

# 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?

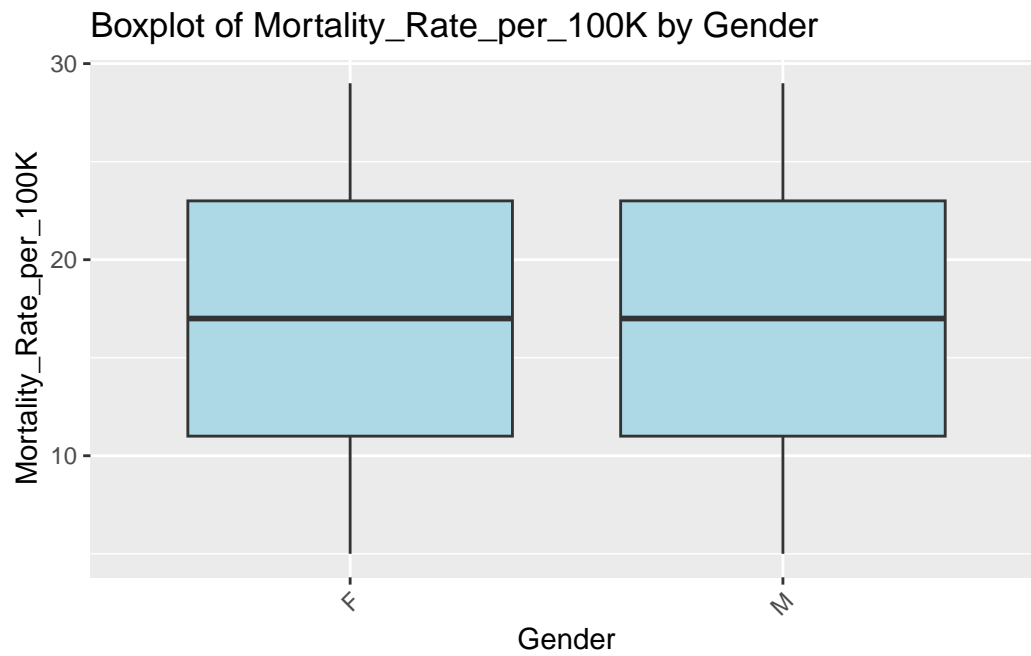
### Analysis

#### Exploratory Data Analysis (EDA)

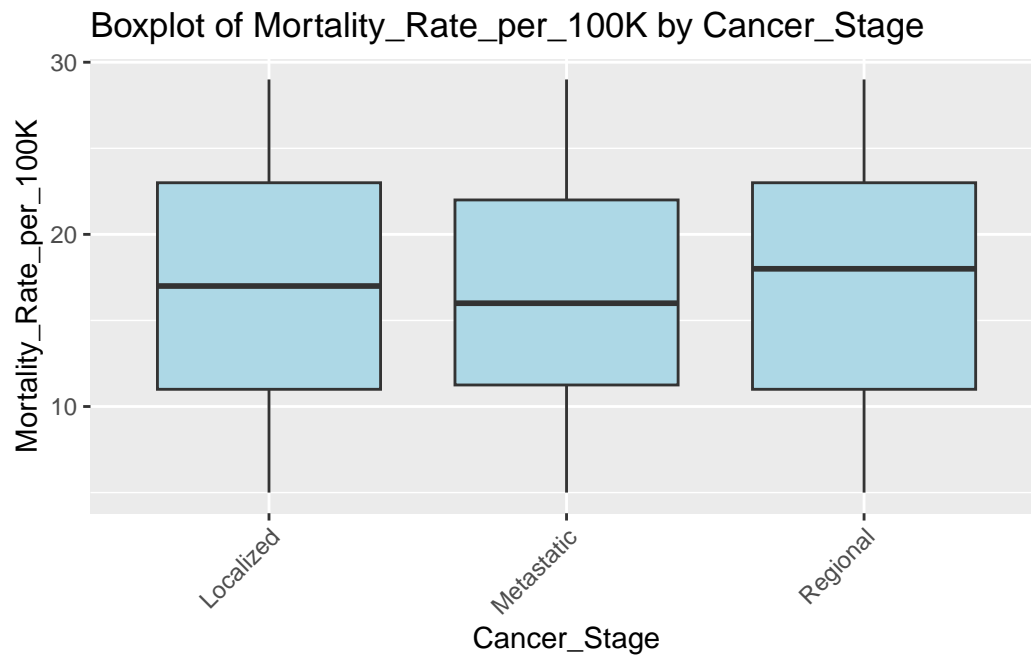
#### Data Visualizations

## Categorical Variables

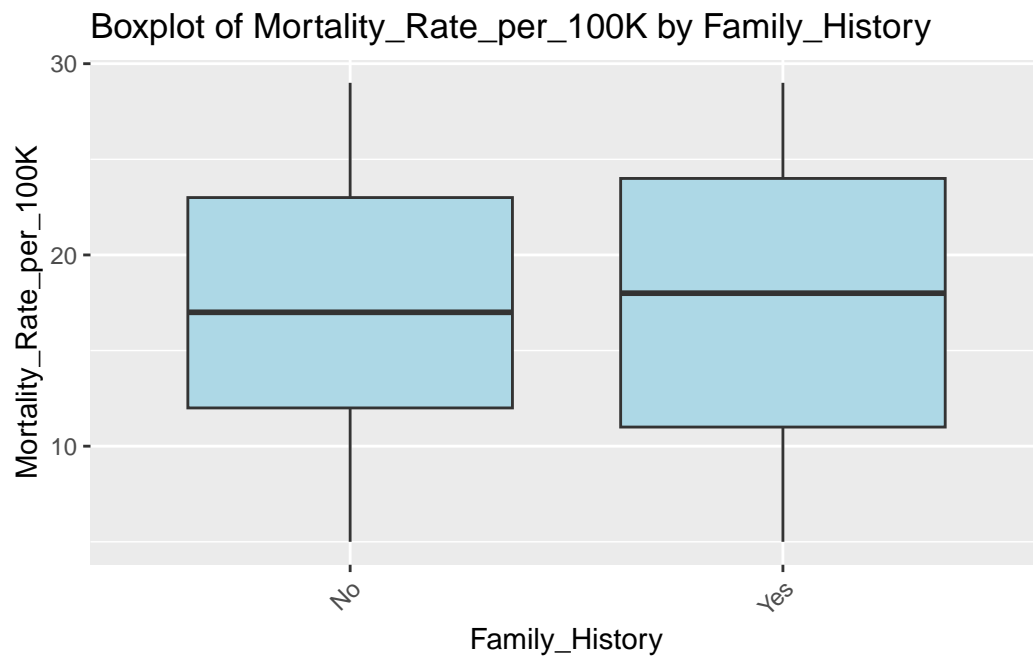
Boxplot: Mortality\_Rate\_per\_100K by Gender



