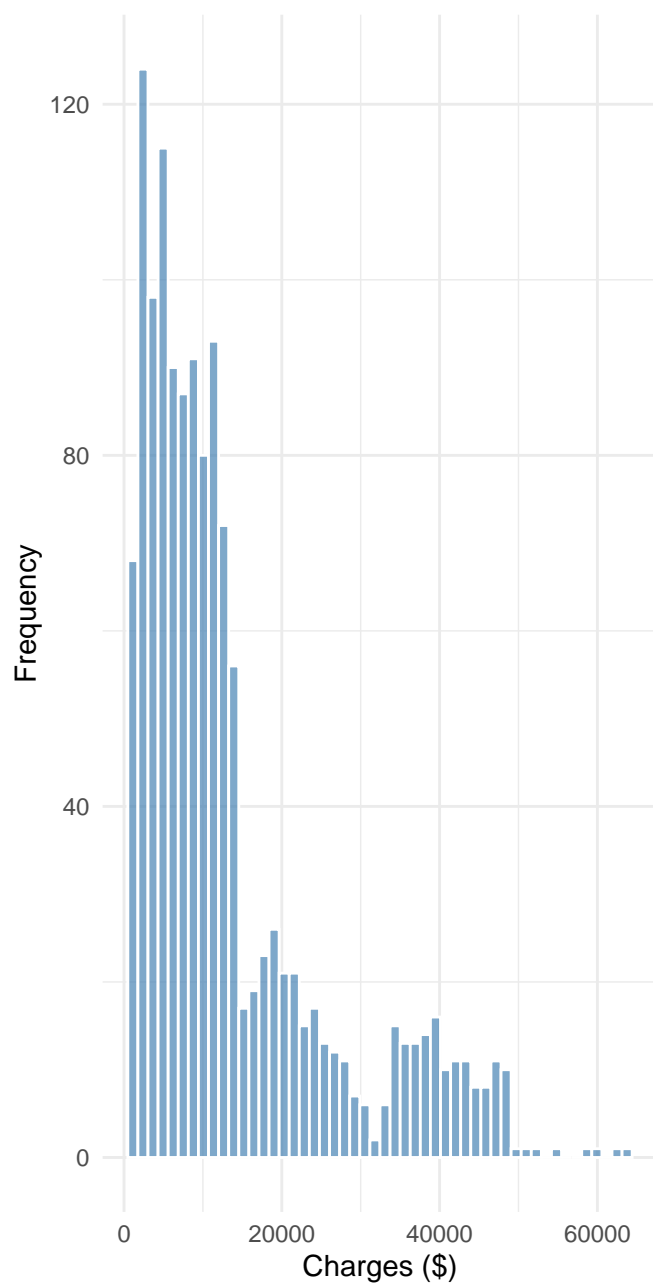
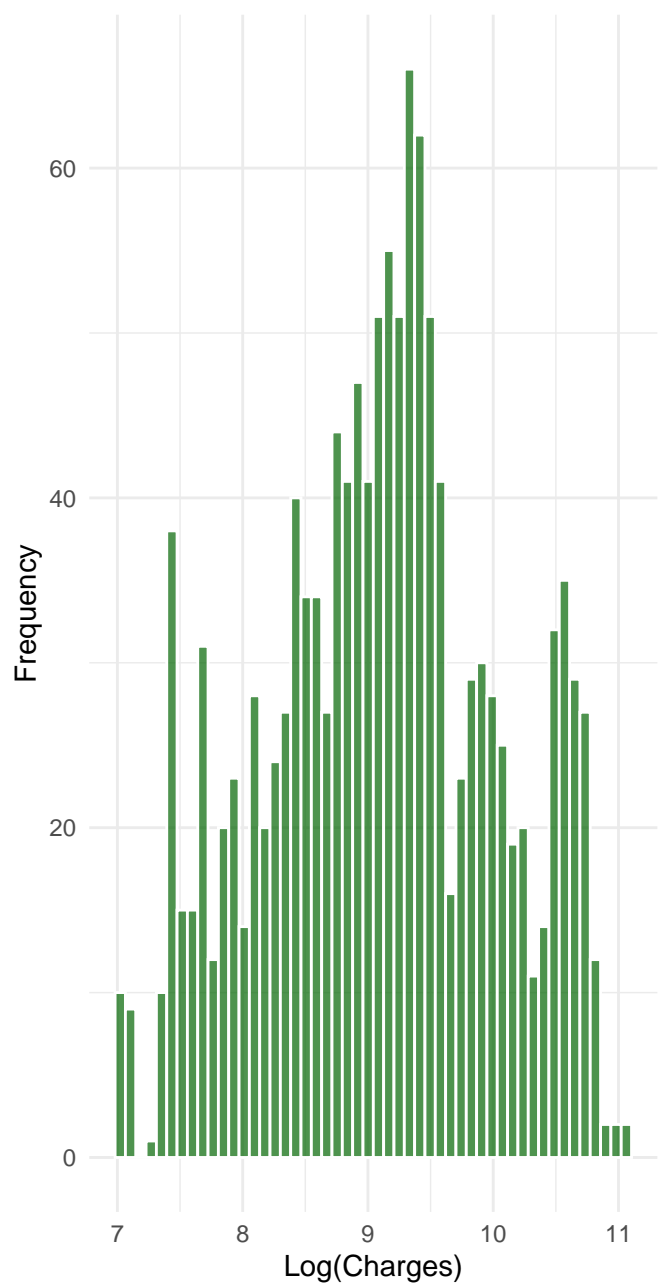


