Associations of Healthcare Costs and Early Detection with Colorectal Cancer Mortality in the USA

Yulin Yuan, Ruohan Sun, Tony Lee

Introduction

Study Objective

Clearly define the research questions or objectives.

Dataset Overview

Describe:

- Where the data comes from
- What each variable represents
- How the data was collected (if known)

Motivation

Why is this question or data important/interesting?

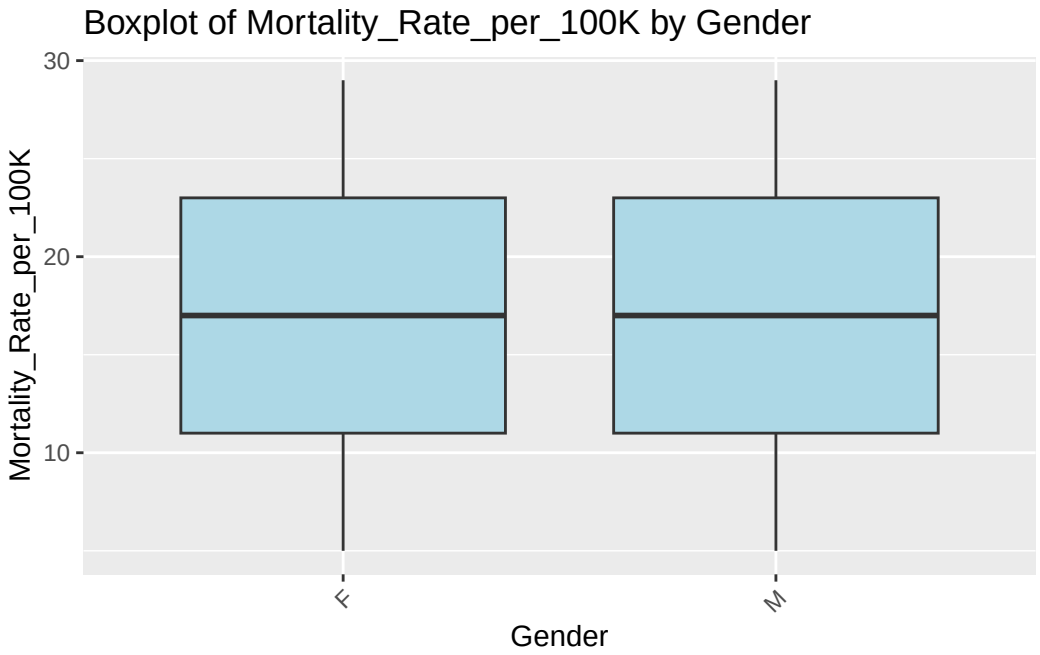
Analysis

Exploratory Data Analysis (EDA)

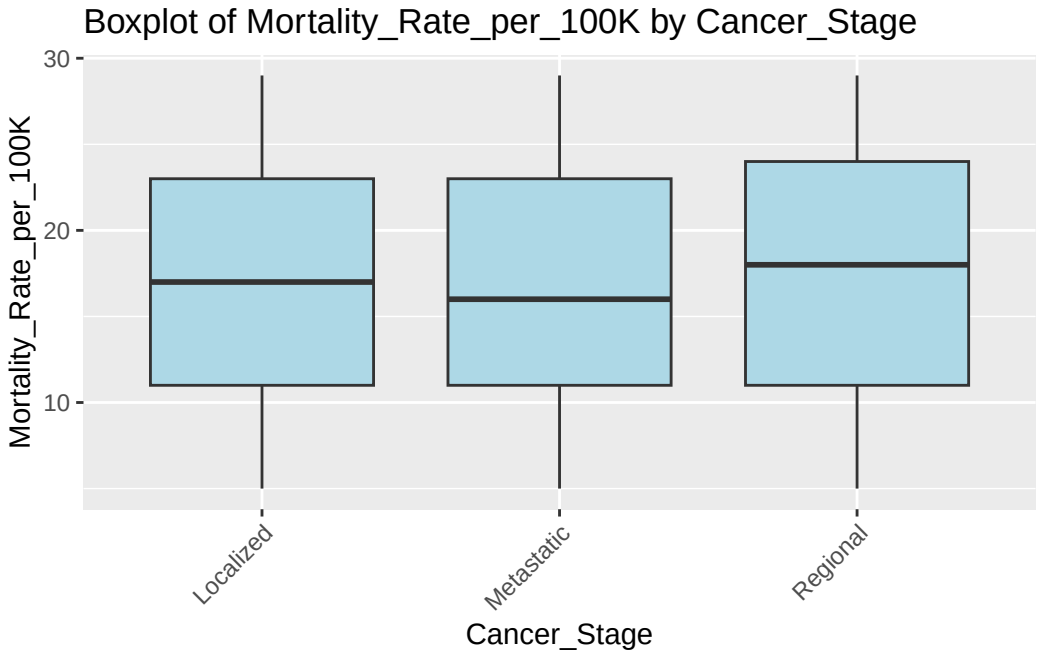
Data Visualizations

Categorical Variables

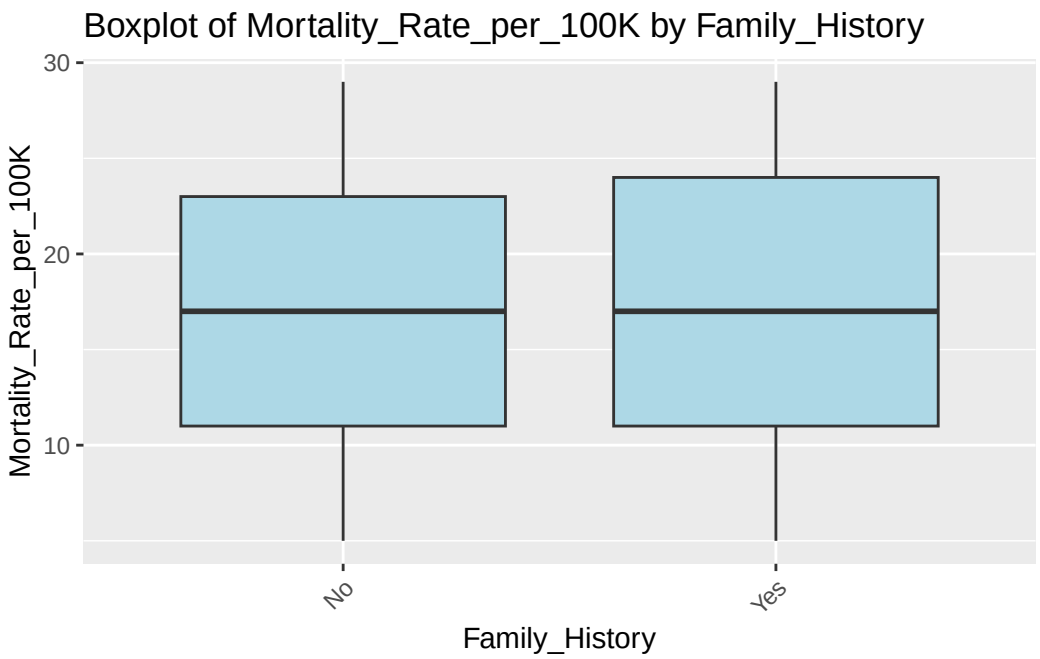
`Boxplot: Mortality_Rate_per_100K by Gender`



