Analysis of Health, Environmental, and Socioeconomic Factors Across Global Populations

BERNABEO Bernard
CAMOTA, King Eluard
SANDATE, Maria del Carmen
SINGSON, Trizzia Ellaine

Dataset Overview

- Covers **188** recognized countries
- Spans a **10-year** period (2012–2021)
- Includes **global health, demographic, economic, and environmental** metrics
- **1,880** total data points (188 countries × 10 years)

Data Source

Source:

Kaggle (Global Health and Development 2012–2021)

Compiled from:

World Bank, World Health Organization (WHO)

Includes:

29 variables covering key global metrics

1. Demographic & Geographic Information
2. Health Indicators
3. Economic Indicators
4. Social Development Indicators

29 variables covering key global metrics

1. **Demographic & Geographic Information**
    a. Country
    b. Total Population
    c. Female Population
    d. Male Population

29 variables covering key global metrics

2. **Health Indicators**
    a. Life Expectancy in Years (Male, Female)
    b. Infant Deaths per 1000 Live Births
    c. Hospital Beds Per 1000 People
    d. Immunization Rate (%)
    e. Tuberculosis Cases per 100,000
    f. Suicide Rate (%)
    g. Obesity Rate (%)
    h. Underweight Rate (%)
    i. Overweight Rate (%)

29 variables covering key global metrics

3. **Economic Indicators**
    a. Unemployment Rate
    b. GDP Per Capita
    c. Sanitary Expense Per GDP
    d. Sanitary Expense Per Capita
    e. Labour Force Total

29 variables covering key global metrics

3. **Social Development Indicators**
    a. Fertility Rate
    b. Urban Population (%)
    c. Alcohol Consumption Per Capita
    d. Water Access (%)
    e. CO_2 Exposure (%)
    f. Safe Water Access (%)

Data Bias

1. **Reporting Bias**
    - Different countries may use varying methodologies or definitions for the same indicators, leading to inconsistencies.
2. **Measurement Bias**
    - Differences in data collection methods between organizations (World Bank vs. WHO) may introduce inconsistencies.

Problem

Global health disparities persist, with significant variations in life expectancy and infant mortality across countries. Factors such as gender, socioeconomic conditions, and environmental influences contribute to these disparities, often leading to poorer health outcomes in low-income or underserved populations. This study analyzes a global health dataset to better understand these factors and identify potential solutions to improve health equity worldwide.

Objective 1

What are the differences in life expectancy between genders when categorized into distinct ranges?

Objective 2

What economic and social factors most influence lower life expectancy in gender?

Objective 3

How do socioeconomic and environmental factors impact infant mortality across countries over a 10-year period?

Data Cleaning Process

Handling Missing Data

a. **Grouping & Average by Country**
   - *<u>Why</u>*: Group data by country (10-year period) to capture country specific trends.
   - Aggregate using the median to reduce the impact of outliers.

b. **Select Relevant Variables**
   - *<u>Why</u>*: Focus only on parameters of interest to align with the specific objective.
   - Avoid dropping data if no parameters are excluded, preserving data integrity.

Systematic Literature Review

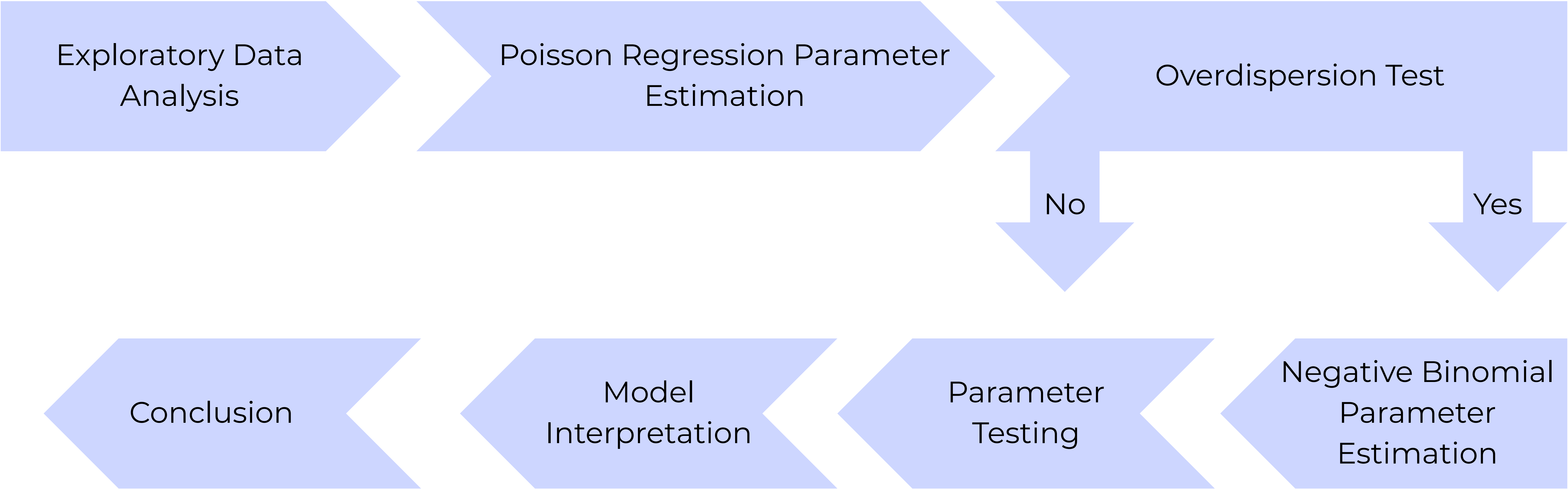
Fadhilah Fitri (2021) begins with **descriptive analysis** to summarize both response and explanatory variables, followed by exploratory analysis to examine correlations between variables.

Modeling Process:

1. **Poisson Regression**: The model parameters are estimated, and the Akaike Information Criterion (AIC) is calculated to evaluate model fit.
2. **Overdispersion Testing**: If the assumption that the mean equals the variance is violated, the study shifts to Negative Binomial Regression to handle overdispersion.
3. **Negative Binomial Regression**: Model parameters are estimated, and regression parameter tests (both simultaneous and partial) are conducted.
4. **Model Comparison**: The AIC values of both models are compared, selecting the model with the lower AIC as the better fit.
5. **Interpretation & Conclusion**: The final model is analyzed, and conclusions and recommendations are made based on the findings.

This statistical approach ensures robust modeling of infant mortality data by addressing overdispersion and selecting the best-fitting model.

Systematic Literature Review



Fadhilah Fitri (2021) Flow Chart

Objective 1:

What are the differences in life expectancy between genders when categorized into distinct ranges?

Objective 1: What are the differences in life expectancy between genders when categorized into distinct ranges?

Categorization of Life Expectancy (Female & Male)

