R_Final_Project_Abed_Turaki

Laila and Nargis

2024-12-02

Contents

List of Figures

Introduction:

This project focuses on analyzing mortality trends and the influencing factors in Nigeria, with a particular emphasis on maternal mortality and its correlation with various socio-economic indicators. Using data spanning several decades, we explore the relationship between public health expenditure and maternal mortality rates, the impact of health-related factors such as fertility rates and life expectancy, and how economic growth and demographic trends have influenced public health outcomes. By applying spatial data exploration techniques, time series analysis, and regression models, this project aims to provide a comprehensive understanding of the dynamics driving maternal health outcomes in Nigeria and to highlight areas for potential intervention and improvement.

Question 1: How did the population growth rate evolve over time in the dataset, and what factors might have contributed to these changes?

```r
library(ggplot2)

# Convert Year to numeric for plotting
dataset2$Year <- as.numeric(dataset2$Year)

# Create the plot
ggplot(dataset2, aes(x = Year)) +
  geom_line(aes(y = Population_male, color = "Male"), linewidth = 1) +
  geom_line(aes(y = Population_female, color = "Female"), linewidth = 1) +
  scale_color_manual(values = c("Male" = "blue", "Female" = "pink")) +
  labs(
    title = "Population of Males and Females Over the Years",
    x = "Year",
    y = "Population",
    color = "Gender"
  ) +
  theme_minimal()
```

```r
#Life expectancy
ggplot(dataset2, aes(x = Year)) +
  geom_line(aes(y = Life_expectancy_atbirth_male, color = "Male"), linewidth = 1) +
  geom_line(aes(y = Life_expectancy_at_birth_female, color = "Female"), linewidth = 1) +
  scale_color_manual(values = c("Male" = "blue", "Female" = "pink")) +
  labs(
    title = "Life Expectancy Of Males and Females Over the Years",
    x = "Year",
    y = "Life Expectancy",
    color = "Gender"
  ) +
  theme_minimal()
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (‘geom_line()‘).
## Removed 1 row containing missing values or values outside the scale range
## (‘geom_line()‘).
```

```r
# Survival to age 65
ggplot(dataset2, aes(x = Year)) +
  geom_line(aes(y = Survival_to_age_65_male, color = "Male"), linewidth = 1) +
  geom_line(aes(y = Survival_to_age_65_female, color = "Female"), linewidth = 1) +
  scale_color_manual(values = c("Male" = "blue", "Female" = "pink")) +
  labs(
    title = "Survival To Age 65 Males and Females Over the Years",
    x = "Year",
    y = "Survival to Age 65",
```

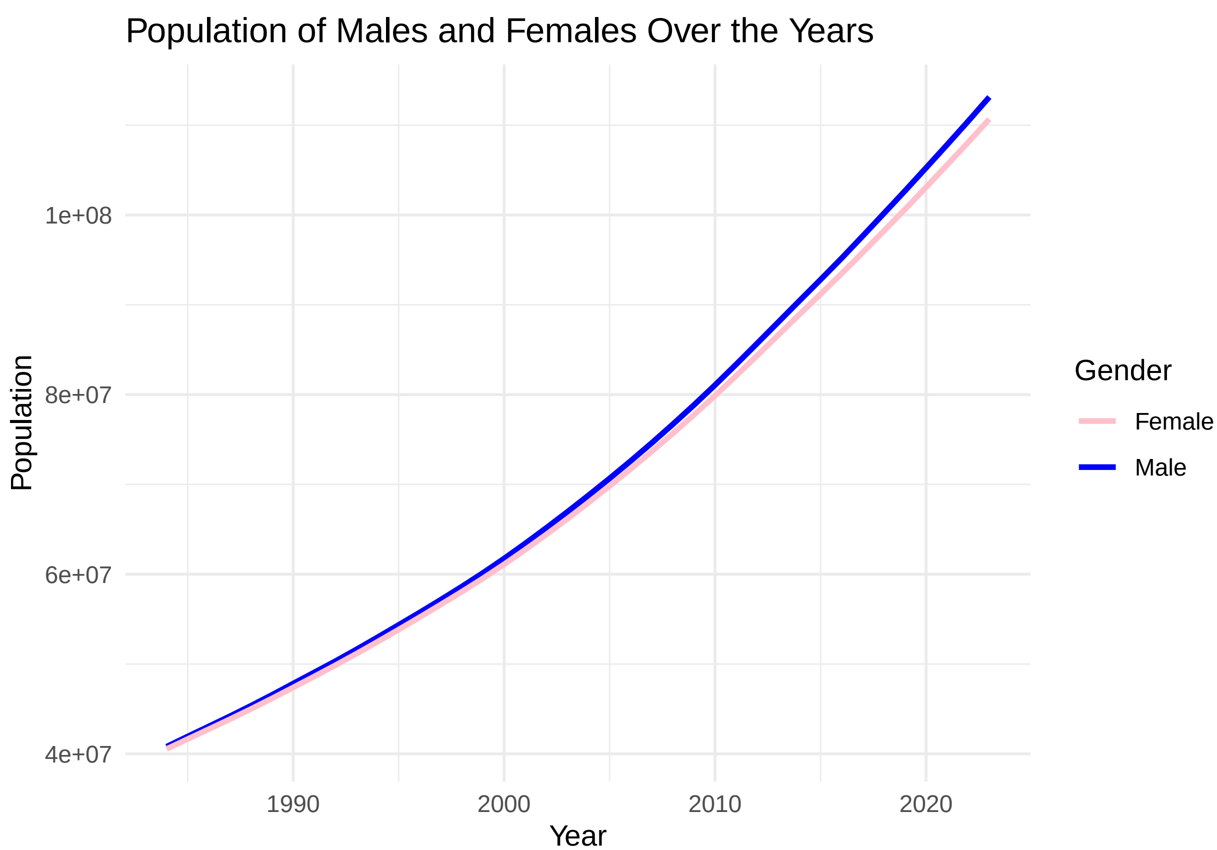

Figure 1: Population of Males and Females Over the Years

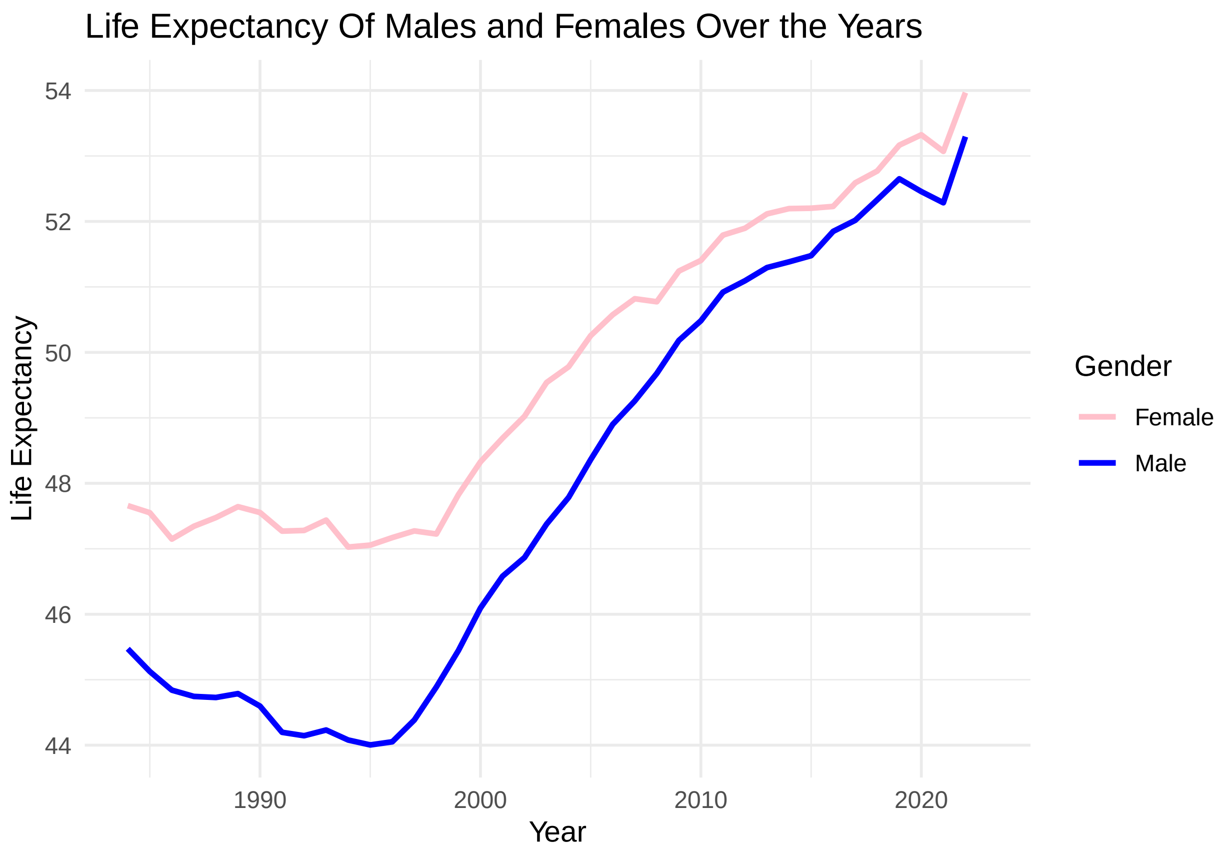

Figure 2: Population of Males and Females Over the Years

```r
    color = "Gender"
  ) +
  theme_minimal()
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_line()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_line()`).
```

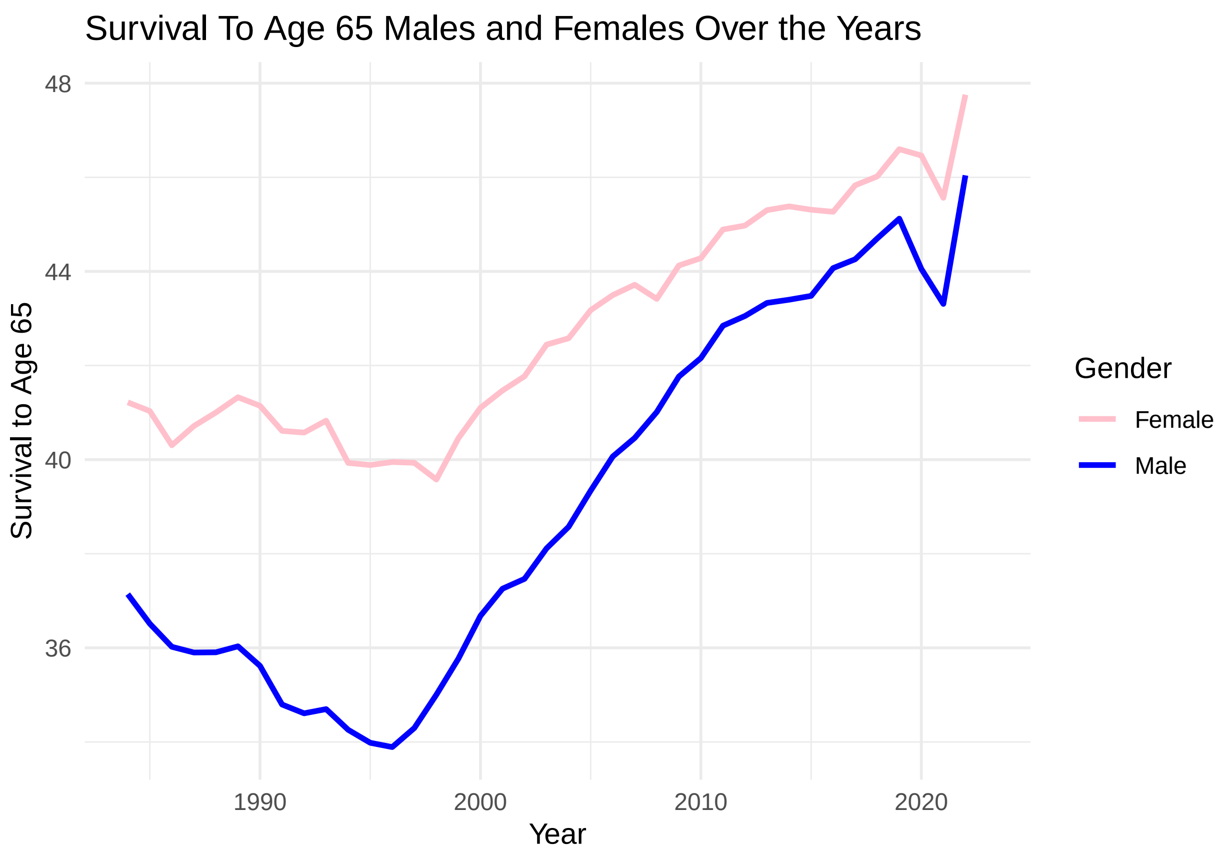

Figure 3: Population of Males and Females Over the Years

```r
# Mortality rate
ggplot(dataset2, aes(x = Year)) +
  geom_line(aes(y = Mortality_rate_adult_male, color = "Male"), linewidth = 1) +
  geom_line(aes(y = Mortality_rate_adult_female, color = "Female"), linewidth = 1) +
  scale_color_manual(values = c("Male" = "blue", "Female" = "pink")) +
  labs(
    title = "Mortality Rate Of Adult Males and Females Over the Years",
    x = "Year",
    y = "Mortality Rate",
    color = "Gender"
  ) +
  theme_minimal()
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_line()`).
## Removed 1 row containing missing values or values outside the scale range
## (`geom_line()`).
```

