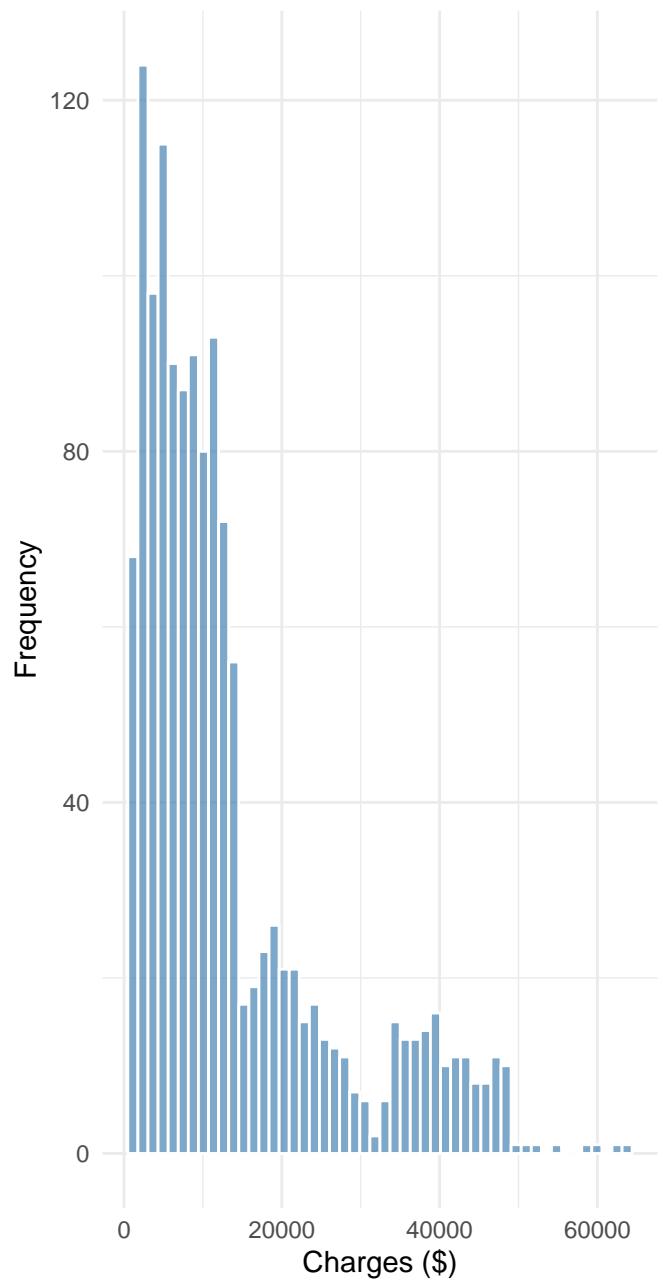
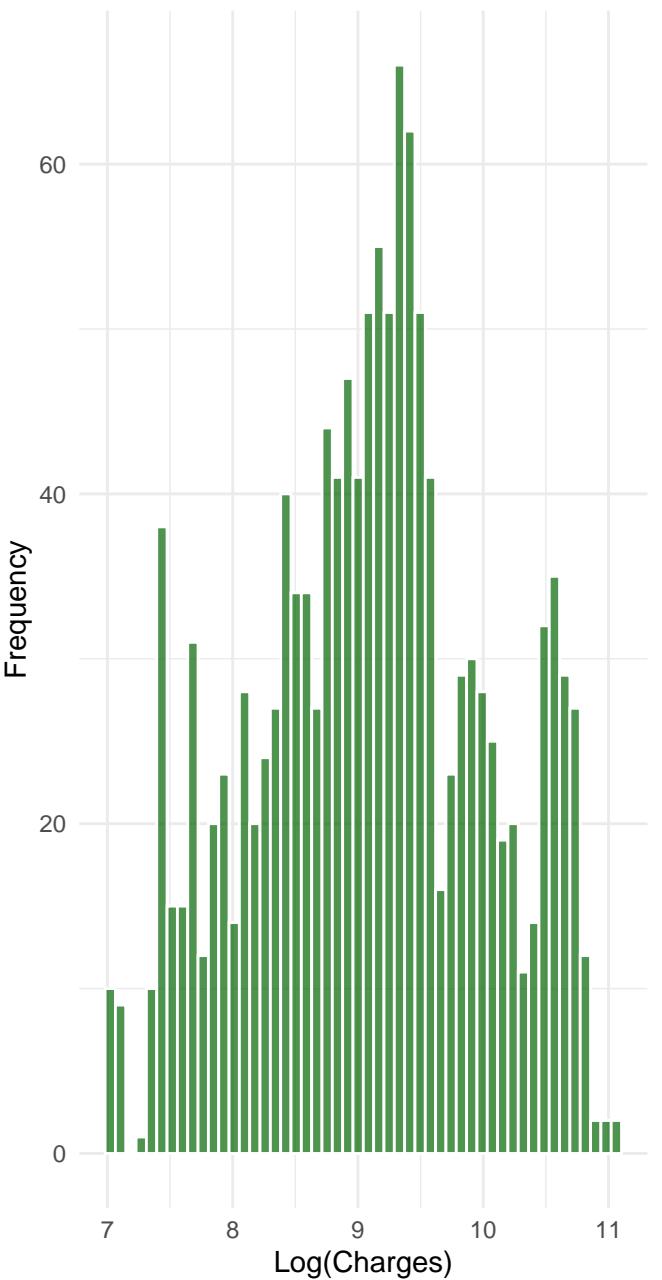


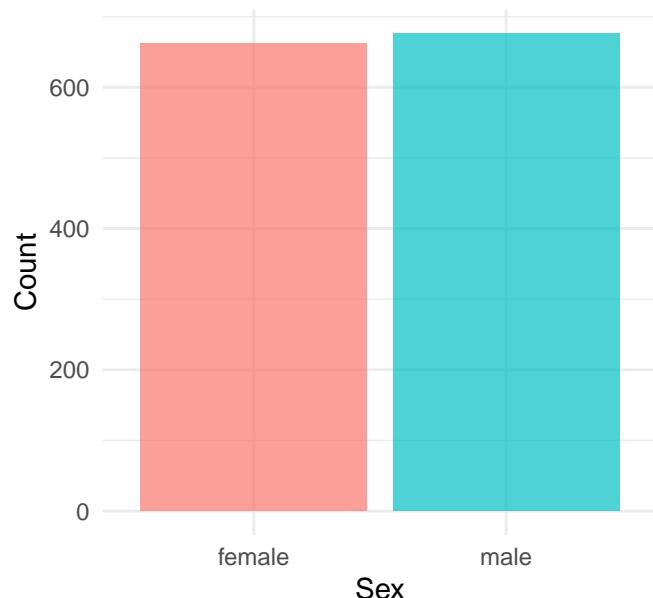
