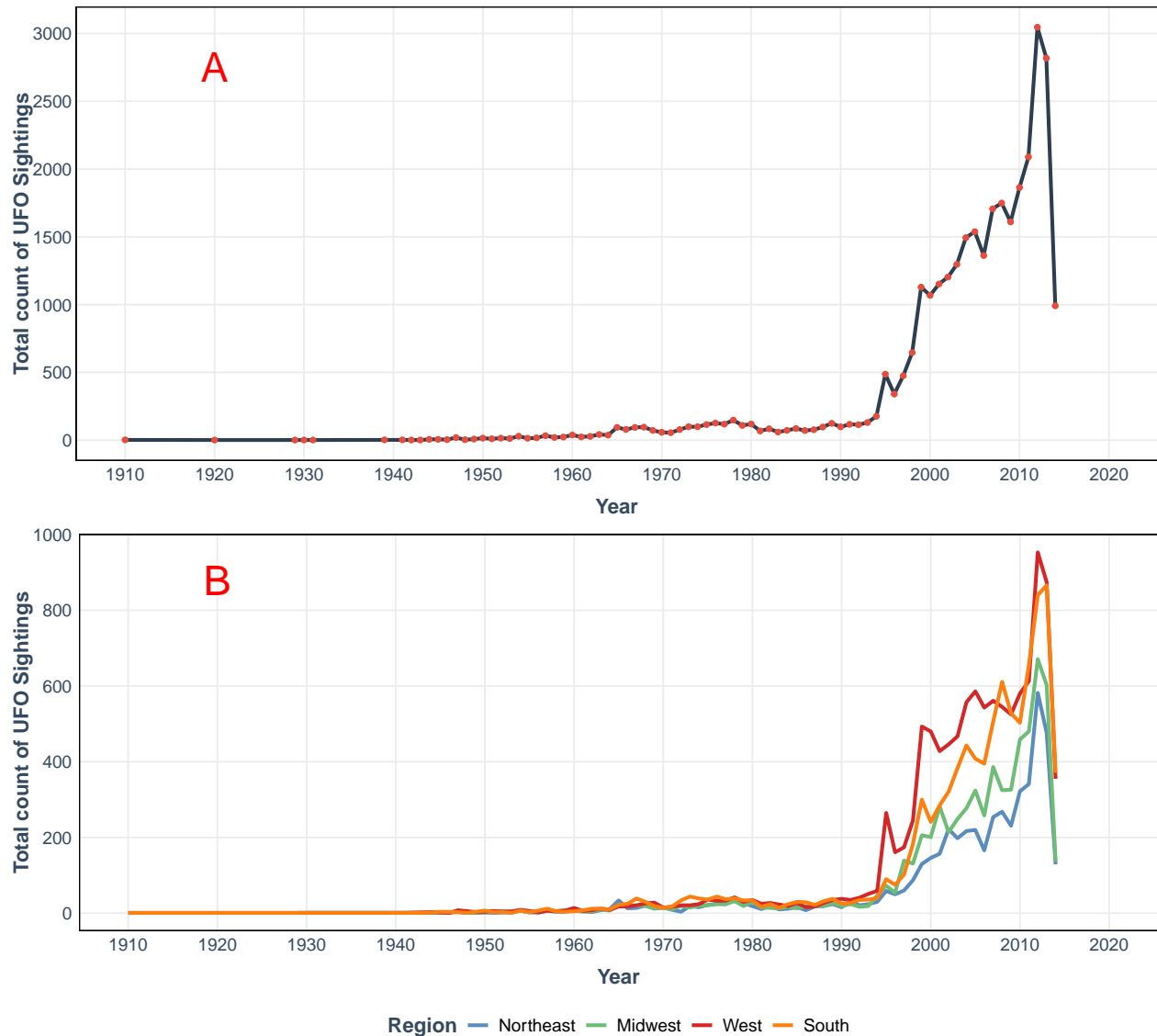