Category	Definition (in years)
Low	< 60
Medium	60 – 74.9
High	≥ 75

Summary Statistics - Median Life Expectancy

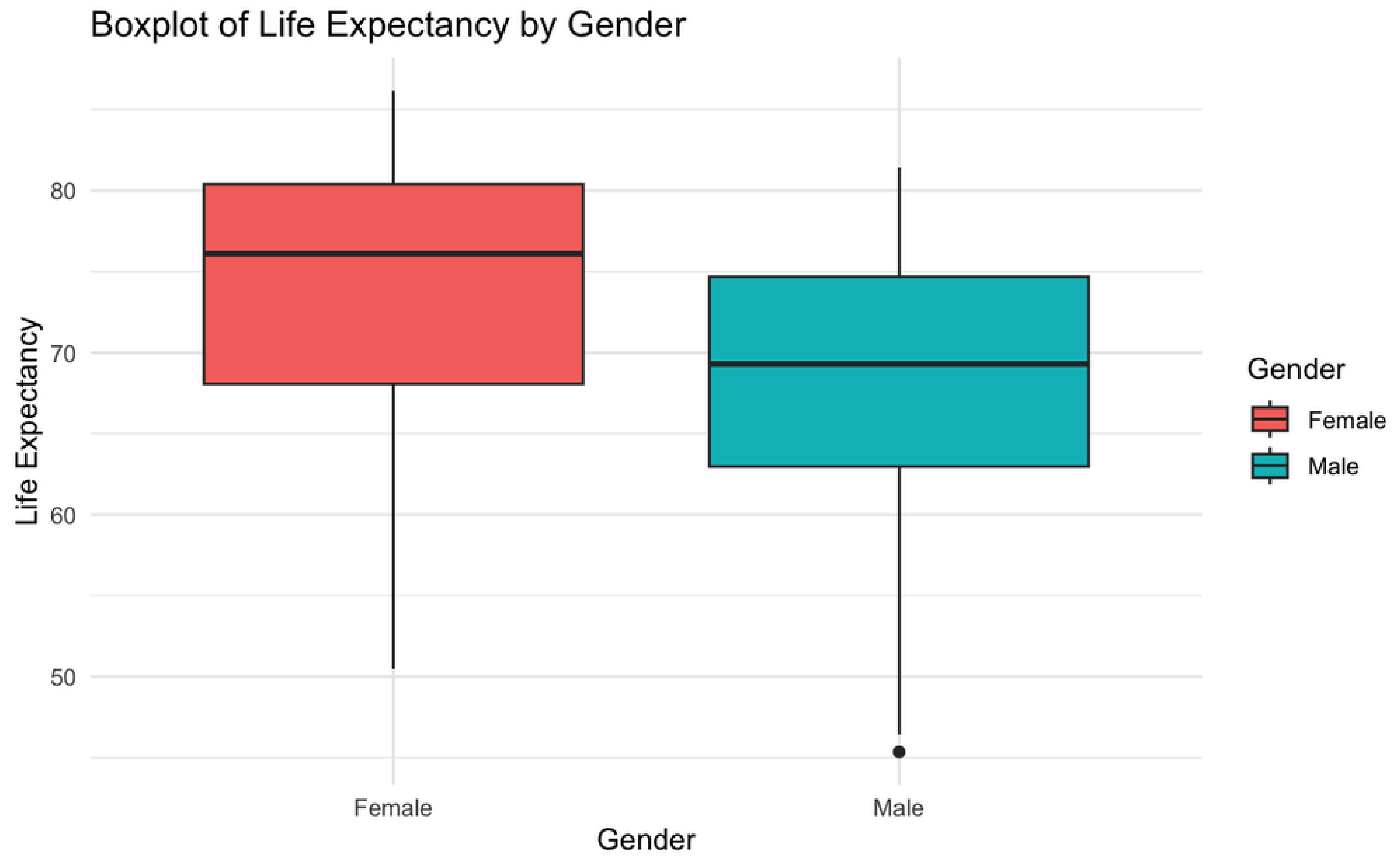
76.09 years

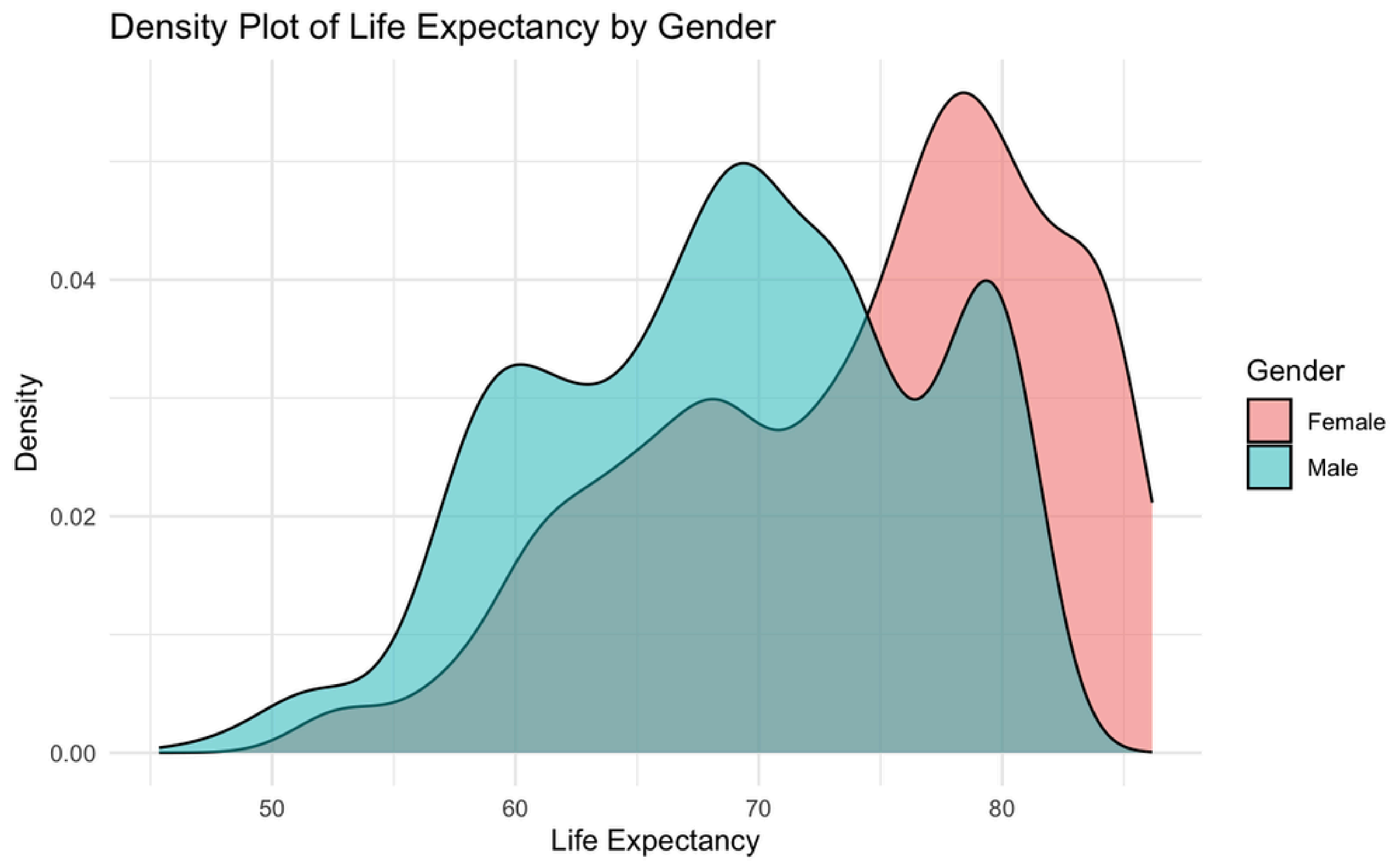
FEMALE

69.30 years

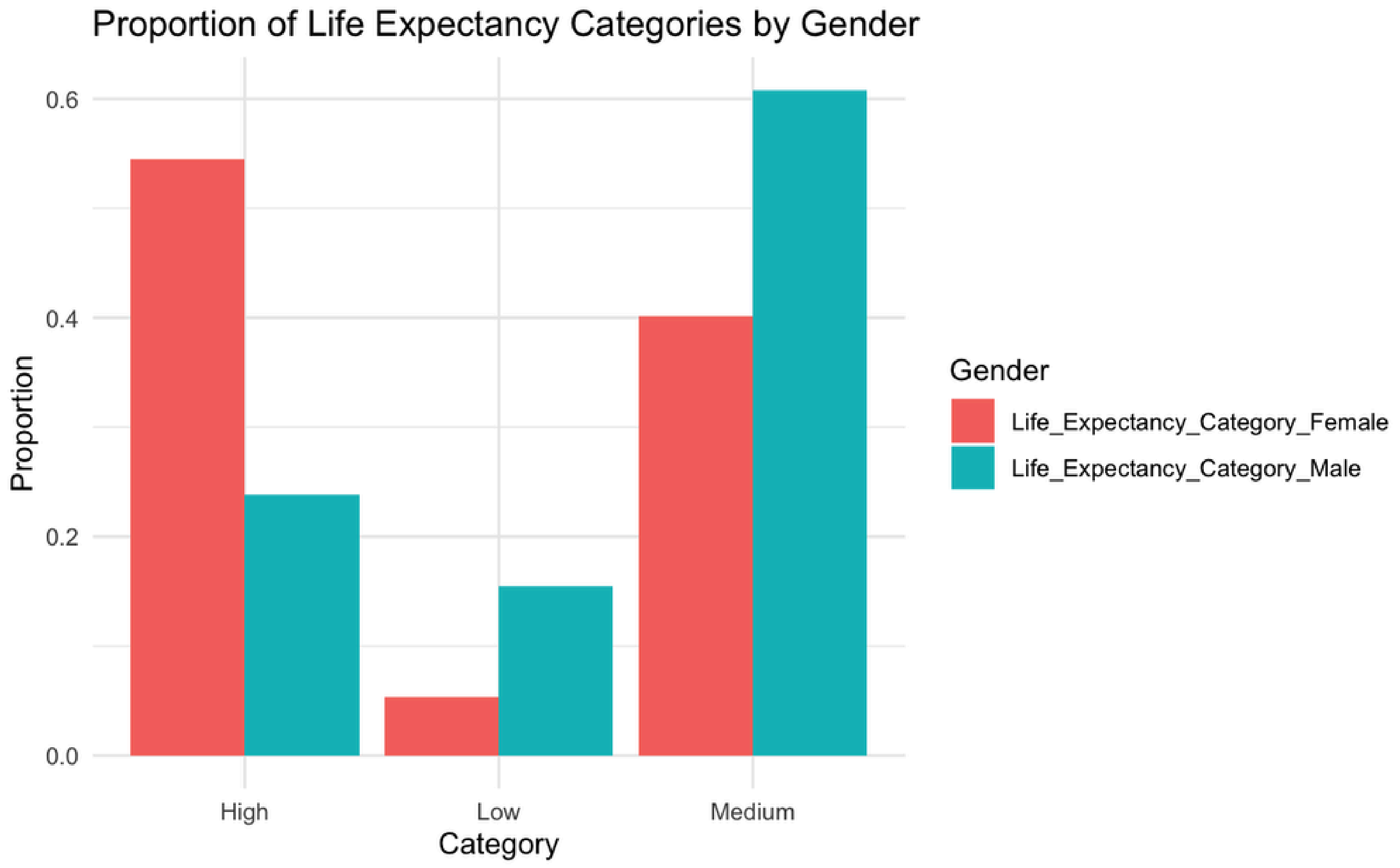
MALE

Objective 1: What are the differences in life expectancy between genders when categorized into distinct ranges?





Objective 1: What are the differences in life expectancy between genders when categorized into distinct ranges?



Frequency Table

Category	Female	Male
High	1002	438
Medium	739	1118
Low	99	284

Females

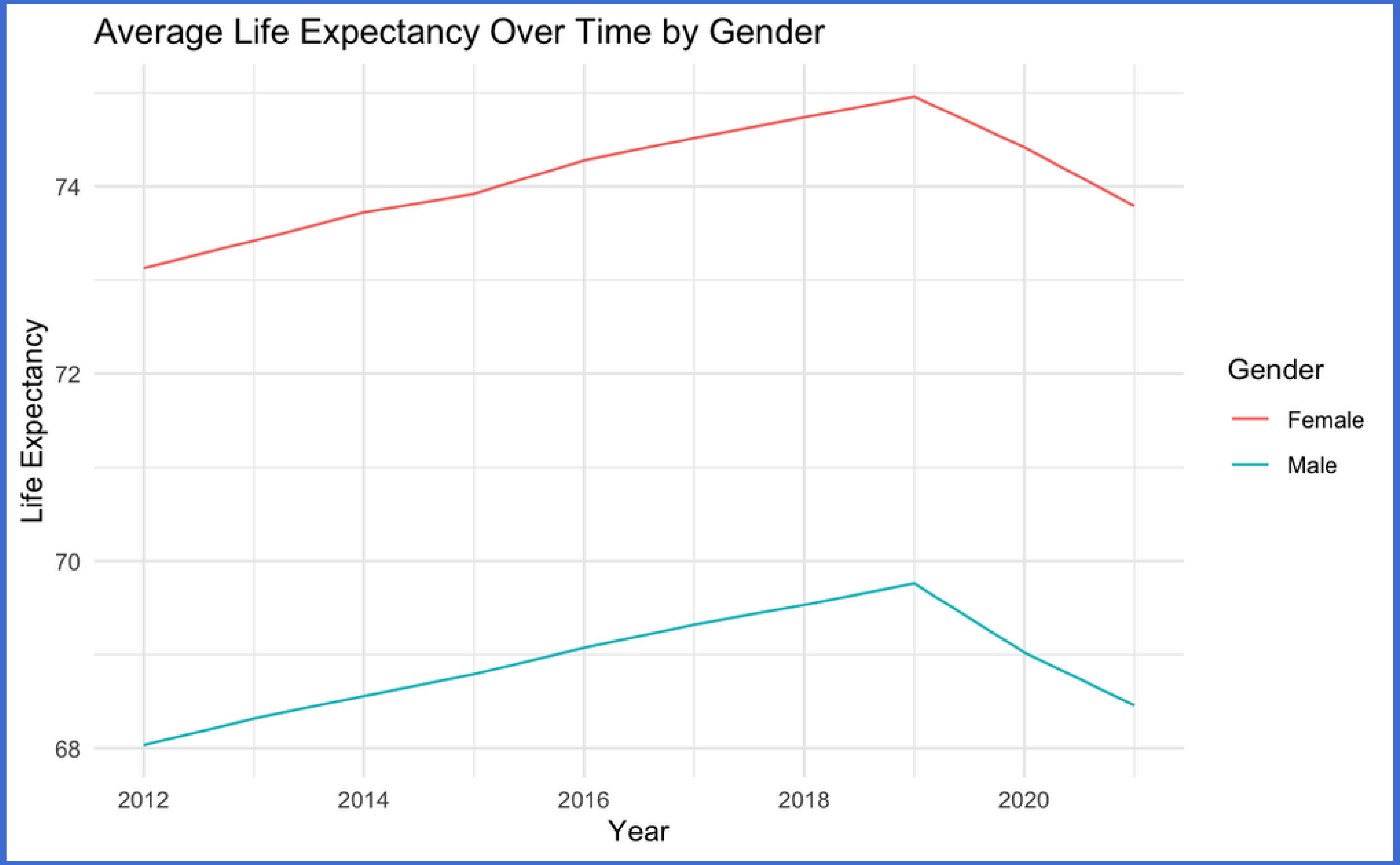
likely to be in the **“High”** life expectancy category.

