Data Analysis Report

1. Topic

Global Suicide Trends & Analysis

- **Focus**: Investigating the patterns, risk factors, and socioeconomic effects of suicide globally.

2. Introduction

The World Health Organization (WHO), a preeminent global public health authority, has recognized suicide as a serious issue that affects people in all geographic and demographic contexts. This study seeks to provide a thorough examination of the patterns, risk factors, and socioeconomic effects of suicide, delving extensively into the global trends in this phenomenon. This initiative aims to identify the fundamental causes of this complicated issue by utilizing the vast resources and data the World Health Organization has made available. It seeks to uncover the intricate relationship between economic prosperity and suicide rates, offering a comparative analysis across diverse economic contexts. By leveraging an extensive dataset encompassing variables like GDP, regional, and demographic details, this study endeavors to elucidate the underlying factors contributing to suicide. The findings are anticipated to offer significant insights, potentially guiding the development of effective preventive strategies and mental health policies. Given that suicide claims the lives of more than **700,000 people a year** [1], it is vital to comprehend these trends. The findings are anticipated to offer significant insights, potentially guiding the development of effective preventive strategies and mental health policies.

3. Previous Research

Previous research on global suicide trends has highlighted several key insights. A systematic analysis for the **Global Burden of Disease Study 2016** [2] showed that while the total number of deaths from suicide increased globally over 27 years, the age-standardized mortality rate decreased significantly, pointing to a complex interplay of demographic factors in suicide rates worldwide. Another study, **Impact of 2008 global economic crisis on suicide** [3] noted that despite overall declining trends in several areas, suicide rates have increased in some countries post the 2008 global financial crisis, indicating economic factors' impact on suicide trends.

4. Research Questions

The project will assist in answering the following questions:

1. Analyzing the Impact of Economic Factors on Global Suicide Rates

The study examines the global link between economic indicators, like GDP per capita, and suicide rates. It compares these rates across countries with varying economic statuses to understand how financial health impacts mental well-being and suicide instances worldwide.

Hypothesis: Economic factors, such as GDP per capita, significantly influence suicide rates in countries, with a complex, potentially inverse relationship. Different economic tiers of countries may exhibit distinct patterns in this correlation. *[Discussed in section 5.9]*

2. What is the Relationship Between Age-Related Factors and Suicide Rates Globally?

This question aims to explore how age-specific factors contribute to suicide rates across different countries and continents.
Hypothesis: Certain age groups are more vulnerable to suicide, influenced by unique socio-economic and psychological pressures that differ with age. *[Discussed in section 5.2, 5.6, 5.9.2]*

3. Analysis of Gender Disparities in Suicide Rates Across Different Socioeconomic Contexts
This question investigates the "gender paradox" in suicide, where despite higher rates of depression among women, men have higher suicide rates. The analysis will delve into how this disparity manifests across various countries and economic backgrounds.
Hypothesis: Cultural, societal, and economic factors contribute to this gender disparity in suicide rates. *[Discussed in section 5.3, 5.7, 5.8, 5.9.2]*

5. Data

5.1 Dataset Description

The dataset used in this project for my study on global suicide trends is sourced from Kaggle, an online community and platform for data scientists and machine learning practitioners. Kaggle provides a wide array of datasets, including the one I need, which encompasses various fields such as country, year, sex, age group, count of suicides, and other socio-economic indicators.

The dataset (Suicide Rates Overview 1985 to 2016) [4] encompasses a variety of fields providing comprehensive insights into global suicide trends.

```
# Loading the data
data <- read.csv("master.csv", check.names = FALSE)
```

5.2 Dataset Sample

country	year	sex	age	suicides_no	population	suicides/100k pop	country-year	HDI for year	gdp_for_year ($)	gdp_per_capita ($)	generation
Albania	1987	male	15-24 years	21	312900	6.71	Albania1987		2,156,624,900	796	Generation X
Albania	1987	male	35-54 years	16	308000	5.19	Albania1987		2,156,624,900	796	Silent
Albania	1987	female	15-24 years	14	289700	4.83	Albania1987		2,156,624,900	796	Generation X
Albania	1987	male	75+ years	1	21800	4.59	Albania1987		2,156,624,900	796	G.I. Generation
Albania	1987	male	25-34 years	9	274300	3.28	Albania1987		2,156,624,900	796	Boomers

Total Rows (Data Points): 27,820
Total Columns (Attributes): 12
Column Names and Data Types:

- **country** (String): Geographic location of the data.
- **year** (Integer): Time frame of the data collection.
- **sex** (String): Gender of individuals (male/female).
- **age** (String): Categorised age brackets.
- **suicides_no** (Integer): Total suicides recorded.
- **population** (Integer): Total population for each subgroup.
- **suicides/100k pop** (Float): Suicides per 100,000 individuals.
- **country-year** (String): A unique identifier for each data point.
- **HDI for year** (Float): Human Development Index for the specified year.

- **gdp_for_year** ($) (String): Country's Gross Domestic Product.
- **gdp_per_capita** ($) (Integer): GDP relative to population size.
- **generation** (String): Defined based on average age group.

Missing Values:

- HDI for year column is missing significant number (19,456) of entries
- All other columns are fully populated.

5.3 Dataset Cleaning

First step involves removing the 'HDI for year' column due to a high number of missing values making it unusable for our analysis. Additionally we renamed a few columns for simplifying the structure and making it more user-friendly for analysis.

```
> sum(is.na(data$`HDI.for.year`))
[1] 19456
```

```
data <- data %>%
  select(-`HDI for year`) %>%
  rename(
    gdp_for_year = `gdp_for_year ($)`,
    gdp_per_capita = `gdp_per_capita ($)`,
    country_year = `country-year`,
    suicides_per_100k = `suicides/100k pop`
  )
```

We want to ensure data consistency across country-year combinations. We should have 12 rows for every county-year combination (6 age bands * 2 genders)

For the year 2016, we noted irregularity with many countries displaying incomplete or missing data. Therefore, we excluded the data for the year 2016.

```
> data %>%
+     group_by(country_year) %>%
+     count() %>%
+     filter(n != 12)
# A tibble: 16 × 2
# Groups:   country_year [16]
   country_year          n
   <chr>             <int>
 1 Armenia2016          10
 2 Austria2016          10
 3 Croatia2016          10
 4 Cyprus2016           10
 5 Czech Republic2016   10
 6 Grenada2016          10
 7 Hungary2016          10
 8 Iceland2016          10
 9 Lithuania2016        10
10 Mauritius2016        10
11 Mongolia2016         10
```

```
12 Netherlands2016       10
13 Qatar2016             10
14 Romania2016           10
15 Sweden2016            10
16 Thailand2016          10
```

```
data <- data %>%
  filter(year != 2016) %>% # Remove rows where the year is 2016
  select(-country_year) # Drop the country_year column since we won't need it now
```

Here we're also focusing on the amount of data available for each country. Countries with only 3 or fewer years of data are removed from the dataset. This ensures we're working with countries that have enough data for a reliable analysis.

```
# Excluding countries with <= 3 years of data:
minimum_years <- data %>%
  group_by(country) %>%
  summarize(rows = n(), years = rows / 12) %>%
  filter(years > 3)
data <- data %>%
  filter(country %in% minimum_years$country)
```

Tidying the dataframe,

```
data <- data %>%
  mutate(
    age = gsub(" years", "", age),    # Remove "years" from the age column
    sex = ifelse(sex == "male", "Male", "Female")  # Standardize the sex column values
  )
```

Now since our analysis involve suicide rates across different continents, we also add continent data is added to the dataset

```
# Getting continent data
data <- data %>%
  mutate(
    continent = countrycode(sourcevar = country,
                            origin = "country.name",
                            destination = "continent")
  )
```

Selected nominal variables **'country'**, **'sex'**, and **'continent'** are converted to **factors** for categorical analysis, while 'age' is ordered from youngest to oldest age groups, and 'generation' is ordered chronologically from oldest to newest generation, streamlining the dataset for structured analysis.

```
# Convert selected nominal variables to factors
# 'country', 'sex', and 'continent' are treated as nominal
data_nominal <- c('country', 'sex', 'continent')
data[data_nominal] <- lapply(data[data_nominal], factor)
```

```r
# Convert 'age' to an ordered factor
data$age <- factor(data$age,
                   ordered = TRUE,
                   levels = c("5-14", "15-24", "25-34", "35-54", "55-74", "75+"))

# Convert 'generation' to an ordered factor
# The levels are ordered chronologically from the oldest to the newest generation
data$generation <- factor(data$generation,
                          ordered = TRUE,
                          levels = c("G.I. Generation",
                                     "Silent",
                                     "Boomers",
                                     "Generation X",
                                     "Millenials",
                                     "Generation Z"))
```

Converting dataset to tibble format for easier handling. Additionally, the global average suicide rate per 100,000 people is calculated as the total number of suicides divided by the total population, then multiplied by 100,000. This metric provides an overview of the average suicide rate across the dataset.

```r
data <- as_tibble(data)

# i) Calculate the global average suicide rate per 100,000 people
global_average <- (sum(as.numeric(data$suicides_no)) / sum(as.numeric(data$population)))
* 100000

# Function to calculate suicide rates
calc_suicide_rate <- function(data) {
  (sum(as.numeric(data$suicides_no)) / sum(as.numeric(data$population))) * 100000
}
```

```r
glimpse(data)
```

```
> glimpse(data)
Rows: 27,492
Columns: 11
$ country           <fct> Albania, Albania, Albania, Albania, Albania, Albania, Albania, A…
$ year              <int> 1987, 1987, 1987, 1987, 1987, 1987, 1987, 1987, 1987, 1987, 1987…
$ sex               <fct> Male, Male, Female, Male, Male, Female, Female, Female, Male, Fe…
$ age               <ord> 15-24, 35-54, 15-24, 75+, 25-34, 75+, 35-54, 25-34, 55-74, 5-14,…
$ suicides_no       <int> 21, 16, 14, 1, 9, 1, 6, 4, 1, 0, 0, 0, 2, 17, 1, 14, 4, 8, 3, 5,…
$ population        <int> 312900, 308000, 289700, 21800, 274300, 35600, 278800, 257200, 13…
$ suicides_per_100k <dbl> 6.71, 5.19, 4.83, 4.59, 3.28, 2.81, 2.15, 1.56, 0.73, 0.00, 0.00…
$ gdp_for_year      <chr> "2,156,624,900", "2,156,624,900", "2,156,624,900", "2,156,624,90…
$ gdp_per_capita    <int> 796, 796, 796, 796, 796, 796, 796, 796, 796, 796, 796, 796, 769,…
$ generation        <ord> Generation X, Silent, Generation X, G.I. Generation, Boomers, G.…
$ continent         <fct> Europe, Europe, Europe, Europe, Europe, Europe, Europe, Europe, …
```

Now we completed all the data cleaning steps, lets move to the exploratory data analysis.

