A2 Assignment - Refugee Analysis

MBAN - Group 7

Team Members:

Abi Joshua GEORGE / Eri YOSHIMOTO / Hakeem GARCIA
Nattida TAVAROJN / Neha NAGABHUSHAN / Weikang YANG

Introduction

Refugee migration has been a significant global issue, with various countries experiencing fluctuations in the number of people seeking asylum. This report analyzes refugee data from multiple countries over a span of years to identify trends, patterns, and key insights. The study involves data cleaning, transformation, visualization, and statistical analysis to gain a deeper understanding of global refugee movements.

Data Preparation and Cleaning

The dataset initially contained raw refugee statistics, which required significant cleaning before analysis.

```r
library(tidyverse)
library(readr)

# read data from database
raw_df <- read_csv("data/A2_refugee_status.csv", col_types = cols(.default = "c"))
print(raw_df)
```

```
## # A tibble: 70 x 11
##    Continent/Country of Natio~1 ‘2006‘ ‘2007‘ ‘2008‘ ‘2009‘ ‘2010‘ ‘2011‘ ‘2012‘
##    <chr>                        <chr>  <chr>  <chr>  <chr>  <chr>  <chr>  <chr>
##  1 Africa                       18,129 17,486 8,943  9,678  13,325 7,693  10,629
##  2 Asia                         10,086 23,564 44,819 58,309 52,695 44,583 44,416
##  3 Europe                       9,615  4,192  2,059  1,693  1,238  996    908
##  4 North America                3,145  2,922  4,177  4,800  4,856  2,930  1,948
##  5 Oceania                      -      -      -      -      -      -      -
##  6 South America                119    54     100    57     126    46     130
##  7 Unknown                      -      -      9      65     1,053  136    148
##  8 Afghanistan                  651    441    576    349    515    428    481
##  9 Angola                       13     4      -      8      -      D      -
## 10 Armenia                      87     29     9      4      D      15     8
## # i 60 more rows
## # i abbreviated name: 1: ‘Continent/Country of Nationality‘
## # i 3 more variables: ‘2013‘ <chr>, ‘2014‘ <chr>, ‘2015‘ <chr>
```

Handling Missing and Inconsistent Values

The dataset contained placeholders for missing values, such as "D," "X," and "-", which were converted to "NA." Numeric values were reformatted by removing commas and converting them into numerical data types.

```
# set cleaned data frame as df in later operation
# Set non-numerical value("D", "X", "-")as "NA"
raw_df[raw_df == "D" | raw_df == "X" | raw_df == "-"] <- NA
# set the value column as numbers value and delete comma and transfer to numeric value
raw_df[ , -1] <- lapply(raw_df[ , -1], function(x) as.numeric(gsub(",", "", x)))

print(raw_df)
```

```
## # A tibble: 70 x 11
##    Continent/Country of Natio~1 `2006` `2007` `2008` `2009` `2010` `2011` `2012`
##    <chr>                         <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
##  1 Africa                        18129  17486   8943   9678  13325   7693  10629
##  2 Asia                          10086  23564  44819  58309  52695  44583  44416
##  3 Europe                         9615   4192   2059   1693   1238    996    908
##  4 North America                  3145   2922   4177   4800   4856   2930   1948
##  5 Oceania                          NA     NA     NA     NA     NA     NA     NA
##  6 South America                   119     54    100     57    126     46    130
##  7 Unknown                          NA     NA      9     65   1053    136    148
##  8 Afghanistan                     651    441    576    349    515    428    481
##  9 Angola                           13      4     NA      8     NA     NA     NA
## 10 Armenia                          87     29      9      4     NA     15      8
## # i 60 more rows
## # i abbreviated name: 1: `Continent/Country of Nationality`
## # i 3 more variables: `2013` <dbl>, `2014` <dbl>, `2015` <dbl>
```

Standardizing Country Names, pivot data, add Year_Date format data

Countries with alternative naming conventions, such as "China, People's Republic" and "Korea, North," were renamed to "China" and "North Korea" for consistency. Also add Year_Data format column for further analysis.

```
library(countrycode)
library(lubridate)

# Remove unwanted categories (Unknown, Other, Total)
remove_list <- c("Unknown", "Other", "Total")
continent_list <- c("Africa", "Asia", "Europe",
                    "North America", "Oceania", "South America")

# Convert Data to Longformat

long_df <- raw_df %>%
  pivot_longer(cols = -`Continent/Country of Nationality`,
               names_to = "Year", values_to = "Total_Refugees")

# Extract Country Data (Exclude Continent & Unwanted Categories)
country_df <- long_df %>%
```

```r
  filter(!(`Continent/Country of Nationality` %in%
             c(continent_list, remove_list))) %>%
  rename(Country = `Continent/Country of Nationality`) %>%
  mutate(
    # Convert Country Names to ISO3 Codes (No Manual Entry Needed!)
    ISO3 = countrycode(Country, "country.name", "iso3c"),

    # Convert Back to Standard Country Names (Handles Variations Automatically)
    Country = countrycode(ISO3, "iso3c", "country.name"),

    # Convert Year to Integer and Date Format
    Year = as.integer(Year),
    Year_Date = ymd(paste0(Year, "-01-01"))  # Convert to YYYY-MM-DD
  ) %>%
  select(Country, Year, Year_Date, Total_Refugees)  # Add Year_Date column

# Search for unmatched country
unmatched_countries <- country_df %>% filter(is.na(Country))

# Print unmatched country name
if (nrow(unmatched_countries) > 0) {
  print("Unmatched Countries (Check for Manual Fix):")
  print(unique(unmatched_countries$Country))
} else {
  print("All countries successfully converted!")
}
```

```
## [1] "All countries successfully converted!"
```

```r
# Check Country DataFrame
print(country_df)
```

```
## # A tibble: 600 x 4
##    Country      Year Year_Date  Total_Refugees
##    <chr>       <int> <date>              <dbl>
##  1 Afghanistan  2006 2006-01-01            651
##  2 Afghanistan  2007 2007-01-01            441
##  3 Afghanistan  2008 2008-01-01            576
##  4 Afghanistan  2009 2009-01-01            349
##  5 Afghanistan  2010 2010-01-01            515
##  6 Afghanistan  2011 2011-01-01            428
##  7 Afghanistan  2012 2012-01-01            481
##  8 Afghanistan  2013 2013-01-01            661
##  9 Afghanistan  2014 2014-01-01            753
## 10 Afghanistan  2015 2015-01-01            910
## # i 590 more rows
```

Extract Continents Information, pivot data, add Year_Date format data

Same as Countries data cleaning, extract continents information from raw_df. then pivot the data. Also add Year_Data format column for further analysis.

```r
# Make Continent List
continent_list <- c("Africa", "Asia", "Europe", "North America", "Oceania", "South America")

# convert long-format
continent_tidy_df <- raw_df %>%
  pivot_longer(cols = -`Continent/Country of Nationality`,
               names_to = "Year",
               values_to = "Total_Refugees")

# Make Continent Dataframe
continent_df <- continent_tidy_df %>%
  filter(`Continent/Country of Nationality` %in% continent_list) %>%
  rename(Continent = `Continent/Country of Nationality`) %>%
  mutate(
    Year = as.integer(Year),
    Year_Date = ymd(paste0(Year, "-01-01"))
    ) %>%
  select(Continent, Year, Year_Date, Total_Refugees)

# Continent dataframe check
print(continent_df)
```

```
## # A tibble: 60 x 4
##    Continent  Year Year_Date  Total_Refugees
##    <chr>     <int> <date>              <dbl>
##  1 Africa     2006 2006-01-01          18129
##  2 Africa     2007 2007-01-01          17486
##  3 Africa     2008 2008-01-01           8943
##  4 Africa     2009 2009-01-01           9678
##  5 Africa     2010 2010-01-01          13325
##  6 Africa     2011 2011-01-01           7693
##  7 Africa     2012 2012-01-01          10629
##  8 Africa     2013 2013-01-01          15984
##  9 Africa     2014 2014-01-01          17501
## 10 Africa     2015 2015-01-01          22492
## # i 50 more rows
```