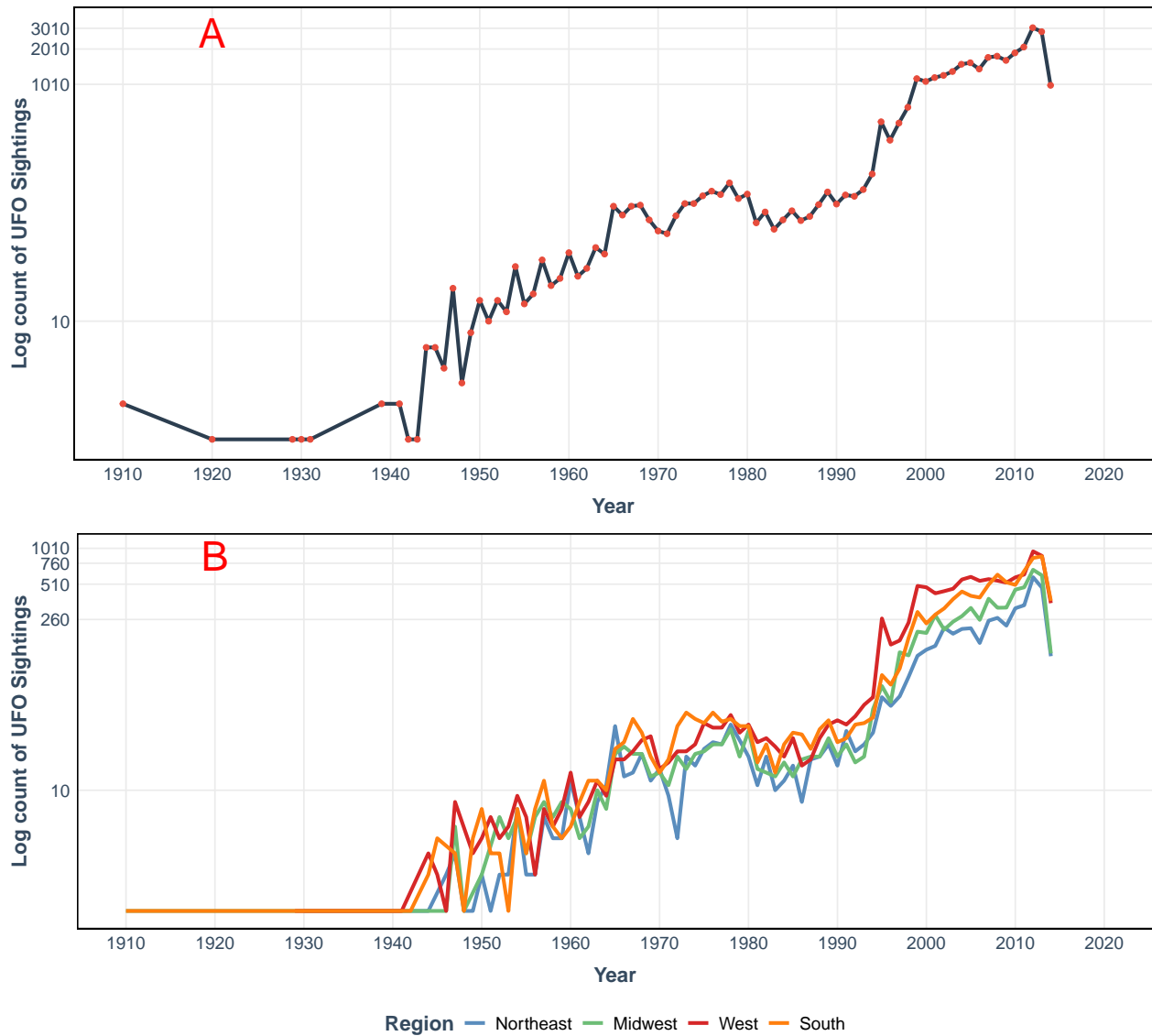


