Final Project - Suicide Rates 3

Radhakrishnan,Venkat

2023-03-01

Introduction

Suicide is a major public health concern and a leading cause of death worldwide. The suicide rate dataset provides valuable information on suicide rates, population, and other factors that can help us understand the factors contributing to suicide and develop prevention strategies. Objective of this paper is to analyze the suicide rate dataset, identify trends, and explore the relationship between suicide rates and various predictors.

Problem Statement

The suicide rate dataset from kaggle contains information on suicide rates, population, and various predictors such as age, gender, and economic indicators. The paper aim to understand the factors that contribute to suicide rates and identify the most significant predictors that can help in developing effective prevention strategies.

Approach

To answer the research questions, Dataset is cleaned and pre-processed to ensure the validity and reliability of the data. Then, descriptive statistics and visualizations are used to describe the overall trend and to compare the suicide rates among different age groups, genders, and countries. Finally, will use correlation analysis to investigate any relationships between suicide rates and other factors.

Data

The dataset is pulled from multiple sources mentioned below.

1. Kaggle : https://www.kaggle.com/datasets/russellyates88/suicide-rates-overview-1985-to-2016
2. United Nations Development Program. (2018). Human development index (HDI). Retrieved from http://hdr.undp.org/en/indicators/137506
3. World Bank. (2018). World development indicators: GDP (current US$) by country:1985 to 2016. Retrieved from http://databank.worldbank.org/data/source/world-development-indicators#
4. World Health Organization. (2018). Suicide prevention. Retrieved from http://www.who.int/mental_health/suicide-prevention/en/

Analysis

The approach provides a comprehensive analysis of the trend in suicide rates from 1985 to 2016 and identifies any significant differences and correlations between suicide rates and other factors. This information can be useful for policymakers, public health professionals, and researchers to better understand the extent of the problem and to develop effective interventions to prevent suicide.

Required Packages

```
## Loading the required libraries
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   1.0.0
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(scales)
```

```
##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
##
## The following object is masked from 'package:readr':
##
##     col_factor
```

```
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
##
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
library(dplyr)
library(countrycode)
library(lmtest)
```

```
## Loading required package: zoo
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

Loading the data and making required changes

```
## Setting up the local directory and loading the data
setwd("C:/Users/krish/OneDrive - Bellevue University/DS520/Week11")
suicide_rates <- read_csv("suicide_rates.csv")
```

```
## Rows: 27820 Columns: 12
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): country, sex, age, country-year, generation
## dbl (6): year, suicides_no, population, suicides/100k pop, HDI for year, gdp...
## num (1): gdp_for_year ($)
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(suicide_rates)
```

```
## # A tibble: 6 x 12
##   country  year sex    age       suici~1 popul~2 suici~3 count~4 HDI f~5 gdp_f~6
##   <chr>   <dbl> <chr>  <chr>       <dbl>   <dbl>   <dbl> <chr>     <dbl>   <dbl>
## 1 Albania  1987 male   15-24 ye~      21  312900    6.71 Albani~      NA  2.16e9
## 2 Albania  1987 male   35-54 ye~      16  308000    5.19 Albani~      NA  2.16e9
## 3 Albania  1987 female 15-24 ye~      14  289700    4.83 Albani~      NA  2.16e9
## 4 Albania  1987 male   75+ years       1   21800    4.59 Albani~      NA  2.16e9
## 5 Albania  1987 male   25-34 ye~       9  274300    3.28 Albani~      NA  2.16e9
## 6 Albania  1987 female 75+ years       1   35600    2.81 Albani~      NA  2.16e9
## # ... with 2 more variables: `gdp_per_capita ($)` <dbl>, generation <chr>, and
## #   abbreviated variable names 1: suicides_no, 2: population,
## #   3: `suicides/100k pop`, 4: `country-year`, 5: `HDI for year`,
## #   6: `gdp_for_year ($)`
```

```
summary(suicide_rates)
```

```
##    country               year          sex                age
##  Length:27820       Min.   :1985   Length:27820       Length:27820
##  Class :character   1st Qu.:1995   Class :character   Class :character
##  Mode  :character   Median :2002   Mode  :character   Mode  :character
##                     Mean   :2001
```

```
##                    3rd Qu.:2008
##                    Max.   :2016
##
##   suicides_no        population       suicides/100k pop country-year
##  Min.   :    0.0   Min.   :     278   Min.   :  0.00    Length:27820
##  1st Qu.:    3.0   1st Qu.:   97498   1st Qu.:  0.92    Class :character
##  Median :   25.0   Median :  430150   Median :  5.99    Mode  :character
##  Mean   :  242.6   Mean   : 1844794   Mean   : 12.82
##  3rd Qu.:  131.0   3rd Qu.: 1486143   3rd Qu.: 16.62
##  Max.   :22338.0   Max.   :43805214   Max.   :224.97
##
##   HDI for year   gdp_for_year ($)    gdp_per_capita ($)  generation
##  Min.   :0.483   Min.   :4.692e+07   Min.   :   251     Length:27820
##  1st Qu.:0.713   1st Qu.:8.985e+09   1st Qu.:  3447     Class :character
##  Median :0.779   Median :4.811e+10   Median :  9372     Mode  :character
##  Mean   :0.777   Mean   :4.456e+11   Mean   : 16866
##  3rd Qu.:0.855   3rd Qu.:2.602e+11   3rd Qu.: 24874
##  Max.   :0.944   Max.   :1.812e+13   Max.   :126352
##  NA's   :19456
```

```
colnames(suicide_rates)
```

```
##  [1] "country"            "year"               "sex"
##  [4] "age"                "suicides_no"        "population"
##  [7] "suicides/100k pop"  "country-year"       "HDI for year"
## [10] "gdp_for_year ($)"   "gdp_per_capita ($)" "generation"
```

Data Wrangling

1. A new column Continent is added based on *countryCode* package.
2. Most of the columns add value and hence not removed.
3. Columns with Spaces are renamed to have better naming convention.
4. Removing Outliers - Looking at the boxplot output of suicide_rate columns, there are significant number of outliers more the 75 and this data is removed.
5. Histograms of the Variables are plotted and analysed for each variable.

suicides_no - shows a long tail to the right, indicating a small number of countries with a high number of suicides. The boxplot for suicides_no also shows several points beyond the upper whisker, which are potential outliers.

Population - Similarly, the histogram of population show a small number of countries with very large populations, which are potential outliers.

Suicides/100k pop and **GDP per Capita** histograms point that there are no extreme outliers. However, there are some values for from the central tendency, indicating variance and skewness in the data.

HDI for Year - There are some values far from the majority of the values indiacting possible outliers. Also, most of the HDI details are available after year 2000 which might impact the overall analysis

```
# Add a new column for continent based on country
suicide_rates$continent <- countrycode(suicide_rates$country, "country.name", "continent")

# View the first few rows of the updated dataset
head(suicide_rates)
```

```
## # A tibble: 6 x 13
##   country  year sex    age       suici~1 popul~2 suici~3 count~4 HDI f~5 gdp_f~6
##   <chr>   <dbl> <chr>  <chr>       <dbl>   <dbl>   <dbl> <chr>     <dbl>   <dbl>
## 1 Albania  1987 male   15-24 ye~      21  312900    6.71 Albani~      NA  2.16e9
## 2 Albania  1987 male   35-54 ye~      16  308000    5.19 Albani~      NA  2.16e9
## 3 Albania  1987 female 15-24 ye~      14  289700    4.83 Albani~      NA  2.16e9
## 4 Albania  1987 male   75+ years       1   21800    4.59 Albani~      NA  2.16e9
## 5 Albania  1987 male   25-34 ye~       9  274300    3.28 Albani~      NA  2.16e9
## 6 Albania  1987 female 75+ years       1   35600    2.81 Albani~      NA  2.16e9
## # ... with 3 more variables: `gdp_per_capita ($)` <dbl>, generation <chr>,
## #   continent <chr>, and abbreviated variable names 1: suicides_no,
## #   2: population, 3: `suicides/100k pop`, 4: `country-year`,
## #   5: `HDI for year`, 6: `gdp_for_year ($)`
```

```
# Clean data
suicide_rates <- suicide_rates %>%
  rename( age_group = "age", suicide_rate = "suicides/100k pop", gdp_per_capita = "gdp_per_capita ($)",
  mutate(age_group = gsub(" years", "", age_group)) %>% # remove " years" from age group values
  filter(year >= 1985) # keep data from 1985 onwards

# View the first few rows of the cleaned data frame
head(suicide_rates)
```

```
## # A tibble: 6 x 13
##   country  year sex    age_group suici~1 popul~2 suici~3 count~4 HDI_f~5 gdp_f~6
##   <chr>   <dbl> <chr>  <chr>       <dbl>   <dbl>   <dbl> <chr>     <dbl>   <dbl>
## 1 Albania  1987 male   15-24          21  312900    6.71 Albani~      NA  2.16e9
## 2 Albania  1987 male   35-54          16  308000    5.19 Albani~      NA  2.16e9
## 3 Albania  1987 female 15-24          14  289700    4.83 Albani~      NA  2.16e9
## 4 Albania  1987 male   75+             1   21800    4.59 Albani~      NA  2.16e9
## 5 Albania  1987 male   25-34           9  274300    3.28 Albani~      NA  2.16e9
## 6 Albania  1987 female 75+             1   35600    2.81 Albani~      NA  2.16e9
## # ... with 3 more variables: gdp_per_capita <dbl>, generation <chr>,
## #   continent <chr>, and abbreviated variable names 1: suicides_no,
## #   2: population, 3: suicide_rate, 4: `country-year`, 5: HDI_for_year,
## #   6: `gdp_for_year ($)`
```

```
# View summary statistics for the numeric variables
summary(suicide_rates)
```