### Distribution of Insurance Charges



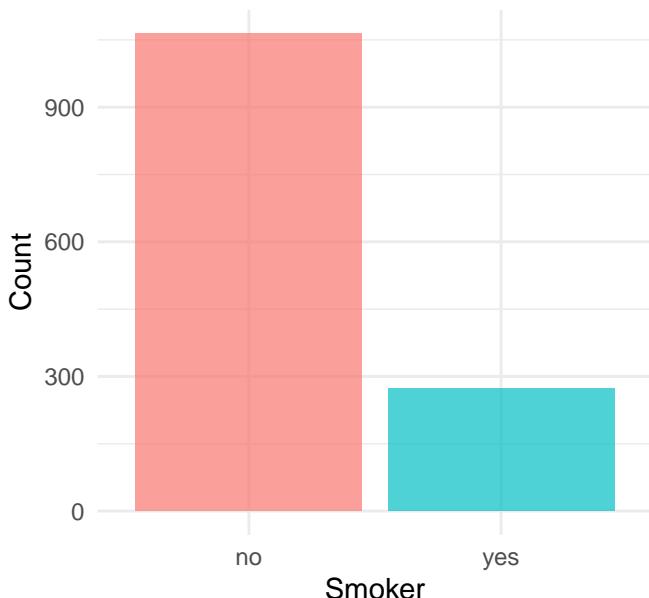
### Distribution of Log-Transformed Charges



