UFO Sightings

Peregrin Took

The UFO Sightings data available on Kaggle from National UFO Reporting Center (NUFORC) has been used for this purpose. This dataset contains over 80,000 reports of UFO sightings over the last century. Source: https://www.kaggle.com/NUFORC/ufo-sightings

UFOs(Unidentified Flying Objects), sometimes referred to as "flying saucers" are unusual or unconventional objects observed in the sky which are often claimed to be Extraterrestrial, or "alien" spacecrafts. Quite naturally, there's a lot of interest, folk-lore and conspiracy theories surrounding UFOs. Although most UFO observations have been later proven to be just misidentified mundane objects or natural henomena, the excitement still seems to be intact surrounding them.

So we take a look at the data to get an idea about the number of UFO sightings across the world, and throughout the years. The dataset contains two csv files, namely "complete.csv" and "scrubbed.csv".The complete data includes entries where the location of the sighting was not found or blank (0.8146%) or have an erroneous or blank time (8.0237%). Since the reports date back to the 20th century, some older data might be obscured. Data contains city, state, time, description, and duration of each sighting. We use the "complete.csv" file to read data into a data fram and remove the "NA" values for City and Datetime. However, we still kepe the data where Country or State value was missing, cause those are still valid sightings.

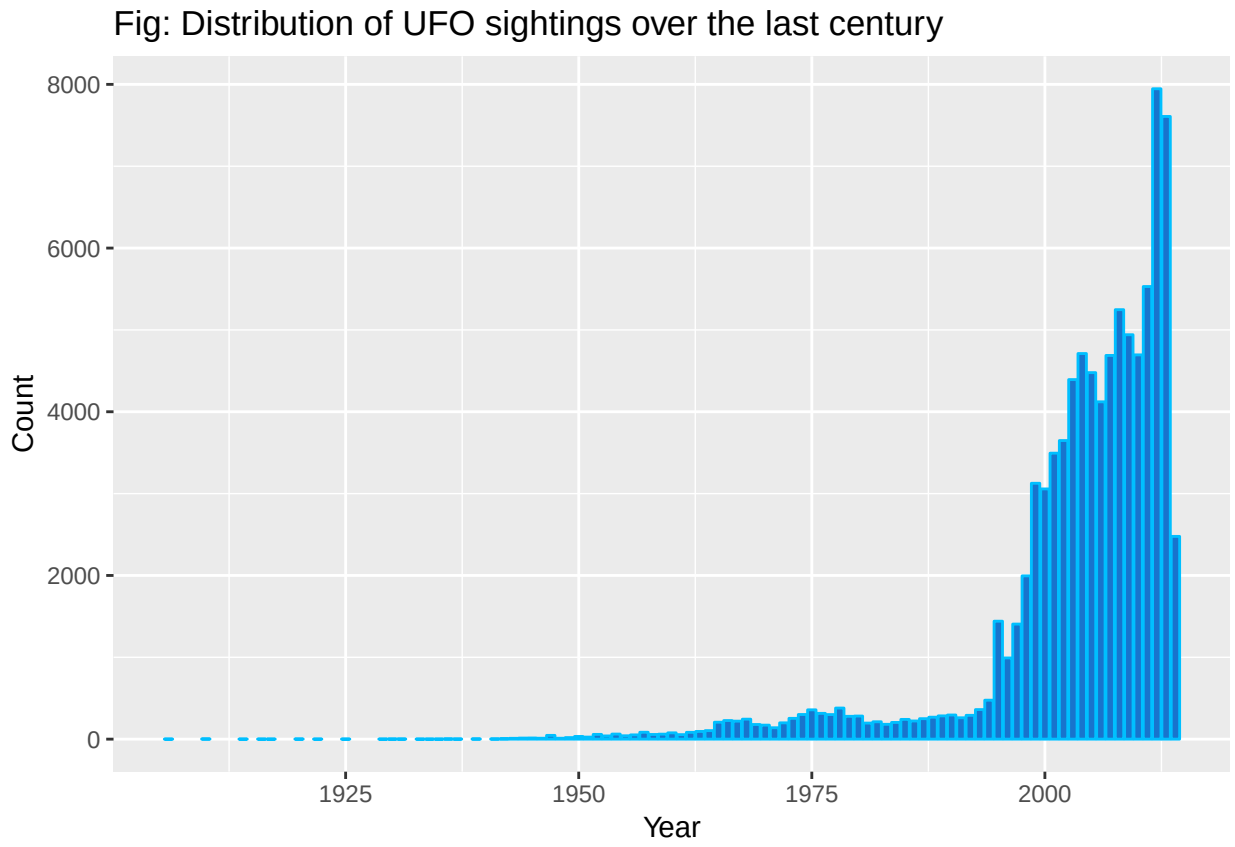
```
ufo_ <- read_csv("complete.csv")
ufo <- as.data.frame(ufo_)

print_ <- ufo %>%
  select(-comments)

head(print_, 10)
```

```