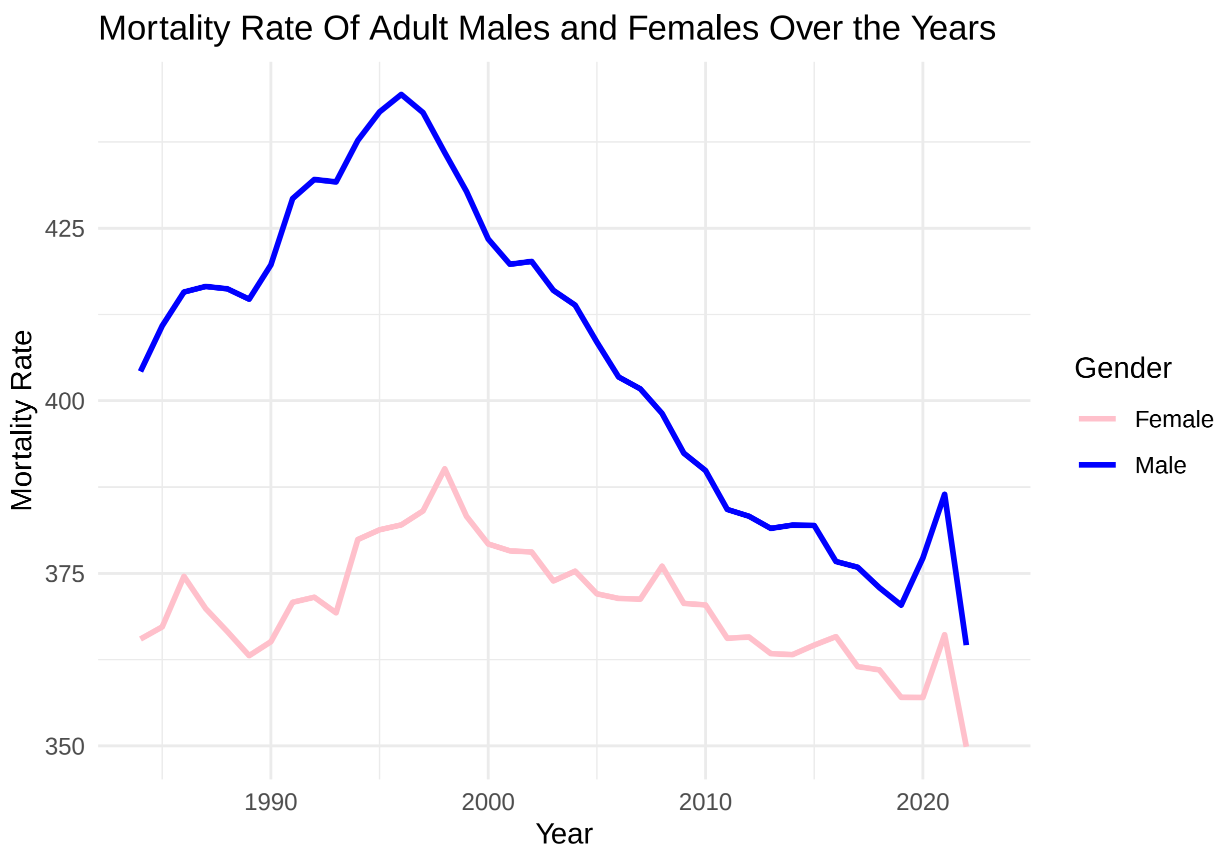

Figure 4: Population of Males and Females Over the Years

Analysis1:Population Growth Rate Evolution

The population growth in Nigeria shows distinct trends for males and females over time. Both male and female populations have been steadily increasing from 1990 to 2020. The male population (represented by the blue line) consistently remains slightly higher than the female population (depicted by the pink line). The growth appears to be exponential, with the rate of increase accelerating over time. Key factors contributing to these changes include high fertility rates, improved healthcare leading to lower infant mortality, better life expectancy, and possibly changes in migration patterns. Additionally, we noticed from the graphs that the life expectancy for females is higher than males. However the gap between them is decreasing along the years. The Mortality rate for males is higher than females, and the same observation the gap is decreasing along the years. Survivals to age 65 years of females are more than males and again the gap is decreasing along the years.

Question2: What trends were observed in the mortality rate ratio over time, and how might this be indicative of broader health and socioeconomic conditions in the country?

```r
# Calculating maternal mortality ratio (MMR) over time in Nigeria
# Load necessary library
library(ggplot2)

# Data Preparation (assuming your dataset is called 'dataset2')
dataset2$Year <- as.numeric(dataset2$Year)

# Filter the data to remove rows with missing values
dataset2_clean <- dataset2 %>%
  select(Year, Maternal_mortality_ratio) %>%
  filter(!is.na(Maternal_mortality_ratio) & !is.na(Year))

# Plotting Maternal Mortality Ratio Over Time
ggplot(dataset2_clean, aes(x = Year, y = Maternal_mortality_ratio)) +
  geom_line(color = "blue") +
  geom_point(color = "red") +  # Add points to show data points clearly
  labs(title = "Maternal Mortality Ratio Over Time in Nigeria",
       x = "Year",
       y = "Maternal Mortality Ratio") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

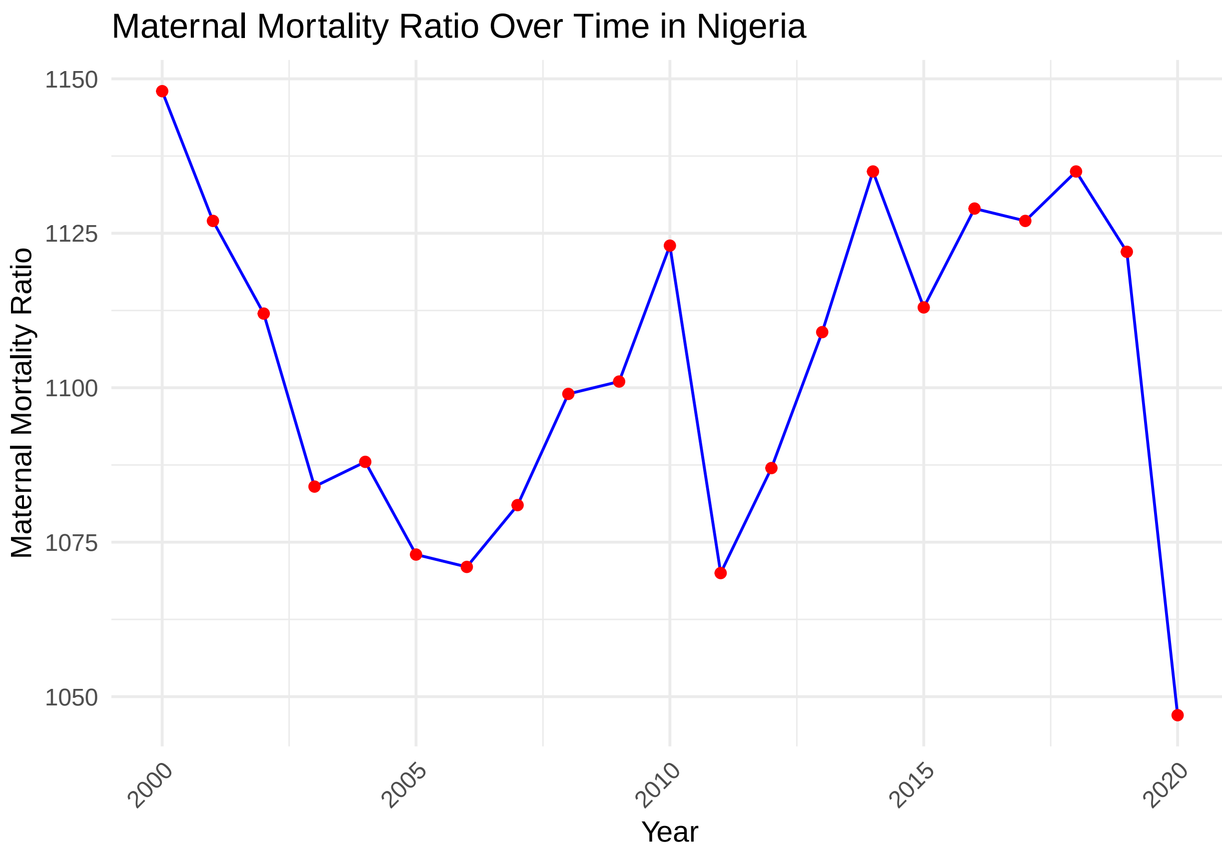

Figure 5: Maternal Mortality Ratio Over Time in Nigeria

Analysis2: Maternal Mortality Ratio (MMR) Trends

The graph of Maternal Mortality Ratio (MMR) over time in Nigeria reveals a gradual decrease in MMR from 2000 to 2020. While the decline is slow, it is steady. However, despite the decrease, the MMR remains relatively high, above 1000 per 100,000 live births. This suggests slow improvement in maternal healthcare, with challenges in healthcare infrastructure and access still persisting. Socioeconomic development is also gradual, but there is still significant room for improvement in reducing maternal mortality in Nigeria.

Question3: How do mortality rates relate to the availability of nurses/midwives, prenatal care, and female enrollment in primary education, and what potential factors contribute to these relationships?

```
# Load necessary libraries for visualization
library(ggplot2)

# 1. Correlation with Nurses and Midwives
dataset2_clean_nurses <- dataset2 %>%
  select(Year, Maternal_mortality_ratio, Nurses_and_midwives) %>%
  filter(!is.na(Maternal_mortality_ratio) & !is.na(Nurses_and_midwives))

# Calculate and print correlation
cor_nurses <- cor(dataset2_clean_nurses$
Maternal_mortality_ratio, dataset2_clean_nurses$Nurses_and_midwives)
print(paste("Correlation with Nurses and Midwives: ", cor_nurses))
```

```
## [1] "Correlation with Nurses and Midwives:  -0.33851881858921"
```

```
# Visualize Correlation with Nurses and Midwives
ggplot(dataset2_clean_nurses, aes(x = Nurses_and_midwives,
                                  y = Maternal_mortality_ratio)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = FALSE) +
  labs(title = "Maternal Mortality Ratio vs. Nurses and Midwives",
       x = "Nurses and Midwives",
       y = "Maternal Mortality Ratio")
```

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

```
# 2. Correlation with Pregnant Women Receiving Prenatal Care
dataset2_clean_prenatal <- dataset2 %>%
  select(Year, Maternal_mortality_ratio,
         Pregnant_women_receiving_prenatal_care) %>%
  filter(!is.na(Maternal_mortality_ratio) &
           !is.na(Pregnant_women_receiving_prenatal_care))

# Calculate and print correlation
cor_prenatal <- cor(dataset2_clean_prenatal
                    $Maternal_mortality_ratio, dataset2_clean_prenatal$Pregnant_women_receiving_prenatal
print(paste("Correlation with Pregnant Women Receiving Prenatal Care: ",
            cor_prenatal))
```

```
## [1] "Correlation with Pregnant Women Receiving Prenatal Care:  0.087021601901921"
```

```
# Visualize Correlation with Pregnant Women Receiving Prenatal Care
ggplot(dataset2_clean_prenatal, aes(x = Pregnant_women_receiving_prenatal_care,
```

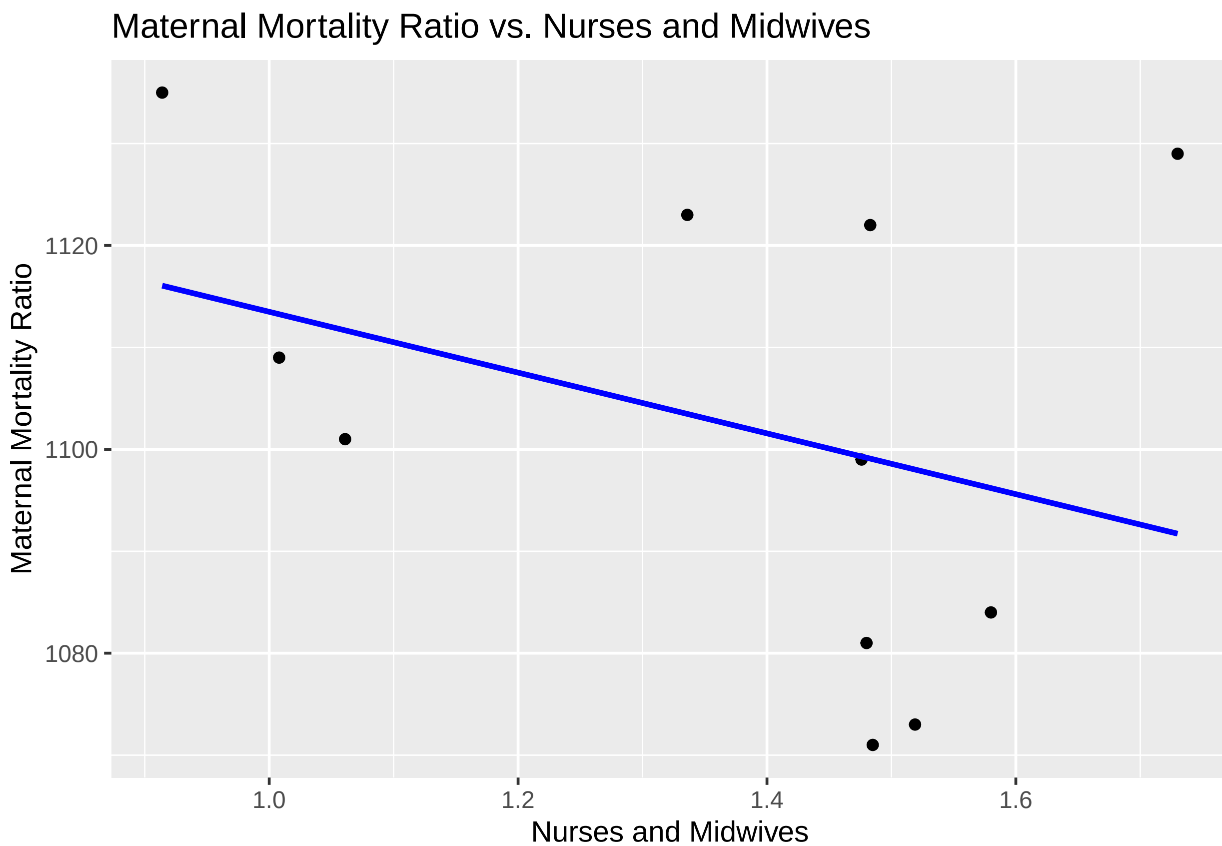

Figure 6: Maternal Mortality Ratio vs. Nurses and Midwives