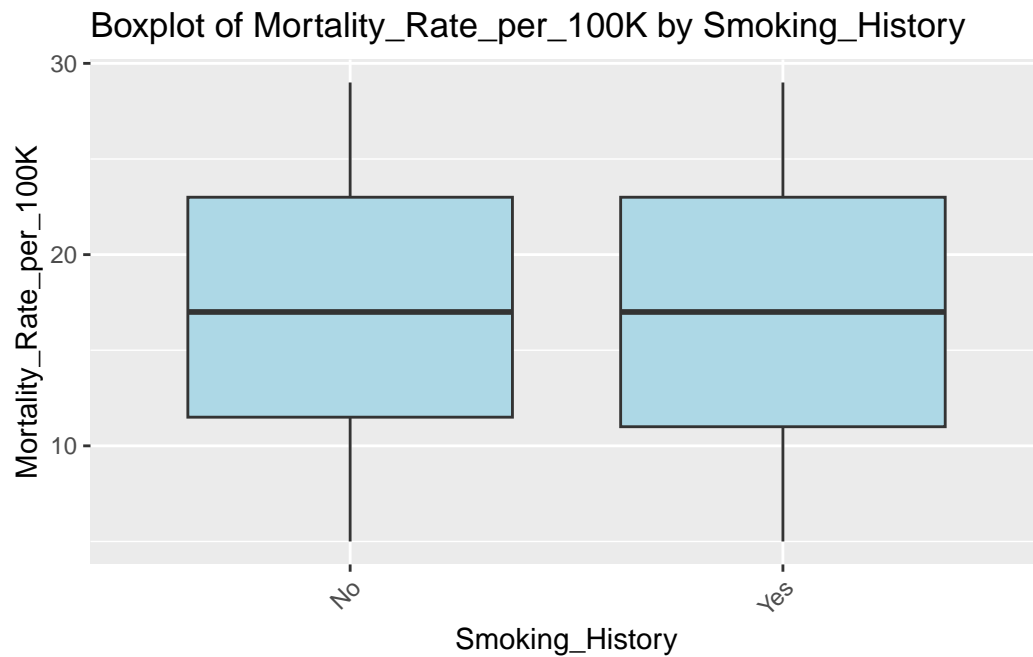
Boxplot: Mortality\_Rate\_per\_100K by Cancer\_Stage



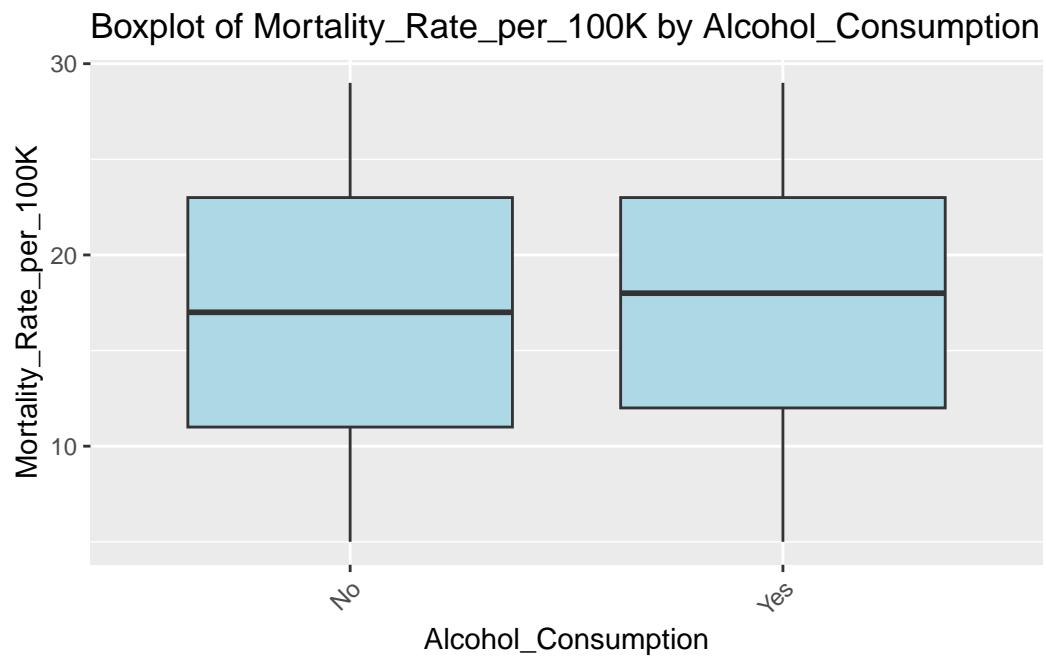
Boxplot: Mortality\_Rate\_per\_100K by Family\_History



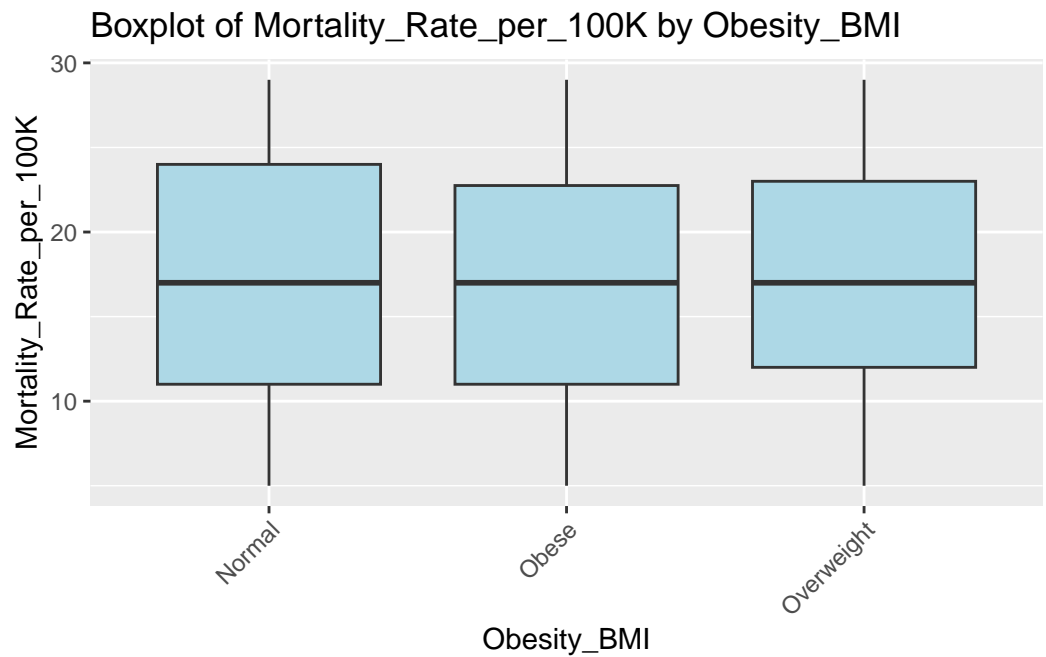
Boxplot: Mortality\_Rate\_per\_100K by Smoking\_History