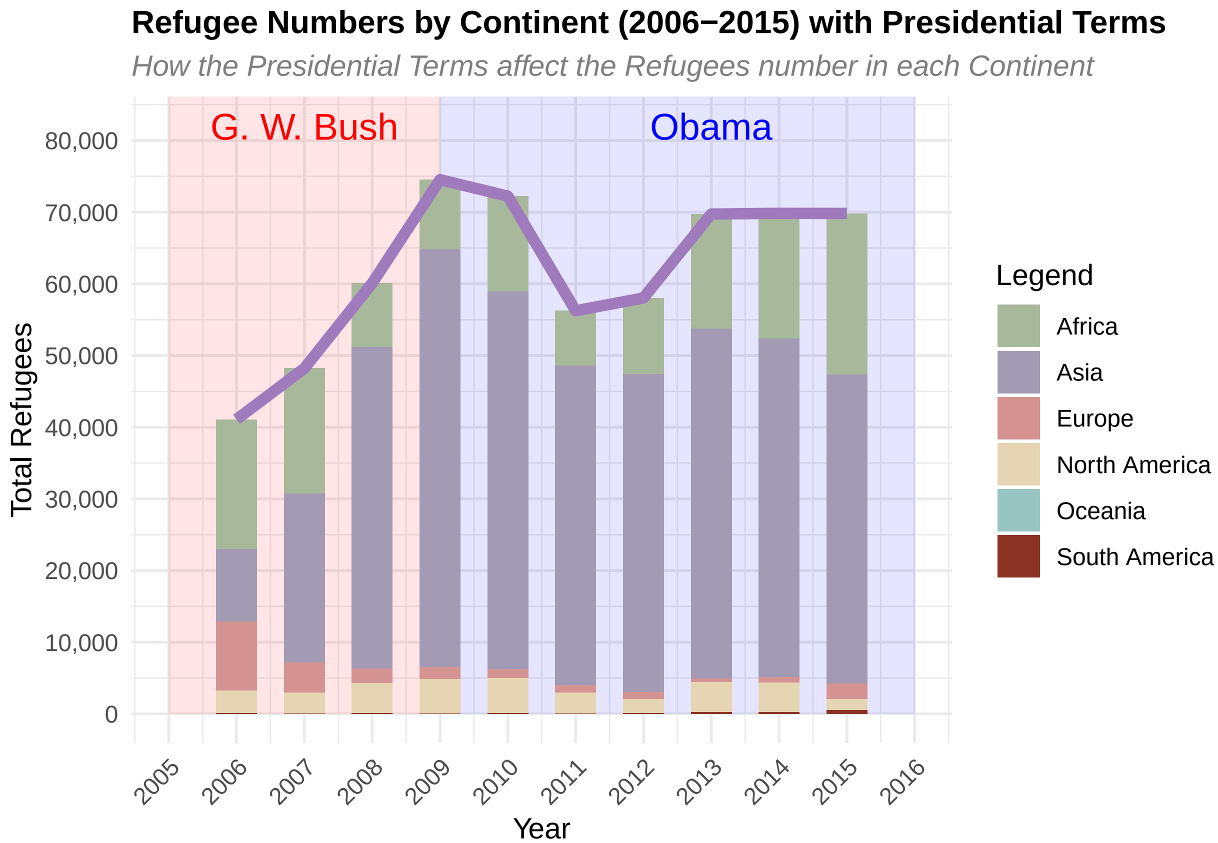
Data Analysis and Visualization

Visualization #1: Refugee Number by Continent (Stacked Bar Chart)

```r
# Load libraries
library(ggplot2)
library(dplyr)
library(scales)
library(svglite)

# Calculate the number of refugees for each continents by year

# Add US President data
```

```r
presidents <- data.frame(
  President = c("G. W. Bush", "Obama"),
  Start_Year = c(2005, 2009),
  End_Year = c(2009, 2016)
)

# Make summary data of year and total refugee number

yearly_refugees <- continent_df %>%
  group_by(Year) %>%
  summarise(Total_Refugees = sum(Total_Refugees, na.rm = TRUE))

#Plotting
p <-ggplot() +

# Presidential data
  geom_rect(aes(xmin = 2005, xmax = 2009, ymin = 0, ymax = Inf),
            fill = "red", alpha = 0.1) +
  geom_rect(aes(xmin = 2009, xmax = 2016, ymin = 0, ymax = Inf),
            fill = "blue", alpha = 0.1) +

# President Labels
  annotate("text", x = 2007, y = max(yearly_refugees$Total_Refugees) * 1.10,
           label = "G. W. Bush", face = "bold", color = "red", size = 5) +
  annotate("text", x = 2013, y = max(yearly_refugees$Total_Refugees) * 1.10,
           label = "Obama", face = "bold", color = "blue", size = 5) +


#  Refugees by continents stacked bar, ignore NA for plotting
  geom_bar(data = continent_df,
           aes(x = Year, y = Total_Refugees, fill = `Continent`),
           stat = "identity", position = "stack", width = 0.6, na.rm = TRUE) +

# Total Refugees line chart
  geom_line(data = yearly_refugees, aes(x = Year, y = Total_Refugees, group = 1),
            color = "#9f7abc", size = 2) +
# color settings
  scale_fill_manual(
    name = "Legend",
    values = c(
      "Africa" = "#a8b89a",
      "Asia" = "#a29bb3",
      "Europe" = "#d49391",
      "North America" = "#e5d5b2",
      "Oceania" = "#98c4c1",
      "South America" = "#8A3324"
    ),
    guide = guide_legend(override.aes = list(alpha = 1))
  ) +

# Axis Settings
  scale_y_continuous(
    labels = scales::comma,
```

```r
    breaks = scales::pretty_breaks(n = 10)) +
  scale_x_continuous(limits = c(2005, 2016), breaks = 2005:2016) +

# Labels and  Theme
  theme_minimal() +
  labs(
    title = "Refugee Numbers by Continent (2006-2015) with Presidential Terms",
    subtitle = "How the Presidential Terms affect the Refugees number in each Continent",
    x = "Year",
    y = "Total Refugees",

  ) +
  theme(
    plot.title = element_text(face = "bold", size = 12),
    plot.subtitle = element_text(size = 11, face = "italic",color = "gray50"),
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.caption = element_text(size = 8, hjust = 0, color = "gray50"),
    legend.position = "right"
  )

# Saving figures

if (!dir.exists("output_plot")) dir.create("output_plot", showWarnings = FALSE)
if (!dir.exists("png")) dir.create("png", showWarnings = FALSE)

ggsave("./output_plot/Refugee Numbers by Continent with Presidential Terms.svg",
       width = 10, height = 6, dpi = 300, device = "svg")
ggsave("./png/Refugee Numbers by Continent with Presidential Terms.png",
       width = 10, height = 6, dpi = 300)

p
```



This illustration represents the flow of refugee admissions into the United States, from 2006 through 2015, by continent of origin. It also overlays U.S. presidential terms on top to explore possible correlation between trends of refugee admission and which political administration it falls under. The stacked bar graph also shows the shifting regional composition of refugees through time, and the background shading distinguishes between the presidencies of George W. Bush and Barack Obama.

Key findings from the chart:

- Refugee admissions increased greatly from 2006 to 2009, at the end of the Bush presidency and the start of the Obama presidency.
- Asia was constantly the leading source region for refugees, Africa is the second source region.
- The number of refugees declined in 2011 but stabilized in subsequent years.
- By overlaying the presidential terms, the chart sparks a debate on whether U.S. refugee policy shifts align with changes in political leadership.

CRAP:
- Contrast: It is shown with different colors used to distinguish continents for clarity of regional differences, while the background shading further contrasts the presidential terms.
- Repetition: Bars follow a uniform format every year; this makes trends easier to identify.
- Alignment: The visualization has kept a clear timeline along the x-axis, matching presidential terms with refugee numbers for ease of interpretation.
- Proximity: The legend is placed close to the visualization, in order to ensure accessibility without

cluttering the main chart.

Kieran Healy's principles:
- Truthfulness: The visualization accurately represents refugee trends using reliable government data.
- Functionality: The stacked bar format allows for a comparative view of continental refugee distribution while maintaining readability.
- Beauty: The chart has a well-balanced color scheme and grid layout which improves clarity.
- Insightfulness: The overlay of presidential terms provides additional context, making viewers consider the role of political leadership in shaping refugee admissions.
- Enlightenment: This visualization encourages critical thinking about policy trends and regional shifts in global refugee movements.

Divide the data by Continenet

```r
# Loar Libraries
library(dplyr)
library(countrycode)

# Step 1 Add continents information to country_df with countrycode()
data_long <- country_df %>%
  mutate(
    Continent = countrycode(Country, "country.name", "continent")
  )

print(data_long)
```

```
## # A tibble: 600 x 5
##    Country      Year Year_Date  Total_Refugees Continent
##    <chr>       <int> <date>              <dbl> <chr>
##  1 Afghanistan  2006 2006-01-01            651 Asia
##  2 Afghanistan  2007 2007-01-01            441 Asia
##  3 Afghanistan  2008 2008-01-01            576 Asia
##  4 Afghanistan  2009 2009-01-01            349 Asia
##  5 Afghanistan  2010 2010-01-01            515 Asia
##  6 Afghanistan  2011 2011-01-01            428 Asia
##  7 Afghanistan  2012 2012-01-01            481 Asia
##  8 Afghanistan  2013 2013-01-01            661 Asia
##  9 Afghanistan  2014 2014-01-01            753 Asia
## 10 Afghanistan  2015 2015-01-01            910 Asia
## # i 590 more rows
```

```r
# Step 2: Manually classify certain countries as North America
data_long <- data_long %>%
  mutate(
    Continent = ifelse
    (Country %in% c("Haiti", "Cuba", "Honduras"), "North America", Continent)
  )

# Step 3: Exclude countries from Oceania and South America
data_long <- data_long %>%
```

```r
  filter(!Continent %in% c("Oceania", "Americas")
         )

# Step 4: Define a function to get top 5 countries by total refugees for each continent
top_5_per_continent <- function(data, continent, rank) {
  data %>%
    filter(Continent == continent) %>%
    group_by(Country) %>%
    summarise(Total_Refugees = sum(Total_Refugees, na.rm = TRUE)) %>%
    arrange(desc(Total_Refugees)) %>%
    slice_head(n = rank) %>%
    inner_join(data, by = "Country") %>%
    select(-Total_Refugees.x)
}

# Step 5: Get the top 5 countries for each continent
asia_data <- top_5_per_continent(data_long, "Asia", 5)
africa_data <- top_5_per_continent(data_long, "Africa", 5)
north_america_data <- top_5_per_continent(data_long, "North America", 3)
europe_data <- top_5_per_continent(data_long, "Europe", 5)
```