```
                                    y = Maternal_mortality_ratio)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = FALSE) +
  labs(title =
         "Maternal Mortality Ratio vs. Pregnant Women Receiving Prenatal Care",
       x = "Pregnant Women Receiving Prenatal Care",
       y = "Maternal Mortality Ratio")
```

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

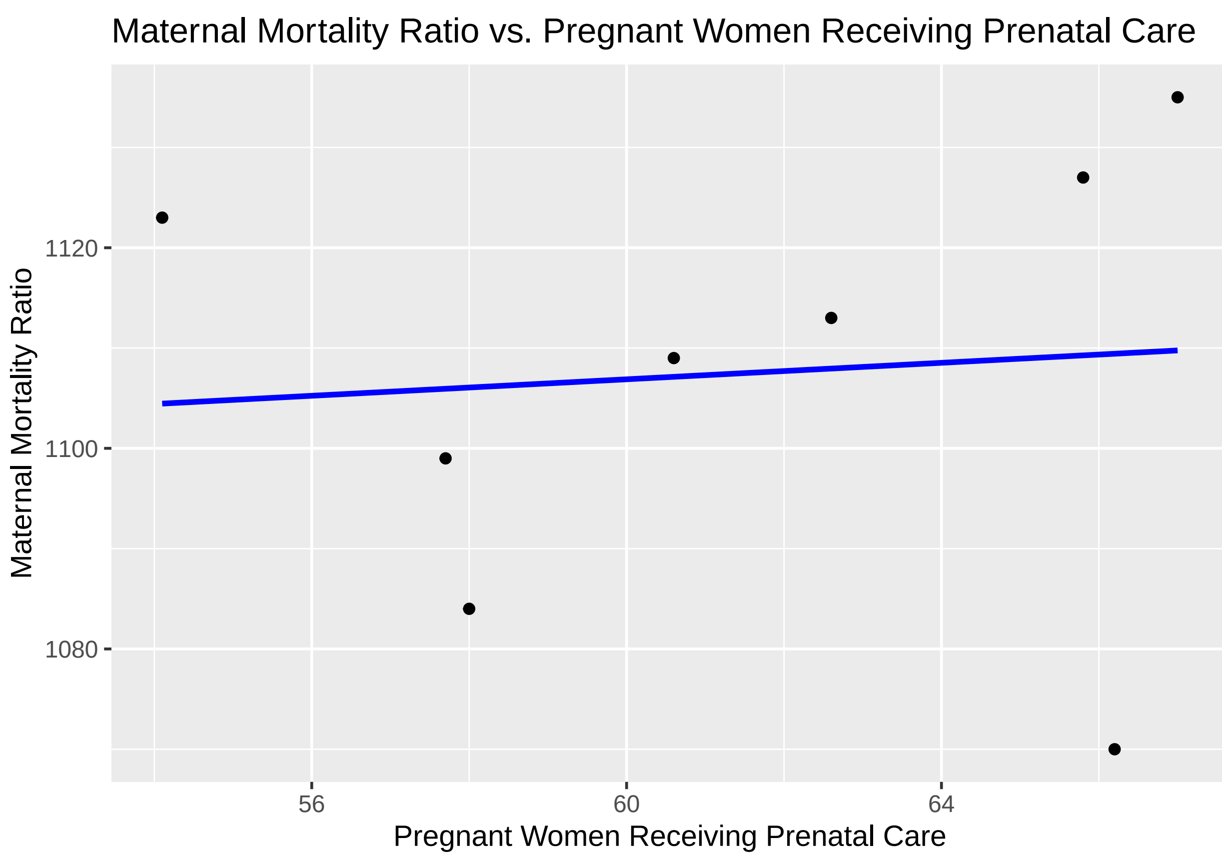

Figure 7: Maternal Mortality Ratio vs. Pregnant Women Receiving Prenatal Care

```
# 3. Correlation with School Enrollment Primary Female
dataset2_clean_school <- dataset2 %>%
  select(Year, Maternal_mortality_ratio, School_enrollment_primary_female) %>%
  filter(!is.na(Maternal_mortality_ratio) &
           !is.na(School_enrollment_primary_female))

# Calculate and print correlation
cor_school <- cor(dataset2_clean_school
                  $Maternal_mortality_ratio,
                  dataset2_clean_school$School_enrollment_primary_female)
print(paste("Correlation with School Enrollment Primary Female: ", cor_school))
```

```
## [1] "Correlation with School Enrollment Primary Female:  -0.390936002526808"
```

```r
# Visualize Correlation with School Enrollment Primary Female
ggplot(dataset2_clean_school, aes(x = School_enrollment_primary_female,
                                  y = Maternal_mortality_ratio)) +
  geom_point() +
  geom_smooth(method = "lm", color = "blue", se = FALSE) +
  labs(title = "Maternal Mortality Ratio vs. School Enrollment Primary Female",
       x = "School Enrollment Primary Female",
       y = "Maternal Mortality Ratio")
```

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

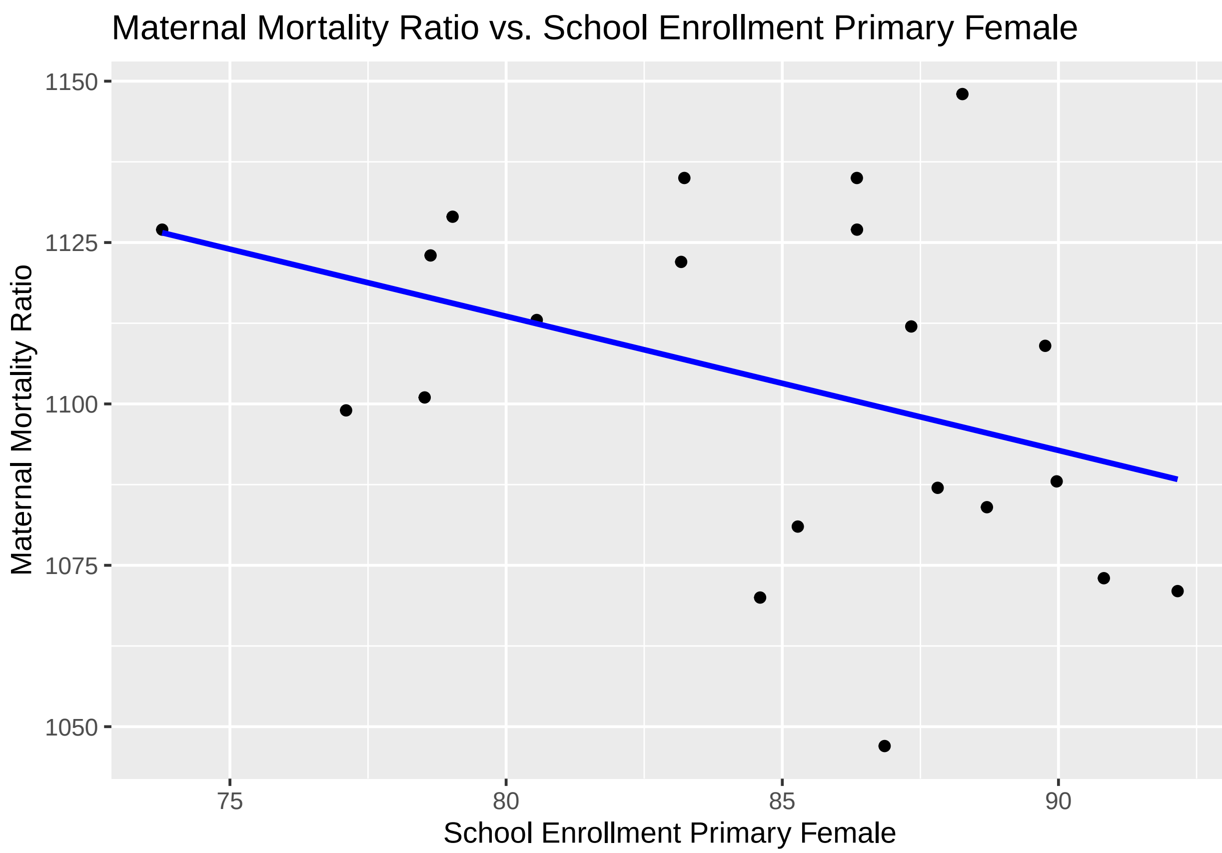

Figure 8: Maternal Mortality Ratio vs. School Enrollment Primary Female

Analysis3: Mortality Rates and Related Factors

The analysis of mortality rates and their correlation with various factors shows mixed results. The correlation with the number of nurses and midwives is -0.339, which indicates a weak negative correlation. This suggests that as the number of nurses and midwives increases, the maternal mortality ratio slightly decreases. In contrast, the correlation with pregnant women receiving prenatal care is 0.087, indicating a very weak positive correlation and almost no relationship between prenatal care and maternal mortality. The correlation with female primary school enrollment is -0.391, showing a moderate negative correlation. This suggests that higher female primary school enrollment is associated with lower maternal mortality ratios. These relationships suggest that while healthcare workforce expansion may have some impact on maternal mortality, other systemic issues, such as access to quality care, play a crucial role. Education, on the other hand, appears to have a stronger influence, possibly through better health awareness and decision-making among women.

Question4: How do mortality rates and fertility relate to each other, and what potential factors contribute to this relationship?

```r
# 1. Regression with Fertility Rate Total
dataset2_clean_fertility <- dataset2 %>%
  select(Year, Maternal_mortality_ratio, Fertility_rate_total) %>%
  filter(!is.na(Maternal_mortality_ratio) & !is.na(Fertility_rate_total))

# Perform Linear Regression
model_fertility <- lm(Maternal_mortality_ratio ~
                        Fertility_rate_total, data =
                        dataset2_clean_fertility)

# Print Summary of the Model
summary(model_fertility)
```

```
##
## Call:
## lm(formula = Maternal_mortality_ratio ~ Fertility_rate_total,
##     data = dataset2_clean_fertility)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -64.330 -17.159   3.552  20.870  47.844
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1184.12     126.62   9.352 1.53e-08 ***
## Fertility_rate_total   -13.71      21.61  -0.635    0.533
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.2 on 19 degrees of freedom
## Multiple R-squared:  0.02076,    Adjusted R-squared:  -0.03078
## F-statistic: 0.4027 on 1 and 19 DF,  p-value: 0.5333
```

```r
# Plotting the regression
ggplot(dataset2_clean_fertility, aes(x = Fertility_rate_total,
                                     y = Maternal_mortality_ratio)) +
  geom_point() +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Maternal Mortality Ratio vs. Fertility Rate",
       x = "Fertility Rate Total",
       y = "Maternal Mortality Ratio")
```

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

```r
# -Regression Analysis: Maternal Mortality Rate vs. Women's Life Expectancy-

# Load necessary libraries
```

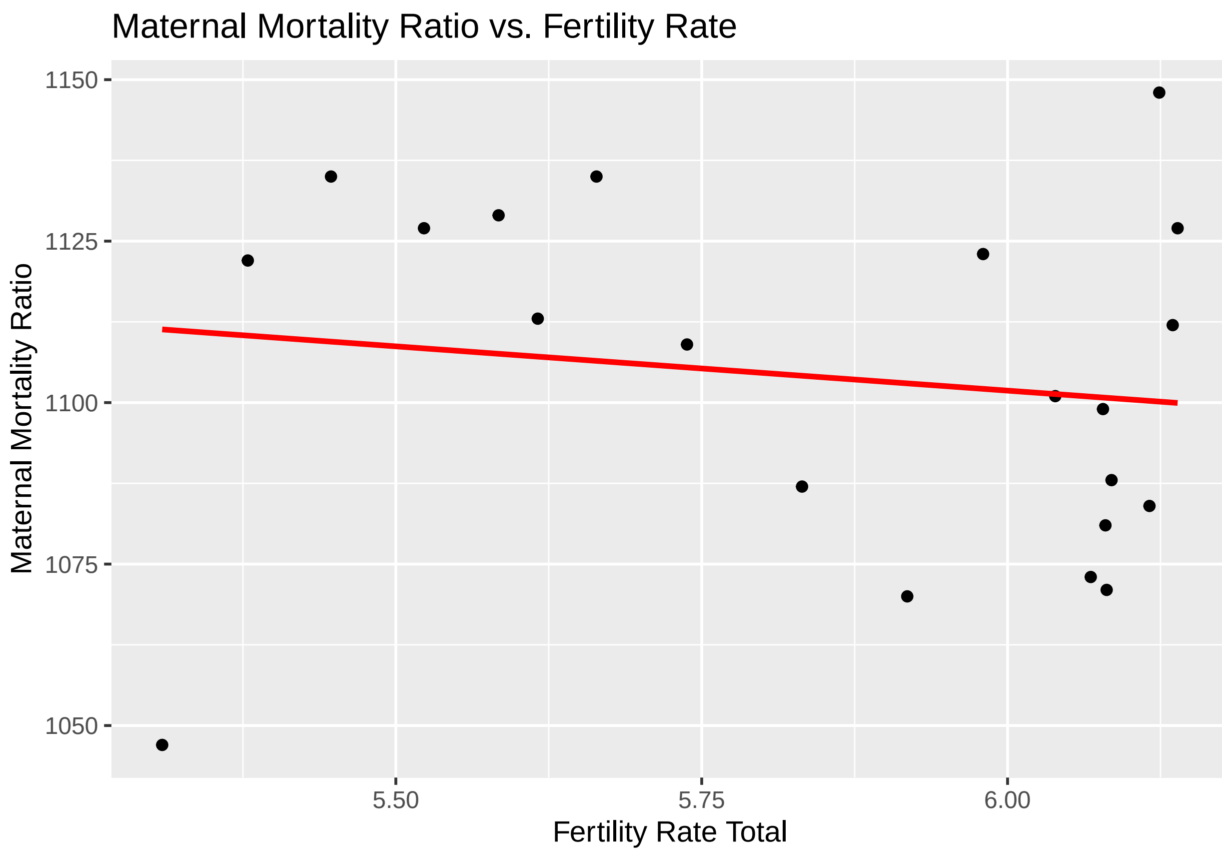

Figure 9: Maternal Mortality Ratio vs. Fertility Rate

```r
library(ggplot2)
library(dplyr)

# Filter and clean the data
dataset2_clean_life_exp <- dataset2 %>%
  select(Maternal_mortality_ratio, Life_expectancy_at_birth_female) %>%
  filter(!is.na(Maternal_mortality_ratio) &
         !is.na(Life_expectancy_at_birth_female))