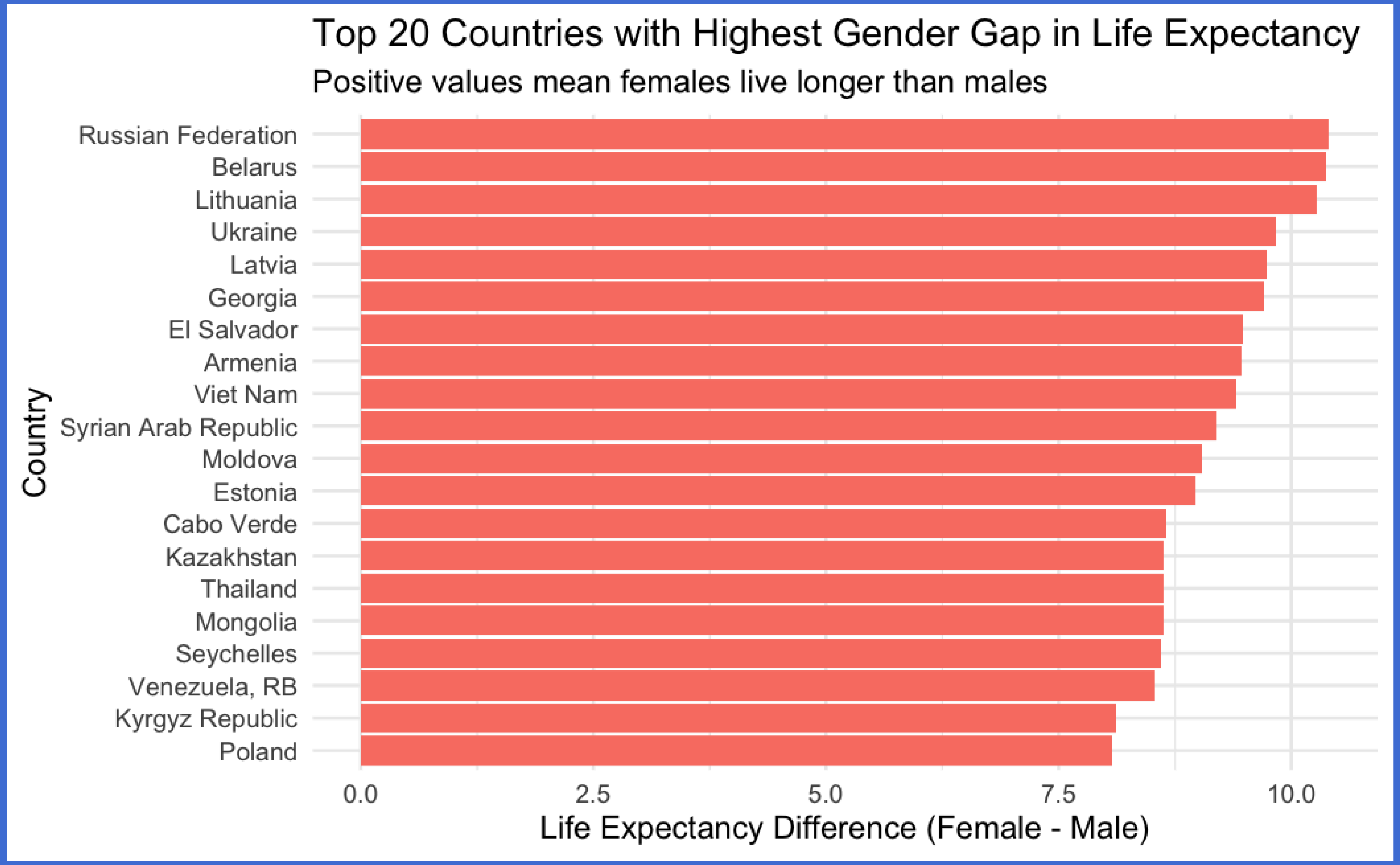
Males

dominate the “**Medium”** and **“Low”** categories.

Gender differences persist even when categorized - **females** have a greater chance of being in the **high longevity group**.

Objective 1: What are the differences in life expectancy between genders when categorized into distinct ranges?





Objective 1: What are the differences in life expectancy between genders when categorized into distinct ranges?

Paired t-test

Are female & male life expectancies different?

- Mean difference ≈ 5.20 years
- p-value < 2.2e-16 → Statistically significant

💡 There's strong evidence that **females** live longer than **males**, and the difference is not due to random chance.

ANOVA

- F = 398.3, p-value < 2e-16 ***

💡 **Gender** is a significant factor affecting life expectancy across the dataset.

Objective 1: What are the differences in life expectancy between genders when categorized into distinct ranges?

Binomial Test

Is ≥ 50% of females in High Life Expentancy category?

💡 A significantly greater-than-50% share of females fall into the High life expectancy category.

Proportion / Score Test

Do female and male populations differ significantly in their likelihood of being in the High category?

💡 p-value < 2.2e-16. There is strong statistical evidence that females are more likely to fall into the High life expectancy group than males.

Objective 1: What are the differences in life expectancy between genders when categorized into distinct ranges?

Chi-square Test

Is there an association between **gender and life expectancy category** (Low/Medium/High)?

💡 Since the p-value (< 2.2e-16) is far below **0.05,** we reject the null hypothesis that gender and life expectancy category are independent.

💡 There is a statistically significant association between gender and life expectancy category.

💡 The distribution of Low, Medium, and High categories differs significantly between males and females.

Result and Interpretation

💡 Gender is a significant determinant of life expectancy. Globally, females are more likely to live longer than males.

Females are significantly more likely to be in the High Life Expectancy category compared to males.

Across all statistical tests (binomial, t-test, chi-square, etc.), the results consistently show a gender gap in favor of females.

On average, females live ~5 years longer than males.

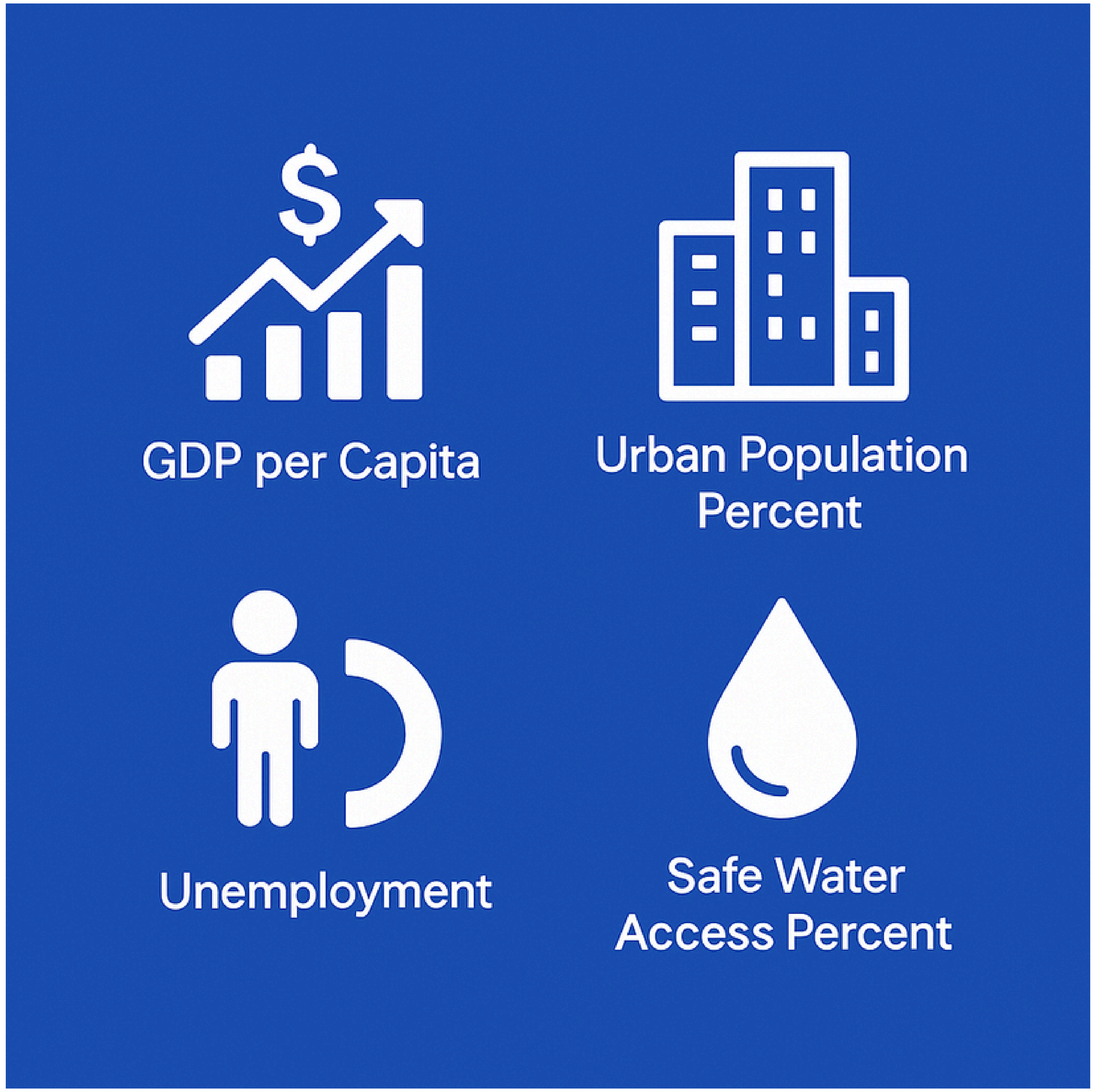
The proportion of females in the High category is 54.5%, while only 25.8% of males fall into this group.

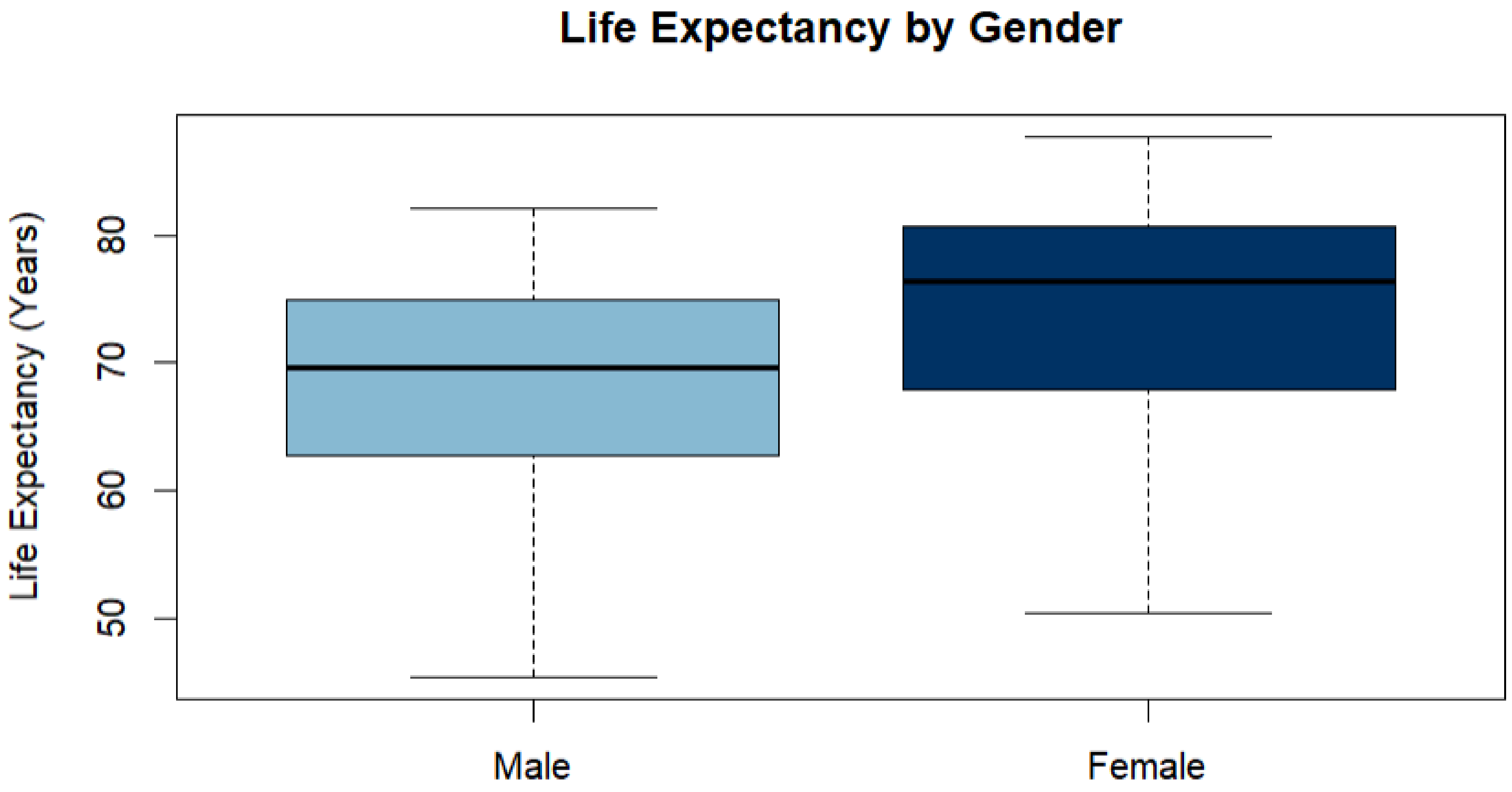
Boxplots & Density plots: Show a consistent advantage in female life expectancy

Objective 2:

What economic and social factors most influence lower life expectancy in gender?