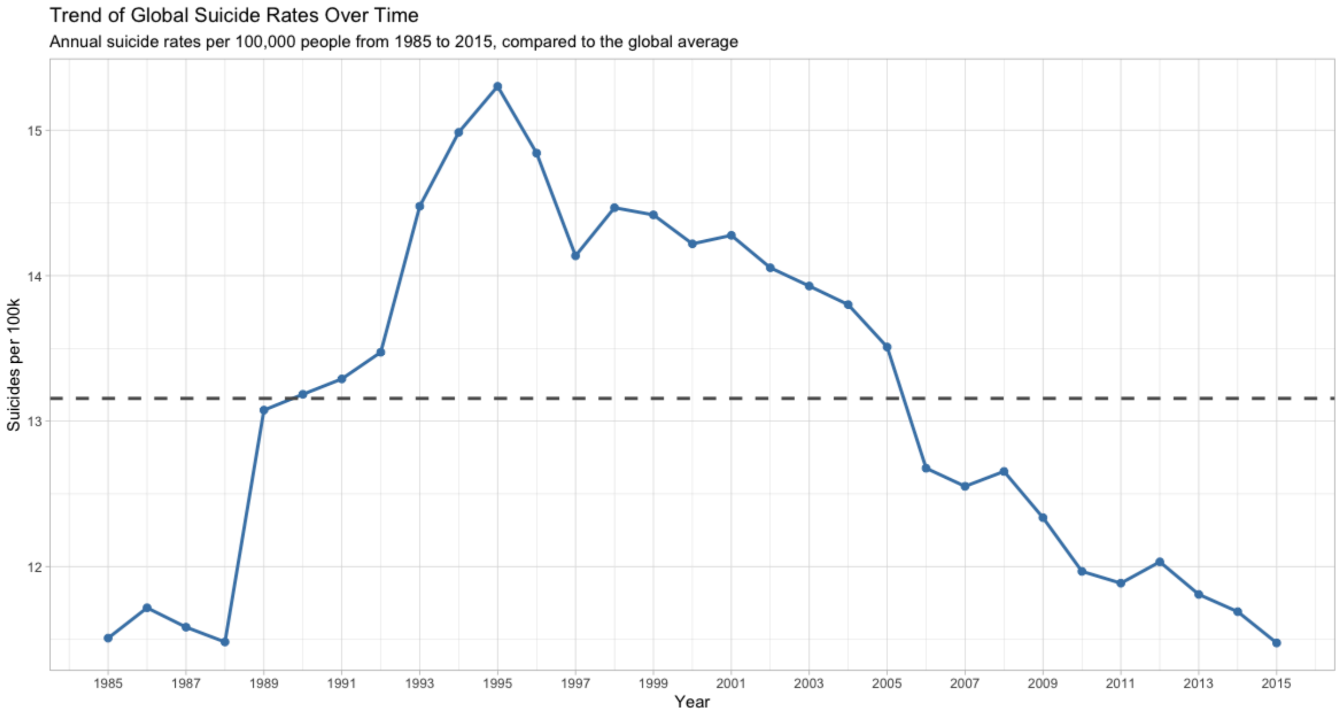
6. Global Analysis

6.1 Global Trend

First, we look into the global trent to provide a foundational understanding of the global suicide trend over three decades, allowing us to identify temporal patterns and set a context for deeper analysis of region-specific or demographic-specific trends that may correlate with these overarching patterns.

```
# Calculate yearly statistics and plot global suicides per 100k over time
yearly_stats <- data %>%
  group_by(year) %>%
  summarize(
    rate_per_100k = calc_suicide_rate(cur_data())
  )
# Plotting the annual suicide rate
ggplot(yearly_stats, aes(x = year, y = rate_per_100k)) +
  geom_line(color = "steelblue", size = 1) + geom_point(color = "steelblue", size = 2) +
  geom_hline(yintercept = global_average, color = "grey35", linetype = 2, size = 1) +
  labs(
    title = "Trend of Global Suicide Rates Over Time",
    subtitle = "Annual suicide rates per 100,000 people from 1985 to 2015, compared to
the global average",
    x = "Year",
    y = "Suicides per 100k") +
  scale_x_continuous(breaks = seq(1985, 2015, 2)) +
  scale_y_continuous(breaks = seq(10, 20))
```

The dashed line is the **global average suicide** rate from **1985 - 2015**: **<u>13.15 deaths</u>** (per 100k, per year).



Analysis of Global Suicide Rates (1985-2015):

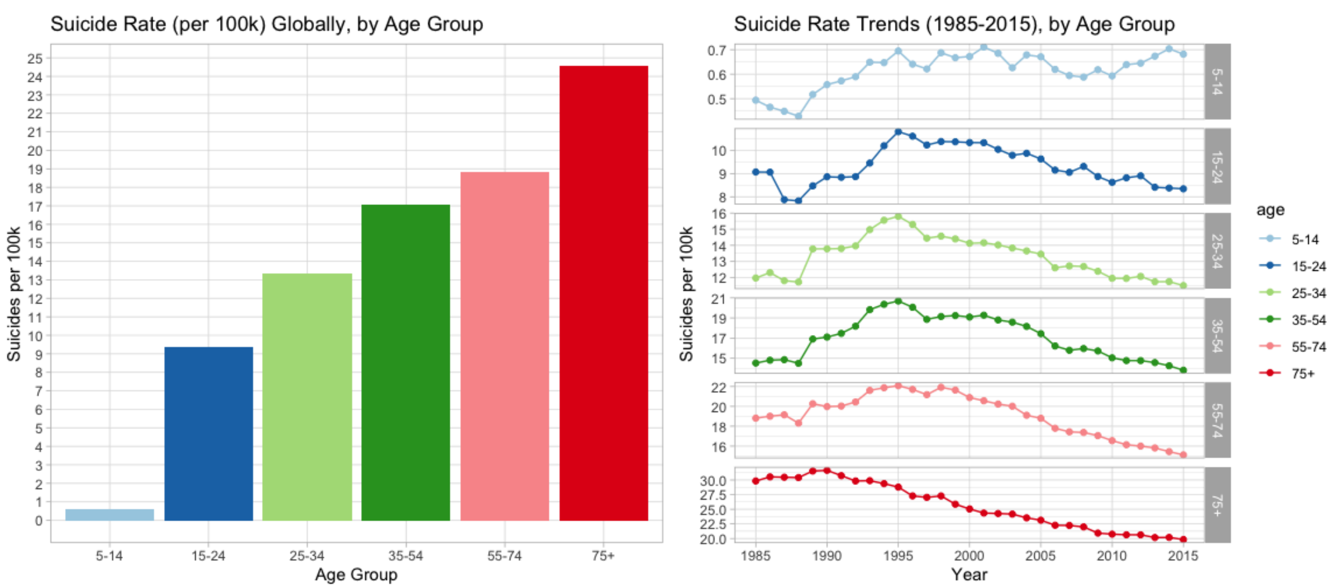
- The highest recorded global suicide rate within this period was in **1995**, with **15.3 deaths** per 100,000 individuals.
- Following the peak in 1995, there has been a consistent decline, culminating in a rate of **11.5** per 100,000 by **2015**, which indicates a **25% decrease** from the peak.
- The downward trend has brought the suicide rates back to levels seen before the rise in the early 1990s.
- The **data from the 1980s** should be interpreted with **caution due to its limited scope**, which may not accurately reflect the global situation at the time.

6.2 By Age

The age-based plots were selected to highlight suicide rates across different life stages and observe how these trends change over years, providing insight into age as a factor in global suicide patterns.

```
# Prepare data for plotting suicide rates by age
age_stats <- data %>%
  group_by(age) %>%
  summarize(suicide_per_100k = calc_suicide_rate(cur_data()))

# Plot global suicides per 100k, categorized by age groups
age_plot <- ggplot(age_stats, aes(x = age, y = suicide_per_100k, fill = age)) +
  geom_bar(stat = "identity") +
  labs(title = "Suicide Rate (per 100k) Globally, by Age Group",
       x = "Age Group",
       y = "Suicides per 100k") +
  theme(legend.position = "none") + scale_fill_brewer(palette = "Paired") +
```

```
  scale_y_continuous(breaks = seq(0, 30, 1), minor_breaks = FALSE)

# Prepare data for plotting trends over time by age
age_time_stats <- data %>%
  group_by(year, age) %>%
  summarize(suicide_per_100k = calc_suicide_rate(cur_data()))

# Plot trends over time for suicide rates by age group
age_time_plot <- ggplot(age_time_stats, aes(x = year, y=suicide_per_100k, color = age)) +
  facet_grid(age ~ ., scales = "free_y") +
  geom_line() +  geom_point() +
  labs(title = "Suicide Rate Trends (1985-2015), by Age Group ",
       x = "Year",
       y = "Suicides per 100k") +
  theme(legend.position = "right") +scale_color_brewer(palette = "Paired") +
  scale_x_continuous(breaks = seq(1985, 2015, 5), minor_breaks = FALSE)

grid.arrange(age_plot, age_time_plot, ncol = 2)
```



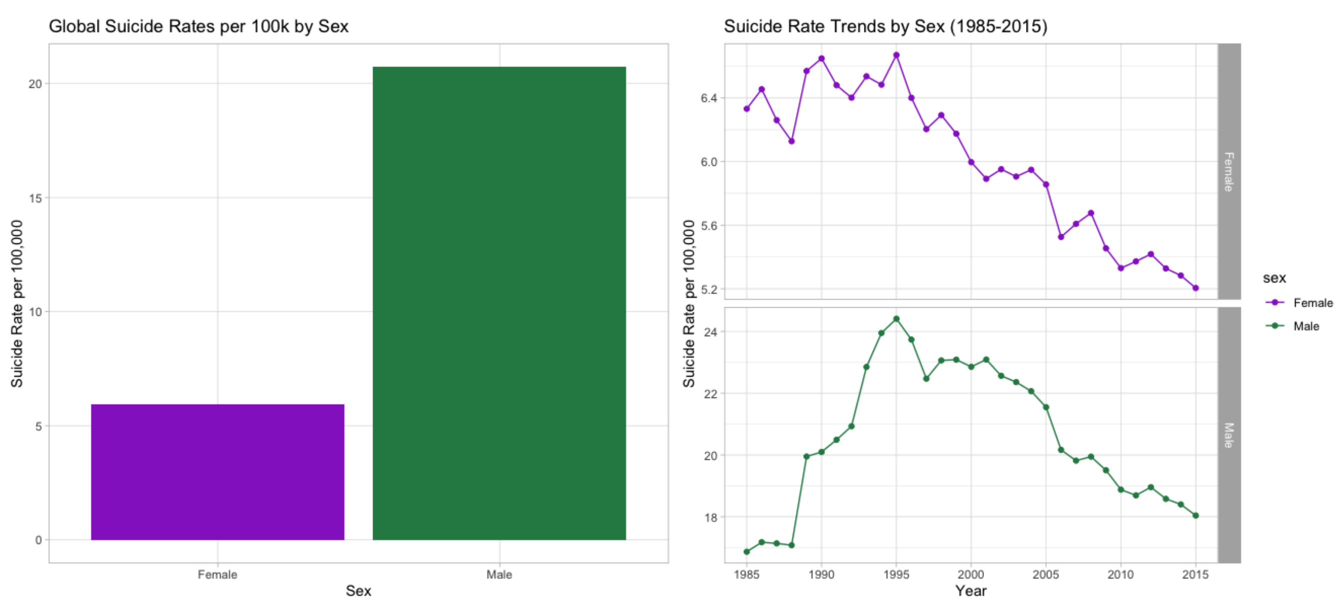
Analysis of Global Suicide Rates by Age Group:

- The data reveals a distinct correlation between age and suicide rates on a global scale, with **higher age groups consistently exhibiting higher suicide rates**.
- Starting from **1995**, there has been a **steady decline** in suicide rates across all age groups beginning from **15 years and older**.
- Individuals aged **75 and older** have experienced a substantial reduction in suicide rates, with a **decrease of approximately 50% since 1990**.
- The suicide rate within the age group, **5-14 years**, has remained relatively unchanged and minimal, with **less than 1 incident per 100,000 people per year**.