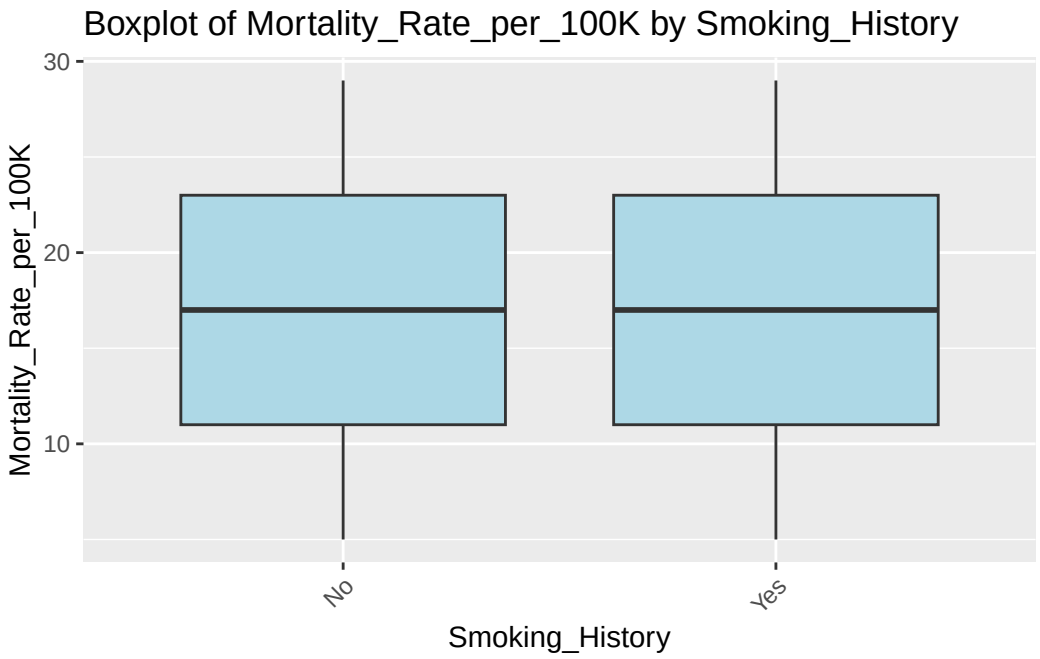
`Boxplot: Mortality_Rate_per_100K by Cancer_Stage`



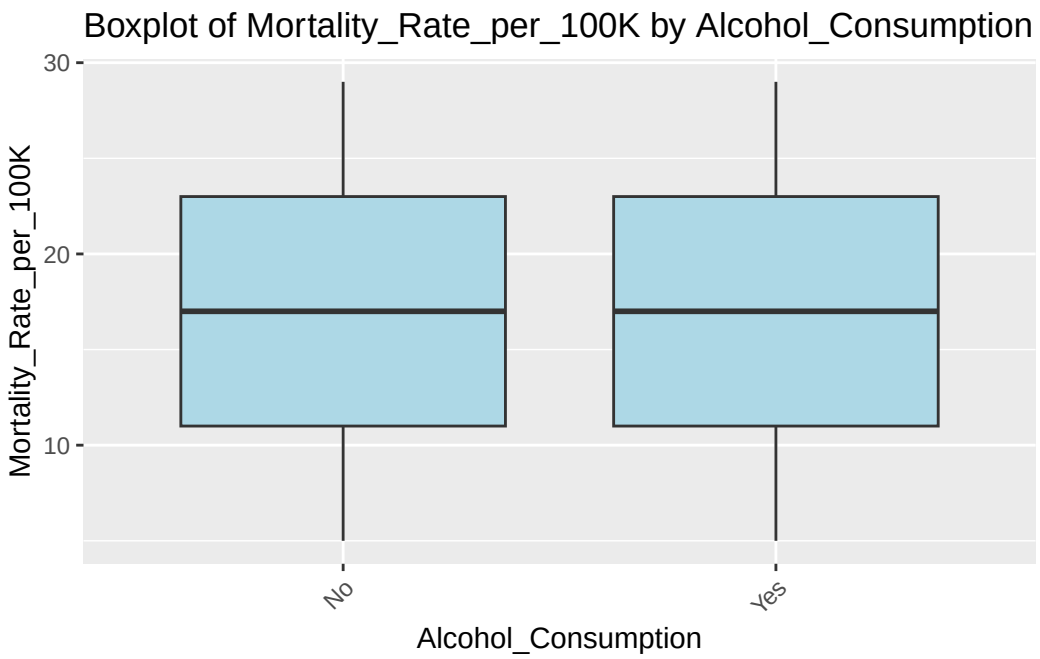
```
Boxplot: Mortality_Rate_per_100K by Family_History
```



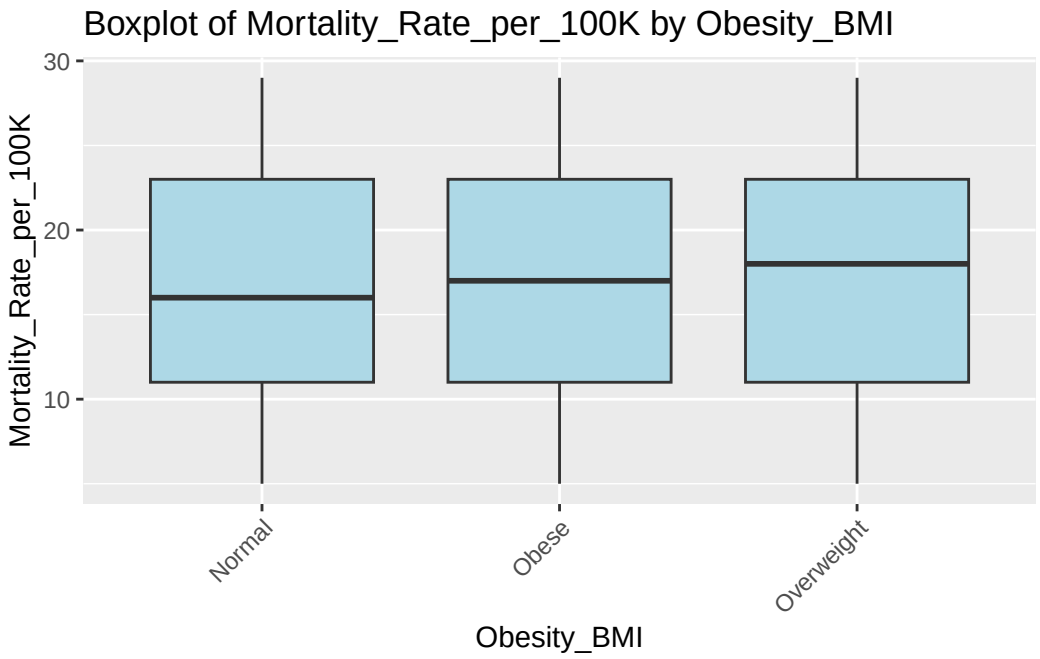

```
Boxplot: Mortality_Rate_per_100K by Smoking_History
```