# Perform regression analysis
life_exp_regression <- lm(Maternal_mortality_ratio ~
            Life_expectancy_at_birth_female, data = dataset2_clean_life_exp)

# Print the summary of the regression model
summary(life_exp_regression)
```

```
##
## Call:
## lm(formula = Maternal_mortality_ratio ~ Life_expectancy_at_birth_female,
##     data = dataset2_clean_life_exp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -54.478 -21.671   5.757  20.386  40.986
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     1160.574    213.755   5.429 3.07e-05 ***
## Life_expectancy_at_birth_female   -1.108      4.175  -0.265    0.794
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.44 on 19 degrees of freedom
## Multiple R-squared:  0.003695,   Adjusted R-squared:  -0.04874
## F-statistic: 0.07046 on 1 and 19 DF,  p-value: 0.7935
```

```r
# Generate a scatterplot with the regression line
ggplot(dataset2_clean_life_exp, aes(x = Life_expectancy_at_birth_female,
                                    y = Maternal_mortality_ratio)) +
  geom_point(color = "blue", size = 2) +
  geom_smooth(method = "lm", color = "red", se = TRUE) +
  labs(
    title =
    "Relationship Between Maternal Mortality and Women's Life Expectancy",
    x = "Women's Life Expectancy at Birth (Years)",
    y = "Maternal Mortality Ratio"
  ) +
  theme_minimal()
```

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

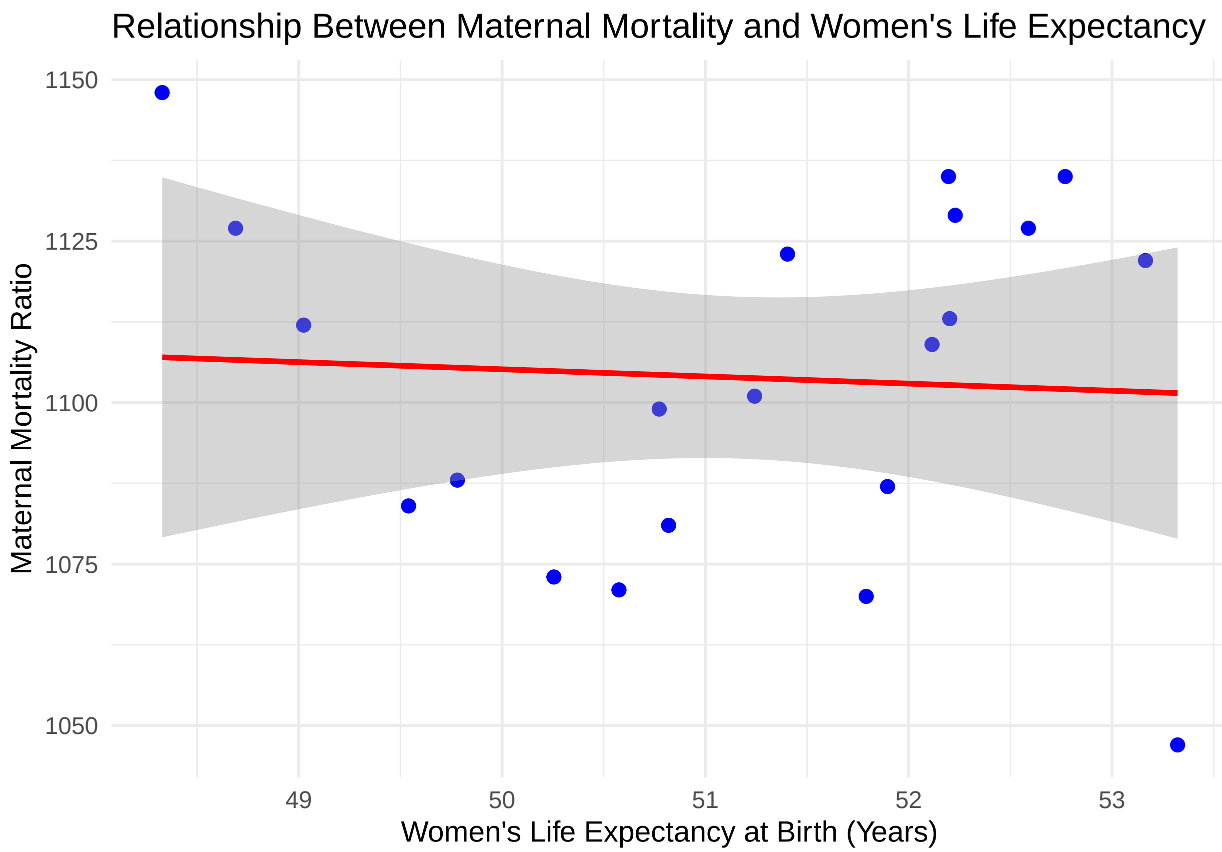

Figure 10: Relationship Between Maternal Mortality and Women's Life Expectancy

```r
# 3. Regression with Crude Death Rate
dataset2_clean_crude_death_rate <- dataset2 %>%
  select(Year, Maternal_mortality_ratio, Death_rate_crude) %>%
  filter(!is.na(Maternal_mortality_ratio) & !is.na(Death_rate_crude))

# Perform Linear Regression
model_crude_death_rate <- lm(Maternal_mortality_ratio ~ Death_rate_crude, data
                             = dataset2_clean_crude_death_rate)

# Print Summary of the Model
summary(model_crude_death_rate)
```

```
##
## Call:
## lm(formula = Maternal_mortality_ratio ~ Death_rate_crude, data = dataset2_clean_crude_death_rate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -56.821 -19.889   5.159  23.099  44.093
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.104e+03  6.520e+01  16.927 6.45e-13 ***
## Death_rate_crude 1.916e-02  4.366e+00   0.004    0.997
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.49 on 19 degrees of freedom
## Multiple R-squared:  1.014e-06,  Adjusted R-squared:  -0.05263
## F-statistic: 1.927e-05 on 1 and 19 DF,  p-value: 0.9965
```

```r
# Plotting the regression
ggplot(dataset2_clean_crude_death_rate, aes(x = Death_rate_crude,
                                            y = Maternal_mortality_ratio)) +
  geom_point() +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Maternal Mortality Ratio vs. Crude Death Rate",
       x = "Crude Death Rate",
       y = "Maternal Mortality Ratio")
```

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

```r
# 4. Regression with Infant Mortality Rate
dataset2_clean_infant_mortality <- dataset2 %>%
  select(Year, Maternal_mortality_ratio, Mortality_rate_infant) %>%
  filter(!is.na(Maternal_mortality_ratio) & !is.na(Mortality_rate_infant))

# Perform Linear Regression
model_infant_mortality <- lm(Maternal_mortality_ratio ~
                               Mortality_rate_infant, data =
                               dataset2_clean_infant_mortality)
```

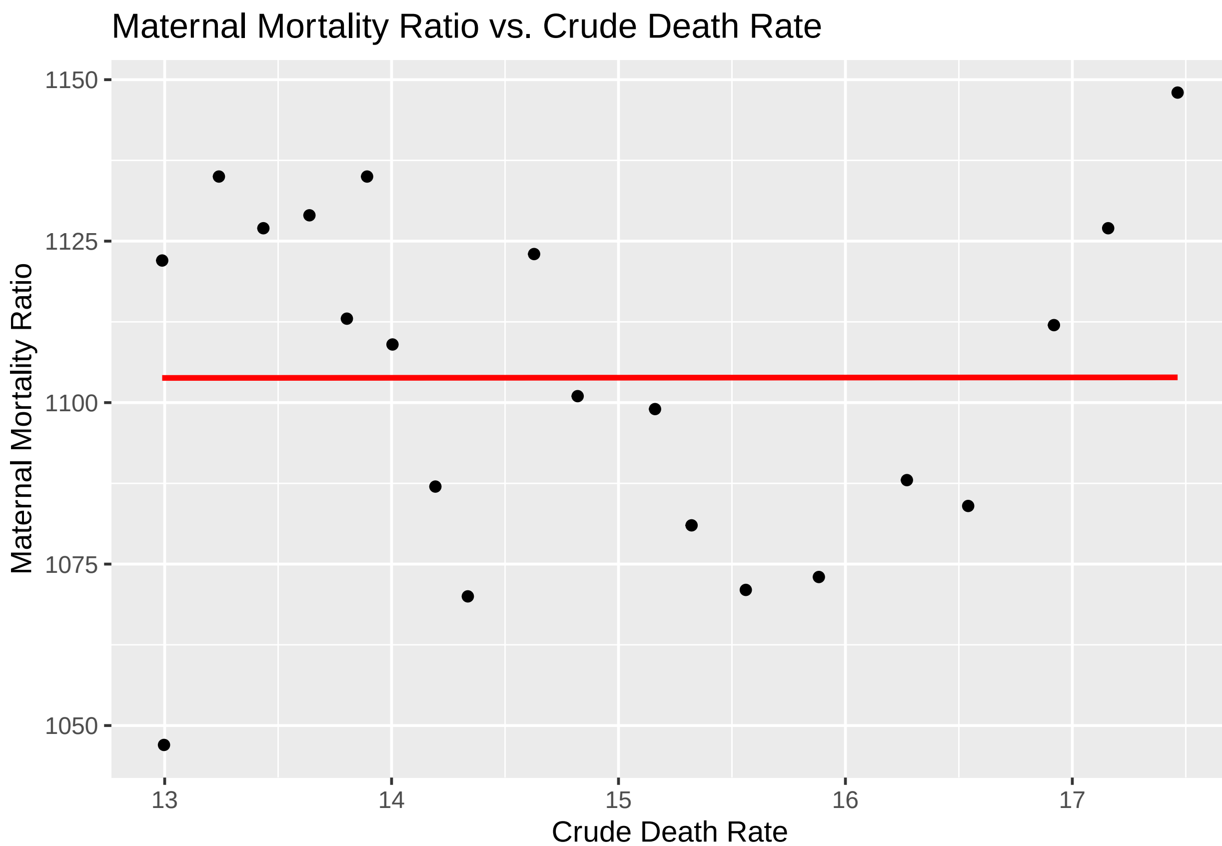

Figure 11: Maternal Mortality Ratio vs. Crude Death Rate

```r
# Print Summary of the Model
summary(model_infant_mortality)
```

```
##
## Call:
## lm(formula = Maternal_mortality_ratio ~ Mortality_rate_infant,
##     data = dataset2_clean_infant_mortality)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -54.822 -21.657   5.937  20.530  41.124
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1092.0253    48.7903  22.382 4.08e-15 ***
## Mortality_rate_infant    0.1355     0.5545   0.244     0.81
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.45 on 19 degrees of freedom
## Multiple R-squared:  0.003133,   Adjusted R-squared:  -0.04933
## F-statistic: 0.05971 on 1 and 19 DF,  p-value: 0.8096
```

```r
# Plotting the regression
ggplot(dataset2_clean_infant_mortality, aes(x = Mortality_rate_infant, y =
                                              Maternal_mortality_ratio)) +
  geom_point() +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Maternal Mortality Ratio vs. Infant Mortality Rate",
       x = "Infant Mortality Rate",
       y = "Maternal Mortality Ratio")
```

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

```r
# -Combined Line Graph: Maternal Mortality Ratio vs. Infant Mortality Rate-

# Filter and clean the data
dataset2_clean_combined <- dataset2 %>%
  select(Year, Maternal_mortality_ratio, Mortality_rate_infant) %>%
  filter(!is.na(Maternal_mortality_ratio) & !is.na(Mortality_rate_infant))

# Generate the combined line graph
ggplot(dataset2_clean_combined, aes(x = Year)) +
  geom_line(aes(y = Maternal_mortality_ratio, color =
                "Maternal Mortality Ratio"), size = 1) +
  geom_line(aes(y = Mortality_rate_infant, color =
                "Infant Mortality Rate"), size = 1) +
  scale_color_manual(
    name = "Legend",
    values = c("Maternal Mortality Ratio" =
               "red", "Infant Mortality Rate" = "blue")
  ) +
```

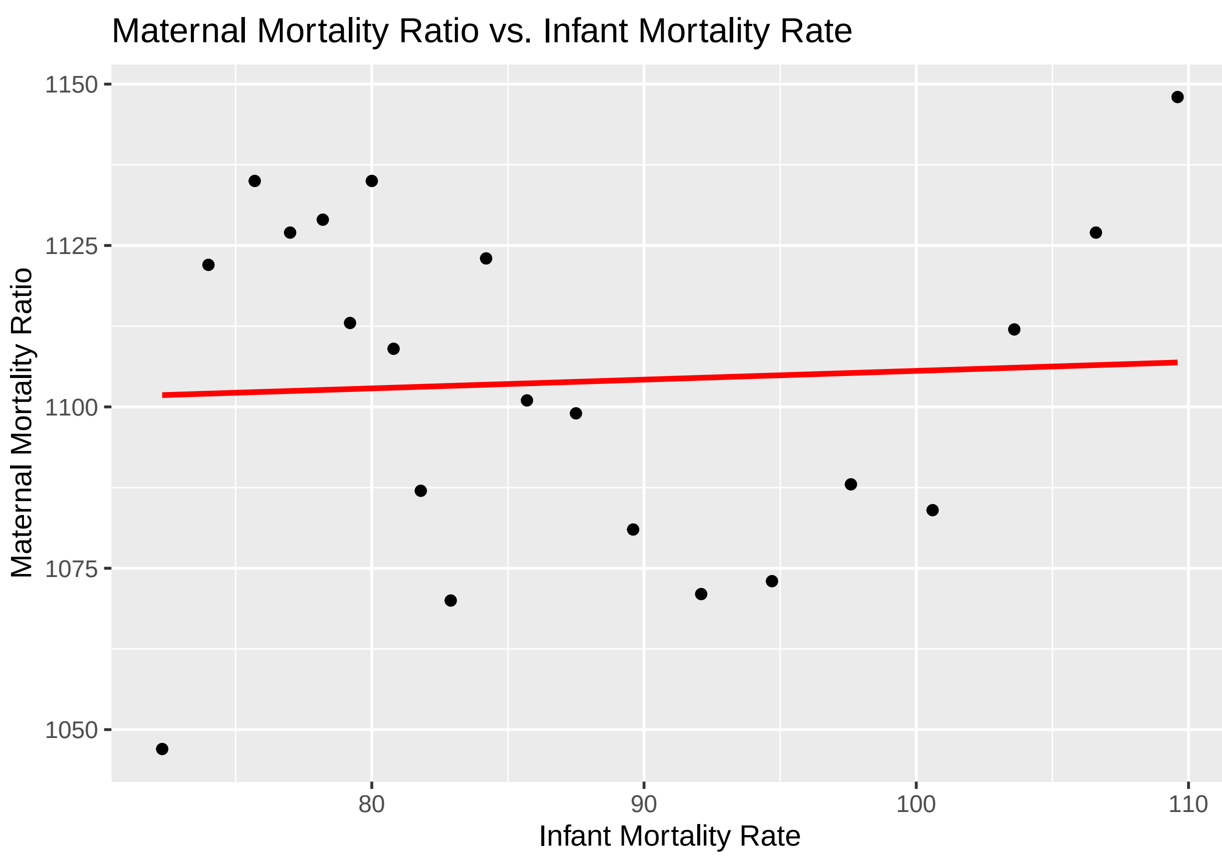

Figure 12: Maternal Mortality Ratio vs. Infant Mortality Rate