Visualization #2: Top 5 Refugee countries by Continent (For all 4 Line Graphs)

```r
# Step 6: Define custom colors for countries
custom_colors <- c(
  # Asia
  "Bhutan" = "#a8b89a", "Iraq" = "#a29bb3", "Burma" = "#d49391",
  "Vietnam" = "#e5d5b2", "Iran" = "#8A3324",

  # Africa
  "Burundi" = "#a8b89a", "Somalia" = "#a29bb3", "Congo" = "#d49391",
  "Sudan" = "#e5d5b2", "Eritrea" = "#8A3324",

  # North America
  "Cuba" = "#a8b89a", "Honduras" = "#a29bb3", "Haiti" = "#d49391",

  # Europe
  "Belarus" = "#a8b89a", "Russia" = "#a29bb3", "Latvia" = "#d49391",
  "Ukraine" = "#e5d5b2", "Moldova" = "#8A3324"
)

# Step 7: Adjust plot dimensions for a single column layout
options(repr.plot.width = 15, repr.plot.height = 30)

# Step 8: Modify the plot_continent function to include custom line colors
plot_continent <- function(data, title) {
  ggplot(data, aes(x = Year_Date, y = Total_Refugees.y,
                   color = Country, group = Country)) +
    geom_line(size = 1) +
    geom_point(size = 3, shape = 16) +
    theme_minimal() +
```

```
    labs(title = title, x = "Year", y = "Number of Refugees", color = NULL,
   ) +
    theme(
      plot.title = element_text(face = "bold", size = 12),
      legend.position = "bottom",
      legend.text = element_text(size = 8),
      plot.caption = element_text(size = 8, hjust = 0, color = "gray50"),
      axis.text.x = element_text(angle = 45, hjust = 1,)
    ) +
    scale_x_date(date_labels = "%Y", date_breaks = "1 year")+
    scale_y_continuous(
      labels = scales::comma,
      breaks = scales::pretty_breaks(n = 5)
    ) +
    scale_color_manual(values = custom_colors)
}

# Step 9: Create plots for each continent
plot_asia <- plot_continent(asia_data, "Top 5 Refugee Countries in Asia")
plot_africa <- plot_continent(africa_data, "Top 5 Refugee Countries in Africa")
plot_north_america <- plot_continent(north_america_data, "Top 5 Refugee Countries in North America")
plot_europe <- plot_continent(europe_data, "Top 5 Refugee Countries in Europe")

# Step 11: Export figures

continent_names <- c("asia", "africa", "north_america", "europe")
plots <- list(plot_asia, plot_africa, plot_north_america, plot_europe)


lapply(seq_along(continent_names), function(i) {
  ggsave(paste0("output_plot/plot_", continent_names[i], ".svg"),
         plots[[i]], width = 10, height = 6, dpi = 300)
})
```

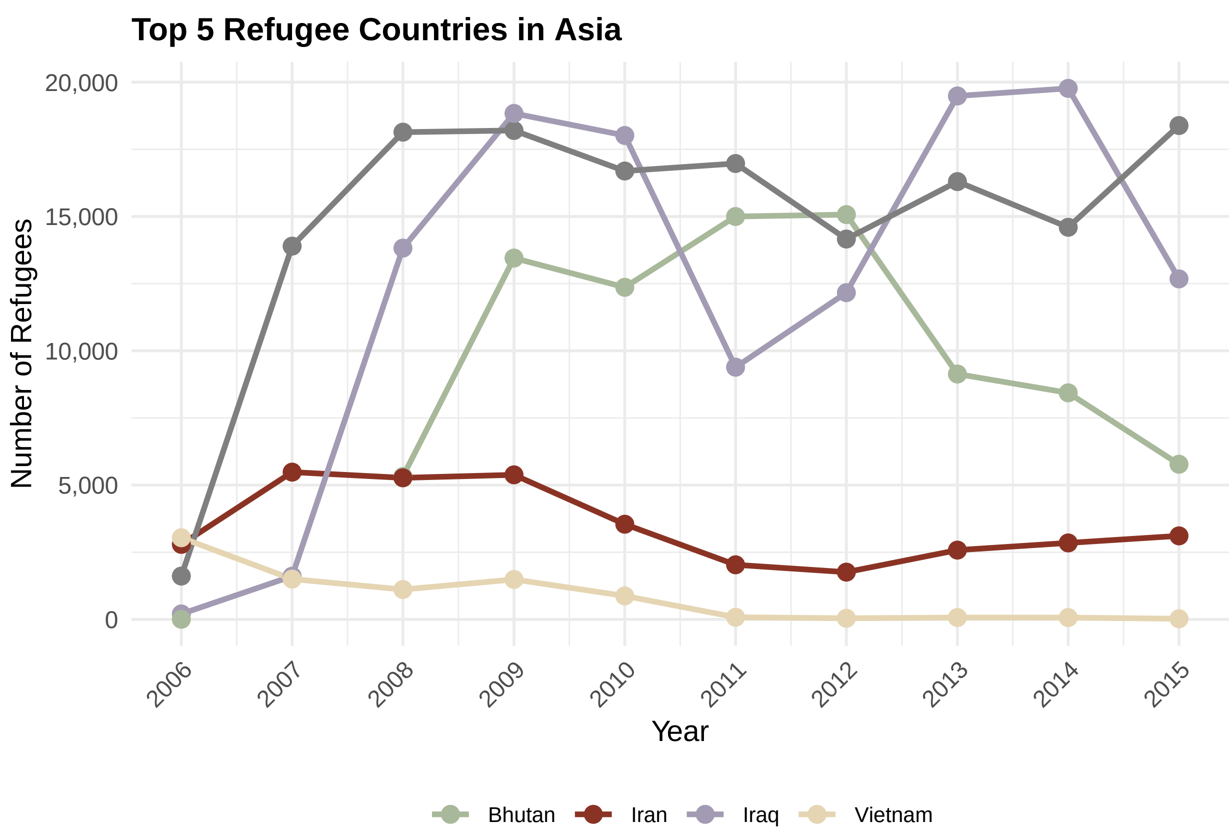
```
## [[1]]
## [1] "output_plot/plot_asia.svg"
##
## [[2]]
## [1] "output_plot/plot_africa.svg"
##
## [[3]]
## [1] "output_plot/plot_north_america.svg"
##
## [[4]]
## [1] "output_plot/plot_europe.svg"
```

```
lapply(seq_along(continent_names), function(i) {
  ggsave(paste0("png/plot_", continent_names[i], ".png"),
         plots[[i]], width = 10, height = 6, dpi = 300)
})
```

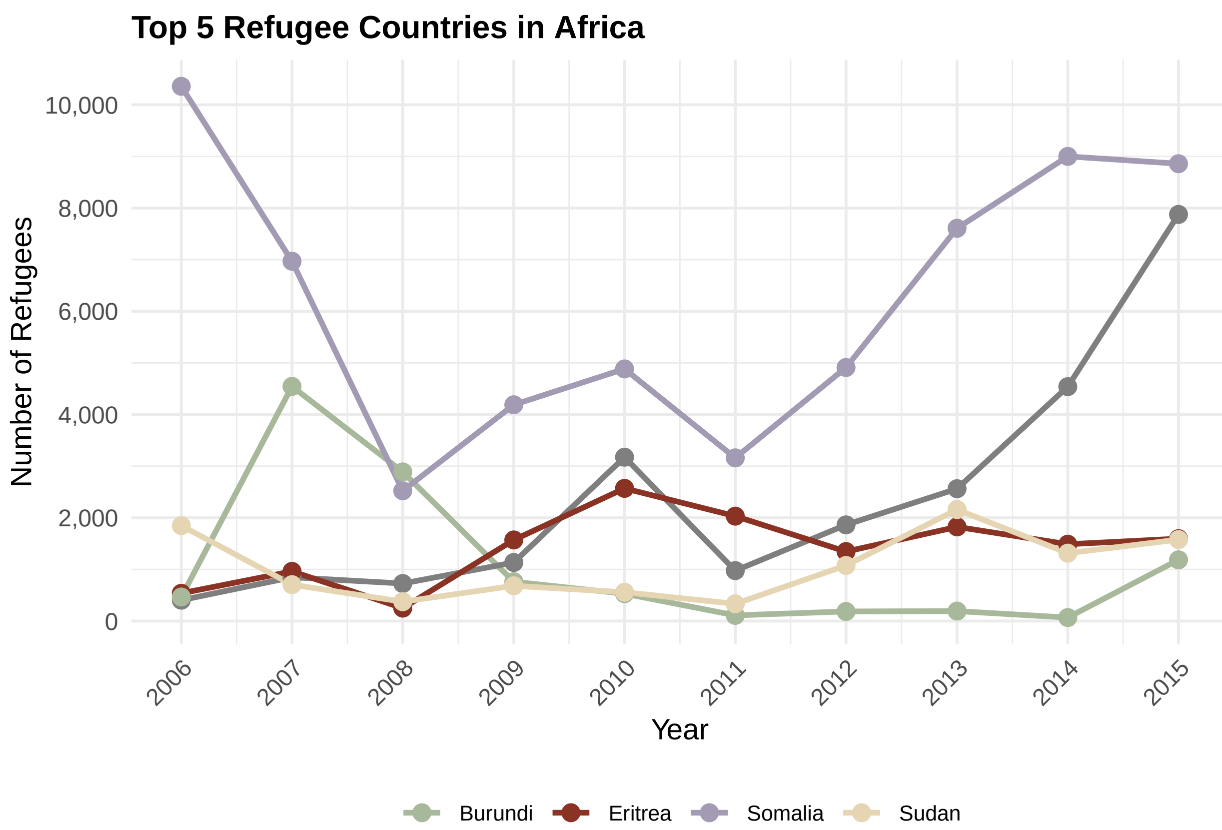
```
## [[1]]
## [1] "png/plot_asia.png"
```

```
##
## [[2]]
## [1] "png/plot_africa.png"
##
## [[3]]
## [1] "png/plot_north_america.png"
##
## [[4]]
## [1] "png/plot_europe.png"
```

