

Viz 3 Advanced Plots

Hayden Ginman

2023-09-11

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
## `summarise()` has grouped output by 'year'. You can override using the  
## `.groups` argument.
```

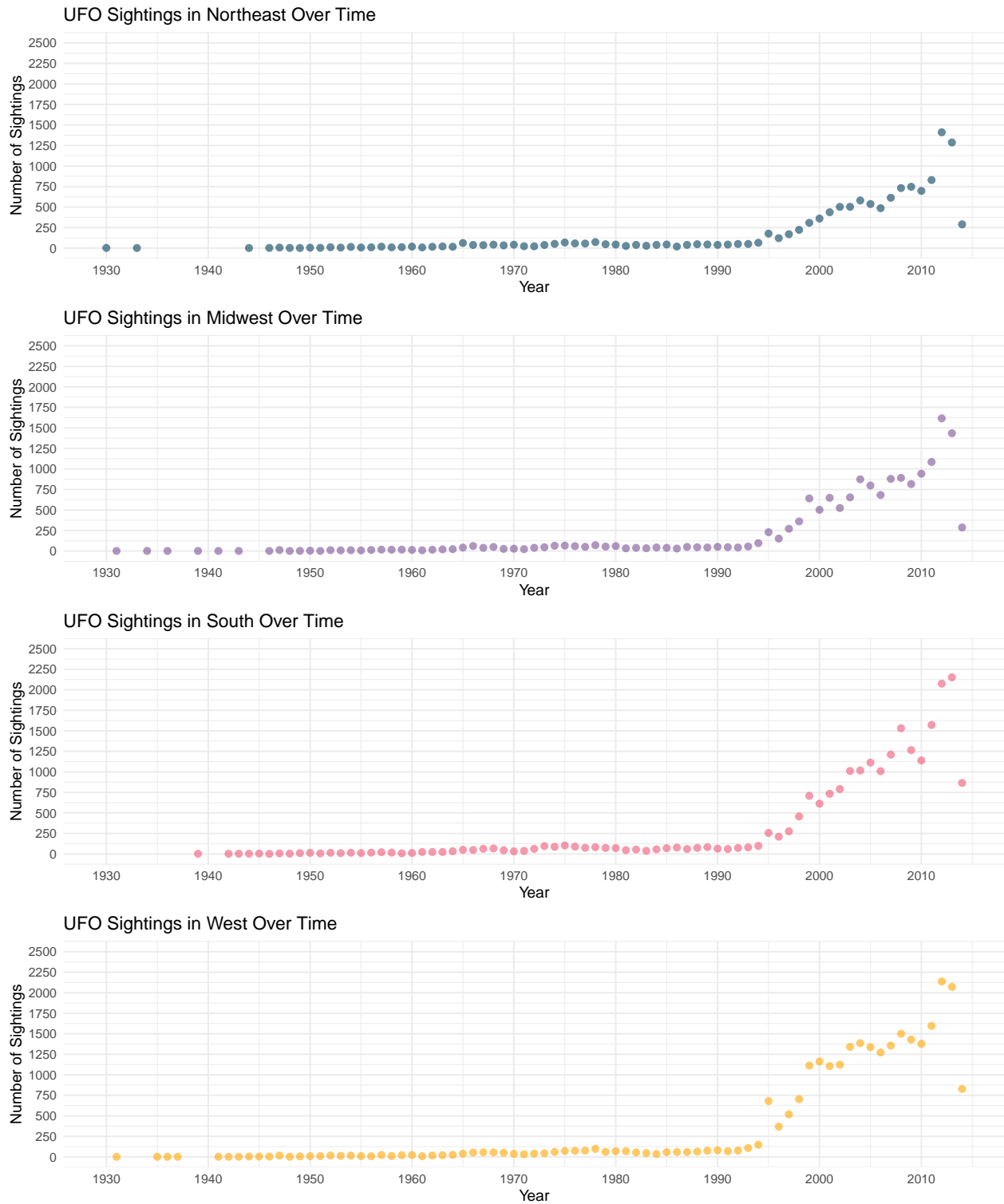


Figure 1: Yearly trends in UFO sightings Across U.S. Regions (1930-2014): This series of visualizations showcases the yearly count of reported UFO sightings across the four major U.S. regions from 1930 to 2014.