6.3 By Gender

In this section, we study the difference in suicide rates between genders and track the change in these trends from 1985 to 2015 to better understand the gender dynamics in global suicide statistics.

```r
# Generate a bar plot of global suicide rates per 100k by sex
suicide_by_gender_plot <- data %>%
  group_by(sex) %>%
  summarize(suicide_rate = calc_suicide_rate(cur_data())) %>%
  ggplot(aes(x = sex, y = suicide_rate, fill = sex)) +
  geom_bar(stat = "identity") +
  labs(title = "Global Suicide Rates per 100k by Sex",
       x = "Sex",
       y = "Suicide Rate per 100,000") + theme(legend.position = "none") +
  scale_fill_manual(values = c("Female" = "darkorchid", "Male" = "seagreen")) +
  scale_y_continuous(breaks = seq(0, 25, 5), minor_breaks = FALSE)

# Generate a line plot of suicide rate trends by sex from 1985-2015
suicide_by_gender_trend_plot <- data %>%
  group_by(year, sex) %>%
  summarize(suicide_rate = calc_suicide_rate(cur_data())) %>%
  ggplot(aes(x = year, y = suicide_rate, color = sex)) +
  facet_grid(sex ~ ., scales = "free_y") + geom_line() + geom_point() +
  labs(title = "Suicide Rate Trends by Sex (1985-2015)",
       x = "Year",
       y = "Suicide Rate per 100,000") +
  theme(legend.position = "right") +
  scale_color_manual(values = c("Female" = "darkorchid", "Male" = "seagreen")) +
  scale_x_continuous(breaks = seq(1985, 2015, 5), minor_breaks = FALSE)

grid.arrange(suicide_by_gender_plot, suicide_by_gender_trend_plot, ncol = 2)
```



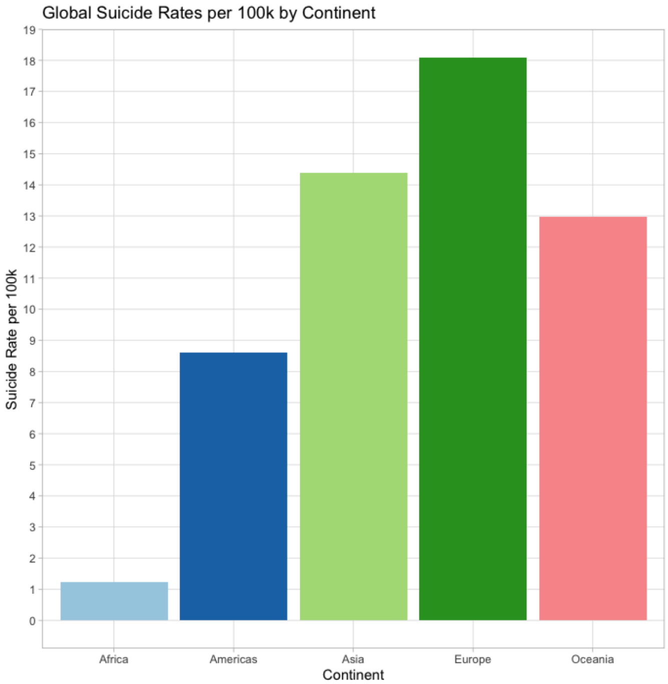
Analysis of Global Suicide Rates by Gender:

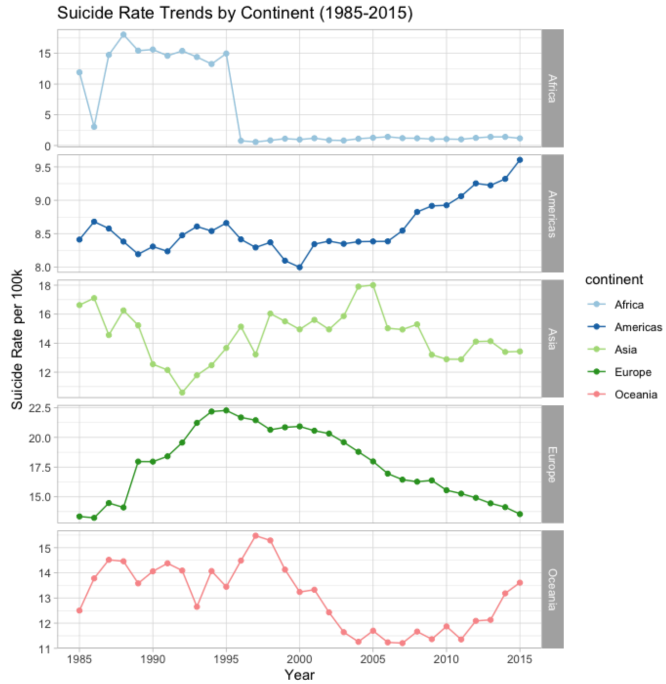
- The global suicide rate among **men** is consistently **higher**, approximately **3.5 times** that of **women**.
- A **peak in suicide rates** for both genders occurred in **1995**, with a general **decline observed subsequently**.
- The male-to-female suicide rate ratio has been fairly **stable** at **3.5 to 1** since the **mid-1990s**.
- However, during the **1980s**, the ratio was lower, at around **2.7 to 1** (male to female).

6.4 By Continent

Here, we examine suicide rates across continents to uncover regional variations and trends, offering a geographical perspective on the global suicide data.

```r
# Generate a bar plot of global suicide rates per 100k by continent
continent_barplot <- data %>%
  group_by(continent) %>%
  summarize(suicide_rate = calc_suicide_rate(cur_data())) %>%
  ggplot(aes(x = continent, y = suicide_rate, fill = continent)) +
  geom_bar(stat = "identity") +
  labs(title = "Global Suicide Rates per 100k by Continent",
       x = "Continent",
       y = "Suicide Rate per 100k") +
  theme(legend.position = "none") +
  scale_fill_brewer(palette = "Paired") +
  scale_y_continuous(breaks = seq(0, 20, 1), minor_breaks = FALSE)

# Generate a line plot of suicide rate trends by continent from 1985-2015
continent_trend_lineplot <- data %>%
  group_by(year, continent) %>%
  summarize(suicide_rate = calc_suicide_rate(cur_data())) %>%
  ggplot(aes(x = year, y = suicide_rate, color = continent)) +
  facet_grid(continent ~ ., scales = "free_y") +
  geom_line() +
  geom_point() +
  labs(title = "Suicide Rate Trends by Continent (1985-2015)",
       x = "Year",
       y = "Suicide Rate per 100k") +
  theme(legend.position = "right") +
  scale_color_brewer(palette = "Paired") +
  scale_x_continuous(breaks = seq(1985, 2015, 5), minor_breaks = FALSE)

grid.arrange(continent_barplot, continent_trend_lineplot, ncol = 2)
```





Analysis of Global Suicide Rates by Continent:

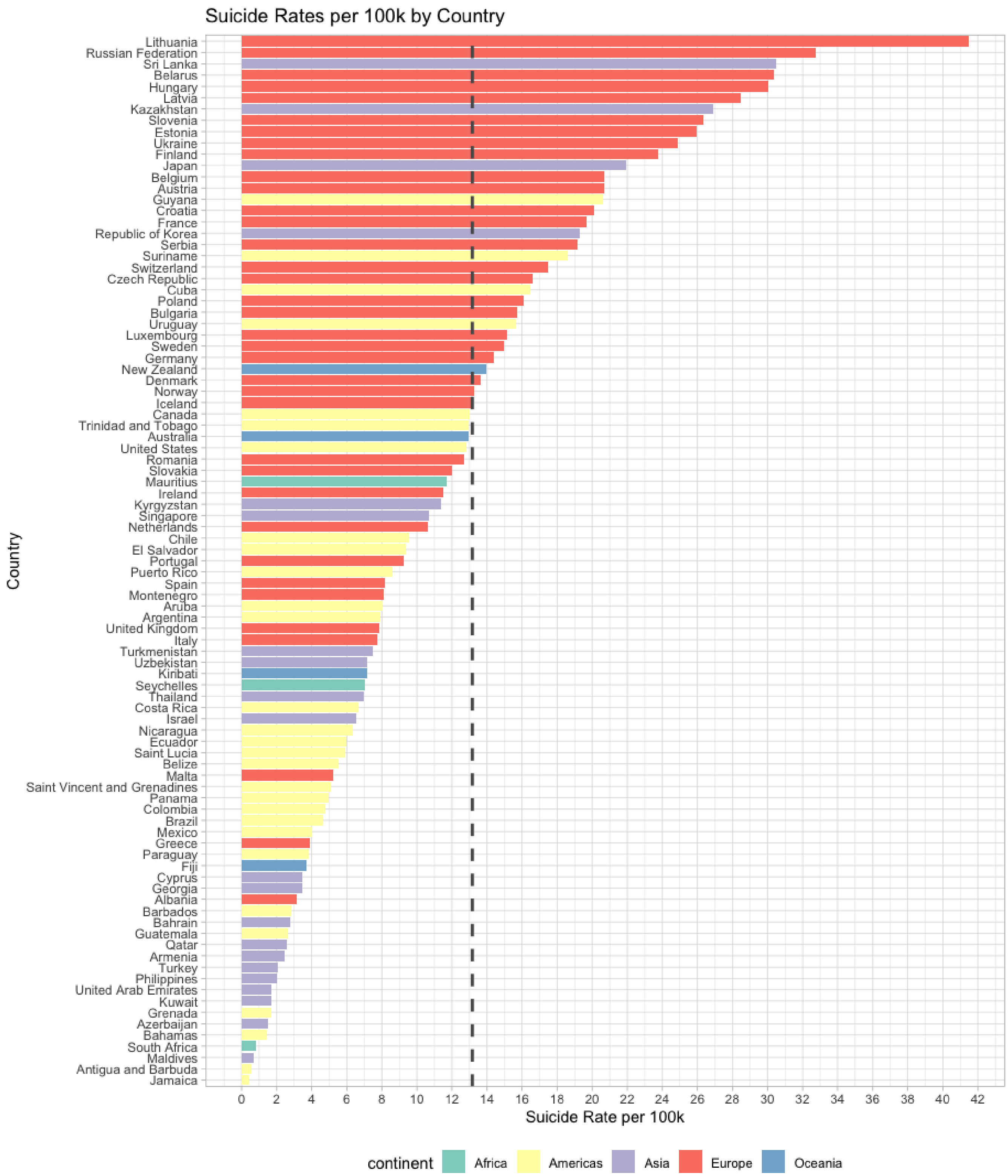
- **Europe** has the **highest** recorded suicide rates but has observed a significant **decrease of around 40%** since **1995**.
- By **2015**, **Europe's** suicide rate converged to **levels comparable to** those of **Asia** and **Oceania**.
- The data for **Africa**, which shows a relatively flat trend line, **may not be fully representative** due to limited reporting, **with only 3 countries contributing data**.
- The trends in **Oceania** and the **Americas** show patterns that are concerning and require further investigation due to their **upward trajectories**.