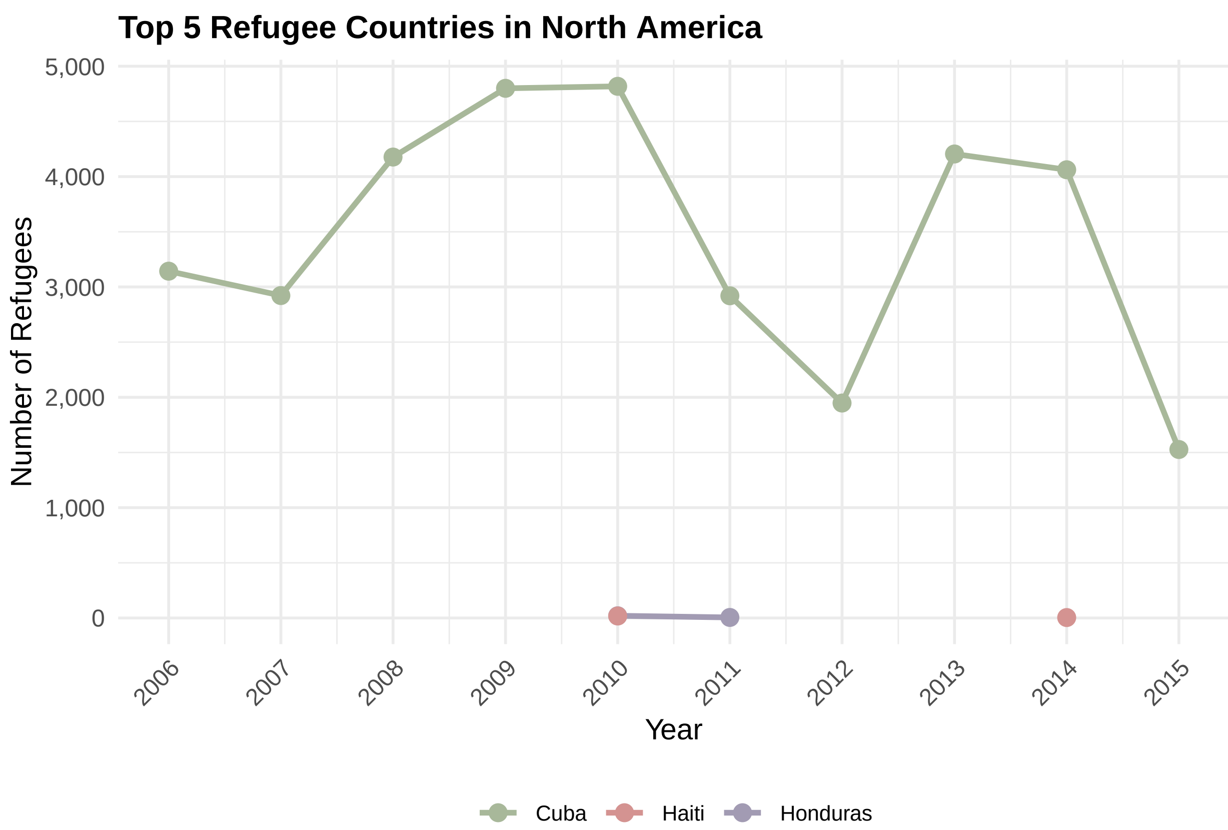
```
# Step 10: Show combined plots with a single column layout
plot_asia
```



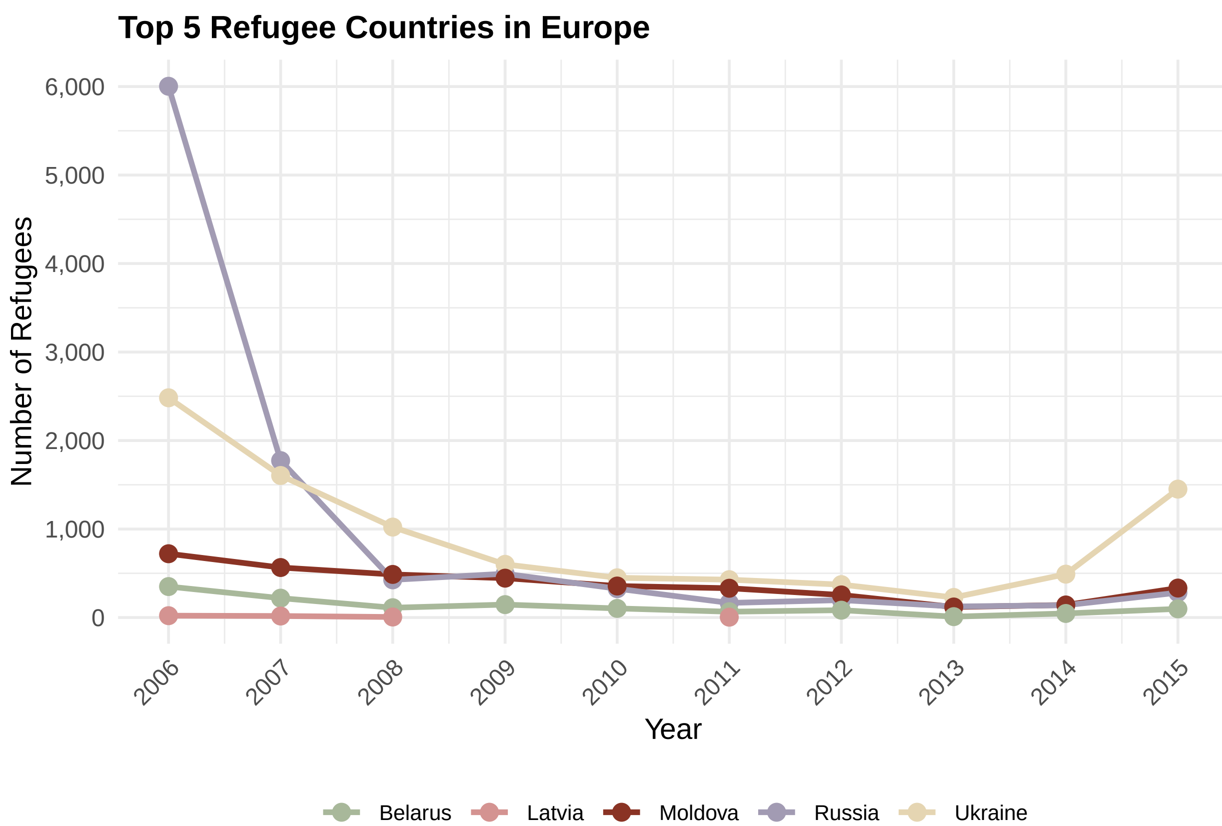

```
plot_africa
```




```
plot_north_america
```


Top 5 Refugee Countries in North America
Number of Refugees
5,000
4,000
3,000
2,000
1,000
0
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
Year
Cuba
Haiti
Honduras

```
plot_europe
```



These graphs represent the top refugee sending countries of Asia, Africa, North America, and Europe for the period from 2006 to 2015. The following line graph outlines refugees from different countries over ye.

Key Findings from the Chart:
- Asia: In the line graph, Iraqi and Burmese refugees show large sending without any reduction. Bhutan's trend line also increased rapidly, peaking in around 2008-2010.
- Africa: Somalia initially provided the largest number of refugees but decreased; Eritrea and Congo, however, had increased numbers towards the end.
- North America: Cuba had significantly higher numbers compared to Haiti and Honduras, yet with a steady though fluctuating trend.
- Europe: Ukraine and Russia had significant initial numbers of refugees, yet general trends are showing a continued decrease.

CRAP:
- Contrast: Every country has its color, and so they would be differentiated. - Repetition: All four graphs follow the same layout, making comparisons easier. - Alignment: The x-axis represents time from 2006 to 2015, making sure chronological order is consistent. - Proximity: Legend of all 4 visualizations are placed near the graphs to increase readability.

Kieran Healy's Principles:
- Simplicity & Clarity: The focus is only on the top five countries per continent and hence it reduces clutter and enhances readability.
- Effective Use of Space: The line graphs provide a clear representation of trends without too many overwhelming details.
- Comparative Analysis: By presenting multiple continents separately, each visualization tells a unique but related story.

