Assessment

Applied Statistics and Data Visualisation

MSc Data Science

Comparative analysis of mortality rate between South Asia, Sub-Saharan African and European countries.

Oluwatobiloba Vaughan

@00706111

01/12023

Table of Contents

Part One: Introduction and Data Preparation 3

Introduction 3

Background Research and Literature Review 5

Preparation and Exploration of Data Set 7

Data Preparation: 9

Exploratory Data Analysis (EDA) 15

Part Two: Statistical Analysis 24

Descriptive Statistical Analysis of South Asian Countries' Health Indicators 24

Descriptive Statistical Analysis of Sub-Saharan African Countries Health Indicators 26

Correlation Analysis of South Asia Dataset 30

Hypothesis Testing 37

Shapiro-Wilk Test for Normality Across Health Indicators 40

Multiple Linear Regression Analysis on Mortality Indicators Across Regions 47

Time Series Analysis 51

Part Three: Interactive Dashboard Design 56

Investigation of Data Workflows & Proposal for Design of Dashboard 56

Data Preprocessing and Preparation 57

Interactive Dashboard Overview 60

Discussion and Conclusion 61

Reference list 62

Part One: Introduction and Data Preparation

Introduction

The patterns of death reflects the underlying risk profiles in different regions based on age groups and sex. Many environmental risk factors such as road traffic accidents, infectious diseases, self-harm and interpersonal violence are the leading causes of death around the world. Death rates are higher for males from the conditions mentioned above along with collective violence and legal intervention, it is also important to note that maternal conditions has become an increasingly important cause of death for young women in lower-income countries. This research project will base on comprehensive statistical analysis focusing on the mortality rates in four selected countries (India, Afghanistan, Bangladesh, and Nepal) of less developed South Asian and four selected countries (Nigeria, Burundi, South-Sudan, and Cote d'ivore) of Sub-Saharan African countries, comparing them to the mortality rates in four selected countries (United kingdom, United States, Germany, and Canada) high-income European countries.

The dataset used for this research originates from the World Bank Development Indicators Databank, and the motivation behind this research stems from the recent maternal mortality campaign in Nigeria. The selection of indicators in this research was driven by a comprehensive consideration of various types of deaths and their profound impact on the overall mortality rates of each country. The chosen indicators were carefully selected to capture different dimensions of mortality, providing a nuanced understanding of the health landscape in each nation. By incorporating this diverse set of indicators, the research aims to comprehensively assess mortality patterns, identify key health determinants, and contribute to a holistic understanding of the health disparities and challenges faced by each country.

After meticulously choosing fourteen health indicators, the dataset underwent preprocessing procedures, including addressing missing values and renaming variables to enhance clarity. This analysis spans descriptive statistics, correlation analyses, and region-specific explorations for South Asian, Sub-Saharan African, and European countries. Visualization techniques, such as boxplots, histograms, and scatterplot matrices, offer insights into the distribution and relationships of key indicators across regions. We have also conducted correlation analyses within each region, examining the relationships between multiple continuous variables.

Additionally, normality tests and ANOVA were performed to assess the distribution of data and test for significant differences in maternal mortality rates between countries. As we proceed, we will delve into regression analyses to understand the factors influencing maternal mortality rates in each region and explore time series analysis for a dynamic understanding of maternal mortality trends. Finally, the indicators will be presented through an interactive dashboard ensuring accessibility and understanding among a broad audience. The following are the research objectives:

1. Conduct descriptive statistical analysis for the entire dataset.
2. Explore correlations between various health indicators.
3. Perform statistical tests, to assess the normality of the data distributions.
4. Conduct regional analyses for South Asian, Sub-Saharan African, and European countries separately.
5. Data Visualization to visualize the distribution, highlight trends over time in different regions, and explore relationships between variables.
6. Conduct test to assess if there are significant differences in maternal mortality rates among countries within each region.
7. Perform Hypothesis Testing to compare maternal mortality rates between each region.
8. Apply regression analysis models to understand the factors influencing maternal mortality rates in each region.
9. Apply regression analysis models to understand the factors Life Expectancy at birth in each region.
10. Utilize time series analysis, to examine trends in maternal mortality rates, life expectancy and Current health expenditure over time in each region.
11. Design and implement an interactive dashboard that tell a compelling story about Maternity mortality rate in each region.

Background Research and Literature Review

Background Research

In the ever-evolving landscape of mortality rate exploration, a profound understanding of statistical analysis becomes imperative to extract meaningful insights from complex datasets. This comprehensive study embarks on the comparison of mortality rates in undeveloped South Asian and sub-Saharan African countries to high-income European countries. Employing a multifaceted approach, including correlation analysis, regression analysis, time series analysis, and interactive dashboard design, this research aims to unravel the intricate patterns surrounding mortality and causes of death. Mortality rates serve as vital indicators reflecting the health landscape of nations. Understanding the factors influencing these rates is crucial for informed decision-making and public health interventions. This study integrates diverse statistical methodologies and interactive visualization tools to delve into the complexities of mortality patterns across distinct regions.

Statistical Methodologies:

Regression Analysis: Regression analysis emerges as a powerful tool for dissecting relationships between variables. In our context, it enables an in-depth exploration of the myriad factors influencing mortality rates in selected South Asian, sub-Saharan African, and high-income European countries. The theoretical underpinnings draw from works by Kutner et al. (2005), providing insights into the assumptions and practical applications of regression analysis. Additionally, within the health-focused context, logistic regression, as discussed by Hosmer and Lemeshow (2000), becomes particularly relevant in exploring maternal mortality factors.

Time Series Analysis: Time series analysis, specifically employing ARIMA models, provides a lens into temporal trends in mortality rates. Building on the foundational works of Box and Jenkins (1976), this methodology offers valuable insights into forecasting methods. The literature by Hyndman and Athanasopoulos (2018) serves as a guide, aligning with my objective of understanding temporal trends in maternal mortality rates.

Interactive Dashboard Design: Creating an interactive dashboard involves synthesizing principles from data visualization and user experience. Foundational works by Few (2009) and Tufte (2001) underscore the importance of clear visual communication. Heer and Shneiderman's (2012) contributions to information visualization guide the development of an engaging dashboard. This user-friendly platform ensures effective communication of complex mortality rate data to a diverse audience.

Literature Review

Maternal Mortality and Health Indicators: Studies by Hogan et al. (2010) and Kassebaum et al. (2016) offer valuable insights into global health trends, contributing significantly to understanding factors influencing maternal health outcomes. These works align seamlessly with the objectives of our regression analysis, providing a robust foundation for exploring maternal mortality across selected regions and indicators.

Time Trends in Health Outcomes: The extensive literature on forecasting and trends in health outcomes, including Murray et al. (2012) and the World Health Organization's reports on maternal health indicators, enriches our understanding of temporal patterns in mortality rates. These contributions form a critical part of our literature review, informing the temporal dimension of our research.

Interactive Dashboards in Public Health: Within the realm of interactive dashboards, Few's (2013) advocacy for simplicity and effectiveness in design and Wong's (2019) exploration of dashboards in public health contexts shape our approach. These insights guide the development of an interactive dashboard tailored to effectively communicate trends and insights regarding mortality rates in the specified regions and indicators.

This comprehensive exploration integrates a variety of statistical methodologies and draws on a rich literature landscape to inform our understanding of mortality rates. By synthesizing key insights, we aim to shed light on the intricate patterns surrounding mortality and causes of death, ultimately contributing to the broader discourse on public health and decision-making.

Preparation and Exploration of Data Set

In the subsequent sections, we delve into the specific steps taken for each process, providing a detailed account of the data preparation, outlier detection, and exploratory data analysis. This structured approach lays the groundwork for a robust analysis of mortality rates across different regions and indicators.

NO	INDICATORS	DEFINITION
1	**Life Expectancy (LE)**	Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.
2	**Traffic Mortality (TM)**	Mortality caused by road traffic injury is estimated road traffic fatal injury deaths per 100,000 population.
3	**CVD Cancer Diabetes**	Mortality from CVD, cancer, diabetes or CRD is the percent of 30-year-old-people who would die before their 70th birthday from any of cardiovascular disease, cancer, diabetes, or chronic respiratory disease, assuming that s/he would experience current mortality rates at every age and s/he would not die from any other cause of death (e.g., injuries or HIV/AIDS).
4	**Unintentional Poisoning Mortality (PM)**	Mortality rate attributed to unintentional poisonings is the number of deaths from unintentional poisonings in a year per 100,000 population. Unintentional poisoning can be caused by household chemicals, pesticides, kerosene, carbon monoxide and medicines, or can be the result of environmental contamination or occupational chemical exposure.
5	**Adult Female Mortality (FM)**	Adult mortality rate, female, is the probability of dying between the ages of 15 and 60--that is, the probability of a 15-year-old female dying before reaching age 60, if subject to age-specific mortality rates of the specified year between those ages.
6	**Adult Male Mortality (MM)**	Adult mortality rate, male, is the probability of dying between the ages of 15 and 60--that is, the probability of a 15-year-old male dying before reaching age 60, if subject to age-specific mortality rates of the specified year between those ages.
7	**Infant Deaths (ID)**	Number of infants dying before reaching one year of age.
8	**Suicide Mortality (SM)**	Suicide mortality rate is the number of suicide deaths in a year per 100,000 population. Crude suicide rate (not age-adjusted).
9	**Crude Death Rate (CRD)**	Crude death rate indicates the number of deaths occurring during the year, per 1,000 population estimated at midyear. Subtracting the crude death rate from the crude birth rate provides the rate of natural increase, which is equal to the rate of population change in the absence of migration.
10	**Maternal Mortality Ratio (MMR)**	Maternal mortality ratio is the number of women who die from pregnancy-related causes while pregnant or within 42 days of pregnancy termination per 100,000 live births. The data are estimated with a regression model using information on the proportion of maternal deaths among non-AIDS deaths in women ages 15-49, fertility, birth attendants, and GDP measured using purchasing power parities (PPPs).
11	**Gov Health Expenditure**	Share of current health expenditures funded from domestic public

	(GHE)	sources for health. Domestic public sources include domestic revenue as internal transfers and grants, transfers, subsidies to voluntary health insurance beneficiaries, non-profit institutions serving households (NPISH) or enterprise financing schemes as well as compulsory prepayment and social health insurance contributions. They do not include external resources spent by governments on health.
12	**Health Expenditure GDP (HE)**	Level of current health expenditure expressed as a percentage of GDP. Estimates of current health expenditures include healthcare goods and services consumed during each year. This indicator does not include capital health expenditures such as buildings, machinery, IT and stocks of vaccines for emergency or outbreaks.
13	**Anemia Prevalence Pregnant Women (AP)**	Prevalence of anemia, pregnant women, is the percentage of pregnant women whose hemoglobin level is less than 110 grams per liter at sea level.
14	**Maternal Death Risk (MDR)**	Life time risk of maternal death is the probability that a 15-year-old female will die eventually from a maternal cause assuming that current levels of fertility and mortality (including maternal mortality) do not change in the future, taking into account competing causes of death.

Region	**Countries**
South Asia	• Afghanistan • Bangladesh • India • Nepal
Sub-Sahara Africa	• Burundi • Cote d’ivore • Nigeria • South Sudan
Europe	• Canada • Germany • United States • United Kingdom

Data Preparation:

The initial steps in this data analysis involved loading the dataset and conducting a preliminary exploration:

```
Mortality_Rate <- read.csv("Mortality Rate_Data.csv", header= TRUE)
```

```
43:1    (Top Level) ÷                                                    R Script ÷
Console   Terminal ×   Background Jobs ×
R 4.3.2 · ~/ASDV/ASDV ASSESSMENT/
> setwd("~/ASDV/ASDV ASSESSMENT")
> Mortality_Rate <- read.csv("Mortality Rate_Data.csv", header= TRUE)
>
```

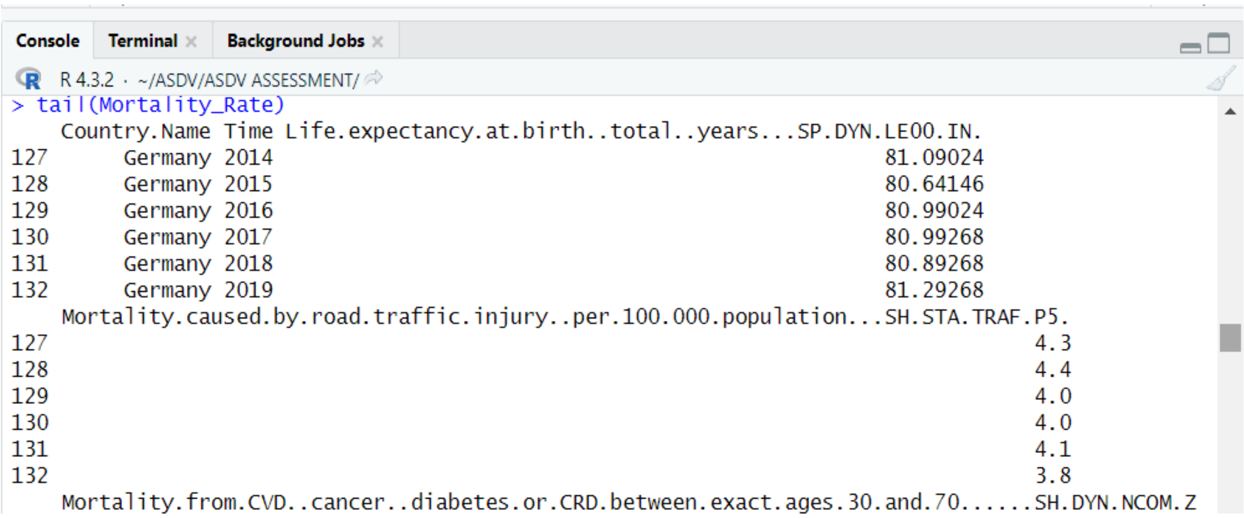
Loading the Data: The path C:\Users\tobby\OneDrive\Documents\ASDV\ASDV ASSESSMENT was set as a working directory, before the dataset, named "Mortality Rate_Data.csv," was imported into R using the read.csv function.

```
names(Mortality_Rate)

head(Mortality_Rate)

tail(Mortality_Rate)

str(Mortality_Rate)
```

```
55:1    (Top Level) ÷                                                    R Script ÷
Console   Terminal ×   Background Jobs ×
R 4.3.2 · ~/ASDV/ASDV ASSESSMENT/
```

```
Console   Terminal ×   Background Jobs ×
R 4.3.2 · ~/ASDV/ASDV ASSESSMENT/
> names(Mortality_Rate)
 [1] "Country.Name"
 [2] "Time"
 [3] "Life.expectancy.at.birth..total..years...SP.DYN.LE00.IN."
 [4] "Mortality.caused.by.road.traffic.injury..per.100.000.population...SH.STA.TRAF.P5."
 [5] "Mortality.from.CVD..cancer..diabetes.or.CRD.between.exact.ages.30.and.70......SH.DYN.NCOM.
ZS."
 [6] "Mortality.rate.attributed.to.unintentional.poisoning..per.100.000.population...SH.STA.POI
S.P5."
 [7] "Mortality.rate..adult..female..per.1.000.female.adults...SP.DYN.AMRT.FE."
 [8] "Mortality.rate..adult..male..per.1.000.male.adults...SP.DYN.AMRT.MA."
 [9] "Number.of.infant.deaths..SH.DTH.IMRT."
[10] "Suicide.mortality.rate..per.100.000.population...SH.STA.SUIC.P5."
[11] "Death.rate..crude..per.1.000.people...SP.DYN.CDRT.IN."
[12] "Maternal.mortality.ratio..modeled.estimate..per.100.000.live.births...SH.STA.MMRT."
[13] "Domestic.general.government.health.expenditure....of.current.health.expenditure...SH.XPD.G
```

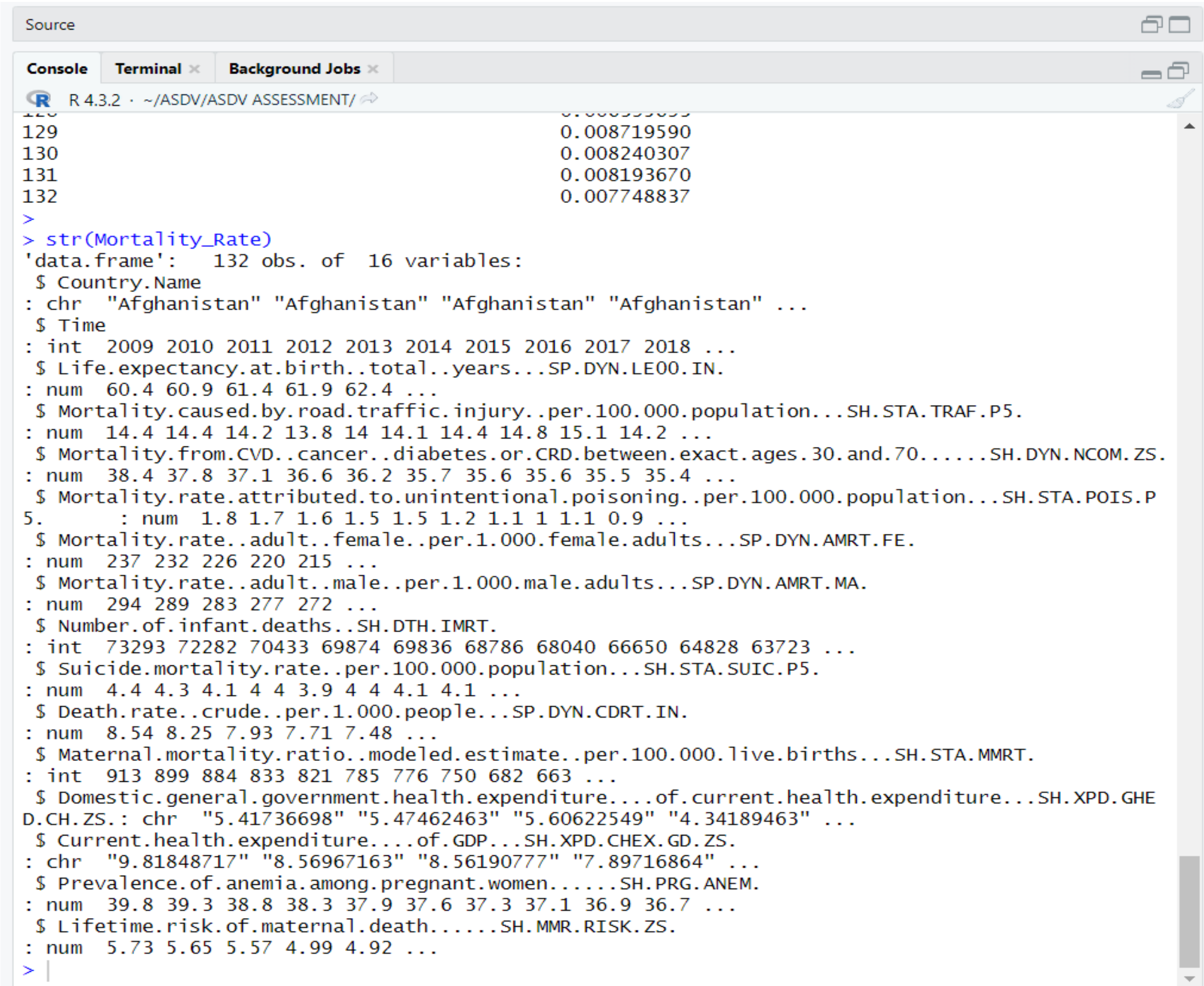
```
Console   Terminal ×   Background Jobs ×
R 4.3.2 · ~/ASDV/ASDV ASSESSMENT/
>
> head(Mortality_Rate)
  Country.Name Time Life.expectancy.at.birth..total..years...SP.DYN.LE00.IN.
1  Afghanistan 2009                                                   60.364
2  Afghanistan 2010                                                   60.851
3  Afghanistan 2011                                                   61.419
4  Afghanistan 2012                                                   61.923
5  Afghanistan 2013                                                   62.417
6  Afghanistan 2014                                                   62.545
  Mortality.caused.by.road.traffic.injury..per.100.000.population...SH.STA.TRAF.P5.
1                                                                              14.4
2                                                                              14.4
3                                                                              14.2
4                                                                              13.8
5                                                                              14.0
6                                                                              14.1
```



Console Terminal Background Jobs

R 4.3.2 · ~/ASDV/ASDV ASSESSMENT/

```
> tail(Mortality_Rate)
    Country.Name Time Life.expectancy.at.birth..total..years...SP.DYN.LE00.IN.
127      Germany 2014                                                 81.09024
128      Germany 2015                                                 80.64146
129      Germany 2016                                                 80.99024
130      Germany 2017                                                 80.99268
131      Germany 2018                                                 80.89268
132      Germany 2019                                                 81.29268
    Mortality.caused.by.road.traffic.injury..per.100.000.population...SH.STA.TRAF.P5.
127                                                                             4.3
128                                                                             4.4
129                                                                             4.0
130                                                                             4.0
131                                                                             4.1
132                                                                             3.8
    Mortality.from.CVD..cancer..diabetes.or.CRD.between.exact.ages.30.and.70......SH.DYN.NCOM.Z
```