Objective 2: What economic and social factors most influence lower life expectancy in gender?





Chi-square test -FEMALE

Factor	Chi-Square Statistic	p-value	Conclusion
GDP per Capita	630.32	< 2.2e-16	Significant association
Urban Population	566.96	< 2.2e-16	Significant association
Unemployment Rate	37.92	7.37e-10	Significant association
Safe Water Access	627.12	< 2.2e-16	Significant association

All four socioeconomic factors show a statistically significant relationship with female life expectancy.

Countries with higher GDP, urbanization, and water access, tend to have higher female life expectancy.

Countries with higher unemployment rates tend to have higher female life expectancy. This result may reflect underlying social or economic factors not directly measured, such as access to healthcare, welfare systems, or regional characteristics.

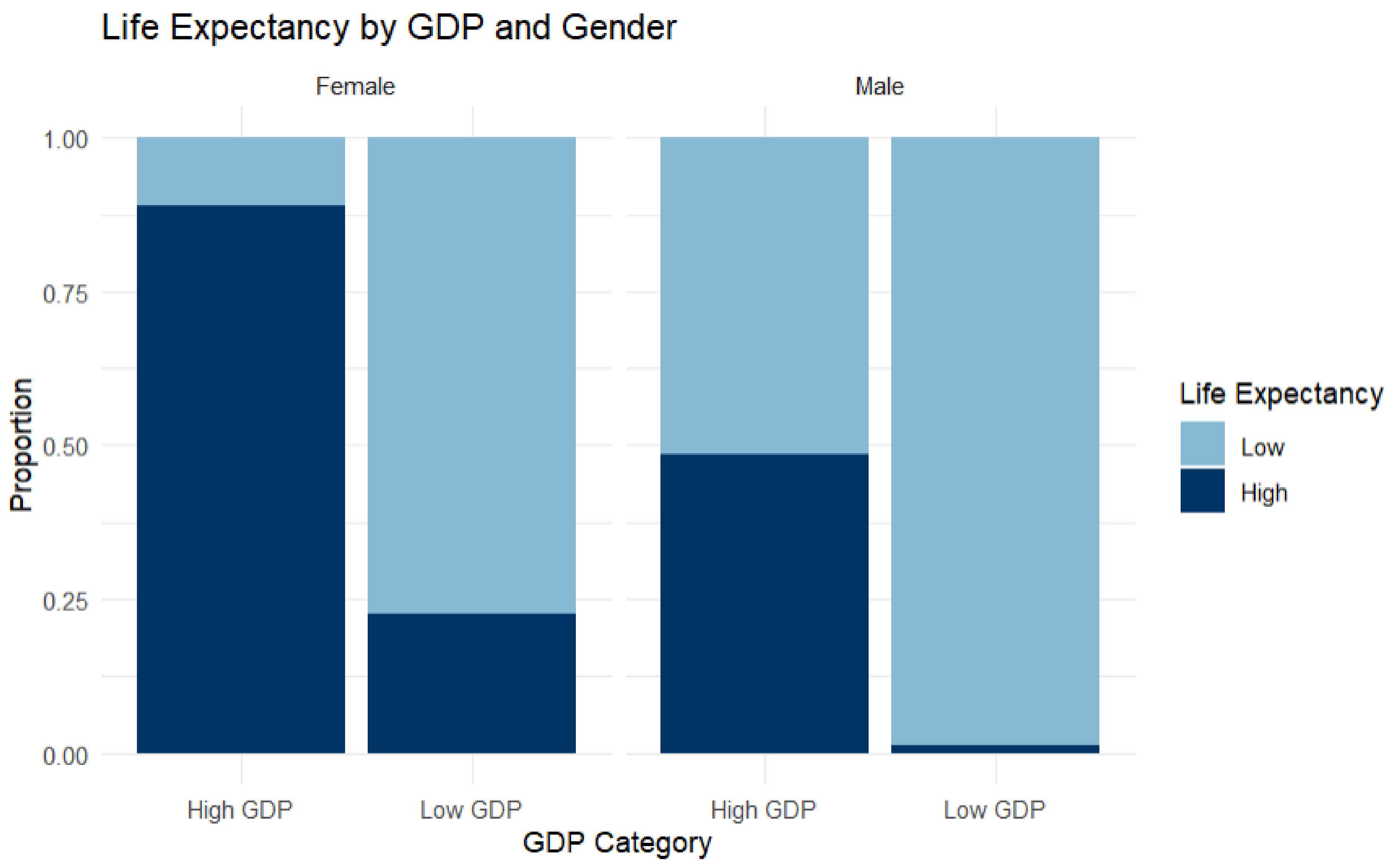
Chi-square test -MALE

Factor	Chi-Square Statistic	p-value	Conclusion
GDP per Capita	457.93	< 2.2e-16	Significant association
Urban Population	331.82	< 2.2e-16	Significant association
Unemployment Rate	0.087	0.7682	No significant association
Safe Water Access	390.01	< 2.2e-16	Significant association

Three of the four factors (GDP, urbanization, and access to safe water) show a strong statistical association with male life expectancy.

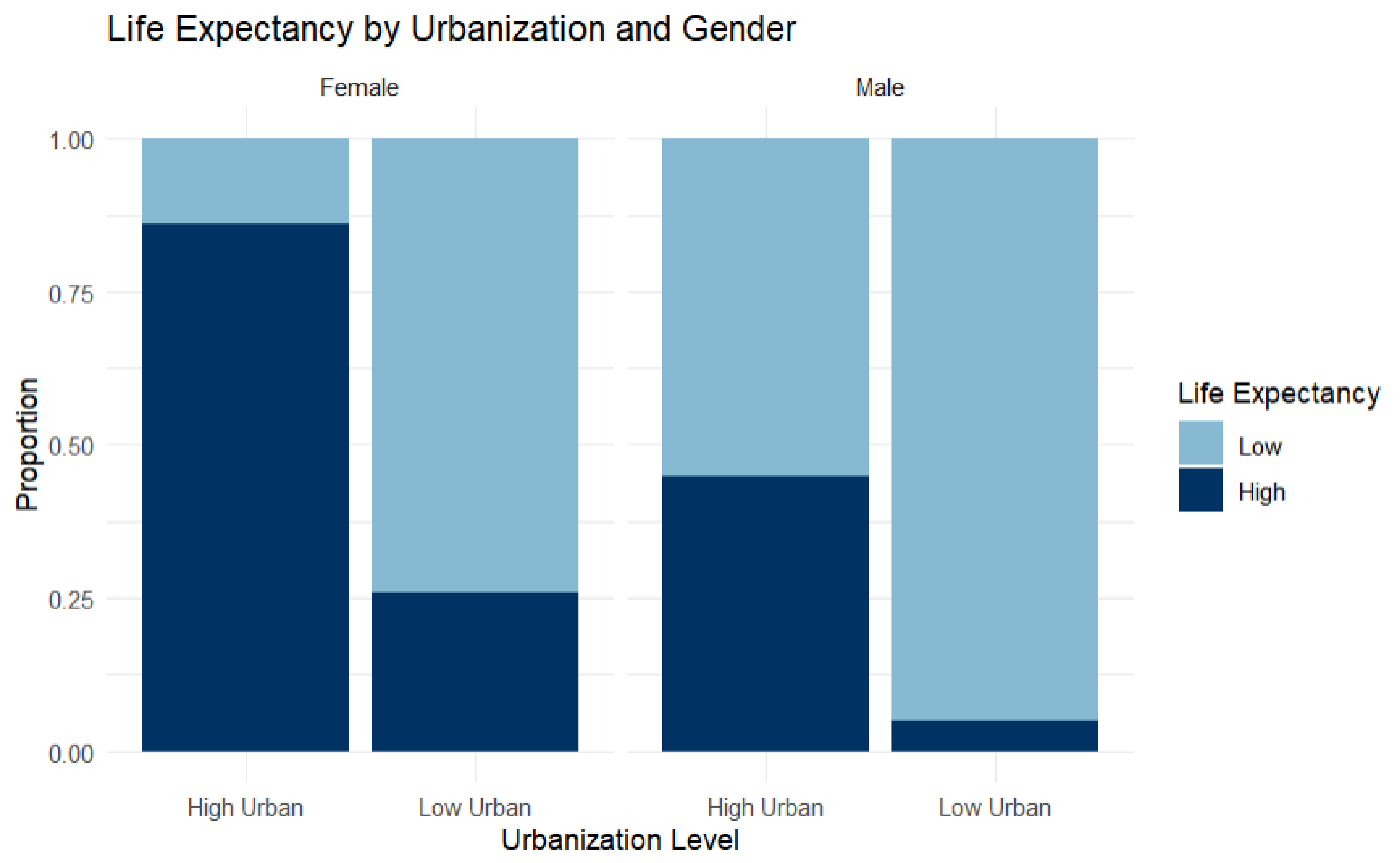
However, unemployment rate does not appear to be related to male life expectancy.

Impact of GDP on Life Expectancy by Gender



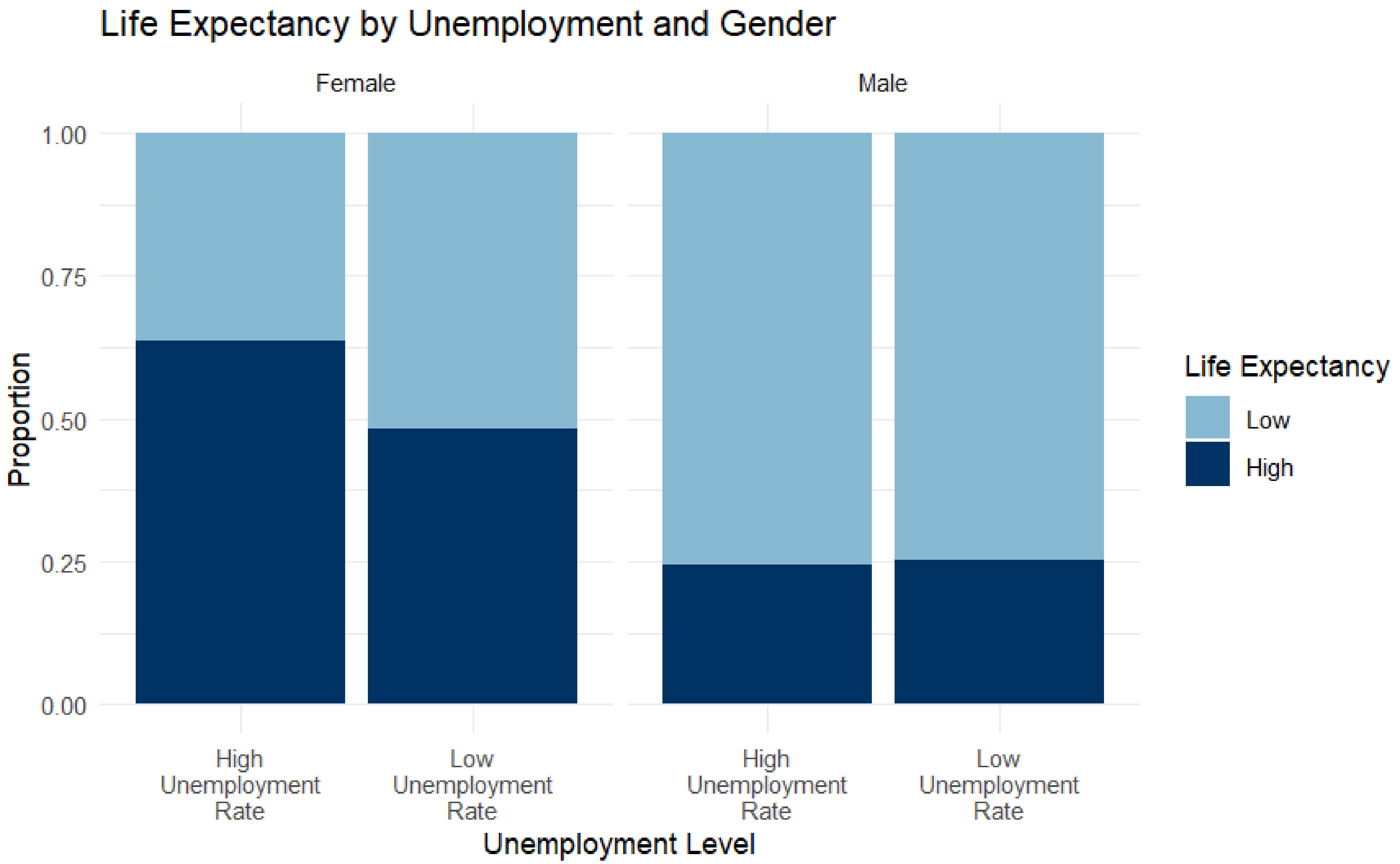
- Male life expeHigher GDP is linked to higher life expectancy in both genders.
- The effect is stronger in females, with a clear GDP-based gap.
- Male life expectancy stays mostly low, even in high GDP countries → possible gender gap.