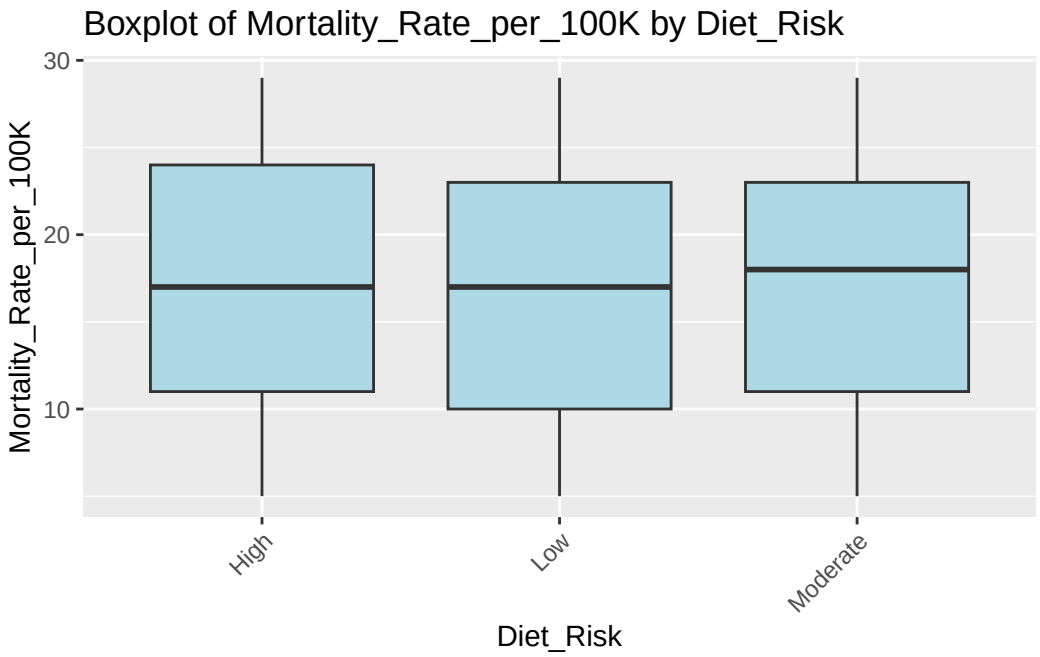
```
Boxplot: Mortality_Rate_per_100K by Alcohol_Consumption
```



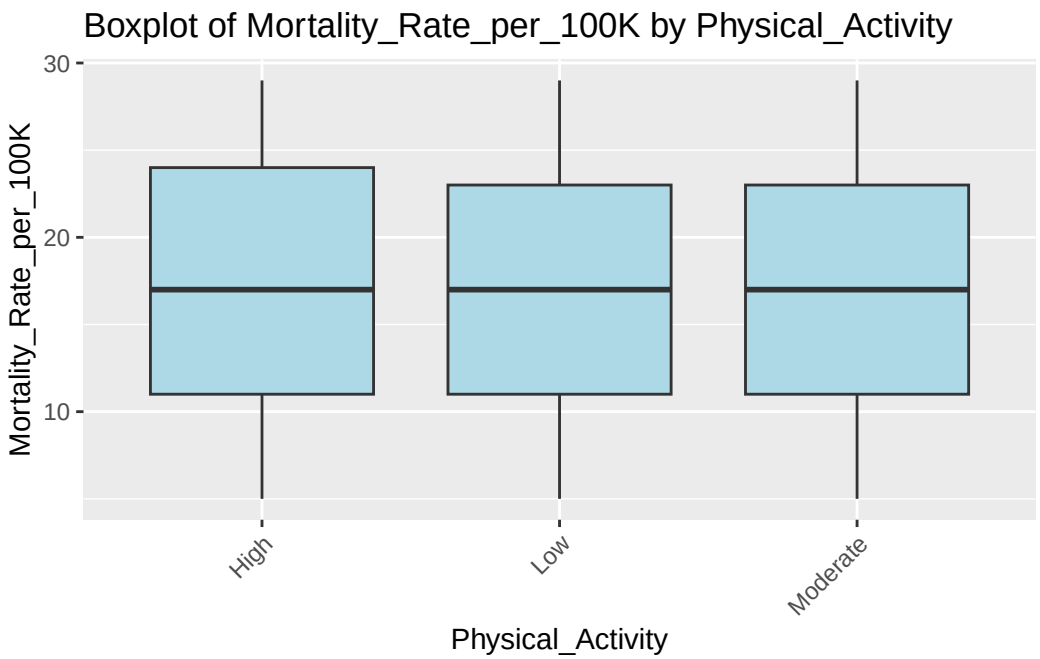

```
Boxplot: Mortality_Rate_per_100K by Obesity_BMI
```



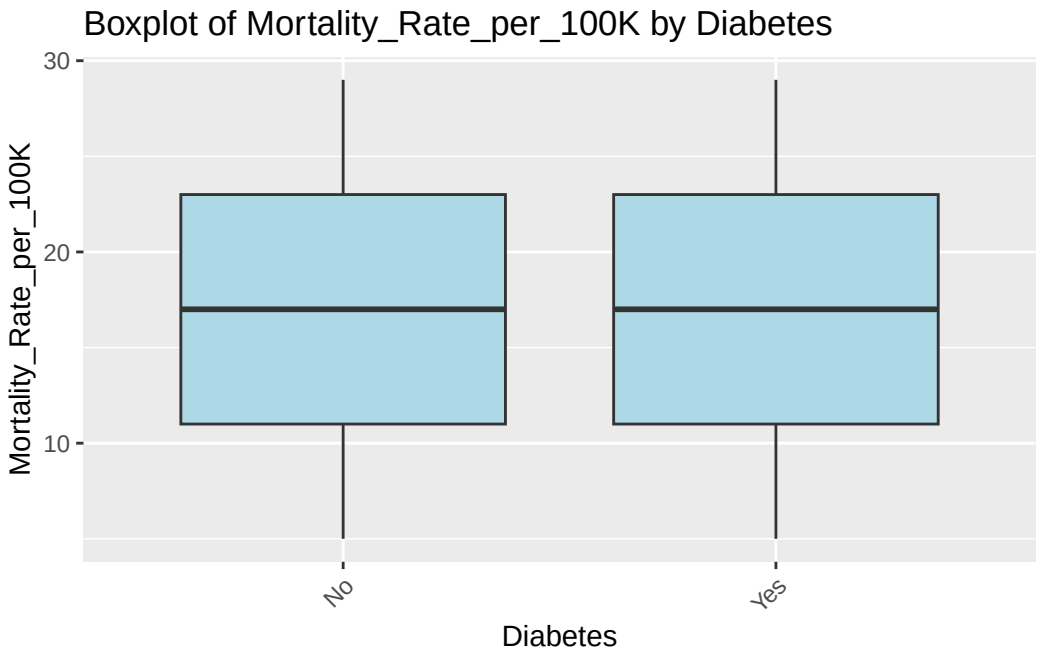
Boxplot: Mortality_Rate_per_100K by Diet_Risk



