

Data Analysis Report

Data Summary

count:

{'Pregnancies': 768.0, 'Glucose': 768.0, 'BloodPressure': 768.0, 'SkinThickness': 768.0, 'Insulin': 768.0, 'BMI': 768.0, 'DiabetesPedigreeFunction': 768.0, 'Age': 768.0, 'Outcome': 768.0}

mean:

{'Pregnancies': 3.8450520833333335, 'Glucose': 120.89453125, 'BloodPressure': 69.10546875, 'SkinThickness': 20.536458333333332, 'Insulin': 79.79947916666667, 'BMI': 31.992578124999998, 'DiabetesPedigreeFunction': 0.47187630208333325, 'Age': 33.240885416666664, 'Outcome': 0.3489583333333333}

std:

{'Pregnancies': 3.3695780626988694, 'Glucose': 31.97261819513622, 'BloodPressure': 19.355807170644777, 'SkinThickness': 15.952217567727637, 'Insulin': 115.24400235133817, 'BMI': 7.884160320375446, 'DiabetesPedigreeFunction': 0.3313285950127749, 'Age': 11.760231540678685, 'Outcome': 0.47695137724279896}

min:

{'Pregnancies': 0.0, 'Glucose': 0.0, 'BloodPressure': 0.0, 'SkinThickness': 0.0, 'Insulin': 0.0, 'BMI': 0.0, 'DiabetesPedigreeFunction': 0.078, 'Age': 21.0, 'Outcome': 0.0}

25%:

{'Pregnancies': 1.0, 'Glucose': 99.0, 'BloodPressure': 62.0, 'SkinThickness': 0.0, 'Insulin': 0.0, 'BMI': 27.3, 'DiabetesPedigreeFunction': 0.24375, 'Age': 24.0, 'Outcome': 0.0}

50%:

{'Pregnancies': 3.0, 'Glucose': 117.0, 'BloodPressure': 72.0, 'SkinThickness': 23.0, 'Insulin': 30.5,

'BMI': 32.0, 'DiabetesPedigreeFunction': 0.3725, 'Age': 29.0, 'Outcome': 0.0}

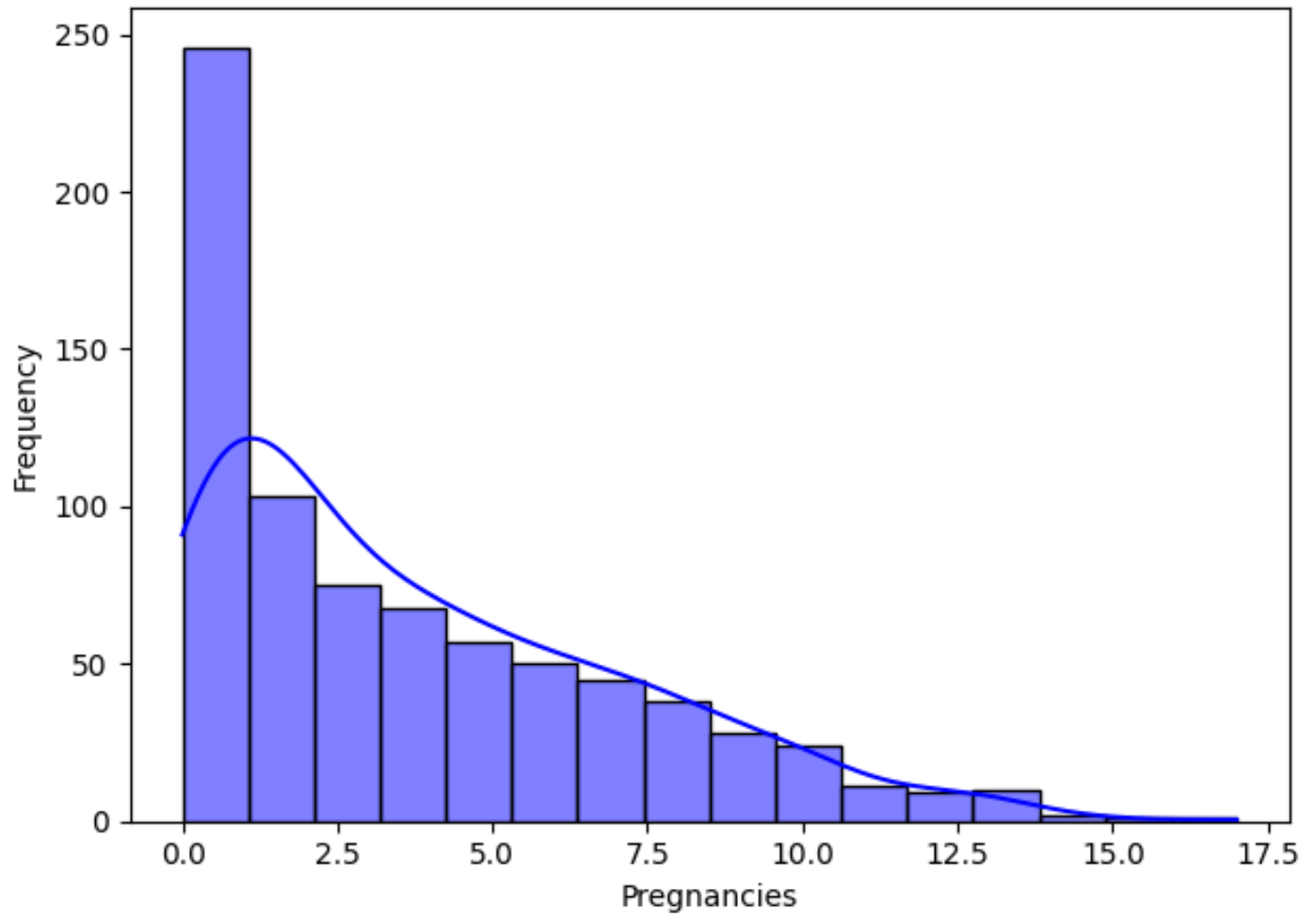
75%:

{'Pregnancies': 6.0, 'Glucose': 140.25, 'BloodPressure': 80.0, 'SkinThickness': 32.0, 'Insulin':
127.25, 'BMI': 36.6, 'DiabetesPedigreeFunction': 0.62625, 'Age': 41.0, 'Outcome': 1.0}

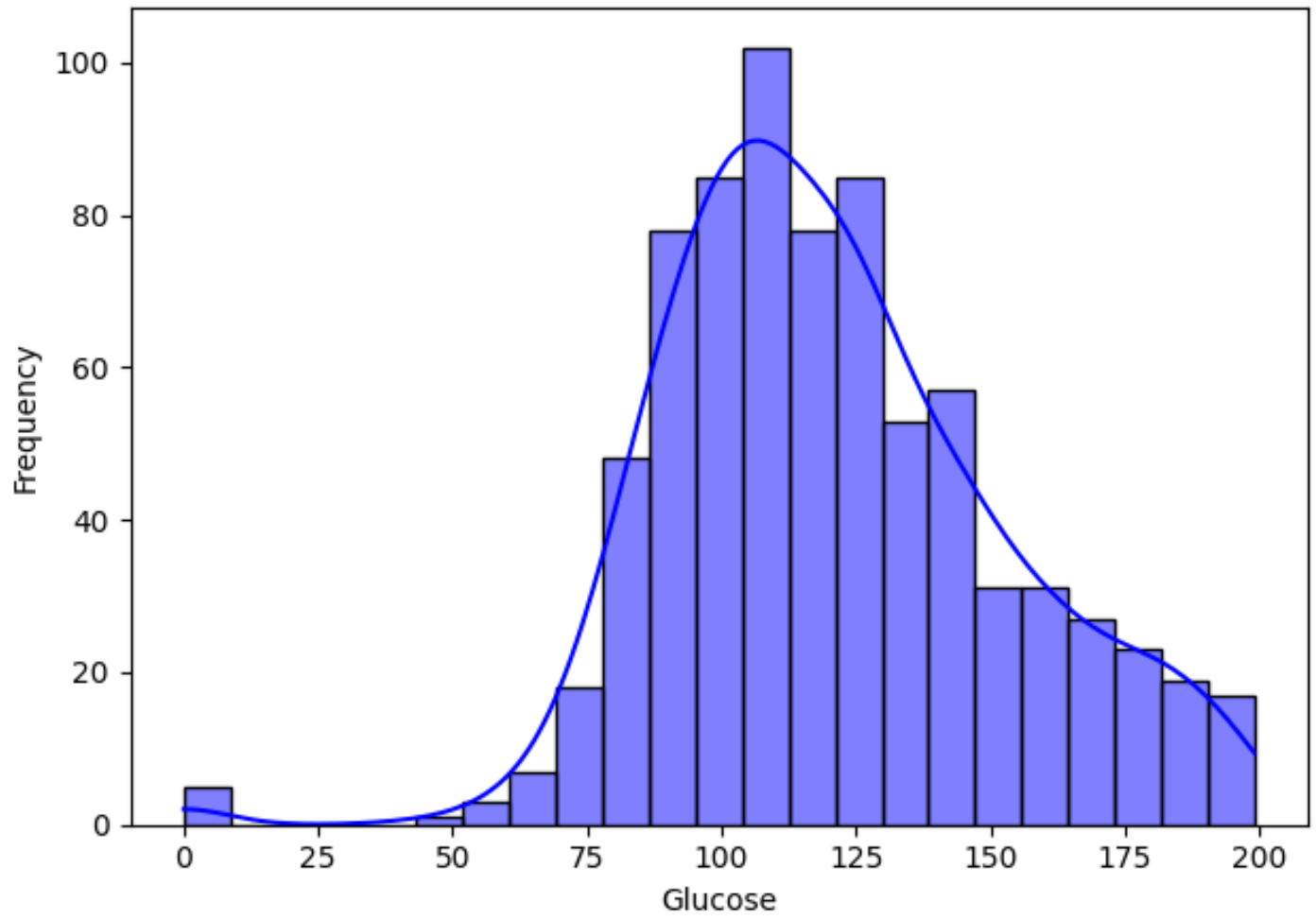
max:

{'Pregnancies': 17.0, 'Glucose': 199.0, 'BloodPressure': 122.0, 'SkinThickness': 99.0, 'Insulin':
846.0, 'BMI': 67.1, 'DiabetesPedigreeFunction': 2.42, 'Age': 81.0, 'Outcome': 1.0}