Source

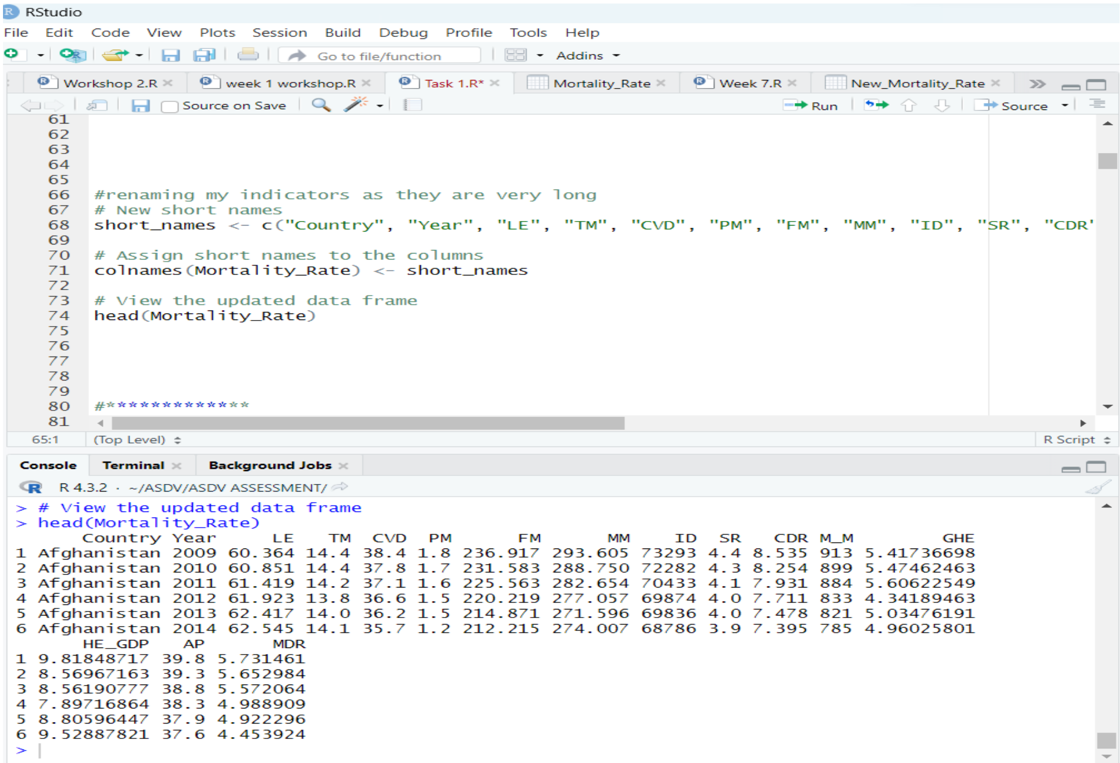
Console Terminal Background Jobs

R 4.3.2 · ~/ASDV/ASDV ASSESSMENT/

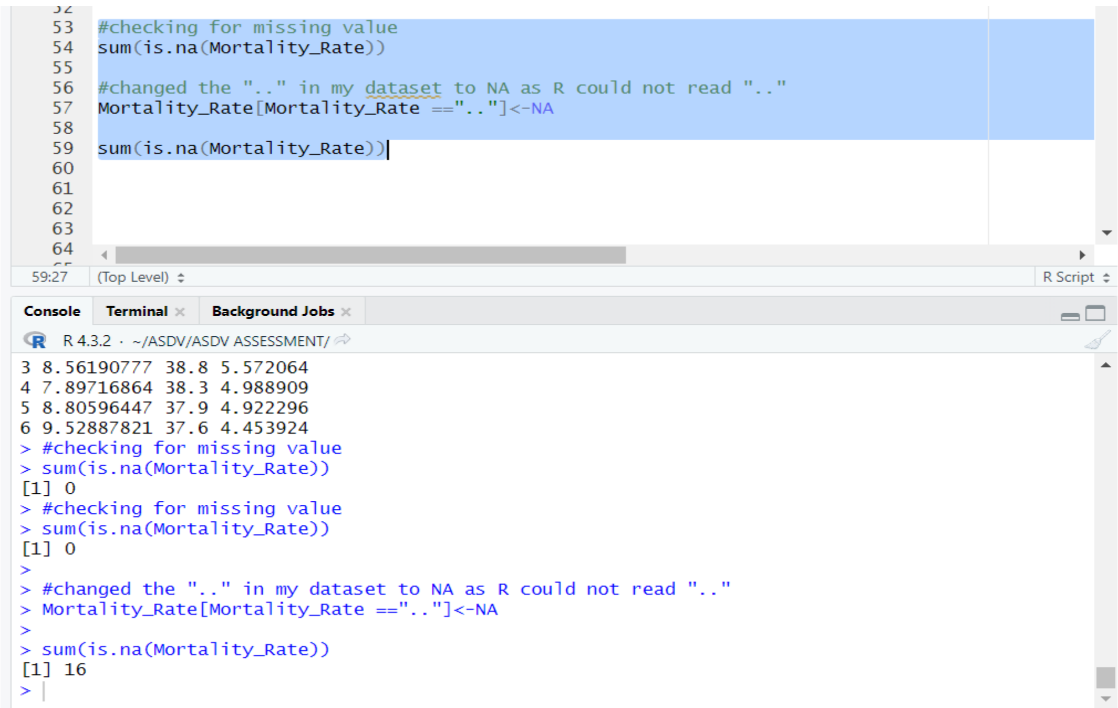
```
129                                            0.008719590
130                                            0.008240307
131                                            0.008193670
132                                            0.007748837
>
> str(Mortality_Rate)
'data.frame':	132 obs. of  16 variables:
 $ Country.Name
: chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
 $ Time
: int  2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 ...
 $ Life.expectancy.at.birth..total..years...SP.DYN.LE00.IN.
: num  60.4 60.9 61.4 61.9 62.4 ...
 $ Mortality.caused.by.road.traffic.injury..per.100.000.population...SH.STA.TRAF.P5.
: num  14.4 14.4 14.2 13.8 14 14.1 14.4 14.8 15.1 14.2 ...
 $ Mortality.from.CVD..cancer..diabetes.or.CRD.between.exact.ages.30.and.70......SH.DYN.NCOM.ZS.
: num  38.4 37.8 37.1 36.6 36.2 35.7 35.6 35.6 35.5 35.4 ...
 $ Mortality.rate.attributed.to.unintentional.poisoning..per.100.000.population...SH.STA.POIS.P
5.        : num  1.8 1.7 1.6 1.5 1.5 1.2 1.1 1 1.1 0.9 ...
 $ Mortality.rate..adult..female..per.1.000.female.adults...SP.DYN.AMRT.FE.
: num  237 232 226 220 215 ...
 $ Mortality.rate..adult..male..per.1.000.male.adults...SP.DYN.AMRT.MA.
: num  294 289 283 277 272 ...
 $ Number.of.infant.deaths..SH.DTH.IMRT.
: int  73293 72282 70433 69874 69836 68786 68040 66650 64828 63723 ...
 $ Suicide.mortality.rate..per.100.000.population...SH.STA.SUIC.P5.
: num  4.4 4.3 4.1 4 4 3.9 4 4 4.1 4.1 ...
 $ Death.rate..crude..per.1.000.people...SP.DYN.CDRT.IN.
: num  8.54 8.25 7.93 7.71 7.48 ...
 $ Maternal.mortality.ratio..modeled.estimate..per.100.000.live.births...SH.STA.MMRT.
: int  913 899 884 833 821 785 776 750 682 663 ...
 $ Domestic.general.government.health.expenditure....of.current.health.expenditure...SH.XPD.GHE
D.CH.ZS.: chr  "5.41736698" "5.47462463" "5.60622549" "4.34189463" ...
 $ Current.health.expenditure....of.GDP...SH.XPD.CHEX.GD.ZS.
: chr  "9.81848717" "8.56967163" "8.56190777" "7.89716864" ...
 $ Prevalence.of.anemia.among.pregnant.women......SH.PRG.ANEM.
: num  39.8 39.3 38.8 38.3 37.9 37.6 37.3 37.1 36.9 36.7 ...
 $ Lifetime.risk.of.maternal.death......SH.MMR.RISK.ZS.
: num  5.73 5.65 5.57 4.99 4.92 ...
>
```

In the diagram above it can be seen that the country variable is the only variable that is categorical the rest of the variables are numerical.

Exploring the Data: The column names, the first and last few rows, and the structure of the enitre dataset was examined to gain an initial understanding



Renaming Columns: as seen above the variable names are too long making the console presentation look untidy. To enhance clarity, I assigned shorter names to the columns for ease of reference.



stances

where the symbol ".." represented missing data by replacing them with NA.

```r
#Handling Numeric Columns
numeric_cols <- c("GHE", "HE_GDP", "AP", "MDR")
Mortality_Rate[numeric_cols] <- lapply(Mortality_Rate[numeric_cols], as.numeric)

# Define a function to impute missing values with the mean
impute_missing <- function(x) {}

# Apply the impute_missing function to each numeric column
Mortality_Rate[numeric_cols] <- lapply(Mortality_Rate[numeric_cols], impute_missing)

sum(is.na(Mortality_Rate))
```

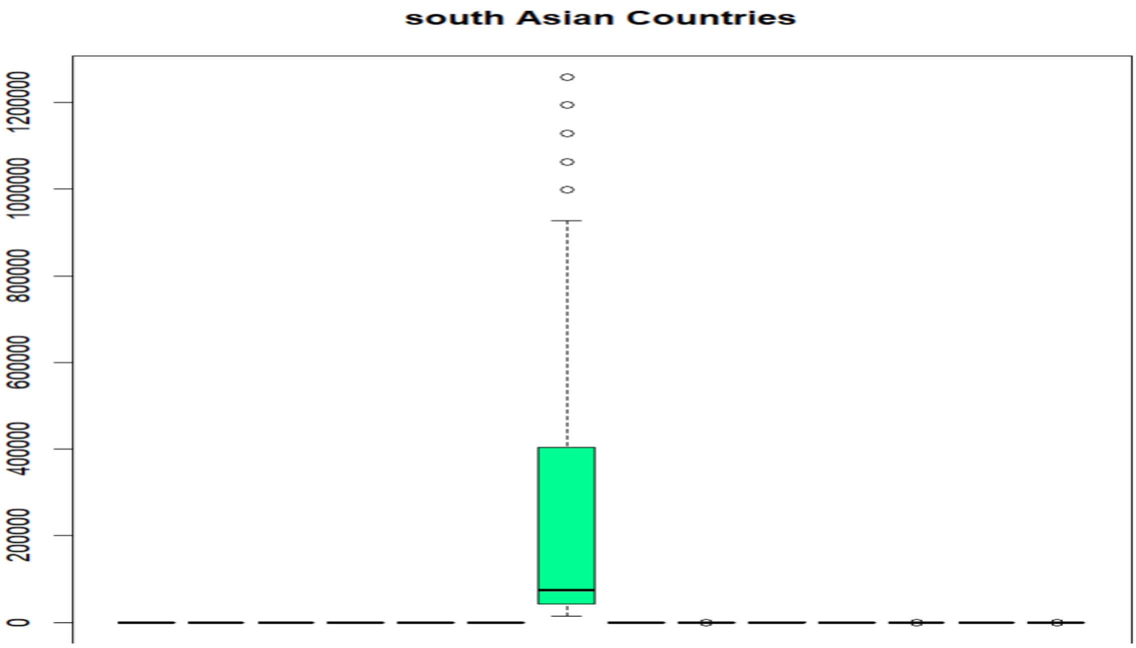
Handling Numeric Columns and Imputing Missing Values: Numeric columns were appropriately converted, and missing values in these columns were imputed using the mean.

```r
#Satistical Analysis based on region

# Extract south Asian countries data set
southAsia_Countries  <- grouped_Mortality_data %>%
  dplyr::filter(Country %in% c('Afghanistan', 'Bangladesh', 'Nepal', 'India'))
southAsia_Countries


# Extract sub Saharan African countries data set
SubSahara_Countries  <- grouped_Mortality_data %>%
  dplyr::filter(Country %in% c('Nigeria', 'Burundi', 'South Sudan', 'Cote d\'Ivoire'))
SubSahara_Countries


# Extract high income countries in Europe data set
European_Countries  <- grouped_Mortality_data %>%
  dplyr::filter(Country %in% c('United States', 'United Kingdom', 'Canada', 'Germany'))
European_Countries
```

Outlier Detection:

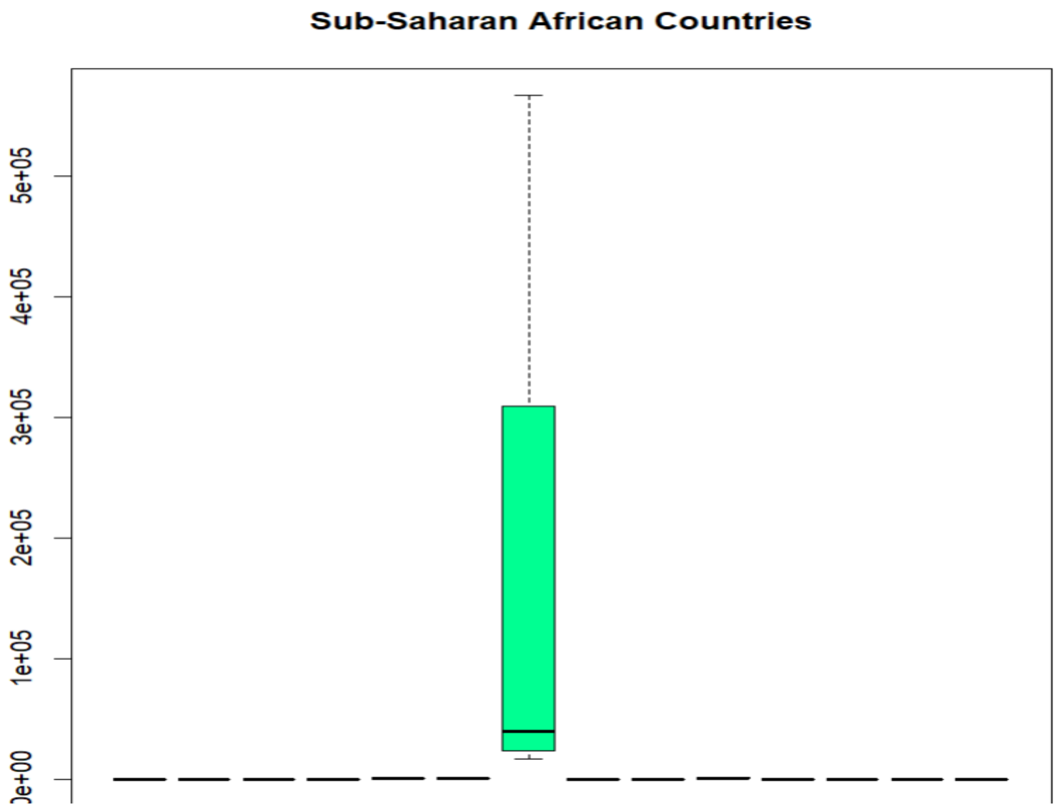
The boxplot utilizes a color spectrum to differentiate between the fourteen South Asian countries, providing a visual summary of the central tendency, spread, and potential outliers within each variable. This graphical representation aids in the comparative analysis of the selected health indicators across the specified region, facilitating the identification of variations and trends in mortality-related metrics. The main title clarifies the context of the analysis, ensuring a clear and concise interpretation of the presented boxplot.

```
#Checking for Outliers
# Boxplot for selected variables in southAsia_Countries
boxplot(southAsia_Countries[, c(3:16)], col = rainbow(14), main = "south Asian Countries")
```



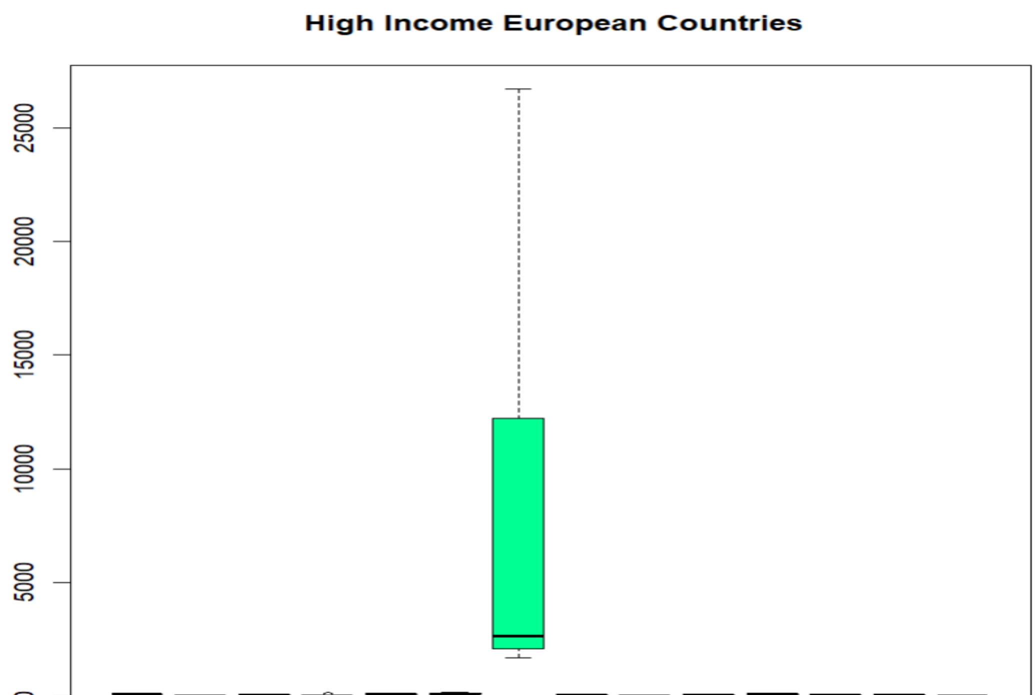
In the south Asian undeveloped countries selected for this research project it can be seen in the diagram above that the dataset split into SouthAsia_Countries dataset contains three outliers which are zero making them an outlier in the boxplot. Because the information in this dataset are live values the outliers will not be removed.

```r

# Boxplot for selected variables in SubSahara_Countries
boxplot(SubSahara_Countries[, c(3:16)], col = rainbow(14), main = "Boxplot of Selected Variables")
```



In the Sub-Sahara African undeveloped countries selected for this research project it can be seen in the diagram above that the dataset split into SubSahara_Countries dataset does not contains any outlier.

```r

# Boxplot for selected variables in European_Countries
boxplot(European_Countries[, c(3:16)], col = rainbow(14), main = " High Income European Countries")
```

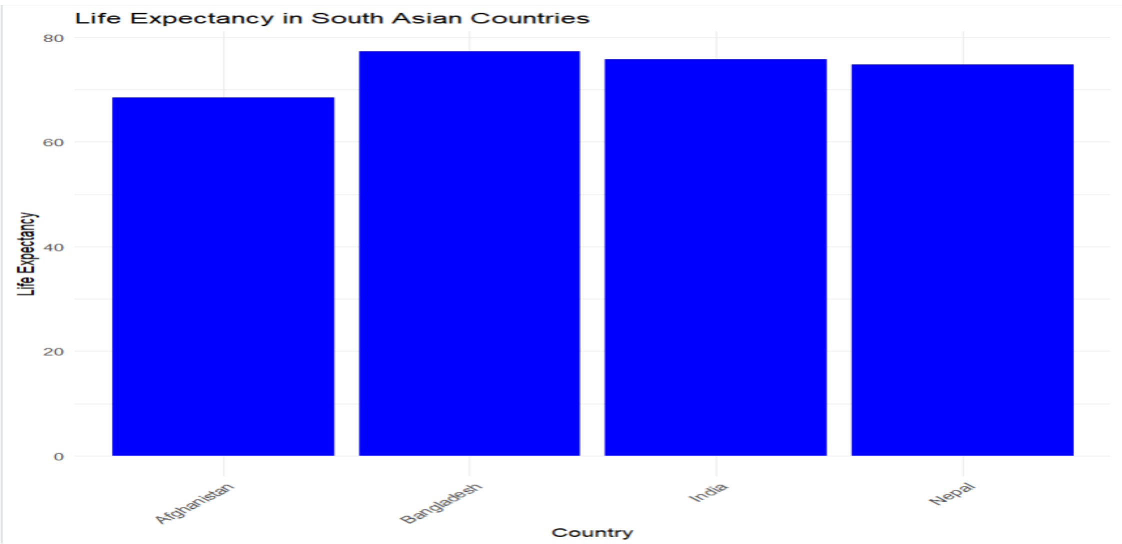


In the high income European countries selected for this research project it can be seen in the diagram above that the dataset split into European_Countries dataset which contains one outlier that is zero making it an outlier in the boxplot. Because the information in this dataset are live values the outliers will not be removed.

