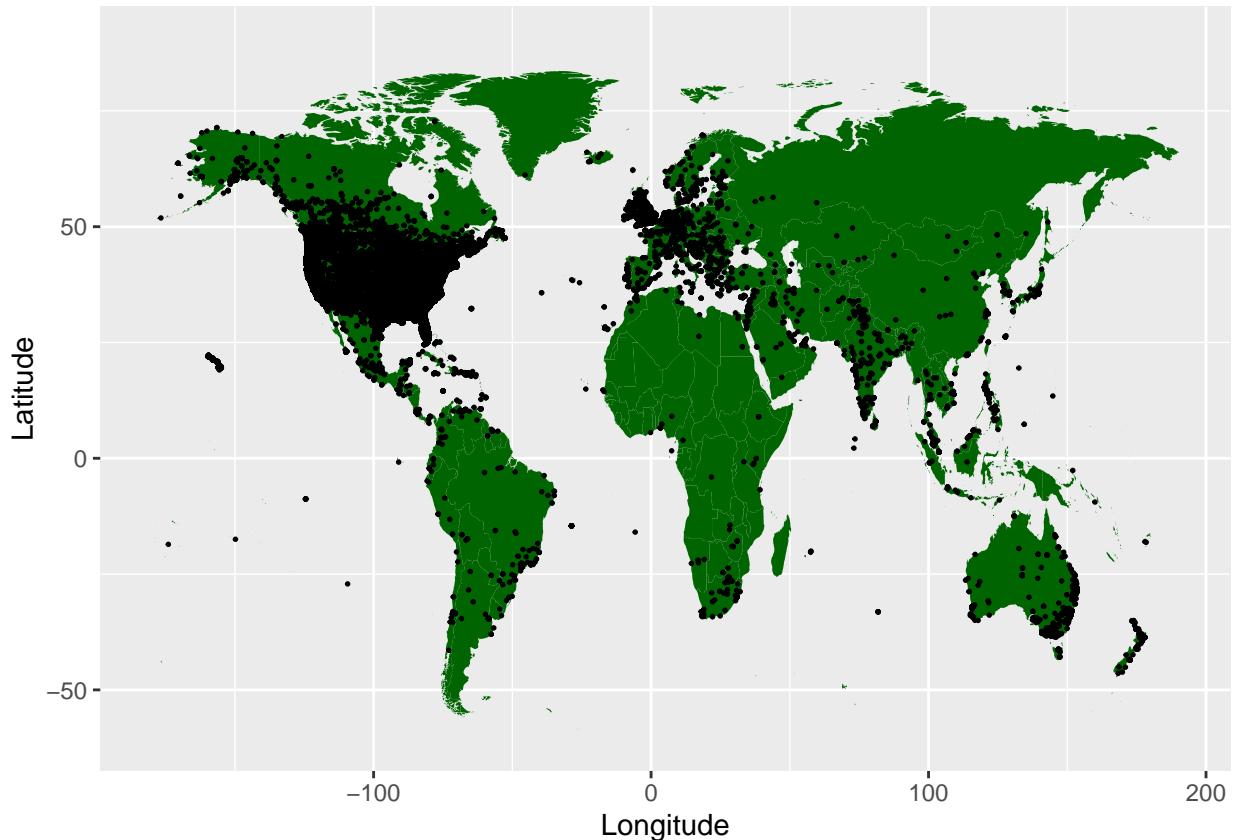
```
map <- map_data("world")

map_plot <- ggplot()+
  geom_polygon(data = map,
    mapping = aes(x= long,
                  y = lat,
                  group = group),
    fill = "darkgreen")+
  geom_point(data = ufo,
    mapping = aes(x = longitude,
                  y = latitude),
    size = 0.3)+
  labs(
    x = "Longitude",
    y = "Latitude"
  )+
  scale_y_continuous(expand = c(0, 0, 0.05, 0))+
  ylim(-60,90)
```

```
## Scale for y is already present.  
## Adding another scale for y, which will replace the existing scale.
```

```
map_plot
```