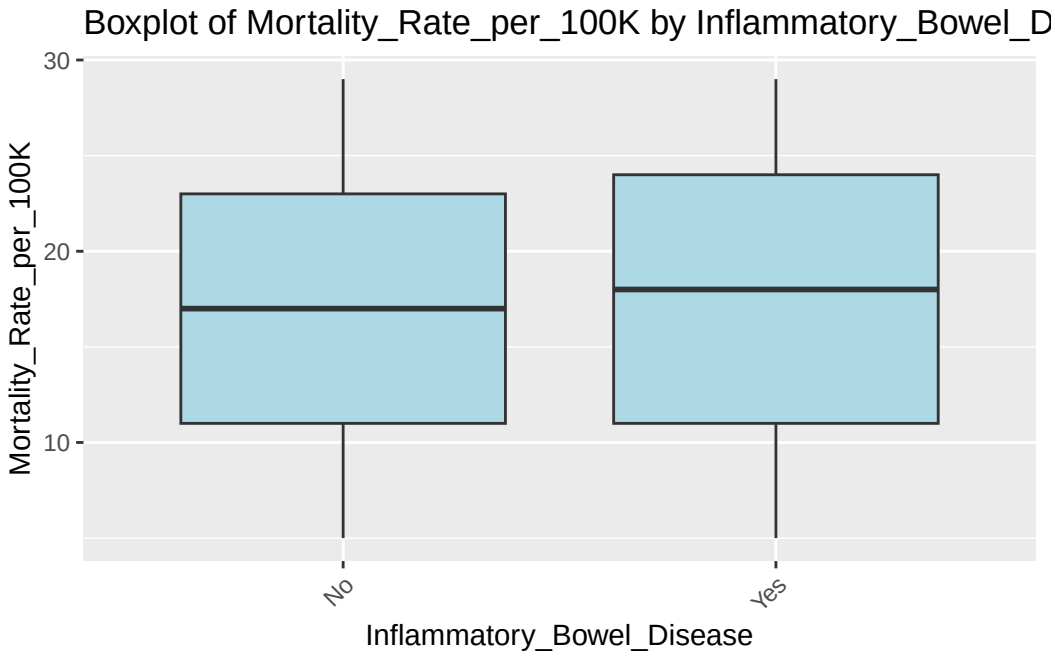
Boxplot: Mortality_Rate_per_100K by Physical_Activity



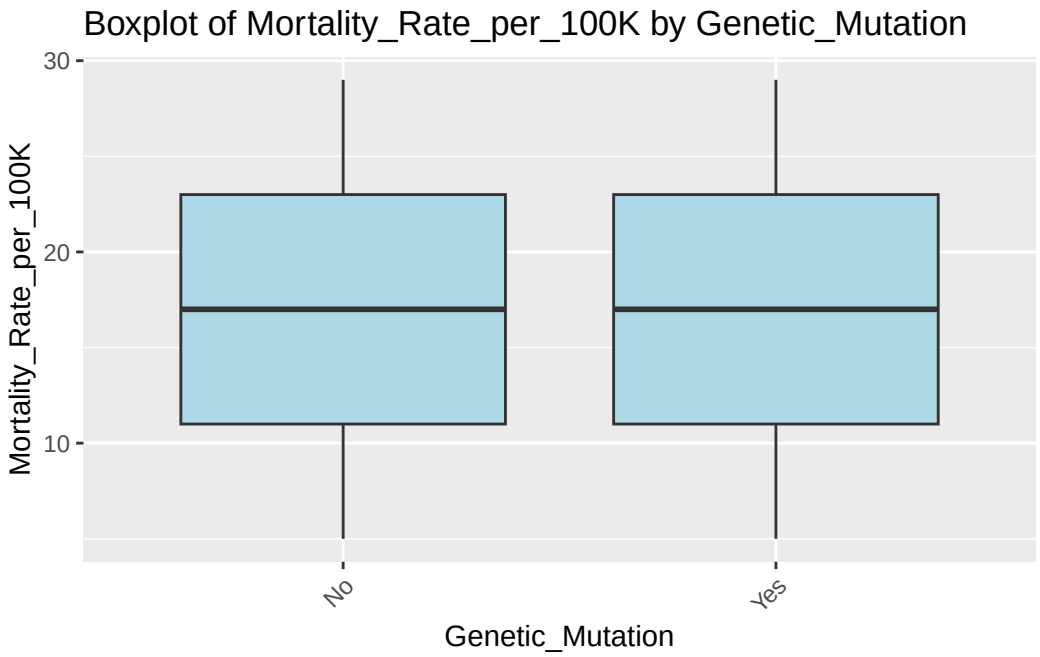
```
Boxplot: Mortality_Rate_per_100K by Diabetes
```



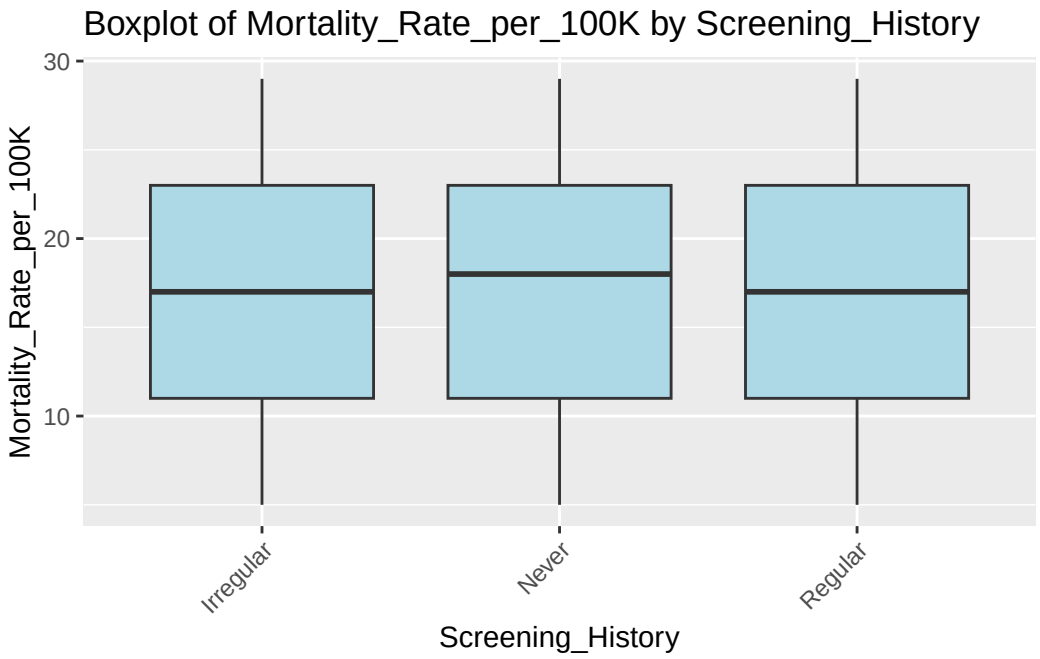

```
Boxplot: Mortality_Rate_per_100K by Inflammatory_Bowel_Disease
```



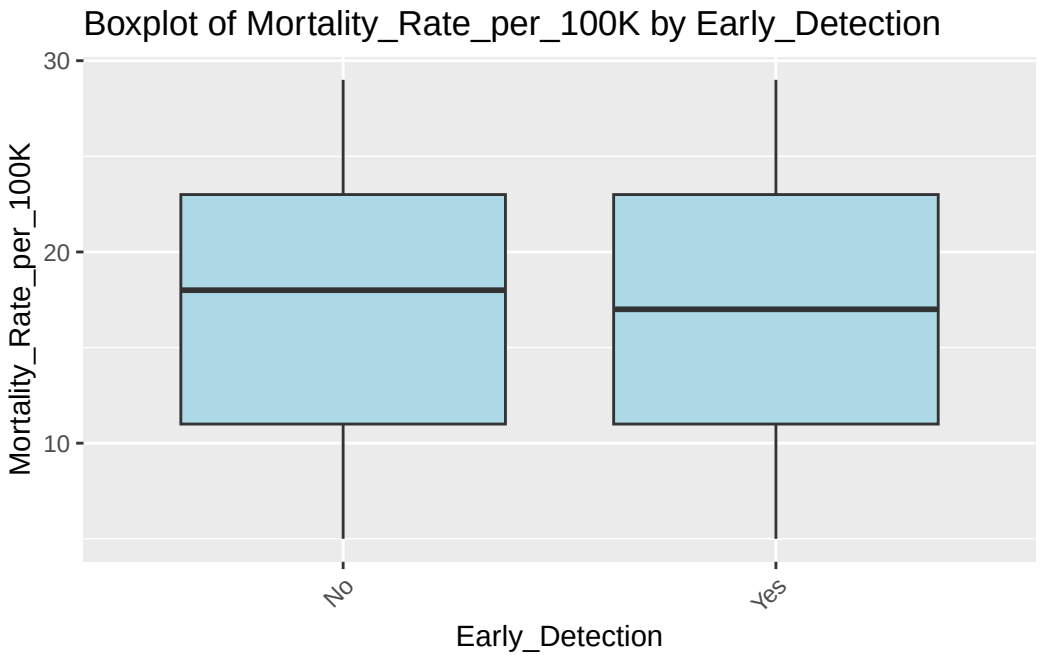
Boxplot: Mortality_Rate_per_100K by Genetic_Mutation