## Visualization - Task 2



**Figure 1.** An examination of UFO sightings spanning from 1910 to 2014. *Panel A* represents a line plot indicating a sharp upward surge in global sightings over time beginning in 1995, while *Panel B* provides a line plot with a regional breakdown within the United States, highlighting differences in sighting frequencies across the Northeast, Midwest, West, and South regions. Only sightings with documented dates were considered, with data sourced from the National UFO Reporting Center dataset on Kaggle.



**Figure 2.** An examination of UFO sightings spanning from 1910 to 2014. *Panel A* represents a line plot that illustrates a global trend in sightings over time, marked by noticeable fluctuations. Meanwhile *Panel B* provides a line plot with a regional breakdown within the United States, highlighting the varying frequencies of UFO sightings across the Northeast, Midwest, West, and South regions. Both panels employ a logarithmic scale on the y-axis for clarity. Only sightings with documented dates were considered, with data sourced from the National UFO Reporting Center dataset on Kaggle.

## Code Used To Make These Plots

```
#Importing the necessary libraries
library(ggplot2)
library(dplyr)
library(lubridate)
library(gridExtra)

##### DATA MANIPULATION CODE #####

# Loading the dataset as 'ufo'
ufo <- read.csv("https://raw.githubusercontent.com/riyanshibohra/data-visualization-using-R/refs/heads/main/ufo.csv")

# Extracting the year from the datetime column and saving it in a new column
ufo$datetime <- as.Date(ufo$datetime, format = "%d/%m/%Y %H:%M")
ufo$Year <- format(ufo$datetime, format = "%Y")

# Removing unnecessary columns from the dataframe
ufo <- ufo %>%
  select(-c(shape, duration..seconds.,
            duration..hours.min., comments, date.posted, latitude, longitude))

# List of US state abbreviations
us_state <- c("al", "ak", "az", "ar", "ca", "co", "ct", "de", "fl", "ga",
              "hi", "id", "il", "in", "ia", "ks", "ky", "la", "me", "md",
              "ma", "mi", "mn", "ms", "mo", "mt", "ne", "nv", "nh", "nj",
              "nm", "ny", "nc", "nd", "oh", "ok", "or", "pa", "ri", "sc",
              "sd", "tn", "tx", "ut", "vt", "va", "wa", "wv", "wi", "wy")

# Fill missing country values based on state
ufo <- ufo %>%
  mutate(country = ifelse(state %in% us_state & (is.na(country) | country == ""),
                          "us", country))

# US States divided by regions
west <- c("ak", "az", "ca", "co", "hi", "id", "mt", "nv", "nm", "or", "ut", "wa", "wy")
midwest <- c("il", "in", "ia", "ks", "mi", "mn", "mo", "ne", "nd", "oh", "sd", "wi")
south <- c("al", "ar", "de", "fl", "ga", "ky", "la", "md", "ms", "nc", "ok", "sc",
           "tn", "tx", "va", "wv")
northeast <- c("ct", "me", "ma", "nh", "nj", "ny", "pa", "ri", "vt")

# Assigning regions
ufo <- ufo %>%
  mutate(Region = case_when(
    state %in% west ~ "West",
    state %in% midwest ~ "Midwest",
    state %in% south ~ "South",
    state %in% northeast ~ "Northeast",
  ))

# Removing rows with NA values for "Region"
ufo <- ufo %>%
  filter(!is.na(Region))
```

```

# Calculating the count of UFO sightings for each unique year
# and storing the results in the 'ufo_yearly' dataframe.
ufo_yearly <- ufo %>%
  group_by(Year) %>%
  summarise(count = n())

# Reorder the factor levels for the 'Region' column geographically
ufo$Region <- factor(ufo$Region, levels = c("Northeast", "Midwest", "West", "South"))

# Removing rows where the column 'year' has NA values.
ufo_yearly <- ufo_yearly %>%
  filter(!is.na(Year))

# Calculating the count of UFO sightings for each unique year and region
# and storing the results in the 'ufo_region_yearly' dataframe.
ufo_region_yearly <- ufo %>%
  group_by(Year, Region) %>%
  summarise(count = n(), .groups = "drop") %>%
  filter(!is.na(Year))

##### PRETTY PLOT CODE #####

# Pretty line plot for ufo_yearly dataframe FOR 1910-2014 ----A
pretty_plotA <- ggplot(ufo_yearly, aes(x = as.numeric(Year), y = count)) +
  # Line
  geom_line(na.rm= TRUE, color = "#2C3E50", size = 1.2) +
  # Points
  geom_point(na.rm= TRUE, color = "#E74C3C", size = 1.5, shape = 19) +
  scale_x_continuous(breaks = seq(1910, 2020, by = 10),
                    limits = c(1910, 2020)) +
  scale_y_continuous(breaks = seq(0, 3000, by = 500)) +
  # Theme
  theme_minimal() +
  theme(
    axis.title.x = element_text(size = 14, face = "bold",
                                color = "#34495E",margin=margin(t=10)),
    axis.title.y = element_text(size = 14, face = "bold",
                                color = "#34495E",margin=margin(t=10)),
    axis.text.x = element_text(angle = 0, hjust = 0.5,
                                size = 12, color = "#34495E",),
    axis.text.y = element_text(size = 12, color = "#34495E",
                                margin=margin(t=10)),
    panel.grid.minor = element_blank()
  ) +
  annotate("text",x=1920,y=2750,
         label = "A",
         col="red",
         size= 10)+
  labs(
    x = "Year",
    y = "Total count of UFO Sightings"
  )