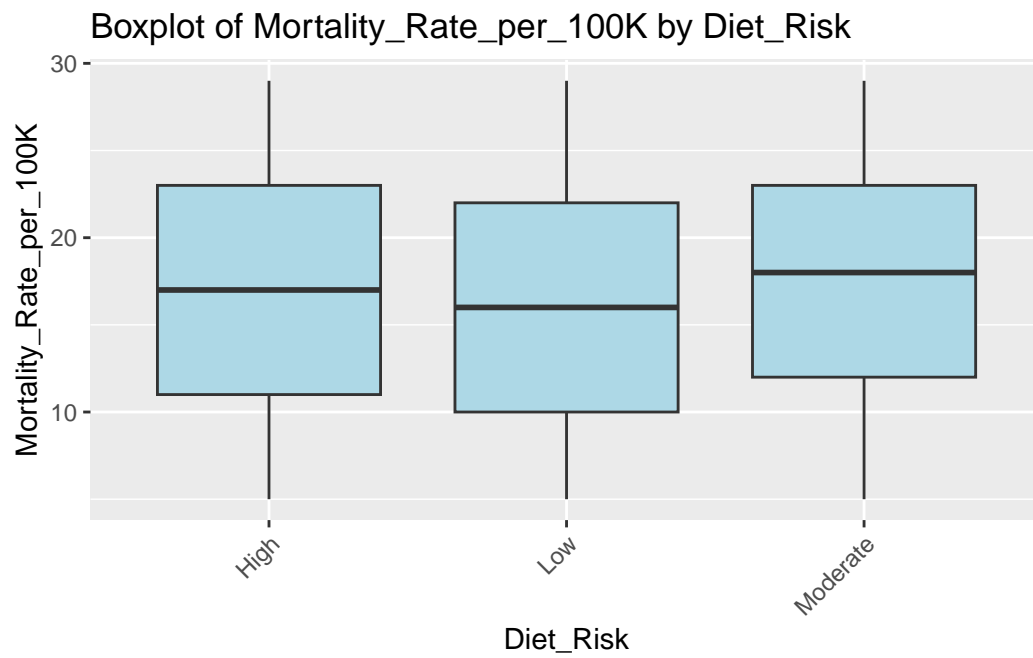
Boxplot: Mortality\_Rate\_per\_100K by Alcohol\_Consumption



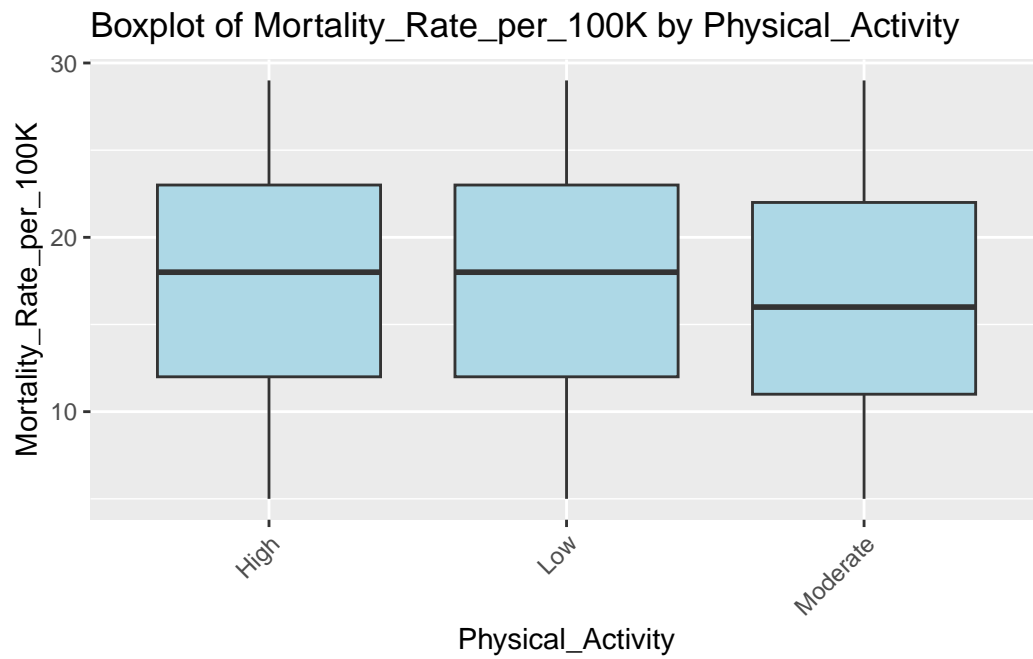
Boxplot: Mortality\_Rate\_per\_100K by Obesity\_BMI



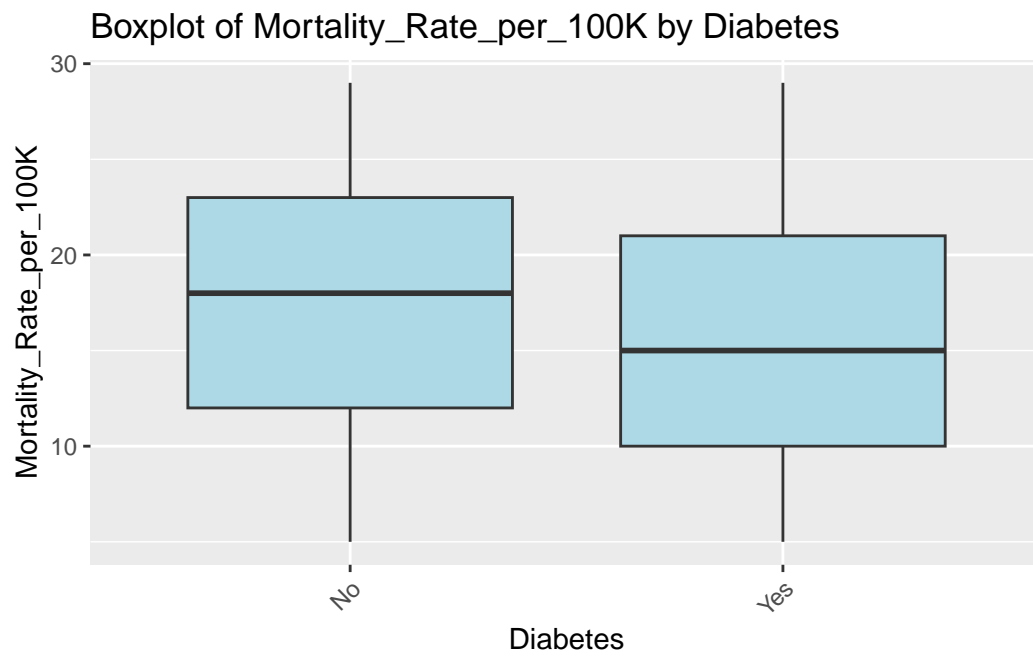
Boxplot: Mortality\_Rate\_per\_100K by Diet\_Risk