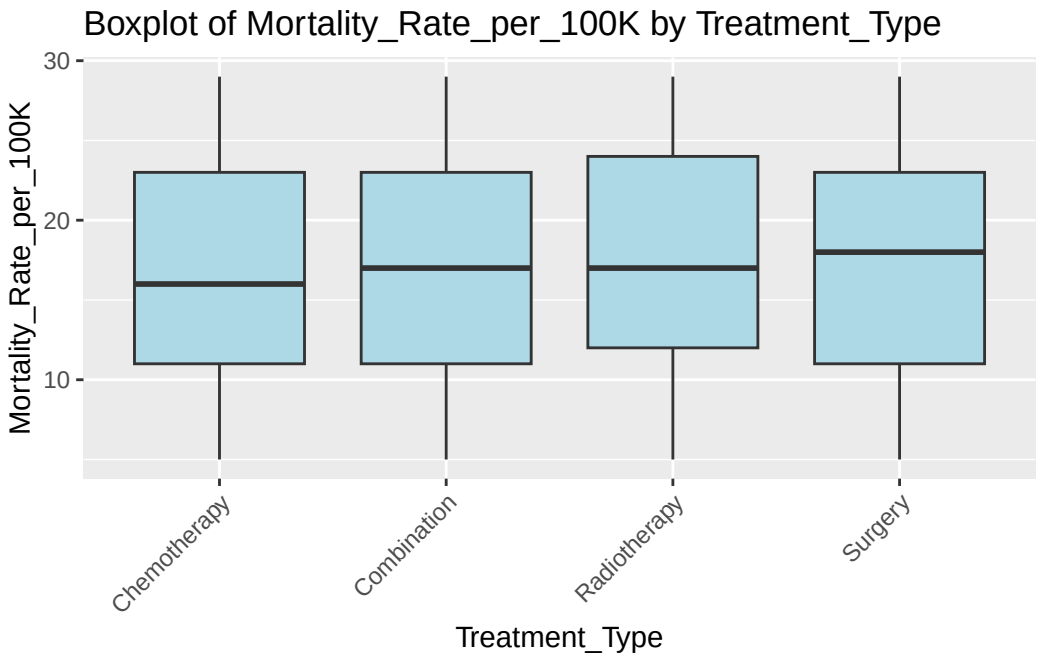
Boxplot: Mortality_Rate_per_100K by Screening_History



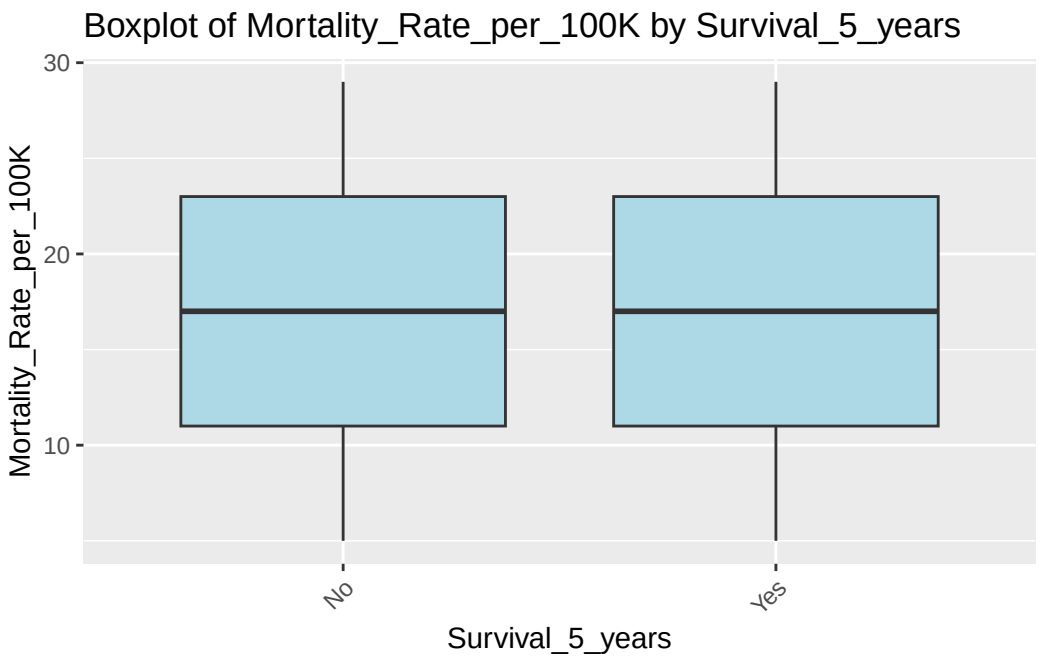
```
Boxplot: Mortality_Rate_per_100K by Early_Detection
```



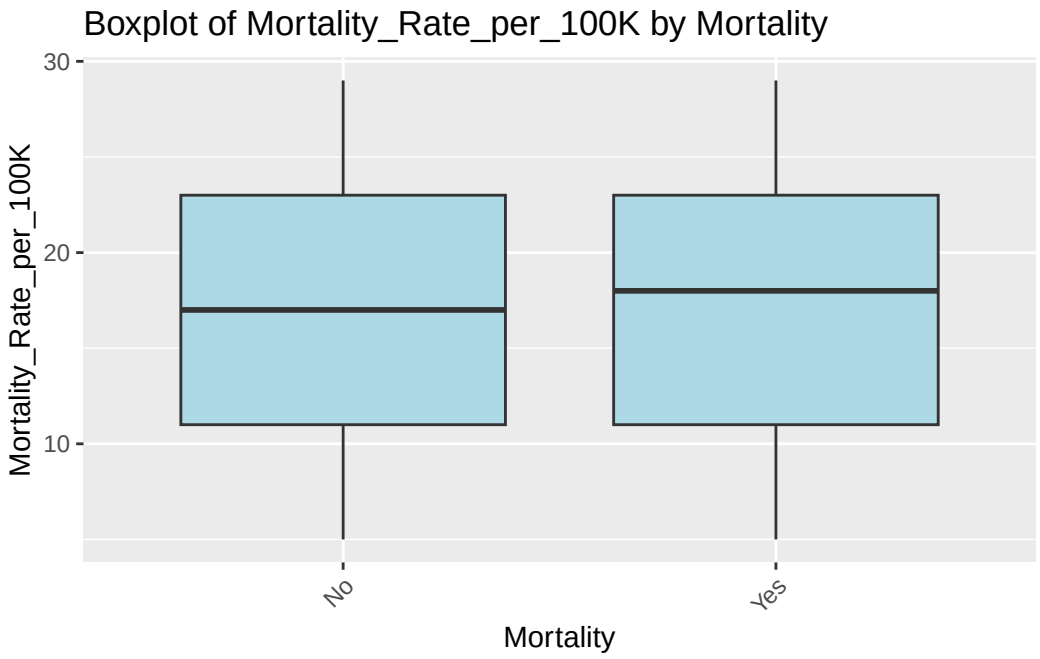

```
Boxplot: Mortality_Rate_per_100K by Treatment_Type
```