Exploratory Data Analysis (EDA)

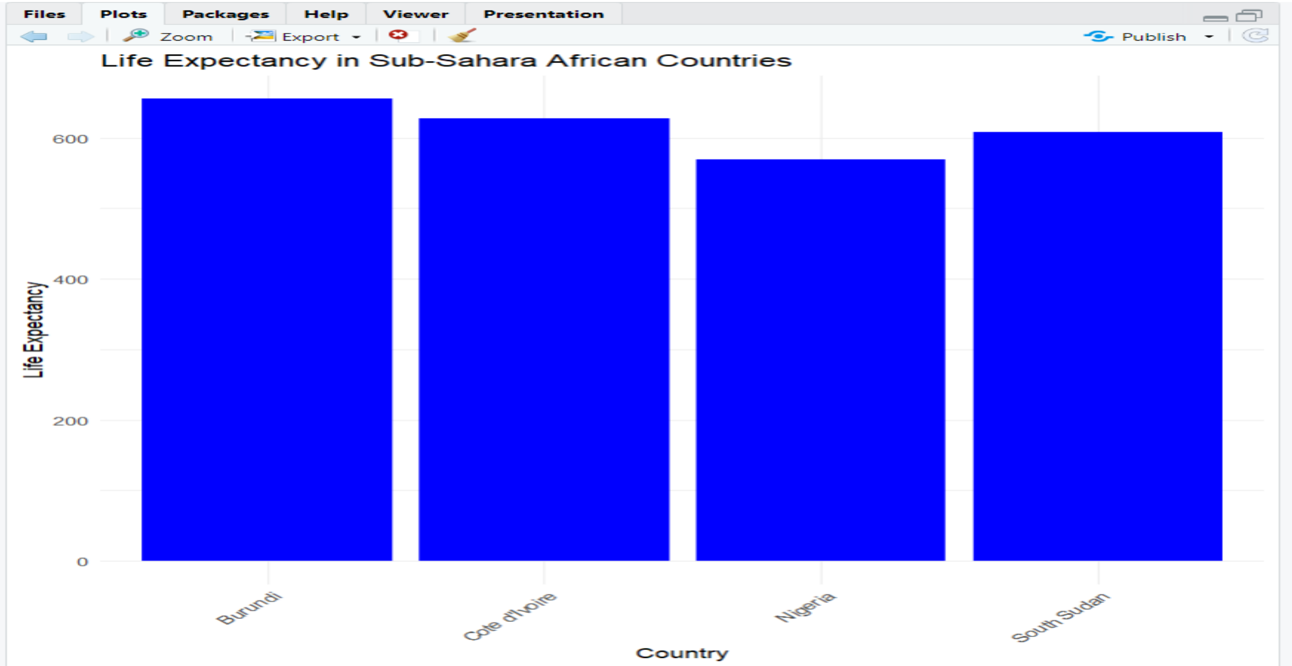
Histograms: This code visualizes and compares life expectancy in different countries using a bar chart. The data is reshaped and plotted using the dplyr and ggplot2 packages. The bar chart is customized with a title, axis labels, and a minimal theme. This visualization helps researchers understand life expectancy trends across different countries.

```
#EDA
# Life Expectancy in southAsia_Countries
library(ggplot2)

# Melt the data for ggplot
melted_data <- southAsia_Countries %>%
  dplyr::select(Country, LE, Year)  # Select the columns needed for plotting

# Plot the bar chart with a single color (blue)
ggplot(melted_data, aes(x = Country, y = LE /10 )) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(title = "Life Expectancy in South Asian Countries",
       x = "Country", y = "Life Expectancy") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



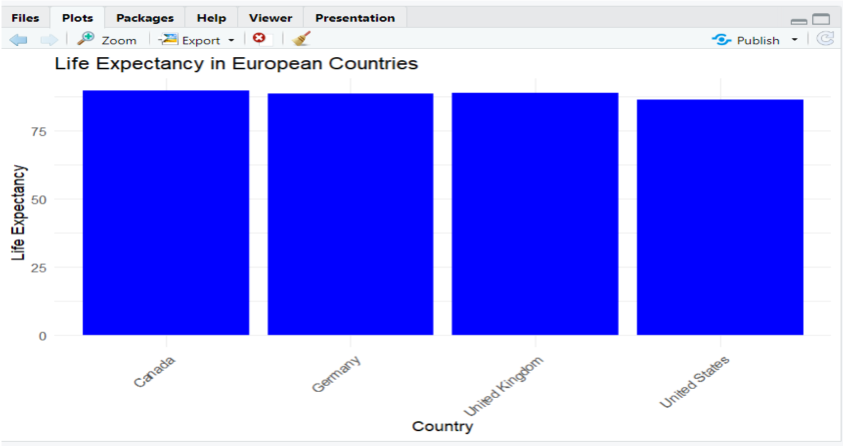
Bangladesh has the highest overall life expectancy among these countries, followed by India and Nepal. Afghanistan has the lowest life expectancy. The chart clearly shows that Bangladesh has the highest life expectancy, followed by India and Nepal, while Afghanistan has the lowest.

```
# Life Expectancy in SubSahara_Countries
# Melt the data for ggplot
melted_data <- SubSahara_Countries %>%
  dplyr::select(Country, LE, Year)  # Select the columns needed for plotting

# Plot the bar chart with a single color (blue)
ggplot(melted_data, aes(x = Country, y = LE)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(title = "Life Expectancy in Sub-Sahara African Countries",
       x = "Country", y = "Life Expectancy") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



As indicated by the presented diagram, Burundi emerges with the highest cumulative life expectancy within this cohort, succeeded by Cote d'Ivoire and South Sudan. In stark contrast, Nigeria registers the lowest overall life expectancy in this particular group. This graphical representation vividly portrays the variations in life expectancy across these countries, underscoring Burundi's prominent position and Nigeria's comparatively lower standing within this group.

```
# Life Expectancy in  High income European Countries
# Melt the data for ggplot
melted_data <- European_Countries %>%
  dplyr::select(Country, LE, Year)  # Select the columns needed for plotting

# Plot the bar chart with a single color (blue)
ggplot(melted_data, aes(x = Country, y = LE/ 10 )) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(title = "Life Expectancy in European Countries",
       x = "Country", y = "Life Expectancy") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

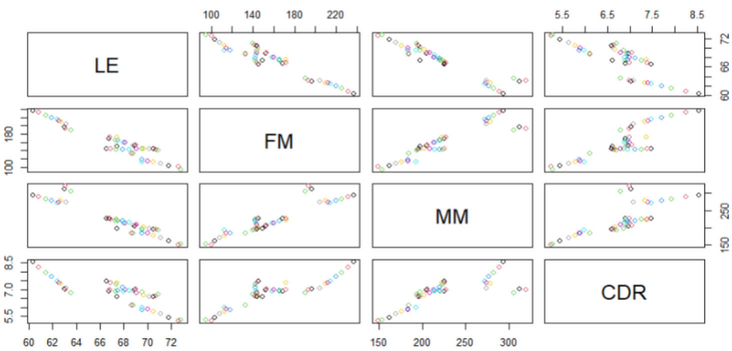


As depicted in the above diagram, Canada leads with the highest overall life expectancy in this category, trailed by Germany and the United Kingdom, which share an equal standing. Conversely, the United States ranks at the bottom in total life expectancy within this group. Notably, this particular group boasts the highest life expectancy across all groups, surpassing even the selected countries in South Asia which is notably higher than the sub-Sahara African group.

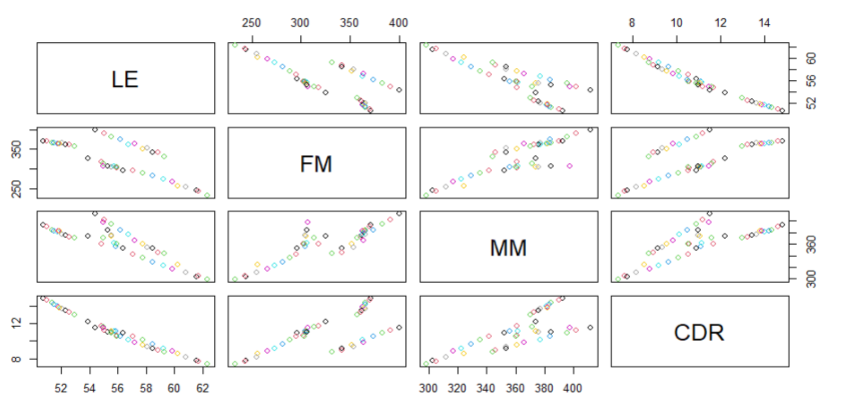
Scatterplot Matrix:

The purpose of the code provided below is to visually explore the relationships and patterns among the selected health-related variables over different years in different region countries. The scatterplot matrix allows researchers to quickly identify potential correlations, trends, or outliers, providing a visual overview of the dataset's multivariate distribution. This type of exploratory visualization is crucial for generating hypotheses and guiding more in-depth analyses in the subsequent stages of the research project.

```
# Scatterplot matrix for southAsia_Countries variables
pairs(southAsia_Countries[, c(3,7,8,11)], col = southAsia_Countries$Year)
```

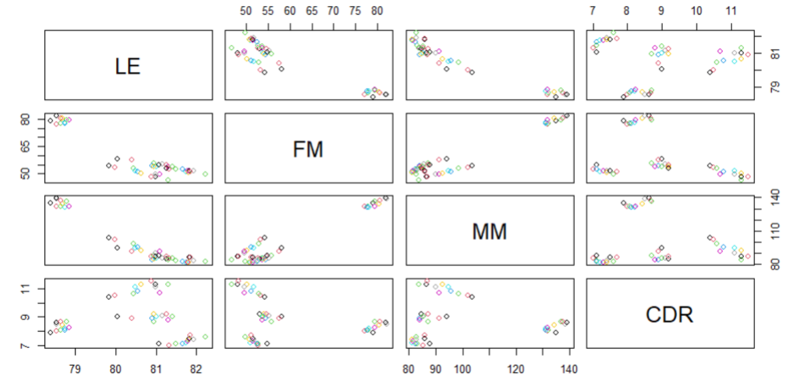


```
# Scatterplot matrix for SubSahara_Countries variables
pairs(SubSahara_Countries[, c(3,7,8,11)], col = SubSahara_Countries$Year)
```



This visualization examines the connection between Life Expectancy, Female Mortality, Male Mortality, and Crude Death Rate variables within the SubSahara_Countries dataset. Upon reviewing the diagram, it indicates the existence of a linear relationship among these variables.

```
pairs(European_Countries[, c(4,5,6,10)], col = European_Countries$Year)
```

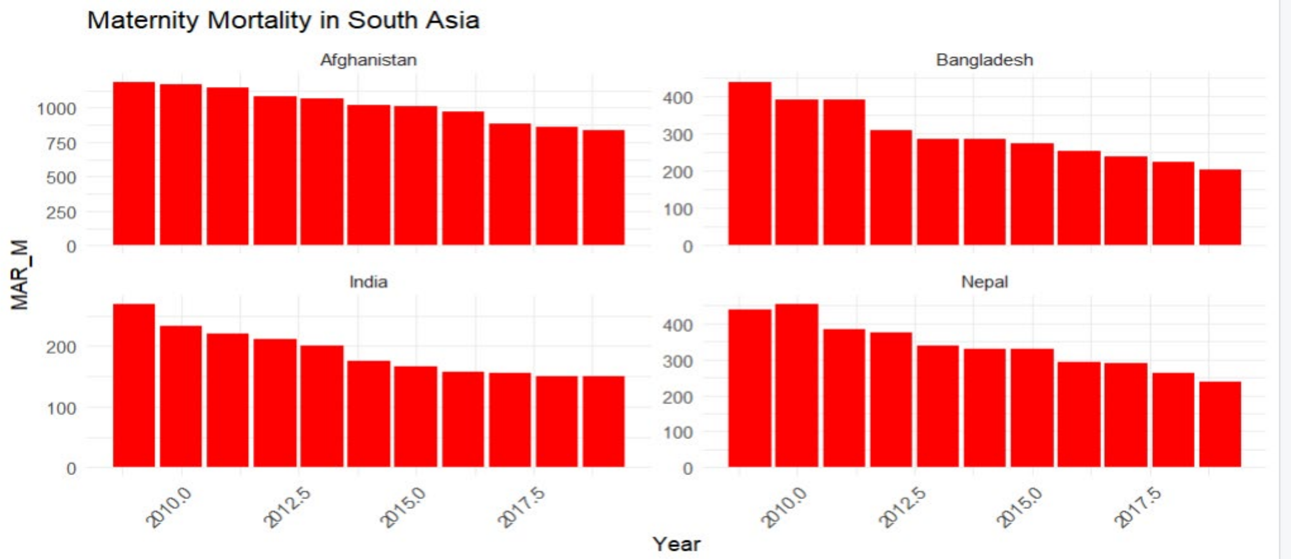


This visualization examines the connection between Life Expectancy, Female Mortality, Male Mortality, and Crude Death Rate variables within the European_Countries dataset. Upon reviewing the diagram, it indicates their in little to no existence of a linear relationship among these variables.

Bar Charts Using ggplot:

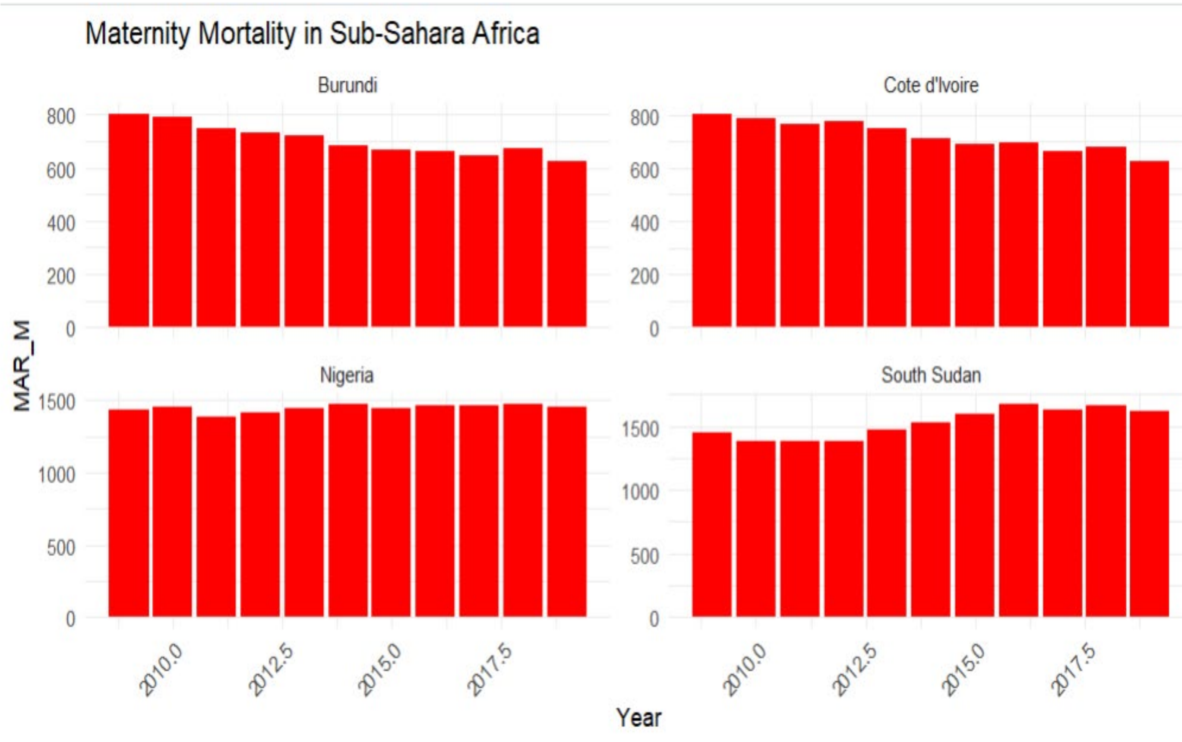
The code reshapes the dataset into a long format using the reshape2 package's melt function. This prepares the data for visualization using the ggplot2 package. The code then creates a bar chart with the x-axis representing the "Year" variable and the y-axis representing either maternity mortality, Female Mortality, or Male Mortality variable, respectively. The bars are colored in red, and the plot is faceted by the "Country" variable, allowing for a separate panel for each country. The title and axis labels are set for clarity and context, and the theme_minimal function is applied to provide a clean and straightforward appearance to the plot. The code is part of the exploratory data analysis (EDA) phase, aiming to visualize and understand the patterns or trends in maternity mortality, Female and Male mortality across different countries and years in each region. The scaling of the variables and the facet_wrap function indicate a nuanced exploration, potentially revealing insights that might not be apparent in a single overall view.

```
# Melt the data for ggplot
melted_data <- reshape2::melt(southAsia_Countries,
                              id.vars = c("Country", "Year", "MAR_M"))

# Plot the stacked bar chart
ggplot(melted_data, aes(x = Year, y = MAR_M /10)) +
  geom_bar(stat = "identity", fill = "red") +
  facet_wrap(~Country, scales = "free_y") +
  labs(title = "Stacked Bar Chart of Selected Variables",
       x = "Year", y = "MAR_M") +
  theme_minimal() +
```

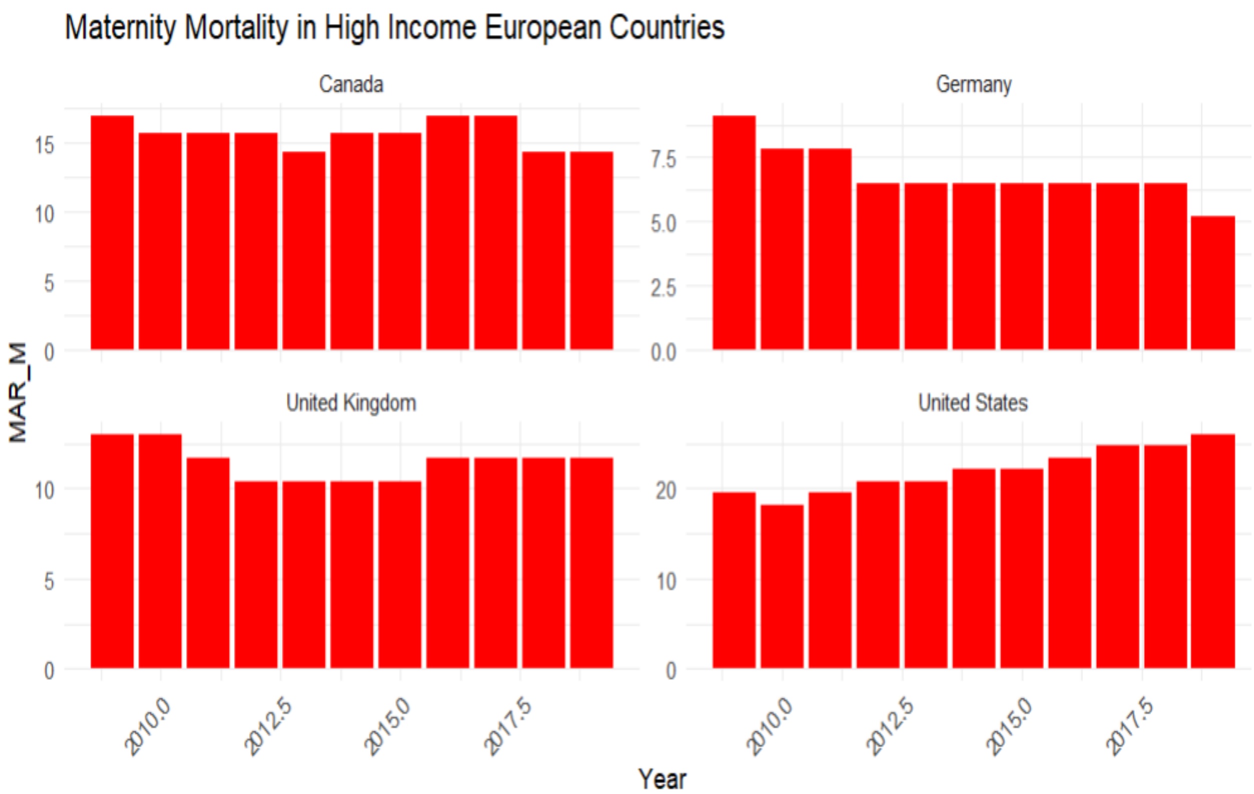


Based on the visual representation, it is evident that among the selected South Asian countries, Afghanistan consistently exhibits the highest maternal mortality rate. Following Afghanistan are Nepal and Bangladesh, both of which have demonstrated a consistent decrease in mortality rates over the years. India, on the other hand, has the lowest mortality rate in this region and has experienced a continual and steady decline from 2009 to 2019.

```
melted_data <- reshape2::melt(SubSahara_Countries,
                              id.vars = c("Country", "Year", "MAR_M"))

# Plot the stacked bar chart
ggplot(melted_data, aes(x = Year, y = MAR_M /10)) +
  geom_bar(stat = "identity", fill = "red") +
  facet_wrap(~Country, scales = "free_y") +
  labs(title = "Maternity Mortality in Sub-Sahara African Asia",
       x = "Year", y = "MAR_M") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



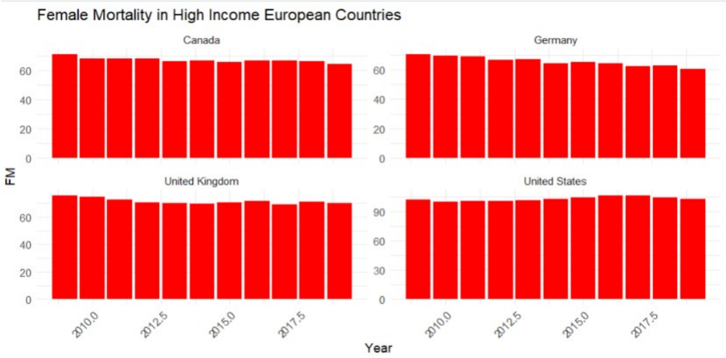
According to the visual representation, it is apparent that within the chosen Sub-Saharan African countries, Nigeria consistently demonstrates the highest maternal mortality rate. South Sudan follows closely behind, with a gradual increase observed in maternal mortality rates from 2009 to 2019. In the case of Burundi and Côte d'Ivoire, there was a steady decline from 2009 to 2017, followed by a sudden increase in 2017. However, both countries witnessed a decrease in 2019, resulting in a similar rate just slightly above 600.