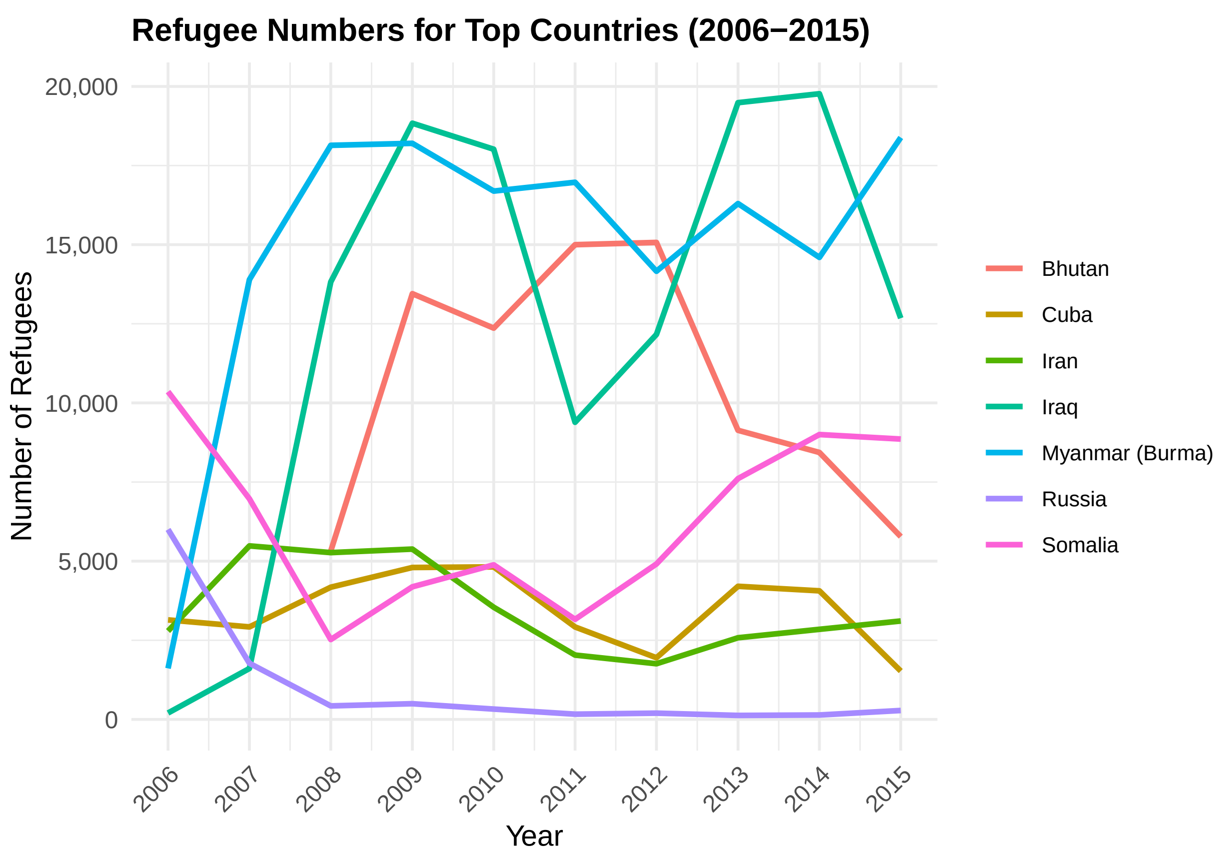
Visualization #3: Refugee Numbers for Top Countries (Line Graph)

```
# Load required libraries
library(dplyr)

# Step 1: Find the top 10 countries by refugees for each year
top10_per_year <- data_long %>%
  group_by(Year) %>%
  slice_max(order_by = Total_Refugees, n = 3) %>%
  ungroup()

# Step 2: Extract all years' data for these top countries
top_countries <- unique(top10_per_year$Country)  # Get unique top countries
all_data_top_countries <- data_long %>%
  filter(Country %in% top_countries)  # Filter all years for top countries

# Step 3: View the data
print(all_data_top_countries)
```

```
## # A tibble: 70 x 5
##    Country  Year Year_Date  Total_Refugees Continent
##    <chr>   <int> <date>              <dbl> <chr>
##  1 Bhutan   2006 2006-01-01              3 Asia
##  2 Bhutan   2007 2007-01-01             NA Asia
##  3 Bhutan   2008 2008-01-01           5320 Asia
##  4 Bhutan   2009 2009-01-01          13452 Asia
##  5 Bhutan   2010 2010-01-01          12363 Asia
##  6 Bhutan   2011 2011-01-01          14999 Asia
##  7 Bhutan   2012 2012-01-01          15070 Asia
##  8 Bhutan   2013 2013-01-01           9134 Asia
##  9 Bhutan   2014 2014-01-01           8434 Asia
## 10 Bhutan   2015 2015-01-01           5775 Asia
## # i 60 more rows
```

```
# Save to a new CSV file if needed
#write.csv(all_data_top_countries, "all_top_countries_refugees.csv", row.names = FALSE)

# Ensure data is loaded into all_data_top_countries
# If not already loaded, reload the dataset
# all_data_top_countries <- read.csv("all_top_countries_refugees.csv")

# Step 4: Create a line plot for all top countries over the years
p00 <- ggplot(all_data_top_countries, aes
              (x = Year_Date, y = Total_Refugees, color = Country, group = Country)) +
  geom_line(size = 1) +  # Add lines for each country
  theme_minimal() +  # Use a minimal theme
  labs(
    title = "Refugee Numbers for Top Countries (2006-2015)",
    x = "Year",
    y = "Number of Refugees",

    color = NULL
```

```r
  ) +
  theme(
    plot.title = element_text(face = "bold", size = 12),
    legend.position = "right",
    legend.text = element_text(size = 8),
    axis.text.x = element_text(angle = 45, hjust = 1) ,
    plot.caption = element_text(size = 8, hjust = 0, color = "gray50")
  ) +
  scale_x_date(date_labels = "%Y",
               breaks = seq(min(all_data_top_countries$Year_Date),
                            max(all_data_top_countries$Year_Date), by = "1 year")
               ) +
  scale_y_continuous(labels = scales::comma)

# Step 5: Export the Figures
ggsave("output_plot/Refugee Numbers for Top Countries 2006-2015.svg",
       p00, width = 8, height = 6, dpi = 300, device = "svg")
ggsave("png/Refugee Numbers for Top Countries 2006-2015.png",
       p00, width = 8, height = 6, dpi = 300)

# Show the Plots
p00
```



This visualization is a line graph showing the refugee population from top contributing countries into the U.S. between the period of 2006-2015 and also shows variation in refugee arrivals from Bhutan, Burma, Cuba, Iraq, Iran, Russia, and Somalia over a period of time.

Key Findings from the Chart:
- Burma and Iraq had the highest influx in refugees, with peaks in 2008 and 2014, respectively.
- Somalia and Bhutan showed steady refugee movement, with a resurgence in Bhutanese refugee numbers after 2011.
- Russia and Iran had comparatively lower numbers, showing a decline over time.

- Cuba's trend remained relatively stable, with minor fluctuations.
- The data suggests that geopolitical events and conflicts significantly influenced refugee resettlement trends.

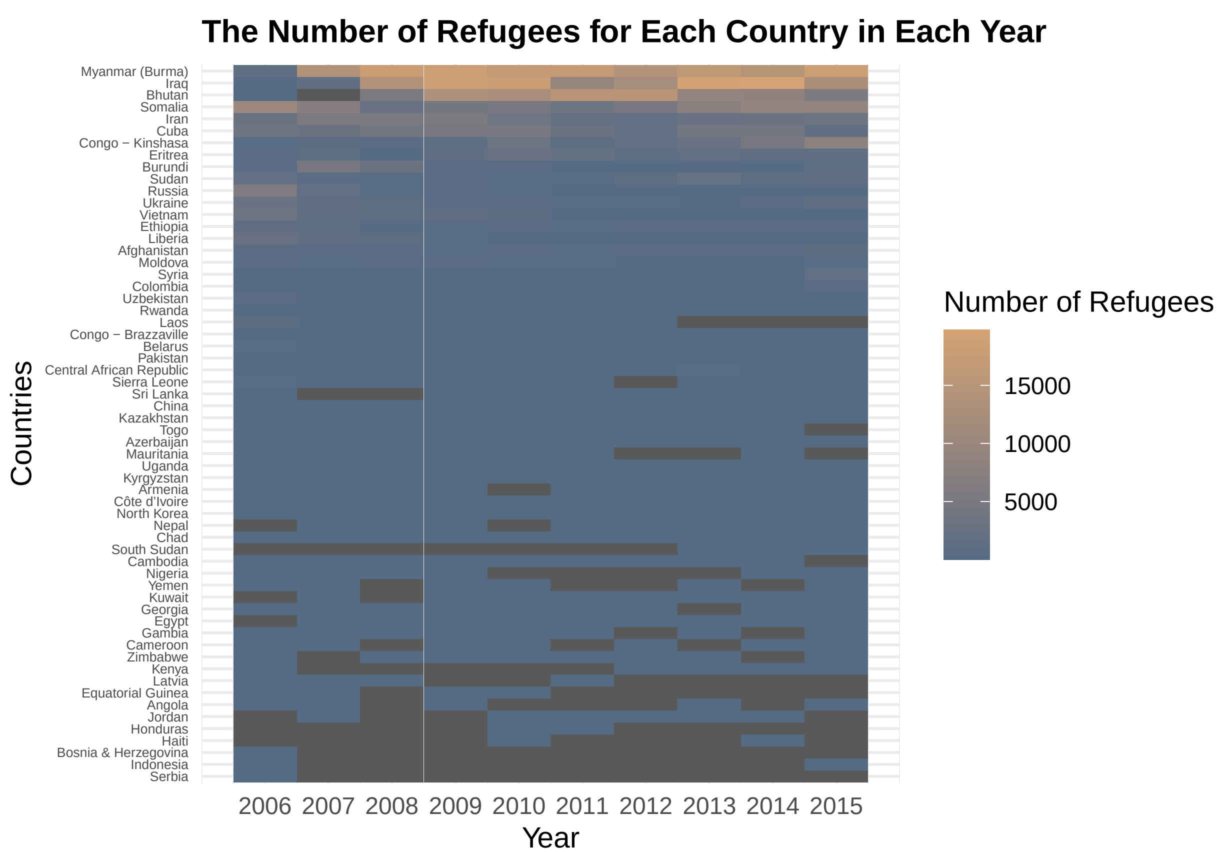
CRAP:
- Contrast: Each country is represented by a unique color, ensuring clear differentiation.
- Repetition: The consistent line thickness and spacing across countries improve readability.
- Alignment: The x-axis represents time, ensuring an easy-to-follow chronological progression.
- Proximity: The legend is placed below the graph, making it easy to reference without crowding the visualization.

Kieran Healy's Principles:
- Simplicity & Clarity: The top refugee-contributing countries are selected for focus, making the chart informative without excessive detail.
- Effective Use of Space: Using the line graph ensures a clear visualization of trends without overlapping data points.
- Comparative Analysis: This chart allows for direct comparisons across countries to understand any variations in refugee movements over time.