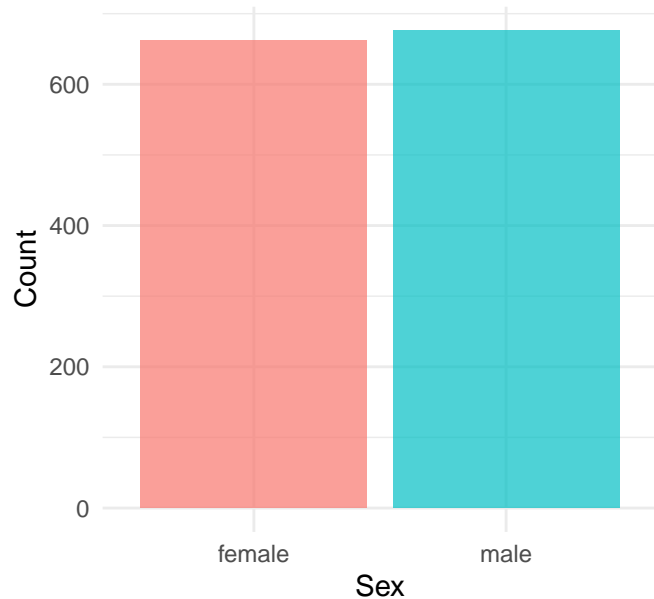
Distribution of Insurance Charges



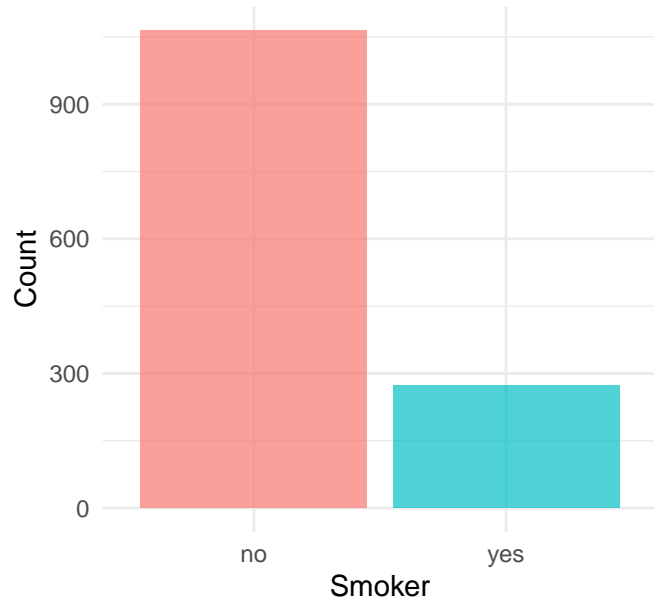
Distribution of Log-Transformed Charges



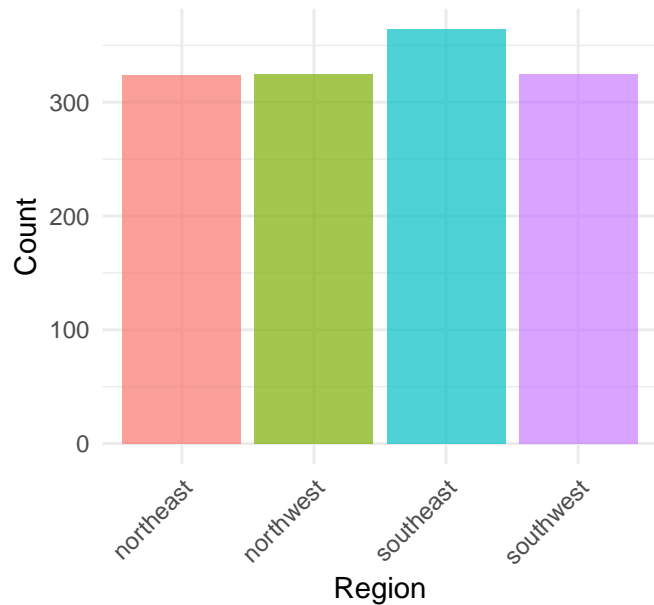