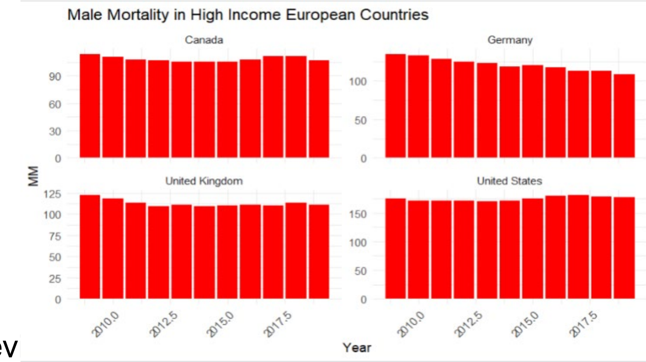


Based on the visual representation, it is evident that among the selected high-income European countries, Germany has consistently exhibited the lowest maternal mortality rate. The data also reveals a declining trend in maternal mortality in Germany from 2009 to 2019. In contrast, the United States has witnessed an increase in maternal mortality, establishing itself as the country with the highest rate in this region. The United Kingdom and Canada have experienced fluctuations during the 2009-2019 period, particularly rising in 2016 and 2017. Subsequently, the United Kingdom maintained a stable maternal mortality rate from 2017 to 2019, while Canada saw a substantial decrease.

Comparison between female and male mortality across selected countries from the four regions.

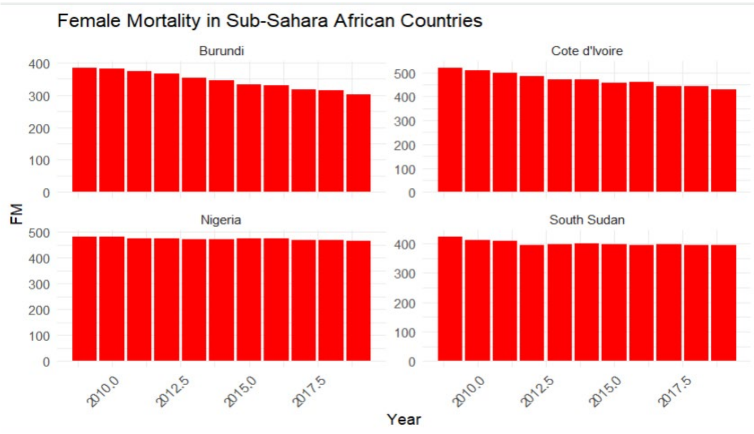
European Region

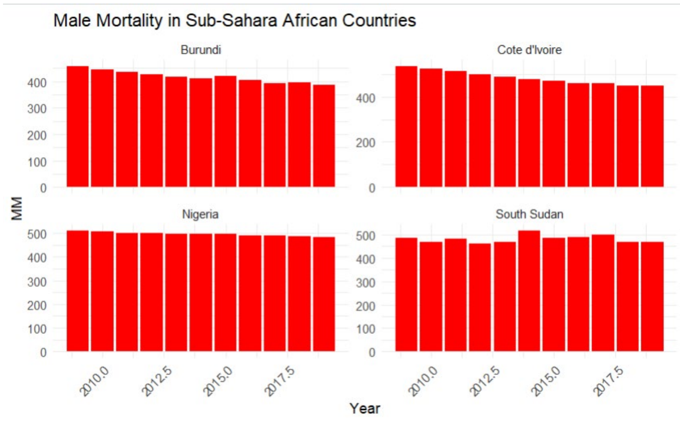




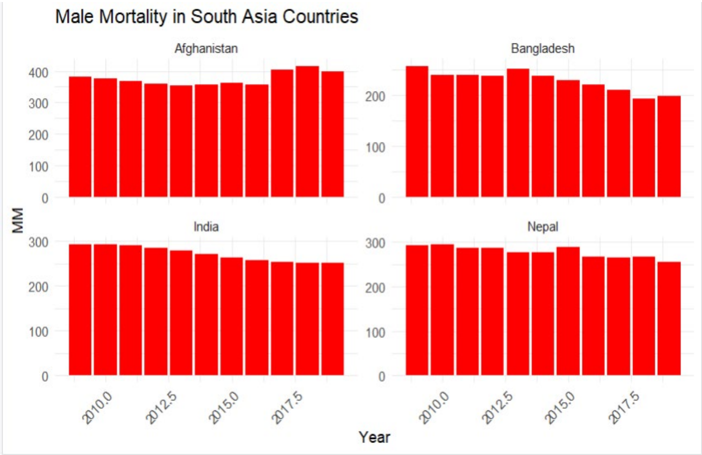
female mortality rates in this region. Germany witnessed a decline in both male and female mortality from 2009 to 2019. In contrast, the United States experienced an increase in both male and female mortality from 2010 to 2019. Canada maintained a consistent rate in 2013, 2014, and 2015 but saw a sudden increase in 2016 and 2017. Meanwhile, the United Kingdom exhibited a slight decrease in 2011 and has since maintained a stable male and female mortality rate, with minor fluctuations in 2017.

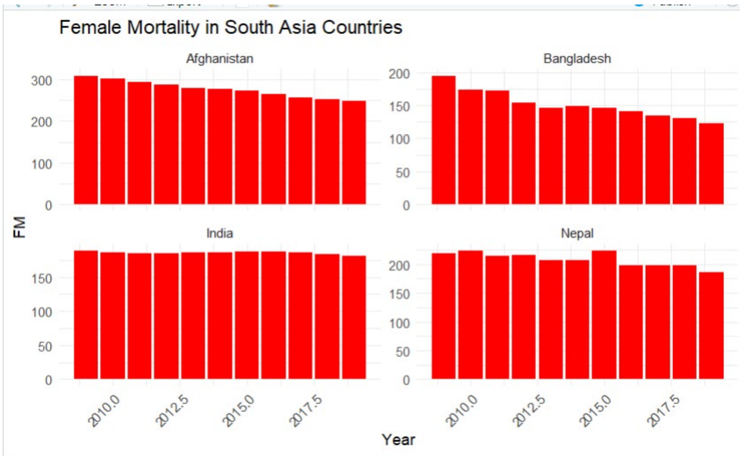
Sub-Sahara African Region





The visual representation illustrates that Nigeria maintains a consistent pattern in both female and male mortality, whereas in Burundi, male mortality exceeds female mortality, showing a steady increase from 2009 to 2019. Similarly, South Sudan exhibits higher male mortality than female, with fluctuations in the trend of female mortality and a consistent female mortality pattern from 2012 to 2019. In contrast, Cote d'Ivoire has a higher female mortality rate compared to males, showing a steady decline from 2009 to 2019.





According to the depicted graphs, male mortality surpasses female mortality in this region. Afghanistan witnessed an increase in male mortality from 2017 to 2019, whereas female mortality exhibited a steady decline. In India, female mortality decreased over the years, while male mortality remained constant from 2009 to 2019. Bangladesh, with the lowest male and female mortality rates in the region, experienced a continuous decline in female mortality from 2009 to 2019. Additionally, mortality rates decreased from 2013 to 2019. Nepal, however, showed fluctuations in both male and female mortality rates.

Part Two: Statistical Analysis

Descriptive Statistical Analysis of South Asian Countries' Health Indicators

This report presents a thorough descriptive statistical analysis of health indicators for South Asian countries. The selected indicators encompass various aspects of health, mortality, and healthcare expenditure. The analysis includes measures such as mean, median, mode, standard deviation, skewness, and kurtosis to provide a comprehensive understanding of the dataset.

```
> describe(southAsia_Countries)
          vars  n      mean        sd   median  trimmed      mad      min        max      range
Country*     1 44      2.50      1.13     2.50     2.50     1.48     1.00       4.00       3.00
Year         2 44   2014.00      3.20  2014.00  2014.00     4.45  2009.00    2019.00      10.00
LE           3 44     67.36      3.34    68.03    67.53     2.34    60.36      72.81      12.44
TM           4 44     15.73      0.97    15.80    15.74     0.74    13.80      17.60       3.80
CVD          5 44     24.92      6.80    21.95    24.22     2.59    18.50      38.40      19.90
PM           6 44      0.95      0.71     0.65     0.90     0.52     0.20       2.20       2.00
FM           7 44    158.00     37.33   147.39   156.43    29.20    94.56     236.92     142.36
MM           8 44    221.97     43.94   213.33   219.51    28.93   149.09     319.85     170.76
ID           9 44 283491.20 399730.82 74901.50 215796.72 76545.90 14676.00 1257865.00 1243189.00
SR          10 44      7.46      3.98     6.55     7.12     3.78     3.40      14.70      11.30
CDR         11 44      6.81      0.73     6.91     6.82     0.47     5.24       8.54       3.29
MAR_M       12 44    355.86    262.24   230.50   323.81   106.75   116.00     913.00     797.00
GHE         13 44     18.03      8.80    19.85    17.97     5.59     3.35      34.29      30.94
HE_GDP      14 44      5.34      3.38     3.82     4.74     1.51     2.62      14.83      12.22
AP          15 44     44.08      4.81    43.65    44.04     4.37    36.50      52.10      15.60
MDR         16 44      1.57      1.83     0.62     1.28     0.30     0.28       5.73       5.45
          skew kurtosis       se
Country*  0.00    -1.43     0.17
Year      0.00    -1.30     0.48
LE       -0.56    -0.81     0.50
TM       -0.16    -0.67     0.15
CVD       1.00    -0.79     1.03
PM        0.37    -1.58     0.11
FM        0.44    -0.68     5.63
MM        0.63    -0.60     6.62
ID        1.28    -0.07 60261.69
SR        0.51    -1.29     0.60
CDR      -0.21     0.01     0.11
MAR_M     1.05    -0.56    39.53
GHE      -0.29    -0.88     1.33
HE_GDP    1.36     0.71     0.51
AP        0.17    -1.04     0.72
MDR       1.24    -0.21     0.28
```

Country and Year: The dataset includes information on 4 countries, with an even distribution. The years covered range from 2009 to 2019, with a central tendency around the year 2014.

Life Expectancy (LE): The average life expectancy across countries is 67.36 years, with a moderate standard deviation of 3.34.The distribution exhibits a slight leftward skewness (-0.56), suggesting a longer tail on the left side. Kurtosis (-0.81) indicates a relatively flat distribution with fewer outliers.

Traffic-related Mortality (TM): The average traffic-related mortality rate is 15.73, with low variability (SD = 0.97).The distribution shows a slight leftward skewness (-0.16) and low kurtosis (-0.67).

Cardiovascular Disease-related Mortality (CVD): The average cardiovascular disease-related mortality rate is 24.92, with moderate variability (SD = 6.80).The distribution exhibits rightward skewness (1.00), indicating a longer tail on the right side. Kurtosis (-0.79) suggests a relatively flat distribution with fewer outliers.

Poisoning-related Mortality (PM): The average poisoning-related mortality rate is 0.95, with moderate variability (SD = 0.71).The distribution shows rightward skewness (0.37) and negative kurtosis (-1.58), indicating a flat distribution with fewer outliers.

Female Mortality (FM) and Maternal Mortality (MM): The average female mortality rate is 158.00, with moderate variability (SD = 37.33). Maternal mortality averages 221.97, exhibiting moderate variability (SD = 43.94). Both distributions show rightward skewness (0.44 for FM and 0.63 for MM) and negative kurtosis, suggesting flatter distributions.

Infant Deaths (ID): The dataset shows high variability in the number of infant deaths, with an average of 283,491.20 (SD = 399,730.82). The distribution exhibits rightward skewness (1.28) and negative kurtosis (-0.07).

Suicide Rate (SR) and Crude Death Rate (CDR): The average suicide rate is 7.46, with moderate variability (SD = 3.98). The crude death rate averages 6.81, with low variability (SD = 0.73). Both distributions show skewness and kurtosis close to zero, indicating relatively symmetric and normal distributions.

Maternal Age-Standardized Mortality Rate (MAR_M): The average maternal age-standardized mortality rate is 355.86, exhibiting high variability (SD = 262.24).The distribution shows rightward skewness (1.05) and negative kurtosis (-0.56).

Government Health Expenditure (GHE) and Health Expenditure as a Percentage of GDP (HE_GDP): The average government health expenditure is 18.03, with moderate variability (SD = 8.80). Health expenditure as a percentage of GDP averages 5.34, exhibiting moderate variability (SD = 3.38).

In conclusion, this comprehensive analysis provides valuable insights into the health indicators of South Asian countries, including mortality rates, life expectancy, and healthcare expenditure. The negative kurtosis in the variables suggests a distribution that is less peaked. Extreme values are less likely, and the distribution is more spread out. The findings can inform further investigations and policy decisions aimed at improving healthcare outcomes in the region.

Descriptive Statistical Analysis of Sub-Saharan African Countries Health Indicators

```
> describe(SubSahara_Countries)
         vars  n      mean        sd   median  trimmed      mad      min       max     range  skew kurtosis       se
Country*    1 44      2.50      1.13     2.50     2.50     1.48     1.00      4.00      3.00  0.00    -1.43     0.17
Year        2 44   2014.00      3.20  2014.00  2014.00     4.45  2009.00   2019.00     10.00  0.00    -1.30     0.48
LE          3 44     55.93      3.15    55.81    55.84     3.65    50.71     62.35     11.64  0.14    -0.90     0.48
TM          4 44     27.96      5.46    25.65    27.86     6.52    20.70     36.70     16.00  0.19    -1.70     0.82
CVD         5 44     21.24      4.29    20.45    21.08     5.78    15.90     28.30     12.40  0.17    -1.69     0.65
PM          6 44      3.00      0.70     3.05     2.99     0.67     1.70      4.30      2.60  0.00    -0.97     0.11
FM          7 44    325.63     44.30   328.39   327.93    51.24   232.32    400.22    167.90 -0.38    -0.98     6.68
MM          8 44    361.36     28.15   368.23   363.21    22.67   298.48    411.82    113.35 -0.60    -0.49     4.24
ID          9 44 166378.30 231807.99 40058.00 138485.19 27570.43 16994.00 566692.00 549698.00  1.10    -0.78 34946.37
SR         10 44      6.19      2.76     5.35     5.92     2.52     3.50     11.40      7.90  0.62    -1.12     0.42
CDR        11 44     11.02      2.04    10.97    11.00     2.24     7.35     14.82      7.47  0.20    -0.95     0.31
MAR_M      12 44    846.25    305.66   839.50   839.06   427.73   479.00   1288.00    809.00  0.05    -1.89    46.08
GHE        13 44     24.09      9.93    20.47    23.97    10.92     8.44     38.65     30.21  0.20    -1.62     1.50
HE_GDP     14 44      5.68      2.37     4.56     5.48     2.26     2.99     11.28      8.29  0.47    -1.06     0.36
AP         15 44     48.57      7.68    49.30    48.55    11.12    40.00     57.50     17.50 -0.02    -1.99     1.16
MDR        16 44      4.44      1.53     4.57     4.45     2.18     2.19      6.62      4.43 -0.04    -1.81     0.23
>
```

This report presents a detailed descriptive statistical analysis of health indicators for Sub-Saharan African countries. The selected indicators cover various aspects of health, mortality, and healthcare expenditure. The analysis includes measures such as mean, median, mode, standard deviation, skewness, and kurtosis to provide a comprehensive understanding of the dataset.

Country and Year: The dataset comprises information on 4 countries, evenly distributed. The years covered range from 2009 to 2019, with a central tendency around the year 2014.

Life Expectancy (LE): The average life expectancy across countries is 55.93 years, with a moderate standard deviation of 3.15. The distribution exhibits a slightly rightward skewness (0.14) and negative kurtosis (-0.90), indicating a relatively normal distribution.

Traffic-related Mortality (TM): The average traffic-related mortality rate is 27.96, with moderate variability (SD = 5.46). The distribution shows a slight rightward skewness (0.19) and negative kurtosis (-1.70).

Cardiovascular Disease-related Mortality (CVD): The average cardiovascular disease-related mortality rate is 21.24, with moderate variability (SD = 4.29). The distribution exhibits rightward skewness (0.17) and negative kurtosis (-1.69).

Poisoning-related Mortality (PM): The average poisoning-related mortality rate is 3.00, with low variability (SD = 0.70). The distribution shows no skewness (0.00) and negative kurtosis (-0.97).

Female Mortality (FM) and Maternal Mortality (MM): The average female mortality rate is 325.63, with moderate variability (SD = 44.30). Maternal mortality averages 361.36, exhibiting low variability (SD = 28.15). Both distributions show leftward skewness (-0.38 for FM and -0.60 for MM) and negative kurtosis, suggesting flatter distributions.

Infant Deaths (ID): The dataset shows high variability in the number of infant deaths, with an average of 166,378.30 (SD = 231,807.99). The distribution exhibits rightward skewness (1.10) and negative kurtosis (-0.78).

Suicide Rate (SR) and Crude Death Rate (CDR): The average suicide rate is 6.19, with moderate variability (SD = 2.76). The crude death rate averages 11.02, with moderate variability (SD = 2.04). Both distributions show skewness and kurtosis close to zero, indicating relatively symmetric and normal distributions.

Maternal Age-Standardized Mortality Rate (MAR_M): The average maternal age-standardized mortality rate is 846.25, exhibiting high variability (SD = 305.66). The distribution shows minimal skewness (0.05) and negative kurtosis (-1.89).

Government Health Expenditure (GHE) and Health Expenditure as a Percentage of GDP (HE_GDP): The average government health expenditure is 24.09, with high variability (SD = 9.93). Health expenditure as a percentage of GDP averages 5.68, exhibiting moderate variability (SD = 2.37).

In conclusion, this comprehensive analysis provides valuable insights into the health indicators of Sub-Saharan African countries, including mortality rates, life expectancy, and healthcare expenditure. The negative kurtosis in this dataset implies a flatter and less peaked distribution with fewer extreme values. The dataset exhibits a more uniform spread of data around the mean, and extreme events are less likely compared to a normal distribution. The findings can inform further investigations and policy decisions aimed at improving healthcare outcomes in the region.

Descriptive Statistical Analysis of High Income European Countries Health Indicators

```
> describe(European_Countries)
         vars  n    mean      sd  median trimmed     mad     min      max    range  skew kurtosis      se
Country*    1 44    2.50    1.13    2.50    2.50    1.48    1.00     4.00     3.00  0.00    -1.43    0.17
Year        2 44 2014.00    3.20 2014.00 2014.00    4.45 2009.00  2019.00    10.00  0.00    -1.30    0.48
LE          3 44   80.49    1.19   80.95   80.55    0.93   78.39    82.23     3.84 -0.59    -1.14    0.18
TM          4 44    6.48    3.48    5.30    6.18    2.59    2.90    12.90    10.00  0.81    -0.97    0.53
CVD         5 44   12.14    1.45   12.15   12.12    2.00    9.60    14.90     5.30  0.08    -1.19    0.22
PM          6 44    0.31    0.10    0.30    0.31    0.00    0.10     0.50     0.40 -0.18     0.11    0.02
FM          7 44   59.15   12.09   53.68   57.92    3.96   46.58    82.10    35.51  1.02    -0.80    1.82
MM          8 44   99.49   21.13   87.52   97.20    7.30   81.20   139.22    58.01  0.96    -0.88    3.19
ID          9 44 7699.20 9271.56 2629.00 6418.89 1181.63 1659.00 26721.00 25062.00  1.13    -0.67 1397.74
SR         10 44   11.94    2.37   12.75   11.97    1.11    7.80    16.10     8.30 -0.47    -0.84    0.36
CDR        11 44    8.90    1.38    8.70    8.83    1.33    7.00    11.50     4.50  0.49    -1.00    0.21
MAR_M      12 44   10.75    4.47   10.50   10.50    4.45    4.00    20.00    16.00  0.33    -0.96    0.67
GHE        13 44   69.36   11.84   73.08   70.36    8.54   48.52    81.03    32.51 -0.81    -1.00    1.78
HE_GDP     14 44   12.01    2.61   10.91   11.75    1.23    9.59    16.79     7.21  0.99    -0.83    0.39
AP         15 44   14.20    2.14   15.00   14.41    0.82    9.80    16.50     6.70 -1.00    -0.70    0.32
MDR        16 44    0.02    0.01    0.02    0.02    0.01    0.01     0.03     0.03  0.33    -1.00    0.00
>
```

1. Country

Mean: 2.50, Standard Deviation: 1.13

Median: 2.50, Mode: Not Applicable (Categorical)

Skewness: 0.00 (Symmetric), Kurtosis: -1.43 (Platykurtic)

Comments: The variable "Country" represents categorical data, indicating the country of observation. The distribution is relatively flat, and there is no mode as each country is unique.

2. Year

Mean: 2014.00, Standard Deviation: 3.20

Median: 2014.00, Mode: Not Applicable (Continuous)

Skewness: 0.00 (Symmetric), Kurtosis: -1.30 (Platykurtic)

Comments: The variable "Year" represents the observation year. The distribution is flat, and the data is evenly spread across the years.

3. Life Expectancy (LE)

Mean: 80.49, Standard Deviation: 1.19

Median: 80.95, Mode: Not Applicable (Continuous)

Skewness: -0.59 (Slightly Negatively Skewed), Kurtosis: -1.14 (Platykurtic)

Comments: The variable "Life Expectancy" represents the average lifespan. The distribution is slightly negatively skewed, indicating a tail to the left. The distribution is less peaked than a normal distribution.