Visualization #4: Number of Refugees for each country in each year (HEATMAP)

```r
# Step 0-1 :
country_order <- country_df %>%
  group_by(Country) %>%
  summarise(
    Total_Refugees_Sum = sum(Total_Refugees, na.rm = TRUE),
    NA_Count = sum(is.na(Total_Refugees)),
    Data_Count = n()
  ) %>%
  mutate(NA_Ratio = NA_Count / Data_Count) %>%
  arrange((Total_Refugees_Sum), NA_Ratio) %>%
  pull(Country)

# Step 0-2:
country_df2 <- country_df %>%
  mutate(Country = factor(Country, levels = country_order))

# Step 1: Plotting HeatMap
p1 <- ggplot(country_df2, aes(x = Year_Date, y = Country, fill = Total_Refugees)) +
  geom_tile(na.rm = TRUE) +
```

```r
  scale_fill_gradient(low = "#556b84", high = "#d4a373", na.value = "gray35") +
  scale_x_date(date_labels = "%Y",
               breaks = seq(as.Date("2006-01-01"),
                            as.Date("2015-01-01"),
                            by = "1 year")) +
  theme_minimal() +
  labs(title = "The Number of Refugees for Each Country in Each Year",
       x = "Year",
       y = "Countries",
       fill = "Number of Refugees",
      ) +
  theme(plot.title = element_text(face = "bold", size = 12),legend.position = "right",
  axis.text.y = element_text(size = 5),
  plot.caption = element_text(size = 8, hjust = 0, color = "gray50"))

# Step 2: Saving plot
ggsave("output_plot/The Number of Refugees for Each Country in Each Year2.svg",
       p1, width = 10, height = 6, dpi = 300, device = "svg")
ggsave("png/The Number of Refugees for Each Country in Each Year2.png",
       p1, width = 10, height = 6, dpi = 300)

# Step 3: Show plot
p1
```



This heat map displays refugee admissions from various countries between the 2006 and 2015 period, where the shading intensity represents the number of refugees. Darker shade indicates fewer refugees, while lighter shades highlight higher refugee numbers.

Key Findings from the Chart:
- This visualization reveals patterns of refugee movement from different countries over time.
- Countries like Burma, Iraq, and Somalia consistently have higher refugee numbers, as indicated by the lighter-colored rows.
- Periods of significant influx align with major geopolitical crises or policy changes.
- Some countries exhibit sporadic peaks, reflecting temporary surges in refugee resettlement.
- Countries with consistently dark shading contribute very few refugees to the U.S.

CRAP:
- Contrast: The gradient color scale enhances readability, showing clear differentiation between high and low refugee numbers.
- Repetition: The consistent structure of country names and time intervals ensures clarity across the data set.
- Alignment: The x-axis represents years, while the y-axis lists countries, which follows a left-to-right, top-to-bottom reading order.
- Proximity: The legend is placed to the right, making sure that the focus is on the heatmap while allowing for easy interpretation.

Kieran Healy's Principles:
- Comparative Focus: This heat map allows an immediate visual comparison between different countries and time periods.
- Effective Use of Space: By using light and dark color gradients, the visualization efficiently describes large amounts of information without overwhelming the viewer.
- Layering & Separation: The color scale helps differentiate refugee intensity and prevents cluttering while also maintaining a structured presentation.
- Visual Hierarchy: Brighter colors naturally draw attention to key data points (countries with the highest refugee numbers).
- Simplicity & Clarity: The heatmap eliminates unnecessary visual clutter, avoiding excessive numerical values while keeping the focus on trends.
- Proportionality: Each row is proportional to the number of countries, ensuring fair representation without distorting information.

Visualization #5 The World Map with Refugees number

```r
# Load Libraries
library(gganimate)
library(sf)
library(rnaturalearth)
library(rnaturalearthdata)
library(gifski)
library(transformr)


# load the world map
world_map <- ne_countries(scale = "medium", returnclass = "sf")

# combine the map data and refugees data
map_data <- world_map %>%
```

```r
  left_join(country_df, by = c("name" = "Country"))

# Plotting
plot_yearly_maps <- function(year) {
    yearly_data <- map_data %>%
    filter(Year_Date == as.Date(paste0(year, "-01-01")))

  ggplot() +
    geom_sf(data = world_map, fill = "gray90", color = "white") +
    geom_sf(data = yearly_data, aes(fill = Total_Refugees), color = "black") +
    scale_fill_gradient(low = "#556b84", high = "#d4a373", na.value = "gray90") +
    theme(plot.title = element_text(face = "bold", size = 12))+
    theme_minimal() +
    labs(title = paste("Global Refugee Map - Year", year), fill = "Number of Refugees")
}

p2006 <- plot_yearly_maps(2006)
p2007 <- plot_yearly_maps(2007)
p2008 <- plot_yearly_maps(2008)
p2009 <- plot_yearly_maps(2009)
p2010 <- plot_yearly_maps(2010)
p2011 <- plot_yearly_maps(2011)
p2012 <- plot_yearly_maps(2012)
p2013 <- plot_yearly_maps(2013)
p2014 <- plot_yearly_maps(2014)
p2015 <- plot_yearly_maps(2015)

years <- 2006:2015
plots <- list(p2006, p2007, p2008, p2009, p2010, p2011, p2012, p2013, p2014, p2015)

lapply(seq_along(years), function(i) {
  ggsave(paste0("output_plot/p", years[i], ".svg"), plots[[i]], width = 8, height = 6, dpi = 300)
})
```

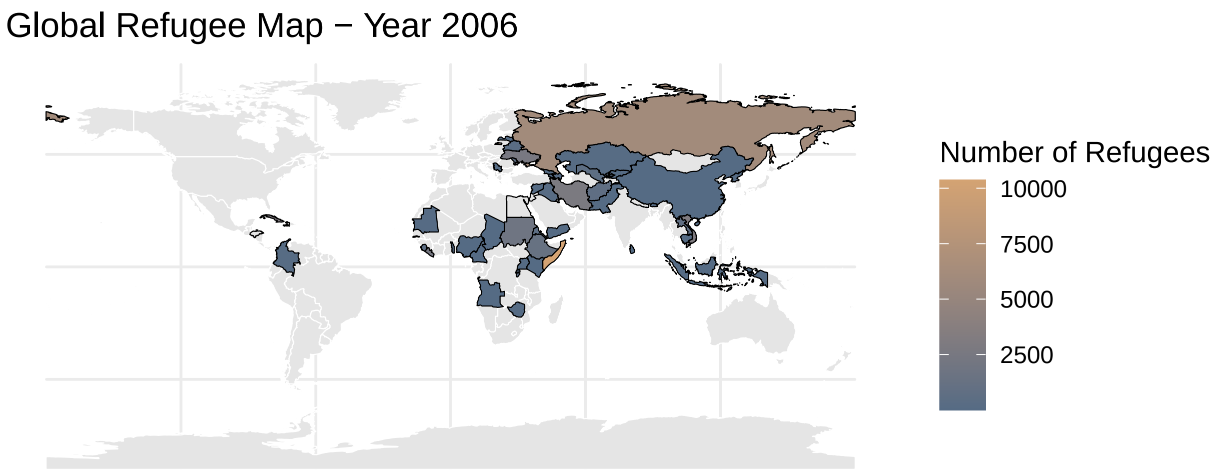
```
## [[1]]
## [1] "output_plot/p2006.svg"
##
## [[2]]
## [1] "output_plot/p2007.svg"
##
## [[3]]
## [1] "output_plot/p2008.svg"
##
## [[4]]
## [1] "output_plot/p2009.svg"
##
## [[5]]
## [1] "output_plot/p2010.svg"
##
## [[6]]
## [1] "output_plot/p2011.svg"
##
## [[7]]
```

```
## [1] "output_plot/p2012.svg"
##
## [[8]]
## [1] "output_plot/p2013.svg"
##
## [[9]]
## [1] "output_plot/p2014.svg"
##
## [[10]]
## [1] "output_plot/p2015.svg"
```

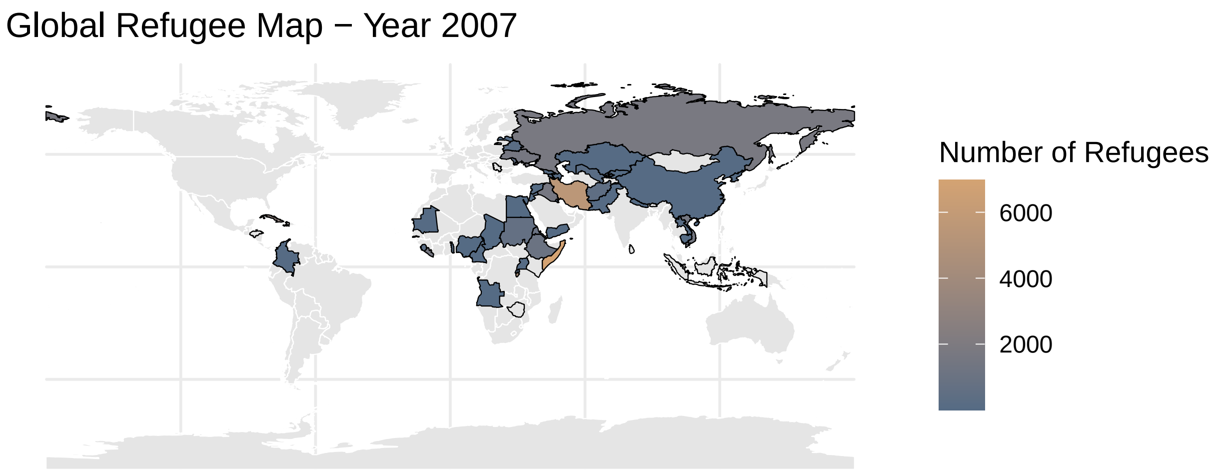
```
lapply(seq_along(years), function(i) {
  ggsave(paste0("png/p", years[i], ".png"), plots[[i]], width = 8, height = 6, dpi = 300)
})
```

```
## [[1]]
## [1] "png/p2006.png"
##
## [[2]]
## [1] "png/p2007.png"
##
## [[3]]
## [1] "png/p2008.png"
##
## [[4]]
## [1] "png/p2009.png"
##
## [[5]]
## [1] "png/p2010.png"
##
## [[6]]
## [1] "png/p2011.png"
##
## [[7]]
## [1] "png/p2012.png"
##
## [[8]]
## [1] "png/p2013.png"
##
## [[9]]
## [1] "png/p2014.png"
##
## [[10]]
## [1] "png/p2015.png"
```