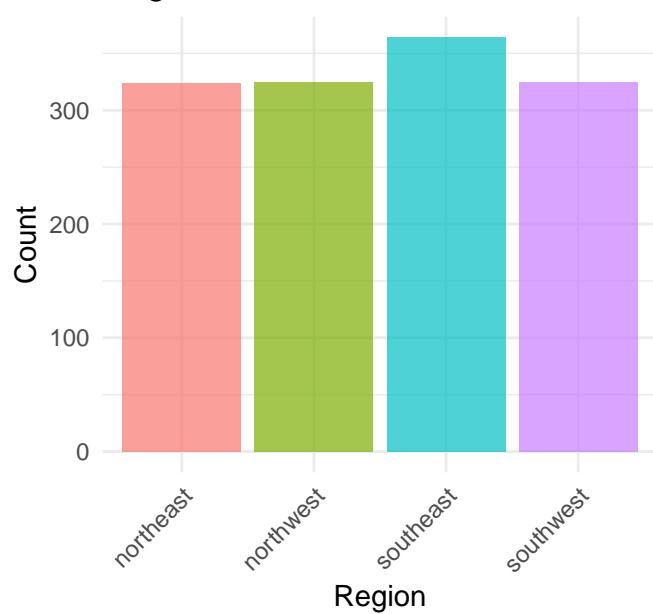
### Gender Distribution



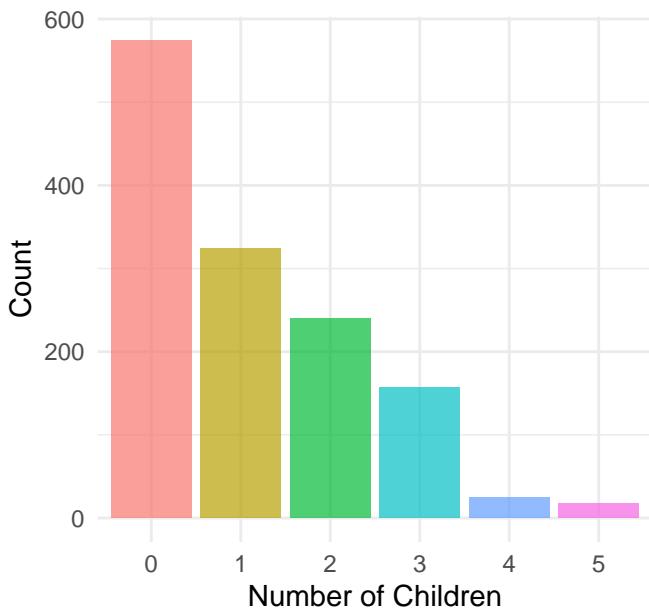
### Smoker Distribution