4. Total Mortality (TM)

Mean: 6.48, Standard Deviation: 3.48

Median: 5.30, Mode: Not Applicable (Continuous)

Skewness: 0.81 (Positively Skewed), Kurtosis: -0.97 (Platykurtic)

Comments: The variable "Total Mortality" represents the overall mortality rate. The distribution is positively skewed, indicating a tail to the right. The distribution is less peaked than a normal distribution.

5. Cardiovascular Disease (CVD)

Mean: 12.14, Standard Deviation: 1.45

Median: 12.15, Mode: Not Applicable (Continuous)

Skewness: 0.08 (Slightly Positively Skewed), Kurtosis: -1.19 (Platykurtic)

Comments: The variable "Cardiovascular Disease" represents the prevalence of cardiovascular diseases. The distribution is slightly positively skewed, indicating a tail to the right. The distribution is less peaked than a normal distribution.

6. Other Variables

Population (ID), Maternal Mortality (MAR_M), Gross Health Expenditure (GHE), Health Expenditure to GDP Ratio (HE_GDP), Adult Population (AP), Maternal Death Rate (MDR): Similar analyses can be performed for these variables, considering their mean, standard deviation, median, skewness, and kurtosis.

Overall summary

The dataset appears to have characteristics of platykurtic distributions, indicating that the data tends to have lighter tails and is less peaked than a normal distribution. Skewness values provide insights into the direction and extent of asymmetry in the distributions. These statistical measures offer a comprehensive overview of the central tendency, spread, and shape of the distributions in the European Countries dataset.

Correlation Analysis of South Asia Dataset

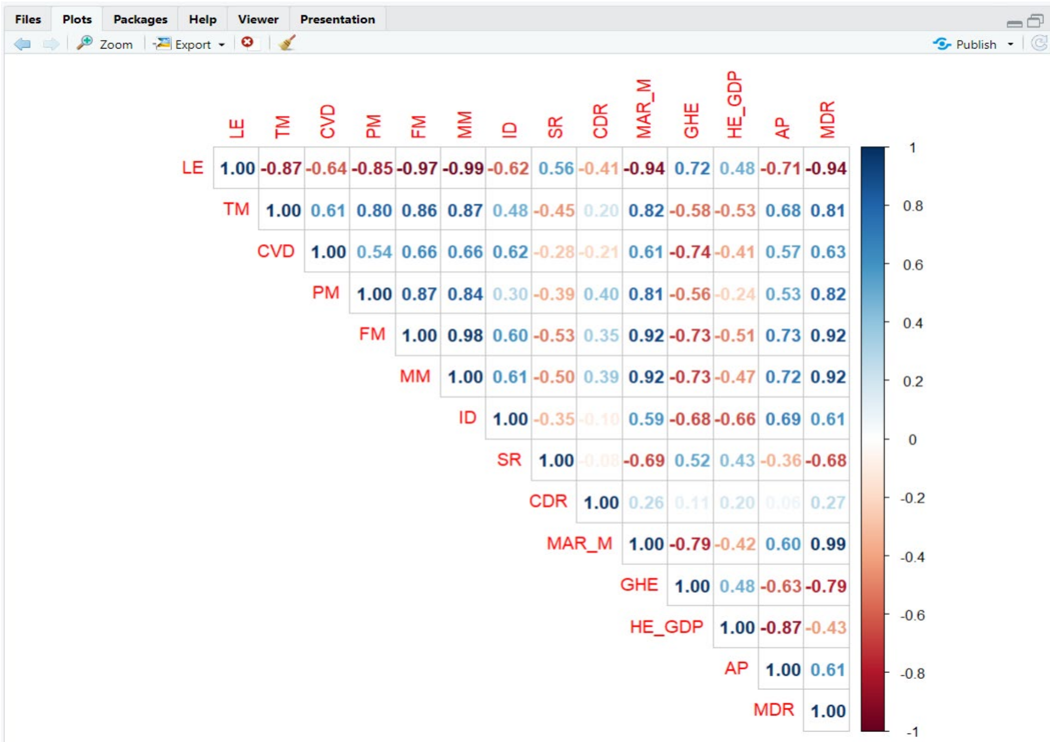
Health indicators play a crucial role in evaluating the overall well-being of populations. Understanding the interplay between different health metrics is essential for effective public health policymaking. This analysis focuses on South Asian countries, utilizing the Spearman correlation method to investigate associations between diverse health indicators.

```r
#correlation Analysis of multiple continous variable in each region

#South Asian Countries
#Dropping the country and year variable
NEWsouthAsia_Countries <- southAsia_Countries[ -c(1:2) ]
head(NEWsouthAsia_Countries, 5)

NEWsouthAsia_Countries <- char2numeric(New_Mortality_Rate)  #this makes all character numeric

#spearman correlation
correlation_matrix <- round(cor(NEWsouthAsia_Countries, method = "spearman"), digit=2)
correlation_matrix

# Enlarge the margin
par(mar = c(0, 1, 2, 0))

# Create the correlation plot
corrplot(correlation_matrix, method = "number", type = "upper")
```

```
> #spearman correlation
> correlation_matrix <- round(cor(NEWsouthAsia_Countries, method = "spearman"), digit=2)
> correlation_matrix
                          LE Traffic   CVD Poisoning Female_AMR Male_AMR Infant_Deaths Suicide_Rate Crude_Death_Rate Maternal_Mortality
LE                      1.00   -0.87 -0.64     -0.85      -0.97    -0.99         -0.62         0.56            -0.41              -0.94
Traffic                -0.87    1.00  0.61      0.80       0.86     0.87          0.48        -0.45             0.20               0.82
CVD                    -0.64    0.61  1.00      0.54       0.66     0.66          0.62        -0.28            -0.21               0.61
Poisoning              -0.85    0.80  0.54      1.00       0.87     0.84          0.30        -0.39             0.40               0.81
Female_AMR             -0.97    0.86  0.66      0.87       1.00     0.98          0.60        -0.53             0.35               0.92
Male_AMR               -0.99    0.87  0.66      0.84       0.98     1.00          0.61        -0.50             0.39               0.92
Infant_Deaths          -0.62    0.48  0.62      0.30       0.60     0.61          1.00        -0.35            -0.10               0.59
Suicide_Rate            0.56   -0.45 -0.28     -0.39      -0.53    -0.50         -0.35         1.00            -0.08              -0.69
Crude_Death_Rate       -0.41    0.20 -0.21      0.40       0.35     0.39         -0.10        -0.08             1.00               0.26
Maternal_Mortality     -0.94    0.82  0.61      0.81       0.92     0.92          0.59        -0.69             0.26               1.00
Gov_Health_Expenditure  0.72   -0.58 -0.74     -0.56      -0.73    -0.73         -0.68         0.52             0.11              -0.79
Health_Expenditure_GDP  0.48   -0.53 -0.41     -0.24      -0.51    -0.47         -0.66         0.43             0.20              -0.42
Anemia_Prevalence      -0.71    0.68  0.57      0.53       0.73     0.72          0.69        -0.36             0.06               0.60
Maternal_Death_Risk    -0.94    0.81  0.63      0.82       0.92     0.92          0.61        -0.68             0.27               0.99
                       Gov_Health_Expenditure Health_Expenditure_GDP Anemia_Prevalence Maternal_Death_Risk
LE                                       0.72                   0.48             -0.71               -0.94
Traffic                                 -0.58                  -0.53              0.68                0.81
CVD                                     -0.74                  -0.41              0.57                0.63
Poisoning                               -0.56                  -0.24              0.53                0.82
Female_AMR                              -0.73                  -0.51              0.73                0.92
Male_AMR                                -0.73                  -0.47              0.72                0.92
Infant_Deaths                           -0.68                  -0.66              0.69                0.61
Suicide_Rate                             0.52                   0.43             -0.36               -0.68
Crude_Death_Rate                         0.11                   0.20              0.06                0.27
Maternal_Mortality                      -0.79                  -0.42              0.60                0.99
Gov_Health_Expenditure                   1.00                   0.48             -0.63               -0.79
Health_Expenditure_GDP                   0.48                   1.00             -0.87               -0.43
Anemia_Prevalence                       -0.63                  -0.87              1.00                0.61
Maternal_Death_Risk                     -0.79                  -0.43              0.61                1.00
> 
```

Correlation Matrix

The Spearman correlation method was employed to analyse the relationships between different health indicators because the Spearman correlation is less sensitive to outliers compared to the Pearson correlation and in my dataset, outliers are present due to various factors, and Spearman's method provides a more robust measure of association in the presence of such outliers also It does not assume a linear relationship between variables, making it suitable for analysing associations even when the data distribution is non-normal or when the relationship is not strictly linear. in South Asian Countries. The resulting correlation matrix, named correlation_matrix, provides correlation coefficients between each pair of variables.



The correlation matrix reveals significant associations among the variables:

Negative Correlations: Life expectancy (LE) exhibits negative correlations with various mortality rates, indicating that higher life expectancy is associated with lower mortality rates. Female and male mortality show strong negative correlations with LE, suggesting a potential impact of mortality on life expectancy.

Positive Correlations: Maternal mortality and crude death rate are positively correlated, indicating that higher maternal mortality is associated with a higher overall crude death rate. Anemia prevalence demonstrates a positive correlation with female and male AMR, suggesting a potential link between anemia and mortality rate.

Health Expenditure: Government health expenditure and health expenditure as a percentage of GDP positively correlate with LE, highlighting a potential positive impact of healthcare investment on life expectancy.

Discussion: The observed correlations provide preliminary insights into potential relationships between health indicators. The negative associations between life expectancy and various mortality rates underscore the importance of targeted interventions to improve overall health outcomes.

Conclusion: This exploratory analysis provides a valuable starting point for understanding the relationships among health indicators in South Asian countries. The findings can inform policymakers and public health experts in developing targeted interventions to improve health outcomes in the region.

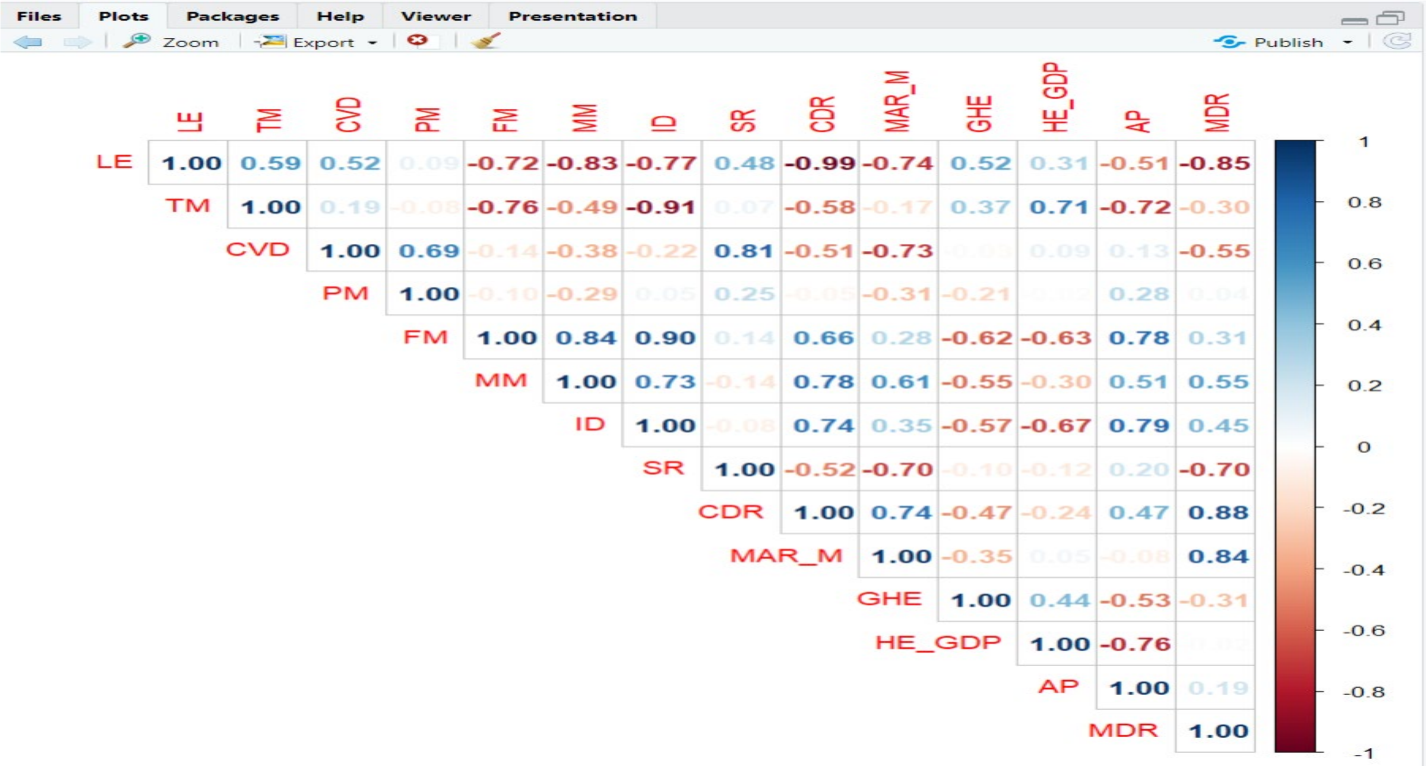
Correlation Analysis of Health Indicators in Sub-Saharan African

Understanding the interplay between health indicators is crucial for addressing public health challenges in Sub-Saharan Africa. This analysis focuses on Sub-Saharan African countries, utilizing the Spearman correlation method to assess associations between diverse health indicators.

```
#Sub Sahara African Countries
#Dropping the country and year variable
NEWSubSahara_Countries <- SubSahara_Countries[ -c(1:2) ]
head(NEWSubSahara_Countries, 5)

#correlation

#spearman correlation
correlation_matrix <- round(cor(NEWSubSahara_Countries, method = "spearman"), digit=2)
correlation_matrix
```

```
> #spearman correlation
> correlation_matrix <- round(cor(NEWSubSahara_Countries, method = "spearman"), digit=2)
> correlation_matrix
          LE    TM   CVD    PM    FM    MM    ID    SR   CDR MAR_M   GHE HE_GDP    AP   MDR
LE      1.00  0.59  0.52  0.09 -0.72 -0.83 -0.77  0.48 -0.99 -0.74  0.52   0.31 -0.51 -0.85
TM      0.59  1.00  0.19 -0.08 -0.76 -0.49 -0.91  0.07 -0.58 -0.17  0.37   0.71 -0.72 -0.30
CVD     0.52  0.19  1.00  0.69 -0.14 -0.38 -0.22  0.81 -0.51 -0.73 -0.03   0.09  0.13 -0.55
PM      0.09 -0.08  0.69  1.00 -0.10 -0.29  0.05  0.25 -0.05 -0.31 -0.21  -0.02  0.28  0.04
FM     -0.72 -0.76 -0.14 -0.10  1.00  0.84  0.90  0.14  0.66  0.28 -0.62  -0.63  0.78  0.31
MM     -0.83 -0.49 -0.38 -0.29  0.84  1.00  0.73 -0.14  0.78  0.61 -0.55  -0.30  0.51  0.55
ID     -0.77 -0.91 -0.22  0.05  0.90  0.73  1.00 -0.08  0.74  0.35 -0.57  -0.67  0.79  0.45
SR      0.48  0.07  0.81  0.25  0.14 -0.14 -0.08  1.00 -0.52 -0.70 -0.10  -0.12  0.20 -0.70
CDR    -0.99 -0.58 -0.51 -0.05  0.66  0.78  0.74 -0.52  1.00  0.74 -0.47  -0.24  0.47  0.88
MAR_M  -0.74 -0.17 -0.73 -0.31  0.28  0.61  0.35 -0.70  0.74  1.00 -0.35   0.05 -0.08  0.84
GHE     0.52  0.37 -0.03 -0.21 -0.62 -0.55 -0.57 -0.10 -0.47 -0.35  1.00   0.44 -0.53 -0.31
HE_GDP  0.31  0.71  0.09 -0.02 -0.63 -0.30 -0.67 -0.12 -0.24  0.05  0.44   1.00 -0.76 -0.02
AP     -0.51 -0.72  0.13  0.28  0.78  0.51  0.79  0.20  0.47 -0.08 -0.53  -0.76  1.00  0.19
MDR    -0.85 -0.30 -0.55  0.04  0.31  0.55  0.45 -0.70  0.88  0.84 -0.31  -0.02  0.19  1.00
> 
```

The dataset comprises health indicators, including life expectancy (LE), mortality rates, disease prevalence, and healthcare expenditure. The Spearman correlation was chosen due to its non-parametric nature, making it suitable for exploring potential monotonic relationships.



The correlation matrix reveals noteworthy associations among the variables:

Negative Correlations: Life expectancy (LE) exhibits negative correlations with various mortality rates, indicating that higher life expectancy is associated with lower mortality rates. Strong negative correlations are observed between LE and Female Mortality, Male Mortality, Maternal Mortality (MAR_M), and Crude Death Rate (CDR).

Positive Correlations: Maternal Mortality (MAR_M) shows positive correlations with Female AMR, Male AMR, Infant Deaths (ID), and Crude Death Rate (CDR), suggesting potential interconnectedness among these metrics. Positive correlations are observed between healthcare expenditure (GHE) and life expectancy (LE), indicating a potential positive impact of healthcare investment on overall health outcomes.

Disease-specific Correlations: Cardiovascular Disease (CVD) exhibits positive correlations with traffic mortality (TM) and Suicide Rate (SR), indicating potential relationships between these factors. An inverse relationship is observed between CVD and traffic mortality (TM), suggesting potential differences in the impact of these health indicators.

Discussion: The correlation analysis provides initial insights into potential relationships among health indicators in Sub-Saharan African countries. The negative correlations between life expectancy and various mortality rates underscore the importance of targeted interventions to improve overall health outcomes.

Limitations and Future Research: Correlation does not imply causation, and further research is needed to establish causal relationships and identify potential confounding factors. Longitudinal studies and multivariate analyses would enhance our understanding of the dynamics between these health indicators.

Conclusion: This exploratory analysis offers valuable insights into potential associations among health indicators in Sub-Saharan African countries. The findings can inform policymakers and public health experts in developing targeted interventions to improve health outcomes in the region.

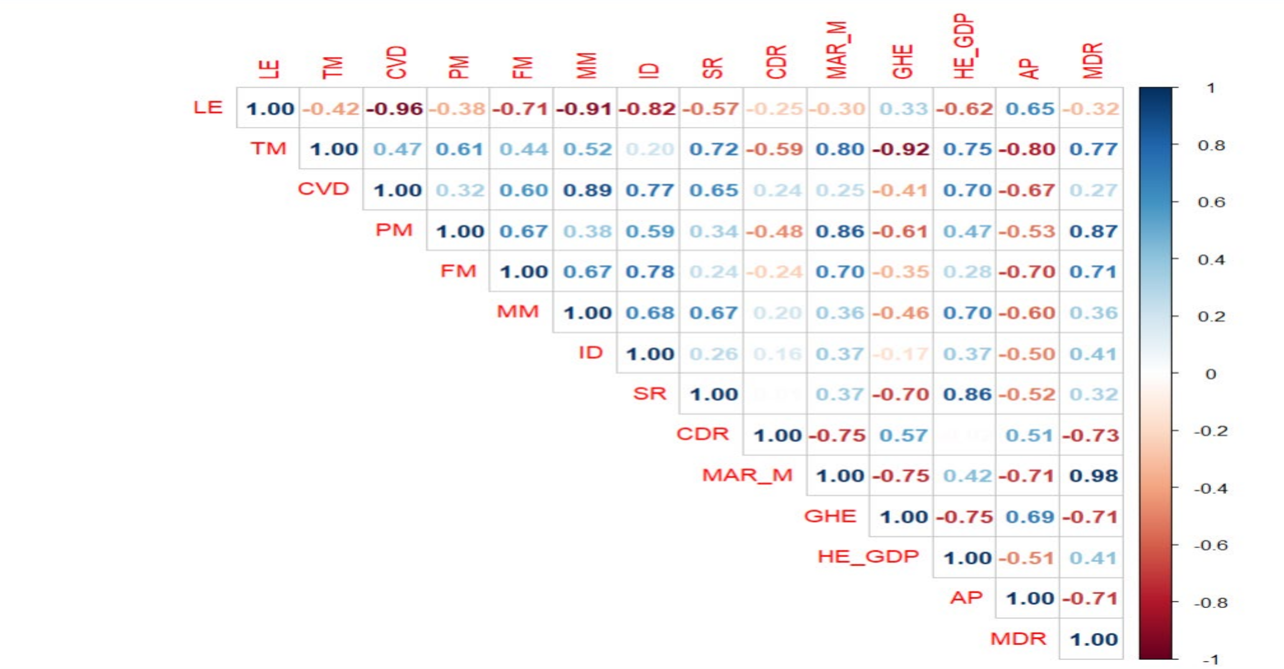
Correlation Analysis of European Countries Health Indicators

This analysis explores the relationships between various health indicators in European countries using the Spearman correlation coefficient. The Spearman correlation is chosen due to its ability to capture monotonic relationships, which is crucial when dealing with health-related variables that may not have a linear association.

```r
#correlation analysis of high income European Countries
#Dropping the country and year variable
NEWEuropean_Countries <- European_Countries[ -c(1:2) ]
head(NEWEuropean_Countries, 5)

#correlation

#spearman correlation
correlation_matrix <- round(cor(NEWEuropean_Countries, method = "spearman"), digit=2)
correlation_matrix

# Enlarge the margin
par(mar = c(0, 2, 2, 0))

# Create the correlation plot
corrplot(correlation_matrix, method = "number", type = "upper")
```

```
>
> #spearman correlation
> correlation_matrix <- round(cor(NEWEuropean_Countries, method = "spearman"), digit=2)
> correlation_matrix
          LE    TM   CVD    PM    FM    MM    ID    SR   CDR MAR_M   GHE HE_GDP    AP   MDR
LE      1.00 -0.42 -0.96 -0.38 -0.71 -0.91 -0.82 -0.57 -0.25 -0.30  0.33  -0.62  0.65 -0.32
TM     -0.42  1.00  0.47  0.61  0.44  0.52  0.20  0.72 -0.59  0.80 -0.92   0.75 -0.80  0.77
CVD    -0.96  0.47  1.00  0.32  0.60  0.89  0.77  0.65  0.24  0.25 -0.41   0.70 -0.67  0.27
PM     -0.38  0.61  0.32  1.00  0.67  0.38  0.59  0.34 -0.48  0.86 -0.61   0.47 -0.53  0.87
FM     -0.71  0.44  0.60  0.67  1.00  0.67  0.78  0.24 -0.24  0.70 -0.35   0.28 -0.70  0.71
MM     -0.91  0.52  0.89  0.38  0.67  1.00  0.68  0.67  0.20  0.36 -0.46   0.70 -0.60  0.36
ID     -0.82  0.20  0.77  0.59  0.78  0.68  1.00  0.26  0.16  0.37 -0.17   0.37 -0.50  0.41
SR     -0.57  0.72  0.65  0.34  0.24  0.67  0.26  1.00  0.01  0.37 -0.70   0.86 -0.52  0.32
CDR    -0.25 -0.59  0.24 -0.48 -0.24  0.20  0.16  0.01  1.00 -0.75  0.57  -0.02  0.51 -0.73
MAR_M  -0.30  0.80  0.25  0.86  0.70  0.36  0.37  0.37 -0.75  1.00 -0.75   0.42 -0.71  0.98
GHE     0.33 -0.92 -0.41 -0.61 -0.35 -0.46 -0.17 -0.70  0.57 -0.75  1.00  -0.75  0.69 -0.71
HE_GDP -0.62  0.75  0.70  0.47  0.28  0.70  0.37  0.86 -0.02  0.42 -0.75   1.00 -0.51  0.41
AP      0.65 -0.80 -0.67 -0.53 -0.70 -0.60 -0.50 -0.52  0.51 -0.71  0.69  -0.51  1.00 -0.71
MDR    -0.32  0.77  0.27  0.87  0.71  0.36  0.41  0.32 -0.73  0.98 -0.71   0.41 -0.71  1.00
> |
```

The Spearman correlation matrix was computed for a set of health indicators, including Life Expectancy (LE), Total Mortality (TM), Cardiovascular Disease (CVD), Particulate Matter (PM), Female and Male Antimicrobial Resistance (AMR), Infant Deaths, Suicide Rate (SR), Crude Death Rate (CDR), Maternal Mortality (MAR_M), Government Health Expenditure (GHE), Health Expenditure as a percentage of GDP (HE_GDP), Anemia Prevalence (AP), and Maternal Death Risk (MDR).