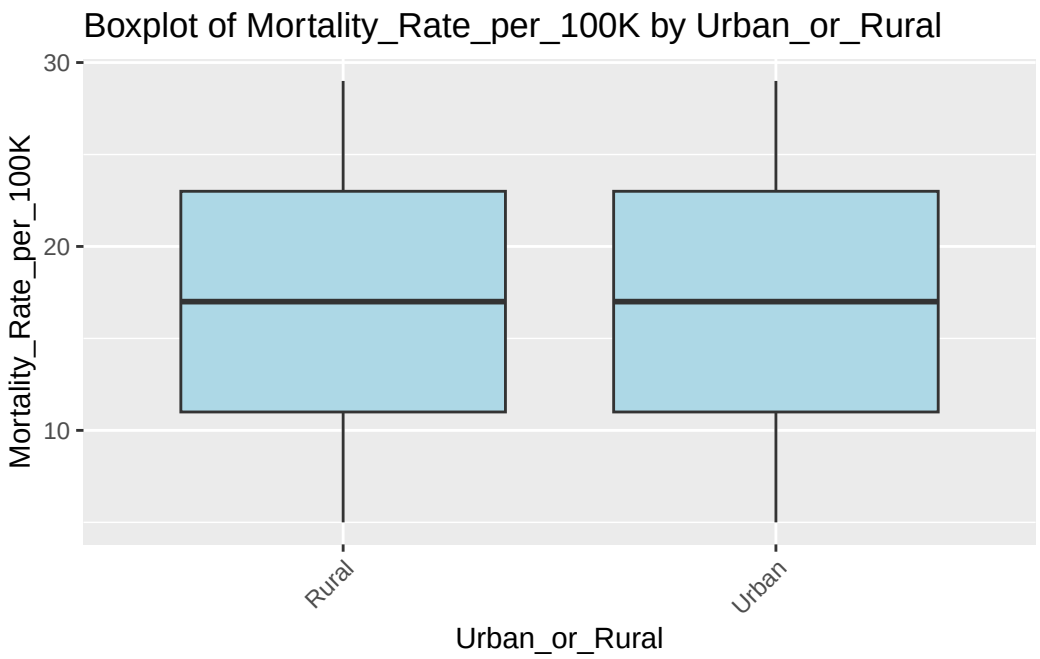
Boxplot: Mortality_Rate_per_100K by Survival_5_years



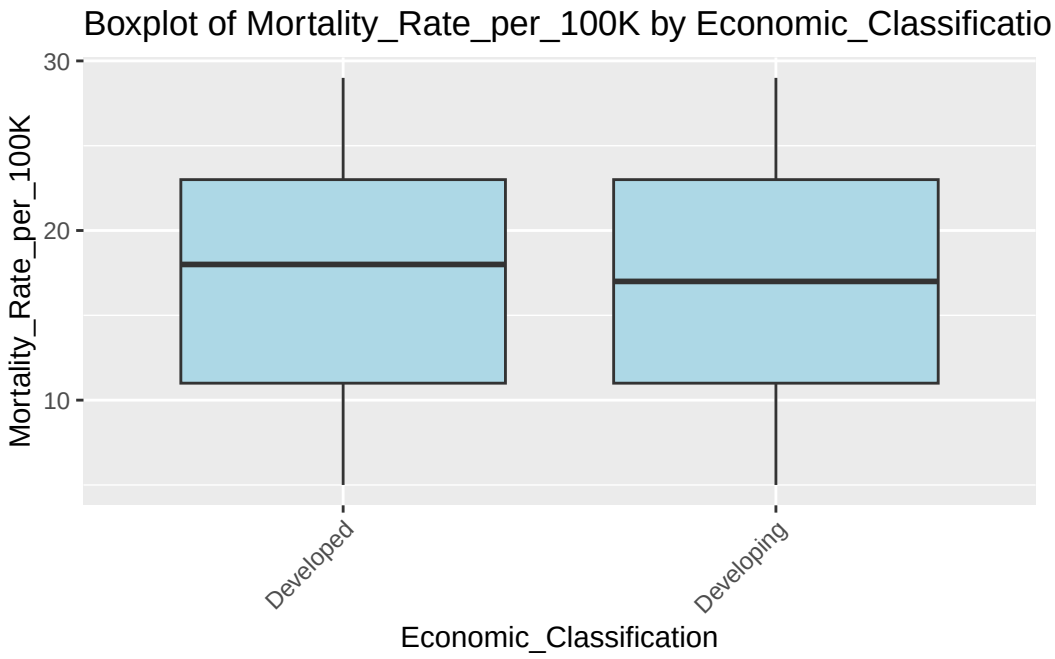
Boxplot: Mortality_Rate_per_100K by Mortality



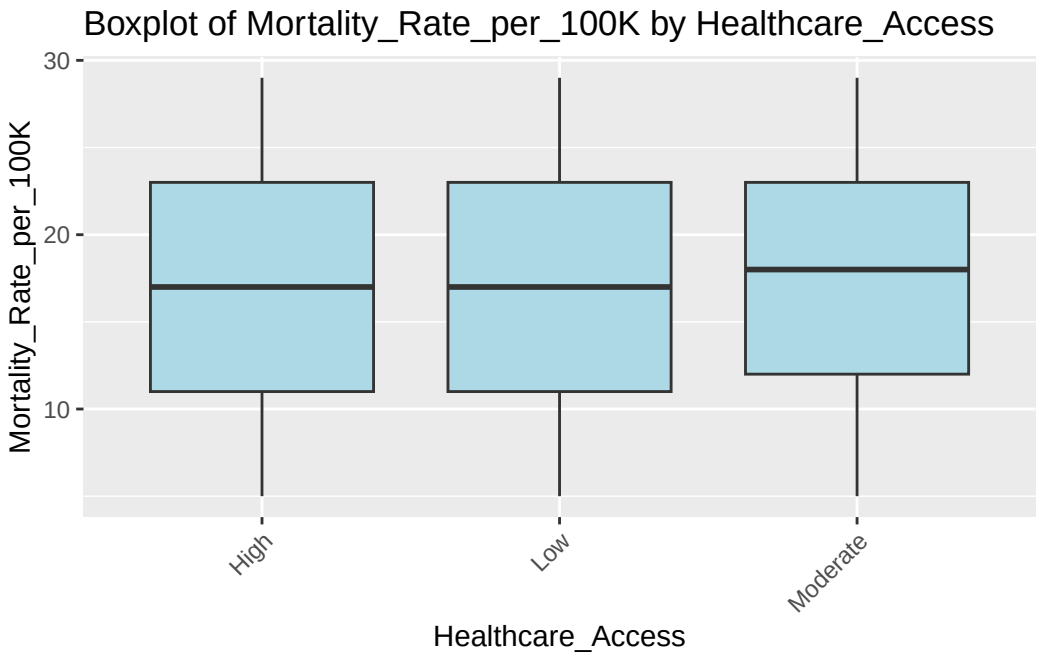
```
Boxplot: Mortality_Rate_per_100K by Urban_or_Rural
```