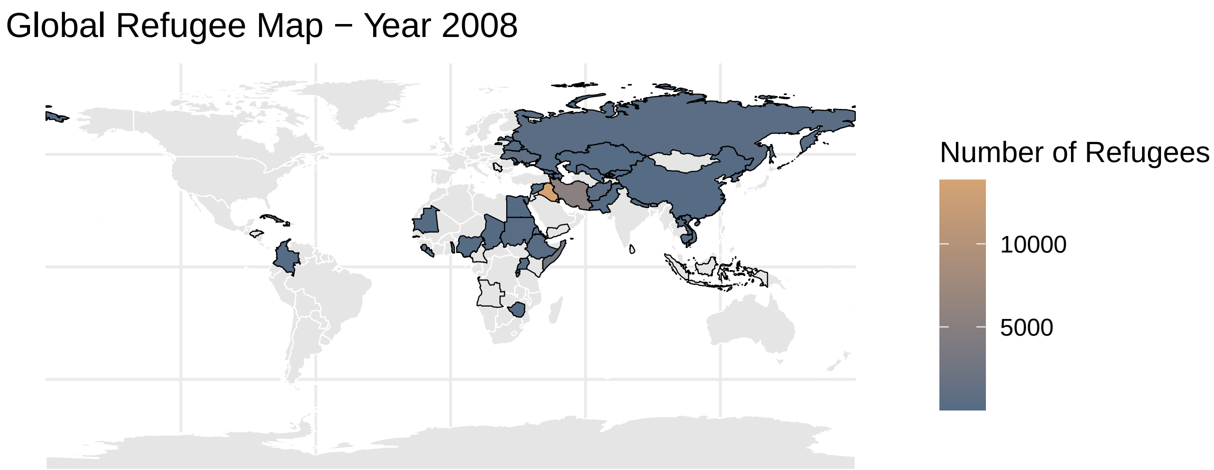
```
p2006
```


Global Refugee Map − Year 2006
Number of Refugees
10000
7500
5000
2500

```
p2007
```

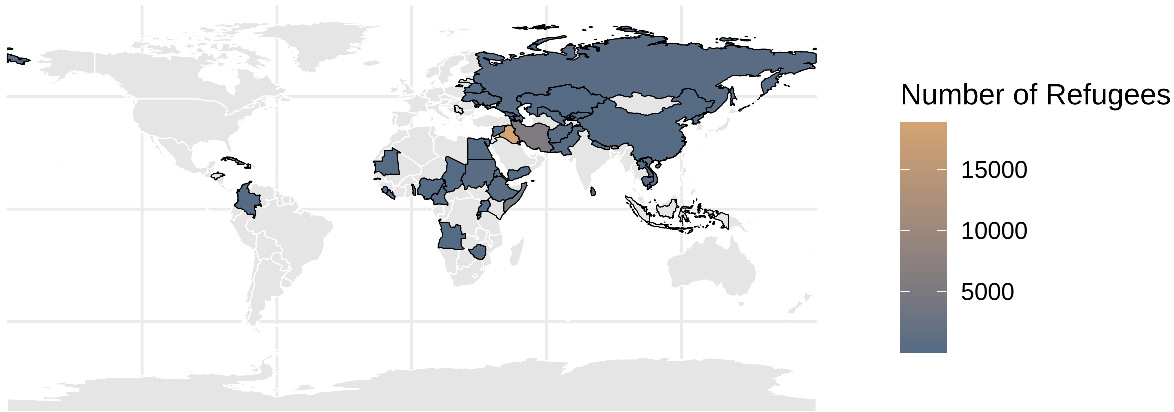

Global Refugee Map − Year 2007
Number of Refugees
6000
4000
2000

```
p2008
```

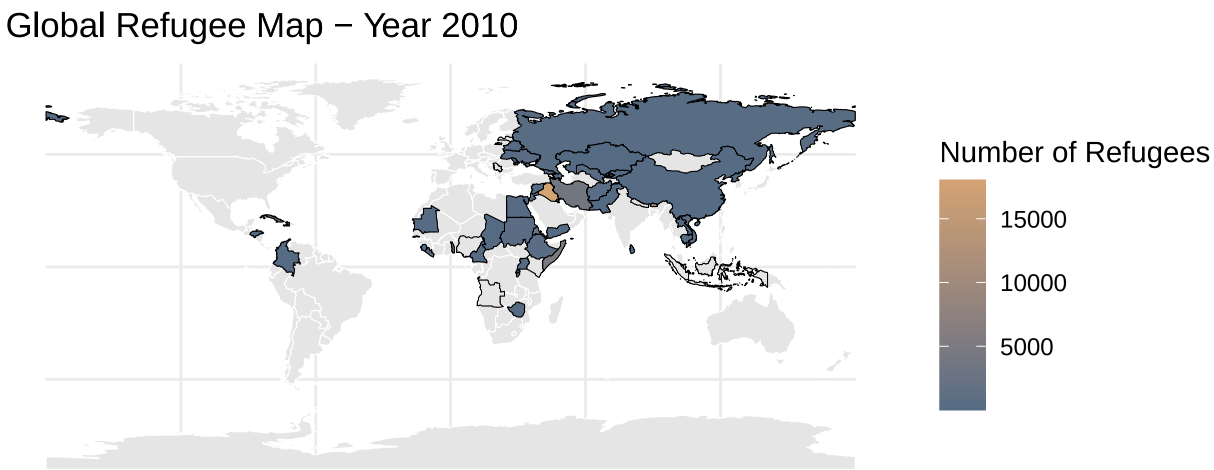



```
p2009
```

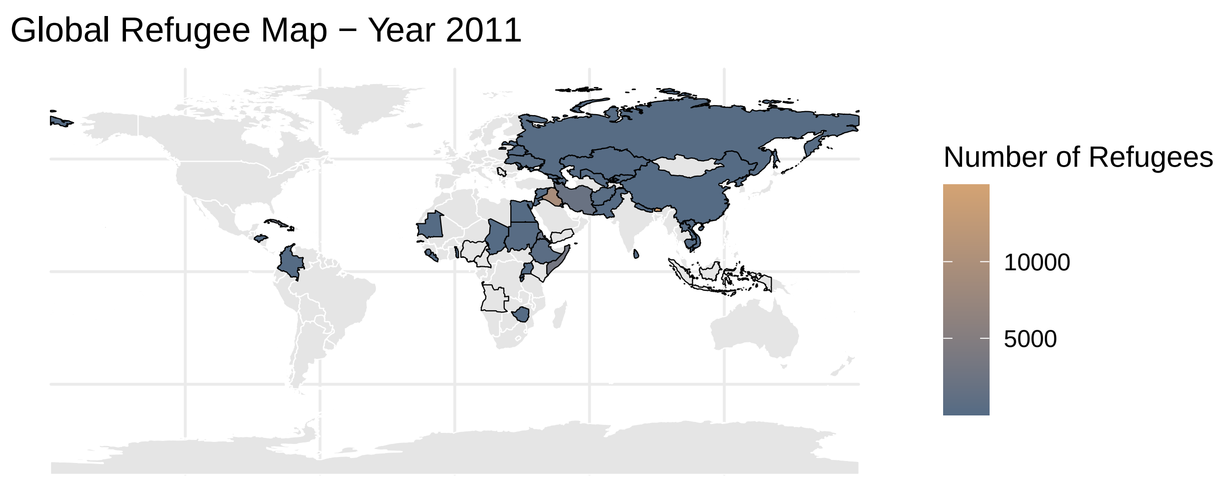
Global Refugee Map − Year 2009



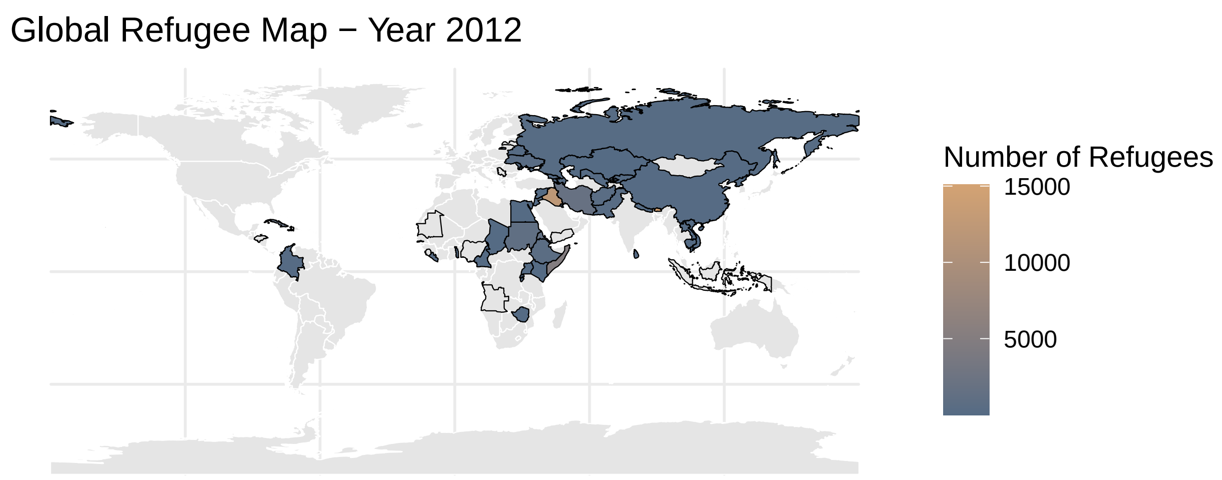

```
p2010
```


Global Refugee Map − Year 2010
Number of Refugees
15000
10000
5000

```
p2011
```