6.5 By Country

A country-specific bar chart was selected to provide a detailed comparison of suicide rates on a country level, highlighting the diversity in suicide trends across different countries and offering a reference point against the global average.

```
# Calculate suicide rates by country and plot
ggplot(data %>%
         group_by(country, continent) %>%
         summarize(n = n(),
                   suicide_rate = calc_suicide_rate(cur_data())) %>%
         arrange(desc(suicide_rate)),
       aes(x = reorder(country, suicide_rate), y = suicide_rate, fill = continent)) +
  geom_bar(stat = "identity") +
  geom_hline(yintercept = global_average, linetype = "dashed", color = "grey35", size =
1) +
  labs(title = "Suicide Rates per 100k by Country",
       x = "Country",
       y = "Suicide Rate per 100k") +
  coord_flip() +
```

```
scale_y_continuous(breaks = seq(0, 45, 2)) +
scale_fill_brewer(palette = "Set3") +
theme(legend.position = "bottom")
```

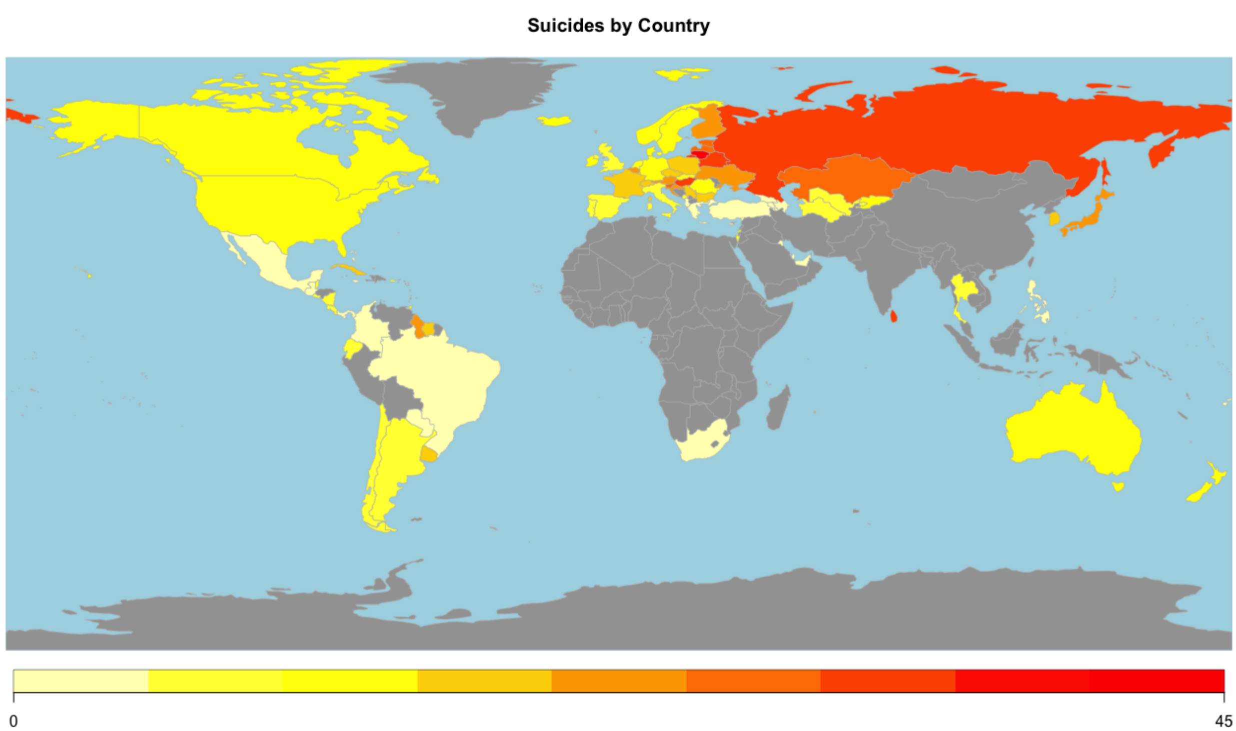


Analysis of Global Suicide Rates by Country:

- Lithuania has the highest suicide rate, **exceeding 41 per 100,000 individuals annually**.
- **European countries** are prominently represented among the nations with **higher suicide rates**, while fewer appear at the lower end of the spectrum.
- The global average suicide rate is marked by a dashed line, allowing for a quick visual comparison of which countries are above or below this benchmark.
- **The arrangement of countries indicates a regional pattern, suggesting potential cultural, economic, or policy influences on suicide rates**

The geographical heatmap below represents suicide rates over the span of our study. It highlights data scarcity in Africa and Asia, and it should be noted that data from 7 countries were omitted due to their incompleteness.

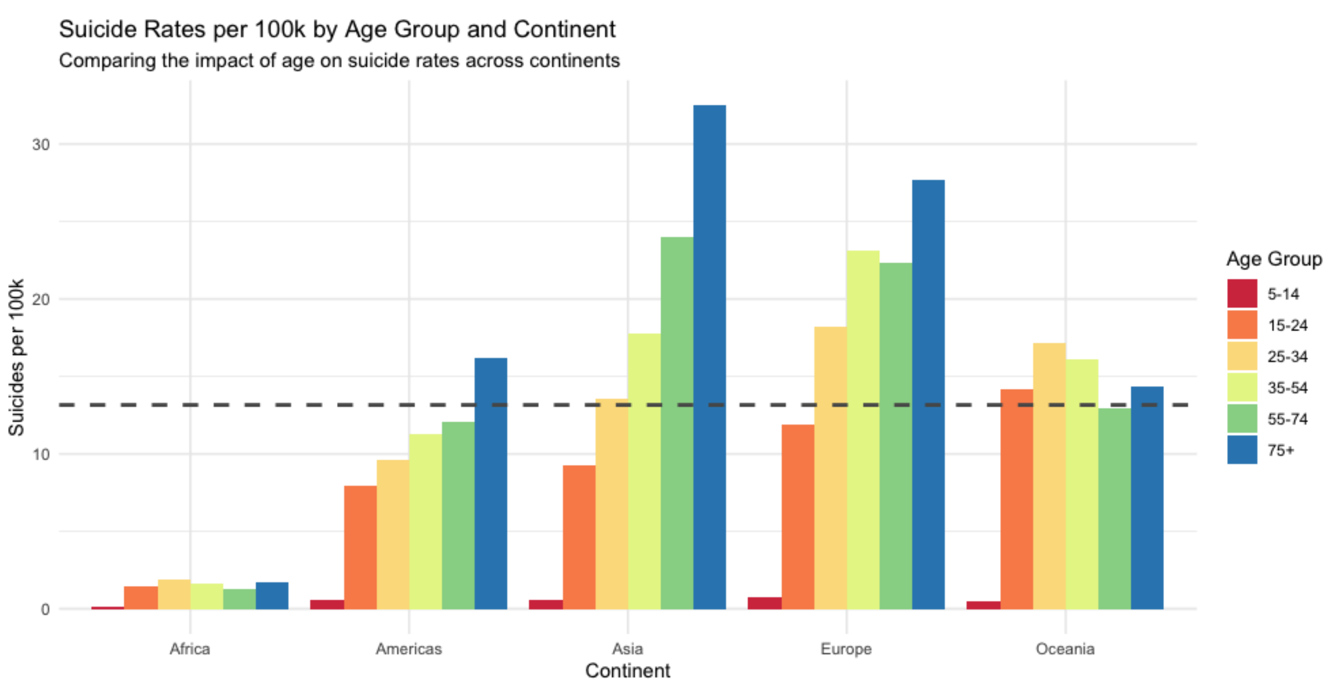
```r
# Summarize suicide rates by country and join to map data
suicide_rates_map <- data %>%
  group_by(country) %>%
  summarize(suicide_rate = calc_suicide_rate(cur_data())) %>%
  joinCountryData2Map(joinCode = "NAME", nameJoinColumn = "country")

# Plot the map with suicide rates
mapParams <- mapCountryData(
  suicide_rates_map,
  nameColumnToPlot = "suicide_rate", mapTitle = "Suicides by Country",
  colourPalette = "heat", oceanCol = "lightblue",
  missingCountryCol = "grey65", catMethod = "pretty",
)
```



6.6 Age differences, by Continent

This section's bar chart compares suicide rates by age group across continents, providing insights into how age-related suicide risks vary in different regional contexts.

```
# Prepare the summarized data for plotting
age_continent_suicides <- data %>%
  group_by(continent, age) %>%
  summarize(
    suicide_per_100k = calc_suicide_rate(cur_data())) %>%
  ungroup()

# Plot the data with ggplot
ggplot(age_continent_suicides, aes(x = continent, y = suicide_per_100k, fill = age)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_brewer(palette = "Spectral") +
  geom_hline(yintercept = global_average, linetype = "dashed", color="grey35", size=1) +
  labs(
    title = "Suicide Rates per 100k by Age Group and Continent",
    subtitle = "Comparing the impact of age on suicide rates across continents",
    x = "Continent",
    y = "Suicides per 100k",
    fill = "Age Group") +
  theme_minimal() + theme(legend.position = "right")
```