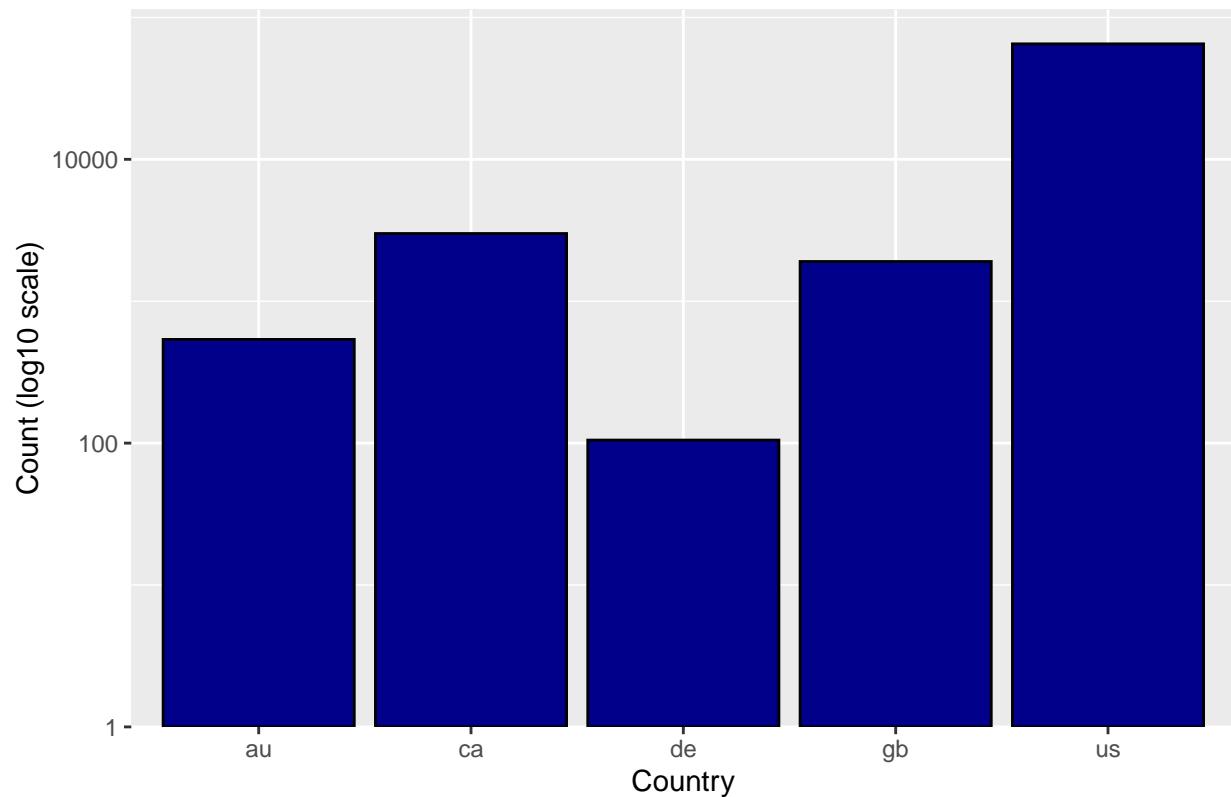
```
## Warning: Removed 2 rows containing missing values or values outside the scale range  
## ('geom_point()').
```



```
### Observations * Most of the sightings are concentrated in North America and in Europe
```

```
modified_data = ufo |> filter(country != "")  
  
ggplot(data = modified_data,  
       mapping = aes(x = country))+  
  geom_bar(fill = "blue4",  
           color = "black",  
           na.rm = T)+  
  scale_y_log10(expand = c(0, 0, 0.05, 0))+  
  labs(title = "Total number of sightings per country",  
       x = "Country",  
       y = "Count (log10 scale)")
```