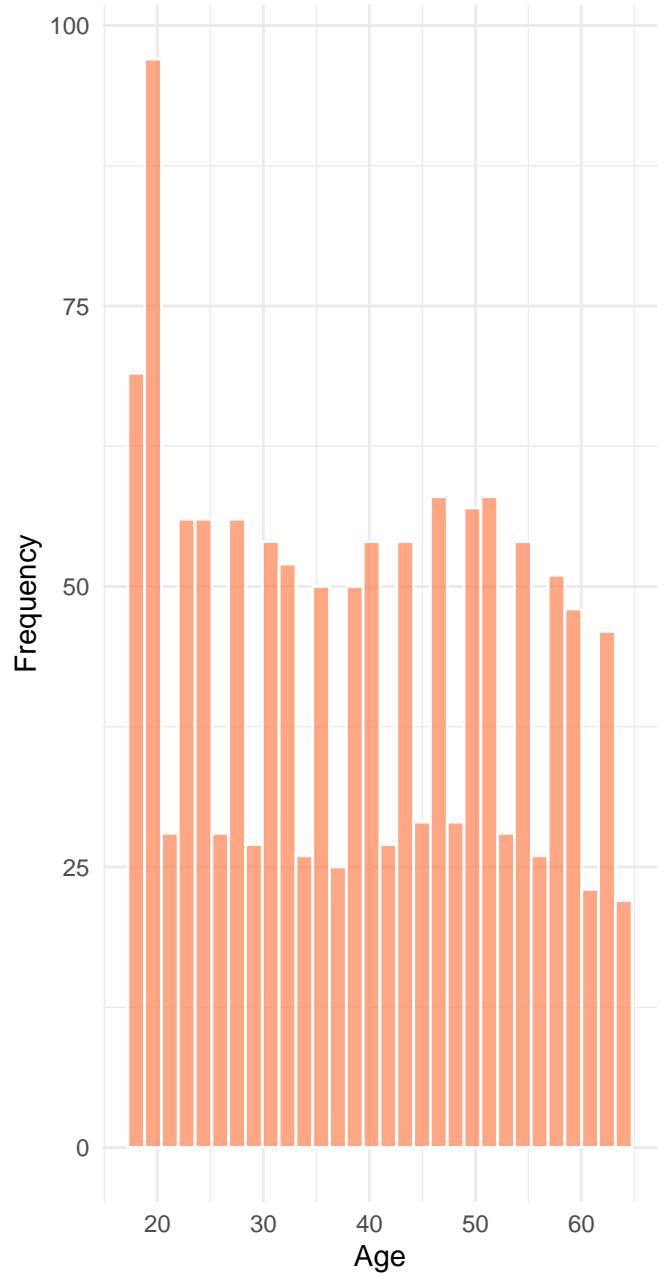
### Region Distribution



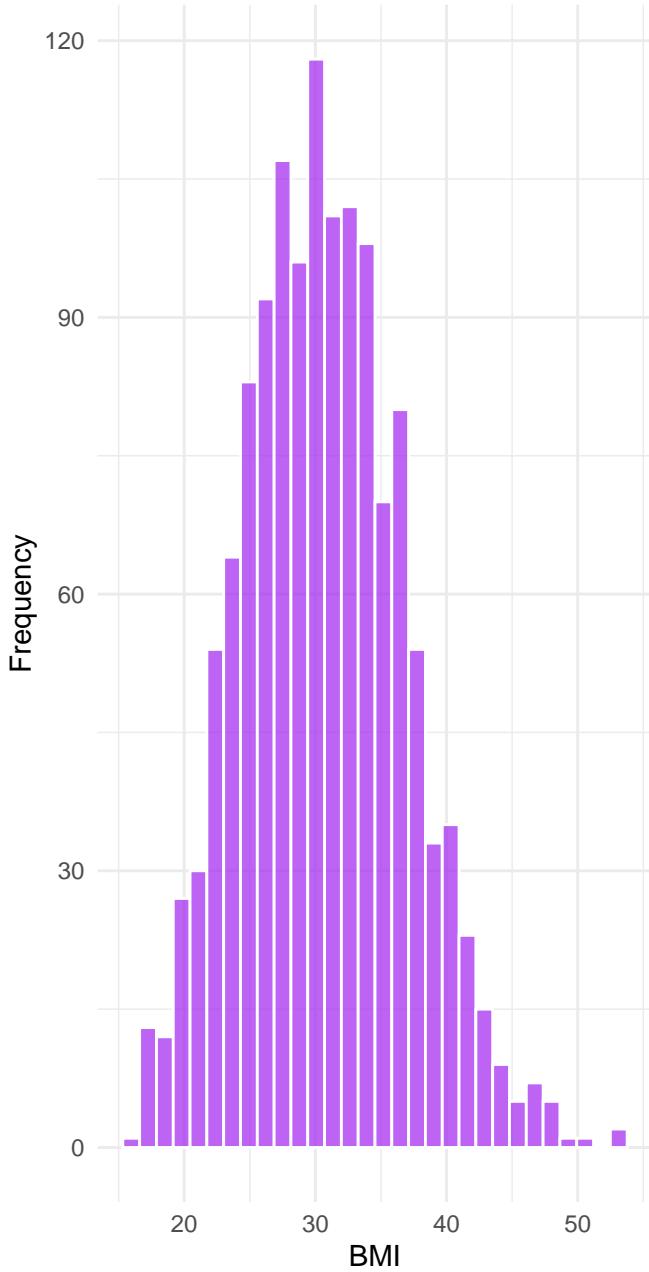
### Children Distribution