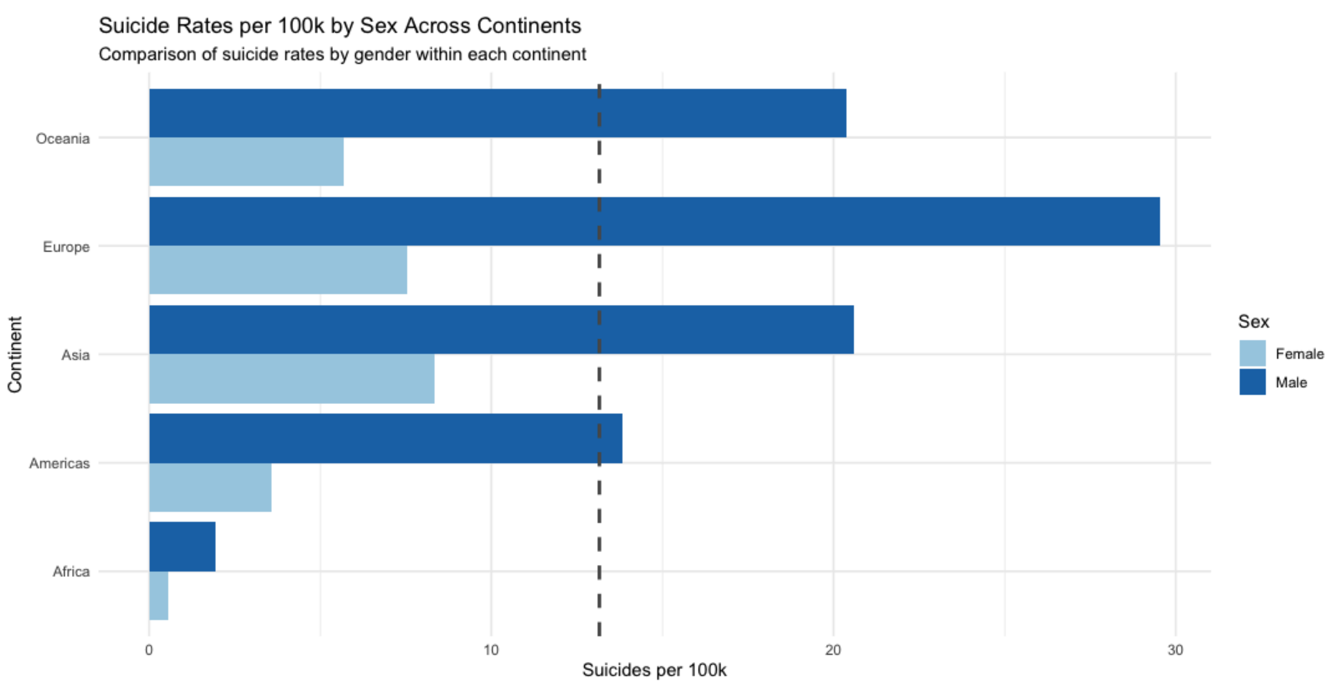
Analysis of Suicide Rates by Age and Continent:

- In the **Americas, Asia, and Europe**, which constitute the majority of the data, there is a clear trend of **increasing suicide rates with advancing age**.
- In contrast, **Oceania** and **Africa** show the highest suicide rates among the **25-34** age group.

6.7 Gender differences, by Continent

This section's visual analysis highlights the disparities in suicide rates between genders within various continents, shedding light on the intersection of geography and gender in understanding global suicide statistics.

```
# Create a bar plot for suicide rates per 100k by gender and continent
ggplot(data %>%
         group_by(continent, sex) %>%
         summarize(suicide_per_100k = calc_suicide_rate(cur_data()), .groups = 'drop'),
       aes(x = continent, y = suicide_per_100k, fill = sex)) +
  geom_bar(stat = "identity", position = "dodge") +
  geom_hline(yintercept = global_average, linetype = "dashed", color = "grey35", size =
1) +
  labs(title = "Suicide Rates per 100k by Sex Across Continents",
       subtitle = "Comparison of suicide rates by gender within each continent",
       x = "Continent",
       y = "Suicides per 100k",
       fill = "Sex") +
  coord_flip() +
  scale_fill_brewer(palette = "Paired") +
  theme_minimal() +
  theme(legend.position = "right")
```



Analysis of Suicide Rates by Sex Across Continents:

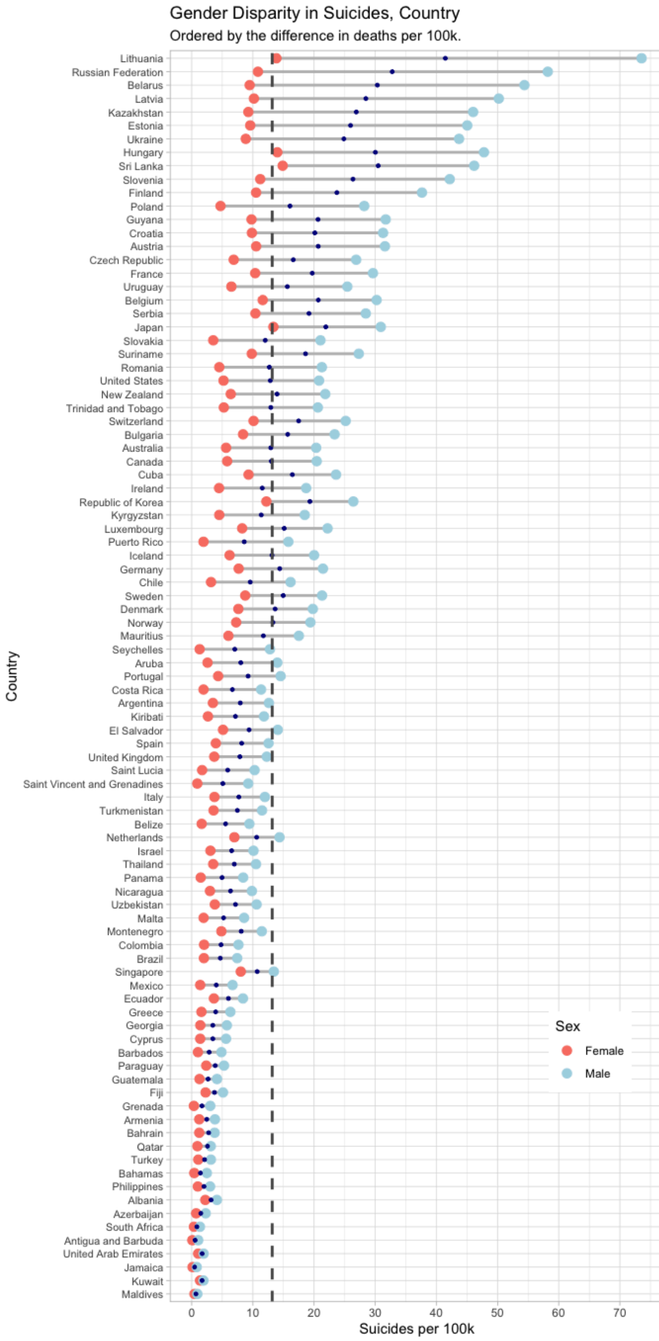
- Between 1985 and 2015, **European males** exhibited the **highest suicide risk**, with rates approximately 30 per 100k annually.
- In Asia, the **male suicide rate** was roughly **2.5 times that of females**, representing the smallest gender disparity in suicide rates within the analyzed regions.
- By comparison, **European males** had a suicide rate nearly **3.9 times higher** than that of **European females**, indicating a more pronounced gender disparity.