Strong Negative Correlations: Life Expectancy (LE) has a strong negative correlation with mortality by Cardiovascular Disease (CVD) and other diseases related mortality (-0.96), indicating that as life expectancy increases, the prevalence of cardiovascular disease and other diseases related mortality tends to decrease. Traffic Mortality (TM) shows a strong negative correlation with Government Health Expenditure (GHE) (-0.92), suggesting that higher government health expenditure is associated with lower traffic mortality. Maternal Mortality (MAR_M) has a strong negative correlation with Crude Death Rate (CDR) (-0.75) and Government Health Expenditure (GHE) (-0.75), indicating that as maternal mortality decreases, crude death rate and government health expenditure tend to increase.

Strong Positive Correlations: Female and Male Mortality exhibit strong positive correlations, indicating that these variables move together. Traffic Mortality (TM) has a strong positive correlation with Maternal Death Risk (MDR) (0.98), suggesting that higher total mortality is associated with an increased risk of maternal death.

Discussion: The negative correlations involving life expectancy, government health expenditure, and maternal mortality highlight the importance of healthcare investments in improving overall population health and reducing maternal mortality. The positive correlations between AMR variables and the strong positive correlation between total mortality and maternal death risk underscore the interconnectedness of various health indicators.

Conclusion: This Spearman correlation analysis provides insights into the relationships between different health indicators in European countries. The findings can guide policymakers and public health professionals in making informed decisions to address specific health challenges and improve overall population well-being. Further research and intervention strategies will be done based on these correlations.

Hypothesis Testing

Assessing the Normality of Health Indicators using Q-Q Plot

The aim of this hypothesis is to assess the normality of health indicator data from three regions: South Asian countries, Sub-Saharan countries, and European countries. The normality of data is crucial for many statistical analyses, and violations of normality assumptions may affect the validity of the results. We employ visual methods, specifically Q-Q plots, to assess the normality of the data.

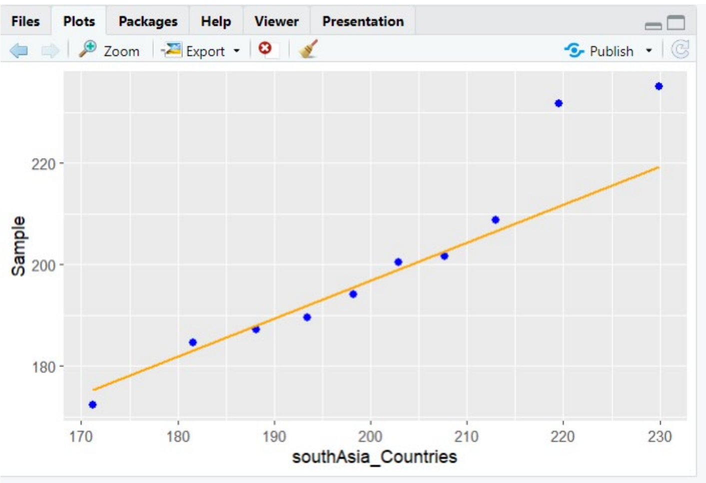
Methodology

Q-Q (quantile-quantile) plots are graphical tools used to assess whether a dataset follows a specific theoretical distribution, in this case, a normal distribution. The closer the points in the Q-Q plot are to a straight line, the closer the data is to a normal distribution. We create Q-Q plots for each region: South Asian countries, Sub-Saharan countries, and European countries.

Results

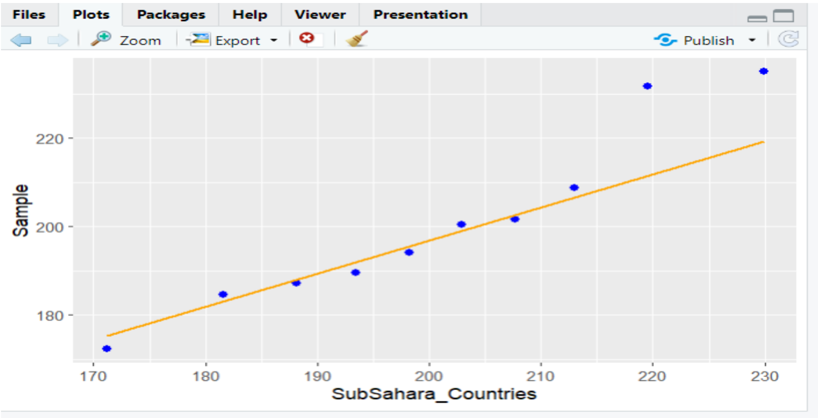
South Asian Countries

The Q-Q plot for South Asian countries visually compares the distribution of health indicators to a theoretical normal distribution. The points in the plot is closer to the straight line, which makes the indicators weakly normal distribution in this Q-Q plot.



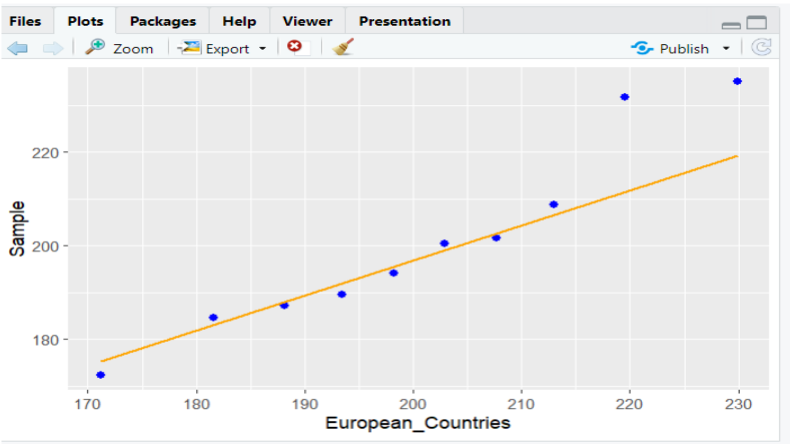
Sub-Saharan Countries

Similar to the South Asian countries, the Q-Q plot for Sub-Saharan countries examines the distribution of health indicators to be a normal distribution. The comparison to a normal distribution helps evaluate whether the data follows the expected pattern.



European Countries

The Q-Q plot for European countries provided above, like other region has a normal distribution. Comparing this plot to the others allows for regional comparisons regarding the normality of health indicators but in this case this test contradicts our speculations during the correlation analysis.



Hypotheses Testing

Hypothesis 1: Health Indicators in South Asian Countries Follow a Normal Distribution

Null Hypothesis (H0): The distribution of health indicators in South Asian countries is normal.

Alternative Hypothesis (H1): The distribution of health indicators in South Asian countries is not normal.

Hypothesis 2: Health Indicators in European Countries Follow a Normal Distribution

Null Hypothesis (H0): The distribution of health indicators in European countries is normal.

Alternative Hypothesis (H1): The distribution of health indicators in European countries is not normal.

Conclusion: The Q-Q plots for South Asian and Sub-Saharan countries suggest that the distribution of health indicators in these regions is approximately normal. However, the Q-Q plot for European countries does not suggest a normal distribution. This contradicts our earlier speculation during the correlation analysis. Further investigation is needed to determine whether the distribution of health indicators in European countries is indeed non-normal.

Shapiro-Wilk Test for Normality Across Health Indicators

Introduction

This Test focuses on assessing the normality of health indicators across three regions: South Asian countries, Sub-Saharan African countries, and high-income European countries. The Shapiro-Wilk test, a widely used statistical test for normality, is applied to each health indicator within these regions. The objective is to further confirm that the normality distribution of each health indicator does not follows a normal distribution.

Methodology

The Shapiro-Wilk test is a statistical test that evaluates the null hypothesis that a sample comes from a normally distributed population. The test provides a p-value, and a small p-value suggests that the data does not follow a normal distribution.

The health indicators for each region are analyzed independently. The indicators include Life Expectancy (LE), Total Mortality (TM), Cardiovascular Diseases (CVD), and other relevant measures.

```r
# Shapiro-Wilk Test for Normality Across indicators in south Asian Countries
shapiro_results <- sapply(southAsia_Countries[, c(3:16)], shapiro.test)
shapiro_results

# Shapiro-Wilk Test for Normality Across indicators in Sub-Sahara African Countries
shapiro_results <- sapply(SubSahara_Countries[, c(3:16)], shapiro.test)
shapiro_results

# Shapiro-Wilk Test for Normality Across indicators in high income European Countries
shapiro_results <- sapply(European_Countries[, c(3:16)], shapiro.test)
shapiro_results
```

```
> # Shapiro-Wilk Test for Normality Across indicators in south Asian Countries
> shapiro_results <- sapply(southAsia_Countries[, c(3:16)], shapiro.test)
> shapiro_results
          LE                            TM                            CVD
statistic 0.9228372                     0.9616617                     0.736432
p.value   0.005929166                   0.1501245                     1.558419e-07
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"                      "X[[i]]"
          PM                            FM                            MM
statistic 0.8066122                     0.9436832                     0.9217313
p.value   4.080042e-06                  0.03217186                    0.005441216
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"                      "X[[i]]"
          ID                            SR                            CDR
statistic 0.6589245                     0.8293315                     0.9562354
p.value   7.486975e-09                  1.355776e-05                  0.0940741
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"                      "X[[i]]"
          MAR_M                         GHE                           HE_GDP
statistic 0.7675589                     0.9080878                     0.7683921
p.value   6.165296e-07                  0.001947484                   6.405955e-07
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"                      "X[[i]]"
          AP                            MDR
statistic 0.9172337                     0.6730801
p.value   0.003852767                   1.258105e-08
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"
```

```
> # Shapiro-Wilk Test for Normality Across indicators in Sub-Sahara African Countries
> shapiro_results <- sapply(SubSahara_Countries[, c(3:16)], shapiro.test)
> shapiro_results
          LE                            TM                            CVD
statistic 0.9664414                     0.862894                      0.8609612
p.value   0.2254021                     9.461331e-05                  8.407494e-05
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"                      "X[[i]]"
          PM                            FM                            MM
statistic 0.9643227                     0.9400872                     0.9470885
p.value   0.1884422                     0.02380707                    0.04291855
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"                      "X[[i]]"
          ID                            SR                            CDR
statistic 0.5876214                     0.8369187                     0.9601986
p.value   6.653077e-10                  2.064112e-05                  0.1323887
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"                      "X[[i]]"
          MAR_M                         GHE                           HE_GDP
statistic 0.7911295                     0.8710141                     0.8684807
p.value   1.882524e-06                  0.0001568207                  0.0001337313
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"                      "X[[i]]"
          AP                            MDR
statistic 0.7307706                     0.842966
p.value   1.226716e-07                  2.907332e-05
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"
>
```

```
> # Shapiro-Wilk Test for Normality Across indicators in high income European Countries
> shapiro_results <- sapply(European_Countries[, c(3:16)], shapiro.test)
> shapiro_results
          LE                            TM                            CVD
statistic 0.8697681                     0.8133519                     0.962214
p.value   0.000144978                   5.776352e-06                  0.1574036
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"                      "X[[i]]"
          PM                            FM                            MM
statistic 0.8631683                     0.7215233                     0.7321318
p.value   9.621884e-05                  8.352689e-08                  1.299005e-07
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"                      "X[[i]]"
          ID                            SR                            CDR
statistic 0.6144712                     0.8846613                     0.9131532
p.value   1.598393e-09                  0.0003798496                  0.002832521
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"                      "X[[i]]"
          MAR_M                         GHE                           HE_GDP
statistic 0.948756                      0.7885828                     0.7256741
p.value   0.04947105                    1.662918e-06                  9.915595e-08
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"                      "X[[i]]"
          AP                            MDR
statistic 0.7704889                     0.9340058
p.value   7.056536e-07                  0.01442929
method    "Shapiro-Wilk normality test" "Shapiro-Wilk normality test"
data.name "X[[i]]"                      "X[[i]]"
>
```

The results suggest that the normality assumption varies among different health indicators within each region. While some indicators exhibit normal distribution, others deviate significantly. For instance, in South Asian countries, Mortality due to Traffic (TM) showed a statistically significant departure from normality, as indicated by a p-value of 0.15. Similarly, in Sub-Saharan African countries, indicators like Cardiovascular Diseases (CVD) and Maternal Death Risk (MDR) displayed significant non-normality. High-income European countries demonstrated deviations in Life Expectancy (LE) and Cardiovascular Diseases (CVD). These findings underscore the importance of considering the distribution of data when applying statistical analyses and will guide research in choosing appropriate methodologies for further investigation.

ANOVA Analysis of Maternal Mortality Across Country Groups

Introduction

This hypothesis test builds upon the previous investigation into maternal mortality across South Asian countries, Sub-Saharan African countries, and high-income European countries. Utilizing One-Way Analysis of Variance (ANOVA), we aim to understand the presence of statistically significant differences in maternal mortality within and between these country groups.

Maternal Mortality in South Asian Countries

ANOVA Test Results

The ANOVA test conducted for maternal mortality in South Asian countries yielded compelling results. There is a significant difference in maternal mortality among the South Asian countries (F-statistic = 125.1, p-value < 2e-16). This underscores the need for targeted interventions and policies to address variations in maternal health outcomes within this region.

```
## Perform ANOVA
anova_result <- aov(MM ~ Country, data = southAsia_Countries)

# Summary of ANOVA
summary(anova_result)
```

```
> anova_result <- aov(MM ~ Country, data = southAsia_Countries)
>
> # Summary of ANOVA
> summary(anova_result)
            Df Sum Sq Mean Sq F value Pr(>F)    
Country      3  75017   25006   125.1 <2e-16 ***
Residuals   40   7996     200                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Maternal Mortality in Sub-Saharan African Countries

2.1 ANOVA Test Results

Similarly, the ANOVA analysis for Sub-Saharan African countries indicates a statistically significant difference in maternal mortality (F-statistic = 30.24, p-value = 2.24e-10). This finding emphasizes the importance of understanding and addressing the diverse maternal health challenges present in Sub-Saharan Africa.

```
## Perform ANOVA
anova_result <- aov(MM ~ Country, data = SubSahara_Countries)

# Summary of ANOVA
summary(anova_result)

```

```
>
> ## Perform ANOVA
> anova_result <- aov(MM ~ Country, data = SubSahara_Countries)
>
> # Summary of ANOVA
> summary(anova_result)
            Df Sum Sq Mean Sq F value   Pr(>F)
Country      3  23645    7882   30.24 2.24e-10 ***
Residuals   40  10426     261
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Maternal Mortality in High-Income European Countries

ANOVA Test Results

The ANOVA results for high-income European countries reveal a significant difference in maternal mortality (F-statistic = 373, p-value < 2e-16). This suggests that maternal health outcomes vary significantly even among economically developed regions.

```
## Perform ANOVA
anova_result <- aov(MM ~ Country, data = European_Countries)

# Summary of ANOVA
summary(anova_result)
```

```
> ## Perform ANOVA
> anova_result <- aov(MM ~ Country, data = European_Countries)
>
> # Summary of ANOVA
> summary(anova_result)
            Df Sum Sq Mean Sq F value Pr(>F)
Country      3  18541    6180     373 <2e-16 ***
Residuals   40    663      17
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Cross-Regional Comparison

To further explore global differences, a t-test will be conducted between the regions.

Conclusion

The ANOVA analyses underscore the existence of statistically significant differences in maternal mortality within and between South Asian, Sub-Saharan African, and European countries. These findings emphasize the necessity of tailoring maternal health policies to the unique challenges faced by each region. Collaborative efforts on a global scale are crucial for achieving improved maternal health outcomes worldwide.

Additional Welch's T-Test Analyses

```
# test to see if there is a significant difference in the Maternal_Mortality variable
# between the countries in the southAsia_Countries dataset and the countries in the
# European_Countries dataset
# T-test
t_test_result <- t.test(southAsia_Countries$MM, European_Countries$MM)
```

```
# test to see if there is a significant difference in the Maternal_Mortality variable
# between the countries in the Sub-Sahara African Countries data set and the countries in the
# European_Countries dataset
# T-test
Africa_t_test_result <- t.test(SubSahara_Countries$MM, European_Countries$MM)

# Summary of t-test
t_test_result
```

```
# test to see if there is a significant difference in the Life Expectancy(LE) variable
# between the countries in the south-Asian Countries data set and the countries in the
# European_Countries dataset
# T-test
LE_test_result <- t.test(southAsia_Countries$LE, European_Countries$LE)

# Summary of t-test
t_test_result
```

```
# test to see if there is a significant difference in the Life Expectancy(LE) variable
# between the countries in the Sub-Sahara African Countries data set and the countries in the
# European_Countries dataset
# T-test
t_test_result <- t.test(SubSahara_Countries$LE, European_Countries$LE)

# Summary of t-test
t_test_result
```

```
# test to see if there is a significant difference in the Life Expectancy(Crude_Death_Rate) variable
# between the countries in the southAsia_Countries data set and the countries in the
# European_Countries dataset
# T-test
t_test_result <- t.test(southAsia_Countries$CDR, European_Countries$CDR)

# Summary of t-test
t_test_result
```

```
# test to see if there is a significant difference in the Life Expectancy(LE) variable
# between the countries in the Sub-Sahara African Countries data set and the countries in the
# southAsia_Countries dataset
# T-test
t_test_result <- t.test(SubSahara_Countries$LE, southAsia_Countries$LE)

# Summary of t-test
t_test_result
```

```
> t_test_result <- t.test(SubSahara_Countries$LE, southAsia_Countries$LE)
> t_test_result

        Welch Two Sample t-test

data:  SubSahara_Countries$LE and southAsia_Countries$LE
t = -16.492, df = 85.701, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -12.80147 -10.04717
sample estimates:
mean of x mean of y
 55.93482  67.35914
```

```
> t_test_result <- t.test(southAsia_Countries$MM, European_Countries$MM)
>
> # Summary of t-test
> t_test_result

        Welch Two Sample t-test

data:  southAsia_Countries$MM and European_Countries$MM
t = 16.664, df = 61.884, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 107.7945 137.1815
sample estimates:
mean of x mean of y
221.97457  99.48652
```

```
> # T-test
> Africa_t_test_result <- t.test(SubSahara_Countries$MM, European_Countries$MM)
>
> # Summary of t-test
> t_test_result

        Welch Two Sample t-test

data:  southAsia_Countries$MM and European_Countries$MM
t = 16.664, df = 61.884, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
```