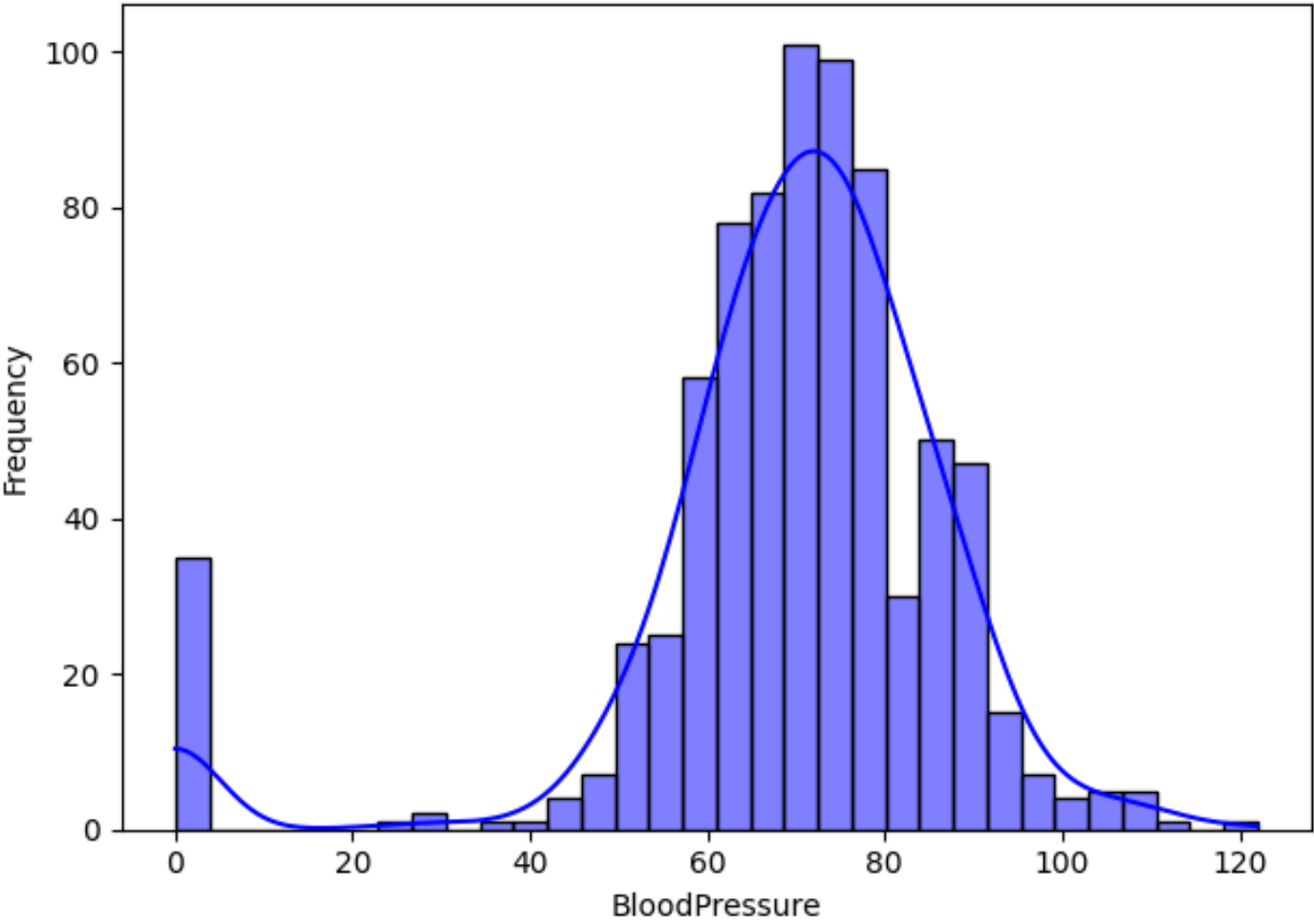
Histogram of Pregnancies



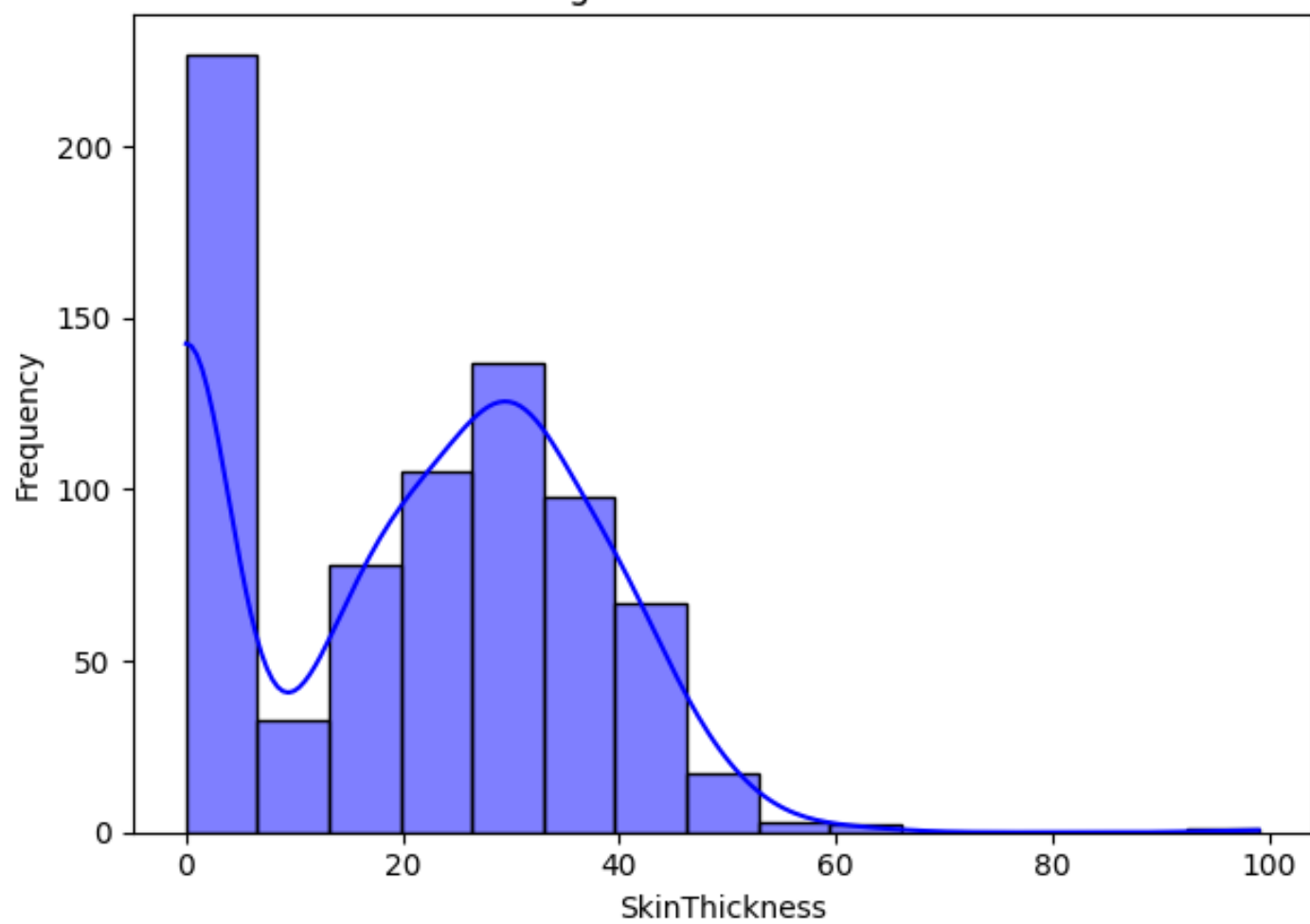
Histogram of Glucose



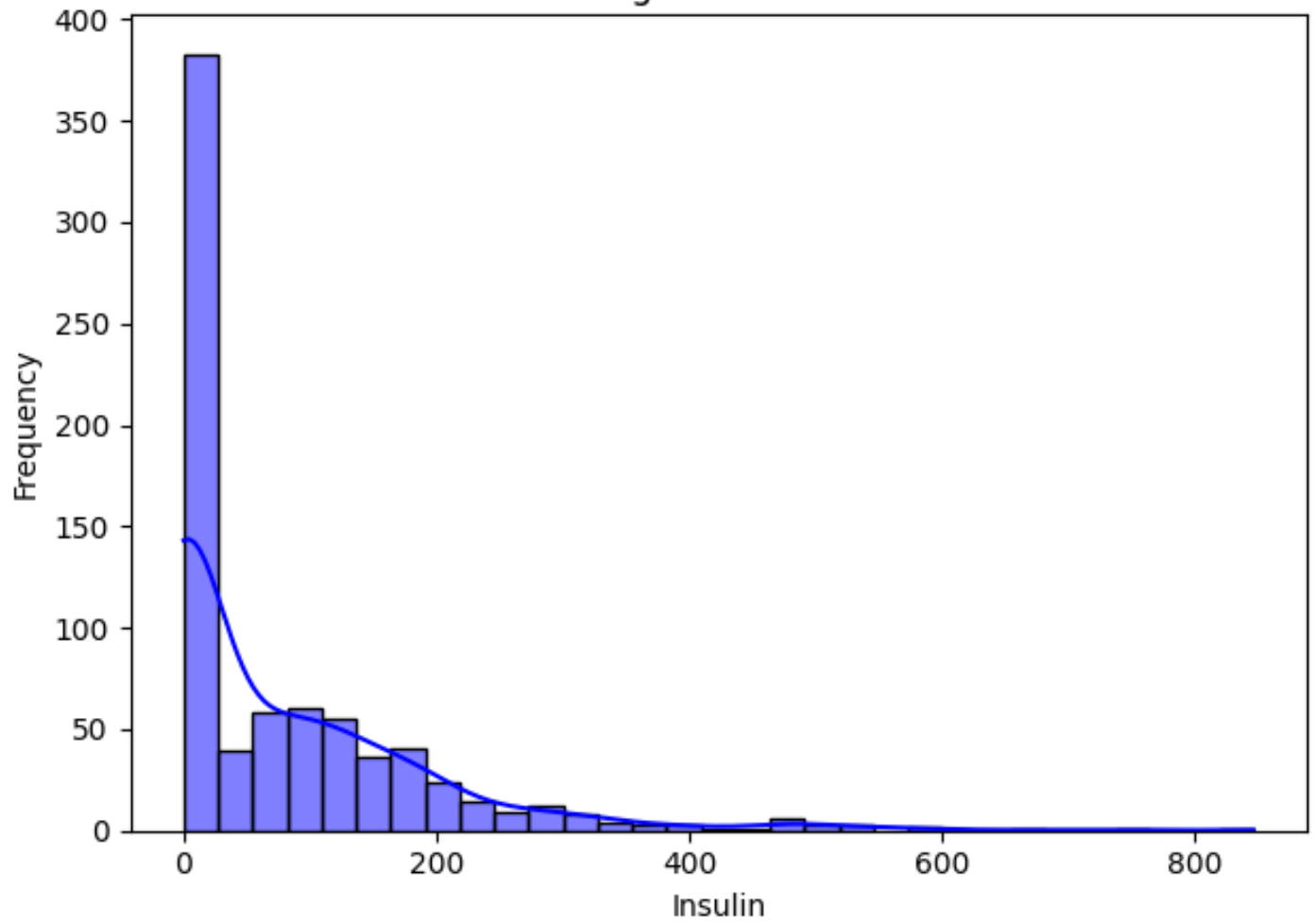
Histogram of BloodPressure