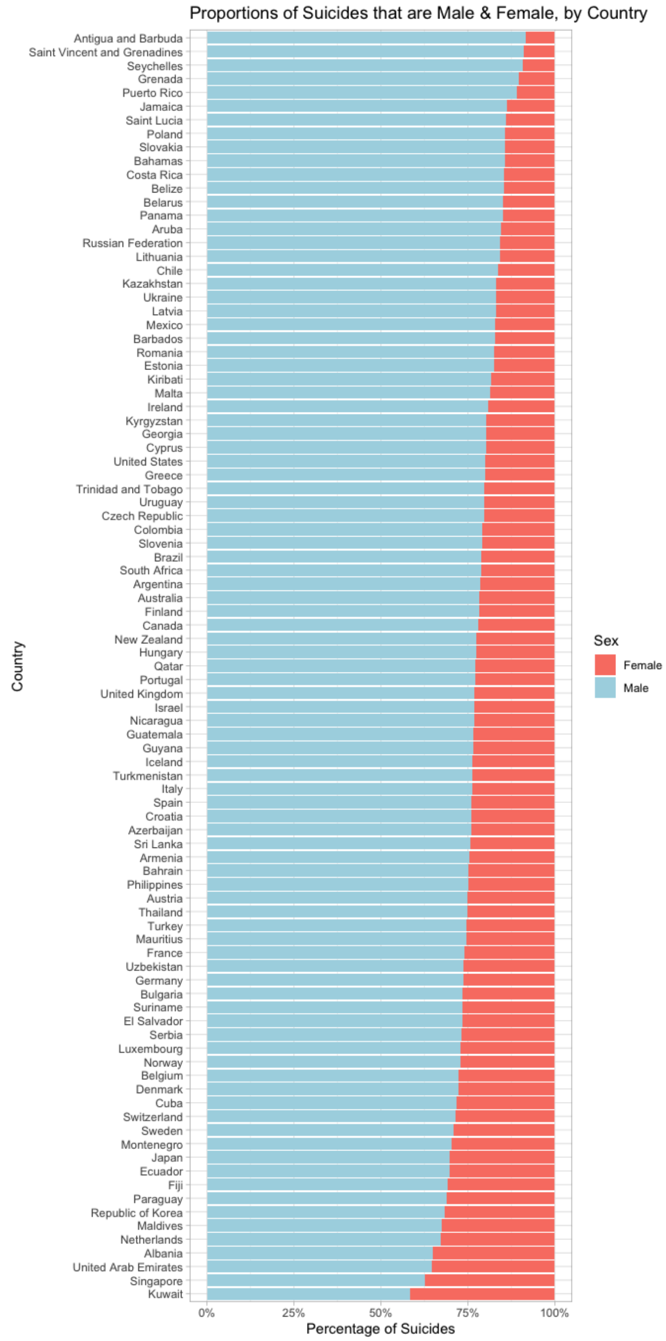
6.8 Gender differences, by Country

The plots in this section were created to analyze gender disparities in suicide rates across different countries, revealing the extent of variation between male and female suicide rates on a national scale and illustrating the proportions of suicides by gender for a comprehensive global comparison.

```
# Calculate suicides per 100,000 for overall data by country and continent
overallSuicideRates <- aggregate(cbind(suicides_no, population) ~ country + continent,
data, sum)
overallSuicideRates$suicide_per_100k <- (as.numeric(overallSuicideRates$suicides_no) /
as.numeric(overallSuicideRates$population)) * 100000

# Calculate suicides per 100,000 by country, continent, and sex
suicideRatesBySex <- aggregate(cbind(suicides_no, population) ~ country + continent +
sex, data, sum)
suicideRatesBySex$suicide_per_100k <- (as.numeric(suicideRatesBySex$suicides_no) /
as.numeric(suicideRatesBySex$population)) * 100000

# Transform data from long to wide format
wideFormat <- reshape(suicideRatesBySex, idvar = c("country", "continent"), timevar =
"sex", direction = "wide")
wideFormat <- wideFormat[order(wideFormat$suicide_per_100k.Male -
wideFormat$suicide_per_100k.Female), ]

# Order countries in both datasets
wideFormat$country <- factor(wideFormat$country, ordered = TRUE, levels =
wideFormat$country)
suicideRatesBySex$country <- factor(suicideRatesBySex$country, ordered = TRUE, levels =
wideFormat$country)

# Plotting the data
gender_disparity_plot <- ggplot() +
  geom_dumbbell(data = wideFormat, aes(y = country, x = suicide_per_100k.Female, xend =
suicide_per_100k.Male),
                color = "grey", size = 1) +
  geom_point(data = suicideRatesBySex, aes(x = suicide_per_100k, y = country, color =
sex),
             size = 3) +
  geom_point(data = overallSuicideRates, aes(x = suicide_per_100k, y = country),
             color = "darkblue", size = 1) +
  geom_vline(xintercept = global_average, linetype = 2, color = "grey35", size = 1) +
```

```
  scale_color_manual(values = c("Female" = "salmon", "Male" = "lightblue")) +
  theme(axis.text.y = element_text(size = 8),
        legend.position = c(0.85, 0.2)) +
  scale_x_continuous(breaks = seq(0, 80, 10)) +
  labs(title = "Gender Disparity in Suicides, by Country",
       x = "Suicides per 100k",
       y = "Country",
       color = "Sex")

wideFormat$Male_Proportion <- wideFormat$suicide_per_100k.Male /
(wideFormat$suicide_per_100k.Female + wideFormat$suicide_per_100k.Male)
wideFormat <- wideFormat[order(wideFormat$Male_Proportion), ]
suicideRatesBySex$country <- factor(suicideRatesBySex$country, ordered = TRUE, levels =
wideFormat$country)

# Plotting the data as stacked bar chart showing the proportion of male and female
suicides
gender_proportion_plot <- ggplot(suicideRatesBySex, aes(x = country, y =
suicide_per_100k, fill = sex)) +
  geom_bar(position = "fill", stat = "identity") +
  scale_fill_manual(values = c("Female" = "salmon", "Male" = "lightblue")) +
  scale_y_continuous(labels = scales::percent) +
  coord_flip() +
  labs(title = "Proportions of Suicides that are Male & Female, by Country",
       x = "Country",
       y = "Percentage of Suicides",
       fill = "Sex")

grid.arrange(gender_disparity_plot, gender_proportion_plot, ncol = 2)
```





Analysis of Global Suicide Rates: A Gender Perspective

- **Men** are **disproportionately represented** in suicide deaths across various countries. This observation is prominent and varied in magnitude but remains a **universal** pattern.
- Notably, while **women are statistically more likely to experience depression and entertain suicidal ideation**, **it is men who have a higher incidence of suicide completion**. This phenomenon constitutes the "**gender paradox**" [3] in suicidal behavior, which underscores the complex **interplay between mental health issues and their fatal outcomes** along gender lines.

- Further analysis reveals that **countries with higher overall suicide rates often exhibit greater gender disparities**. It suggests that **societal**, **cultural**, or perhaps **access-related** factors may exacerbate the risk of suicide among men in these regions.

In addressing the **gender paradox**, it's crucial to **recognize that men may face barriers to seeking help** due to **societal expectations** or **stigma**. These barriers can contribute to the underreporting of mental health struggles and a lack of engagement with support systems, ultimately leading to higher suicide rates.

6.9 Economic Influence on Suicide Rates: The GDP Factor

6.9.1 As a country gets richer, does it's suicide rate decrease?

In most countries, there's a **strong relationship** between the passage of **time** and the **increase in GDP per capita**, meaning that as years progress, the GDP per capita tends to rise in a linear manner.

```
# Calculate mean GDP per capita for each country and year,
# then calculate the correlation between year and GDP per capita for each country
country_year_gdp_corr <- data %>%
  group_by(country, year) %>%
  summarize(gdp_per_capita = mean(gdp_per_capita), .groups = 'drop') %>%
  group_by(country) %>%
  summarize(year_gdp_correlation = cor(year, gdp_per_capita), .groups = 'drop')

# Calculate the mean of the year and GDP per capita correlations across all countries
mean_correlation <- mean(country_year_gdp_corr$year_gdp_correlation, na.rm = TRUE)
cat("The mean correlation is:", mean_correlation, "\n")
```

Console:

```
> cat("The mean correlation is:", mean_correlation, "\n")
The mean correlation is: 0.8780947
```

I computed the pearson correlations between **'year'** and **'GDP per capita'** for each country and summarized the findings:

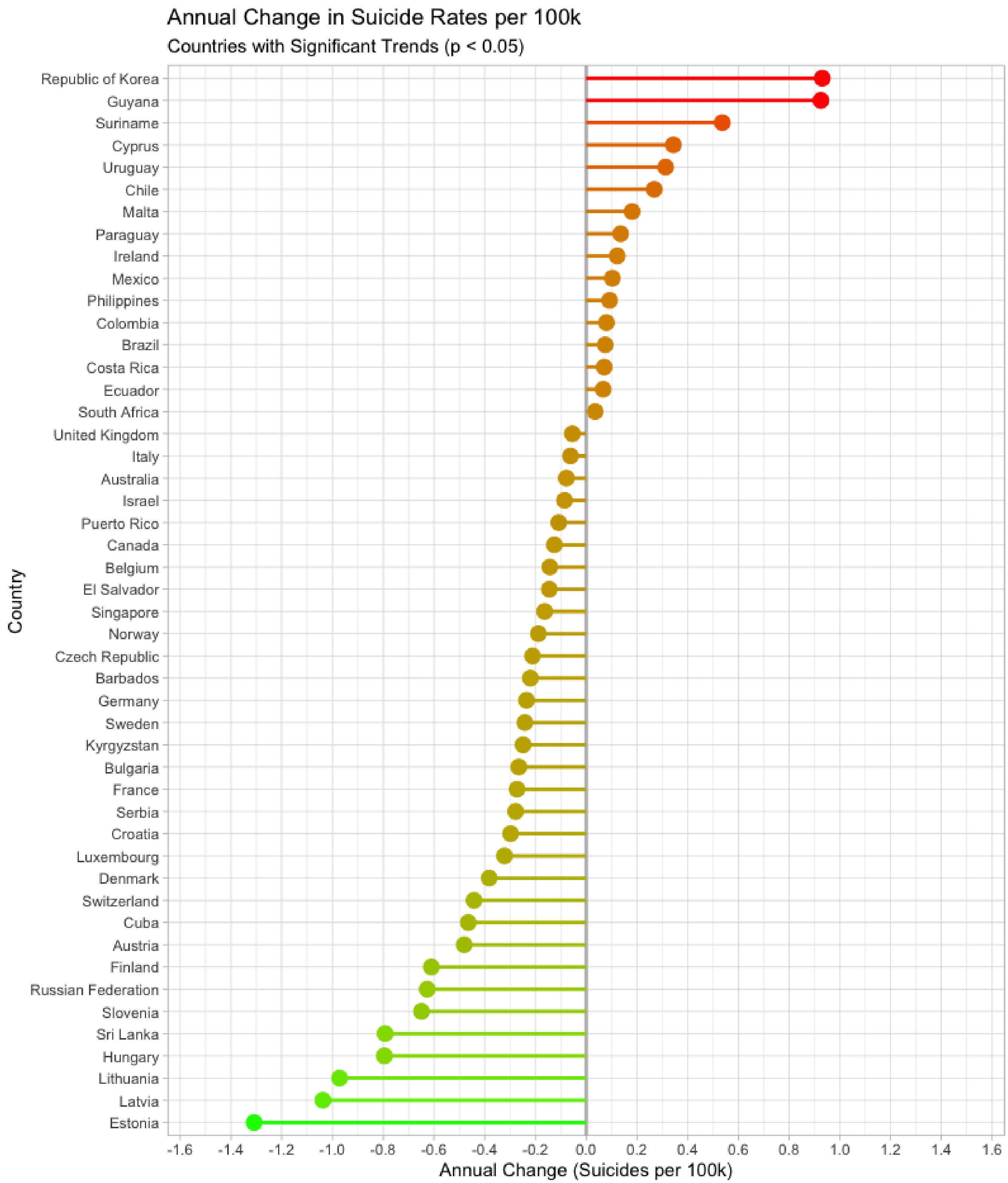
The **average correlation is 0.878**, showing a very **strong positive** linear association.

This suggests that examining the **rise in wealth per individual within a country** is almost equivalent to investigating whether a **country's suicide rate increases over time**.

We can analyze the changes in suicide rates over time within each country. Rather than graphically representing the rates of all 93 countries across time, I applied a basic linear regression to the data of each country. From this, I selected countries with a 'year' **p-value** (adjusted for multiple comparisons) that is **less than 0.05**.

Put simply, I identify countries where the suicide rate shows a linear trend of either increase or decrease over time.

```r
# Summarize data by country and year
country_summary <- data %>%
  group_by(country, year) %>%
  summarize(
    total_suicides = sum(suicides_no),
    total_population = sum(population),
    suicide_rate = (total_suicides / total_population) * 100000,
    avg_gdp_per_capita = mean(gdp_per_capita)
  )

# Fit linear models and tidy the data
trend_analysis <- country_summary %>%
  ungroup() %>%
  nest(data = -country) %>%
  # for each item in 'data', fit a linear model
  mutate(
    lm_fit = map(data, ~ lm(suicide_rate ~ year, data = .)),
    lm_results = map(lm_fit, broom::tidy)
  ) %>%
  unnest(lm_results)

# Filter significant trends and adjust p-values
sig_trends <- trend_analysis %>%
  filter(term == "year") %>%
  mutate(p_adj = p.adjust(p.value, method = "holm")) %>%
  filter(p_adj < .05) %>%
  arrange(estimate) %>%
  mutate(country = factor(country, levels = country, ordered = TRUE))

# Plotting  trends
ggplot(sig_trends, aes(x = country, y = estimate, color = estimate)) +
  geom_point(size = 4) +
  geom_hline(yintercept = 0, color = "grey", size = 1) +
  scale_color_gradient(low = "green", high = "red") +
  geom_segment(aes(y = 0, xend = country, yend = estimate), size = 1) +
  labs(
    title = "Annual Change in Suicide Rates per 100k",
    subtitle = "Countries with Significant Trends (p < 0.05)",
    x = "Country", y = "Annual Change (Suicides per 100k)"
  ) +
  scale_y_continuous(breaks = seq(-2, 2, 0.2), limits = c(-1.5, 1.5)) +
  theme(legend.position = "none") + coord_flip()
```



Insights delivered from the above visualization:

- Approximately **half** of the surveyed countries exhibit **linear changes** in suicide rates over time.
- About **two-thirds**, (32/48) countries, are showing a **decrease** in suicide rates.
- This overall **downtrend** across several nations suggests a potentially **positive shift** in the global landscape of **mental health** and **suicide prevention** efforts.