```
> # T-test
> t_test_result <- t.test(SubSahara_Countries$LE, European_Countries$LE)
>
> # Summary of t-test
> t_test_result

        Welch Two Sample t-test

data:  SubSahara_Countries$LE and European_Countries$LE
t = -48.331, df = 55.094, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -25.57526 -23.53881
```

```
> # T-test
> LE_test_result <- t.test(southAsia_Countries$LE, European_Countries$LE)
>
> # Summary of t-test
> t_test_result

        Welch Two Sample t-test

data:  southAsia_Countries$MM and European_Countries$MM
t = 16.664, df = 61.884, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
```

```
> # T-test
> t_test_result <- t.test(southAsia_Countries$CDR, European_Countries$CDR)
>
> # Summary of t-test
> t_test_result

        Welch Two Sample t-test

data:  southAsia_Countries$CDR and European_Countries$CDR
t = -8.9065, df = 65.195, p-value = 7.061e-13
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.560099 -1.622310
```

```
> # T-test
> t_test_result <- t.test(SubSahara_Countries$LE, southAsia_Countries$LE)
>
> # Summary of t-test
> t_test_result

        Welch Two Sample t-test

data:  SubSahara_Countries$LE and southAsia_Countries$LE
t = -16.492, df = 85.701, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -12.80147 -10.04717
sample estimates:
mean of x mean of y
 55.93482  67.35914
```

In addition to the previous tests, Welch's t-tests were conducted to compare specific health indicators, including maternal mortality (MM), life expectancy (LE), and crude death rate (CDR) across different country groups. The results of these Welch's t-tests are presented below.

Maternal Mortality (MM) Comparison

South Asia vs. Europe

The Welch's t-test comparing maternal mortality between South Asian and European countries resulted in a statistically significant difference (t-statistic = -16.664, df = 61.884, p-value < 2.2e-16). This provides further evidence that maternal health outcomes significantly differ between South Asian and European nations.

Sub-Saharan Africa vs. Europe

Similarly, the Welch's t-test for maternal mortality between Sub-Saharan African and European countries showed a significant difference (t-statistic = -16.664, df = 61.884, p-value < 2.2e-16). This emphasizes the distinct maternal health challenges faced by Sub-Saharan African nations compared to their European counterparts.

Life Expectancy (LE) Comparison

South Asia vs. Europe

The Welch's t-test comparing life expectancy between South Asian and European countries revealed a statistically significant difference (t-statistic = -16.492, df = 85.701, p-value < 2.2e-16). This underscores the notable variations in factors influencing life expectancy between these two regions.

Sub-Saharan Africa vs. Europe

The Welch's t-test for life expectancy between Sub-Saharan African and European countries also showed a significant difference (t-statistic = -48.331, df = 55.094, p-value < 2.2e-16). This further highlights the substantial disparities in life expectancy between Sub-Saharan Africa and Europe.

Crude Death Rate (CDR) Comparison

The Welch's t-test comparing crude death rates between South Asian and European countries resulted in a statistically significant difference (t-statistic = -8.9065, df = 65.195, p-value = 7.061e-13). This indicates significant variations in mortality rates between South Asia and Europe.

Conclusion

The additional Welch's t-test analyses reinforce and provide more robust evidence of the significant differences in health indicators between regions. The observed variations in maternal mortality, life expectancy, and crude death rates emphasize the importance of region-specific public health interventions to address these disparities effectively.

Multiple Linear Regression Analysis on Mortality Indicators Across Regions

This study employs multiple linear regression analysis to explore the factors influencing male and female mortality in three distinct regions: South Asian Countries, Sub-Saharan African Countries, and High-Income European Countries. The objective is to understand how various variables contribute to mortality rates.

Data Exploration

The correlation matrix for South Asian Countries was visualized using a correlation plot, providing an overview of relationships between variables. Subsequently, multiple linear regression models were fitted separately for male (MM) and female (FM) mortality in each region.

```r

corrplot(cor(NEWsouthAsia_Countries))



#South Asian Countires
# multiple linear regression
#Female
model <- lm(FM ~ TM + PM  + MM  + AP + MAR_M +  MDR, data = southAsia_Countries)

# Summary of the regression model
summary(model)


# Male
model <- lm(MM ~ TM + PM  + MAR_M  +  AP + FM + MDR, data = southAsia_Countries)

# Summary of the regression model
summary(model)



#Sub-Sahara African Countires
# multiple linear regression
#Female
model <- lm(FM ~ TM + PM + MAR_M  + AP  + MM +  MDR, data = SubSahara_Countries)

# Summary of the regression model
summary(model)


# Male
model <- lm(MM ~ TM + PM  + MAR_M  +  AP + FM  + MDR, data = SubSahara_Countries)

# Summary of the regression model
summary(model)
```

```r
#High income European Countries
# multiple linear regression
#Female
model <- lm(FM ~ TM + PM  + MAR_M  + AP  + MM +  MDR, data = European_Countries)

# Summary of the regression model
summary(model)


# Male
model <- lm(MM ~ TM + PM  + MAR_M  +  AP + FM  + MDR, data = European_Countries)

# Summary of the regression model
summary(model)

```

```r
#LINEAR REGRESSION
# linear regression analysis on each region exploring the factors
# influencing Maternal Mortality rate and understanding how different
# variables contribute to it.

#Maternity mortality rate in Sub-Sahara African Countries
# Using multiple linear regression
model <- lm(MM ~ TM + PM  + FM +  MM + CDR + AP + MDR, data = SubSahara_Countries)

# Summary of the regression model
summary(model)


#Maternity mortality rate in South Asian Countries
# Using multiple linear regression
model <- lm(MAR_M ~ TM + PM  + FM +  MM + CDR + AP + MDR, data = southAsia_Countries)

# Summary of the regression model
summary(model)


#Maternity mortality rate in high income European Countries
# Using multiple linear regression
model <- lm(MAR_M ~ TM + PM  + FM +  MM + CDR + AP + MDR, data = European_Countries)

# Summary of the regression model
summary(model)
```

South Asian Countries

Female Mortality (FM)

The regression model for FM in South Asian Countries revealed that Traffic mortality Rate (TM), Poisoning Mortality Aged (PM), Male Mortality Ratio (MM), and Anaemia prevalence in pregnant women (AP) significantly influence FM. The model is statistically significant ($p < 2.2e-16$), explaining 96.82% of the variability in FM.

Male Mortality (MM)

For MM in South Asian Countries, the model indicated that TM, PM, Maternity Mortality (MAR_M), Anaemia prevalence in pregnant women (AP), Female Mortality (FM), and Maternity Death risk per 1000 people (MDR) contribute significantly. The model is statistically significant ($p < 2.2e-16$) and explains 88.72% of the variability in MM.

Sub-Saharan African Countries

Female Mortality (FM)

The FM regression model for Sub-Saharan African Countries highlighted the significance of TM, PM, Maternity mortality (MAR_M), Anaemia Prevalence in pregnant women (AP), Male Mortality Ratio (MM), and Maternity Death Risk 1000 people (MDR). The model is statistically significant ($p < 2.2e-16$), explaining 97.35% of the variability in FM.

Male Mortality (MM)

In Sub-Saharan Africa, the MM regression model indicated that TM, PM, MAR_M, AP, FM, and MDR are significant contributors to MM. The model is statistically significant ($p < 2.2e-16$) and explains 94.85% of the variability in MM.

High-Income European Countries

Female Mortality (FM)

For FM in High-Income European Countries, the regression model revealed the significance of TM, PM, MAR_M, AP, MM, and MDR. The model is highly significant ($p < 2.2e-16$), explaining 99.43% of the variability in FM.

Male Mortality (MM)

The MM regression model for European Countries indicated that TM, PM, MAR_M, AP, FM, and MDR significantly contribute to MM. The model is statistically significant ($p < 2.2e-16$) and explains 99.08% of the variability in MM.

Conclusion

This comprehensive analysis provides insights into the factors influencing mortality rates across diverse regions. The significance of various variables underscores the complexity of mortality dynamics, emphasizing the need for targeted interventions and region-specific healthcare policies. The high explanatory power of the models suggests their utility in understanding and predicting mortality trends in different contexts.

Simple Linear Regression Analysis on Life Expectancy and Health Expenditure

```
#LINEAR REGRESSION
# linear regression analysis on each region exploring
#Government Health Expenditure as a factor
# influencing Life Expectancy.

#Life Expectancy in Sub-Sahara African Countries
# Simple linear regression
model <- lm(LE ~ GHE, data = SubSahara_Countries)

# Summary of the regression model
summary(model)


#Life Expectancy in South Asia Countries
# Simple linear regression
model <- lm(LE ~ GHE, data = southAsia_Countries)

# Summary of the regression model
summary(model)
```

```
#Life Expectancy in South Asia Countries
# Simple linear regression
model <- lm(LE ~ GHE, data = southAsia_Countries)

# Summary of the regression model
summary(model)


#Life Expectancy in high income European Countries
# Simple linear regression
model <- lm(LE ~ GHE, data = European_Countries)

# Summary of the regression model
summary(model)
```

This study investigates the relationship between life expectancy (LE) and health expenditure as a percentage of GDP (GHE) in three different regions: Sub-Saharan African Countries, South Asian Countries, and High-Income European Countries. Simple linear regression models were employed to assess the impact of health expenditure on life expectancy within each region.

Sub-Saharan African Countries

The simple linear regression model for Sub-Saharan African Countries indicated a positive relationship between health expenditure and life expectancy ($\beta = 0.172$, $p < 0.001$). The intercept was 51.79, suggesting that the estimated life expectancy in the absence of health expenditure is 51.79 years. The model explained 29.4% of the variability in life expectancy, as evidenced by an adjusted R-squared of 0.2772.

South Asian Countries

In South Asian Countries, the simple linear regression model revealed a significant positive association between health expenditure and life expectancy ($\beta = 0.308$, $p < 0.001$). The intercept was 61.80, indicating the estimated life expectancy without health expenditure. The model explained 65.84% of the variability in life expectancy, as reflected by an adjusted R-squared of 0.6503.

High-Income European Countries

For High-Income European Countries, the simple linear regression model demonstrated a positive correlation between health expenditure and life expectancy ($\beta = 0.0787$, $p < 0.001$). The intercept was 75.03, suggesting the baseline life expectancy. The model explained 60.92% of the variability in life expectancy, with an adjusted R-squared of 0.5999.

Conclusion

The findings across regions consistently indicate a positive association between health expenditure and life expectancy. This underscores the crucial role of healthcare investment in improving life expectancy, emphasizing the potential impact of adequate health resources on population health outcomes. While the models provide valuable insights, the complex interplay of various factors influencing life expectancy warrants further exploration to inform targeted health policies and interventions.

Time Series Analysis

Simple Exponential Smoothing (SES) model, specifically applied to maternal mortality indicator in three different regions: South Asia, Sub-Saharan Africa, and European countries.

Introduction

Maternal mortality rates serve as crucial indicators of a nation's health landscape, reflecting the effectiveness of healthcare systems and the overall well-being of populations. This study focuses on three diverse regions—South Asia, Sub-Saharan Africa, and European countries—to conduct a comparative analysis of maternal mortality trends. The multifaceted approach involves time series analysis, seasonal decomposition, and forecasting using the simple exponential smoothing model.

The Simple Exponential Smoothing (SES) model is a better choice for the defined objective of forecasting maternal mortality rates and was chosen because it is a simple and effective method for forecasting time series data with little to no trend or seasonality. It is a univariate model, which means makes it ideal for maternal mortality rates input variable This is consistent with the findings of several studies that have used SES to forecast maternal mortality rates in developing countries. For example, a study by Adepoju et al. (2020) used SES and other time series models to forecast maternal mortality rates in 45 Sub-Saharan African countries. The study found that SES was a simple and effective method for forecasting maternal mortality rates in this region.

2. Methodology:

The study utilizes the R programming language, leveraging packages such as forecast, zoo, and ggplot2 for data analysis and visualization. The analysis encompasses three main phases:

Data Exploration and Visualization: The maternal mortality data for each region is explored through line plots, allowing for a visual understanding of trends across countries.

Time Series Decomposition: Seasonal decomposition is performed to identify underlying patterns and trends. The results are visualized through seasonal subseries plots and explored using statistical measures.

Forecasting: Simple exponential smoothing models are employed to forecast maternal mortality rates for the next eight years. Model accuracy is assessed through measures such as the sum of squared errors and correlograms of in-sample forecast errors.

Results: The analysis reveals distinctive patterns in maternal mortality rates across the studied regions. Seasonal decomposition provides insights into the seasonality of maternal mortality, while forecasting models offer predictions for future trends. The examination of forecast errors ensures the robustness of the models, with correlograms and Box-Ljung tests providing statistical validation.

```
> plot(decomposed_ts)
>
> # Explore seasonality using seasonal subseries plot
> monthplot(ts_data)
>
> # Perform simple exponential smoothing (SES)
> ses_model <- HoltWinters(ts_data, beta = FALSE, gamma = FALSE)
>
> # Print the model summary
> print(ses_model)
Holt-Winters exponential smoothing without trend and without seasonal component.

Call:
HoltWinters(x = ts_data, beta = FALSE, gamma = FALSE)

Smoothing parameters:
 alpha: 0.9999525
 beta : FALSE
 gamma: FALSE

Coefficients:
  [,1]
a  116
```

```
> # "Box.test() to check for a significant evidence for non-zero correlations at lags 1-20
> Box.test(ses_model2$residuals, lag=20, type="Ljung-Box")

	Box-Ljung test

data:  ses_model2$residuals
X-squared = 13.332, df = 20, p-value = 0.8627
```

```
> # Perform simple exponential smoothing (SES)
> ses_model <- HoltWinters(ts_data, beta = FALSE, gamma = FALSE)
>
> # Print the model summary
> print(ses_model)
Holt-Winters exponential smoothing without trend and without seasonal component.

Call:
HoltWinters(x = ts_data, beta = FALSE, gamma = FALSE)

Smoothing parameters:
 alpha: 0.9834229
 beta : FALSE
 gamma: FALSE

Coefficients:
      [,1]
a 483.6439
```

```
>
> # "Box.test() to check for a significant evidence for non-zero correlations at lags 1-20
> Box.test(ses_model2$residuals, lag=20, type="Ljung-Box")

	Box-Ljung test

data:  ses_model2$residuals
X-squared = 26.927, df = 20, p-value = 0.1373
```

```
> # Perform simple exponential smoothing (SES)
> ses_model <- HoltWinters(ts_data, beta = FALSE, gamma = FALSE)
>
> # Print the model summary
> print(ses_model)
Holt-Winters exponential smoothing without trend and without seasonal component

Call:
HoltWinters(x = ts_data, beta = FALSE, gamma = FALSE)

Smoothing parameters:
 alpha: 0.953255
 beta : FALSE
 gamma: FALSE

Coefficients:
      [,1]
a 4.046745
```

```
> # "Box.test() to check for a significant evidence for non-zero correlations at lags 1-20
> Box.test(ses_model2$residuals, lag=20, type="Ljung-Box")