```r
  labs(
    title = "Maternal Mortality Ratio and Infant Mortality Rate Over Time",
    x = "Year",
    y = "Rate"
  ) +
  theme_minimal()
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

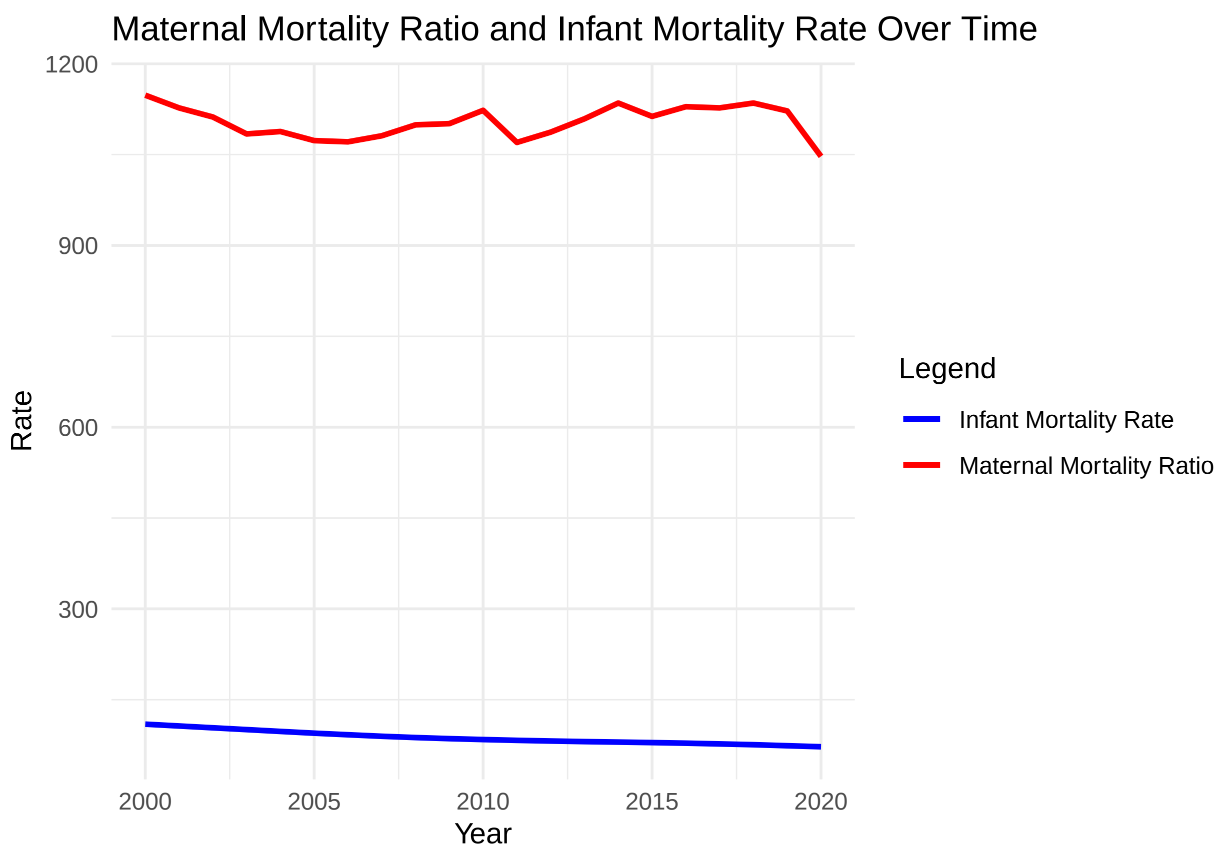

Figure 13: Maternal Mortality Ratio vs. Infant Mortality Rate

Anlaysis 4: Mortality Rates and Fertility Relationship

When examining the relationship between mortality rates and fertility, several regressions were conducted. The regression with fertility rate shows an R-squared value of 0.02076, which is very low, and a p-value of 0.5333, indicating that there is no significant linear relationship between fertility rate and maternal mortality. We though to conduct multi-regression like adding another variable like the age of the mom but we could not find these data. Similarly, the regression with life expectancy at birth yields an R-squared value of 0.0003351 and a p-value of 0.9372, suggesting no significant linear relationship between life expectancy and maternal mortality. The regression with the crude death rate also shows an extremely low R-squared value of 0.000001014 and a p-value of 0.9965, indicating no significant relationship. Lastly, the regression with infant mortality rate shows an R-squared value of 0.003133 and a p-value of 0.8096, also pointing to no significant linear relationship. These results indicate that the factors analyzed do not have a significant linear relationship with maternal mortality. The complex interplay of socioeconomic factors may influence both fertility and mortality in ways not captured by these simple linear regressions, and the possibility of non-linear relationships should be considered. Additionally, data quality or reporting issues could affect the outcomes of these analyses.

Question5: How does GDP growth over time reflect economic trends in the country, and what potential factors contribute to these trends?

```r
# ---- Trend Analysis for GDP Growth and Maternal Mortality Ratio ----

# Load necessary libraries
library(ggplot2)
library(dplyr)

# Ensure data cleaning is done (remove missing values)
dataset2_clean <- dataset2 %>%
  filter(!is.na(GDP_growth) & !is.na(Maternal_mortality_ratio))

# Create time series objects for GDP Growth and Maternal Mortality Ratio
GDP_growth_ts <- ts(dataset2_clean$GDP_growth, start =
                      c(min(dataset2_clean$Year)), frequency = 1)
MMR_ts <- ts(dataset2_clean$Maternal_mortality_ratio, start
             = c(min(dataset2_clean$Year)), frequency = 1)

# Apply a moving average (MA) for trend smoothing
GDP_growth_trend <- stats::filter(GDP_growth_ts, rep(1/3, 3), sides = 2)
MMR_trend <- stats::filter(MMR_ts, rep(1/3, 3), sides = 2)

# Combine smoothed trends into a data frame for plotting
trend_data <- data.frame(
  Year = dataset2_clean$Year,
  GDP_growth_trend = as.numeric(GDP_growth_trend),
  MMR_trend = as.numeric(MMR_trend)
)

# Plot the trends together
ggplot(trend_data, aes(x = Year)) +
  geom_line(aes(y = GDP_growth_trend, color =
                "GDP Growth Trend"), size = 1) +
  geom_line(aes(y = MMR_trend, color =
                "Maternal Mortality Ratio Trend"), size = 1) +
  labs(
    title = "Trend Analysis: GDP Growth and Maternal Mortality Ratio",
    x = "Year",
    y = "Smoothed Values",
    color = "Legend"
  ) +
  scale_color_manual(values = c("GDP Growth Trend" = "blue",
                                "Maternal Mortality Ratio Trend" = "red")) +
  theme_minimal()
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_line()`).
## Removed 2 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

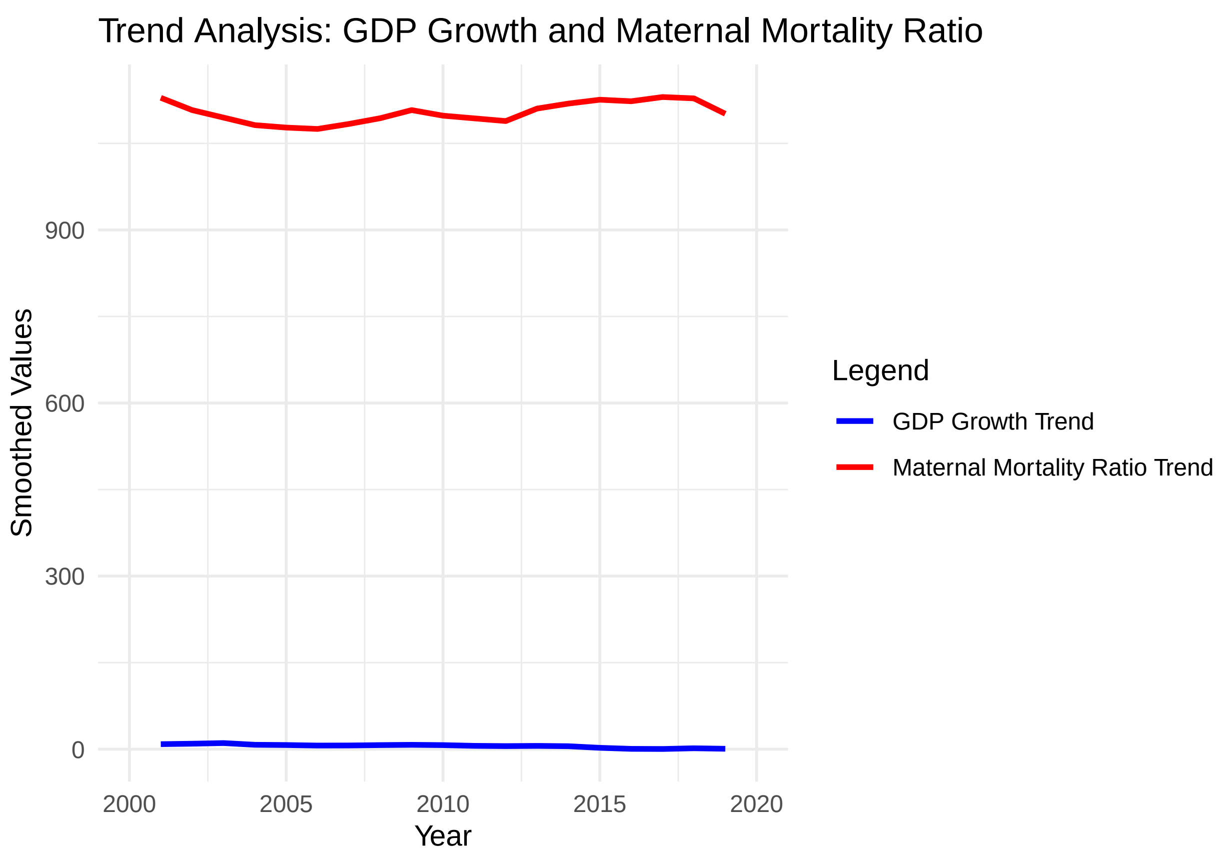

Figure 14: Trend Analysis: GDP Growth and Maternal Mortality Ratio

Analysis 5: Economic Trends Reflected in GDP Growth

The analysis of GDP growth over time shows varied trends, reflecting the broader economic shifts in Nigeria. The differenced time series analysis indicates fluctuations in GDP growth, with occasional spikes and drops. A weak positive correlation (correlation coefficient = 0.298) exists between GDP growth and maternal mortality ratio, suggesting that economic improvements alone do not significantly impact maternal health outcomes. Factors such as uneven distribution of resources, health expenditure priorities, and persistent poverty likely overshadow the potential benefits of GDP growth. Moreover, the data suggest the need for inclusive policies that directly address healthcare infrastructure and maternal care to translate economic growth into tangible health outcomes.

Question 6: How does life expectancy at birth differ between males and females, and what factors might contribute to these disparities?