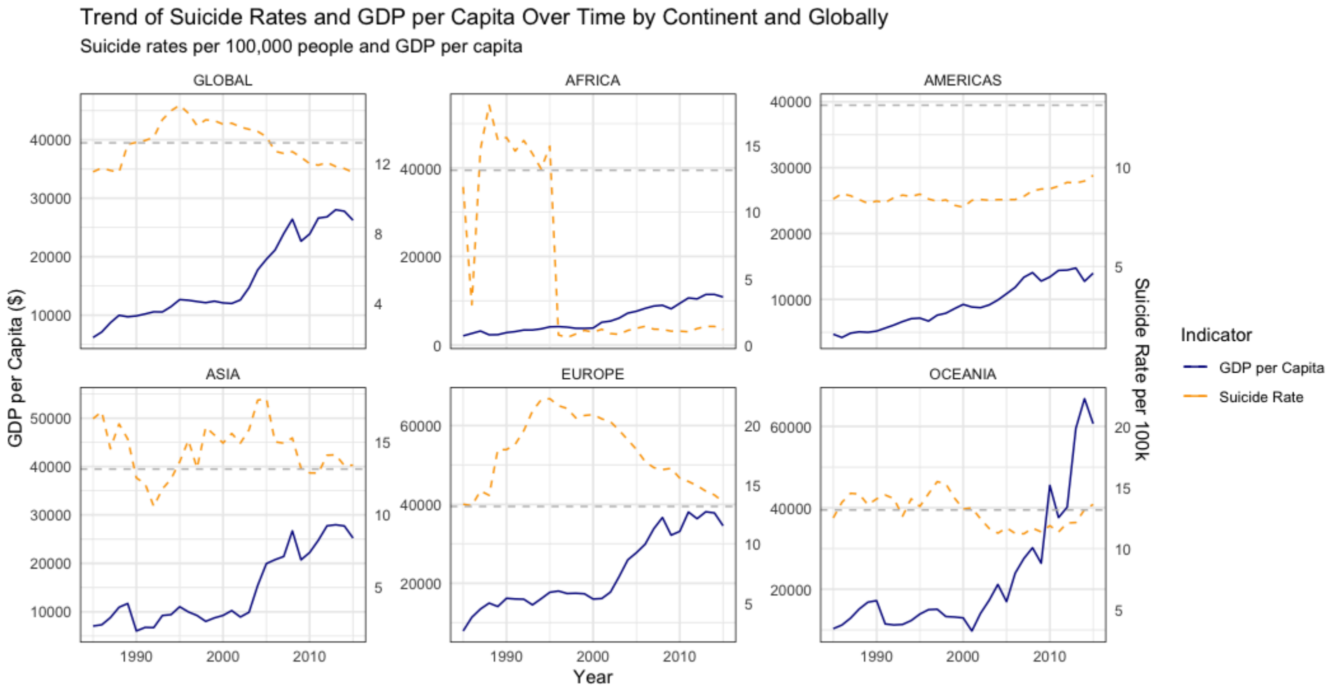
6.9.2 Economic Prosperity and Mental Well-being: A Global and Continental Analysis

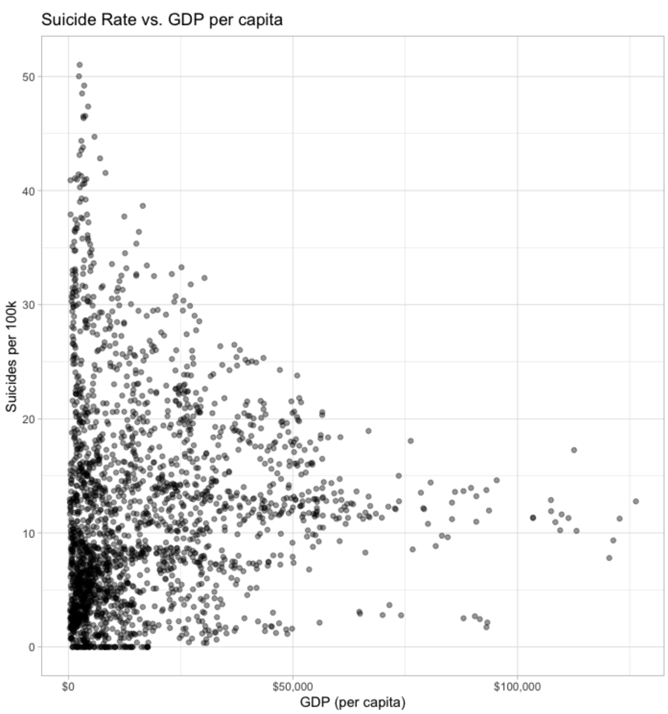
The assumption is that there might be an **inverse relationship between GDP per capita and suicide rates**, as financial prosperity is often considered a factor in overall well-being. Although GDP per capita does not provide a comprehensive measure of an individual's wealth, it serves as the available metric for this analysis. We will examine the data to determine if this hypothesis holds true.

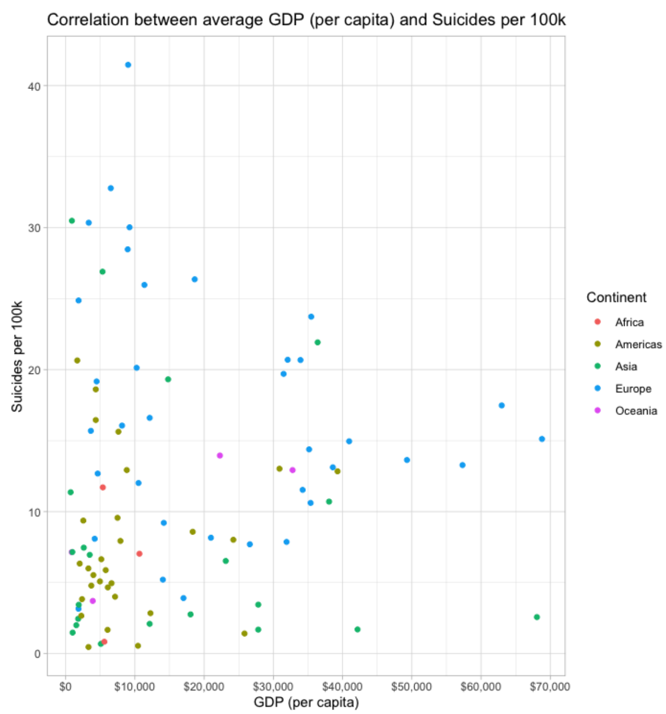
```
# Calculate global stats
global_stats <- data %>%
  group_by(year) %>%
  summarize(
    suicide_rate = calc_suicide_rate(cur_data()),
    avg_gdp_per_capita = mean(gdp_per_capita, na.rm = TRUE)) %>%
  ungroup()

# Add a 'Global' category to the global_stats
global_stats$continent <- 'GLOBAL'

# Combine the global stats with the continent stats
combined_stats <- data %>%
  group_by(year, continent) %>%
  summarize(
    suicide_rate = calc_suicide_rate(cur_data()),
    avg_gdp_per_capita = mean(gdp_per_capita, na.rm = TRUE)) %>%
  ungroup() %>%
  bind_rows(global_stats)

# Reorder so that 'Global' comes first
combined_stats$continent <- toupper(combined_stats$continent)
combined_stats$continent <- factor(combined_stats$continent,
                                   levels = c('GLOBAL',
unique(combined_stats$continent[combined_stats$continent != 'GLOBAL'])))

# Determine the scaling factor for all data combined
max_gdp <- max(combined_stats$avg_gdp_per_capita, na.rm = TRUE)
max_suicide_rate <- max(combined_stats$suicide_rate, na.rm = TRUE)
scaling_factor <- max_suicide_rate / max_gdp

# Create the plots
ggplot(combined_stats, aes(x = year)) +
  geom_line(aes(y = avg_gdp_per_capita, color = "GDP per Capita"), linetype=1) +
  geom_line(aes(y = suicide_rate / scaling_factor, color = "Suicide Rate"), linetype=2) +
  geom_hline(yintercept = global_average / scaling_factor, linetype = "dashed", color =
"grey") +
  labs(title = "Trend of Suicide Rates and GDP per Capita Over Time by Continent and
Globally",
       subtitle = "Suicide rates per 100,000 people and GDP per capita",
       x = "Year", y = "GDP per Capita ($)",
       color = "Indicator") +
  scale_y_continuous(
```

```
    sec.axis = sec_axis(~ . * scaling_factor, name = "Suicide Rate per 100k")
  ) +
  facet_wrap(~ continent, scales = 'free_y') +
  theme_minimal() + theme(panel.border = element_rect(colour = "black", fill=NA,
size=0.5)) + scale_color_manual(values = c("darkblue", "orange"))
```

```
# Create a plot comparing suicide rate against GDP per capita
suicide_rate_plot <- data %>%
  group_by(country, gdp_per_capita) %>%
  summarize(suicide_rate = calc_suicide_rate(cur_data()), .groups = 'drop') %>%
  ggplot(aes(x = gdp_per_capita, y = suicide_rate)) +
  scale_x_continuous(labels = scales::dollar_format(prefix = "$")) + geom_point(alpha =
0.4) +
  labs(
    title = "Suicide Rate vs. GDP per capita",
    x = "GDP (per capita)",
    y = "Suicides per 100k")

# Compute the mean GDP per capita and suicide rates per 100k and create the plot
mean_gdp_suicide_rate_plot <- data %>%
  group_by(country, continent) %>%
  summarize(
    suicide_per_100k = calc_suicide_rate(cur_data()),
    gdp_per_capita = mean(gdp_per_capita, na.rm = TRUE),
    .groups = 'drop') %>%
  ggplot(aes(x = gdp_per_capita, y = suicide_per_100k, color = continent)) + geom_point()
+
  scale_x_continuous(labels = scales::dollar_format(prefix = "$"), breaks = seq(0, 70000,
10000)) +
  labs(
    title = "Correlation between average GDP (per capita) and Suicides per 100k",
    x = "GDP (per capita)",
    y = "Suicides per 100k",
    color = "Continent")

grid.arrange(suicide_rate_plot, mean_gdp_suicide_rate_plot, ncol = 2)
```







Analysis of Economic Indicators and Suicide Rates: A Multifaceted Perspective

The interrelation of suicide rates with GDP per capita across the globe presents a multifaceted narrative:

- **Economic Growth vs. Mental Health**: An overarching trend suggests that **rising GDP per capita** is generally accompanied by **falling suicide rates**, yet this trend is not uniform across the board, indicating other variables at play.