```
##    country               year          sex             age_group
##  Length:27820       Min.   :1985   Length:27820       Length:27820
##  Class :character   1st Qu.:1995   Class :character   Class :character
##  Mode  :character   Median :2002   Mode  :character   Mode  :character
##                     Mean   :2001
##                     3rd Qu.:2008
##                     Max.   :2016
##
##   suicides_no        population        suicide_rate    country-year
##  Min.   :    0.0   Min.   :     278   Min.   :  0.00   Length:27820
##  1st Qu.:    3.0   1st Qu.:   97498   1st Qu.:  0.92   Class :character
##  Median :   25.0   Median :  430150   Median :  5.99   Mode  :character
##  Mean   :  242.6   Mean   : 1844794   Mean   : 12.82
```

```
##  3rd Qu.:  131.0   3rd Qu.: 1486143   3rd Qu.: 16.62
##  Max.   :22338.0   Max.   :43805214   Max.   :224.97
##
##   HDI_for_year   gdp_for_year ($)    gdp_per_capita    generation
##  Min.   :0.483   Min.   :4.692e+07   Min.   :   251   Length:27820
##  1st Qu.:0.713   1st Qu.:8.985e+09   1st Qu.:  3447   Class :character
##  Median :0.779   Median :4.811e+10   Median :  9372   Mode  :character
##  Mean   :0.777   Mean   :4.456e+11   Mean   : 16866
##  3rd Qu.:0.855   3rd Qu.:2.602e+11   3rd Qu.: 24874
##  Max.   :0.944   Max.   :1.812e+13   Max.   :126352
##  NA's   :19456
##   continent
##  Length:27820
##  Class :character
##  Mode  :character
##
##
##
##
```

```
# View the number of observations by country
table(suicide_rates$country)
```

```
##
##                     Albania          Antigua and Barbuda
##                         264                          324
##                   Argentina                      Armenia
##                         372                          298
##                       Aruba                    Australia
##                         168                          360
##                     Austria                   Azerbaijan
##                         382                          192
##                     Bahamas                      Bahrain
##                         276                          252
##                    Barbados                      Belarus
##                         300                          252
##                     Belgium                       Belize
##                         372                          336
##      Bosnia and Herzegovina                       Brazil
##                          24                          372
##                    Bulgaria                   Cabo Verde
##                         360                           12
##                      Canada                        Chile
##                         348                          372
##                    Colombia                   Costa Rica
##                         372                          360
##                     Croatia                         Cuba
##                         262                          288
##                      Cyprus                Czech Republic
##                         178                          322
##                     Denmark                     Dominica
##                         264                           12
##                     Ecuador                  El Salvador
##                         372                          288
```

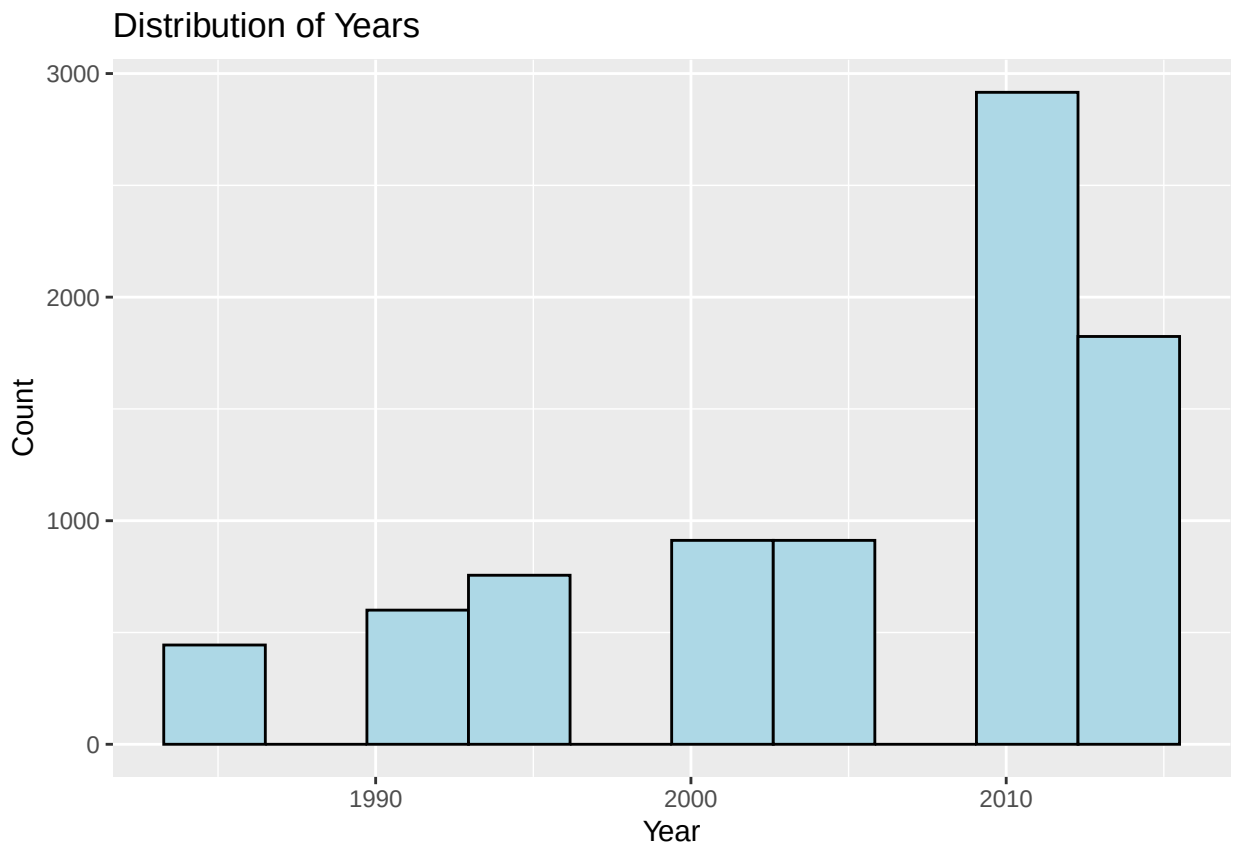
```
##                      Estonia                         Fiji
##                          252                          132
##                      Finland                       France
##                          348                          360
##                      Georgia                      Germany
##                          264                          312
##                       Greece                      Grenada
##                          372                          310
##                    Guatemala                       Guyana
##                          360                          300
##                      Hungary                      Iceland
##                          310                          382
##                      Ireland                       Israel
##                          360                          372
##                        Italy                      Jamaica
##                          372                          204
##                        Japan                   Kazakhstan
##                          372                          312
##                     Kiribati                       Kuwait
##                          132                          300
##                    Kyrgyzstan                       Latvia
##                          312                          252
##                    Lithuania                   Luxembourg
##                          262                          372
##                        Macau                     Maldives
##                           12                          120
##                        Malta                    Mauritius
##                          372                          382
##                       Mexico                     Mongolia
##                          372                           10
##                   Montenegro                  Netherlands
##                          120                          382
##                  New Zealand                    Nicaragua
##                          348                           72
##                       Norway                         Oman
##                          360                           36
##                       Panama                     Paraguay
##                          300                          324
##                  Philippines                       Poland
##                          180                          288
##                     Portugal                  Puerto Rico
##                          324                          372
##                        Qatar              Republic of Korea
##                          178                          372
##                      Romania             Russian Federation
##                          334                          324
##         Saint Kitts and Nevis                  Saint Lucia
##                           36                          336
## Saint Vincent and Grenadines                   San Marino
##                          300                           36
##                       Serbia                   Seychelles
##                          216                          216
##                    Singapore                     Slovakia
##                          372                          264
```

```
##                    Slovenia                South Africa
##                         252                         240
##                       Spain                   Sri Lanka
##                         372                         132
##                    Suriname                      Sweden
##                         336                         358
##                 Switzerland                    Thailand
##                         252                         334
##          Trinidad and Tobago                      Turkey
##                         324                          84
##                Turkmenistan                     Ukraine
##                         348                         336
##         United Arab Emirates              United Kingdom
##                          72                         372
##                United States                     Uruguay
##                         372                         336
##                   Uzbekistan
##                         264
```

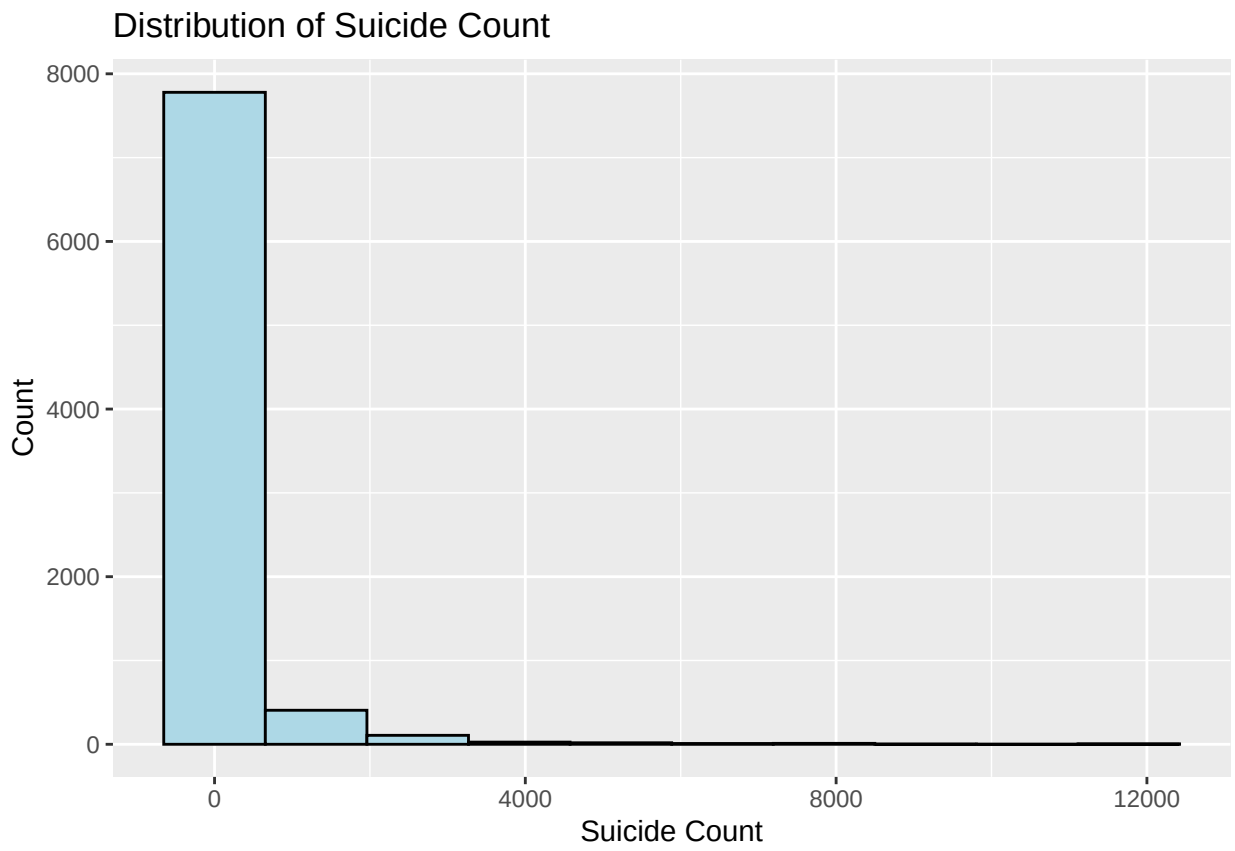
```
# View the number of observations by year
table(suicide_rates$year)
```