Life Expectancy by Urbanization and Gender



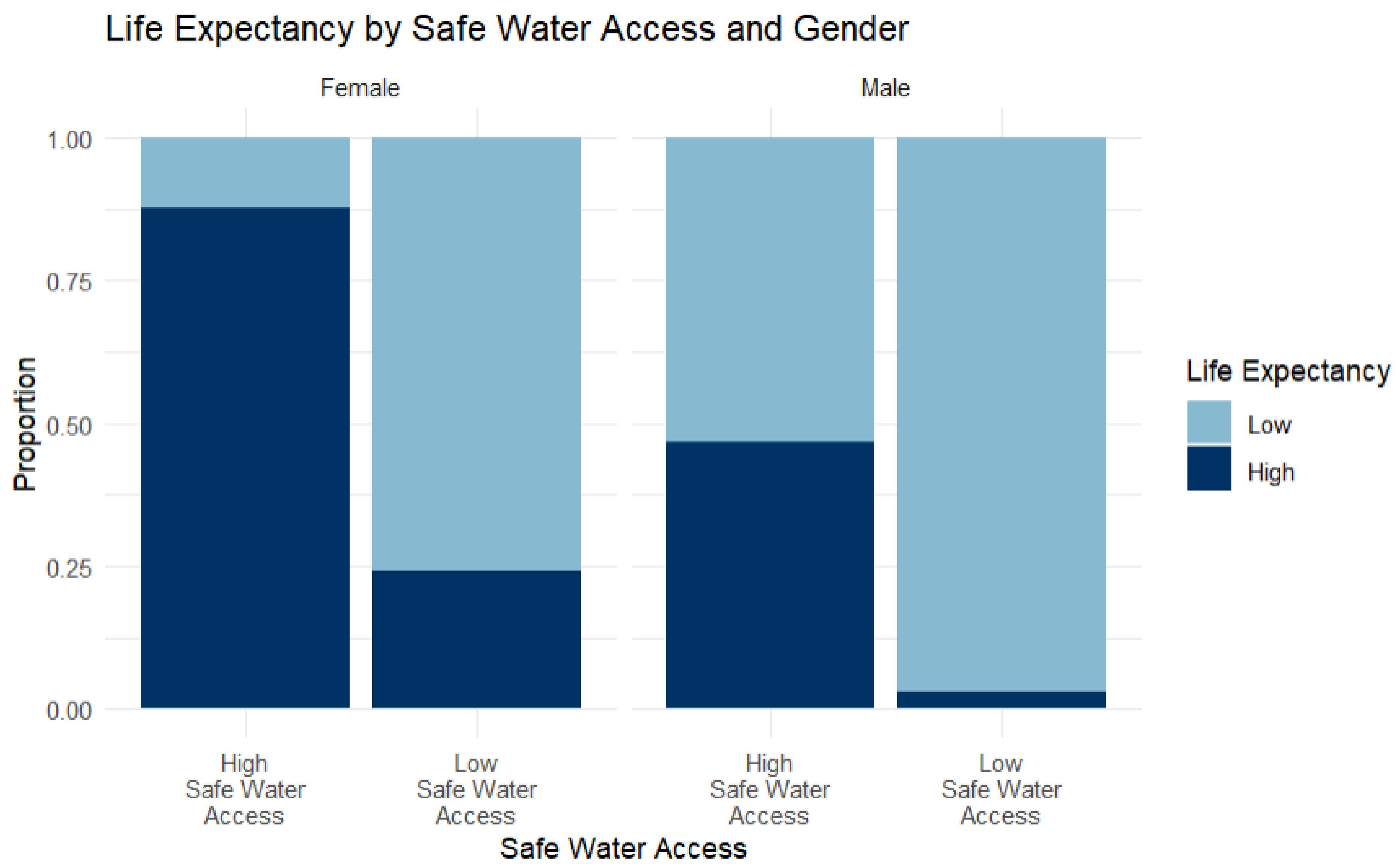
- High urbanization is linked to higher life expectancy, especially in females.
- Women in high urban countries show the highest proportion of long life expectancy.
- For males, the effect exists but is less pronounced.
- Low urbanization is associated with lower life expectancy in both genders, more so for males.

Life Expectancy by Unemployment and Gender



- For females, high unemployment is surprisingly linked to a higher proportion of high life expectancy.
- Low unemployment in females shows a more balanced split between high and low life expectancy.
- For males, life expectancy is low in both unemployment categories — no major difference.
- Suggests that unemployment rate has little or no impact on male life expectancy

Life Expectancy by Safe Water Access



- Female life expectancy is much higher in countries with high access to safe water.
- For males, the difference is also clear, though less dramatic.
- Low access to safe water is strongly associated with low life expectancy in both genders.
- Safe water access appears to be one of the strongest predictors of life expectancy across genders.

Logistic Regression: Predictors of Female Life Expectancy

$$\log\left(\frac{\text{P(High Life Expectancy)}}{\text{1−P(High Life Expectancy}}\right) = 3.5012 - 1.4576 \cdot \text{Low GDP} - 2.2627 \cdot \text{Low Urban} - 2.4548 \cdot \text{Low Water Access} + 0.2587 \cdot \text{Low Unemployment}$$

Effect of GDP on Female Life Expectancy

- Holding all other variables constant, the odds of having high female life expectancy in countries with low GDP are approximately 77% lower than in countries with high GDP. Strong evidence of a GDP effect ($p < 2e-16$).

Logistic Regression: Predictors of Female Life Expectancy

$$\log\left(\frac{\text{P(High Life Expectancy)}}{1-\text{P(High Life Expectancy}}\right) = 3.5012 - 1.4576\cdot\text{Low GDP} - 2.2627\cdot\text{Low Urban}$$
$$-2.4548\cdot\text{Low Water Access} + 0.2587\cdot\text{Low Unemployment}$$

Effect of Urbanization on Female Life Expectancy

- Countries with low urbanization have about 90% lower odds of high female life expectancy compared to highly urbanized countries, after controlling for other variables. Strong urbanization effect (p = 1.15e-15).

Logistic Regression: Predictors of Female Life Expectancy

$$\log\left(\frac{\text{P(High Life Expectancy)}}{\text{1−P(High Life Expectancy}}\right) = 3.5012 - 1.4576 \cdot \text{Low GDP} - 2.2627 \cdot \text{Low Urban}$$
$$- 2.4548 \cdot \text{Low Water Access} + 0.2587 \cdot \text{Low Unemployment}$$

Effect of Urbanization on Female Life Expectancy

- The odds of high female life expectancy in countries with poor access to safe water are about 91% lower than those with good access, controlling for GDP, urbanization, and unemployment. Very strong effect ($p < 2e{-16}$).

Logistic Regression: Predictors of Female Life Expectancy

$$\log\left(\frac{\text{P(High Life Expectancy)}}{\text{1−P(High Life Expectancy}}\right) = 3.5012 - 1.4576\cdot\text{Low GDP} - 2.2627\cdot\text{Low Urban}$$
$$-2.4548\cdot\text{Low Water Access} + 0.2587\cdot\text{Low Unemployment}$$

Effect of Unemployment on Female Life Expectancy

- Unemployment effect is not statistically significant ($p = 0.139$)

Logistic Regression: Predictors of Male Life Expectancy

$$\log\left(\frac{P(\text{High Life Expectancy})}{1-P(\text{High Life Expectancy})}\right) = 0.3486-2.6683\cdot\text{Low GDP}-1.8758\cdot\text{Low Urban}$$
$$-2.4044\cdot\text{Low Water Access}+0.7705\cdot\text{Low Unemployment}$$

Effect of GDP on Male Life Expectancy

- Holding all other variables constant, countries with low GDP have 93% lower odds of high male life expectancy compared to those with high GDP. Strong GDP effect (p =5.79e-15).

Logistic Regression: Predictors of Male Life Expectancy

$$\log\left(\frac{P(\text{High Life Expectancy})}{1-P(\text{High Life Expectancy})}\right) = 0.3486 - 2.6683 \cdot \text{Low GDP} - 1.8758 \cdot \text{Low Urban} - 2.4044 \cdot \text{Low Water Access} + 0.7705 \cdot \text{Low Unemployment}$$

Effect of Urbanization on Male Life Expectancy

- Countries with low urbanization have 85% lower odds of high male life expectancy than highly urbanized countries, adjusting for other variables. Strong urbanization effect ($p < 2e\text{-}16$).

Logistic Regression: Predictors of Male Life Expectancy

$$\log\left(\frac{P(\text{High Life Expectancy})}{1-P(\text{High Life Expectancy})}\right)=0.3486-2.6683\cdot\text{Low GDP}-1.8758\cdot\text{Low Urban}$$
$$-2.4044\cdot\text{Low Water Access}+0.7705\cdot\text{Low Unemployment}$$

Effect of Urbanization on Male Life Expectancy

- Countries with poor access to safe water have 91% lower odds of high male life expectancy than countries with good access. Very strong effect ($p < 2e\text{-}16$).