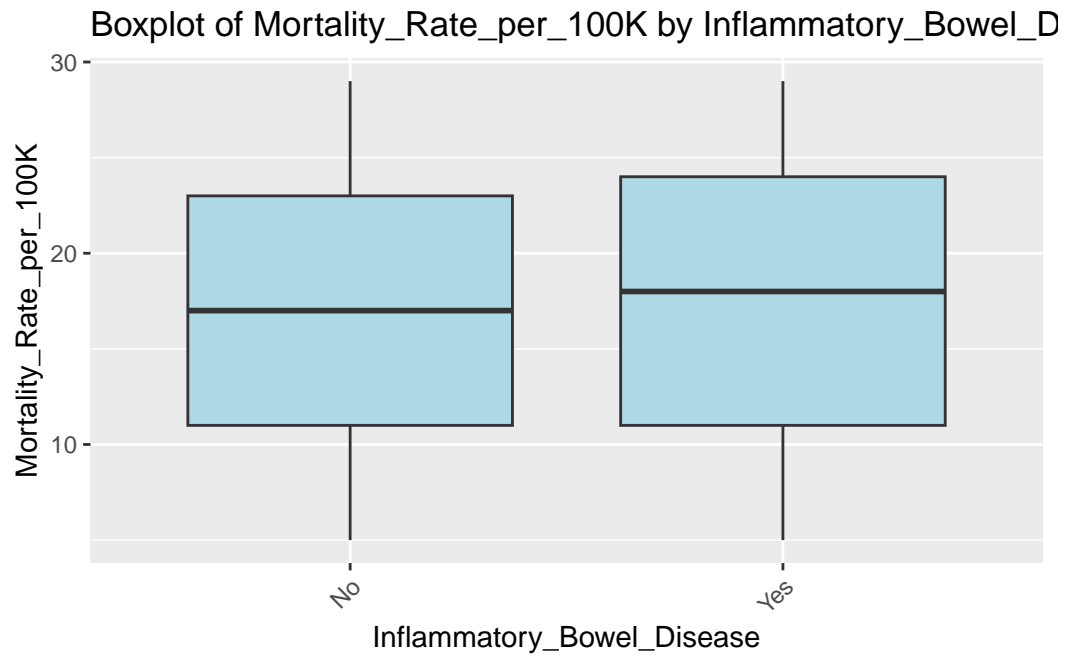
Boxplot: Mortality\_Rate\_per\_100K by Physical\_Activity



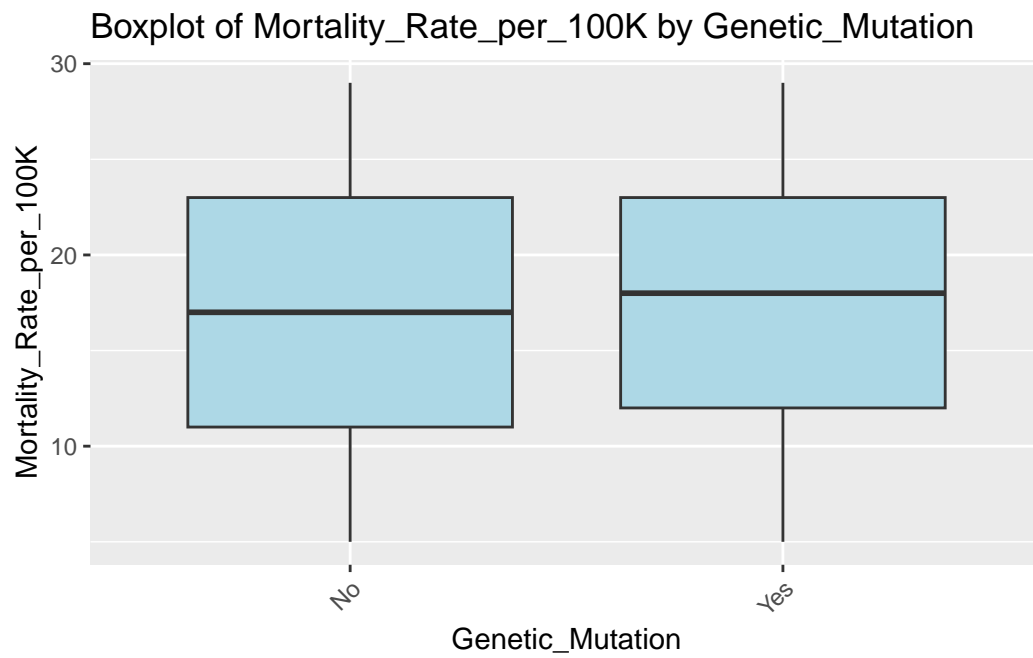
Boxplot: Mortality\_Rate\_per\_100K by Diabetes



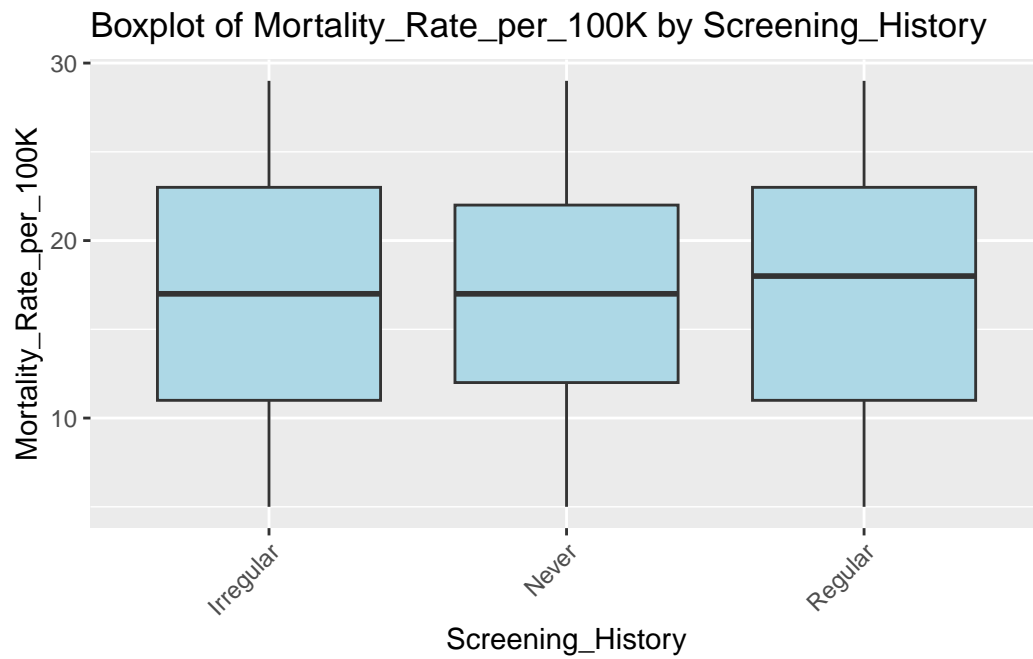
Boxplot: Mortality\_Rate\_per\_100K by Inflammatory\_Bowel\_Disease



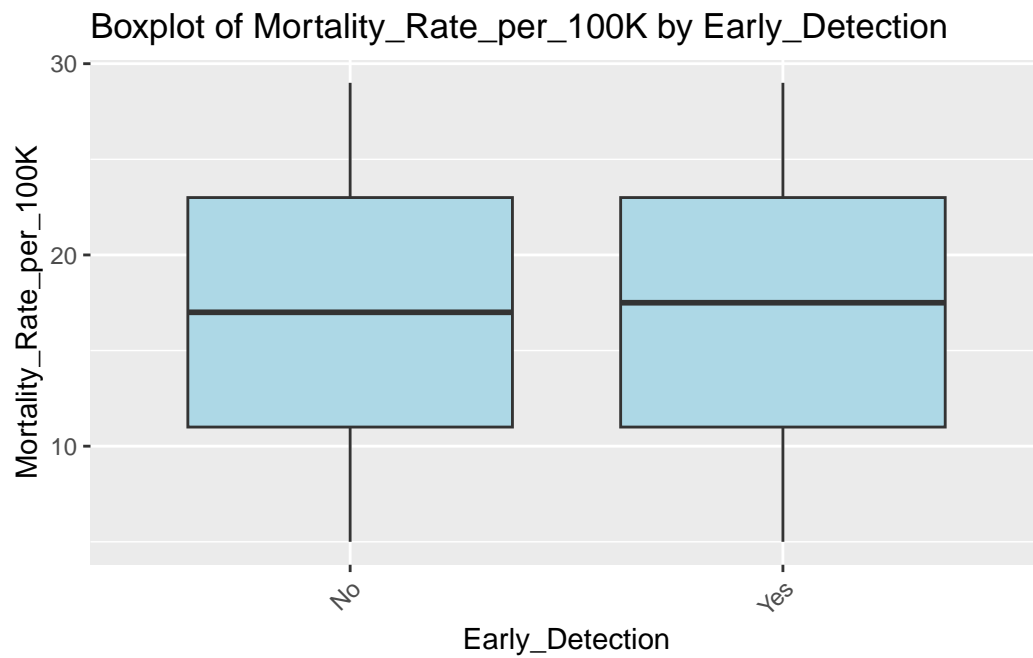
Boxplot: Mortality\_Rate\_per\_100K by Genetic\_Mutation