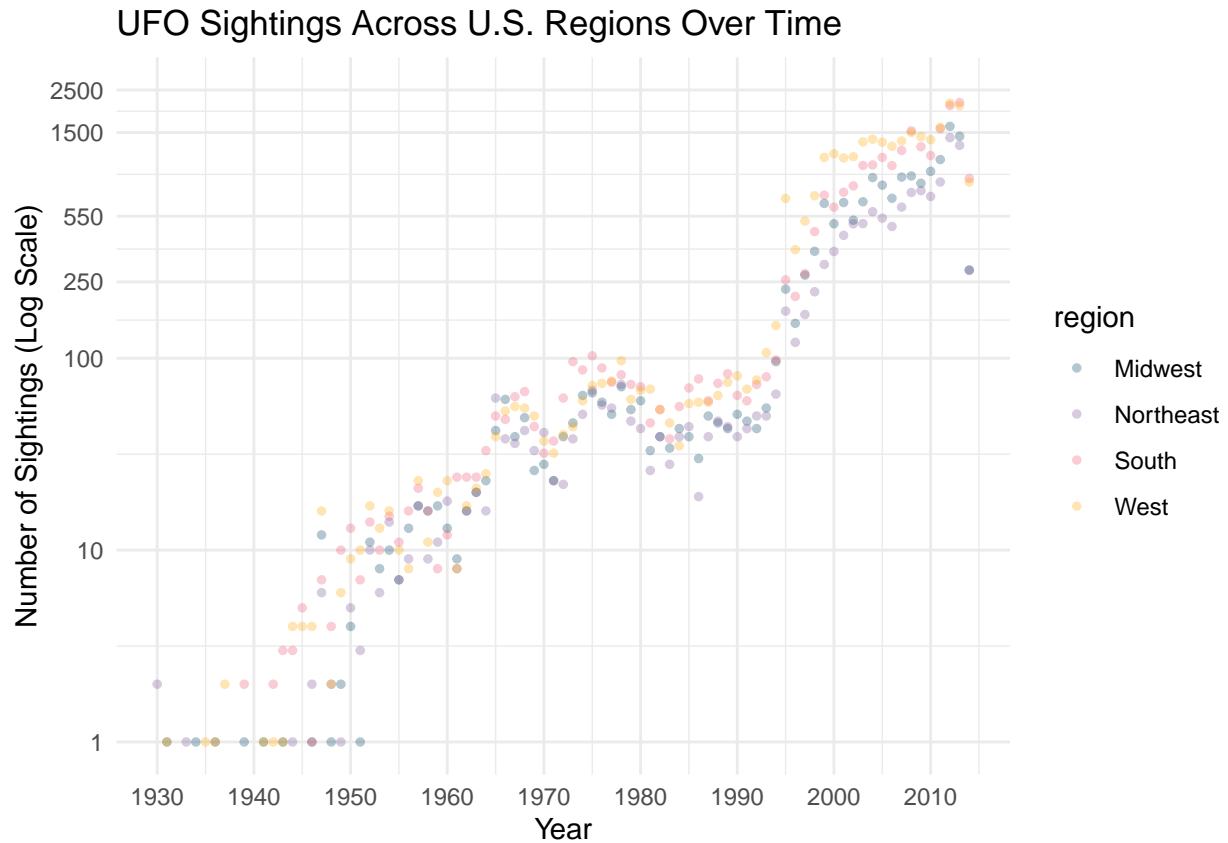


Figure 2: Yearly trends in UFO sightings Across U.S. Regions (1930-2014): This visualization showcases the yearly count of reported UFO sightings across the four major U.S. regions from 1930 to 2014 on a single plot. The Y-axis has been transformed using a logarithmic scale and has irregular breaks that make the data more difficult to interpret. Size of data points is small and opacity was reduced to make overlapping data point more difficult to interpret