Logistic Regression: Predictors of Male Life Expectancy

$$\log\left(\frac{P(\text{High Life Expectancy})}{1-P(\text{High Life Expectancy})}\right)=0.3486-2.6683\cdot\text{Low GDP}-1.8758\cdot\text{Low Urban}$$
$$-2.4044\cdot\text{Low Water Access}+0.7705\cdot\text{Low Unemployment}$$

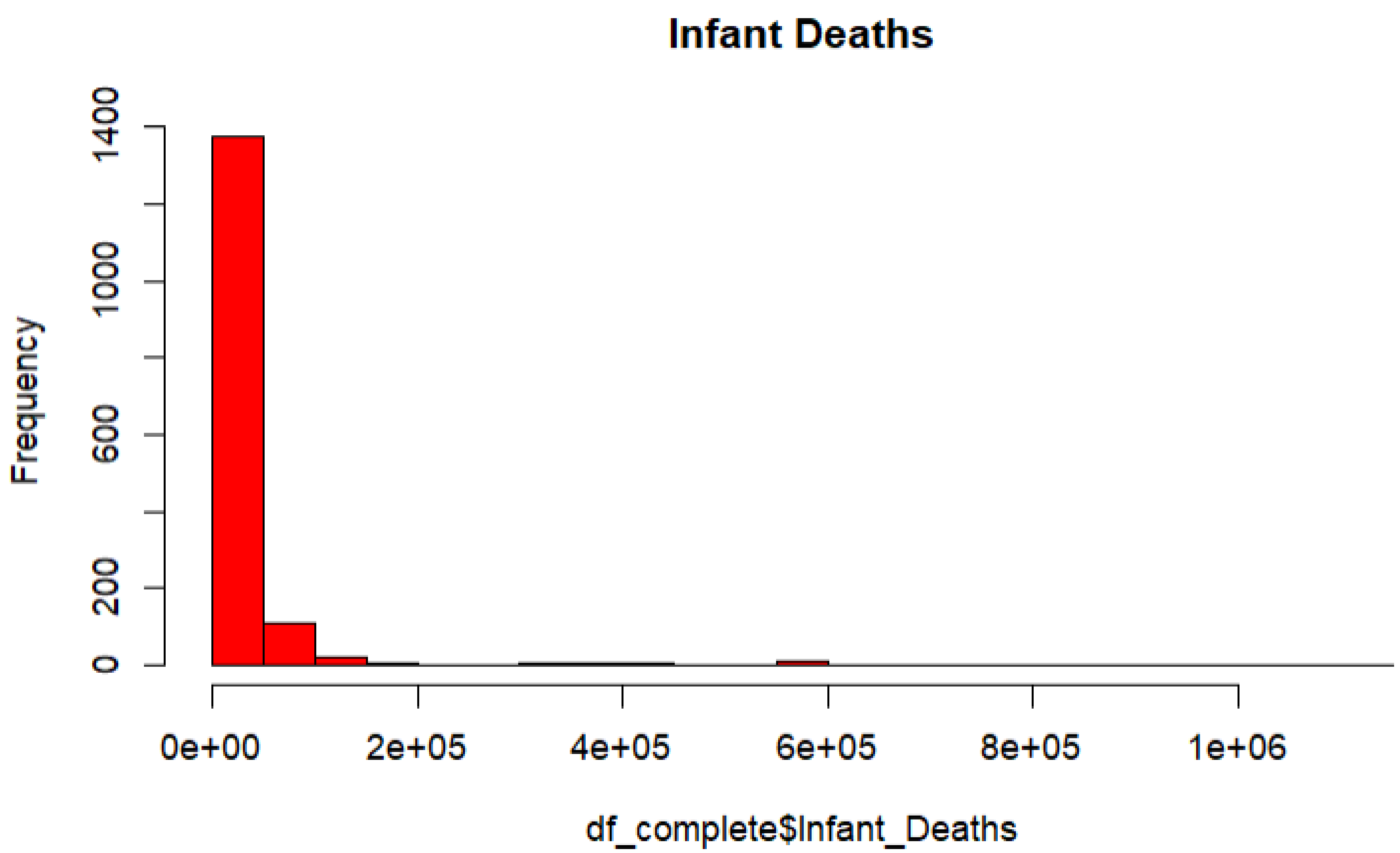
Effect of Unemployment on Male Life Expectancy

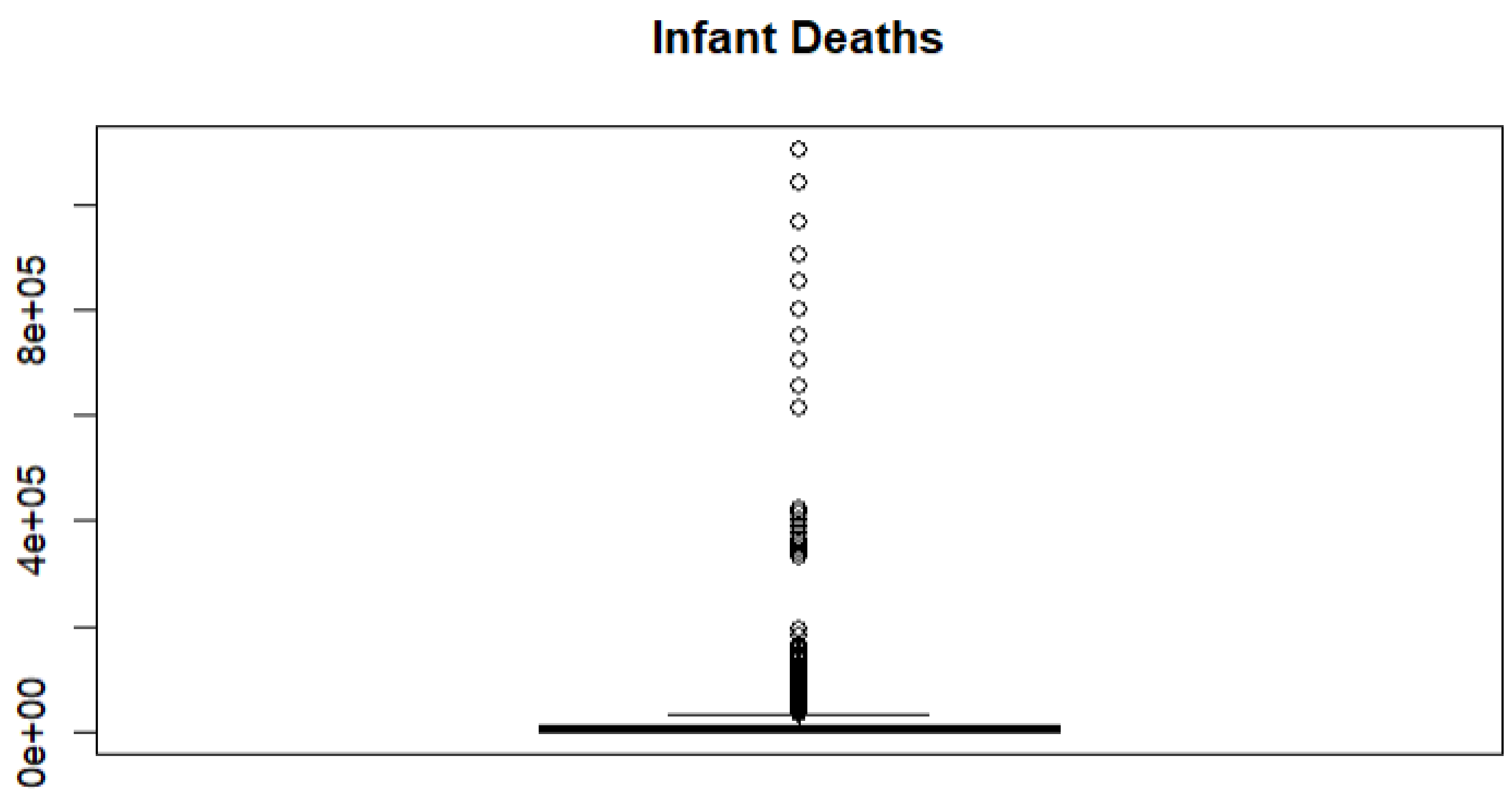
- Countries with low unemployment have more than twice the odds (OR ≈ 2.16) of high male life expectancy compared to those with high unemployment. This time, the effect is statistically significant (p = 3.00e-06)

Objective 3:

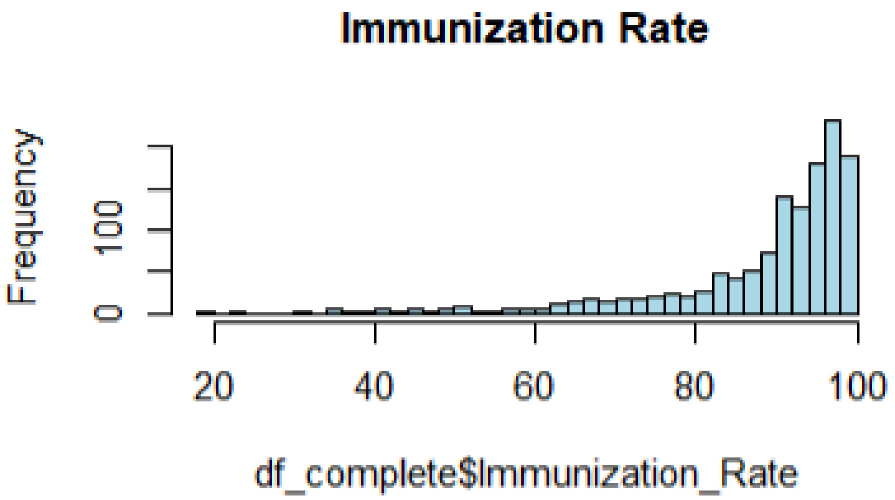
How do socioeconomic and environmental factors impact infant mortality across countries over a 10-year period?

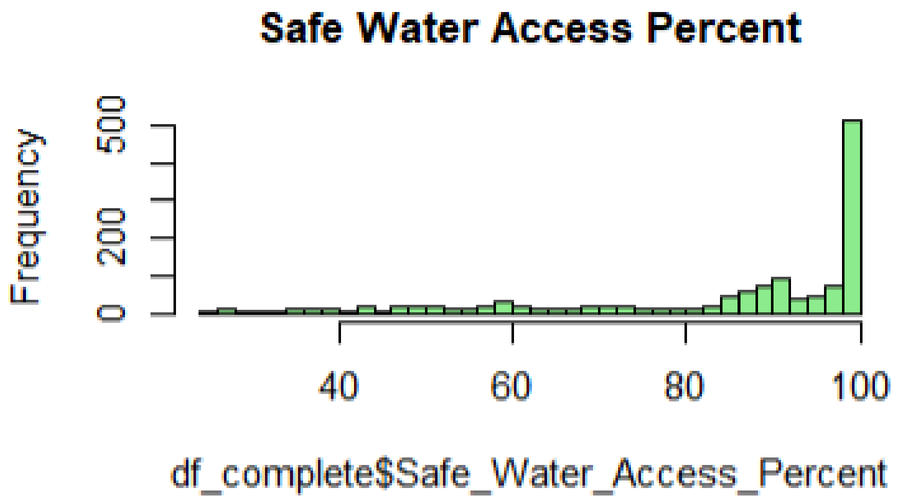
Objective 3: How do socioeconomic and environmental factors impact infant mortality across countries over a 10-year period?