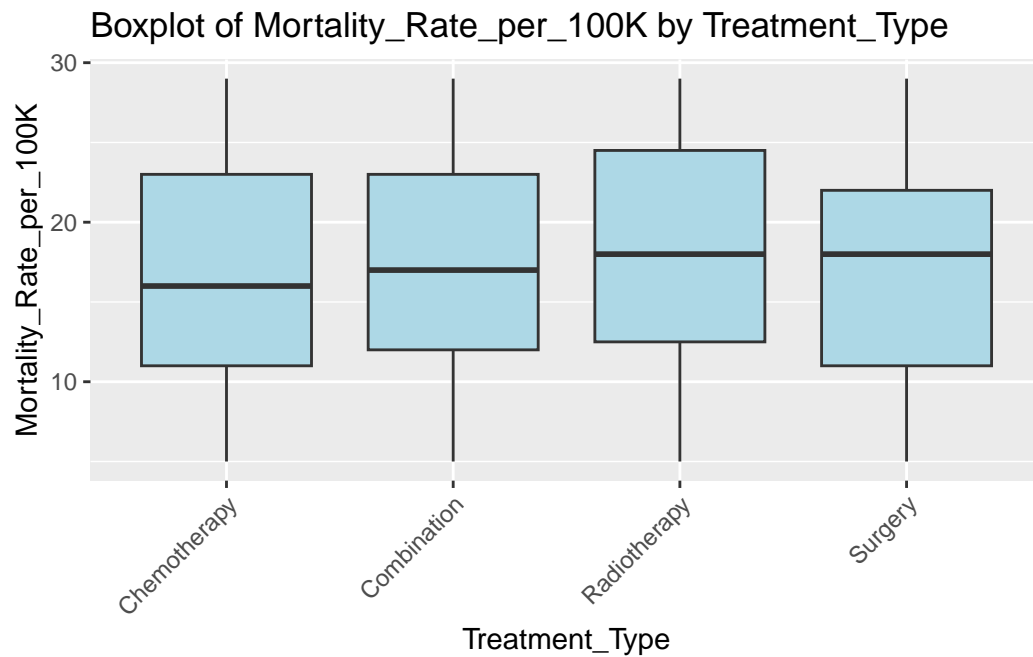
Boxplot: Mortality\_Rate\_per\_100K by Screening\_History