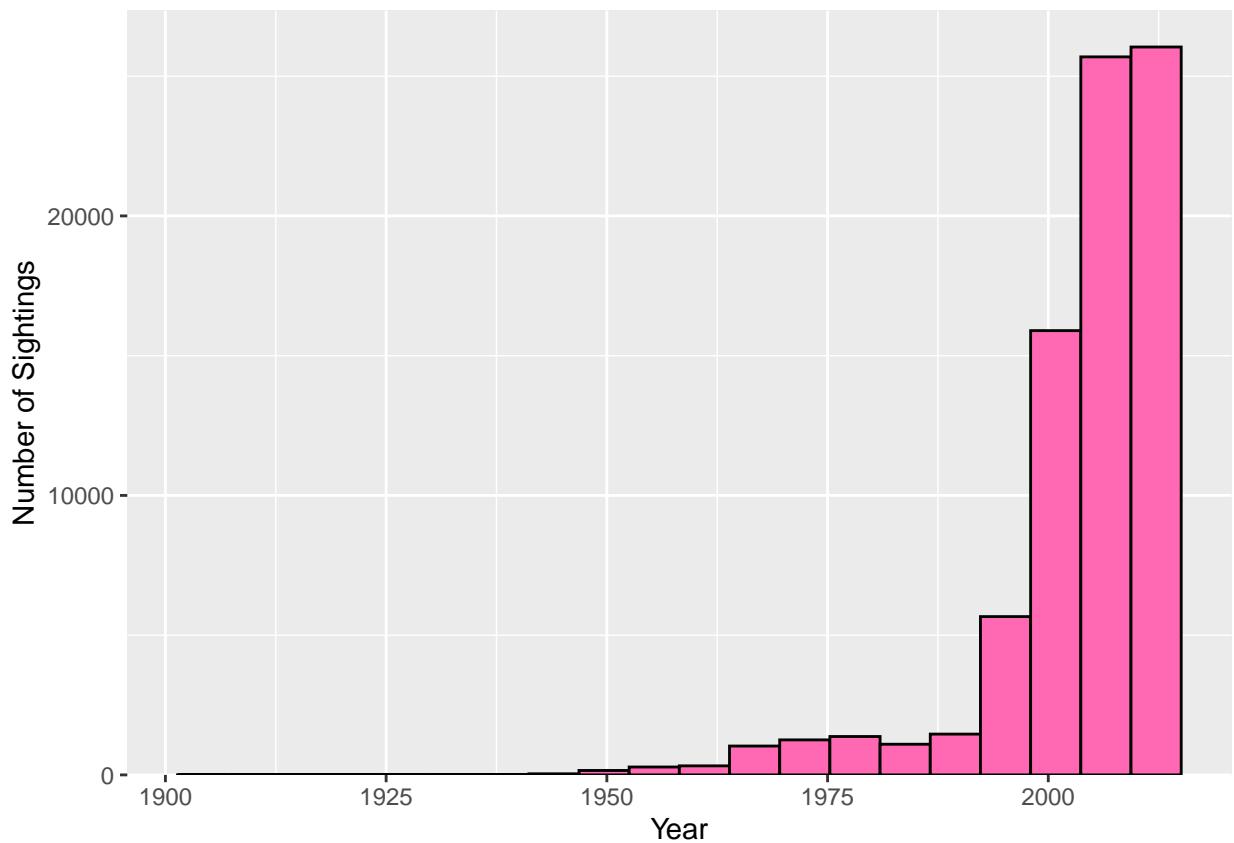
Total number of sightings per country



Observations * This graph supports the fact that the US, Canada, and Great Britain contain most of the sightings

Bar graph of sightings per year

```
ggplot( data = ufo,
        mapping = aes(x = year))+
  geom_histogram(fill = "hotpink",
                 color = "black",
                 bins = 20)+
  labs(
    x = "Year",
    y = "Number of Sightings"
  )+
  scale_y_continuous(expand = c(0, 0, 0.05, 0))
```



```
by_month <- ggplot( data = ufo,
                     mapping = aes(x = month))+
  geom_bar(fill = "purple",
            color = "black")+
  labs(
    x = "Month",
    y = "Number of Sightings"
  )+
  scale_y_continuous(expand = c(0, 0, 0.05, 0))+
  theme(
    axis.text = element_text( angle = 90
      )
  )
```

```
ufo_2000<- ufo|>
  filter(year >= 2000)
```

```
ggplot( data = ufo_2000,
        mapping = aes(x = month))+
  geom_bar(fill = "purple",
            color = "black")+
  labs(
    x = "Month",
    y = "Number of Sightings"
  )+
```

```

scale_y_continuous(expand = c(0, 0, 0.05, 0))+  

theme(  

  axis.text = element_text( angle = 90  

    )  

)+  

  facet_wrap(~ year)

```