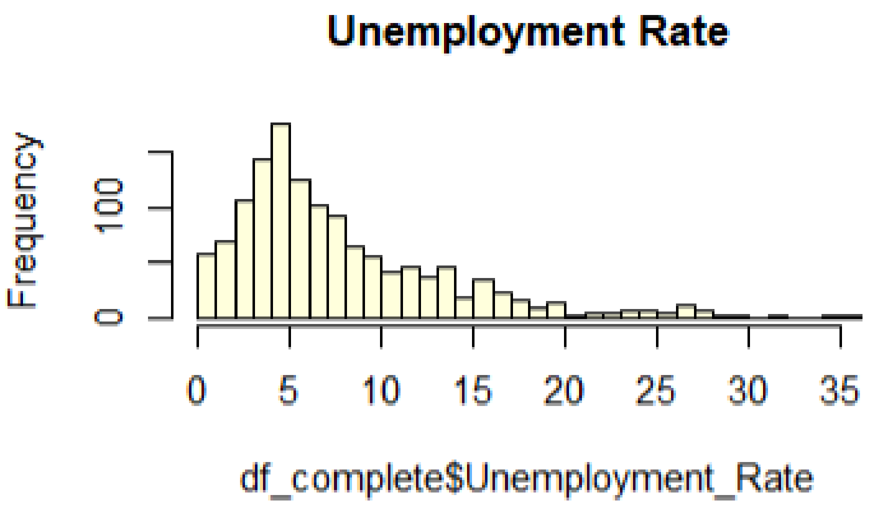


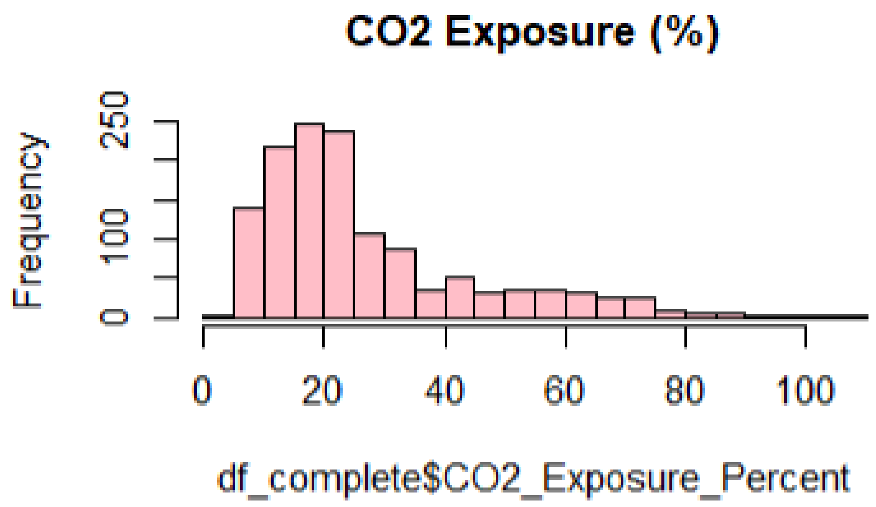


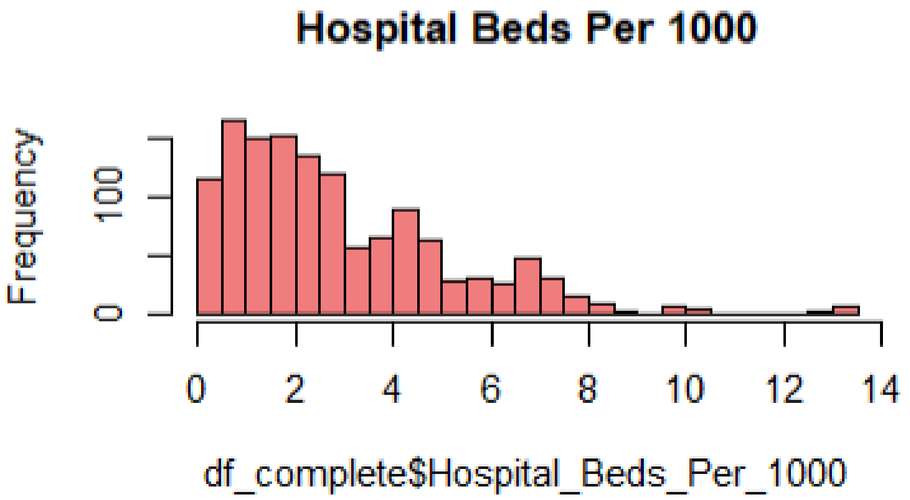
Objective 3: How do socioeconomic and environmental factors impact infant mortality across countries over a 10-year period?