Boxplot: Mortality\_Rate\_per\_100K by Early\_Detection

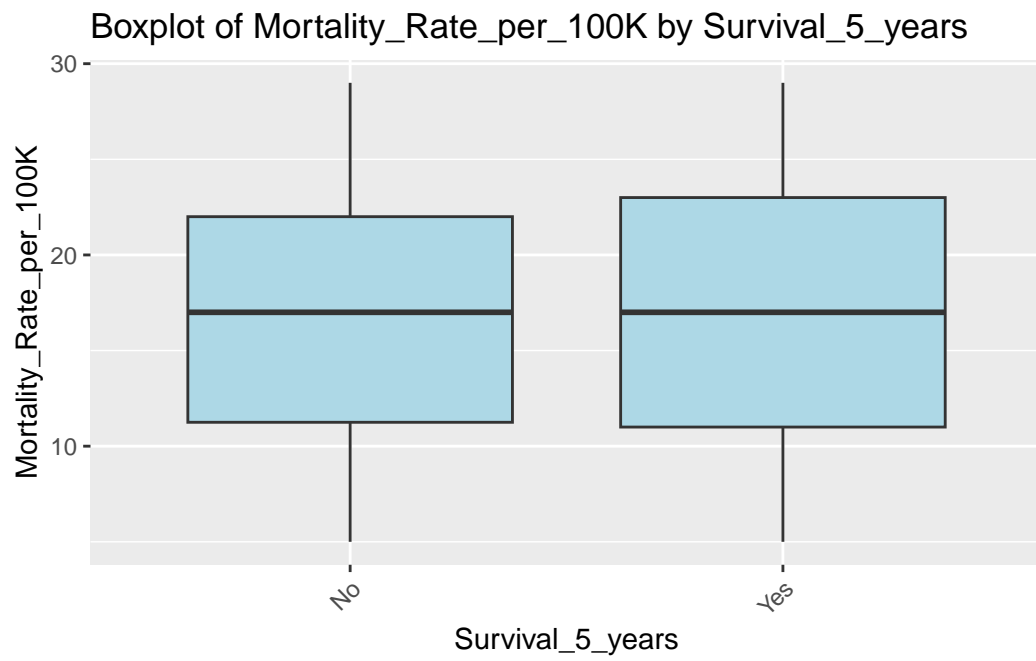


Boxplot: Mortality\_Rate\_per\_100K by Treatment\_Type

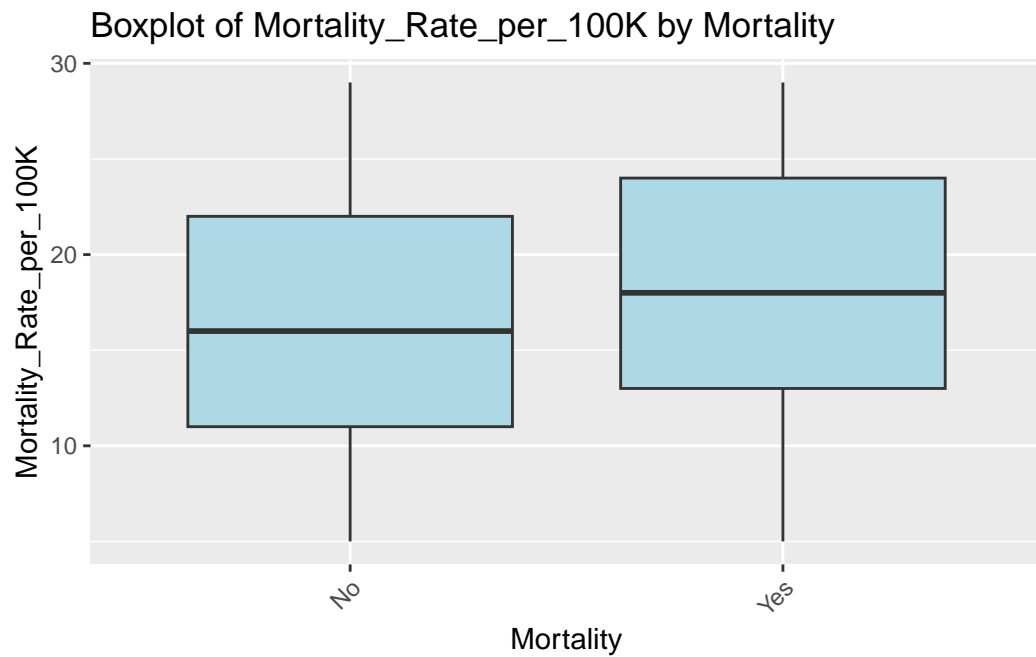




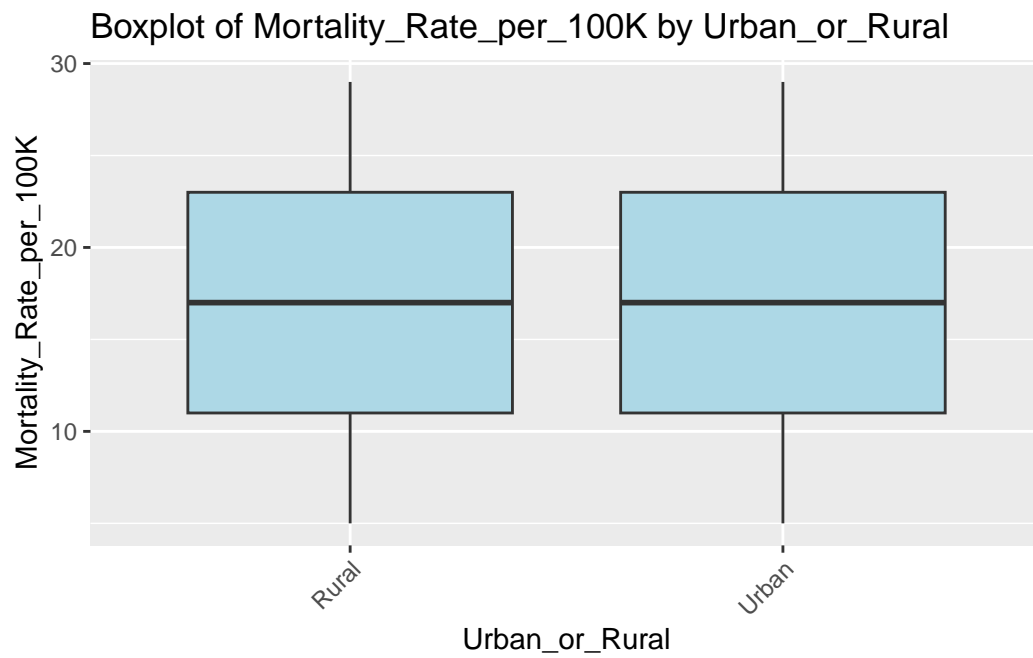
Boxplot: Mortality\_Rate\_per\_100K by Survival\_5\_years



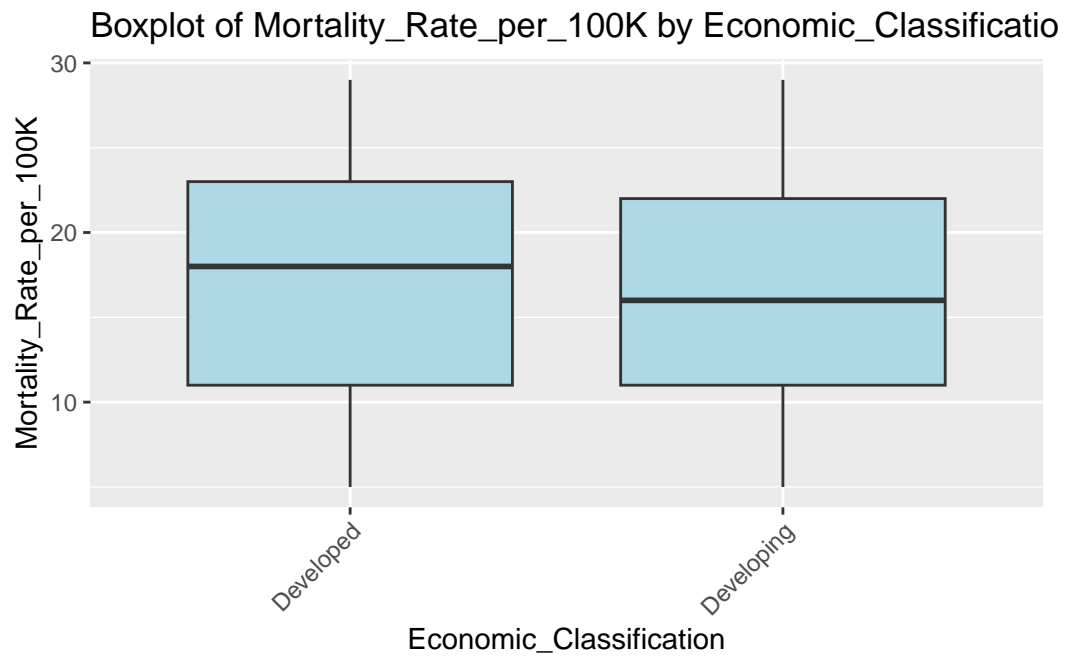
Boxplot: Mortality\_Rate\_per\_100K by Mortality



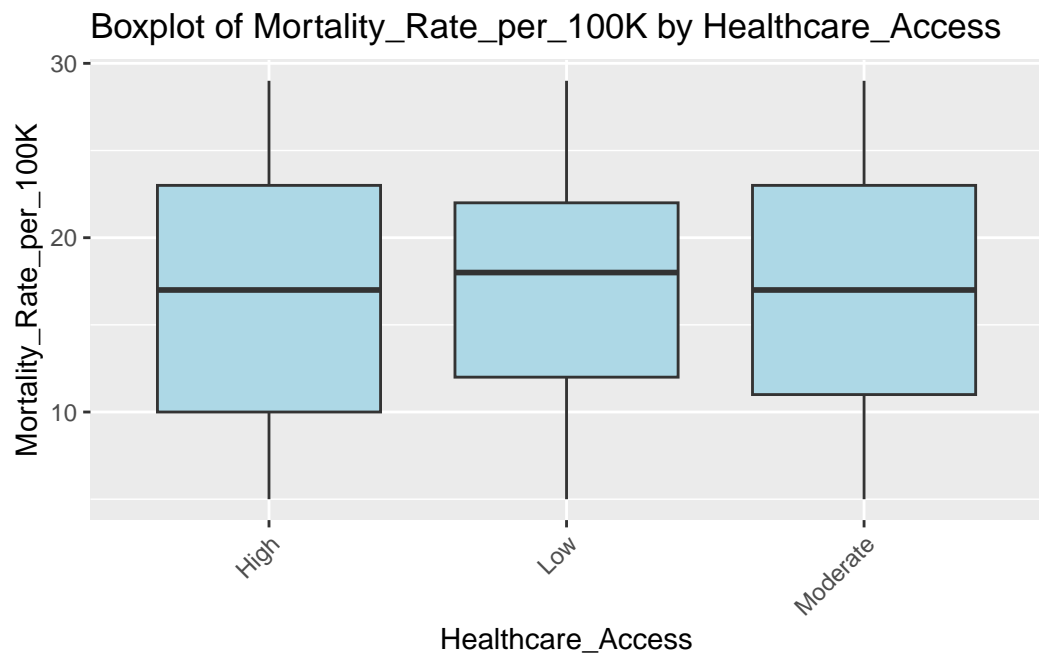
Boxplot: Mortality\_Rate\_per\_100K by Urban\_or\_Rural