Gender Distribution



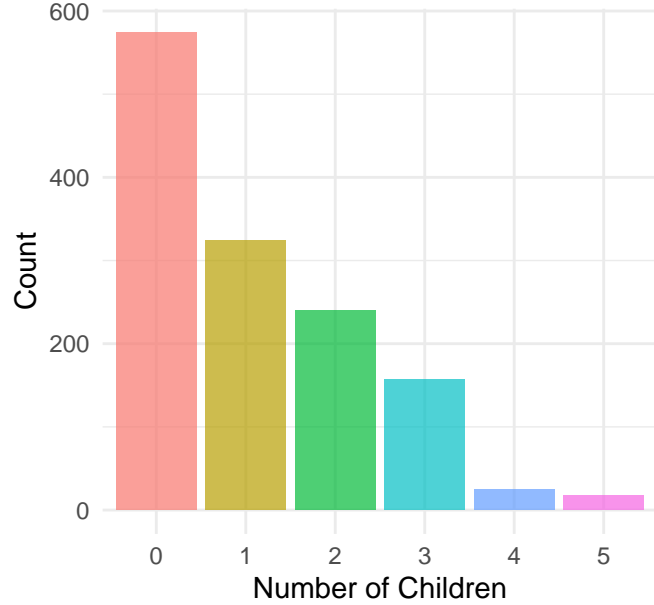
Smoker Distribution



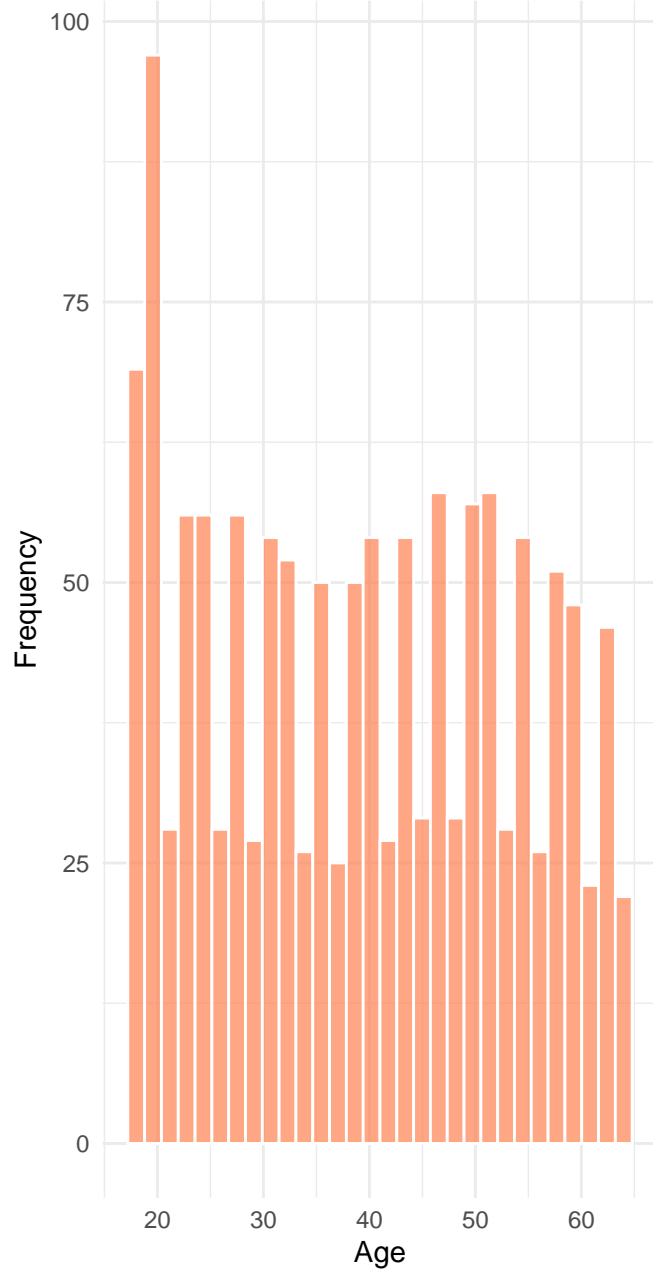
Region Distribution



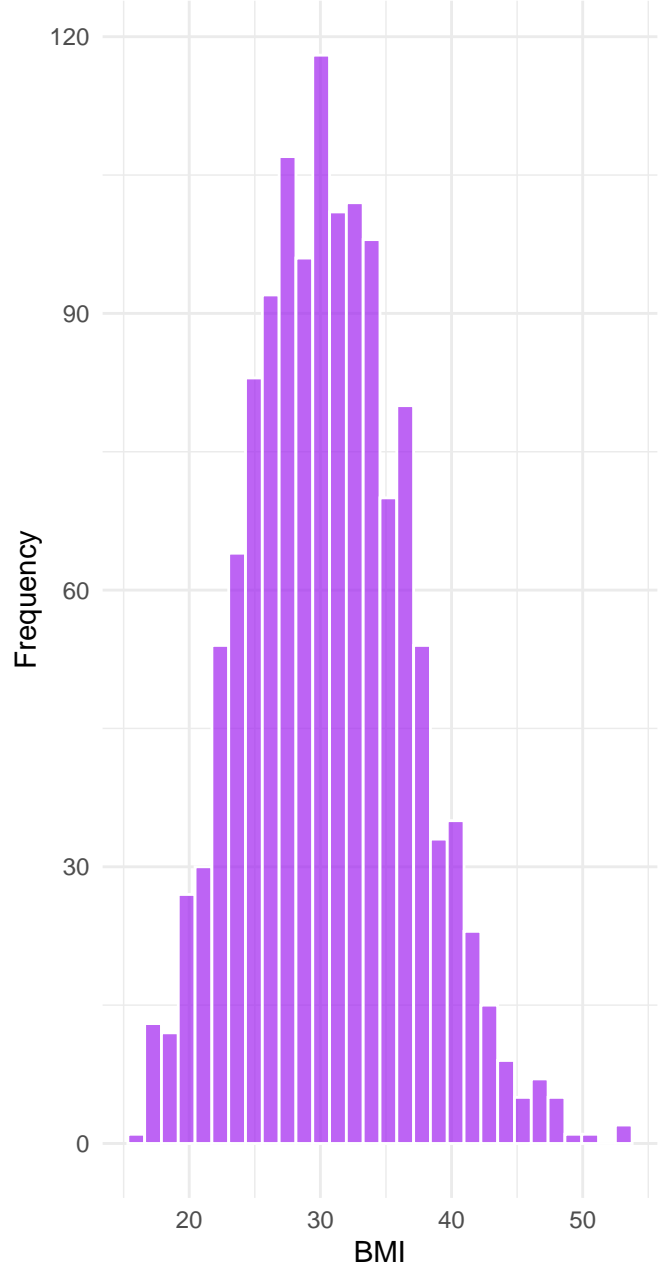