##            datetime                  city state country    shape
## 1  10/10/1949 20:30           san marcos    tx      us cylinder
## 2  10/10/1949 21:00         lackland afb    tx    <NA>    light
## 3  10/10/1955 17:00 chester (uk/england)  <NA>      gb   circle
## 4  10/10/1956 21:00                 edna    tx      us   circle
## 5  10/10/1960 20:00              kaneohe    hi      us    light
## 6  10/10/1961 19:00              bristol    tn      us   sphere
## 7  10/10/1965 21:00   penarth (uk/wales)  <NA>      gb   circle
## 8  10/10/1965 23:45              norwalk    ct      us     disk
## 9  10/10/1966 20:00            pell city    al      us     disk
## 10 10/10/1966 21:00             live oak    fl      us     disk
##    duration (seconds) duration (hours/min) date posted   latitude
## 1                2700           45 minutes   4/27/2004 29.8830556
## 2                7200              1-2 hrs  12/16/2005   29.38421
## 3                  20           20 seconds   1/21/2008       53.2
## 4                  20             1/2 hour   1/17/2004 28.9783333
## 5                 900           15 minutes   1/22/2004 21.4180556
## 6                 300            5 minutes   4/27/2007 36.5950000
## 7                 180         about 3 mins   2/14/2006  51.434722
## 8                1200           20 minutes   10/2/1999 41.1175000
## 9                 180            3  minutes   3/19/2009 33.5861111
## 10                120      several minutes   5/11/2005 30.2947222
##      longitude
```

```
## 1   -97.941111
## 2   -98.581082
## 3    -2.916667
## 4   -96.645833
## 5  -157.803611
## 6   -82.188889
## 7    -3.180000
## 8   -73.408333
## 9   -86.286111
## 10  -82.984167
```

The "coments" column has been removed for printing this table, as it contains large strings, printing outside of page.