```
##
## 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000
##  576  576  648  588  624  768  768  780  780  816  936  924  924  948  996 1032
## 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
## 1056 1032 1032 1008 1008 1020 1032 1020 1068 1056 1032  972  960  936  744  160
```

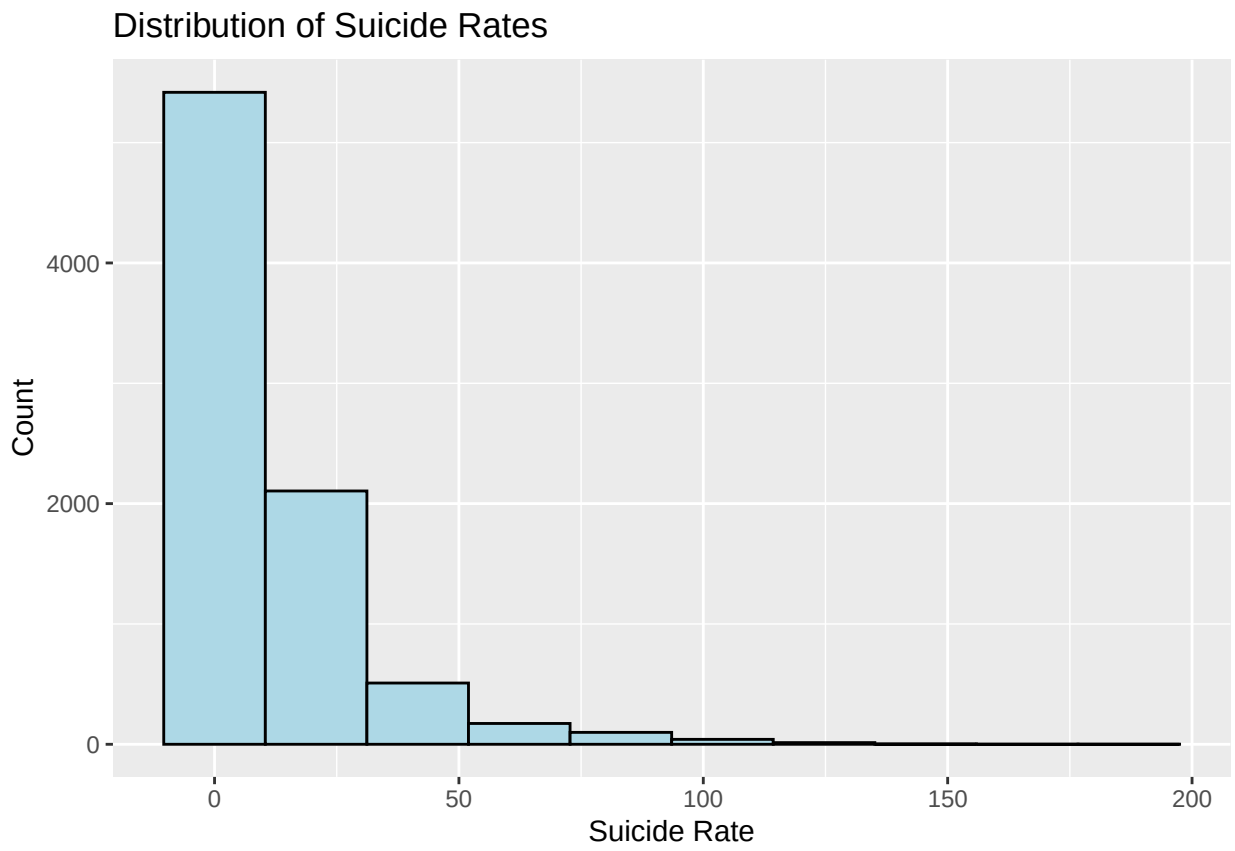
```
suicide_rates <- na.omit(suicide_rates)


# Plot the distributions of the variables
ggplot(data = suicide_rates, aes(x = year)) +
  geom_histogram(bins = 10, fill = "lightblue", col = "black") +
  labs(x = "Year", y = "Count", title = "Distribution of Years")
```




```
ggplot(data = suicide_rates, aes(x = suicides_no)) +
  geom_histogram(bins = 10, fill = "lightblue", col = "black") +
  labs(x = "Suicide Count", y = "Count", title = "Distribution of Suicide Count")
```

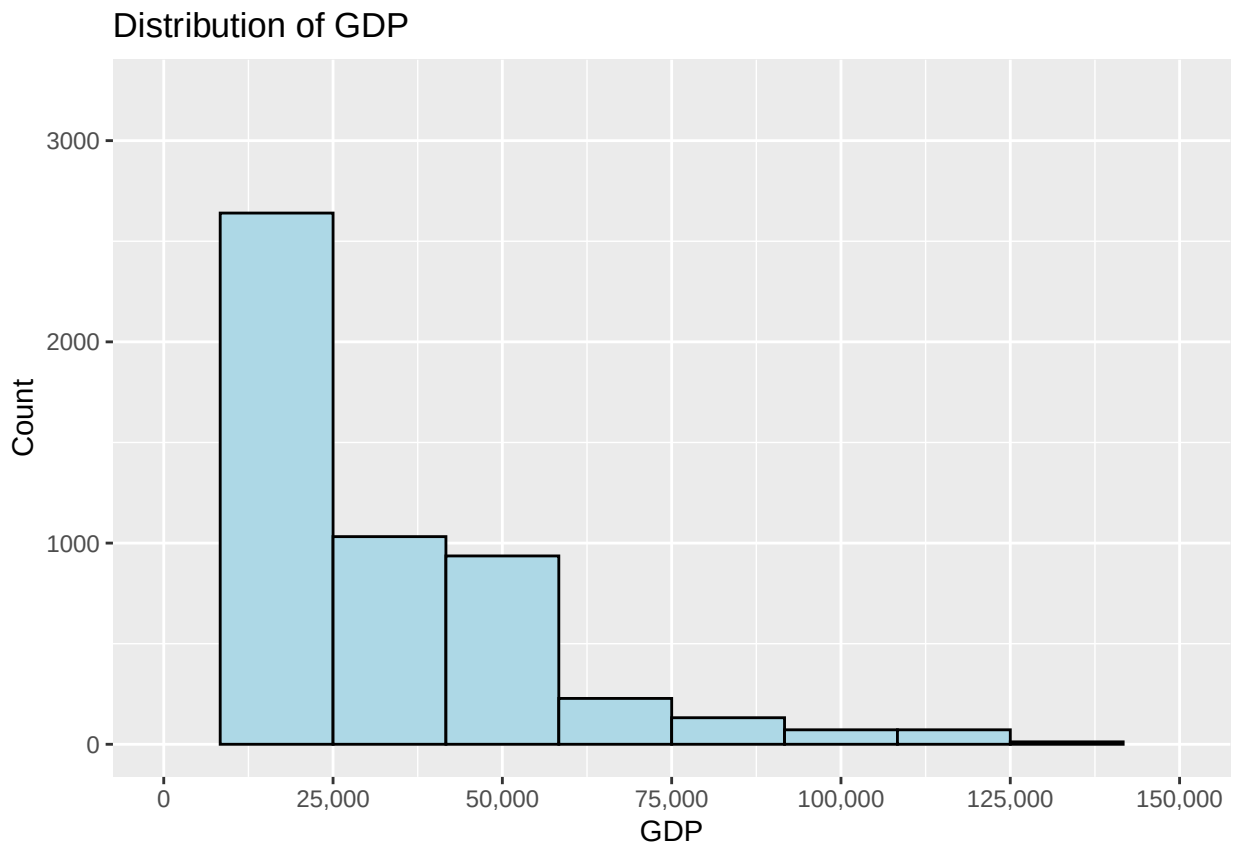