```

```

pretty_plotA <- pretty_plotA +
  annotation_custom(
    grob = grid::rectGrob(gp = grid::gpar(fill = NA, lwd = 2)),
    xmin = -Inf, xmax = Inf, ymin = -Inf, ymax = Inf
  )

# Pretty line plot for ufo_region_yearly dataframe FOR 1910-2014 ----B
pretty_plotB <- ggplot(ufo_region_yearly, aes(x = as.numeric(Year), y = count, color = Region)) +
  # Line
  geom_line(size = 1.2, aes(group = Region), na.rm = TRUE) +
  scale_x_continuous(breaks = seq(1910, 2020, by = 10), limits = c(1910, 2020)) +
  scale_y_continuous(breaks = seq(0, 1000, by = 200)) +
  scale_color_manual(values = c("Midwest" = "#6DBD76", "South" = "#ff7f0e",
                                "Northeast" = "#5A8DBD", "West" = "#d62728")) +

  # Theme
  theme_minimal() +
  theme(
    axis.title.x = element_text(size = 14, face = "bold",
                                color = "#34495E",margin=margin(t=10)),
    axis.title.y = element_text(size = 14, face = "bold",
                                color = "#34495E",margin=margin(t=20)),
    axis.text.x = element_text(angle = 0, hjust = 0.5,
                                size = 12, color = "#34495E"),
    axis.text.y = element_text(size = 12, color = "#34495E"),
    panel.grid.minor = element_blank(),
    legend.position = "bottom",
    legend.text = element_text(size = 12),
    legend.title = element_text(size = 14,color = "#34495E",face = "bold"),
  ) +
  annotate("text",x=1920,y=880,
          label = "B",
          col="red",
          size= 10)+
  labs(
    x = "Year",
    y = "Total count of UFO Sightings",
    color = "Region"
  )
pretty_plotB <- pretty_plotB +
  annotation_custom(
    grob = grid::rectGrob(gp = grid::gpar(fill = NA, lwd = 2)),
    xmin = -Inf, xmax = Inf, ymin = -Inf, ymax = Inf
  )

grid.arrange(pretty_plotA, pretty_plotB, ncol = 1, heights = c(1, 1))

##### ILLUSION PLOT CODE #####

# Illusion line plot FOR 1910-2014 ----A
illusion_plotA <- pretty_plotA +
  scale_y_continuous(trans = 'log10', breaks = seq(10, 3500, by = 1000)) +
  labs(
    y = "Log count of UFO Sightings"
  )

```

```

illusion_plotA <- illusion_plotA +
  annotation_custom(
    grob = grid::rectGrob(gp = grid::gpar(fill = NA, lwd = 2)),
    xmin = -Inf, xmax = Inf, ymin = -Inf, ymax = Inf
  )

# Illusion line plot FOR 1910-2014 ----B
illusion_plotB <- pretty_plotB +
  scale_y_continuous(trans = 'log10', breaks = seq(10, 1010, by = 250)) +
  labs(
    y = "Log count of UFO Sightings"
  )
illusion_plotB <- illusion_plotB +
  annotation_custom(
    grob = grid::rectGrob(gp = grid::gpar(fill = NA, lwd = 2)),
    xmin = -Inf, xmax = Inf, ymin = -Inf, ymax = Inf
  )

grid.arrange(illusion_plotA, illusion_plotB, ncol = 1)

```