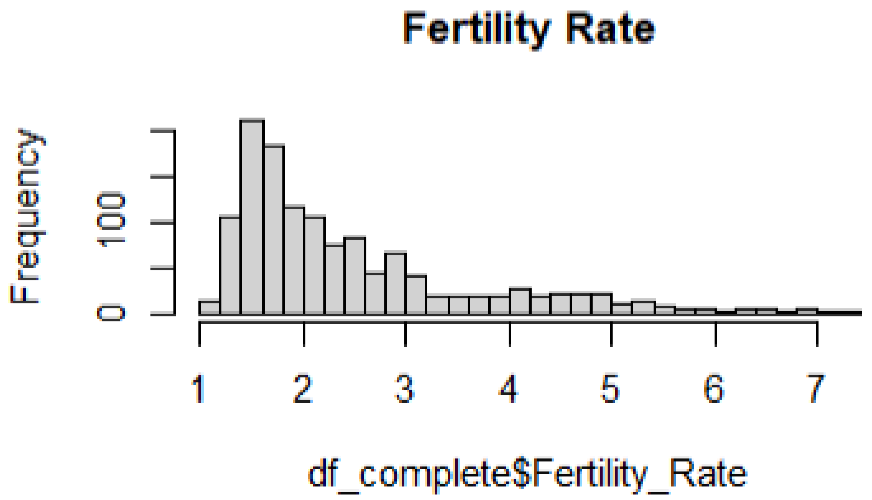


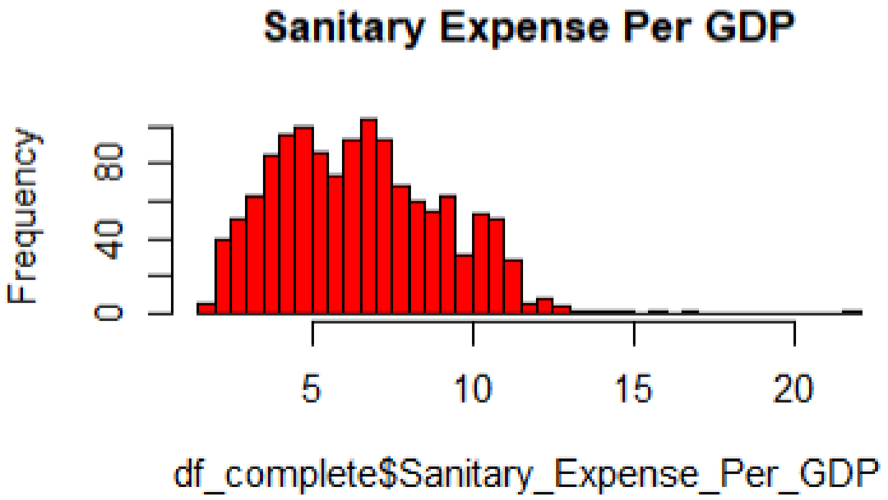




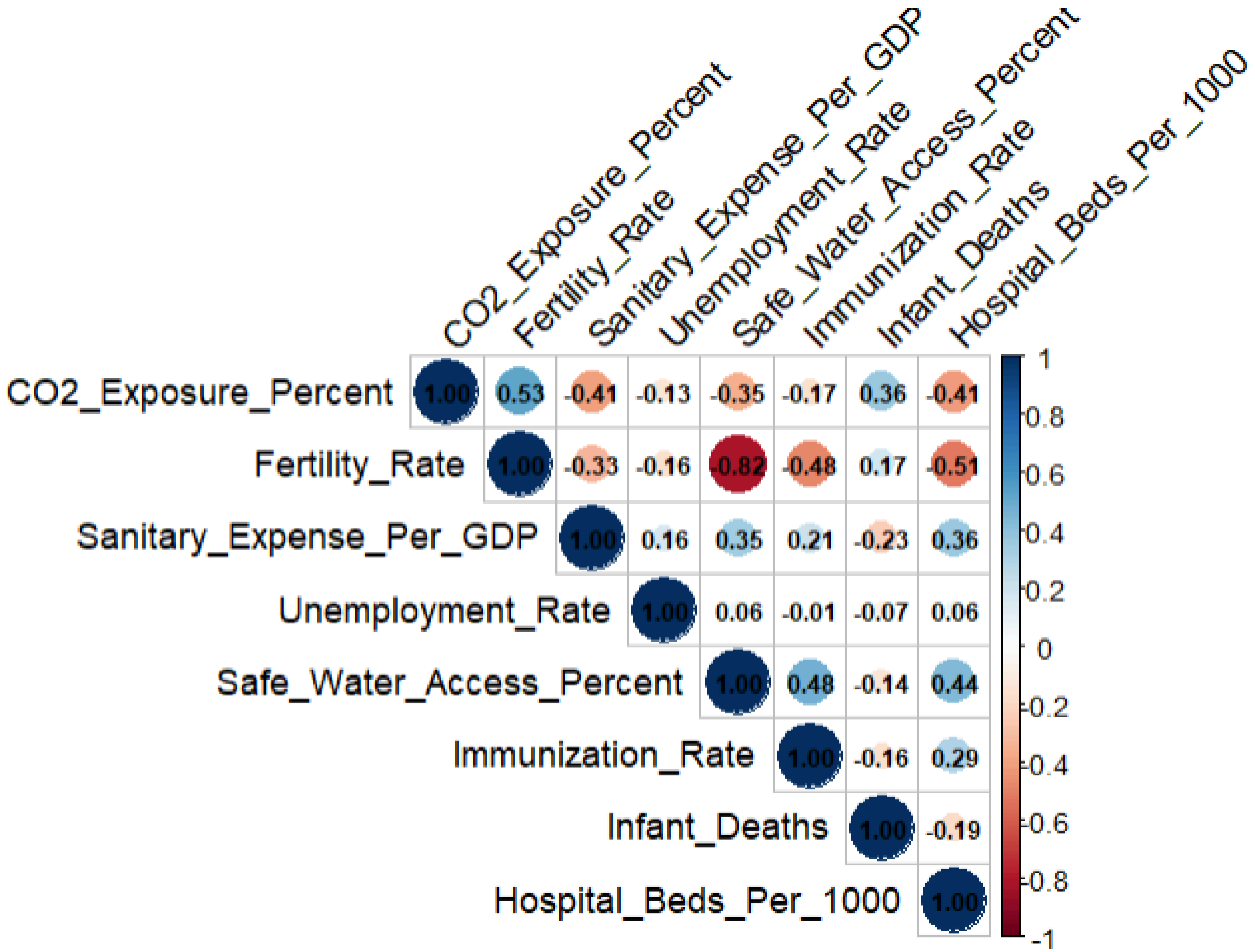






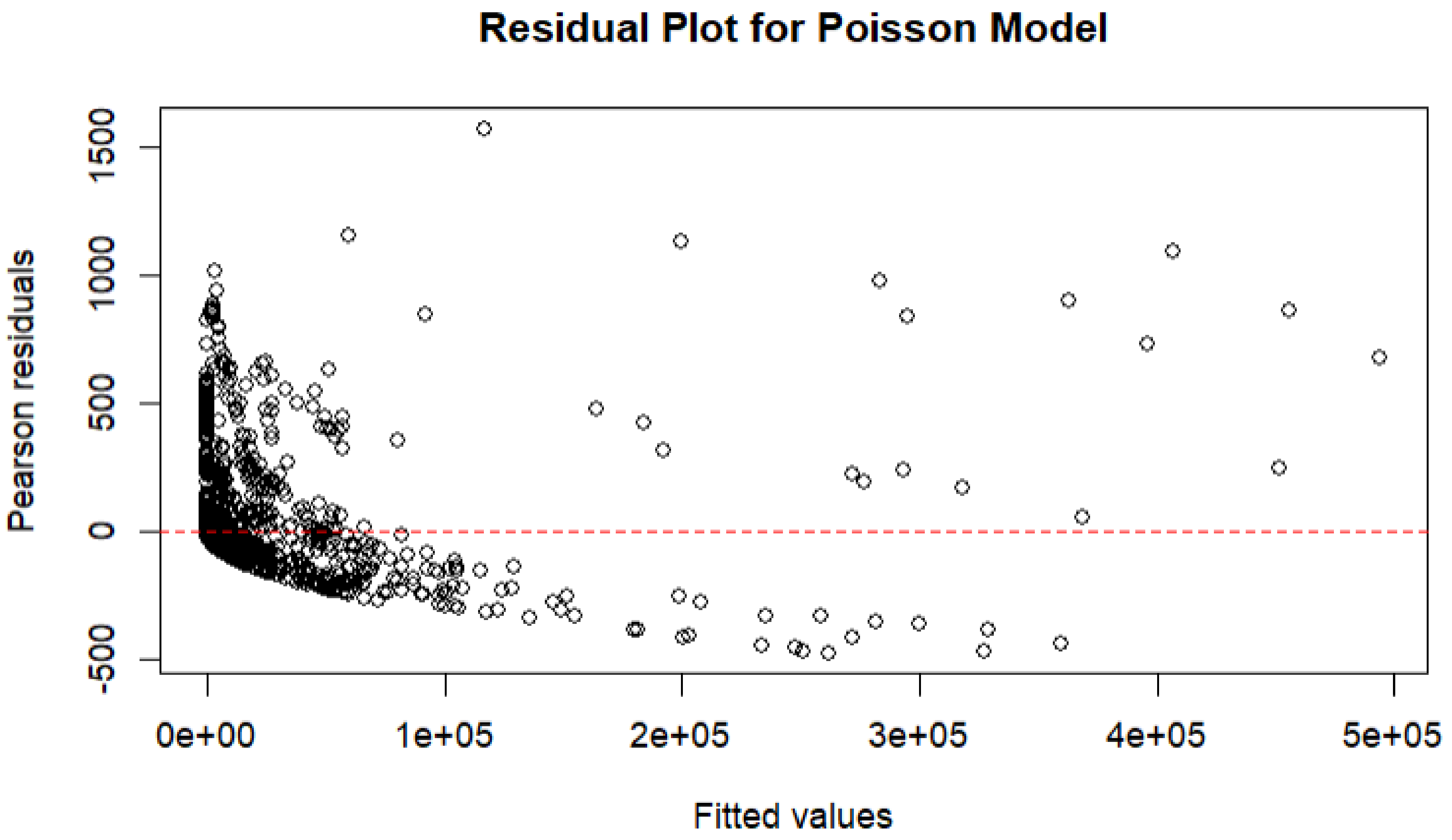


Objective 3: How do socioeconomic and environmental factors impact infant mortality across countries over a 10-year period?



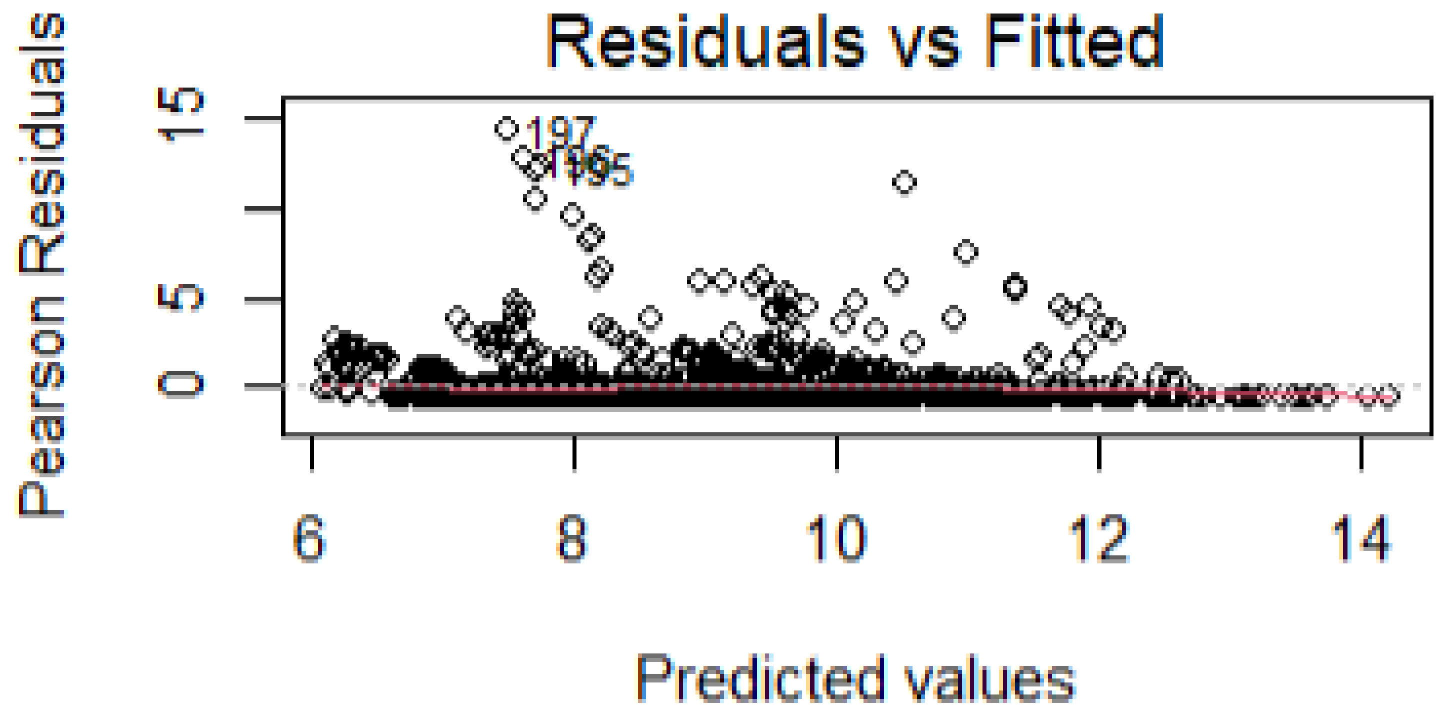
Poisson Regression Model

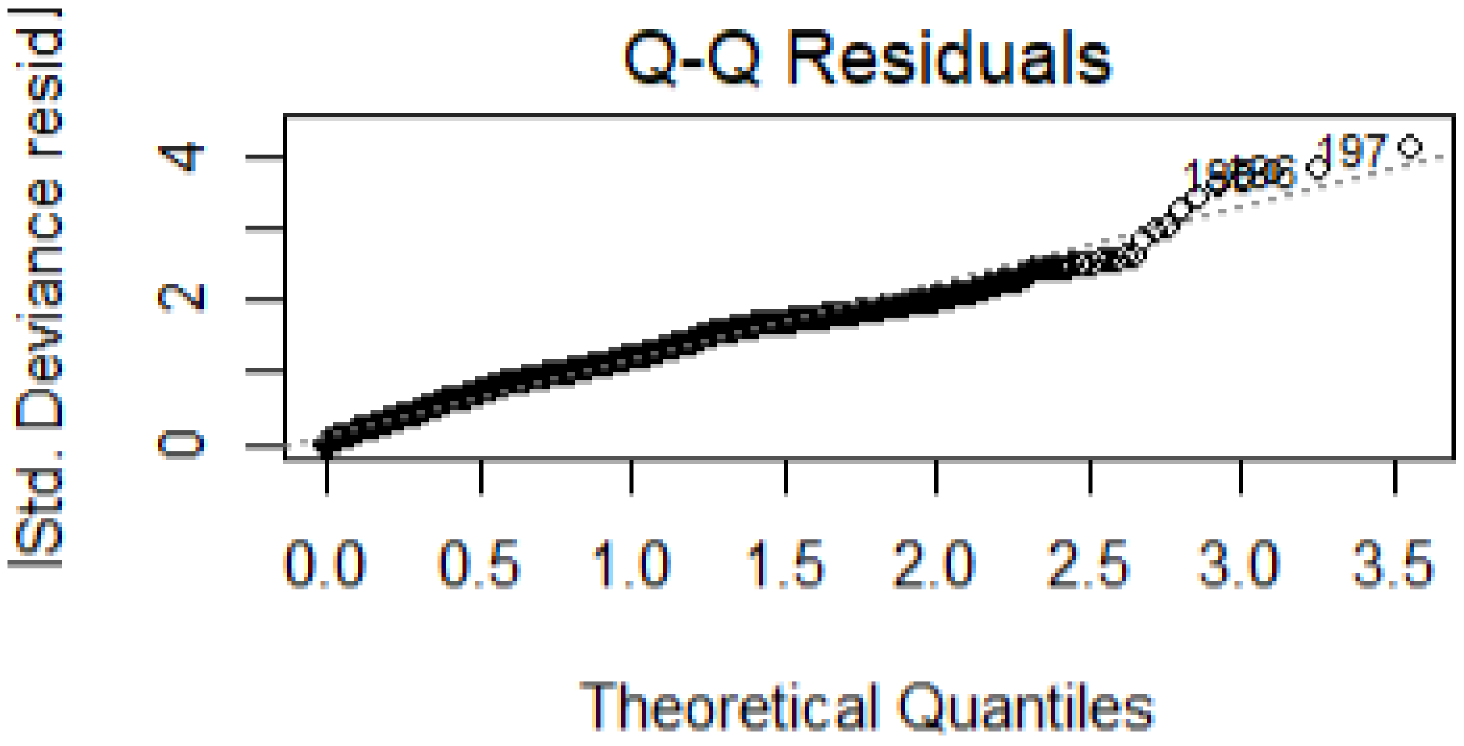
Coefficients	Estimate	p-value
Intercept	14.20	<2e-16 ***
CO2 Exposure (%)	0.043	<2e-16 ***
Fertility Rate	-0.332	<2e-16 ***
Sanitary Expense Per GDP	-0.216	<2e-16 ***
Unemployment Rate	0.0006	<2e-16 ***
Safe Water Access (%)	-0.012	<2e-16 ***
Immunization Rate	-0.028	<2e-16 ***
Hospital Beds	-0.177	<2e-16 ***



Negative Binomial Regression Model

Coefficients	Estimate	p-value
Intercept	15.46	<2e-16 ***
CO2 Exposure (%)	0.049	<2e-16 ***
Fertility Rate	-0.24	0.0002 ***
Sanitary Expense Per GDP	-0.098	5.46e-08 ***
Unemployment Rate	-0.074	<2e-16 ***
Safe Water Access (%)	-0.025	1.22e-11 ***
Immunization Rate	-0.04	<2e-16 ***
Hospital Beds	-0.137	1.25e-10 ***





Objective 3: How do socioeconomic and environmental factors impact infant mortality across countries over a 10-year period?

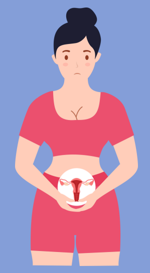
Model Comparison

Model	Dispersion	AIC
Poisson Model	57,568.31	52,279,540
Negative Binomial Model	1.98	25,415.35

Result and Interpretation

💡 CO_2 exposure, fertility rate, sanitary expense per GDP, unemployment rate, safe water access, immunization rate, and hospital beds are significant predictors of infant mortality. These findings highlight the critical role of environmental quality, economic conditions, and healthcare infrastructure in shaping infant mortality rates globally.

A 1% increase in CO2 exposure leads to a 5% increase in infant deaths per 1000 live births.



A 1-unit increase in fertility rate correlates with a 19.8 % decrease in infant deaths per 1000 live births.

A 1% increase in sanitation spending (as a percentage of GDP) leads to 9.4% decrease in infant deaths per 1000 live births.

A 1% increase in unemployment correlates with a 7.1% decrease in infant deaths per 1000 live births.

A 1% increase in access to safe water reduces infant deaths per 1000 live births by 2.2%.

A 1% increase in immunization coverage reduces infant deaths per 1000 live births by 4%.

A 1% unit increase in hospital beds per 1000 population leads to a 12.7% decrease in infant deaths per 1000 live births.

Limitation of the Study

- **Limited Predictive Power:** Current model explains only 45% of infant deaths, indicating missing key factors.

Recommendations

- **Expand Predictors:** Include other variables that could affect infant mortality.
- **Add Interaction Terms:** Analyze how combined factors (e.g., CO2 exposure & Fertility Rate) impact infant mortality.

Thank You

References

- Fadhilah Fitri. (2021). Infant Mortality Case: An Application of Negative Binomial Regression in order to Overcome Overdispersion in Poisson Regression. https://doi.org/10.1007/s40615-017-0381-x

Contribution

25%	25%	25%	25%
Bernard Bernabeo	King Eluard Camota	Maria del Carmen Sandate	Trizzia Ellaine Singson