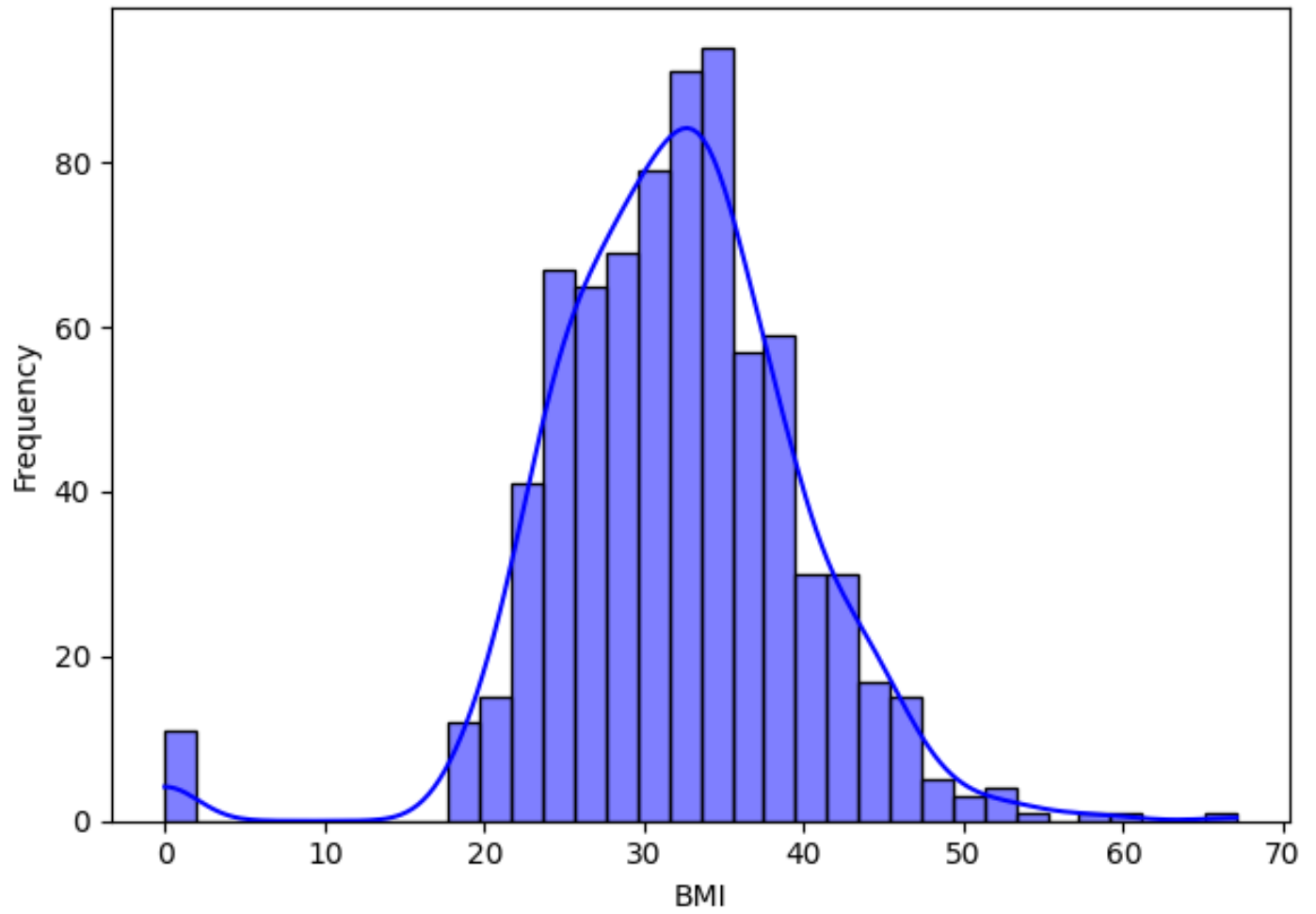
Histogram of SkinThickness



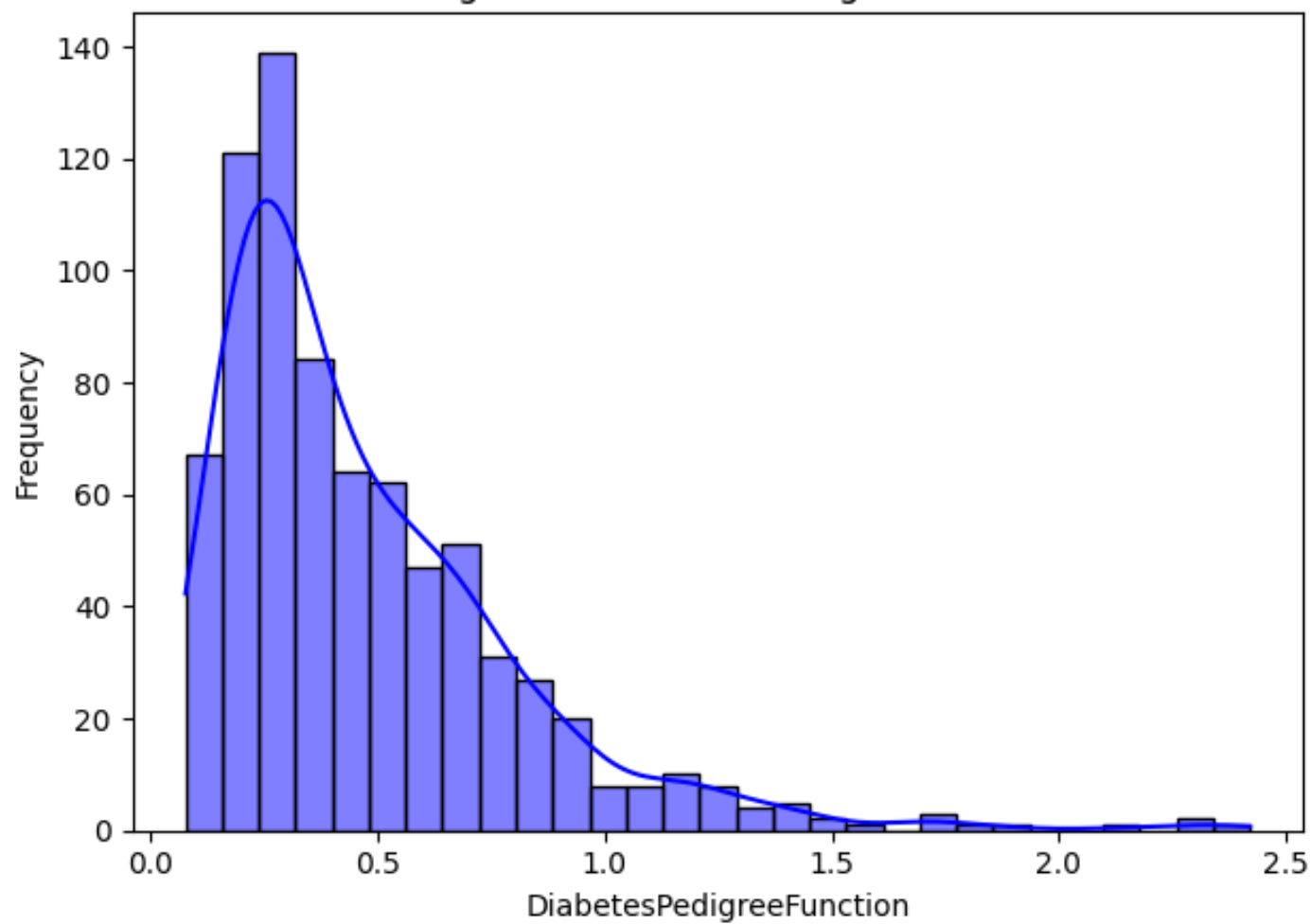
Histogram of Insulin



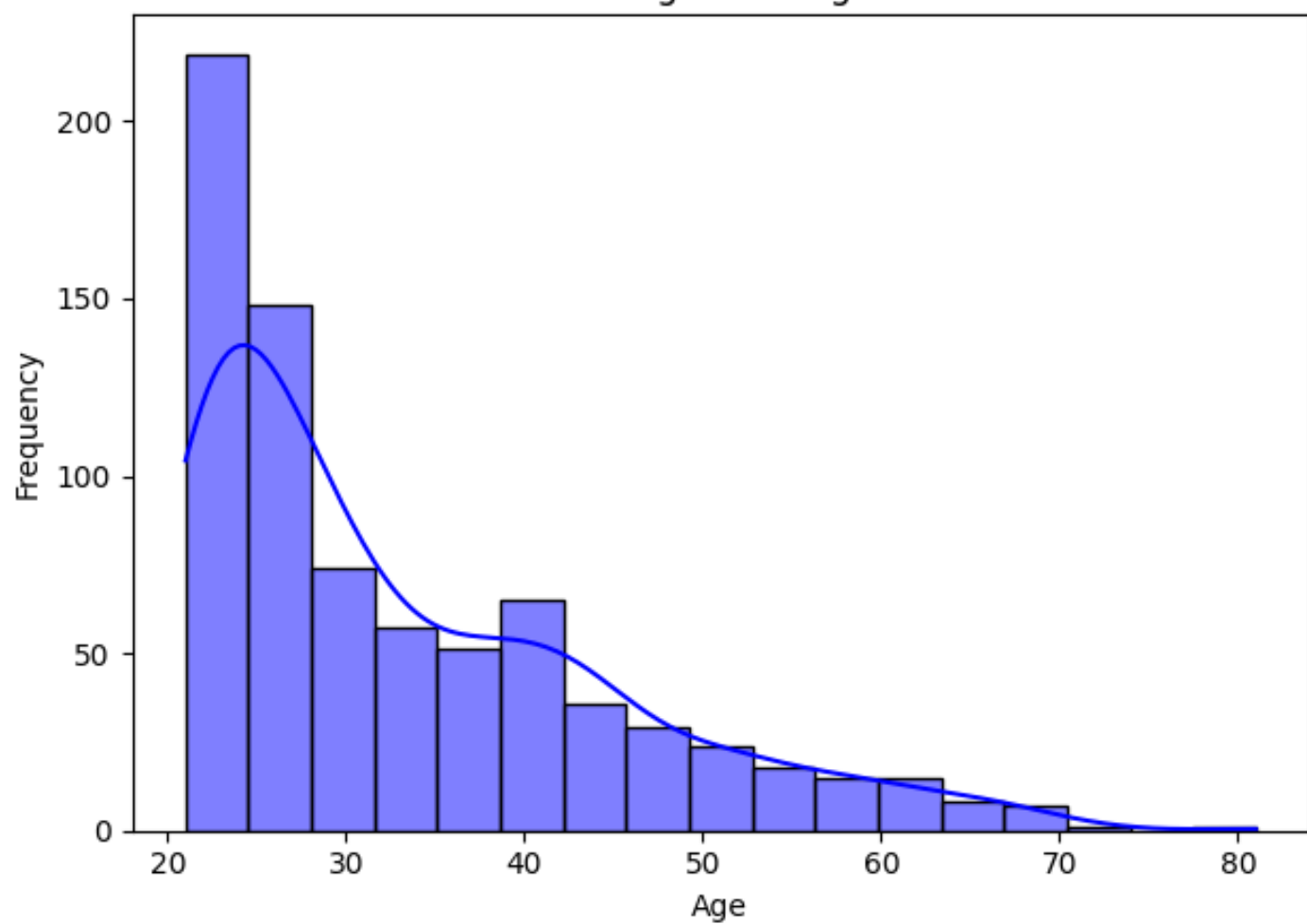
Histogram of BMI



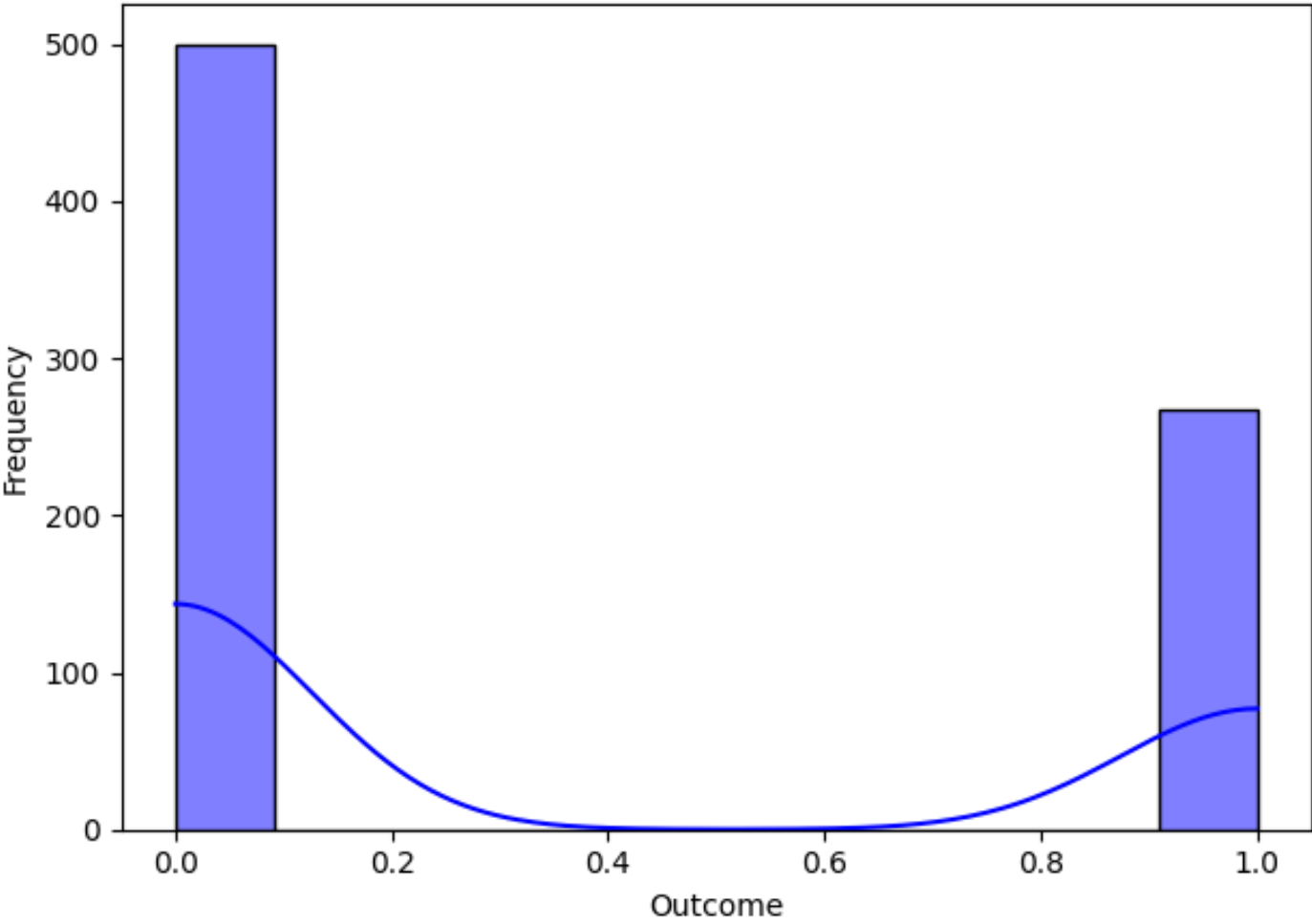