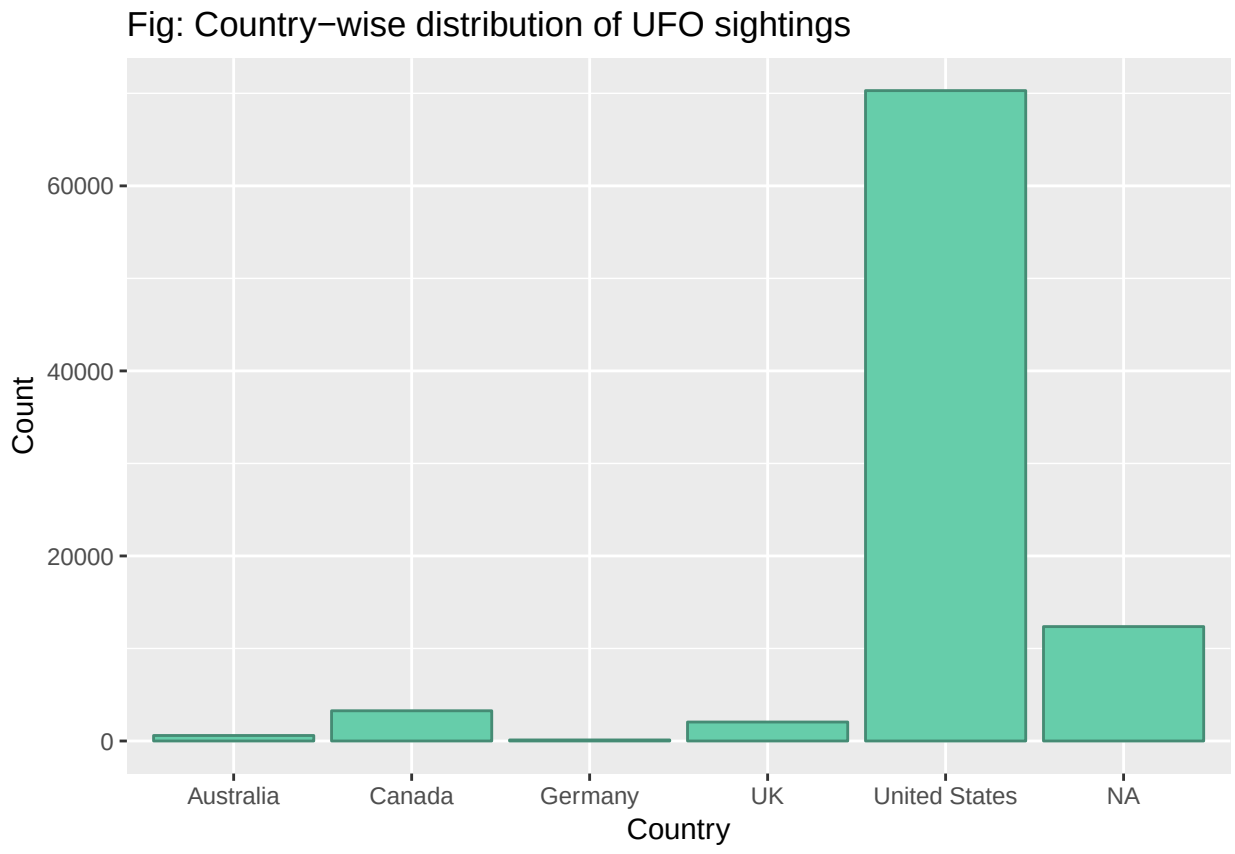
```
ufo %>%
  mutate(dates = year(mdy_hm(datetime))) %>%
  group_by(dates) %>%
  ggplot() +
  geom_bar(aes(x = dates), color = "deepskyblue", fill = "dodgerblue3") +
  labs(x = "Year", y = "Count",
       title = "Fig: Distribution of UFO sightings over the last century")
```



If we plot the number of sightings reported over the last century, we can see that before 1950 this number is negligible, however, we can see a rapid increase in the numbers after around 1985. This could be attributed to easier access to communication channels with the advancement of technology, and also the fact that

NUFORC was established in 1974, and since then has provided a 24-hour hotline phone number for people to report UFO activity that is currently going on in their area