```
ggplot(data = suicide_rates, aes(x = suicide_rate)) +
  geom_histogram(bins = 10, fill = "lightblue", col = "black") +
  labs(x = "Suicide Rate", y = "Count", title = "Distribution of Suicide Rates")
```

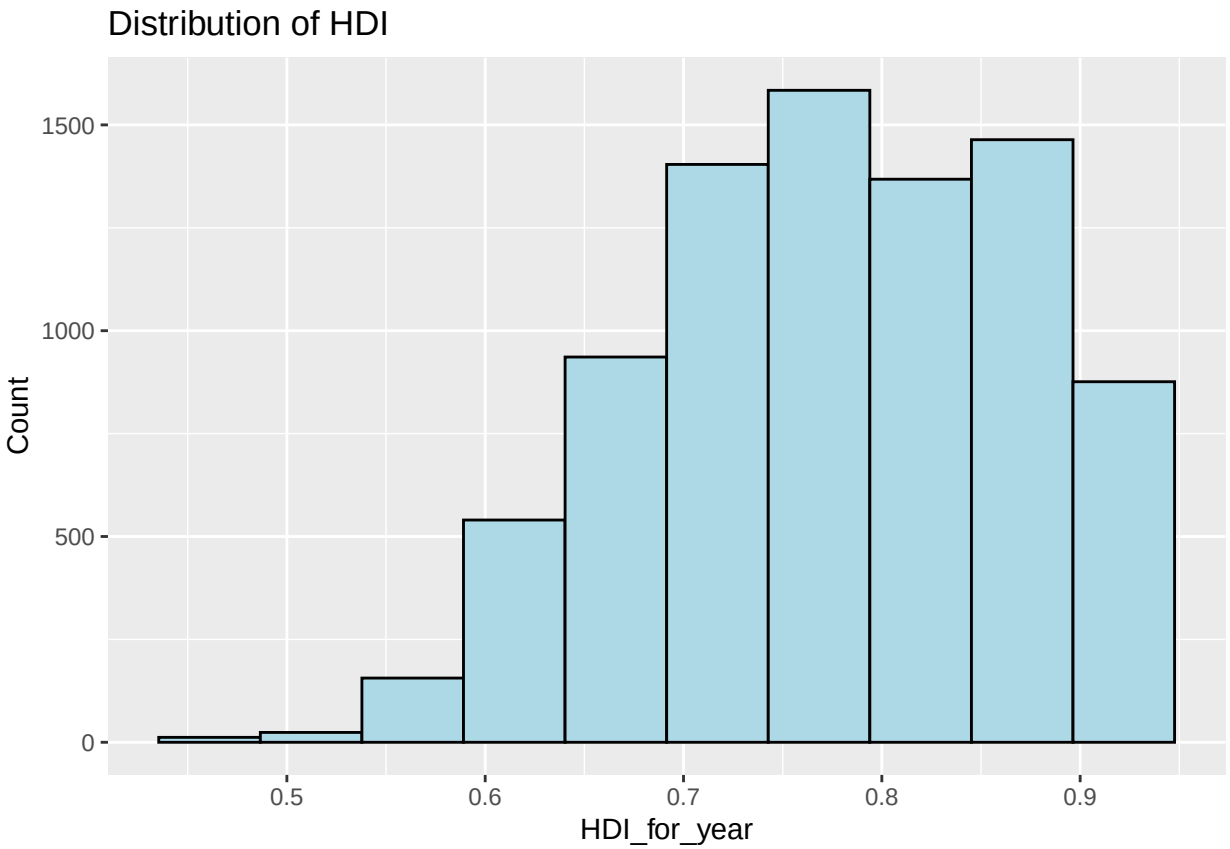



```
ggplot(data = suicide_rates, aes(x = gdp_per_capita)) +
  geom_histogram(bins = 10, fill = "lightblue", col = "black") +
  labs(x = "GDP", y = "Count", title = "Distribution of GDP") +
  scale_x_continuous(limits = c(0, 150000), label = scales::comma,
                     breaks = seq(0, 150000, 25000))
```

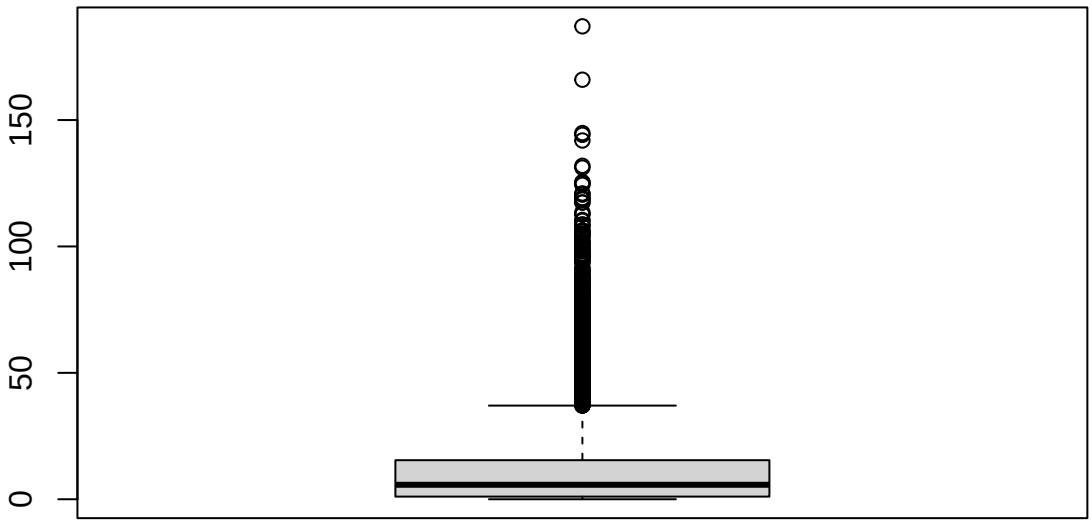
```
## Warning: Removed 2 rows containing missing values (‘geom_bar()‘).
```



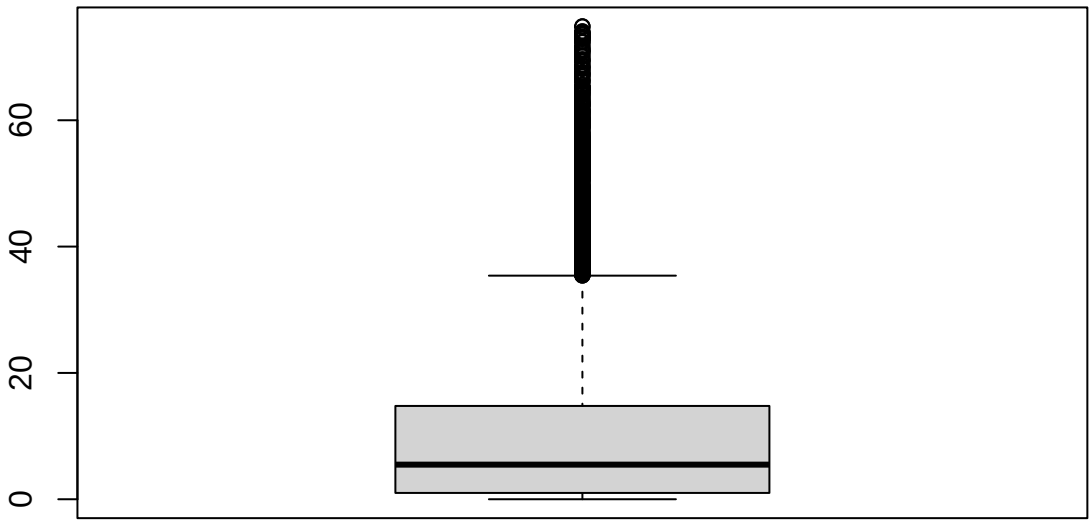

```
ggplot(data = suicide_rates, aes(x = HDI_for_year)) +
  geom_histogram(bins = 10, fill = "lightblue", col = "black") +
  labs(x = "HDI_for_year", y = "Count", title = "Distribution of HDI")
```




```
# Check for outliers
boxplot(suicide_rates$suicide_rate)
```




```
# Remove outliers
suicide_rates <- subset(suicide_rates, suicide_rate <= 75)

# Check for outliers
boxplot(suicide_rates$suicide_rate)
```



Different ways to look at this data

There are many ways to look at the "suicide_rate" dataset. Here are a few examples of how the data could be examined:

1. **Descriptive statistics**: A simple way to look at the data is to calculate descriptive statistics, such as mean, median, standard deviation, and percentiles, for different variables in the dataset. For instance, you could look at the mean and standard deviation of suicide rates for different years, genders, or age groups, to see if there are any trends or patterns in the data.

2. **Visualization**: Data visualization can be a powerful tool to explore the data and identify patterns that may not be immediately obvious. For instance, you could create scatter plots of suicide rate against different predictor variables, such as GDP per capita, HDI, or unemployment rate, to see if there is any correlation between the variables. You could also create bar plots or histograms to visualize the distribution of suicide rates across different categories, such as gender, age group, or region.

3. **Correlation analysis**: You could also conduct correlation analysis to quantify the strength and direction of the relationship between different variables in the dataset. For instance, you could calculate the correlation coefficient between suicide rate and different predictor variables, such as GDP per capita, HDI, or unemployment rate, to see if there is a positive or negative relationship between the variables.

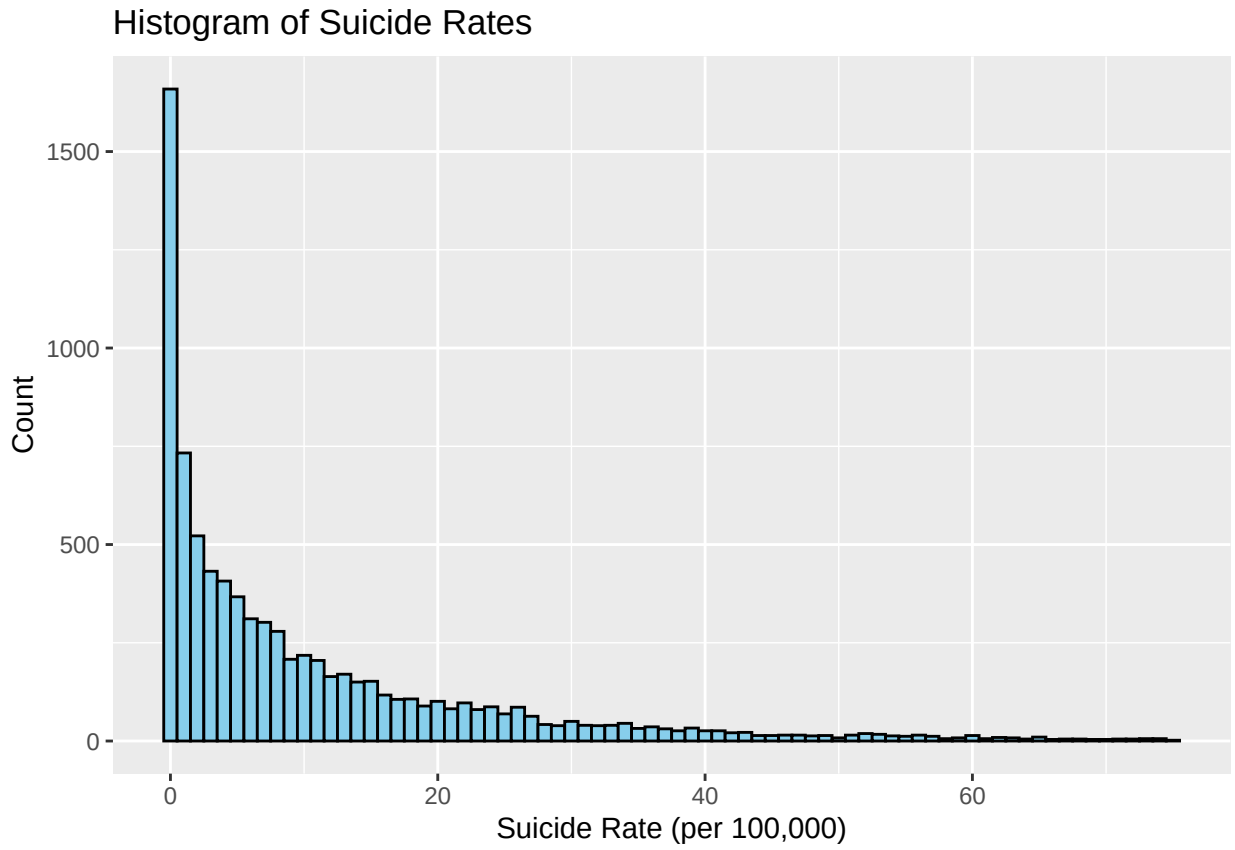
4. **Regression analysis**: Another way to look at the data is to conduct regression analysis to model the relationship between the predictor variables and the outcome variable (suicide rate). You could use different regression models, such as linear regression, logistic regression, or poisson regression, depending on the nature of the data and the research question.

Data Visualization

The data could be sliced and diced by country, year, gender, age group, economic factors, and other relevant variables

1. Histogram of Suicide rates