Histogram of DiabetesPedigreeFunction



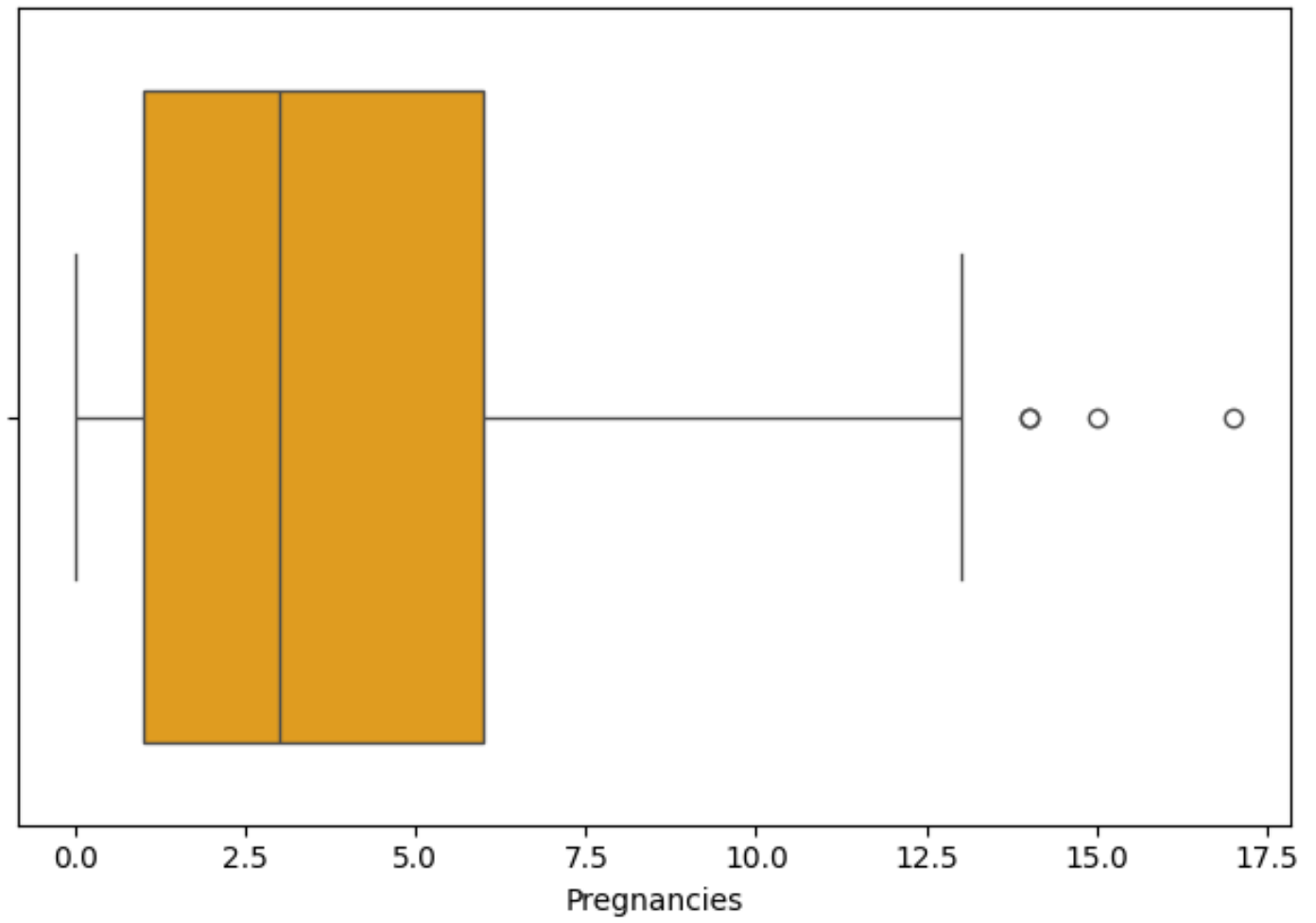
Histogram of Age



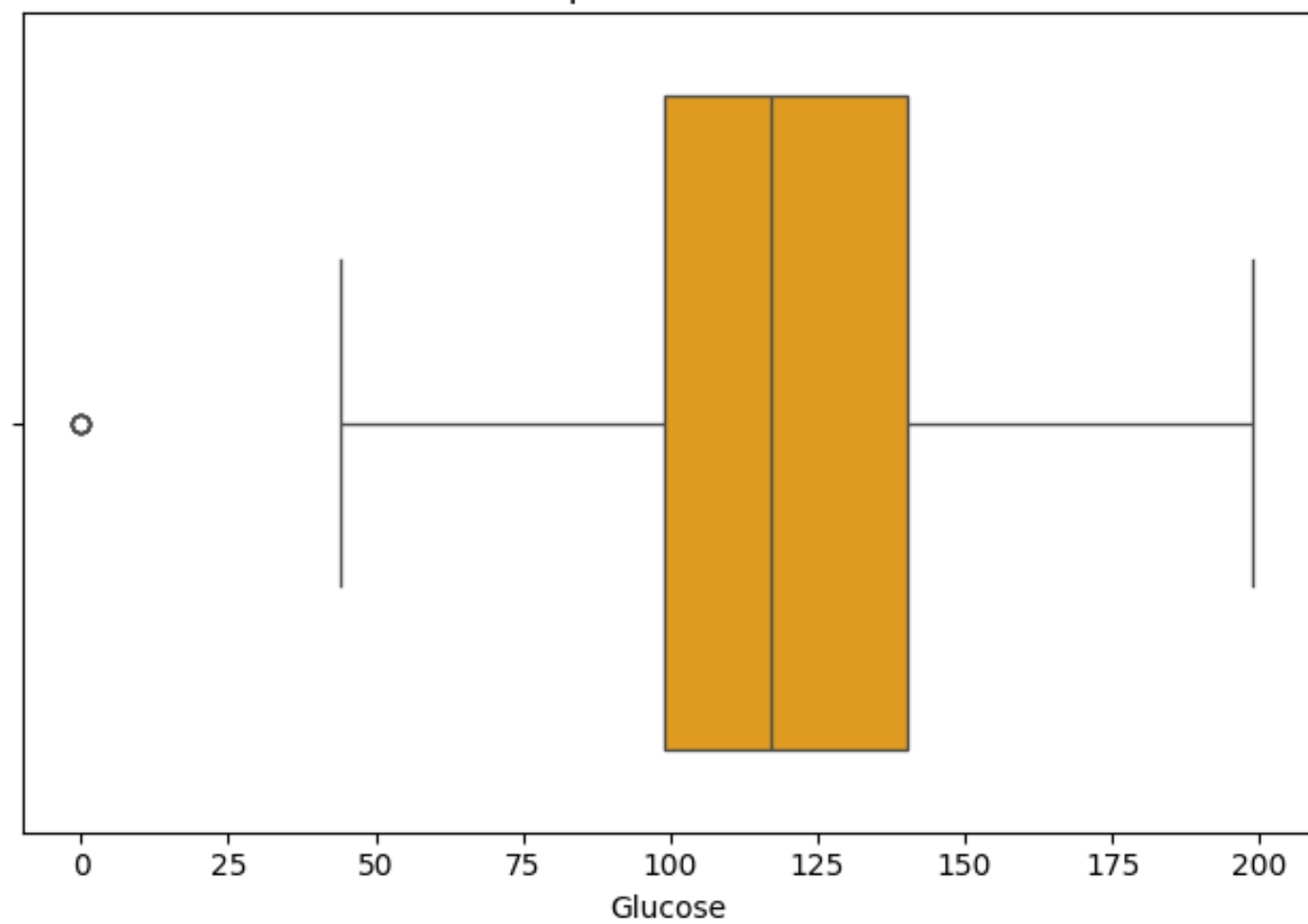
Histogram of Outcome



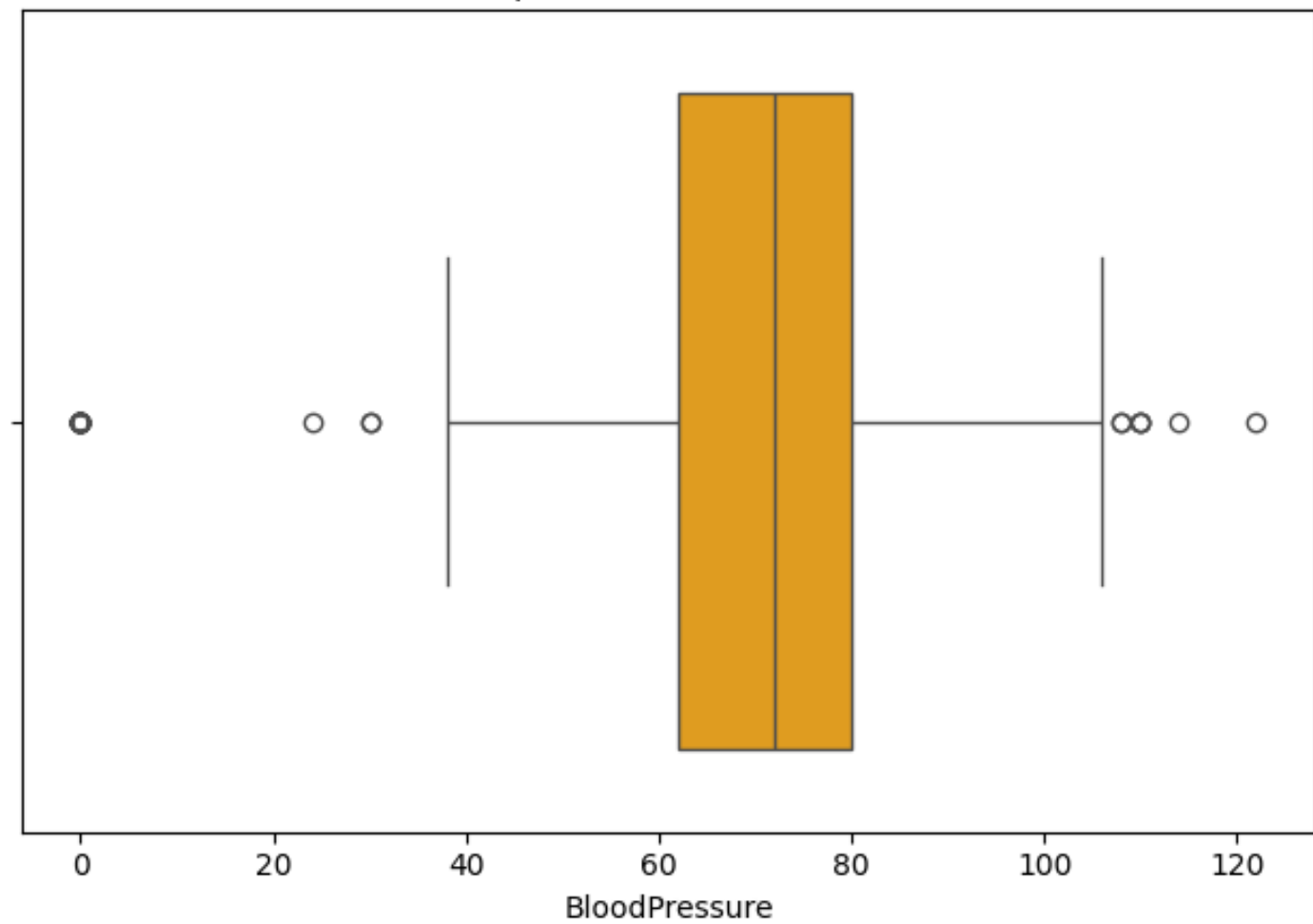