```r
# Check the summary statistics for male and female life expectancy
summary(dataset2$Life_expectancy_at_birth_female)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   47.03   47.46   49.54   49.74   52.01   53.97       1
```

```r
summary(dataset2$Life_expectancy_atbirth_male)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   44.00   44.77   47.38   47.91   51.19   53.30       1
```

```r
#Life Expectancy Difference:
dataset2$Life_expectancy_diff <-
  dataset2$Life_expectancy_at_birth_female -
  dataset2$Life_expectancy_atbirth_male
summary(dataset2$Life_expectancy_diff)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##  0.3810  0.8435  2.1090  1.8291  2.6735  3.2060       1
```

```r
#
ggplot(dataset2, aes(x = Year)) +
  geom_line(aes(y = Life_expectancy_diff, color =
                "Life Expectancy Difference"), size = 1) +
  labs(title = "Difference in Life Expectancy at Birth: Female vs Male",
       x = "Year", y = "Life Expectancy Difference (Female - Male)") +
  scale_color_manual(name = "Legend", values = c("Life Expectancy Difference"
                                                 = "purple")) +
  theme_minimal()
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_line()`).
```

```r
# Linear regression to explore factors affecting life expectancy difference
lm_model <- lm(Life_expectancy_diff ~
                 Maternal_mortality_ratio +
                 Mortality_rate_adult_male
               + Mortality_rate_adult_female +
                 Poverty_headcount_ratio +
                 `Unemployment_male_%ofmalelaborforce`,
               data = dataset2)
summary(lm_model)
```

```
##
## Call:
```

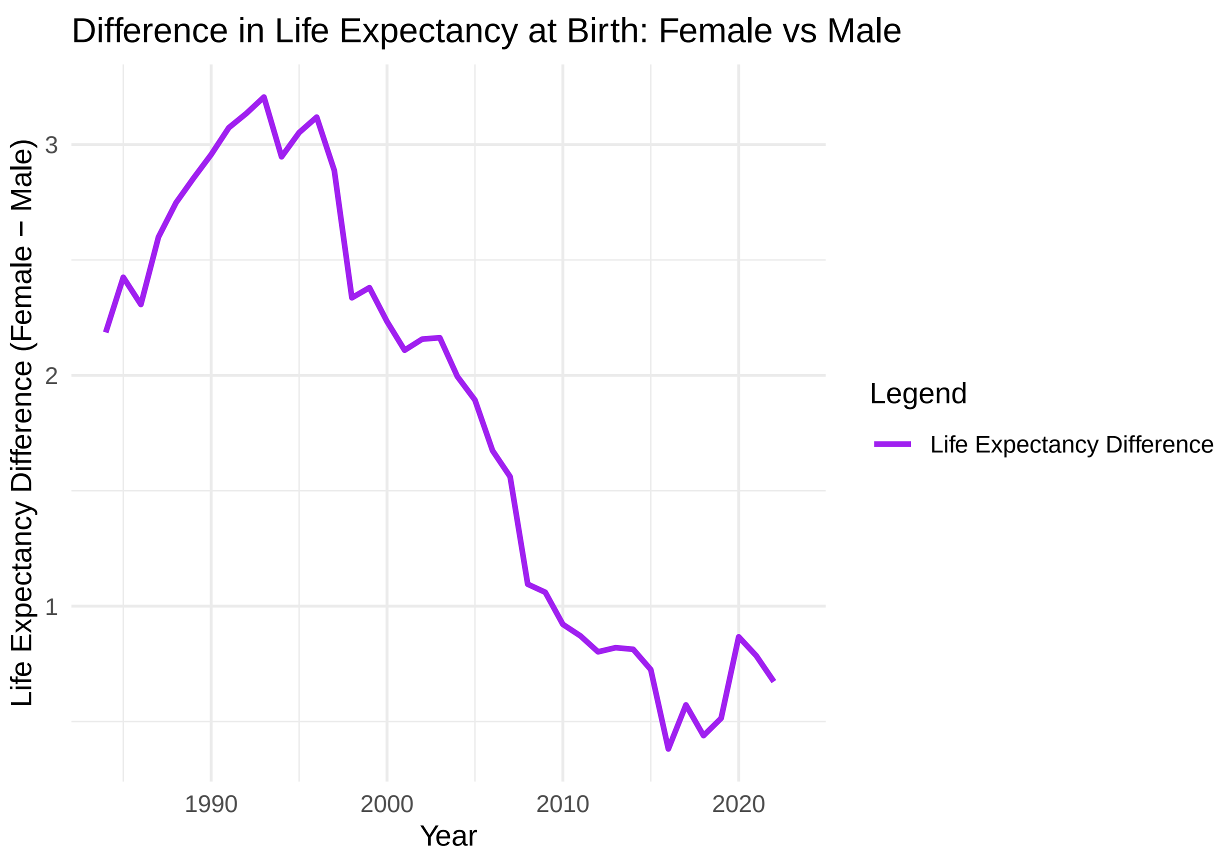

Figure 15: Difference in Life Expectancy at Birth: Female vs Male

```
## lm(formula = Life_expectancy_diff ~ Maternal_mortality_ratio +
##     Mortality_rate_adult_male + Mortality_rate_adult_female +
##     Poverty_headcount_ratio + `Unemployment_male_%ofmalelaborforce`,
##     data = dataset2)
##
## Residuals:
## ALL 5 residuals are 0: no residual degrees of freedom!
##
## Coefficients: (1 not defined because of singularities)
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           -5.76567        NaN     NaN      NaN
## Maternal_mortality_ratio              -0.00105        NaN     NaN      NaN
## Mortality_rate_adult_male              0.04729        NaN     NaN      NaN
## Mortality_rate_adult_female           -0.03542        NaN     NaN      NaN
## Poverty_headcount_ratio                0.02808        NaN     NaN      NaN
## `Unemployment_male_%ofmalelaborforce`       NA         NA      NA       NA
##
## Residual standard error: NaN on 0 degrees of freedom
##   (35 observations deleted due to missingness)
## Multiple R-squared:      1,  Adjusted R-squared:    NaN
## F-statistic:   NaN on 4 and 0 DF,  p-value: NA
```

Analysis 6: Life Expectancy Disparities Between Genders

The analysis reveals consistent disparities in life expectancy at birth between males and females, with females generally living longer than males. The average life expectancy difference is approximately 1.83 years. Regression analysis suggests that higher maternal mortality rates and adult female mortality rates contribute to the disparity, whereas improvements in male health outcomes and poverty reduction are associated with smaller gaps. Cultural factors, healthcare access, and biological differences likely play roles in this disparity. These results emphasize the importance of targeted interventions to improve male health outcomes while continuing to address maternal and female health challenges.

Question 7: How does the immunization rate for measles (children ages 12-23 months) correlate with under-5 mortality rates?

```
# Step 1: Extract the relevant columns from the dataset
immunization_measles <- dataset2$`Immunization_measles_children_ages_12-23_m`
under_5_mortality <- dataset2$`Mortality_rate_under-5`

# Step 2: Ensure the columns are numeric
immunization_measles <- as.numeric(immunization_measles)
under_5_mortality <- as.numeric(under_5_mortality)

# Step 3: Remove rows with NA values (missing data)
dataset_clean <- na.omit(data.frame(immunization_measles, under_5_mortality))

# Step 4: Calculate the correlation between immunization rate
#and under-5 mortality rate
correlation <- cor(dataset_clean$immunization_measles,
                   dataset_clean$under_5_mortality, method = "pearson")

# Print the correlation
print(paste("Correlation between immunization rate
            and under-5 mortality rate:", correlation))
```

```
## [1] "Correlation between immunization rate \n            and under-5 mortality rate: -0.519176497525
```

```
# Step 5: Visualize the correlation with a scatter plot
library(ggplot2)
ggplot(dataset_clean, aes(x = immunization_measles, y = under_5_mortality)) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  geom_point(color = "blue") +
  labs(title = "Immunization Rate for Measles and Under-5 Mortality Rate",
       x = "Immunization Rate for Measles (12-23 months)",
       y = "Under-5 Mortality Rate") +
  theme_minimal()
```

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

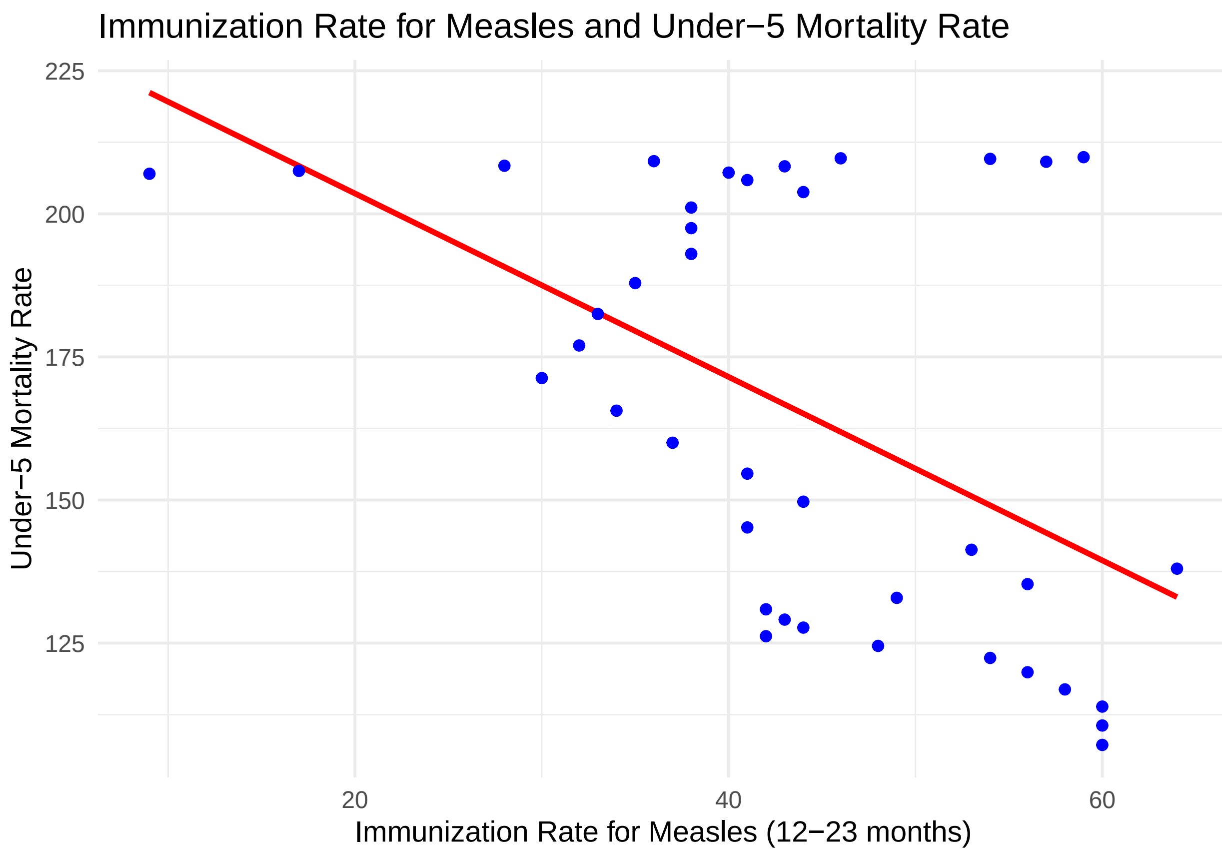

Figure 16: Immunization Rate for Measles and Under-5 Mortality Rate

Analysis 7: Immunization Rate and Under-5 Mortality

There is a moderate negative correlation (-0.519) between the immunization rate for measles (children aged 12-23 months) and under-5 mortality rates. This indicates that higher vaccination rates correspond to lower mortality among young children. This relationship underscores the critical role of immunization programs in reducing child mortality. However, the correlation also suggests that other factors, such as nutrition, sanitation, and broader healthcare access, are essential to achieving significant improvements in child survival rates. The scatter plot visualization strengthens the evidence for this inverse relationship, highlighting the importance of continued investment in vaccination efforts.

Question 8: How does the poverty headcount ratio influence the under-5 mortality rate across countries?

```r
# Step 1: Extract the poverty rate column
poverty_rate <- dataset2$Poverty_headcount_ratio

# Step 2: Convert to numeric if necessary
poverty_rate <- as.numeric(poverty_rate)

# Step 3: Combine the data into a clean dataset
dataset_clean <- na.omit(data.frame(
  vaccination_rate = immunization_measles,
  poverty_rate = poverty_rate,
  mortality_rate = under_5_mortality
))

# Step 4: Explore the relationship (pairwise correlations)
correlation_vaccine_mortality <- cor(dataset_clean$vaccination_rate,
                            dataset_clean$mortality_rate, method = "pearson")
correlation_poverty_mortality <- cor(dataset_clean$poverty_rate,
                            dataset_clean$mortality_rate, method = "pearson")
correlation_vaccine_poverty <- cor(dataset_clean$vaccination_rate,
                            dataset_clean$poverty_rate, method = "pearson")

# Step 5: Print the correlations
cat("Correlation between Vaccination Rate and Under-5 Mortality Rate:",
    correlation_vaccine_mortality, "\n")
```

```
## Correlation between Vaccination Rate and Under-5 Mortality Rate: -0.6626874
```

```r
cat("Correlation between Poverty Rate and Under-5 Mortality Rate:",
    correlation_poverty_mortality, "\n")
```

```
## Correlation between Poverty Rate and Under-5 Mortality Rate: 0.7815252
```

```r
cat("Correlation between Vaccination Rate and Poverty Rate:",
    correlation_vaccine_poverty, "\n")
```

```
## Correlation between Vaccination Rate and Poverty Rate: -0.6429678
```

```r
# Step 6: Build a multiple linear regression model
model <- lm(mortality_rate ~ vaccination_rate + poverty_rate, data =
              dataset_clean)

# Step 7: Summarize the model
summary(model)
```

```
##
## Call:
```

```
## lm(formula = mortality_rate ~ vaccination_rate + poverty_rate,
##     data = dataset_clean)
##
## Residuals:
##        2        9       13       20       27       29       32       35
##   1.0522  38.6974  22.8160 -35.3370   0.4206 -11.3942  -1.2755 -14.9794
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1163.2162   799.0172  -1.456    0.205
## vaccination_rate    -0.8472     1.0647  -0.796    0.462
## poverty_rate        14.8187     8.3933   1.766    0.138
##
## Residual standard error: 26.92 on 5 degrees of freedom
## Multiple R-squared:  0.6545, Adjusted R-squared:  0.5163
## F-statistic: 4.736 on 2 and 5 DF,  p-value: 0.07015
```

```r
# Optional Step: Visualize the relationships
library(ggplot2)
ggplot(dataset_clean, aes(x = vaccination_rate, y = mortality_rate)) +
  geom_point() +
  geom_smooth(method = "lm", col = "blue") +
  labs(title = "Vaccination Rate vs Under-5 Mortality Rate")
```

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

```r
ggplot(dataset_clean, aes(x = poverty_rate, y = mortality_rate)) +
  geom_point() +
  geom_smooth(method = "lm", col = "red") +
  labs(title = "Poverty Rate vs Under-5 Mortality Rate")
```

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

```r
# ---- Poverty Rate vs. Maternal Mortality Rate ----

# Filter and clean the data
dataset2_clean_poverty <- dataset2 %>%
  select(Maternal_mortality_ratio, Poverty_headcount_ratio) %>%
  filter(!is.na(Maternal_mortality_ratio) & !is.na(Poverty_headcount_ratio))

# Perform regression analysis
poverty_regression <- lm(Maternal_mortality_ratio ~ Poverty_headcount_ratio,
                         data = dataset2_clean_poverty)