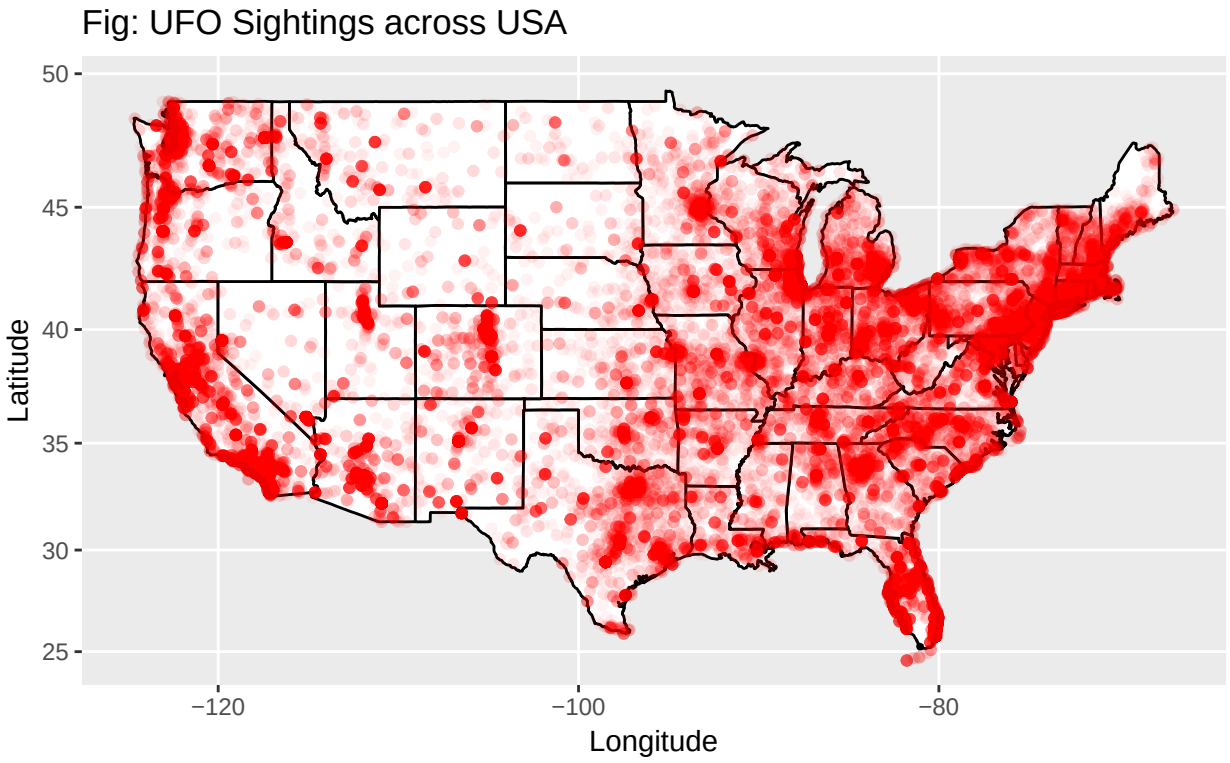
```
ufo %>%
  filter(!is.na(datetime),
         !is.na(city)) %>%
  mutate(country = recode(country,
                          "au" = "Australia",
                          "ca" = "Canada",
                          "gb" = "UK",
                          "de" = "Germany",
                          "us" = "United States"
                          )) %>%
  ggplot() +
  geom_bar(aes(x = country), color = "aquamarine4", fill = "aquamarine3") +
  labs(x = "Country", y = "Count",
       title = "Fig: Country-wise distribution of UFO sightings")
```



It's quite clear from this barplot that United States has more reported sightings than any other country. Hence it's worth checking out the distribution of sightings across United States.

```
ufo_us <- ufo %>%
  filter(country == "us") %>%
  filter(`duration (seconds)` > 0) %>%
```

```
  filter(latitude < 50 & latitude > 23  )
states <- map_data("state")
ggplot() +
  geom_polygon(data = states, aes(x = long, y = lat, group = group), color = "black",
               fill = "white") +
  geom_point(data = ufo_us, aes(x = as.numeric(longitude), y = as.numeric(latitude)),
             alpha = 0.05,
             color = "red")+
  coord_map()+
  labs(x = "Longitude", y = "Latitude", title = "Fig: UFO Sightings across USA")
```



If we plot the locations on the USA map where UFO sightings were reported. The density of the points is highest on the eastern side, i.e., most of the sigtings seem to have been reported from the Eastern parts of USA.