Boxplot of Pregnancies



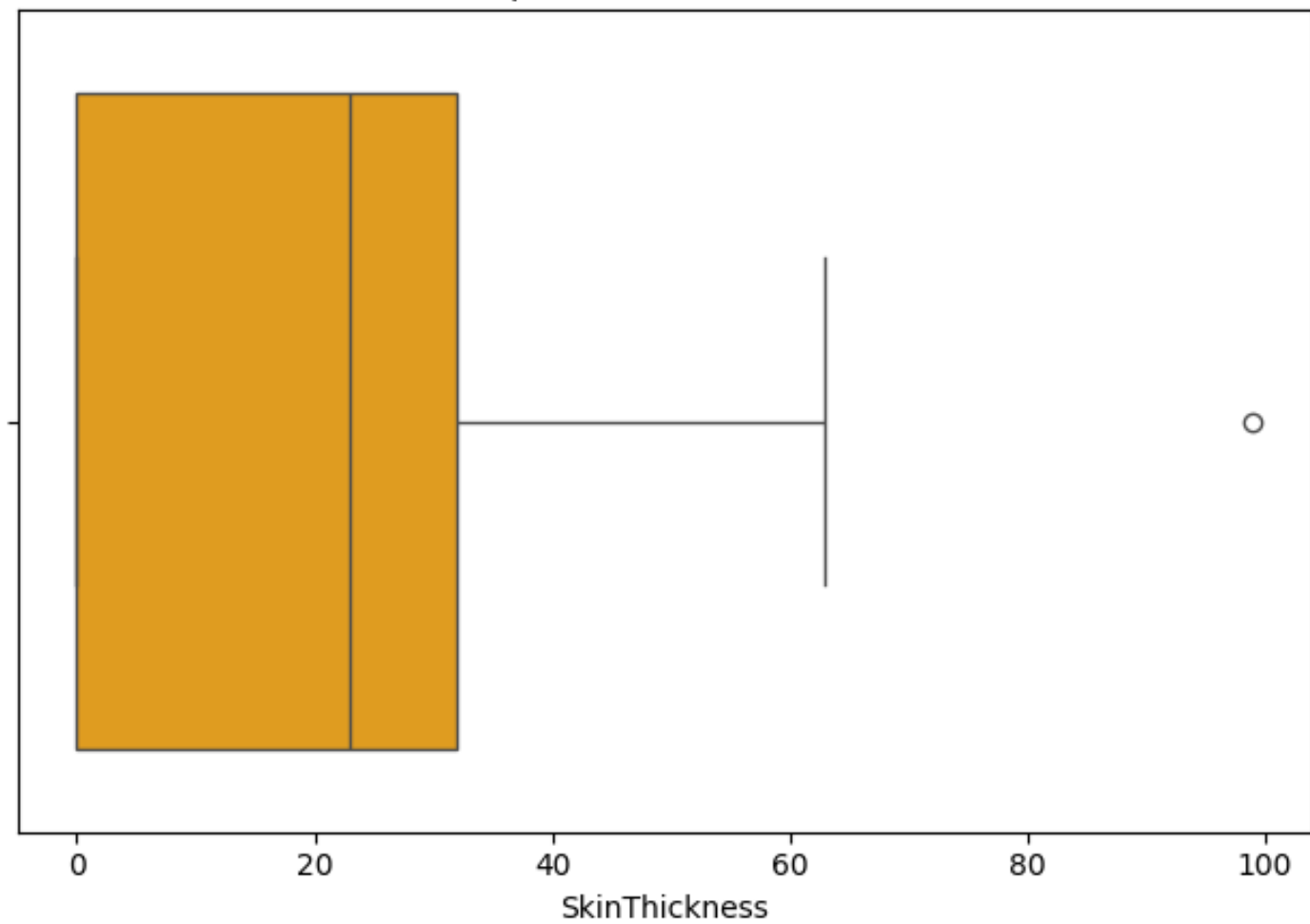
Boxplot of Glucose



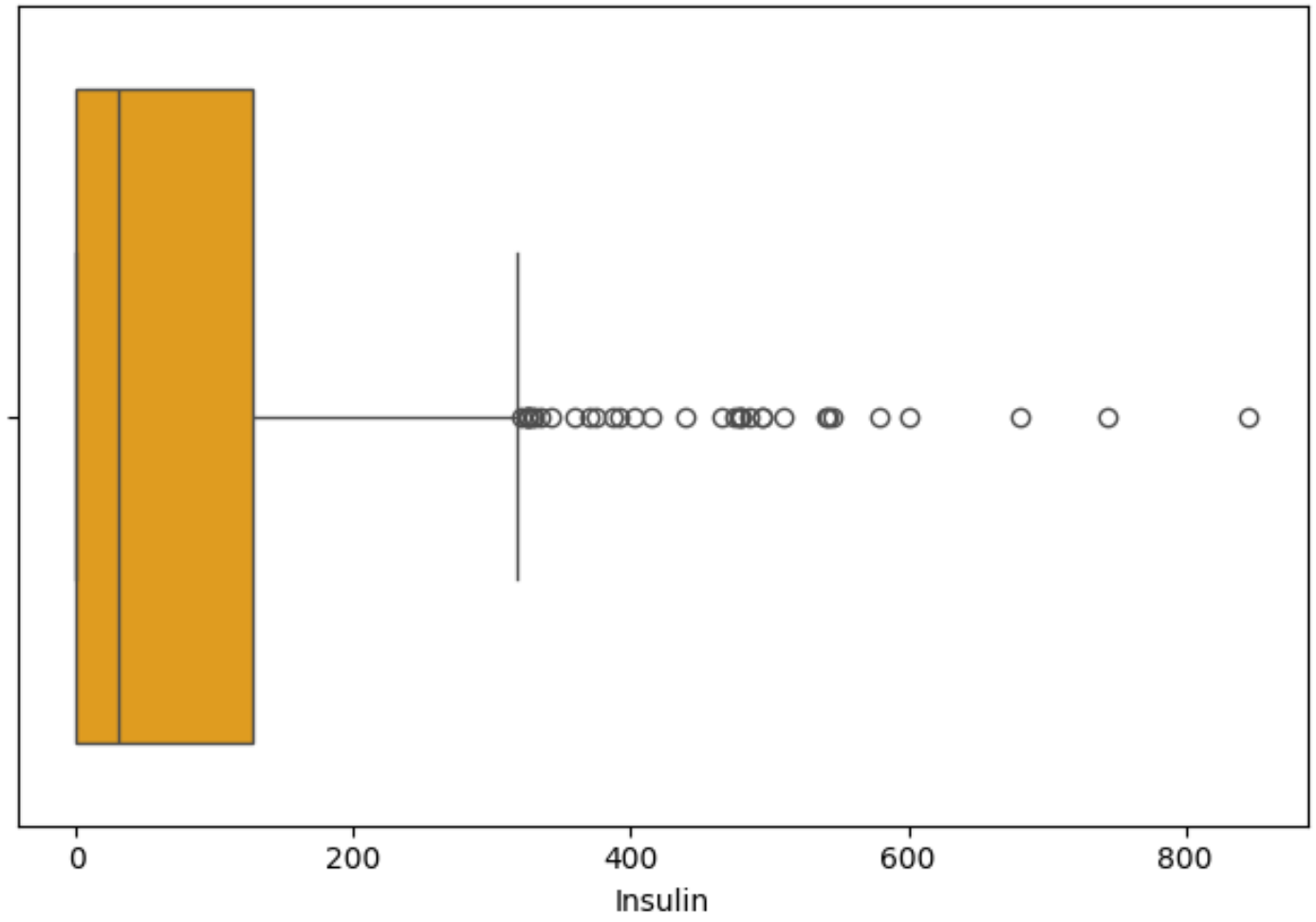
Boxplot of BloodPressure



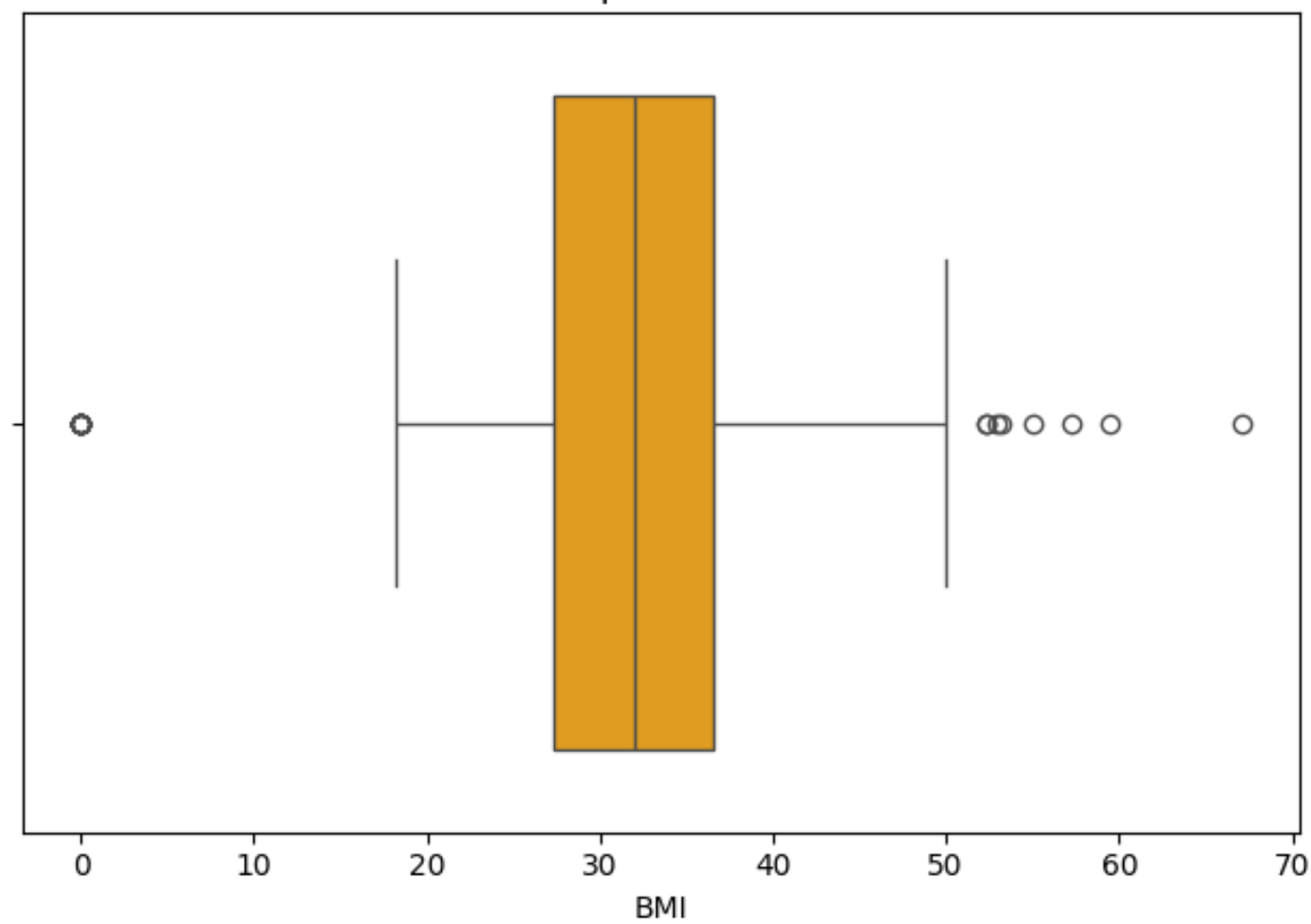
Boxplot of SkinThickness