Children Distribution



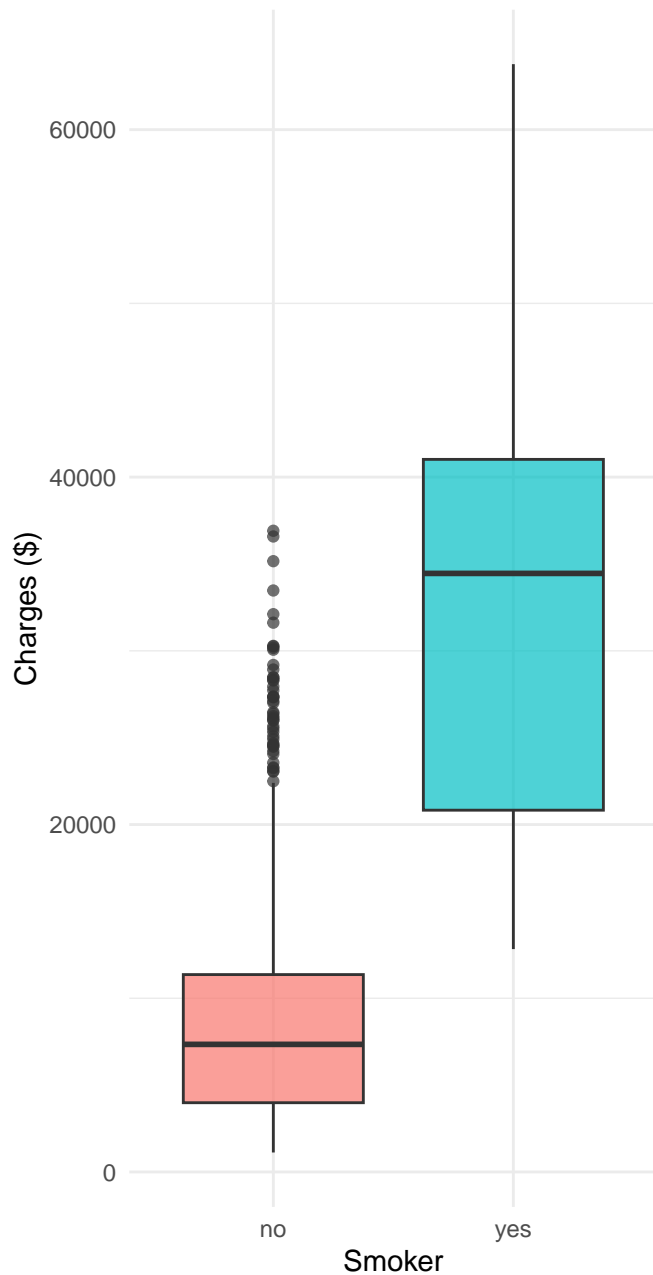
Age Distribution



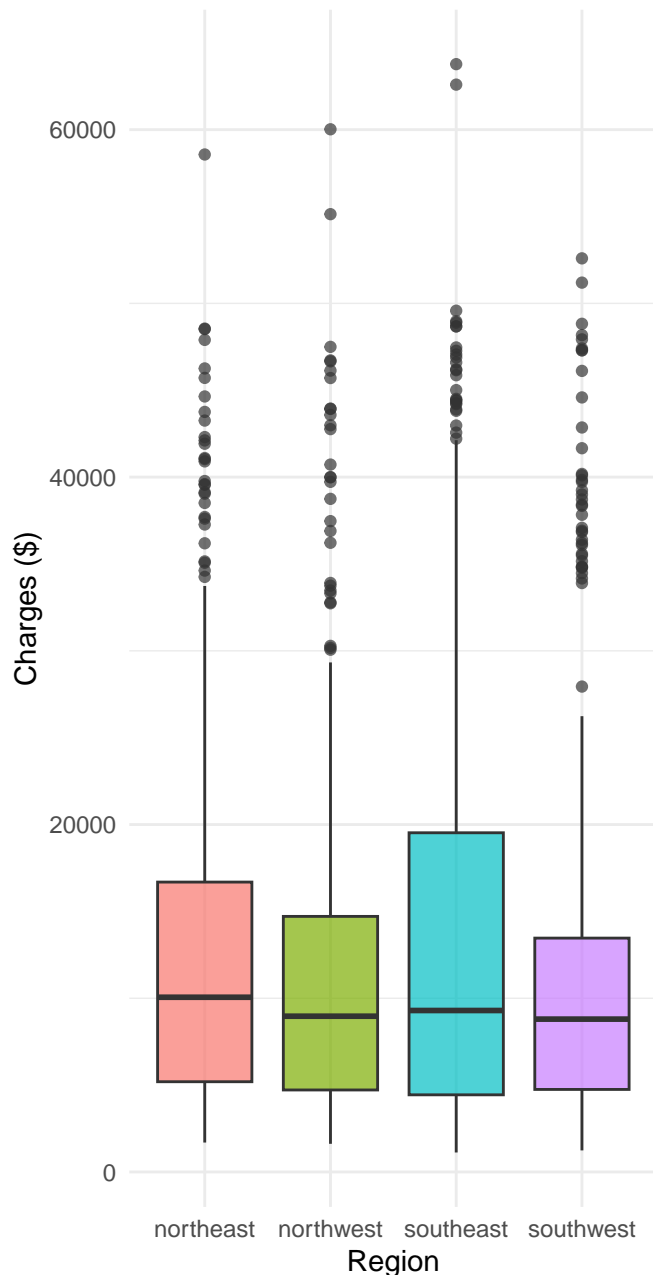
BMI Distribution



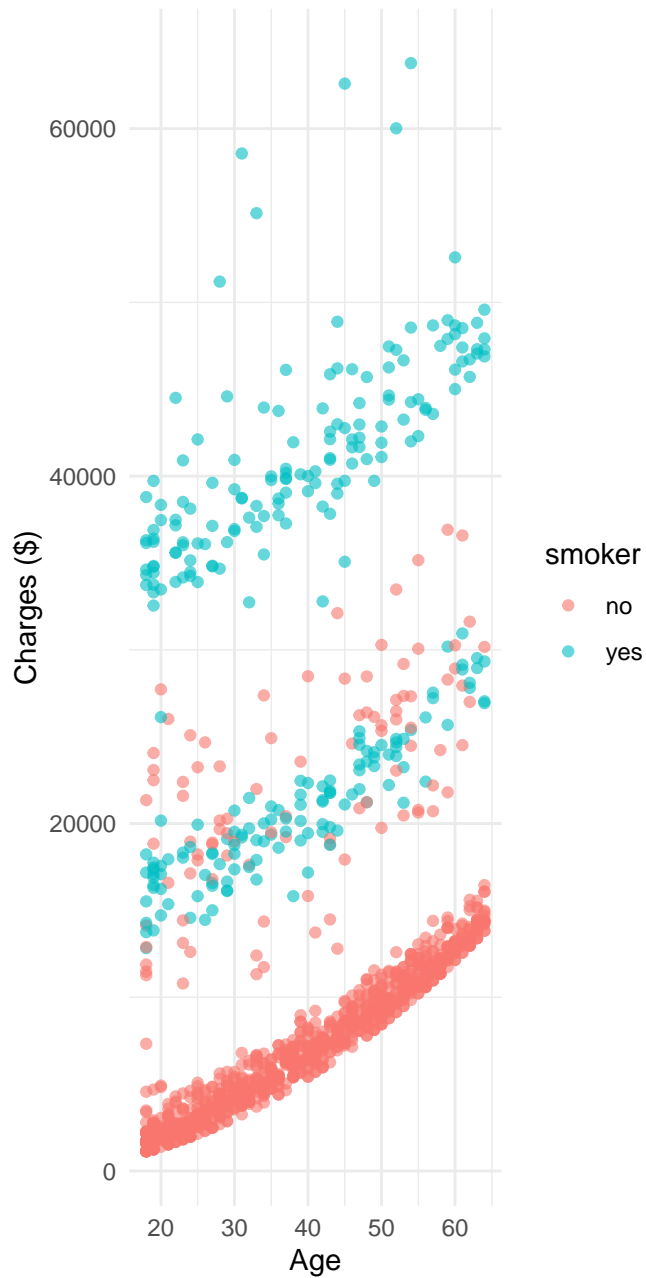