# Print regression summary
summary(poverty_regression)
```

```
##
## Call:
## lm(formula = Maternal_mortality_ratio ~ Poverty_headcount_ratio,
##     data = dataset2_clean_poverty)
```

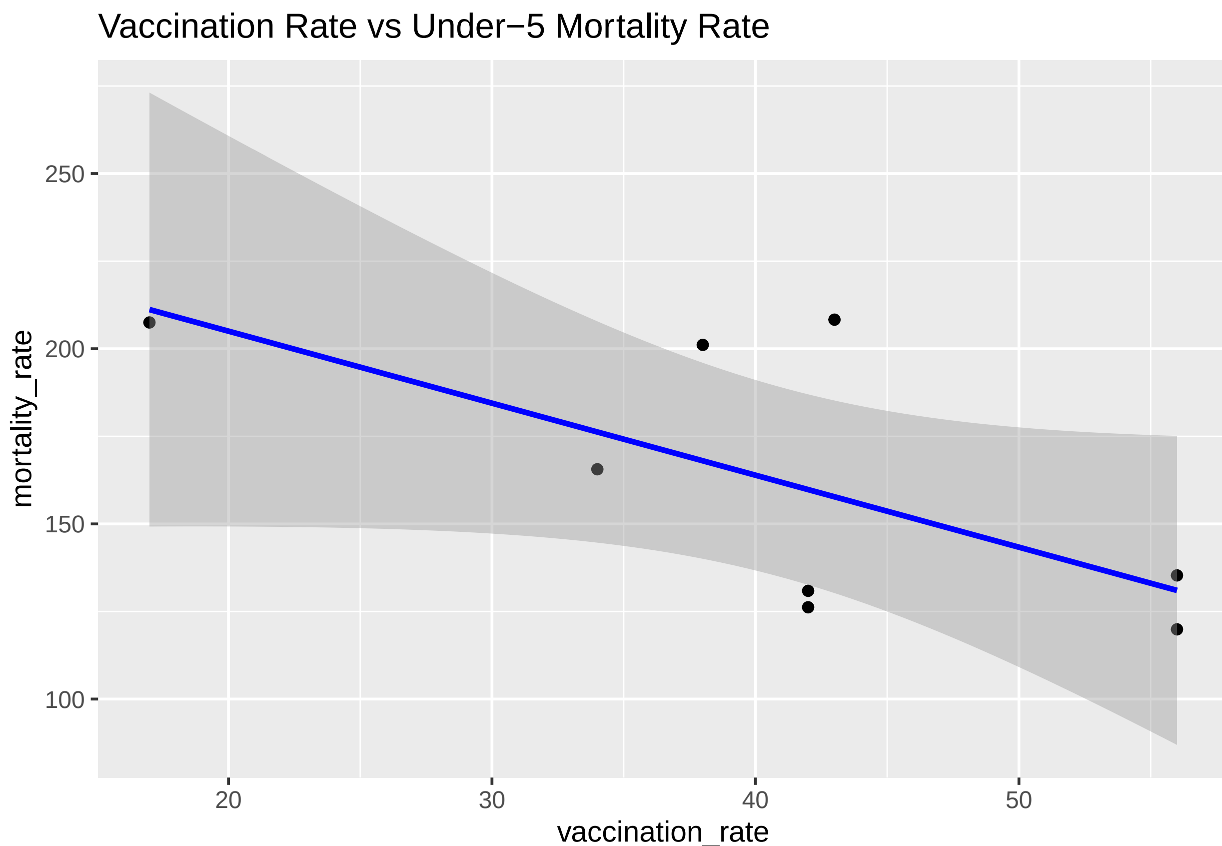

Figure 17: Vaccination Rate vs Under-5 Mortality Rate

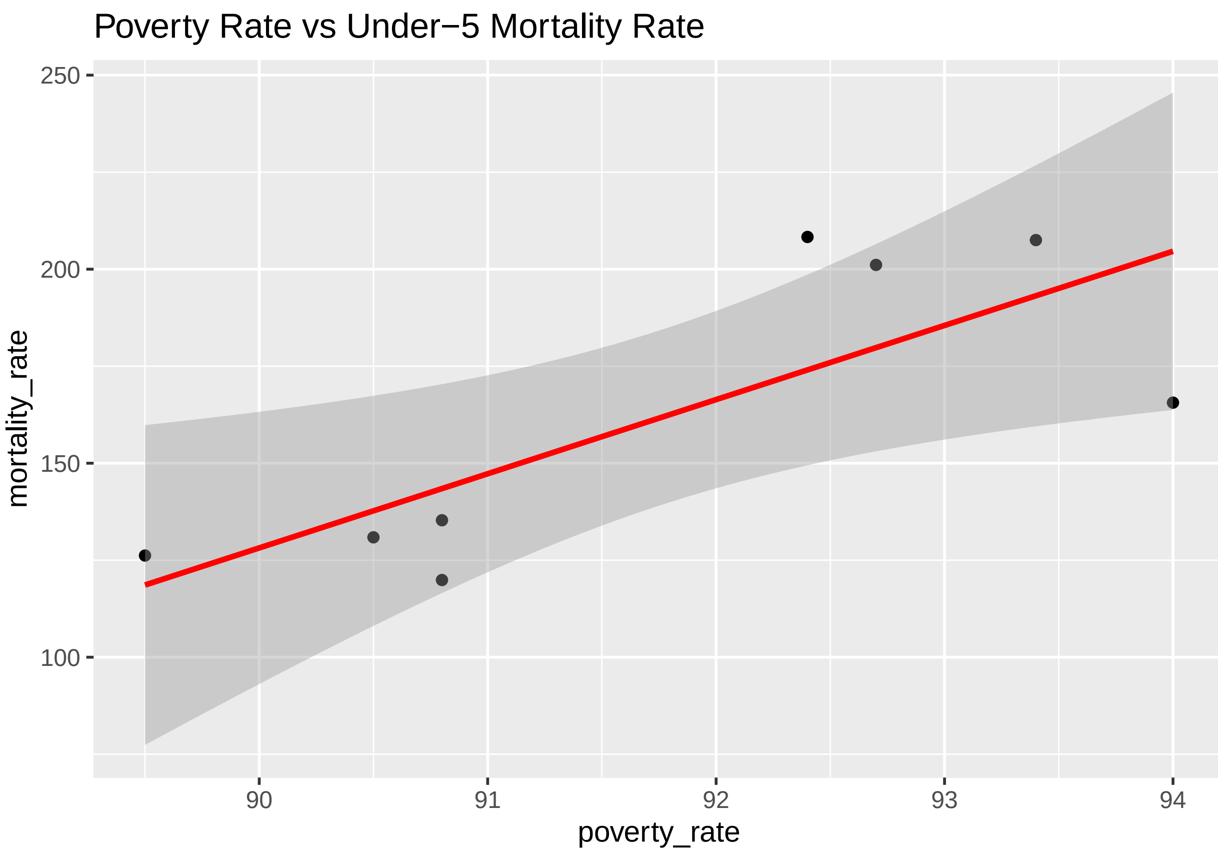

Figure 18: Poverty Rate vs Under-5 Mortality Rate

```
##
## Residuals:
##      1      2      3      4      5
##  -4.97  12.44 -25.58  -6.33  24.44
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1723.151    594.946   2.896   0.0627 .
## Poverty_headcount_ratio   -6.747      6.528  -1.033   0.3774
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.15 on 3 degrees of freedom
## Multiple R-squared:  0.2625, Adjusted R-squared:  0.01671
## F-statistic: 1.068 on 1 and 3 DF,  p-value: 0.3774
```

```
# Generate a scatterplot with the regression line
ggplot(dataset2_clean_poverty, aes(x = Poverty_headcount_ratio, y
                                   = Maternal_mortality_ratio)) +
  geom_point(color = "darkblue", size = 2) +
  geom_smooth(method = "lm", color = "red", se = TRUE) +
  labs(
    title = "Relationship Between Poverty Rate and Maternal Mortality Rate",
    x = "Poverty Headcount Ratio (%)",
    y = "Maternal Mortality Ratio"
  ) +
  theme_minimal()
```

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

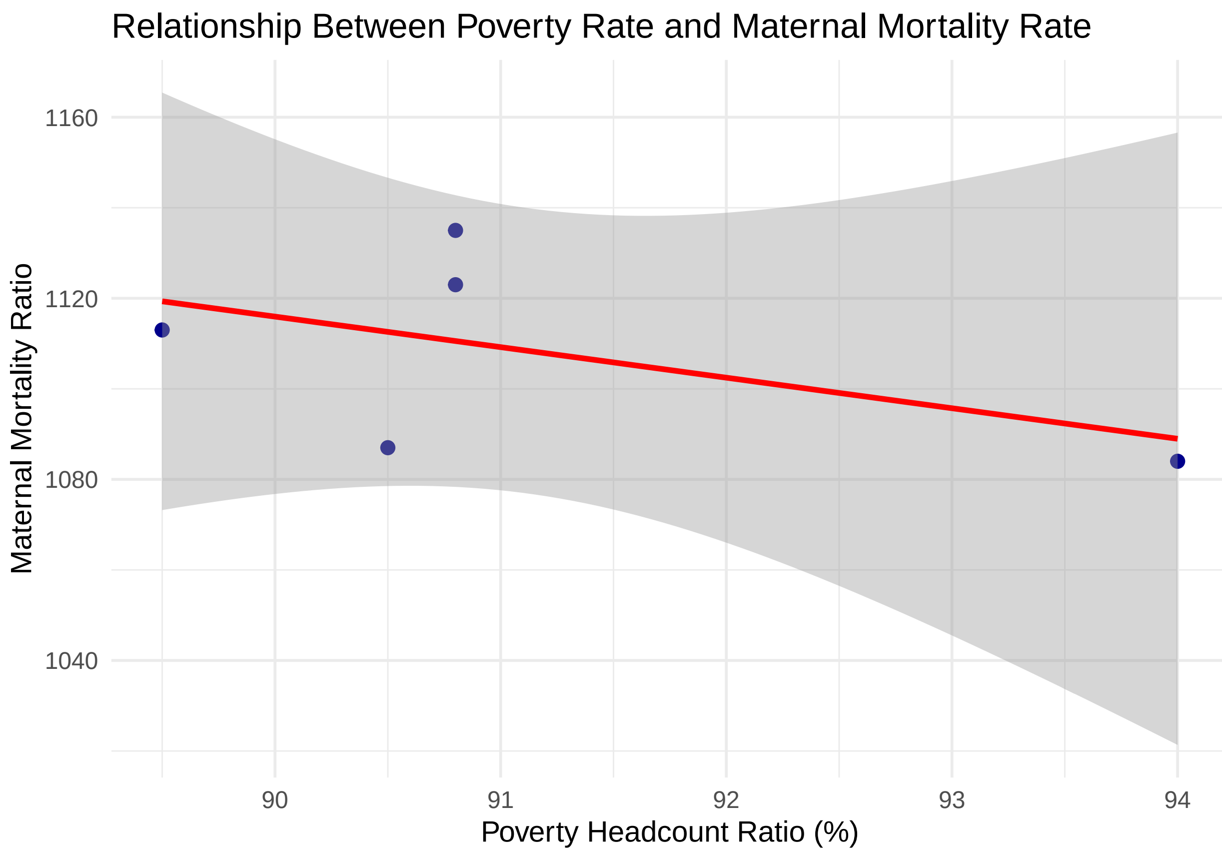

Figure 19: Relationship Between Poverty Rate and Maternal Mortality Rate

Analysis 8; Poverty Headcount Ratio and Under-5 Mortality

The analysis shows a strong positive correlation (0.782) between poverty headcount ratios and under-5 mortality rates, suggesting that higher poverty levels are associated with increased child mortality. Conversely, there is a moderate negative correlation (-0.662) between vaccination rates and under-5 mortality rates, demonstrating the mitigating impact of healthcare access. The regression model incorporating poverty and vaccination rates explains about 65% of the variability in under-5 mortality rates, with poverty showing a stronger influence. These results highlight the pressing need to address poverty and its cascading effects on health while bolstering immunization coverage to protect the most vulnerable populations.

question 9: How does public expenditure on health relate to maternal mortality rates over time, and what insights can be drawn from the trends and correlation between these two variables?

```
# ---- Load Required Libraries ----
library(ggplot2)  # For visualizations
library(dplyr)    # For data manipulation

# ---- Data Cleaning and Preprocessing ----

# Step 1: Filter the dataset to keep rows with no missing values in relevant columns
dataset2_clean_health <- dataset2 %>%
  filter(!is.na(Maternal_mortality_ratio) & !is.na(Domestic_general_government_health_expenditure_per_ca

# Step 2: Replace any missing health expenditure values with the median as a precautionary step
dataset2$Domestic_general_government_health_expenditure_per_capita[is.na(dataset2$Domestic_general_gover
  median(dataset2$Domestic_general_government_health_expenditure_per_capita, na.rm = TRUE)

# Step 3: Check the structure of the dataset (optional step for debugging)
str(dataset2_clean_health)
```

```
## tibble [21 x 74] (S3: tbl_df/tbl/data.frame)
##  $ Country_Name                                                                              : chr [1:2
##  $ Year                                                                                      : num [1:2
##  $ Population_male                                                                           : num [1:2
##  $ Population_female                                                                         : num [1:2
##  $ Population_total                                                                          : num [1:2
##  $ Urban_population                                                                          : num [1:2
##  $ Rural_population                                                                          : num [1:2
##  $ Population_growth                                                                         : num [1:2
##  $ Urban_population_growth                                                                   : num [1:2
##  $ Rural_population_growth                                                                   : num [1:2
##  $ Population_density                                                                        : num [1:2
##  $ Age_dependency_ratio                                                                      : num [1:2
##  $ Age_dependency_ratio_young                                                                : num [1:2
##  $ Age_dependency_ratio_old                                                                  : num [1:2
##  $ Number_deaths_ages_5-9_years                                                              : num [1:2
##  $ Mortality_rate_adult_male                                                                 : num [1:2
##  $ Mortality_rate_adult_female                                                               : num [1:2
##  $ Mortality_rate_under-5                                                                    : num [1:2
##  $ Mortality_rate_under-5_male                                                               : num [1:2
##  $ Mortality_rate_under5_female                                                              : num [1:2
##  $ Mortality_rate_infant                                                                     : num [1:2
##  $ Mortality_rate_infant_male                                                                : num [1:2
##  $ Mortality_rate_infant_female                                                              : num [1:2
##  $ Life_expectancy_atbirth                                                                   : num [1:2
##  $ Life_expectancy_at_birth_female                                                           : num [1:2
##  $ Life_expectancy_atbirth_male                                                              : num [1:2
##  $ Immunization_measles_children_ages_12-23_m                                                : num [1:2
##  $ Fertility_rate_total                                                                      : num [1:2
##  $ Adolescent_fertility_rate                                                                 : num [1:2
```