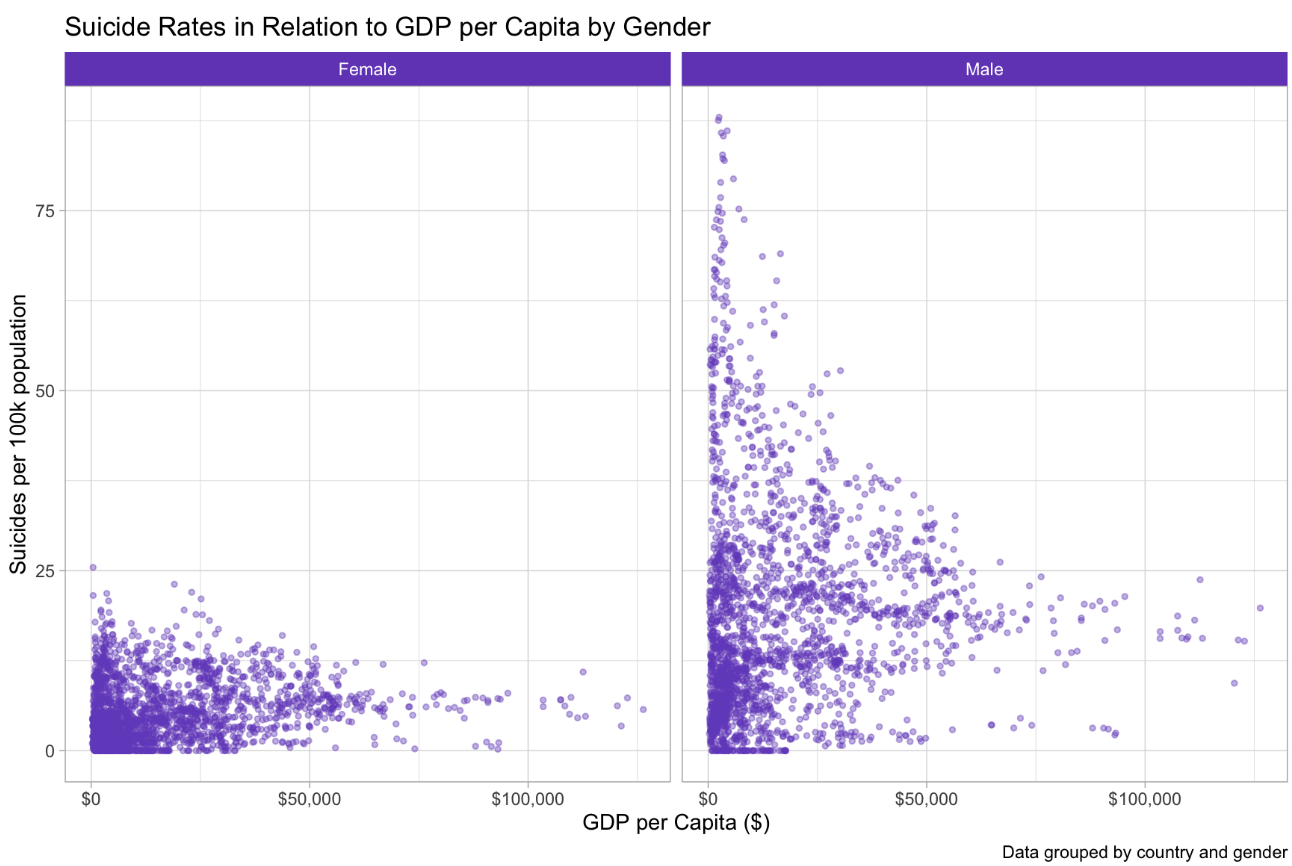
- **Dispersion of Data**: There's a wide dispersion of suicide rates at lower levels of GDP per capita, suggesting variability in how economic factors might influence suicide rates in less wealthy countries.
- **High GDP - Low Suicide Rate Trend**: As **GDP per capita increases**, there appears to be a cluster of data points with **lower suicide rates**, suggesting that higher GDP per capita might be associated with lower suicide rates. This would support the hypothesis of a negative correlation, at least in the upper tiers of GDP per capita. For example, **Oceania** (purple dots) has fewer data points, but these seem to indicate high GDP per capita with lower suicide rates.
- **Lower GDP, Higher Suicide Rates**: Some continents, like **Africa** (red dots), show a concentration of countries with lower GDP per capita and generally higher suicide rates compared to other continents.
- **Asia's Distribution**: Asian countries (cyan dots) show a broad distribution across GDP levels, with generally lower suicide rates, especially at higher GDP per capita levels.
- **Diminishing Returns on Wealth**: There appears to be a saturation point beyond which increased GDP does not equate to lower suicide rates, hinting at a complex dynamic where additional wealth may not yield further mental health benefits.
- **Outliers Spotlighting Other Influences**: Outliers with high suicide rates at high GDP levels underscore that economic indicators do not singularly dictate mental health status, and that cultural or systemic factors may have significant impacts.

In summary, while there is some correlation between GDP per capita and suicide rates, the relationship is layered, complex, and non-linear. This analysis underscores the imperative to delve beyond economic indicators to fully grasp the determinants of mental health and to explore the multifaceted drivers of suicide rates globally.

The visualization on the following page explores the correlation between GDP per capita and age-related suicide trends, and gender:

- There appears to be an **age-related** trend in suicide rates in relation to GDP per capita, with the data suggesting that suicide rates vary **more** significantly at **lower GDP** levels among **older age groups**.
- The relationship between **gender** and **suicide rates** in the context of GDP per capita reveals that **men** have a **noticeably higher** suicide rate at lower GDP levels compared to women.


Suicide Rates by Age Group in relation to GDP per Capita
5-14
15-24
25-34
35-54
55-74
75+
Suicide Rate per 100k population
125
100
75
50
25
0
$0
$50,000
$100,000
GDP per capita ($)
Data grouped by country and age


Suicide Rates in Relation to GDP per Capita by Gender
Female
Male
Suicides per 100k population
75
50
25
0
$0
$50,000
$100,000
GDP per Capita ($)
Data grouped by country and gender

fh272

7. Conclusion

Our comprehensive analysis of global suicide trends over three decades has unearthed significant insights into the patterns and drivers of suicide rates across various demographics. The data has highlighted a notable decline in global suicide rates, with particular emphasis on the reduction in higher age groups and the consistent gender disparity favoring male suicides.

i. **Global Trends and Economic Correlations**: A key finding is the correlation between economic prosperity and suicide rates. While a general trend of decreasing suicide rates accompanies economic growth, the relationship is intricate, suggesting that economic factors alone cannot fully explain suicide trends.

ii. **Age and Gender Dynamics**: The study highlights how age and gender differently affect suicide rates. Older people are seeing a notable drop in suicide rates. However, there's still a major issue with more men committing suicide than women, known as the "gender paradox." This points out the importance of having mental health support that specifically addresses these differences.

iii. **Regional and Continental Variances**: The study also reveals significant regional differences. Europe's high suicide rates have shown a remarkable decline, while trends in Oceania and the Americas warrant further investigation. This variability underscores the influence of cultural, economic, and policy factors on suicide rates.

iv. **Implications and Future Directions**: The findings of this report suggest a complex interplay of economic, social, and psychological factors in determining suicide rates. There is a need for holistic approaches in suicide prevention that consider these multifaceted influences. Future research should explore the nuances of these relationships further, particularly in regions with limited data.

In conclusion, while progress has been made, the battle against global suicide requires ongoing effort, research, and policy adaptation to address its multifaceted nature effectively.

8. Future Scope

Here are some potential future studies that could be pursued to expand on the topic of global suicide trends and analysis:

Incorporating Additional Socioeconomic and Demographic Factors: Future studies could delve deeper by including more comprehensive socioeconomic and demographic features such as **Human Development Index (HDI)**, **generational differences**, **education levels**, and e**mployment status**. This would allow for a more nuanced understanding of the complex interplay between various factors and suicide rates.

Detailed Geographical Analysis: A focused analysis at the continent or country level could reveal specific root causes and trends within geographical regions. This would involve examining local cultural, economic, and policy factors that influence suicide rates, providing insights into targeted prevention strategies.

Updating and Extending the Data Range: Expanding the dataset to include more recent years (2016-2023) would be crucial for understanding current trends and changes in global suicide rates. This could also involve analyzing the impact of recent global events, such as the COVID-19 pandemic, on suicide trends.

Agent-Based Simulation for Predictive Analysis: Implementing agent-based modeling [5] to simulate and predict factors influencing suicide can provide a dynamic understanding of how individual behaviors and societal factors interact to affect suicide rates. This approach would allow for the testing of various hypotheses in a controlled virtual environment.

Digital Phenotyping and Real-Time Suicide Risk Assessment: Employing digital phenotyping [6] and real-time monitoring to assess suicide risk can offer immediate insights into individual mental states and behaviors. This approach can potentially enable timely interventions and personalized care strategies.

Suicide Rates Machine Learning and Prediction: Utilizing machine learning models to predict suicide rates based on demographic, socioeconomic, and health data can provide valuable insights for prevention strategies.

9. References

[1] World Health Organization Fact Sheet, https://www.who.int/news-room/fact-sheets/detail/suicide
[2] Global Burden of Disease Study 2016, https://pubmed.ncbi.nlm.nih.gov/31339847
[3] Impact of 2008 global economic crisis on suicide, https://pubmed.ncbi.nlm.nih.gov/24046155/
[4] Suicide Rates Overview 1985 to 2016, https://www.kaggle.com/datasets/russellyates88/suicide-rates-overview-1985-to-2016
[4] Gender differences in suicide, https://en.wikipedia.org/wiki/Gender_differences_in_suicide
[5] Agent-based modeling, https://www.pnas.org/doi/10.1073/pnas.082080899
[6] Digital phenotyping of suicidal thoughts, https://pubmed.ncbi.nlm.nih.gov/29637663/