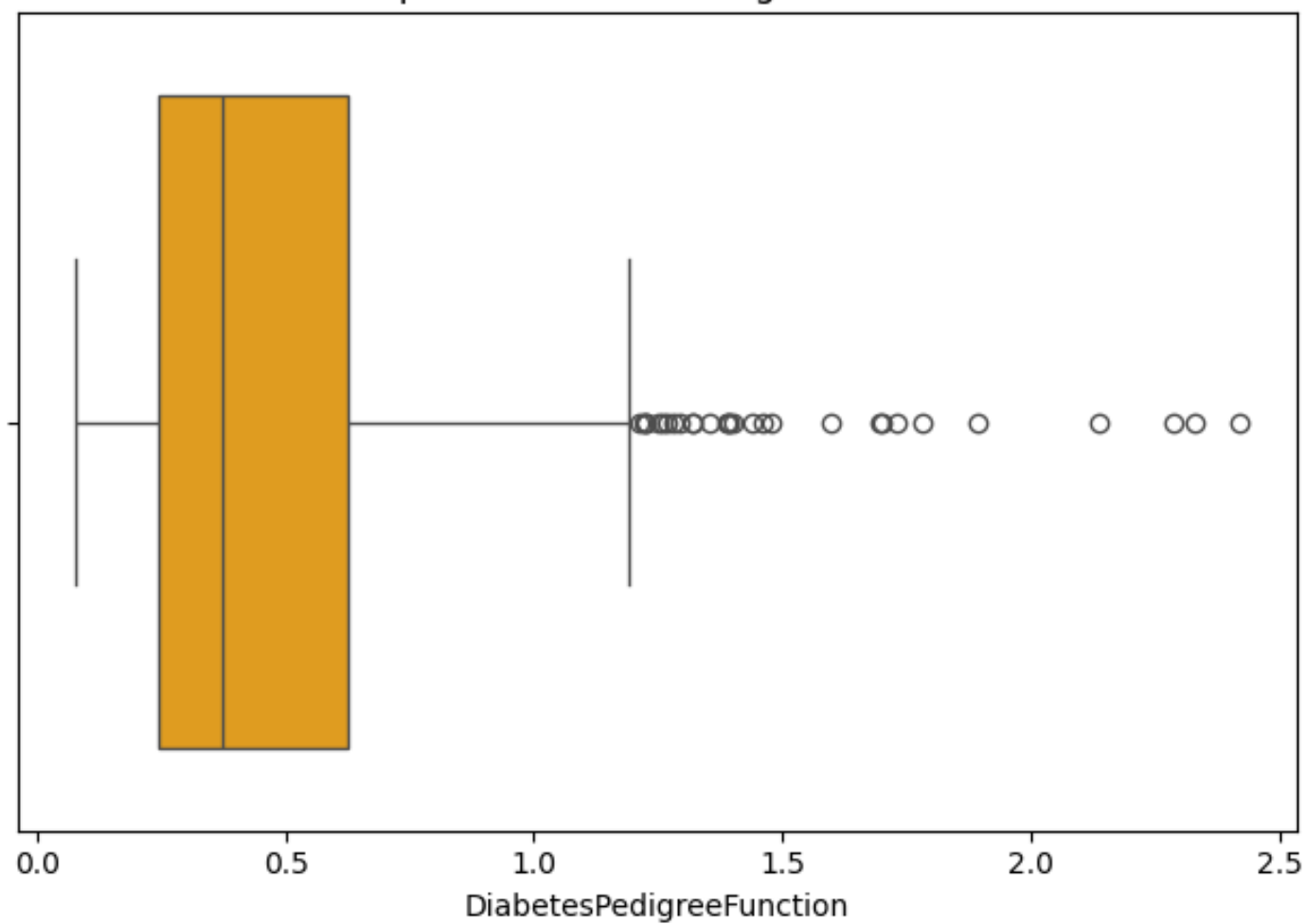
Boxplot of Insulin



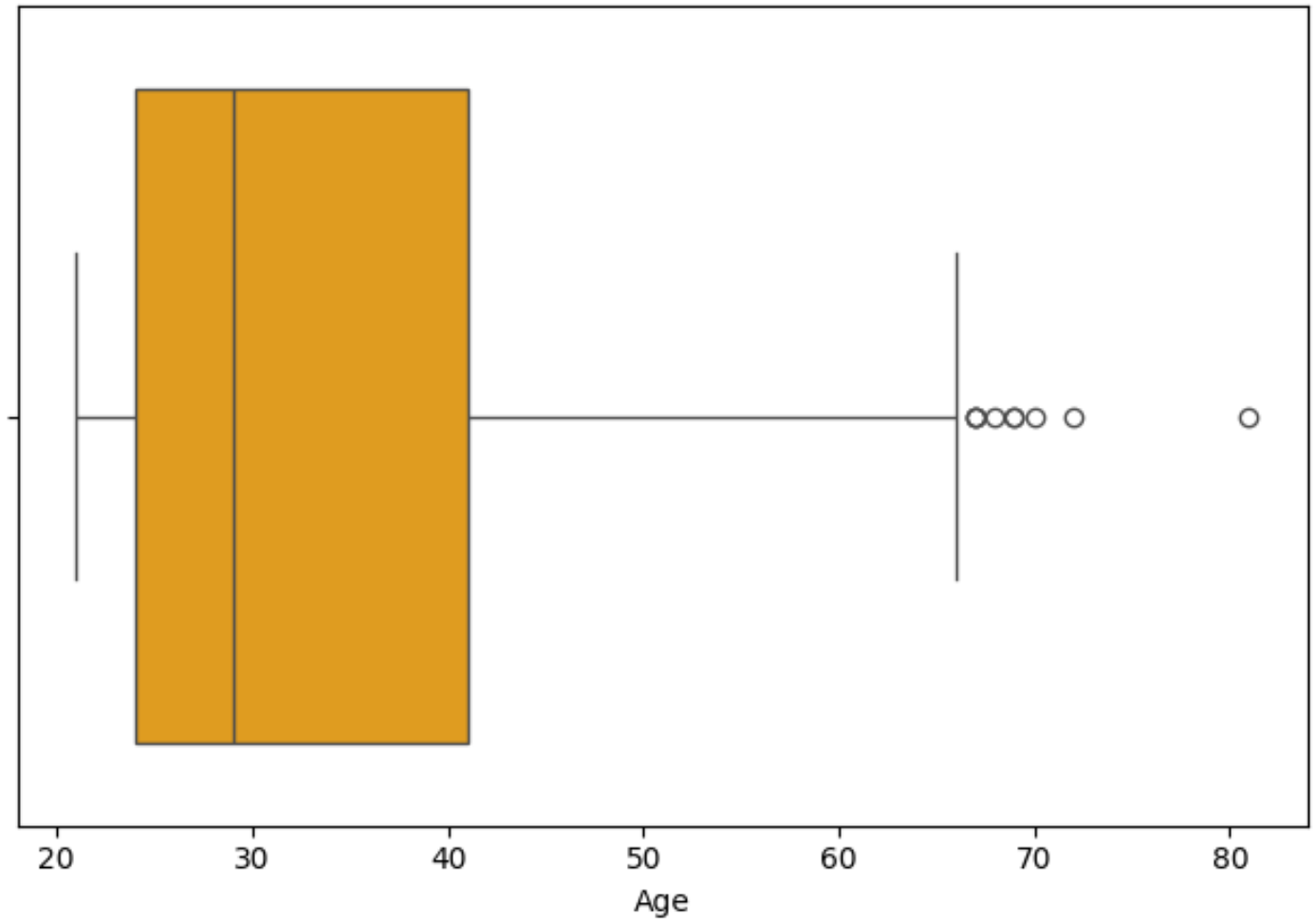
Boxplot of BMI



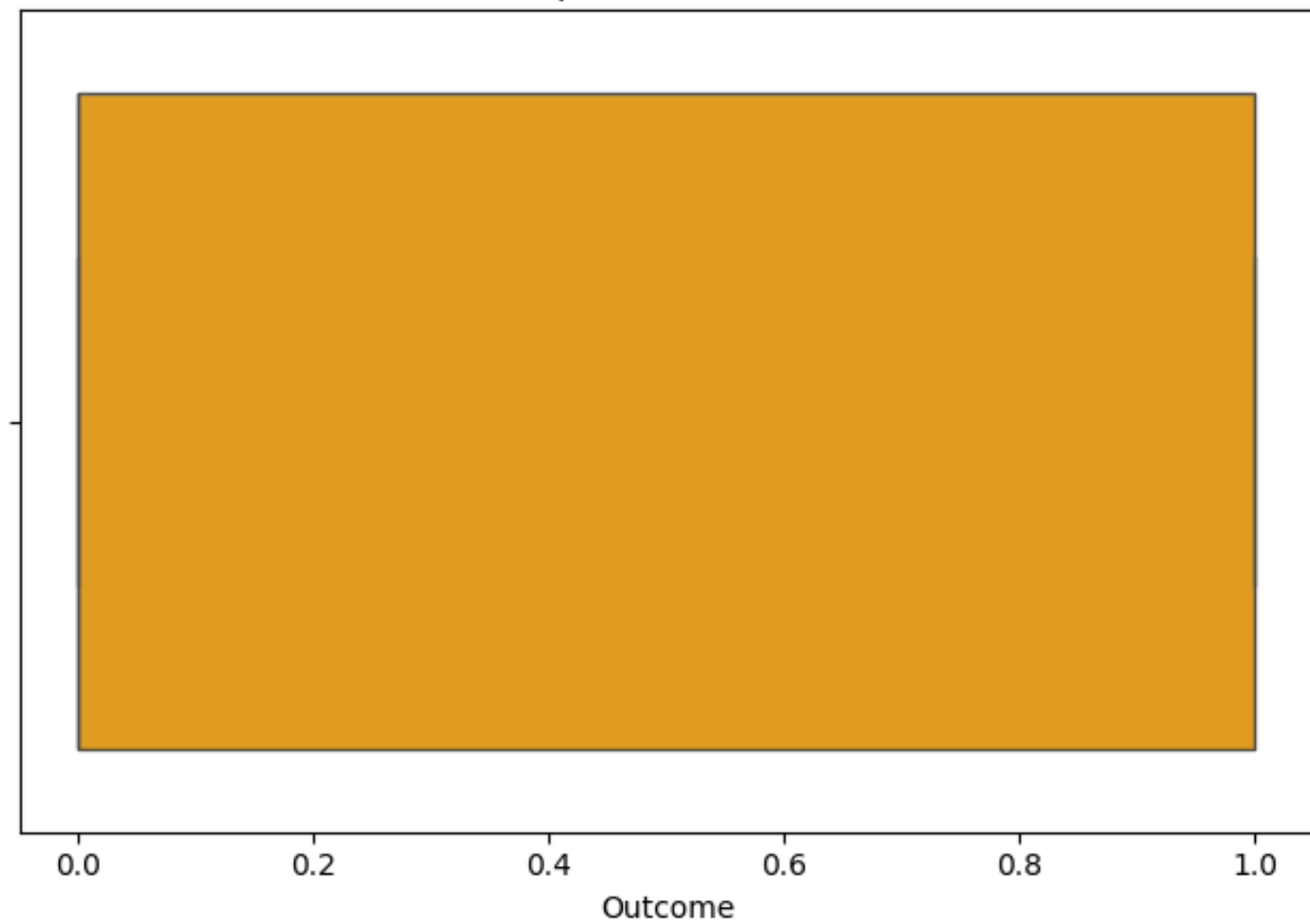
Boxplot of DiabetesPedigreeFunction



Boxplot of Age