```
##  $ Sex_ratio_birth                                                                       : num [1::
##  $ Pregnant_women_receiving_prenatal_care                                                : num [1::
##  $ Survival_to_age_65_female                                                             : num [1::
##  $ Survival_to_age_65_male                                                               : num [1::
##  $ Death_rate_crude                                                                      : num [1::
##  $ Birth_rate_crude                                                                      : num [1::
##  $ Number_of_neonatal_deaths                                                             : num [1::
##  $ Number_of_infant_deaths                                                               : num [1::
##  $ Number_of_deaths_ages_10-14_years                                                     : num [1::
##  $ Mortality_from_CVD_cancer_diabetes_CRD_ages_30-70                                     : num [1::
##  $ Mortality_rate_attributed_tounintentional_poisoning_male                              : num [1::
##  $ Mortality_CVD-cancer-diabetes_ages_30-70_ female                                      : num [1::
##  $ Probability_of_dying_among_youth_ages_ 20-24_years                                    : num [1::
##  $ Probability_of_dying_among_children_ages_5-9_years                                    : num [1::
##  $ Maternal_mortality_ratio                                                              : num [1::
##  $ Mortality_rate_neonatal                                                               : num [1::
##  $ Suicide_mortality_rate_female                                                         : num [1::
##  $ Cause_of_death_by _communicable_diseases _and_nutritionconditions                     : num [1::
##  $ Lifetime_risk_of_maternal_death                                                       : num [1::
##  $ Prevalence_ of_anemia_among_pregnant_women                                            : num [1::
##  $ Nurses_and_midwives                                                                   : num [1::
##  $ Domestic_private_health _expenditure                                                  : num [1::
##  $ Domestic_general_government_health_expenditure_per_capita                             : num [1::
##  $ School_enrollment_primary_female                                                      : num [1::
##  $ Primary_education_pupils                                                              : num [1::
##  $ School_enrollment_primary_secondary                                                   : num [1::
##  $ GDP_per_capita_growth                                                                 : num [1::
##  $ GDP_growth                                                                            : num [1::
##  $ GDP                                                                                   : num [1::
##  $ Unemployment_male_%ofmalelaborforce                                                   : num [1::
##  $ Refugee_population_bycountryoforigin                                                  : num [1::
##  $ Refugee_population_bycountryofasylum                                                  : num [1::
##  $ Population_in_largest_city                                                            : num [1::
##  $ Proportion_population_pushedbelow_povertyline_byout-of-pockethealthcareexpenditure     : num [1::
##  $ Proportion_population_pushedfurtherbelow_povertylinebyout-of-pockethealthcareexpenditure: num [1::
##  $ Total_greenhouse_gas_emissions                                                        : num [1::
##  $ Children_employment_%_of_childrenages7-14                                             : num [1::
##  $ Poverty_headcount_ratio                                                               : num [1::
##  $ People_using_safely_managed_sanitation_rural_%ofruralpopulation                       : num [1::
##  $ Diarrhea_treatment_%ofchildrenunder5                                                  : num [1::
##  $ Female_genital_mutilation_prevalence_%                                                : num [1::
##  $ Use_insecticide-treated-bed-nets_%ofunder-5population                                 : num [1::
##  $ Immunization_DPT_%ofchildrenages12-23months                                           : num [1::
##  $ People_using_safely_managed_drinkingwaterservices_%ruralpopulation                    : num [1::
##  $ Life_expectancy_diff                                                                  : num [1::
```

```
# ---- Linear Regression Analysis ----

# Perform regression analysis: Maternal Mortality Rate vs. Health Expenditure
health_expenditure_regression <- lm(
  Maternal_mortality_ratio ~ Domestic_general_government_health_expenditure_per_capita,
  data = dataset2_clean_health
)
```

```r
# Print the summary of the regression model
summary(health_expenditure_regression)
```

```
##
## Call:
## lm(formula = Maternal_mortality_ratio ~ Domestic_general_government_health_expenditure_per_capita,
##     data = dataset2_clean_health)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -58.845 -15.521   1.308  15.800  29.077
##
## Coefficients:
##                                                            Estimate Std. Error
## (Intercept)                                               1163.4919    21.7365
## Domestic_general_government_health_expenditure_per_capita   -2.2202     0.7872
##                                                           t value Pr(>|t|)
## (Intercept)                                                 53.53   <2e-16 ***
## Domestic_general_government_health_expenditure_per_capita   -2.82   0.0109 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.08 on 19 degrees of freedom
## Multiple R-squared:  0.2951, Adjusted R-squared:  0.258
## F-statistic: 7.954 on 1 and 19 DF,  p-value: 0.01093
```

```r
# ---- Scatter Plot with Regression Line ----

# Create a scatter plot with a regression line
ggplot(dataset2_clean_health, aes(x = Domestic_general_government_health_expenditure_per_capita, y = Ma
  geom_point(color = "blue", size = 2) +  # Scatterplot points
  geom_smooth(method = "lm", color = "red", se = TRUE) +  # Regression line with confidence interval
  labs(
    title = "Relationship Between Maternal Mortality Rate and Health Expenditure",
    x = "Health Expenditure Per Capita (USD)",
    y = "Maternal Mortality Ratio (Per 100,000 Live Births)"
  ) +
  theme_minimal()
```

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

```r
# ---- Trend Analysis for Health Expenditure ----

# Step 1: Create a time series for health expenditure
health_expenditure_ts <- ts(
  dataset2_clean_health$Domestic_general_government_health_expenditure_per_capita,
  start = c(min(dataset2_clean_health$Year)),
  frequency = 1  # Annual data
)

# Step 2: Visualize the trend without decomposition
ggplot(dataset2_clean_health, aes(x = Year, y = Domestic_general_government_health_expenditure_per_capi
```

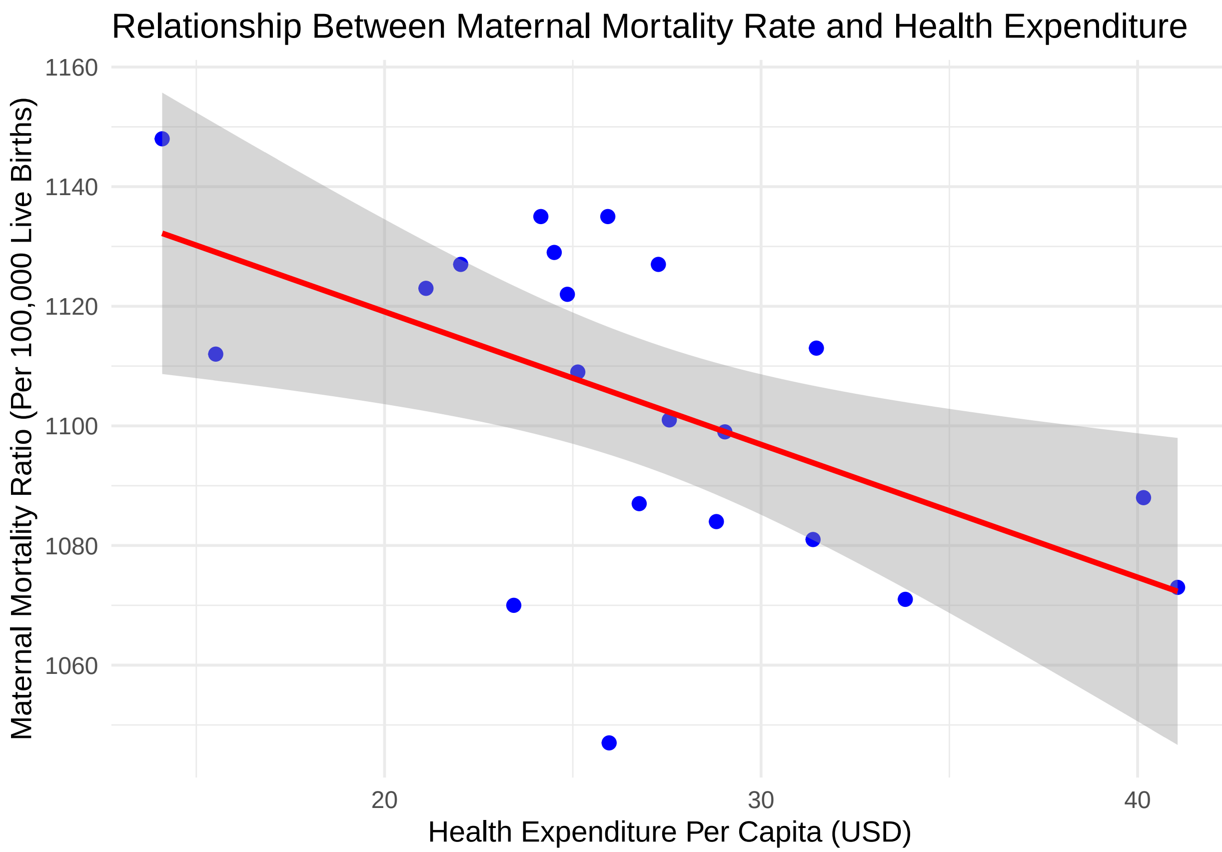

Figure 20: Relationship Between Maternal Mortality Rate and Health Expenditure

```r
  geom_line(color = "blue", size = 1) +  # Line plot for health expenditure
  geom_point(color = "red", size = 2) +  # Highlight individual data points
  labs(
    title = "Trend of Health Expenditure Over Time",
    x = "Year",
    y = "Health Expenditure Per Capita (USD)"
  ) +
  theme_minimal()
```

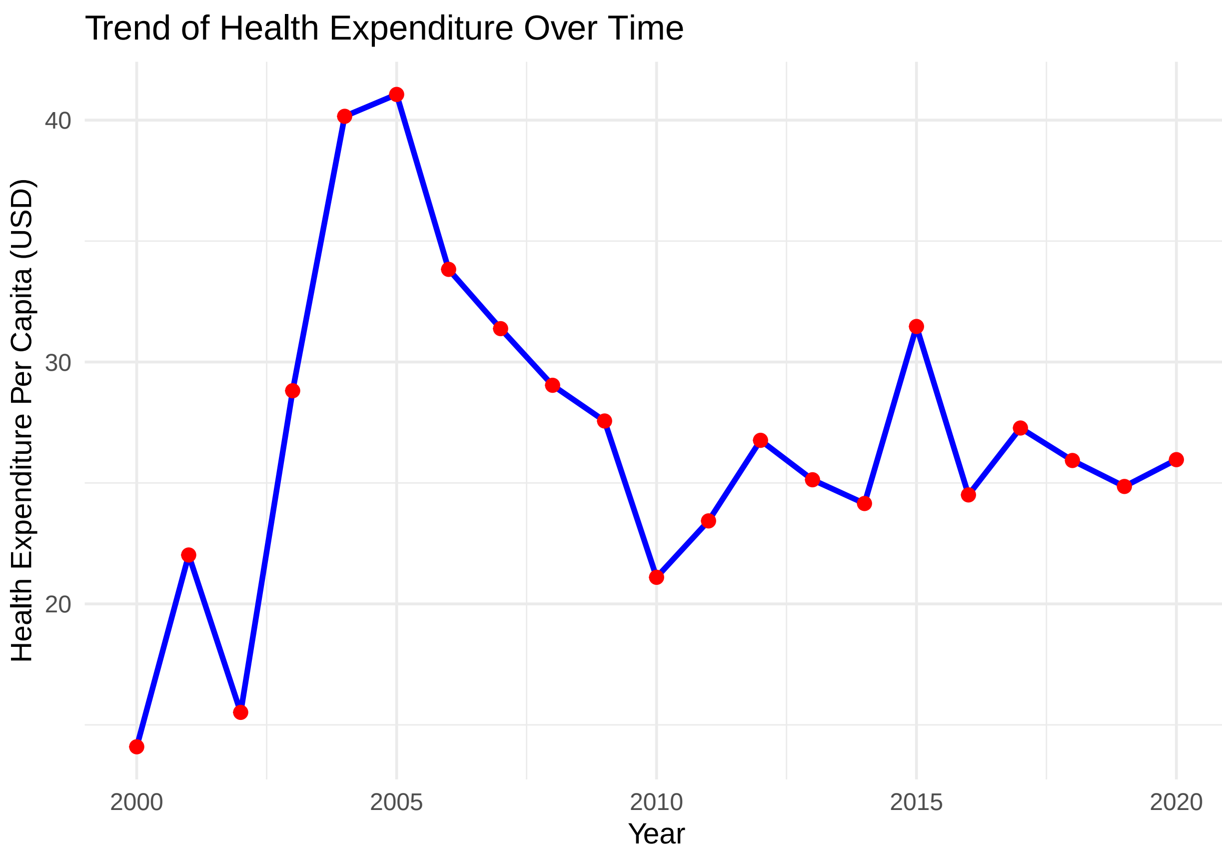

Figure 21: Trend of Health Expenditure Over Time

```r
# ---- Analysis ----

# Use the regression results and trend plot to interpret how public expenditure on health relates to ma
```

Conclusion

The analysis of mortality trends and influencing factors in Nigeria highlights the complexities of health outcomes in a rapidly evolving socio-economic environment. While some progress has been observed, such as the decline in maternal mortality and improvements in life expectancy, critical challenges remain. High maternal mortality rates, stark gender disparities in life expectancy, and the significant impact of poverty on under-5 mortality underscore systemic gaps in healthcare access and quality, as well as broader social inequities. Our analysis shows that indicators such as education levels, immunization rates, poverty headcount ratios, and healthcare workforce availability significantly contribute to mortality rates in Nigeria, either directly or indirectly.

To drive meaningful change, Nigeria must prioritize investments in healthcare infrastructure, education, and poverty alleviation. Efforts to strengthen immunization programs, improve maternal and child healthcare, and ensure equitable distribution of resources are essential to reducing mortality rates further. Integrating economic growth with targeted health and social policies will be crucial to overcoming these challenges and fostering sustainable improvements in population health.