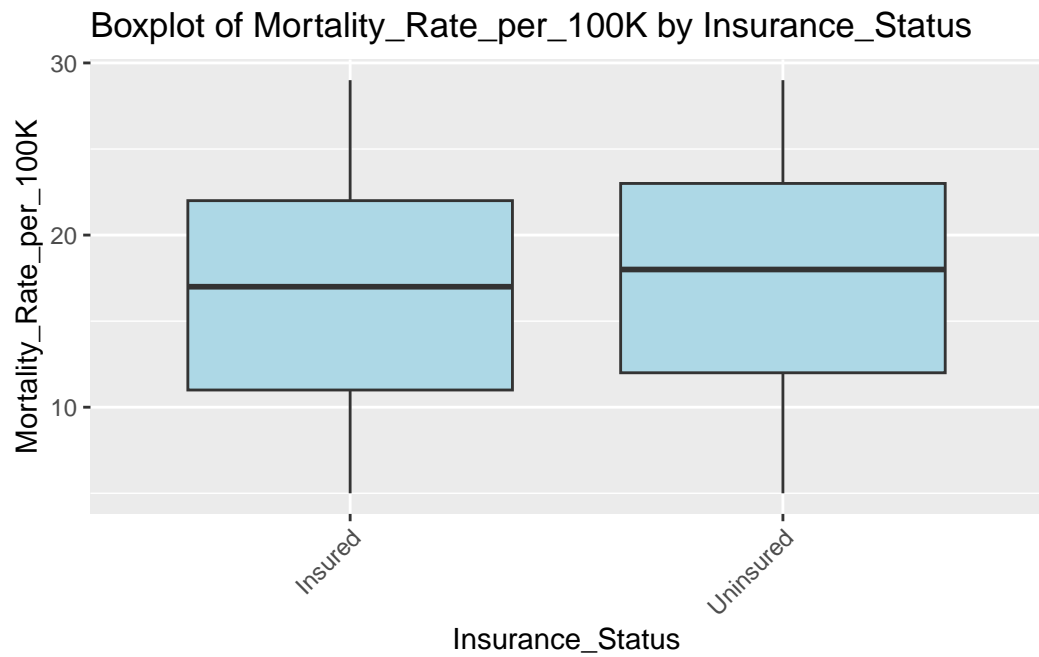
Boxplot: Mortality\_Rate\_per\_100K by Economic\_Classification



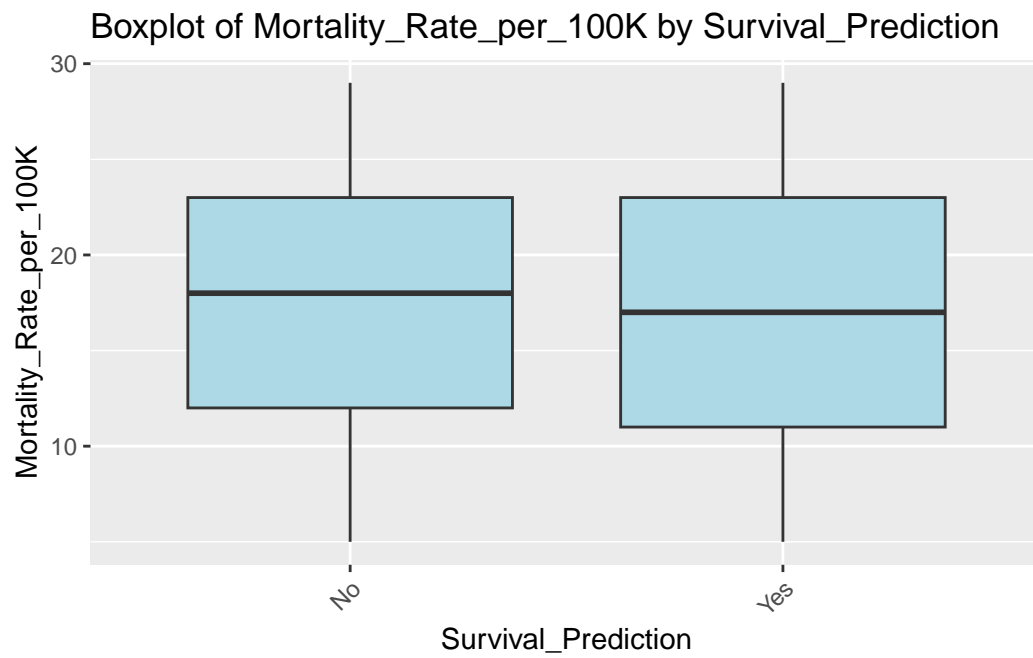
Boxplot: Mortality\_Rate\_per\_100K by Healthcare\_Access