Boxplot of Outcome



Model Results

Model: Logistic Regression

accuracy: 0.7359

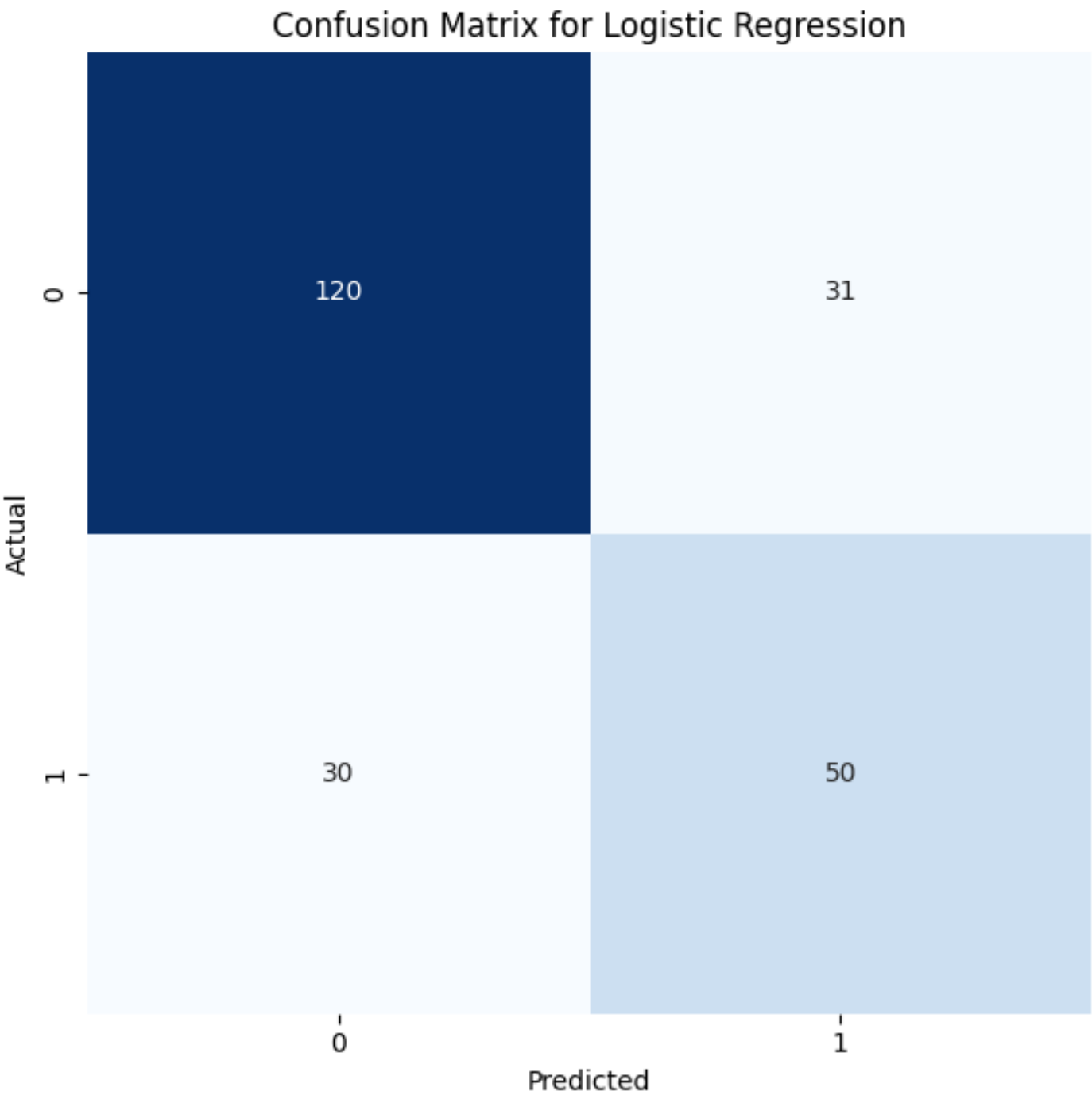
Model: Random Forest

accuracy: 0.7359

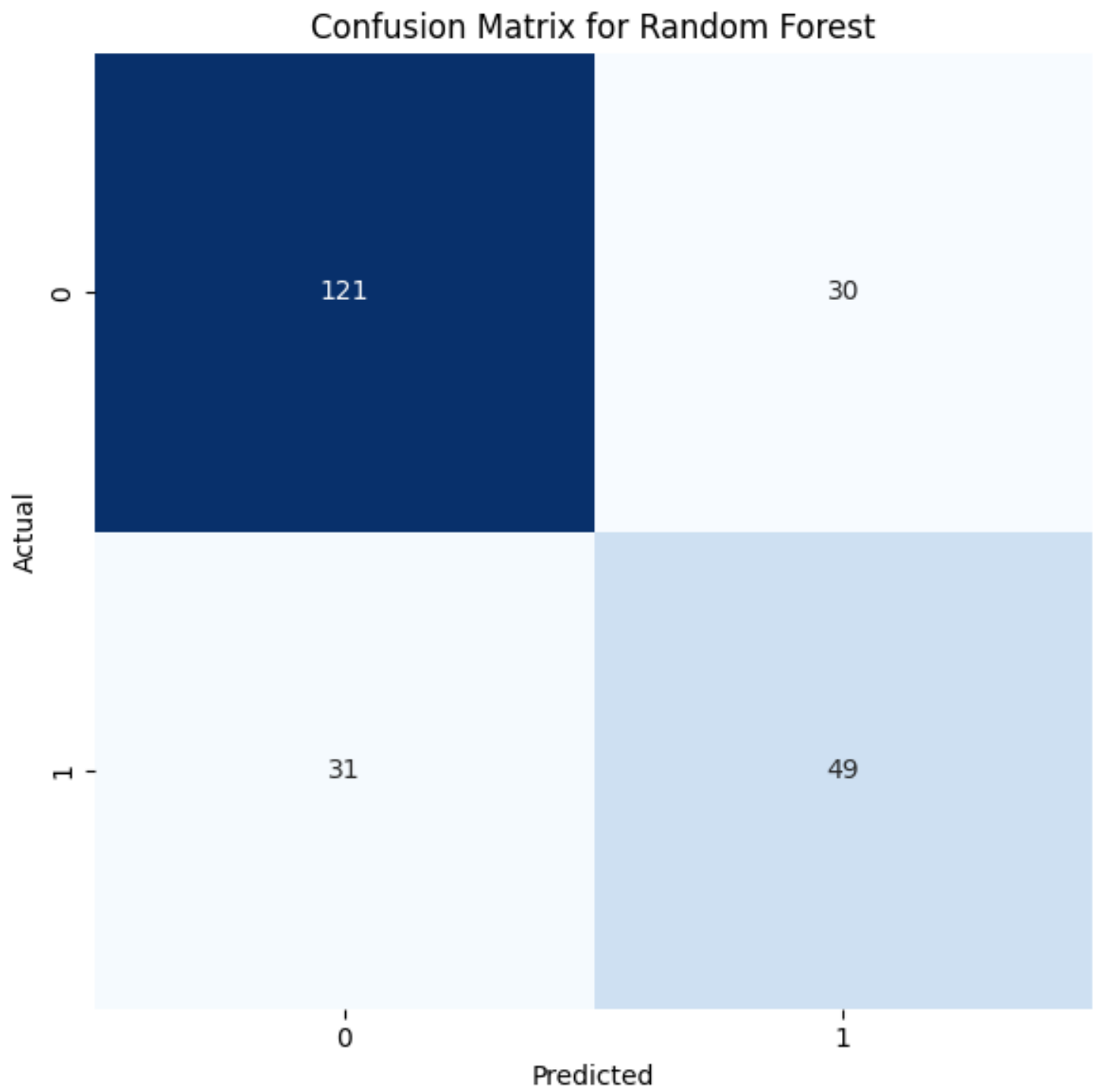
Model: Support Vector Machine