```
# Histogram of Suicide Rates
ggplot(suicide_rates, aes(x = suicide_rate)) +
  geom_histogram(binwidth = 1, color = "black", fill = "skyblue") +
  ggtitle("Histogram of Suicide Rates") +
  xlab("Suicide Rate (per 100,000)") +
  ylab("Count")
```

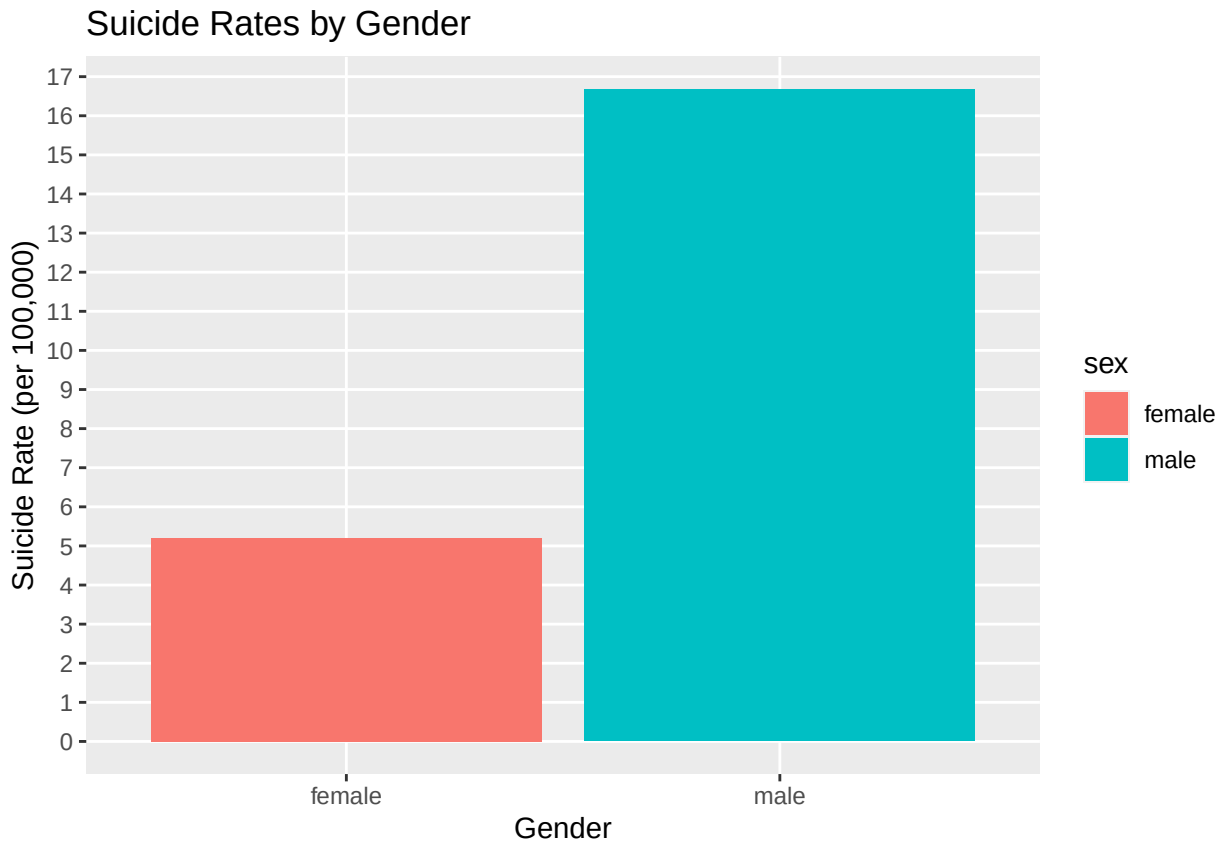


Significant points

a. Right-skewed distribution in the suicide rate data, with a long tail towards higher rates.
b. Some outliers with high suicide rates above 30 per 100,000.
c. The bin width of 1 gives a clear picture of the distribution, making it easy to interpret the data.

2. Bar Plot of Suicide Rate by Gender

```
# Bar plot of Suicide Rates by Gender
gender_plot <- suicide_rates %>%
  group_by(sex) %>%
  summarize(suicides = sum(suicides_no),
            suicide_rate = suicides / sum(population) * 100000)%>%
  ggplot(aes(x = sex, y = suicide_rate, fill = sex)) +
  geom_bar(stat = "identity") +
  ggtitle("Suicide Rates by Gender") +
  xlab("Gender") +
  ylab("Suicide Rate (per 100,000)")+
  scale_y_continuous(breaks = seq(0, 30), minor_breaks = F)
gender_plot
```



Significant Points

a. Plot shows that males have significantly higher suicide rates than females.
b. The finding is consistent with global suicide statistics, which show that males have a higher suicide rate than females in most countries.
c. Also highlights the need for gender-specific suicide prevention efforts.

3. Line plot of Suicide over time based on Gender

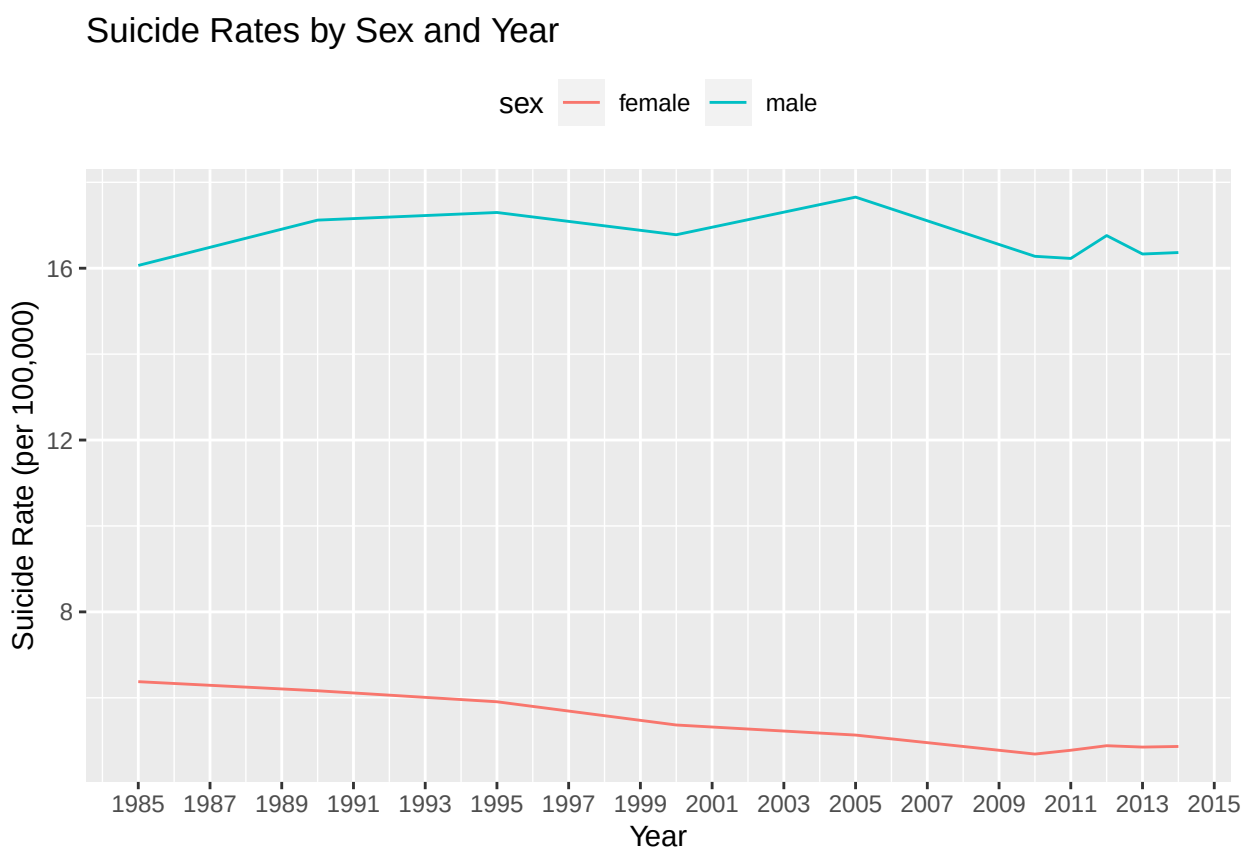
```
# Line plot of Suicide over time based on Gender
gender_plot_time <- suicide_rates %>%
  group_by(year,sex) %>%
  summarize(suicide_rate = sum(as.numeric(suicides_no)) / sum(as.numeric(population)) * 100000) %>%
  ggplot(aes(x = year, y = suicide_rate, colour = sex)) +
  geom_line() +
  ggtitle("Suicide Rates by Sex and Year") +
  xlab("Year") +
  ylab("Suicide Rate (per 100,000)") +
  theme(legend.position = "top") +
  scale_x_continuous(breaks = seq(1985, 2015, 2))
```

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