Charges by Smoker Status



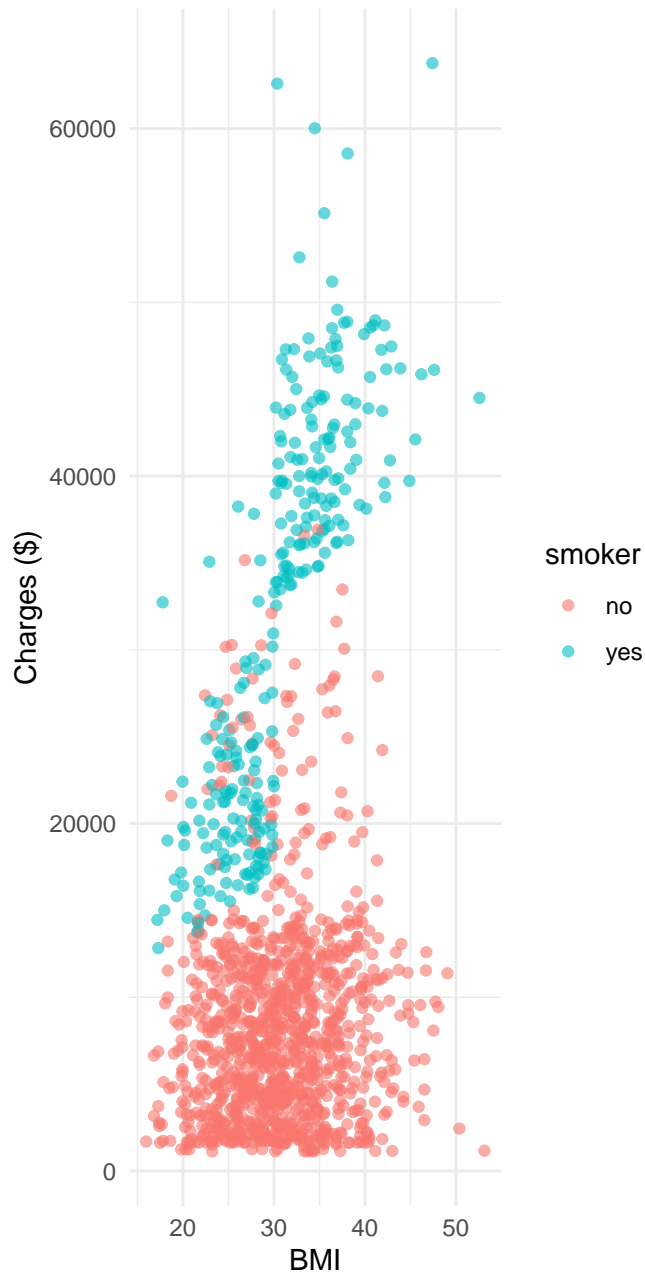
Charges by Region



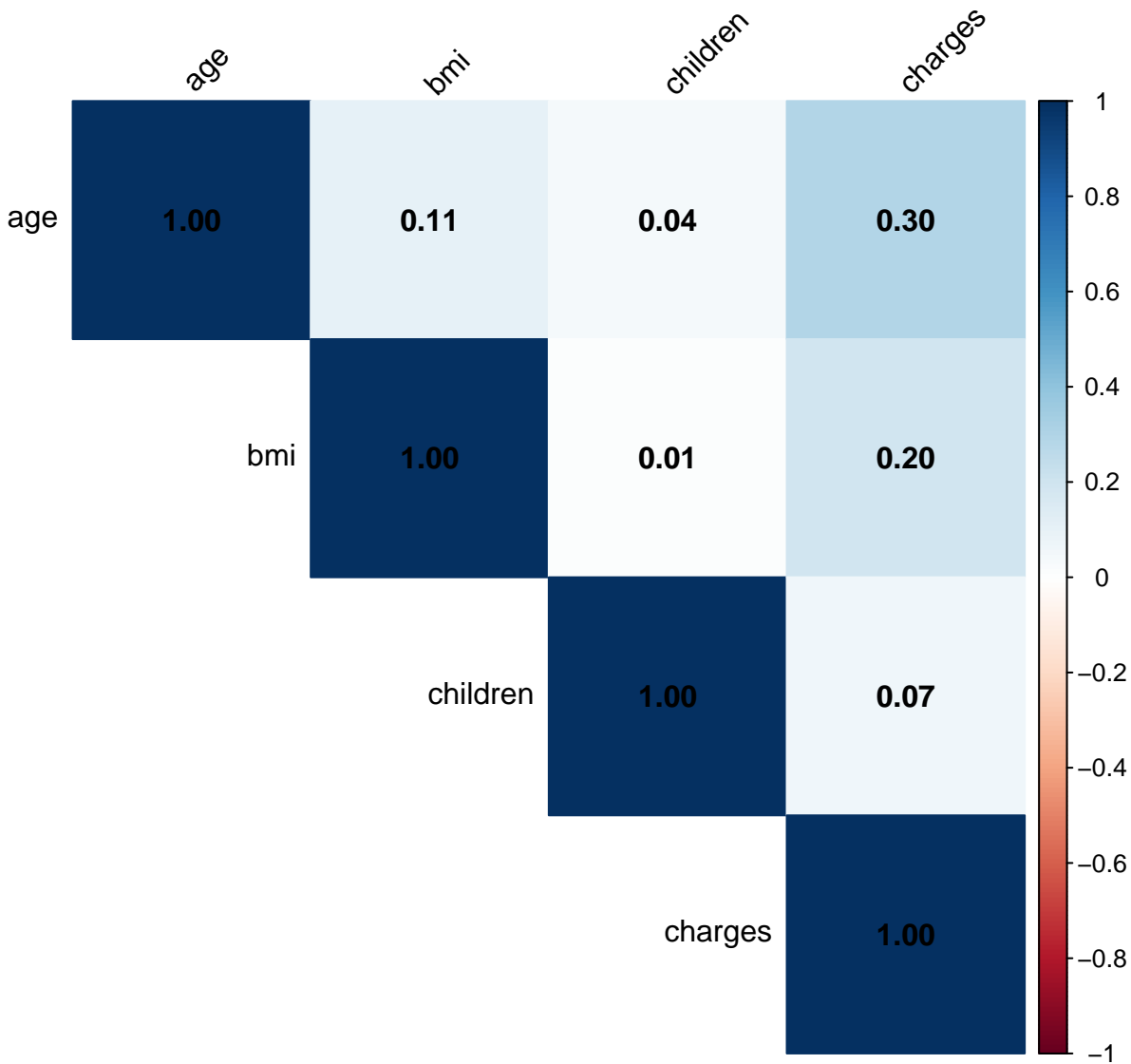
Age vs Charges (by Smoker)



BMI vs Charges (by Smoker)