accuracy: 0.7359

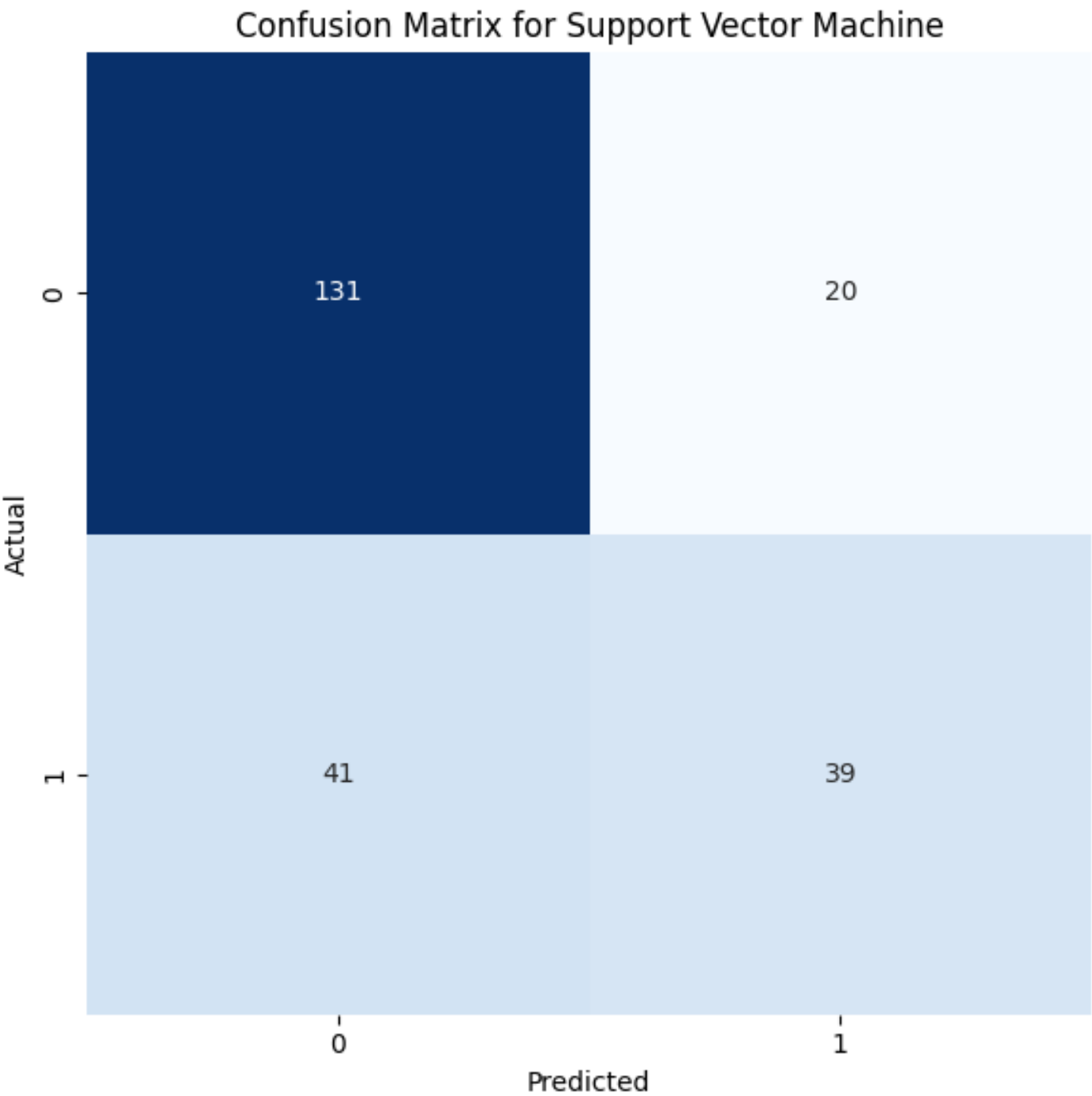
Confusion Matrix for Logistic Regression



Confusion Matrix for Random Forest



Confusion Matrix for Support Vector Machine



Model Comparison