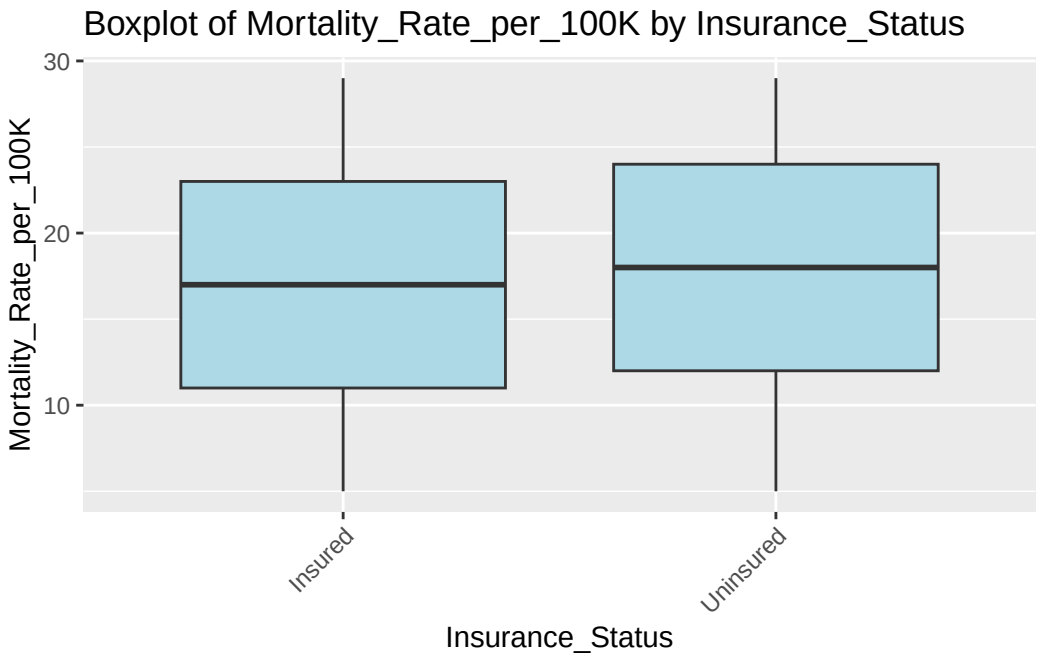

```
Boxplot: Mortality_Rate_per_100K by Economic_Classification
```



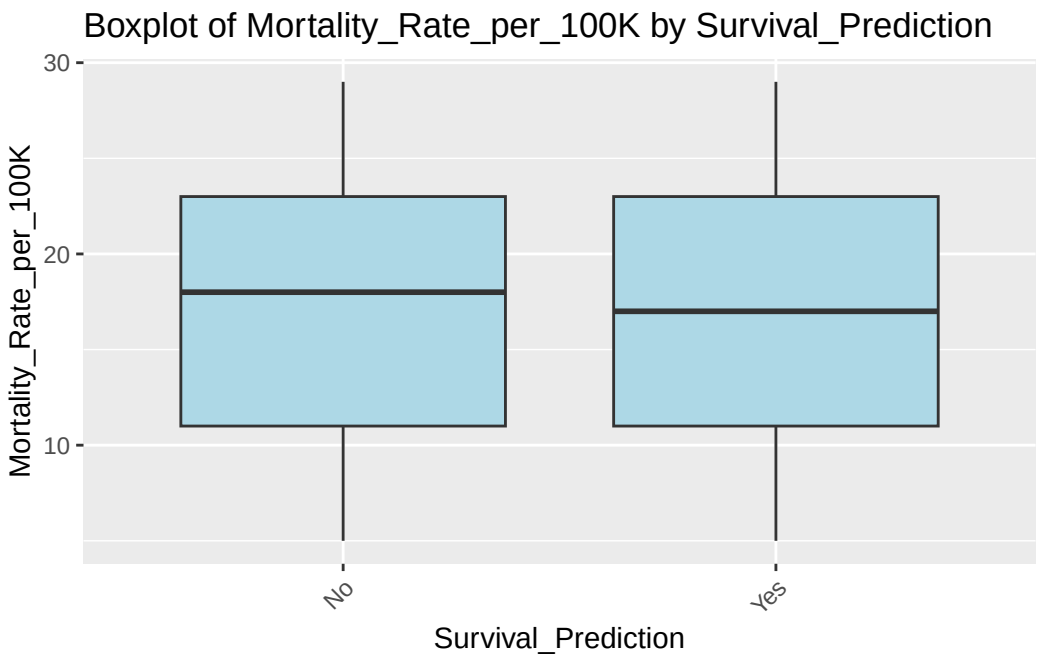
Boxplot: Mortality_Rate_per_100K by Healthcare_Access