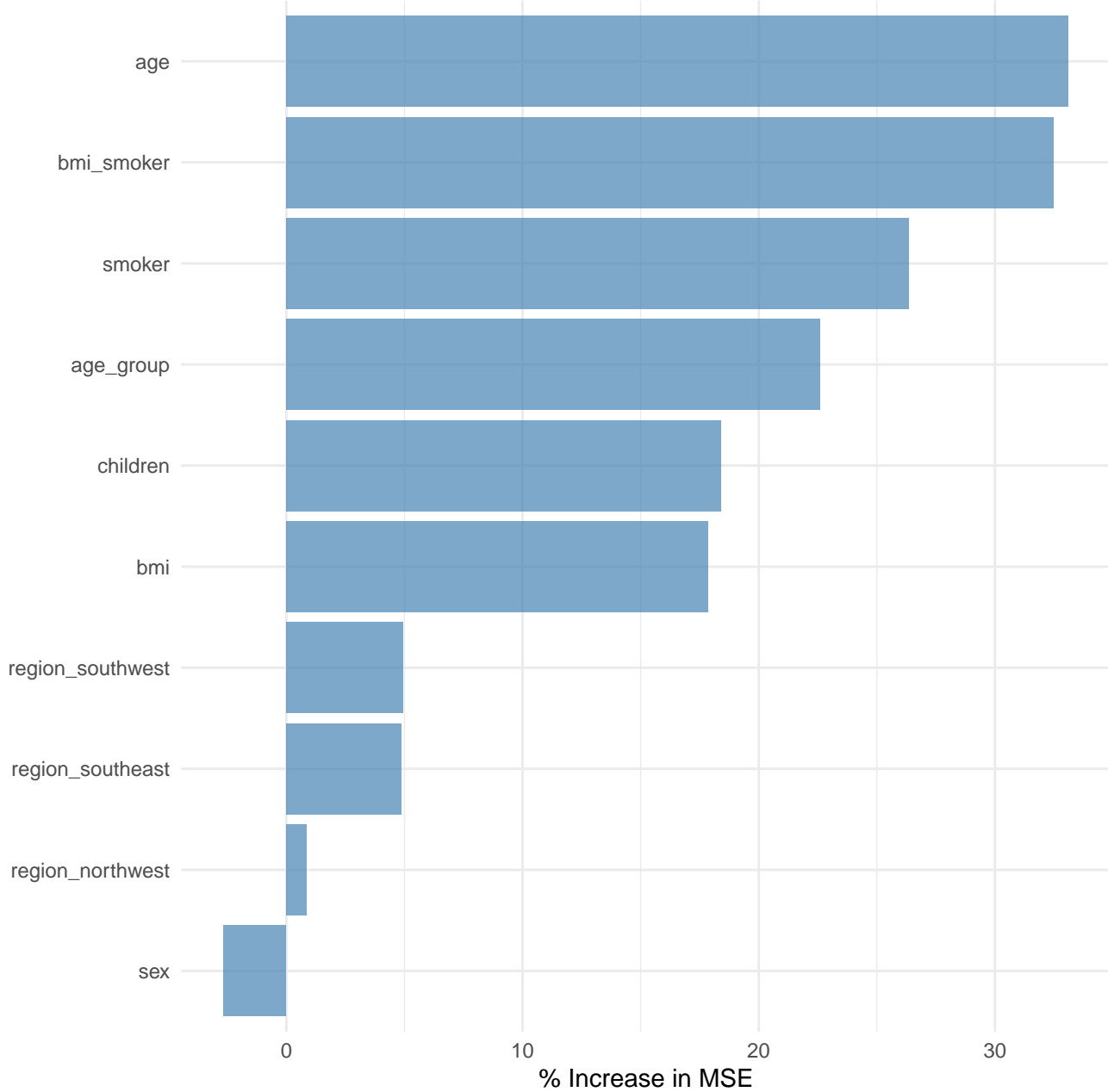


Correlation Matrix



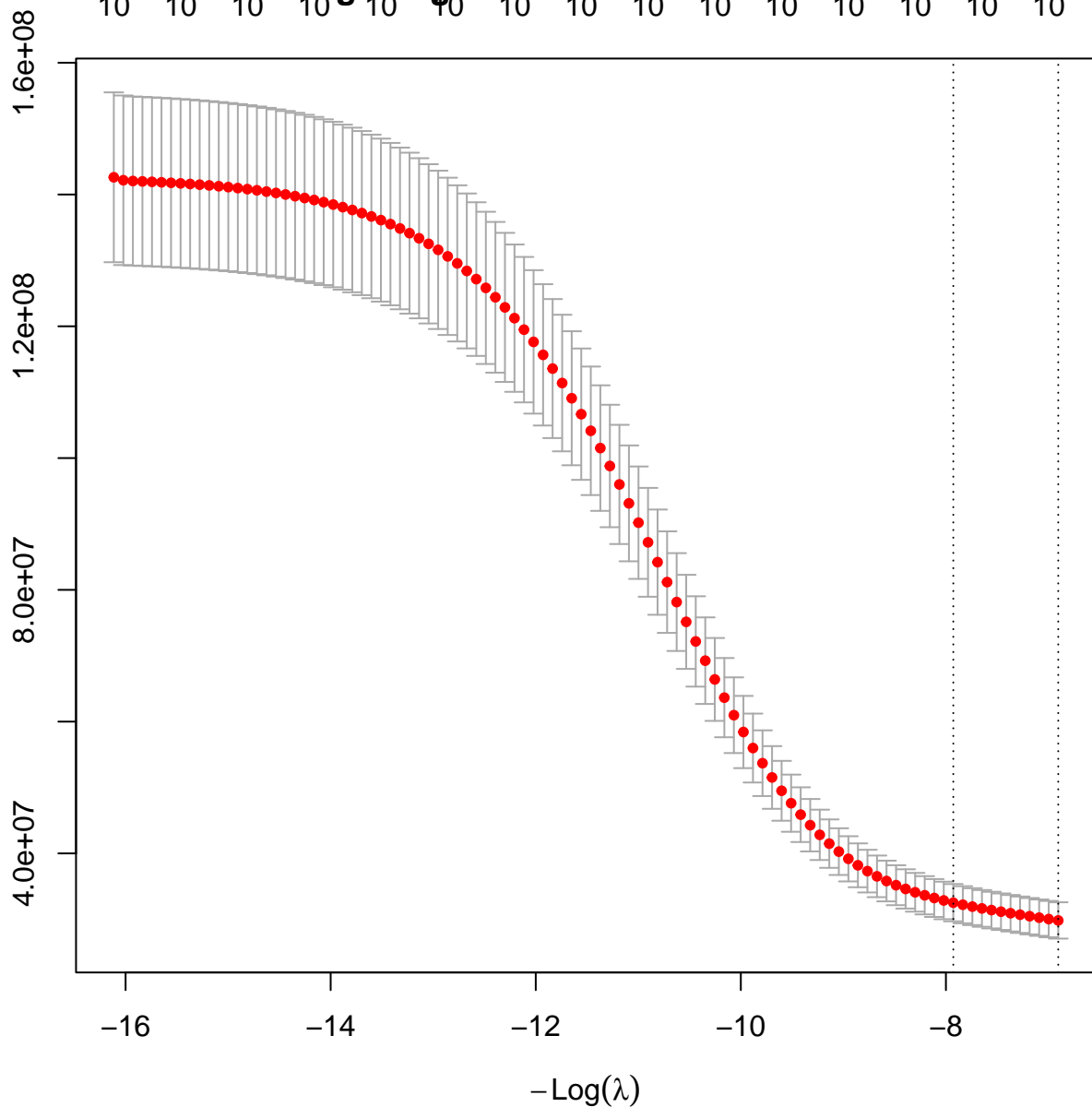
Random Forest – Feature Importance

Feature



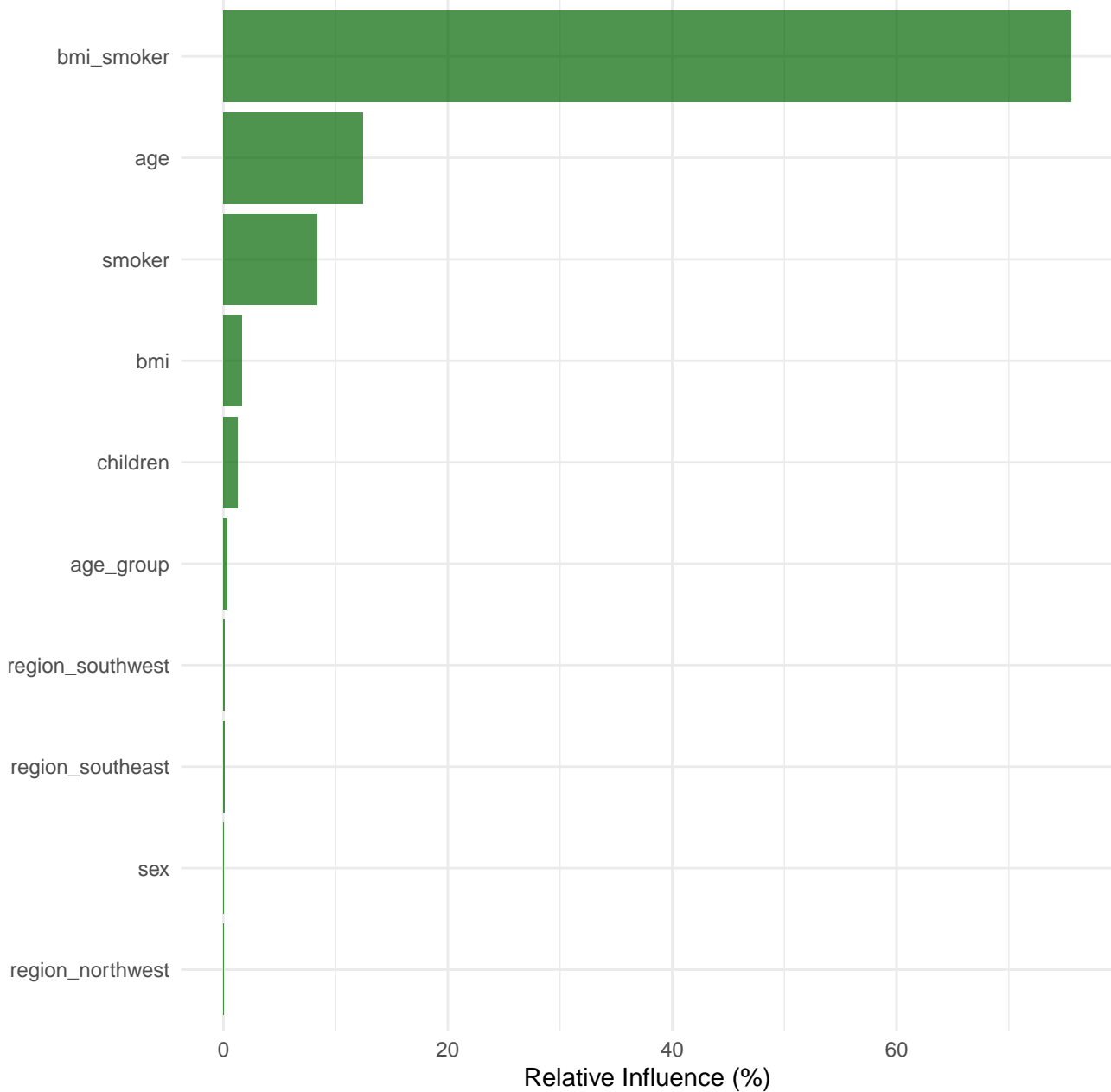
Ridge Regression – Cross-Validation