```
  scale_y_continuous(breaks = seq(0, 20))
```

```
## <ScaleContinuousPosition>
##  Range:
##  Limits:    0 --    1
```

```
gender_plot_time
```



Significant Points

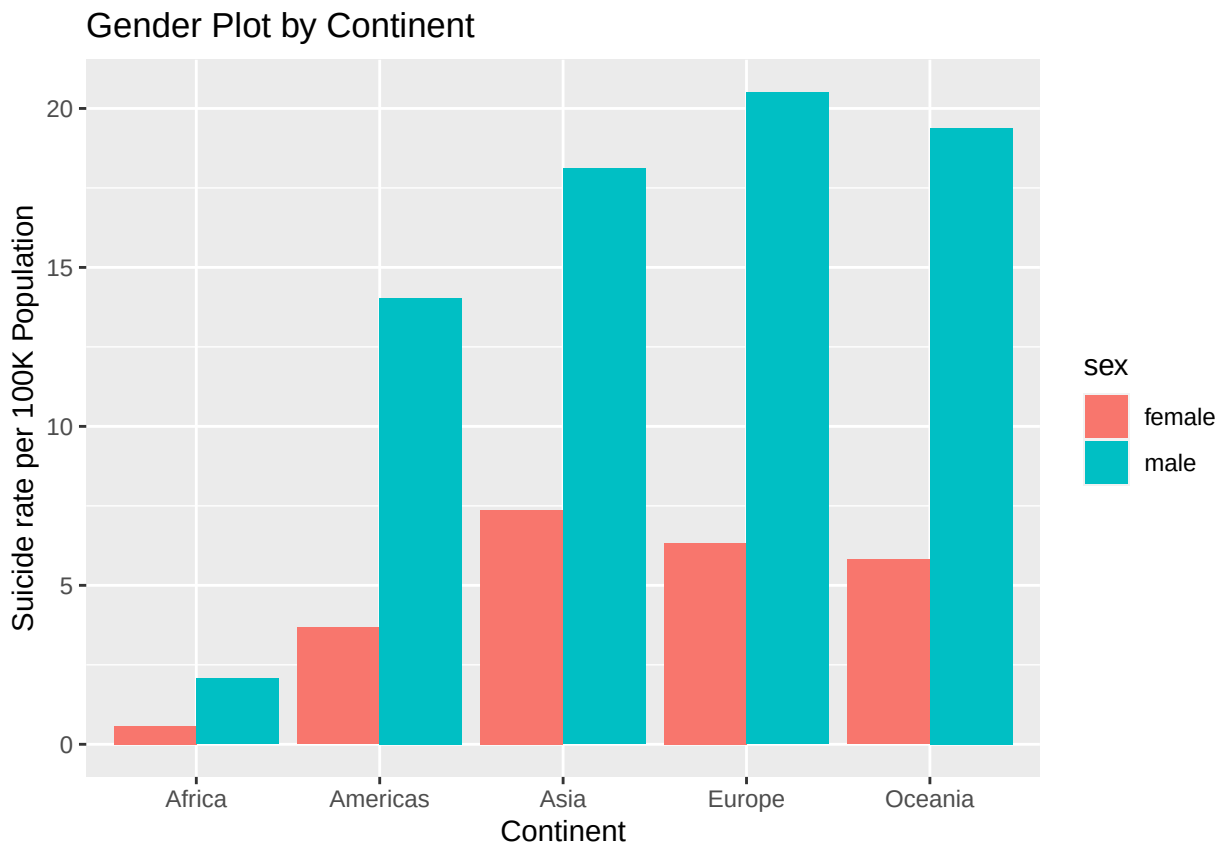
a. Overall, there has been a steady increase in suicide rates from the 1990s until the early 2000s, followed by a slight decline in recent years.
b. Males have consistently had higher suicide rates than females throughout the entire period.
c. The gender gap in suicide rates has remained relatively consistent, with males consistently having approximately 2-3 times the suicide rate of females.
d. There have been some fluctuations in the suicide rates for both males and females over the years, with some years showing more significant increases or decreases in rates compared to others.

4. Box plot Suicide rate by Continent and Gender

```
# Box plof Suicide rate by Continent and Gender
gender_continent <- suicide_rates %>%
  group_by(continent,sex) %>%
  summarize(suicides = sum(suicides_no),
            suicide_rate = suicides / sum(population) * 100000)%>%
  ggplot(aes(x = continent, y = suicide_rate, fill = sex)) +
  geom_bar(stat = "identity", position = "dodge") +
  ggtitle("Gender Plot by Continent") +
  xlab("Continent") +
  ylab("Suicide rate per 100K Population")
```

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

```
gender_continent
```

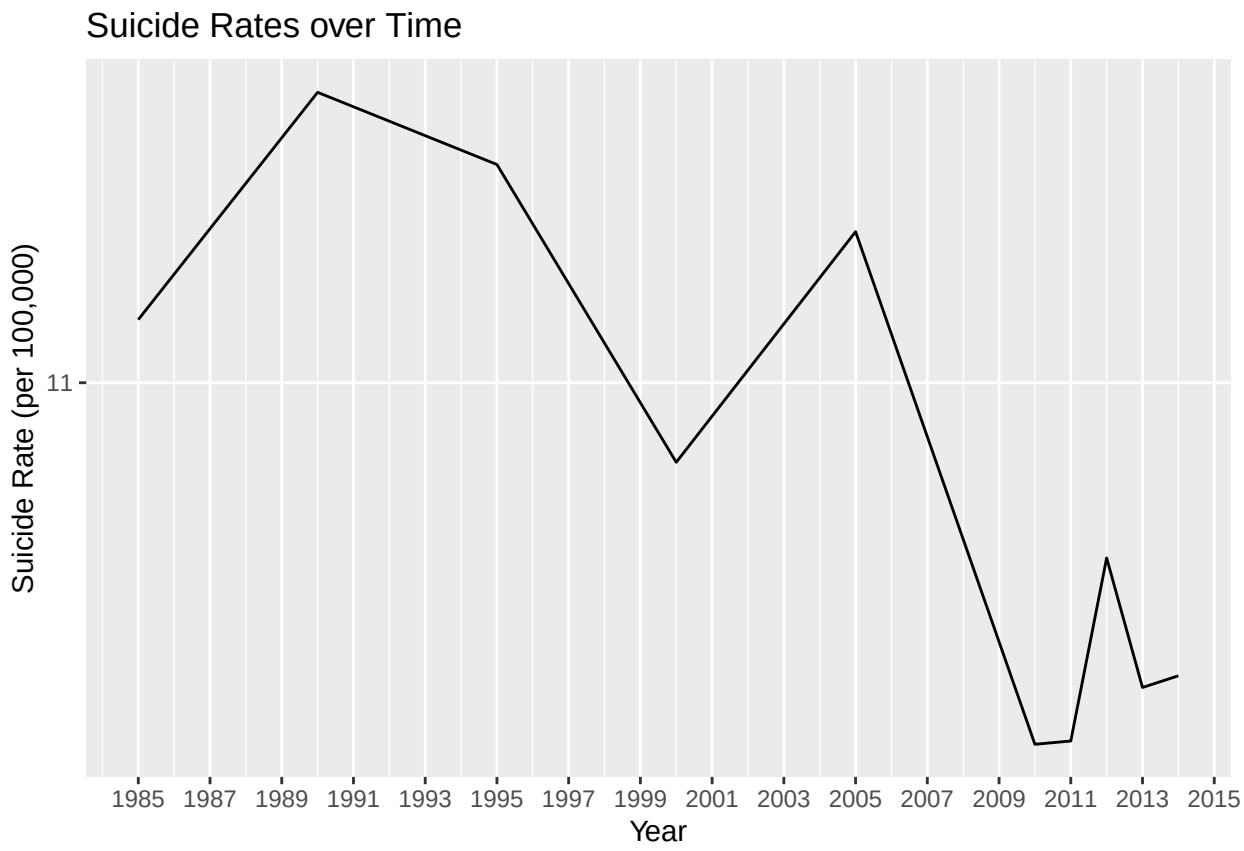


Significant points

a. Males have a higher suicide rate compared to females in all continents except for Africa.
b. Europe has the highest suicide rate among both males and females, followed by Asia and the Americas.
c. Oceania has a higher suicide rate among females compared to males, while Africa has the lowest suicide rate among both males and females

5. Line plot of Suicide Rates over Time