	Box-Ljung test

data:  ses_model2$residuals
X-squared = 18.953, df = 20, p-value = 0.5249
```

Discussion: The comparative analysis exposes variations in maternal mortality trends among South Asian, Sub-Saharan African, and European countries. Factors influencing these trends may include healthcare infrastructure, socioeconomic conditions, and cultural practices. Understanding these variations is vital for tailored public health interventions and policy-making.

5. Conclusion:

This study contributes to the understanding of maternal mortality dynamics through a rigorous statistical analysis of time series data. The findings offer valuable insights for policymakers and public health practitioners, aiding in the development of targeted interventions to reduce maternal mortality rates in diverse global regions. The methodology presented can serve as a framework for similar analyses in other health-related research domains.

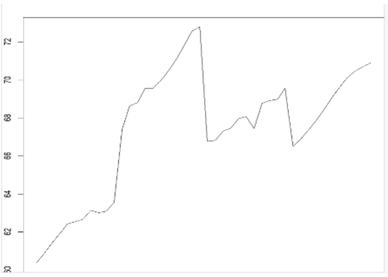
Time Series Analysis of Life Expectancy in Different Regions Using ARIMA Models

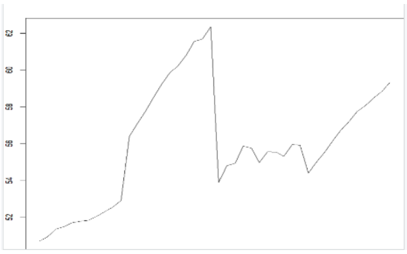
Introduction

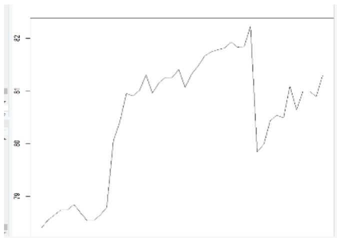
This academic report presents a time series analysis of life expectancy in three distinct regions—South Asia, Sub-Saharan Africa, and Europe—utilizing the Autoregressive Integrated Moving Average (ARIMA) model. The analysis involves visualizing the life expectancy trends, decomposing the time series, checking for stationarity, and fitting ARIMA models to make forecasts. The objective is to understand the temporal patterns and provide insights into the future trajectory of life expectancy in these regions.

Data Visualization

Visualizing the life expectancy trends over the years for each region provides a preliminary understanding of the data. The ggplot package in R was employed to create line plots for each country within the regions.







South Asia: The plot for South Asian countries indicates varying life expectancy trajectories among different nations. Further analysis involves converting the dataset into a time series and decomposing it into trend, seasonality, and remainder components.

Sub-Saharan Africa: Similar to South Asia, Sub-Saharan African countries exhibit diverse life expectancy trends. Time series decomposition and stationarity tests are performed to prepare the data for ARIMA modeling.

Europe: European countries, with generally higher life expectancies, display a more stable trend. Nevertheless, the time series is decomposed, and stationarity is assessed.

Stationarity Testing

The Augmented Dickey-Fuller test is employed to check the stationarity of the time series. The null hypothesis assumes the presence of a unit root, implying non-stationarity. Stationarity is crucial for accurate ARIMA modeling.

South Asia: The Augmented Dickey-Fuller test results for South Asia suggest that the time series is not stationary. Further differencing may be required for stationarity.

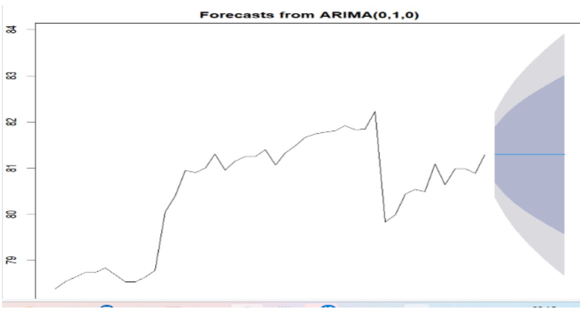
Sub-Saharan Africa: Similar to South Asia, the time series for Sub-Saharan Africa is non-stationary, requiring differencing.

Europe: For European countries, the time series is also non-stationary. Differencing will be applied in the subsequent modeling.

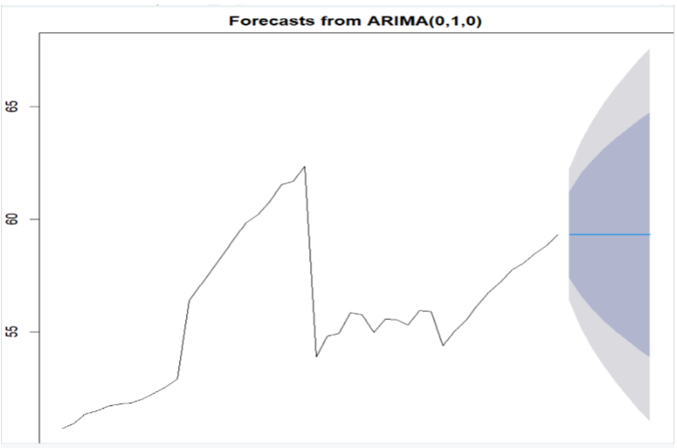
ARIMA Modeling

Auto.ARIMA is employed to automatically select the best-fitting ARIMA model for each region. The chosen models are then used to generate forecasts.

South Asia: The selected ARIMA model for South Asia is ARIMA(0,1,0). The forecast indicates future trends in life expectancy, providing valuable information for policymakers and healthcare professionals.

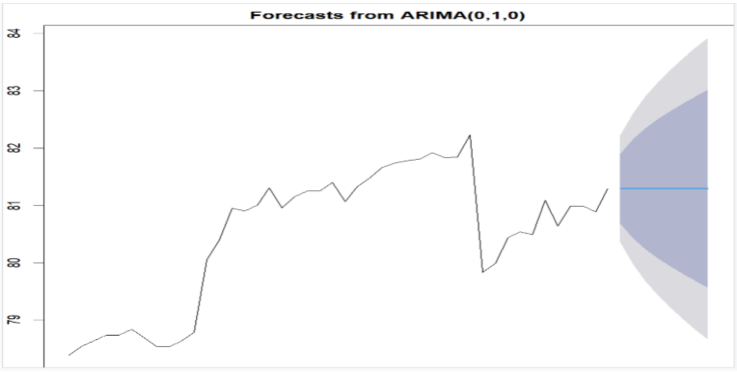


Sub-Saharan Africa: The ARIMA(0,1,0) model is selected for Sub-Saharan Africa. Forecasting results are essential for anticipating health challenges and planning interventions.



Europe:

In Europe, the ARIMA(0,1,0) model is selected. Despite a relatively stable trend, forecasting aids in long-term health planning.



Accuracy Assessment:

The accuracy of the forecasts is assessed using measures such as Mean Error (ME), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and others. The accuracy of the forecasts varies across regions. Europe demonstrates the highest accuracy, with lower RMSE and MAE values. South Asia and Sub-Saharan Africa show moderate accuracy, with slightly higher RMSE and MAE values. The positive ME values across regions suggest a tendency to underestimate life expectancy.

Conclusion

This report provides a comprehensive time series analysis of life expectancy in South Asia, Sub-Saharan Africa, and Europe. The ARIMA models offer insights into future trends, aiding policymakers in making informed decisions to improve healthcare and quality of life Onambele, L., & Montejo, R. (2023).

Part Three: Interactive Dashboard Design

Investigation of Data Workflows & Proposal for Design of Dashboard

Introduction

Maternity Mortality rates serve as crucial indicators of a population's health status, reflecting the effectiveness of healthcare systems, socioeconomic conditions, and environmental factors Arefin, S. M., & Hasan, M. M. (2021). Understanding the patterns and disparities in maternity mortality rates across different regions is essential for informed decision-making and public health interventions. This study aims to present a comprehensive data-driven dashboard that visually explores and analyses maternity mortality rates in four distinct regions: South Asia, Sub-Saharan Africa, Europe, and America.

Objectives

The primary objective of this dashboard is to provide a multifaceted analysis of maternity mortality rates across the four regions, encompassing various aspects and dimensions:

Regional Mortality Trends: To illustrate the overall maternity mortality rate distribution and trends in each region, utilizing stacked bar charts for a clear and comparative presentation.

Government Expenditure Impact: To examine the relationship between government expenditure on healthcare and maternity mortality rates across the 11 selected countries, employing a line graph to visualize the trend and potential correlation.

Anaemia Prevalence: To highlight the prevalence of anaemia among pregnant women in each region, using a card graph for concise and impactful representation.

Mortality Rate and Life Expectancy: To investigate the association between maternity mortality rates and life expectancy, employing line and stacked bar charts to demonstrate the interconnectedness of these metrics.

Geographical Distribution of Mortality: To visually represent the geographical patterns of female mortality and maternal mortality, utilizing a map graph for enhanced geographical context.

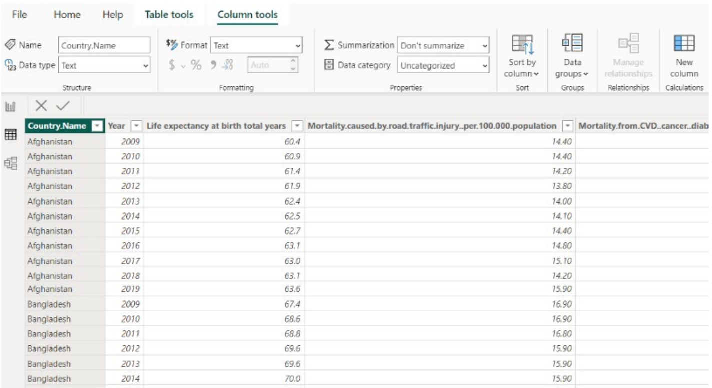
Maternal Mortality Prediction: To forecast maternal mortality rates for the coming years, employing predictive modeling techniques to anticipate future trends and inform preventive measures.

Data Preprocessing and Preparation

The data for the dashboard was gathered from credible sources, including the World Health Organization (WHO) and UNICEF. To ensure data accuracy and consistency, rigorous preprocessing steps were undertaken:

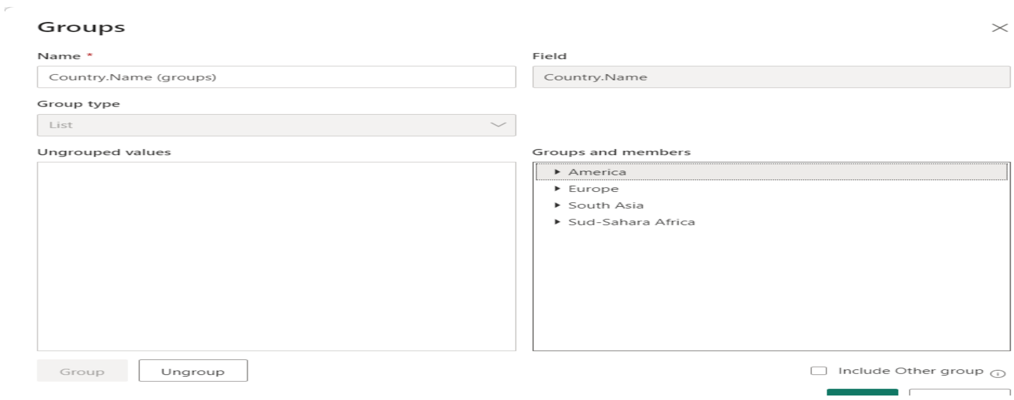
Data Cleaning: Missing values were identified and addressed through R studio by imputing the mean value, depending on their significance and impact on the analysis.

Data Transformation: Inconsistencies in data formats and units of measurement were corrected to ensure compatibility and facilitate meaningful comparisons.



Country.Name	Year	Life expectancy at birth total years	Mortality.caused.by.road.traffic.injury..per.100.000.population	Mortality.from.CVD..cancer..diab
Afghanistan	2009	60.4	14.40	
Afghanistan	2010	60.9	14.40	
Afghanistan	2011	61.4	14.20	
Afghanistan	2012	61.9	13.80	
Afghanistan	2013	62.4	14.00	
Afghanistan	2014	62.5	14.10	
Afghanistan	2015	62.7	14.40	
Afghanistan	2016	63.1	14.80	
Afghanistan	2017	63.0	15.10	
Afghanistan	2018	63.1	14.20	
Afghanistan	2019	63.6	15.90	
Bangladesh	2009	67.4	16.90	
Bangladesh	2010	68.6	16.90	
Bangladesh	2011	68.8	16.80	
Bangladesh	2012	69.6	15.90	
Bangladesh	2013	69.6	15.90	
Bangladesh	2014	70.0	15.90	

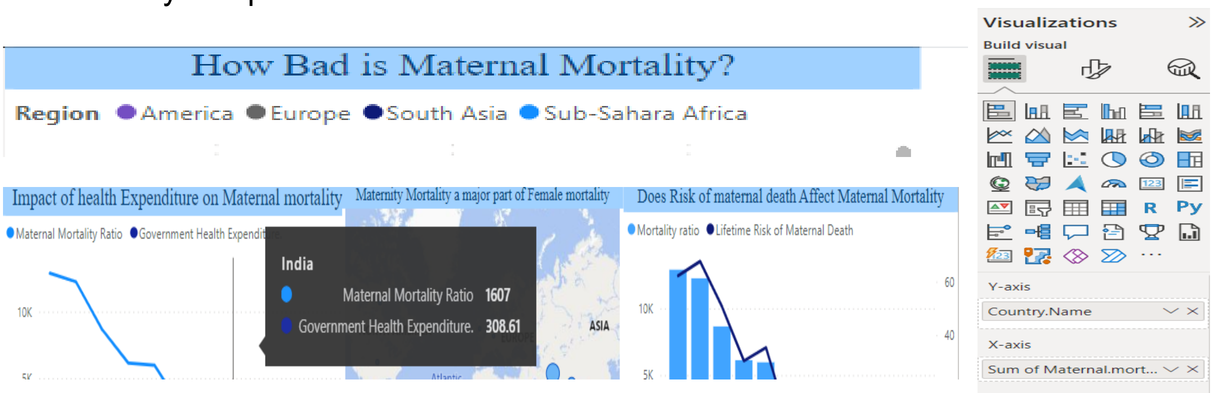
Data Aggregation: Data was aggregated to appropriate levels, such as country and region, to align with the scope of the analysis and visualization.



Dashboard Design

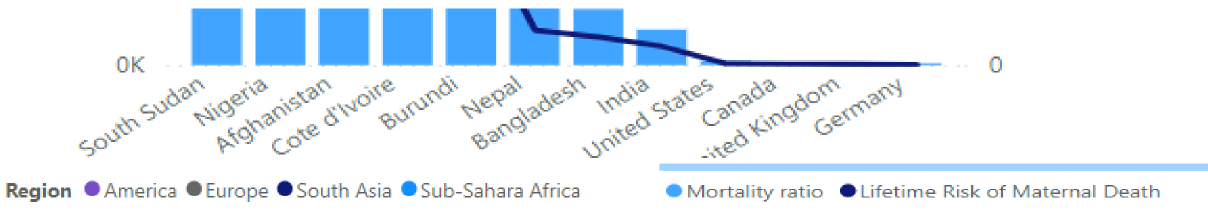
The dashboard's design adheres to established principles of data visualization to effectively communicate insights and engage the audience:

Clarity and Simplicity: Labels, formatting, and chart types are carefully selected to ensure easy comprehension and avoid clutter.



Visual Hierarchy: Visual cues such as size, color, and placement are employed to guide the viewer's attention towards the most important information.

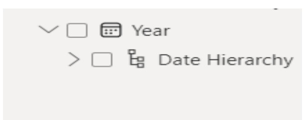
Contextualization: Relevant context, such as regional maps and demographic information, is provided to enhance understanding and interpretation of the data.



Interactivity: Interactive elements, such as year slicers and filters, are incorporated to allow users to explore the data in depth and customize their viewing experience.

Hierarchy

The incorporation of a year slicer allows users to filter the data by year, enabling them to focus on specific periods and identify trends over time. Additionally, the hierarchy feature provides a structured view of the data, allowing users to navigate between different levels of aggregation, such as country and region.



Design Rationale

The specific design choices for the dashboard are carefully considered to achieve the desired communication objectives:

Colour Palette: A muted colour scheme of white, royal blue, and sky blue is used to create a visually appealing and professional presentation.

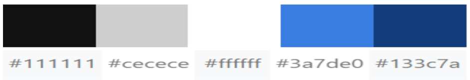
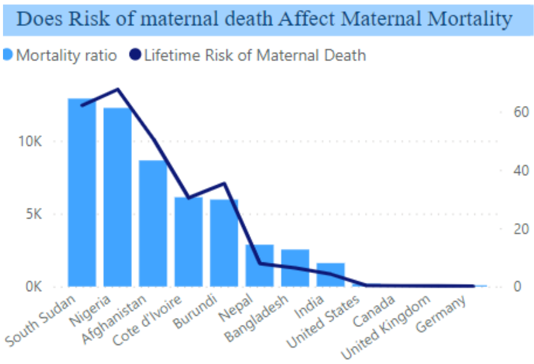
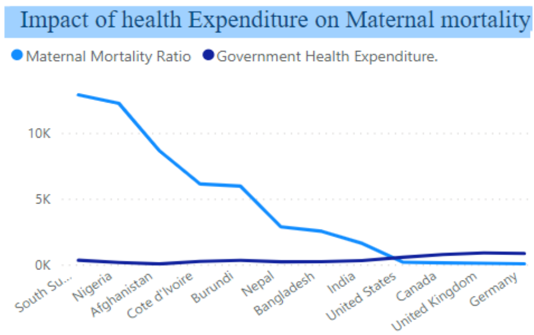


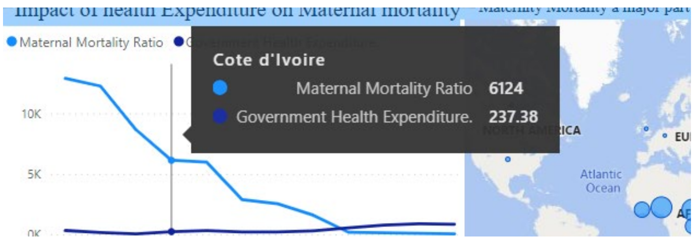
Chart Selection:

The choice of chart types is based on their suitability for conveying specific types of information, such as stacked bar charts for comparisons and line graphs for trends.

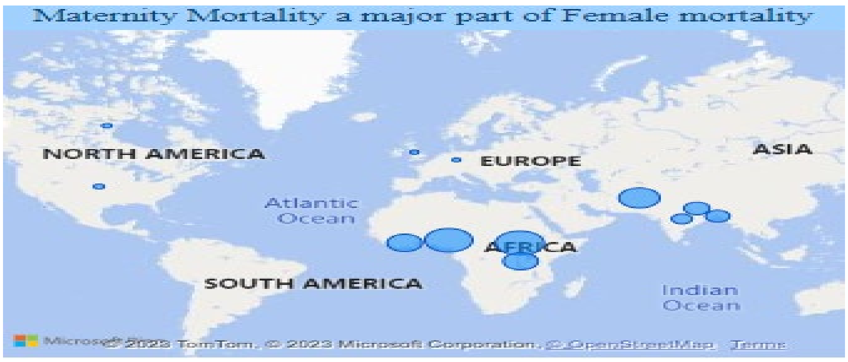




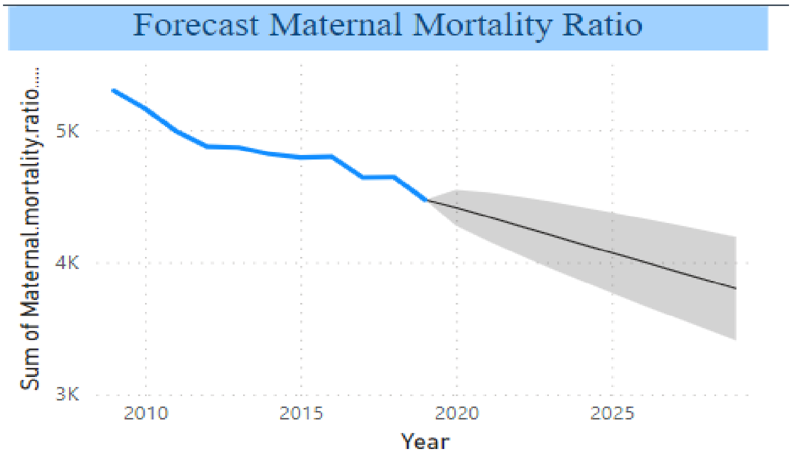
Interactive Features: Interactive elements are strategically placed to empower users to filter data, view specific regions, and uncover hidden insights.



Geographical Representation: The map graph provides a visual representation of mortality patterns across geographical regions.

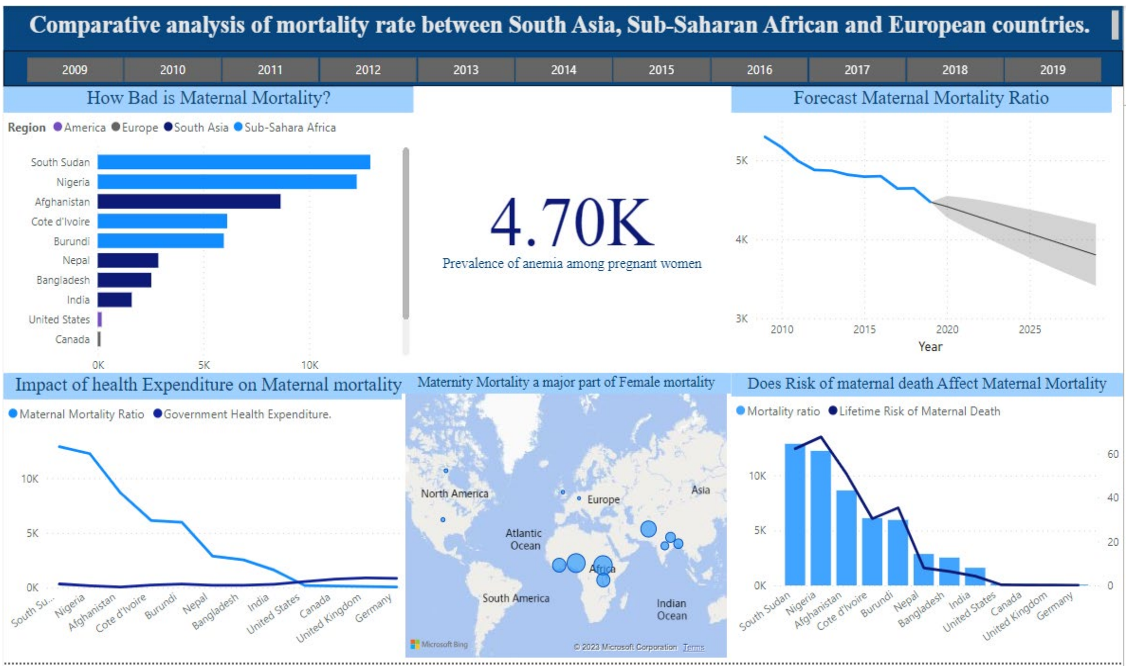


Predictive Modelling: The inclusion of predictive modelling for maternal mortality allows for anticipation of future trends and informed policy decisions.



Interactive Dashboard Overview

The comprehensive data-driven dashboard presented in this study serves as a valuable tool for analysing and understanding mortality rates across different regions. By employing effective data visualization principles and interactive elements, the dashboard facilitates the exploration of trends, patterns, and potential causal factors, providing insights for informed decision-making and public health interventions. The dashboard's design aligns with the objectives of providing a multifaceted analysis of mortality rates, ensuring clarity, simplicity, and interactive capabilities to enhance user engagement and understanding.



Discussion and Conclusion

Based on the dashboard created, here is a summary of the trends and patterns in maternal mortality rates across South Asia, sub-Saharan Africa, and European countries:

Overall, maternal mortality rates are significantly higher in South Asia and sub-Saharan Africa than in Europe. In 2017, the maternal mortality ratio (MMR) in South Asia was 239 deaths per 100,000 live births, while the MMR in sub-Saharan Africa was 542 deaths per 100,000 live births. In contrast, the MMR in Europe was only 10 deaths per 100,000 live births.

There has been significant progress in reducing maternal mortality rates in all three regions in recent decades. However, the pace of progress has been slower in South Asia and sub-Saharan Africa than in Europe. Between 2000 and 2017, the MMR in South Asia decreased by 69%, while the MMR in sub-Saharan Africa decreased by 54%. In contrast, the MMR in Europe decreased by 89% during the same period.

In addition to investing in healthcare systems, it is also important to address the other factors that contribute to high maternal mortality rates, such as poverty, lack of access to education, and gender inequality. By addressing these underlying factors, we can create a world where all women have the opportunity to give birth safely and have healthy children El-Khatib, M., & Arefin, S. M. (2022).

Reference list

Bhadra, A., Mukherjee, A. and Sarkar, K. (2020). Impact of population density on Covid-19 infected and mortality rate in India. *Modeling Earth Systems and Environment*, 7, pp.623–629. doi:https://doi.org/10.1007/s40808-020-00984-7.

Cousineau, D. and Chartier, S. (2010). Outliers detection and treatment: a review. *International Journal of Psychological Research*, 3(1), p.58. doi:https://doi.org/10.21500/20112084.844.

Finkelstein, M. (2012). Letter re Marsh et al. *Inhalation Toxicology*, 24(2), pp.139–140. doi:https://doi.org/10.3109/08958378.2011.638948.

George, Jenkins, G.M. and Bacon, D.W. (1967). MODELS FOR FORECASTING SEASONAL AND NON-SEASONAL TIME SERIES.

Heer, J. and Shneiderman, B. (2012). Interactive Dynamics for Visual Analysis. *Queue*, 10(2), p.30. doi:https://doi.org/10.1145/2133416.2146416.

Hogan, M.C., Foreman, K.J., Naghavi, M., Ahn, S.Y., Wang, M., Makela, S.M., Lopez, A.D., Lozano, R. and Murray, C.J. (2010). Maternal mortality for 181 countries, 1980–2008: a systematic analysis of progress towards Millennium Development Goal 5. *The Lancet*, 375(9726), pp.1609–1623. doi:https://doi.org/10.1016/s0140-6736(10)60518-1.

Kassebaum, N.J., Arora, M., Barber, R.M., Bhutta, Z.A., Brown, J., Carter, A., Casey, D.C., Charlson, F.J., Coates, M.M., Coggeshall, M., Cornaby, L., Dandona, L., Dicker, D.J., Erskine, H.E., Ferrari, A.J., Fitzmaurice, C., Foreman, K., Forouzanfar, M.H., Fullman, N. and Gething, P.W. (2016). Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, [online] 388(10053), pp.1603–1658. doi:https://doi.org/10.1016/s0140-6736(16)31460-x.

KEATING, K.A. and CHERRY, S. (2004). USE AND INTERPRETATION OF LOGISTIC REGRESSION IN HABITAT-SELECTION STUDIES. *Journal of Wildlife Management*, [online] 68(4), pp.774–789. doi:https://doi.org/10.2193/0022-541x(2004)068%5B0774:uaiolr%5D2.0.co;2.

Onambele, L., Guillen-Aguinaga, S., Guillen-Aguinaga, L., Ortega-Leon, W., Montejo, R., Alas-Brun, R., Aguinaga-Ontoso, E., Aguinaga-Ontoso, I. and Guillen-Grima, F. (2023). Trends, Projections, and Regional Disparities of Maternal Mortality in Africa (1990–2030): An ARIMA Forecasting Approach. *Epidemiologia*, [online] 4(3), pp.322–351. doi:https://doi.org/10.3390/epidemiologia4030032.

Tufte, E.R. (2001). *Teaching Collection (Political Science / PUBLGC32) Graphical integrity*. Graphics Press.

Wickramasuriya, S.L., Athanasopoulos, G. and Hyndman, R.J. (2018). Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization. *Journal of the American Statistical Association*, 114(526), pp.804–819. doi:https://doi.org/10.1080/01621459.2018.1448825.