Mean-Squared Error

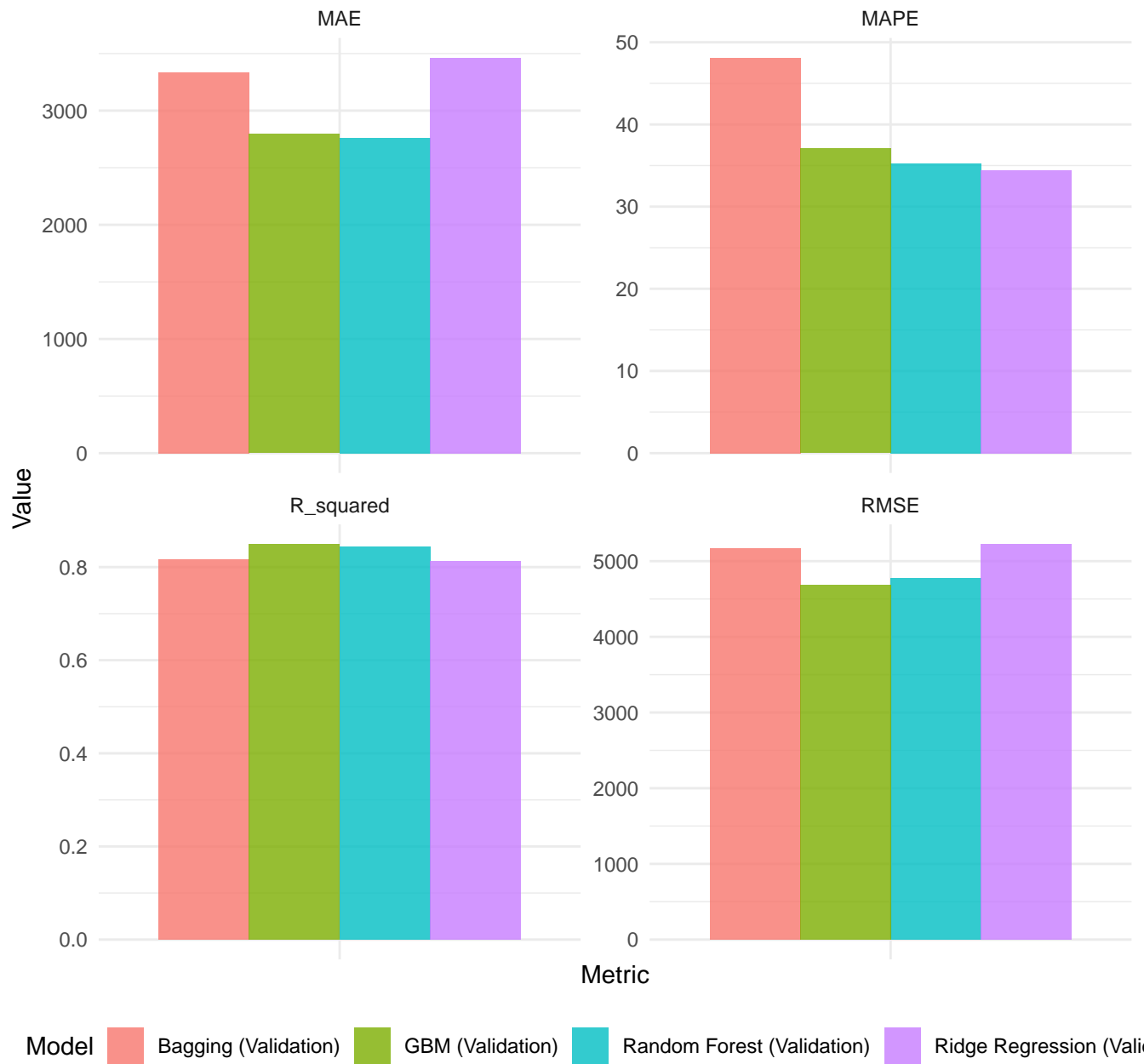


Gradient Boosting – Feature Importance

Feature

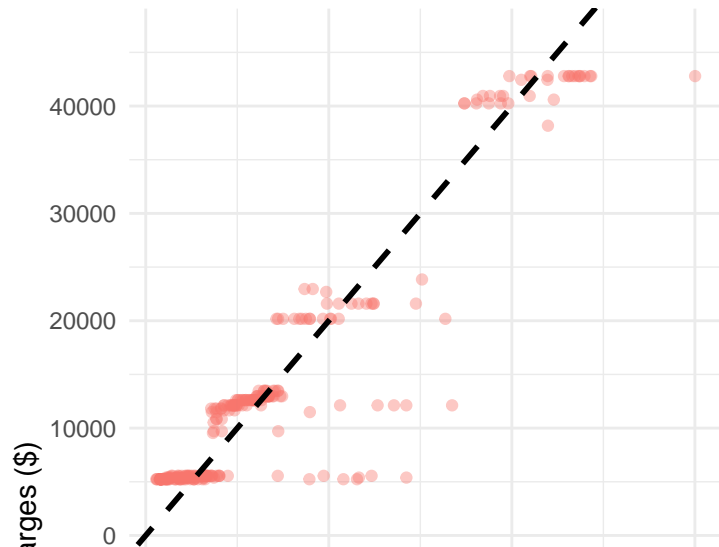


Model Comparison on Validation Set (4 Models)