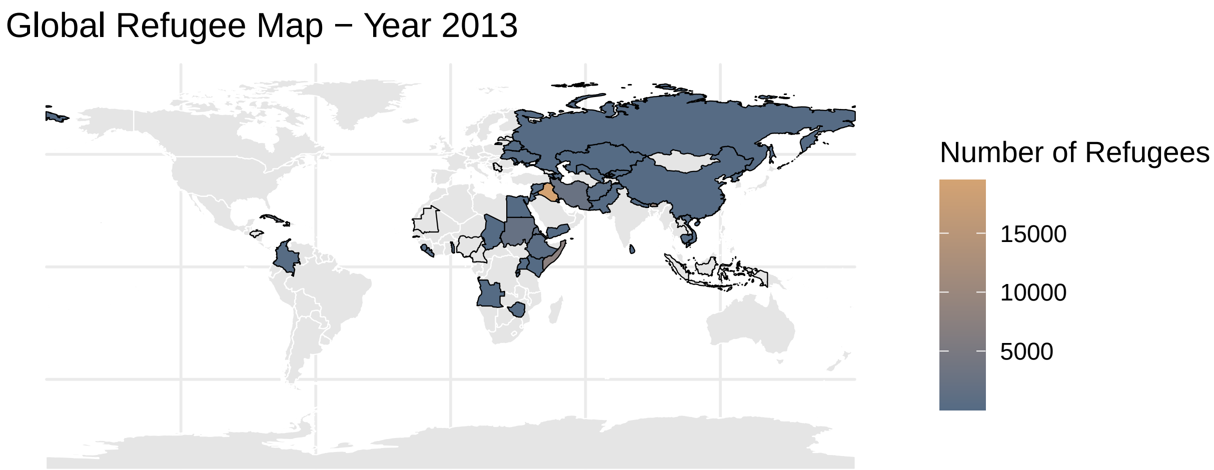



```
p2012
```

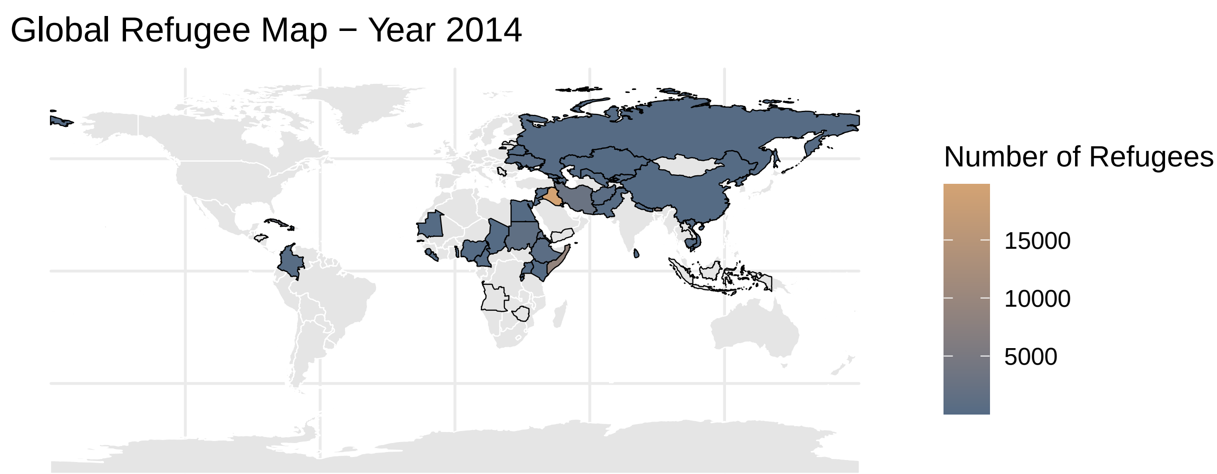


Global Refugee Map − Year 2012
Number of Refugees
15000
10000
5000

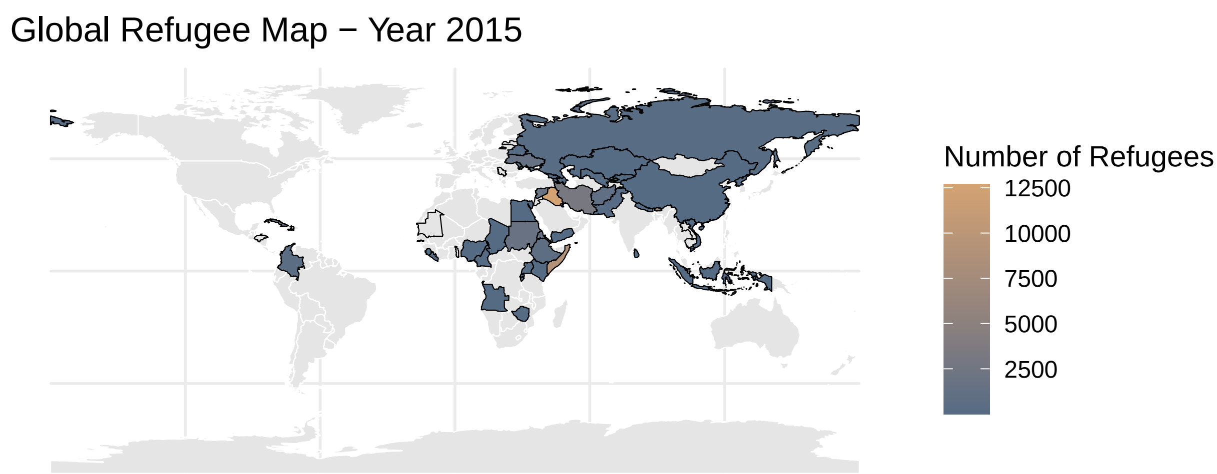

```
p2013
```


Global Refugee Map − Year 2013
Number of Refugees
15000
10000
5000

```
p2014
```


Global Refugee Map − Year 2014
Number of Refugees
15000
10000
5000

```
p2015
```



This animated global refugee map visualizes the number of refugees admitted into the U.S. from various countries for year-by-year periods in the range between 2006 to 2015. This map highlights the geographic origins of refugees and describes how refugee admissions changed over time.

Key Findings from the Chart:
- Continuous refugee flows from conflict-prone regions: Countries in the Middle East, for instance, Iraq and Syria; Africa-for example, Somalia, Sudan, DR Congo; Asia-for example, Burma, Bhutan, have remained consistent sources of refugees.
- Temporal fluctuation in the number of refugees across years is depicted by different shades for different years, sometimes mirroring geopolitical events, changing U.S. immigration policies, and humanitarian crises.
- Regional Patterns: Some regions appear to retain the darker shade continuously, such as the Middle East and parts of Africa, signaling a sustained outflow of refugees.
- Spikes and declines by notable amplitudes: The flux of refugees from a specific country speaks to changes in policy or increased conflations like for example, refugees from Syria and Iraq increased after 2011.

CRAP:
- Contrast: There is a variation in refugee number presentation; a gradient color scale provides darker shade for higher values. This way, areas that have high refugee admissions really pop.

- Repetition: The color scale and map layout are the same in every frame for consistency and ease of comparison through time. The legend is in the same position to maintain user familiarity.

- Alignment: The map, title, and legend are aligned, making the layout clean and professional. The timeline progression is logical and, therefore, easy to interpret through the animation.

- Proximity: The legend is placed close to the visualization, without overlapping it, for the viewers to reach the color gradient without hassle in case they are looking into the trends. Logical grouping of countries is followed by keeping the geographical relevance intact.

Kieran Healy's Principles:
- Truthful: No distortion of data, and there are no misleading visuals. There is one-to-one correspondence in color intensity with regard to the number of refugees.
- Functional: The animated format supports the story-telling as the user has a sense of observing changes through time rather than seeing a static snapshot. Color scale is effective without numerical labels required for each and every country on relative magnitudes.
- Beautiful: The visual aesthetic is pleasing on the eyes and has an appropriate color palette. The animation element is engaging without being distractive.
- Elucidating: The visualization shows the hidden trends in how geopolitical crises surge refugees from specific countries into lighter shade. Users can, at a glance, identify which regions constantly contribute to the high number of refugees.

below code is for generate the GIF (we placed the individual picture above incase the GIF doesn't work) library(gganimate)

p <- ggplot() + geom_sf(data = world_map, fill = "gray90", color = "white") + # the world map geom_sf(data = map_data, aes(fill = Total_Refugees), color = "black") + # data refugees scale_fill_gradient(low = "#556b84", high = "#d4a373", na.value = "gray90") + theme_minimal() + labs(title = "Global Refugees data Map (Year: {frame_time})", fill = "Number of Refugees", x = " ", y ="") + transition_time(as.integer(format(Year_Date,"%Y"))) + # change the plot by year ease_aes('linear') # change smoothly

anim <- animate(p, duration = 20, fps = 20, width = 800, height = 500, renderer = gifski_renderer()) anim_save("refugees_map_smooth.gif", animation = anim)

Conclusion

This analysis provides a comprehensive look at global refugee movements, highlighting key trends, high-risk regions, and changes over time. The combination of data cleaning, visualization, and statistical analysis allows for an in-depth understanding of the factors driving refugee migration, which can support future policy-making and humanitarian efforts.