```

#Step 1 - Load libraries into coding environment
options(repos = c(CRAN = "https://cloud.r-project.org/"))
#install.packages("gridExtra")
library("dplyr")
library("reshape2")
library("lubridate")
library("ggplot2")
library("gridExtra")

#Step 2 - Load cleaned UFO data into coding environment
# isolate the year in datetime column and make a new column called year

UFO <- read.csv("UFO Clean1.csv")

UFO$year <- format(as.Date(UFO$datetime, format = "%m/%d/%Y"), "%Y")

#Step 3 - Group states into regions

regions <- list(
  Northeast = c('me', 'nh', 'vt', 'ma', 'ri', 'ct', 'ny', 'pa', 'nj', 'dc'),
  Midwest = c('nd', 'sd', 'ne', 'ks', 'mo', 'ia', 'mn', 'wi', 'il', 'in', 'oh', 'mi'),
  South = c('tx', 'ok', 'ar', 'la', 'ky', 'tn', 'ms', 'al', 'wv', 'md', 'de', 'va', 'nc', 'sc', 'ga', 'fl'),
  West = c('id', 'mt', 'wy', 'nv', 'ut', 'co', 'az', 'nm', 'ak', 'wa', 'or', 'ca', 'hi')
)

#Step 4 - Assign states to a region

UFO <- UFO %>%
  mutate(
    region = case_when(
      state %in% regions$Northeast ~ "Northeast",
      state %in% regions$Midwest ~ "Midwest",
      state %in% regions$South ~ "South",
      state %in% regions$West ~ "West",
      TRUE ~ NA_character_
    )
  )

#Step 5 - Group data by year and region

UFO$year <- as.numeric(as.character(UFO$year))

UFO_summary <- UFO %>%
  filter(!is.na(region)) %>%
  group_by(year, region) %>%
  summarise(count = n())

#Step 5 - Make a scatter plot that depicts UFO sighting over time for each region

plot_Northeast <- UFO_summary %>%
  filter(region == "Northeast") %>%
  ggplot(aes(x = year, y = count)) +

```

```

geom_point(color = "#003f5c", size = 2, alpha = 0.6) +
labs(title = "UFO Sightings in Northeast Over Time",
     x = "Year",
     y = "Number of Sightings") +
scale_x_continuous(limits = c(1930, 2014), breaks = seq(1930, 2014, by = 10)) +
scale_y_continuous(limits = c(0, 2500), breaks = seq(0, 2500, by = 250)) +
theme_minimal()

plot_Midwest <- UFO_summary %>%
  filter(region == "Midwest") %>%
  ggplot(aes(x = year, y = count)) +
  geom_point(color = "#7a5195", size = 2, alpha = 0.6) +
  labs(title = "UFO Sightings in Midwest Over Time",
       x = "Year",
       y = "Number of Sightings") +
  scale_x_continuous(limits = c(1930, 2014), breaks = seq(1930, 2014, by = 10)) +
  scale_y_continuous(limits = c(0, 2500), breaks = seq(0, 2500, by = 250)) +
  theme_minimal()

plot_South <- UFO_summary %>%
  filter(region == "South") %>%
  ggplot(aes(x = year, y = count)) +
  geom_point(color = "#ef5675", size = 2, alpha = 0.6) +
  labs(title = "UFO Sightings in South Over Time",
       x = "Year",
       y = "Number of Sightings") +
  scale_x_continuous(limits = c(1930, 2014), breaks = seq(1930, 2014, by = 10)) +
  scale_y_continuous(limits = c(0, 2500), breaks = seq(0, 2500, by = 250)) +
  theme_minimal()

plot_West <- UFO_summary %>%
  filter(region == "West") %>%
  ggplot(aes(x = year, y = count)) +
  geom_point(color = "#ffa600", size = 2, alpha = 0.6) +
  labs(title = "UFO Sightings in West Over Time",
       x = "Year",
       y = "Number of Sightings") +
  scale_x_continuous(limits = c(1930, 2014), breaks = seq(1930, 2014, by = 10)) +
  scale_y_continuous(limits = c(0, 2500), breaks = seq(0, 2500, by = 250)) +
  theme_minimal()

```

#Step 6 - Combine plots of each region using gridExtra

```
grid.arrange(plot_Northeast, plot_Midwest, plot_South, plot_West, ncol = 1)
```

#Step 5 - Create an illusion plot

```

combined_plot <- ggplot(UFO_summary, aes(x = year, y = count, color = region)) +
  geom_point(size = 1, alpha = 0.3) +
  labs(title = "UFO Sightings Across U.S. Regions Over Time",
       x = "Year",
       y = "Number of Sightings (Log Scale)") +
  scale_x_continuous(limits = c(1930, 2014), breaks = seq(1930, 2014, by = 10)) +
  scale_y_log10(limits = c(1, 2500), breaks = c(1, 10, 100, 250, 550, 1500, 2500)) +

```

```
scale_color_manual(values = c("#003f5c", "#7a5195", "#ef5675", "#ffa600")) +  
theme_minimal()  
  
print(combined_plot)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.