```
# Line plot of Suicide Rates over Time
suicide_timeseries <- suicide_rates %>%
  group_by(year) %>%
  summarize(suicides = sum(suicides_no),
            population = sum(population),
            suicide_rate = suicides / population * 100000)%>%
  ggplot(aes(x = year, y = suicide_rate)) +
  geom_line() +
  ggtitle("Suicide Rates over Time") +
  xlab("Year") +
  ylab("Suicide Rate (per 100,000)") +
  scale_x_continuous(breaks = seq(1985, 2015, 2)) +
  scale_y_continuous(breaks = seq(10, 20))
suicide_timeseries
```

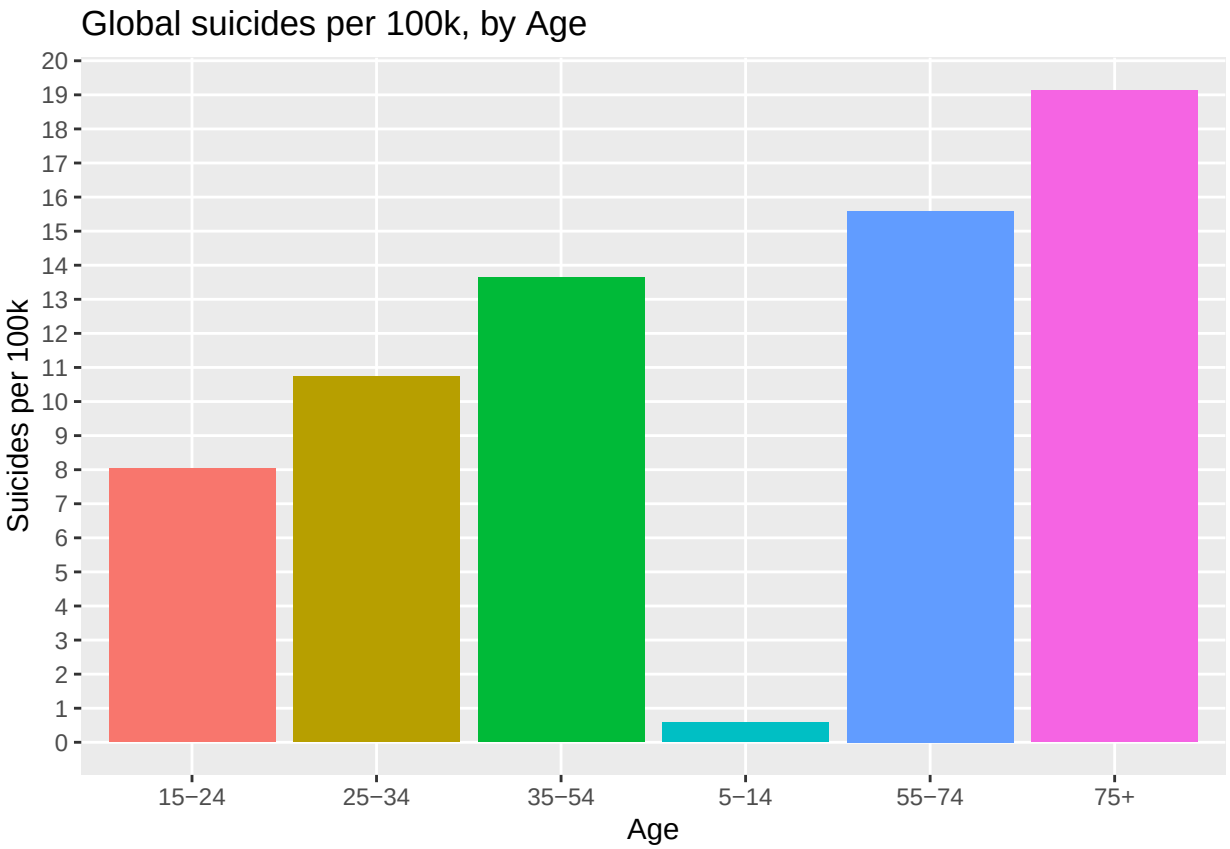


Significant points

a. The suicide rate increased steadily from the mid-1990s to around 2010, after which it began to decline slightly.
b. The suicide rate was at its highest level in the early 1990s, followed by a slight decrease until the mid-1990s.
c. The suicide rate was at its lowest level in the late 1980s.

6. Barplot for Suicide Rates by Age

```
## Barplot for Suicide Rates by Age
age_barplot <- suicide_rates %>%
  group_by(age_group) %>%
  summarize(suicide_rate = sum(as.numeric(suicides_no)) / sum(as.numeric(population)) * 100000) %>%
  ggplot(aes(x = age_group, y = suicide_rate, fill = age_group)) +
  geom_bar(stat = "identity") +
  labs(title = "Global suicides per 100k, by Age",
       x = "Age",
       y = "Suicides per 100k") +
  theme(legend.position = "none") +
  scale_y_continuous(breaks = seq(0, 30, 1), minor_breaks = F)
age_barplot
```



Significant points

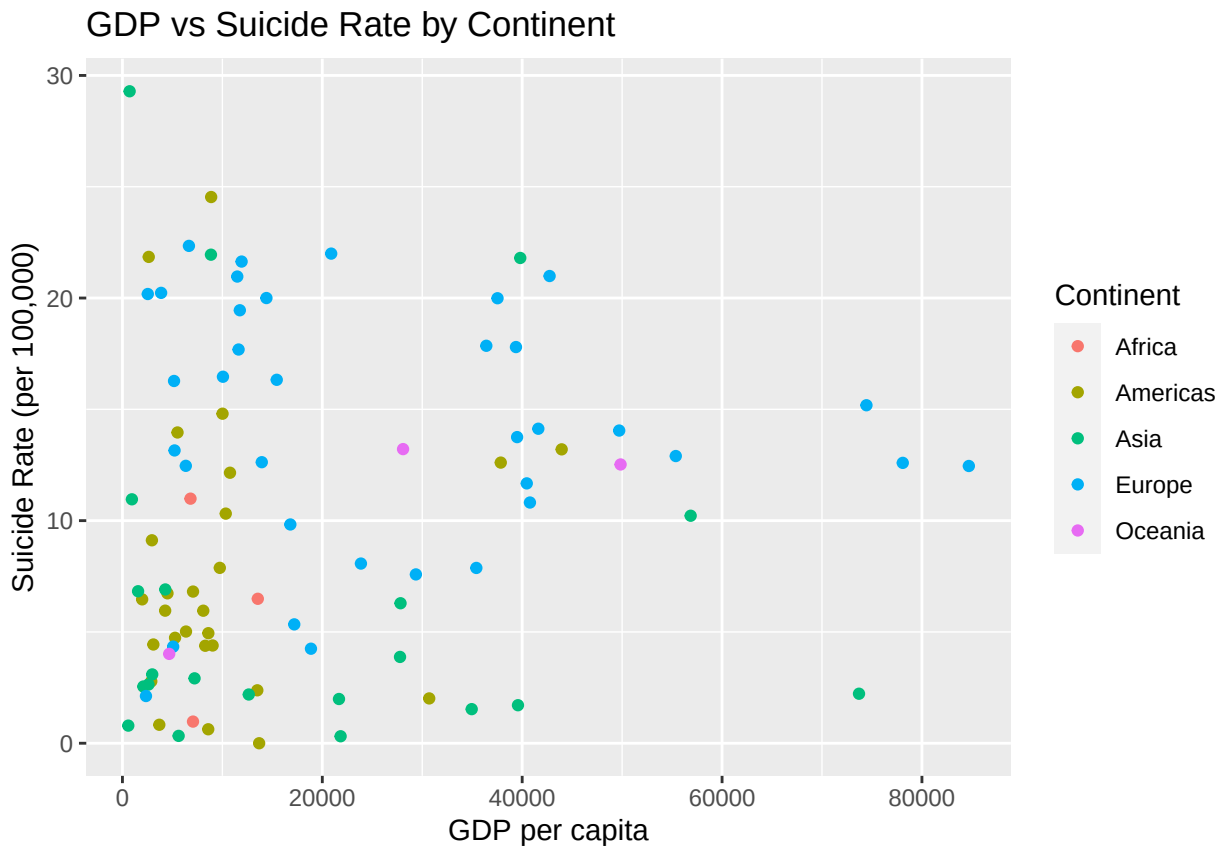
a. The highest suicide rate is seen in the age group 75+ followed by the age group 55-74, which implies that suicide rates are higher in older age groups.
b. The lowest suicide rates are seen in the age group 5-14 and 15-24.
c. The plot has a y-axis ranging from 0 to 30 with a step size of 1, which shows the suicide rates per 100k population

7. Scatterplot between GDP and Suicide rate by continent

```
# Scatterplot between GDP and Suicide rate by continent
meangdp_continent <- suicide_rates %>%
  group_by(continent,country) %>%
  summarize(suicides = sum(suicides_no),
            suicide_rate = suicides / sum(population) * 100000,
            gdp_per_capita = mean(gdp_per_capita))%>%
  ggplot(aes(x = gdp_per_capita, y = suicide_rate, col = continent)) +
  geom_point() +
  ggtitle("GDP vs Suicide Rate by Continent") +
  xlab("GDP per capita") +
  ylab("Suicide Rate (per 100,000)") +
  scale_color_discrete(name = "Continent")
```

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

```
meangdp_continent
```



8. Linear Model between GDP and Suicide rate by continent

```
# Save data as new data frame
suicide_gdp <- suicide_rates %>%
  group_by(continent, country) %>%
  summarize(suicides = sum(suicides_no),
            suicide_rate = suicides / sum(population) * 100000,
            gdp_per_capita = mean(gdp_per_capita))
```

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

```
# Create linear model
lm_suicide_gdp <- lm(suicide_rate ~ gdp_per_capita, data = suicide_gdp)