Boxplot: Mortality_Rate_per_100K by Insurance_Status

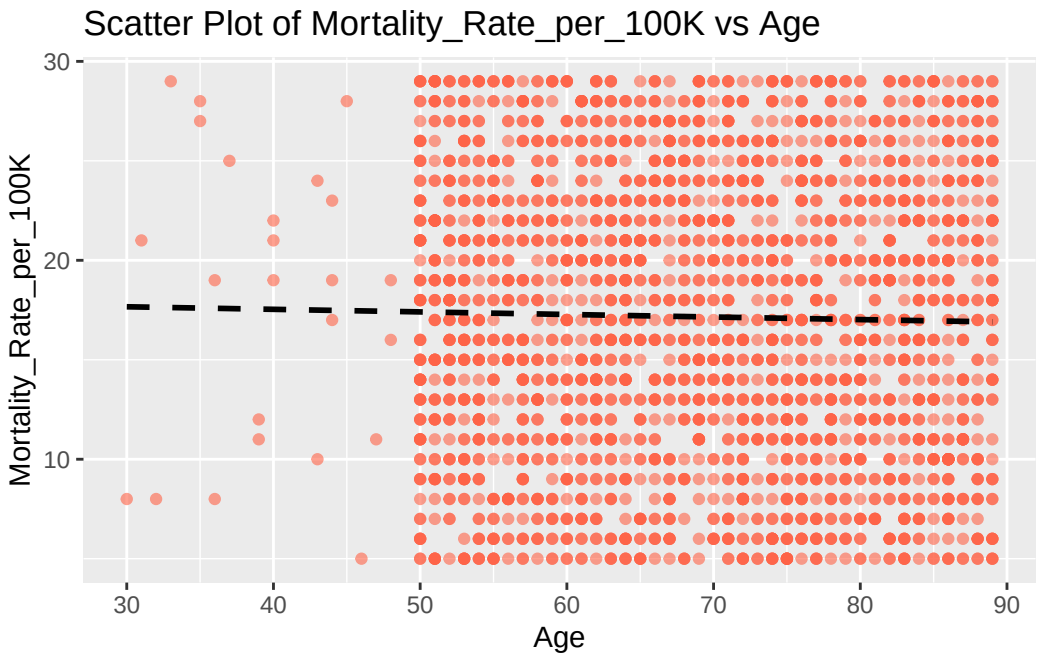


Boxplot: Mortality_Rate_per_100K by Survival_Prediction

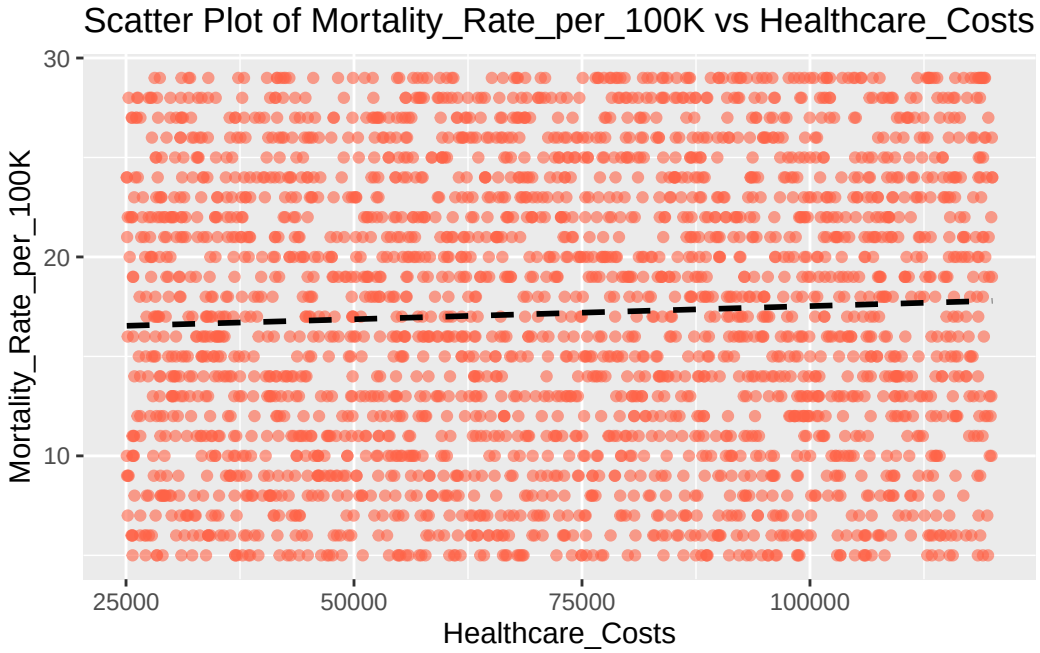


Continuous Variabls

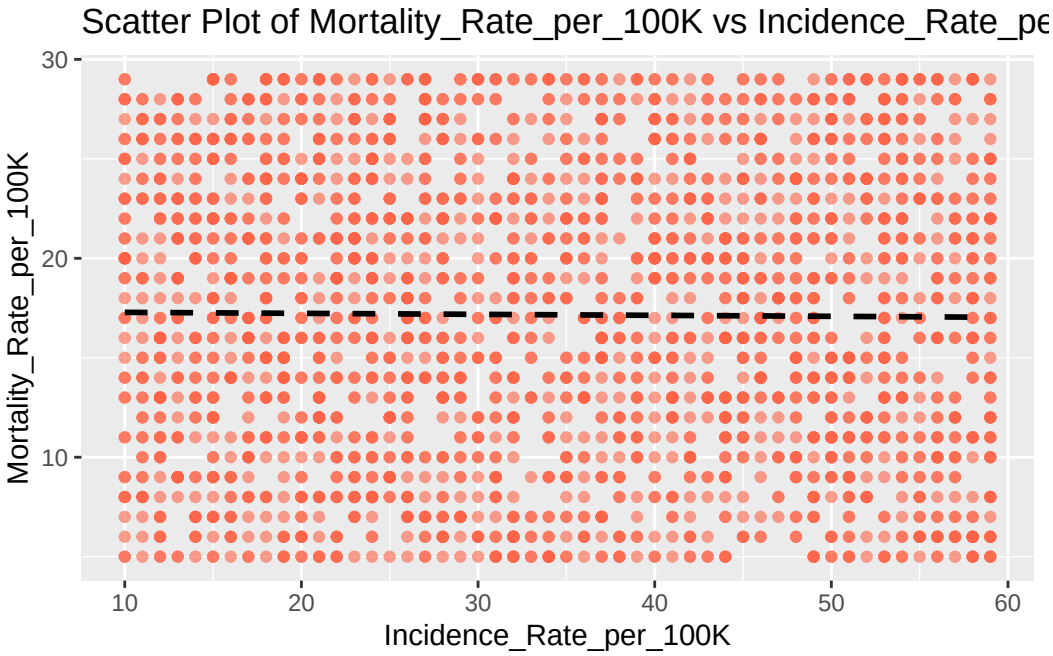
Scatter plot: Mortality_Rate_per_100K vs Age



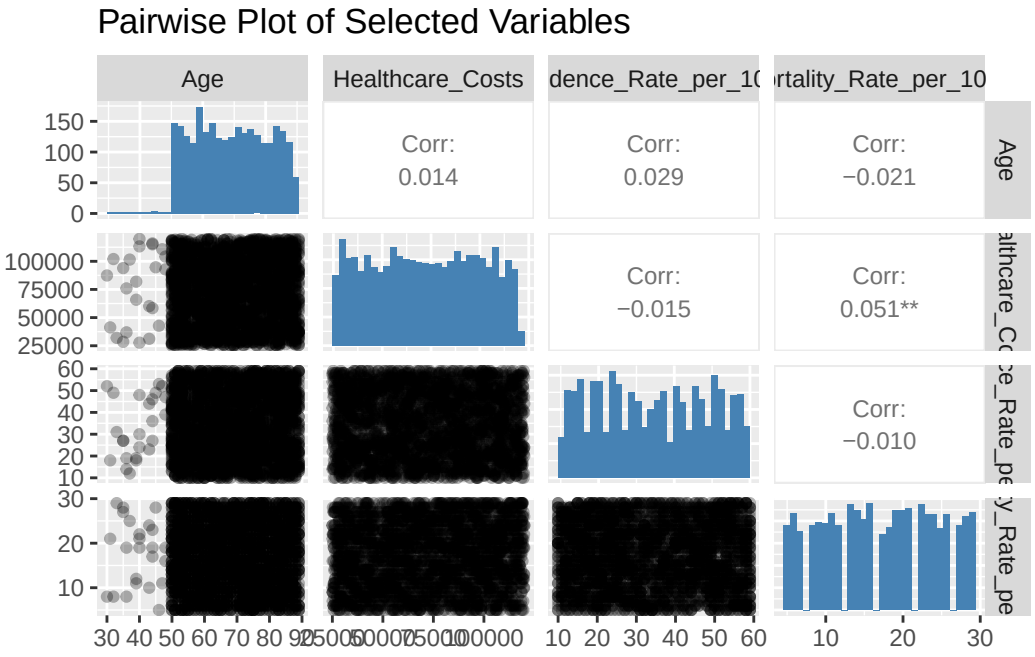
Scatter plot: Mortality_Rate_per_100K vs Healthcare_Costs



Scatter plot: Mortality_Rate_per_100K vs Incidence_Rate_per_100K



Summary Correlations



Interpretation of Findings

Pattern, trends, suggested operations

Model Choice and Reasoning

Linear Regression

explain why choose this model based on EDA and Data description

Data Transformations

Model Selection

Statistical Analysis

Results Interpretation

inference results

Conclusion

Main Findings

Interpreting result in real-world context, careful about causality

Limitations

- Discuss possible sources of bias, limitations in data, model assumptions
- Suggest improvements or next steps

Potential Further research

Mention anything interesting you found that doesn't fit elsewhere

Appendix

- Full regression output
- Extra plots or tables not essential to the main body
- Model selection steps