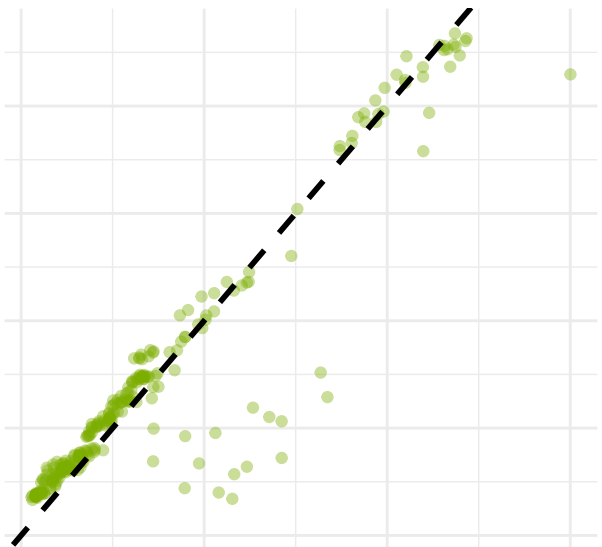


Actual vs Predicted Charges – All Models

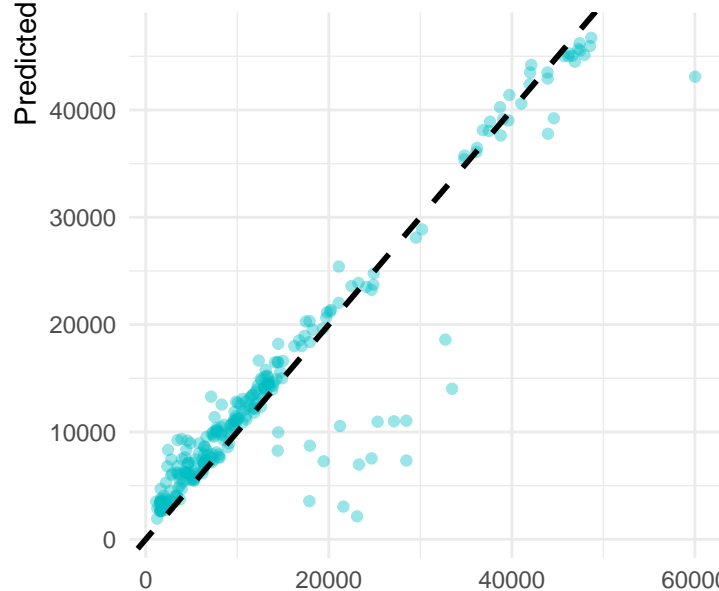
Bagging



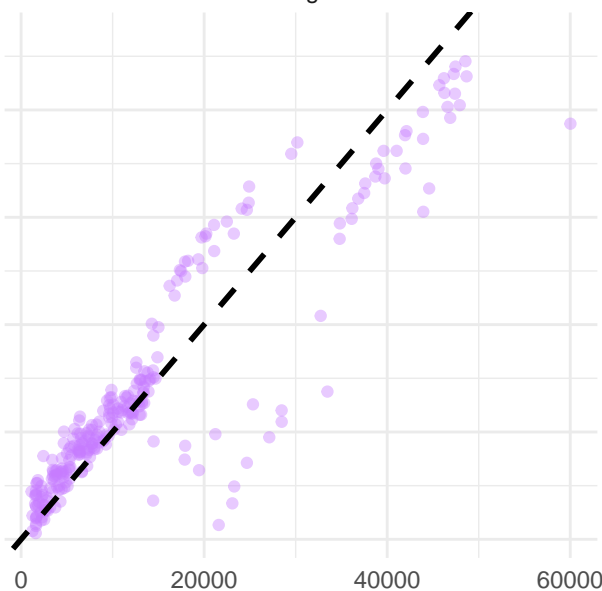
GBM



RandomForest



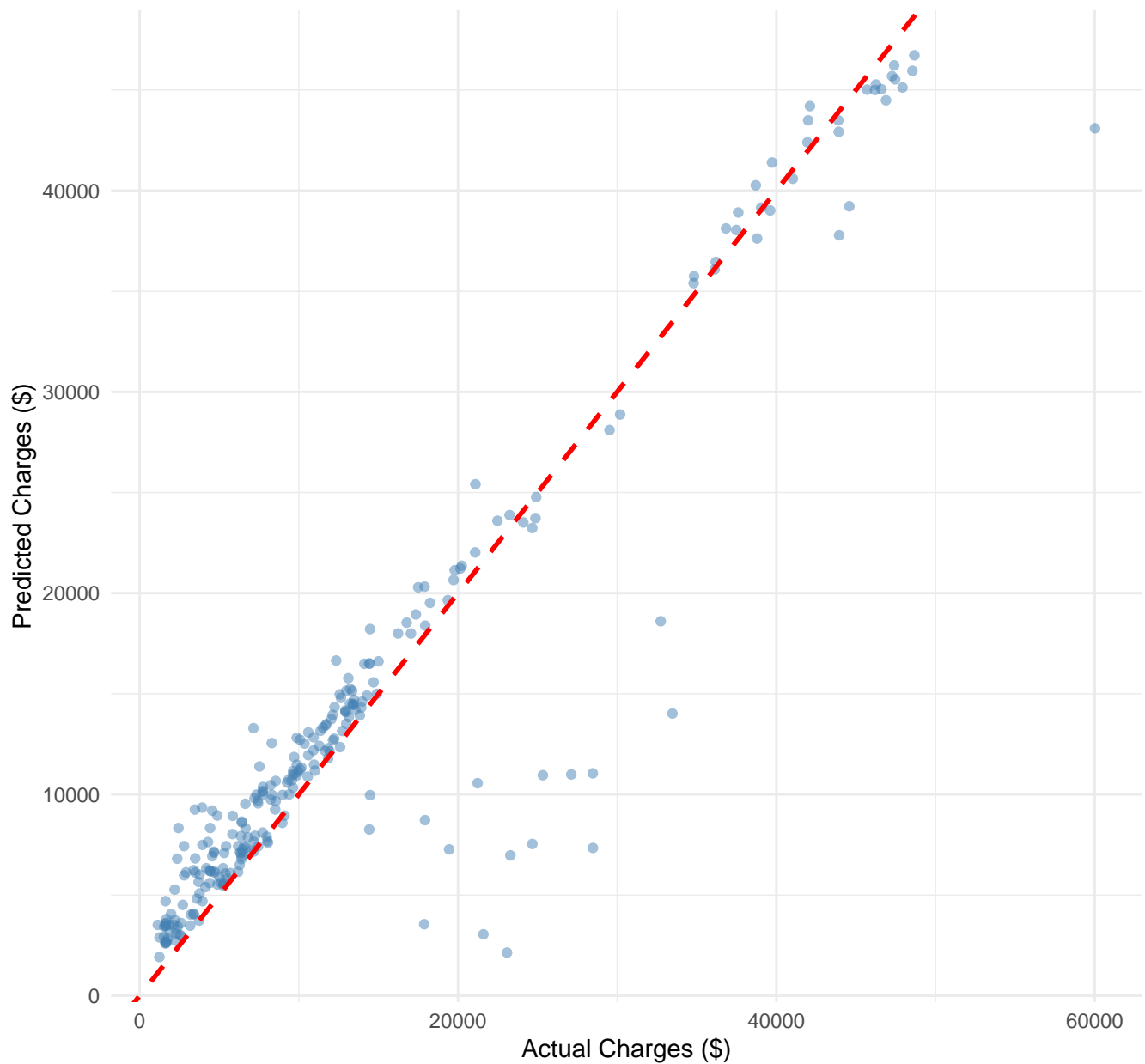
Ridge



Actual Charges (\$)

Actual vs Predicted Charges – Random Forest (Test)

R-squared: 0.884 | RMSE: \$ 4312.33



Residual Plot – Random Forest (Test)



Final Model Comparison (Test Set)