Boxplot: Mortality\_Rate\_per\_100K by Insurance\_Status

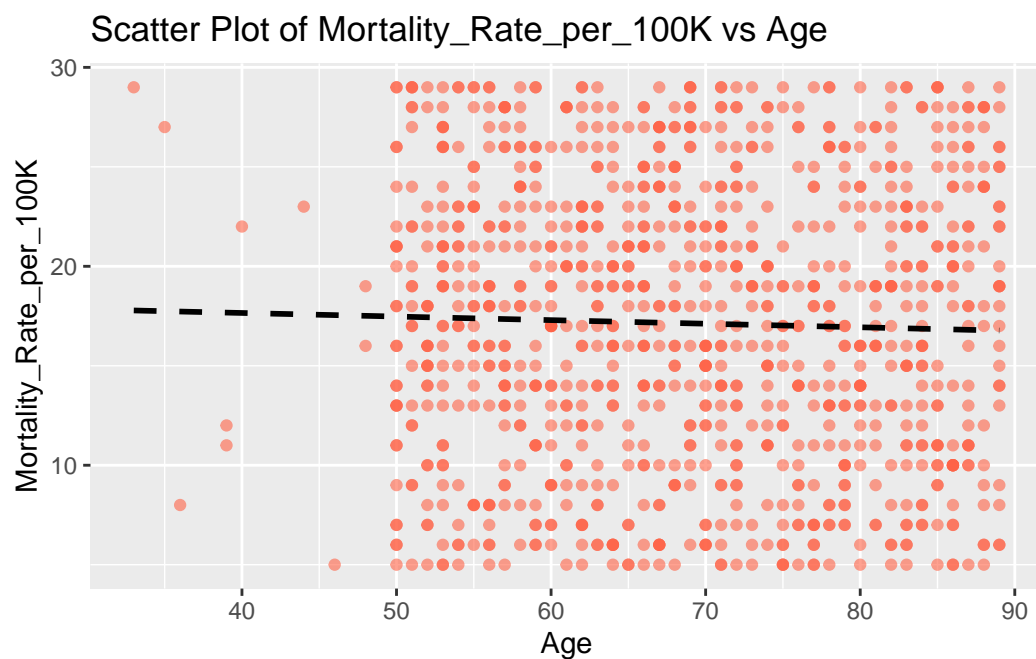


Boxplot: Mortality\_Rate\_per\_100K by Survival\_Prediction

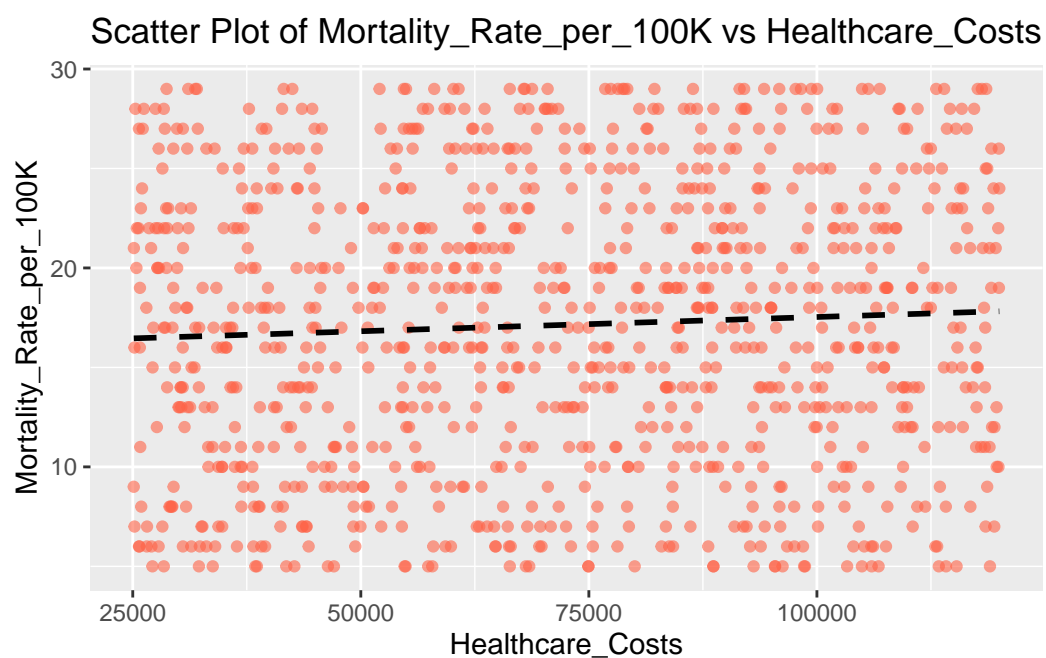


### Continuous Variables

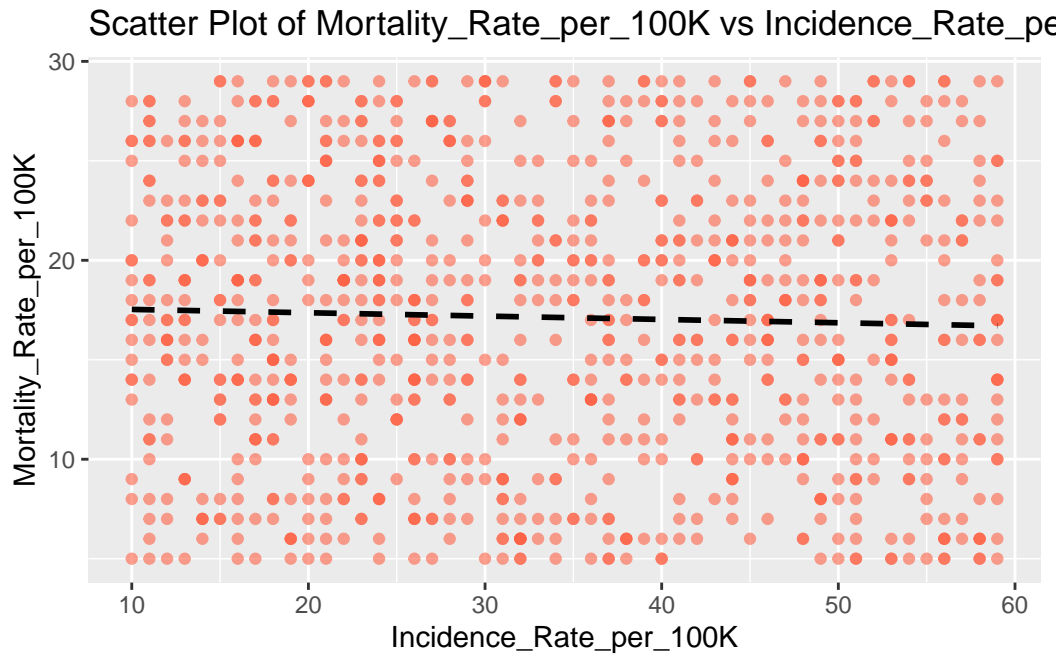
Scatter plot: Mortality\_Rate\_per\_100K vs Age



Scatter plot: Mortality\_Rate\_per\_100K vs Healthcare\_Costs

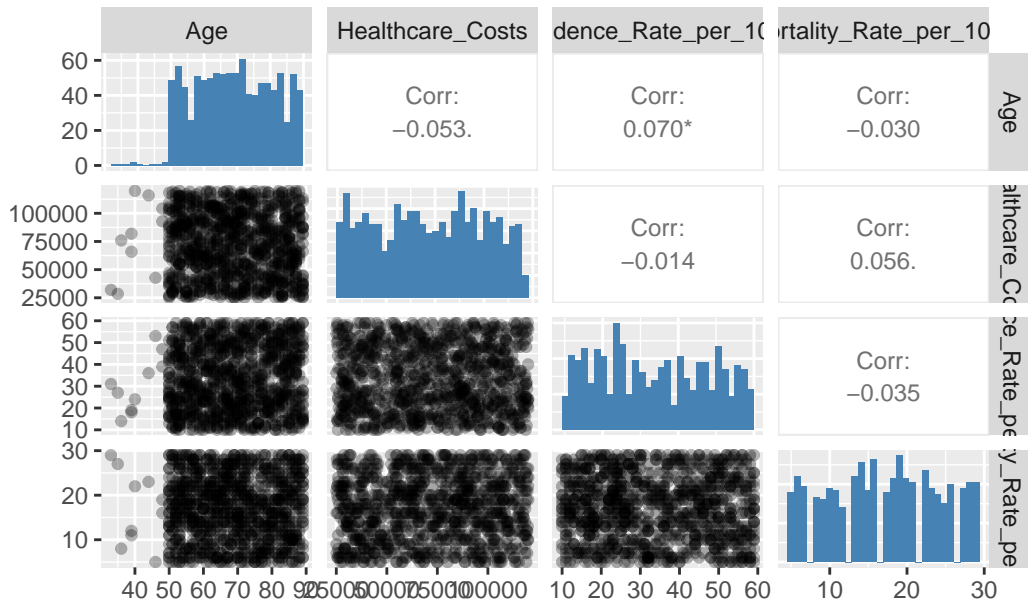


Scatter plot: Mortality\_Rate\_per\_100K vs Incidence\_Rate\_per\_100K



### Summary Correlations

#### Pairwise Plot of Selected Variables



### 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