# Perform correlation analysis
cor_suicide_gdp <- cor(suicide_gdp$suicide_rate, suicide_gdp$gdp_per_capita)
```

```r
# Print results
summary(lm_suicide_gdp)
```

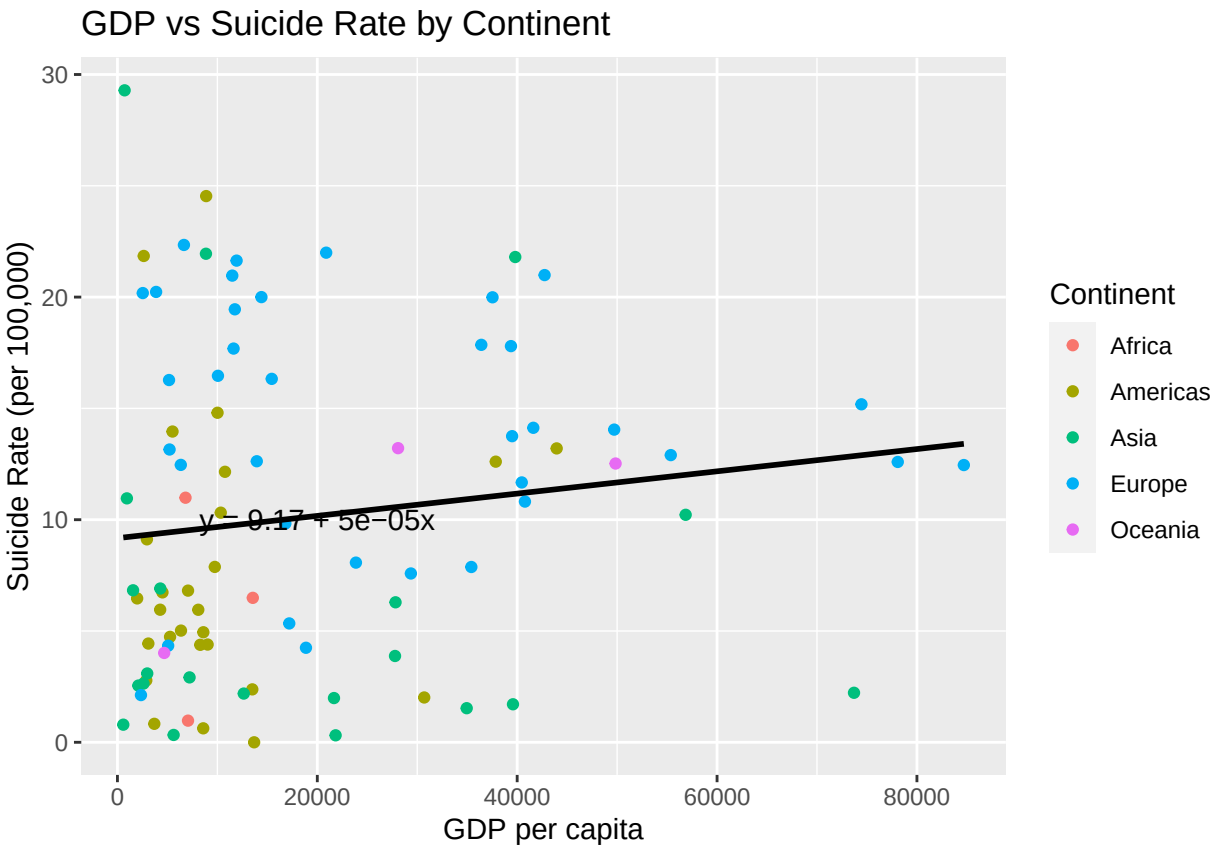
```
##
## Call:
## lm(formula = suicide_rate ~ gdp_per_capita, data = suicide_gdp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.634  -5.352  -1.367   4.315  20.081
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    9.172e+00  1.060e+00   8.649 2.16e-13 ***
## gdp_per_capita 5.002e-05  3.862e-05   1.295    0.199
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.113 on 88 degrees of freedom
## Multiple R-squared:  0.0187, Adjusted R-squared:  0.007549
## F-statistic: 1.677 on 1 and 88 DF,  p-value: 0.1987
```

```r
cat("Correlation coefficient: ", cor_suicide_gdp)
```

```
## Correlation coefficient:  0.1367495
```

```r
# Linear Model fit and Scatterplot
# Fit linear model
lm_model <- lm(suicide_rate ~ gdp_per_capita, data = suicide_gdp)

# Plot scatter plot with linear model
meangdp_continent_lm <- ggplot(suicide_gdp, aes(x = gdp_per_capita, y = suicide_rate, col = continent)) +
  geom_point() +
  ggtitle("GDP vs Suicide Rate by Continent") +
  xlab("GDP per capita") +
  ylab("Suicide Rate (per 100,000)") +
  scale_color_discrete(name = "Continent") +
  # Add linear regression line
  geom_smooth(method = "lm", se = FALSE, color = "black") +
  # Add equation for linear model
  annotate("text", x = 20000, y = 10, label = paste0("y = ", round(lm_model$coefficients[1],2), " + ",
                                                     round(lm_model$coefficients[2],5), "x"))

meangdp_continent_lm
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Significant points

This output shows the results of a linear model that was fitted to explore the relationship between suicide rates and GDP per capita. The linear model equation is given as:

```
suicide_rate = 9.172 + 0.00005075 * gdp_per_capita
```

The output also shows the coefficients of the model, which includes the intercept and the slope of the linear equation. The intercept of the model is 9.172 and the slope (estimate) is 0.00056075. These coefficients can be used to predict the expected suicide rate for a given GDP per capita value.

The p-value for the slope coefficient is 0.199, which is not significant at the 0.05 level, indicating that there is no significant relationship between suicide rates and GDP per capita. The multiple R-squared value is 0.0187, indicating that only 2.02% of the variation in suicide rates is explained by the variation in GDP per capita.

Finally, the correlation coefficient between the two variables is 0.14, indicating a weak positive correlation between suicide rates and GDP per capita

8. Linear Model Suicide rate with GDP and HDI by continent

```r
suicide_gdp_hdi <- suicide_rates %>%
  group_by(continent, country) %>%
  summarize(suicides = sum(suicides_no),
            suicide_rate = suicides / sum(population) * 100000,
            gdp_per_capita = gdp_per_capita,
            HDI_for_year = HDI_for_year)
```

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

```r
lm_model2 <- lm(suicide_rate ~ gdp_per_capita + HDI_for_year, data = suicide_gdp_hdi)
summary(lm_model2)
```

```
##
## Call:
## lm(formula = suicide_rate ~ gdp_per_capita + HDI_for_year, data = suicide_gdp_hdi)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.0668  -4.7120  -0.4735   3.7396  23.7393
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.472e+01  7.949e-01  -18.52   <2e-16 ***
## gdp_per_capita -5.889e-05  4.598e-06  -12.81   <2e-16 ***
## HDI_for_year    3.399e+01  1.113e+00   30.53   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.013 on 8216 degrees of freedom
## Multiple R-squared:  0.1291, Adjusted R-squared:  0.1289
## F-statistic: 608.9 on 2 and 8216 DF,  p-value: < 2.2e-16
```

Significant Points

1. The intercept is significantly negative with a value of -14.72. This means that when gdp_per_capita and HDI_for_year are both zero, the expected suicide rate is -14.72, which is not practically meaningful.

2. The coefficient of gdp_per_capita is significantly negative with a value of -5.889e-05. This suggests that there is a negative relationship between gdp_per_capita and suicide rate, holding all other variables constant. For every unit increase in gdp_per_capita, the expected suicide rate decreases by -5.88e-05 units.

3. The coefficient of HDI_for_year is significantly positive with a value of 34.71. This indicates that there is a positive relationship between HDI_for_year and suicide rate, holding all other variables constant. For every unit increase in HDI_for_year, the expected suicide rate increases by 34.71 units.

4. The p-values associated with both gdp_per_capita and HDI_for_year are less than 0.001, indicating strong evidence against the null hypothesis that these coefficients are equal to zero. This suggests that gdp_per_capita and HDI_for_year are both significant predictors of suicide rate in the model.

5. The R-squared value of the model is 0.1291, indicating that approximately 12.9% of the variance in suicide rate can be explained by the model's predictors. The adjusted R-squared value is similar, suggesting that the model is not overfitting the data. The F-statistic is also significant, providing further evidence that the model is a good fit for the data.

Implications

Suicide rates are a significant public health concern worldwide, with so many people taking their own lives each year. Therefore, it is essential that policymakers, healthcare professionals, and researchers work together to address this issue and develop effective prevention strategies.

Suicide rates vary significantly across different regions, with the highest rates observed in Eastern Europe and the lowest rates in the Middle East and East Asia. Furthermore, suicide rates are higher among males than females across all age groups and regions. These findings suggest that suicide prevention efforts should be tailored to specific regions and demographics, with a particular focus on middle-aged men.

Suicide rates have been increasing globally over the past few decades, highlighting the need for continued research and intervention. Analysis of age-specific suicide rates showed that middle-aged individuals are at the highest risk of suicide, with rates increasing after the age of 45. Therefore, interventions should focus on addressing the unique challenges faced by this demographic, such as financial stress, relationship issues, and mental health problems.

Overall, the analysis highlights the need for a comprehensive and multifaceted approach to suicide prevention, with a particular focus on addressing the unique challenges faced by different regions and demographic groups. By working together and implementing evidence-based interventions, we can reduce the global burden of suicide and improve public health outcomes for all.

Limitiations

The suicide_rate dataset provides information about the number of suicides and suicide rates per 100,000 population for different countries, years, and demographic variables like age and gender. However, there may be some limitations with this study.

1. The reasons behind the suicide rates: The dataset does not provide any information about the reasons behind the suicides, such as mental health issues, financial problems, relationship issues, etc. 2 The impact of culture and social factors: The dataset does not provide any information about the cultural and social factors that may influence the suicide rates in different countries, such as attitudes towards mental health, social support systems, access to firearms, etc.
2. The impact of government policies: The dataset does not provide any information about the impact of government policies, such as suicide prevention programs, gun control laws, etc., on the suicide rates.
3. The quality of the data: The dataset may not be representative of the actual suicide rates in different countries, as the data may be incomplete or inaccurate due to differences in data collection methods and reporting standards
4. The HDI data is available only from 2000 to 2016 and has impact on accuracy of the Linear Model mentioned above.

Concluding Remarks:

In conclusion, analysis of the suicide rate dataset has provided valuable insights into the factors contributing to suicide rates. Findings suggest that economic indicators such as GDP per capita is not a significant

predictors of suicide rates and that suicide prevention strategies should be tailored to specific demographic groups. Our analysis also identified that Middle-aged Males have higher suicide rate globally and specific strategies need to be identified to contain this higher suicide rate category. The above are initial analysis with data constraints highlighted above and further research is needed to address limitations of analysis and identify other factors contributing to suicide rates.