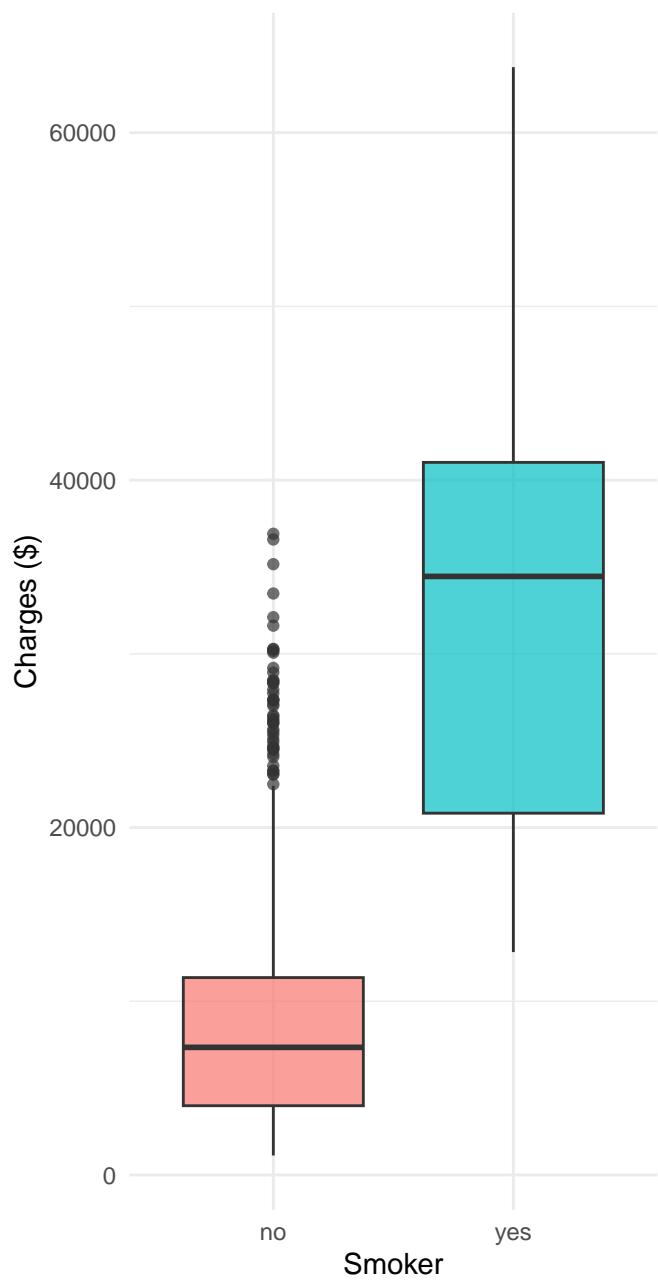
### Age Distribution



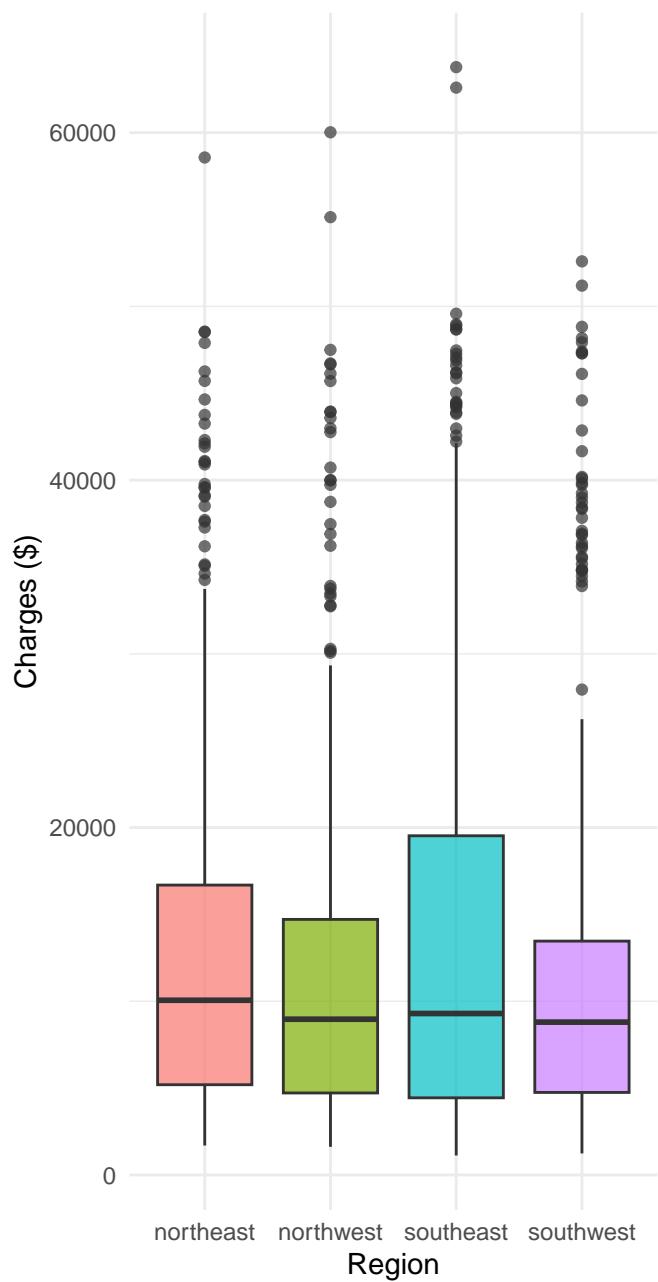
### BMI Distribution



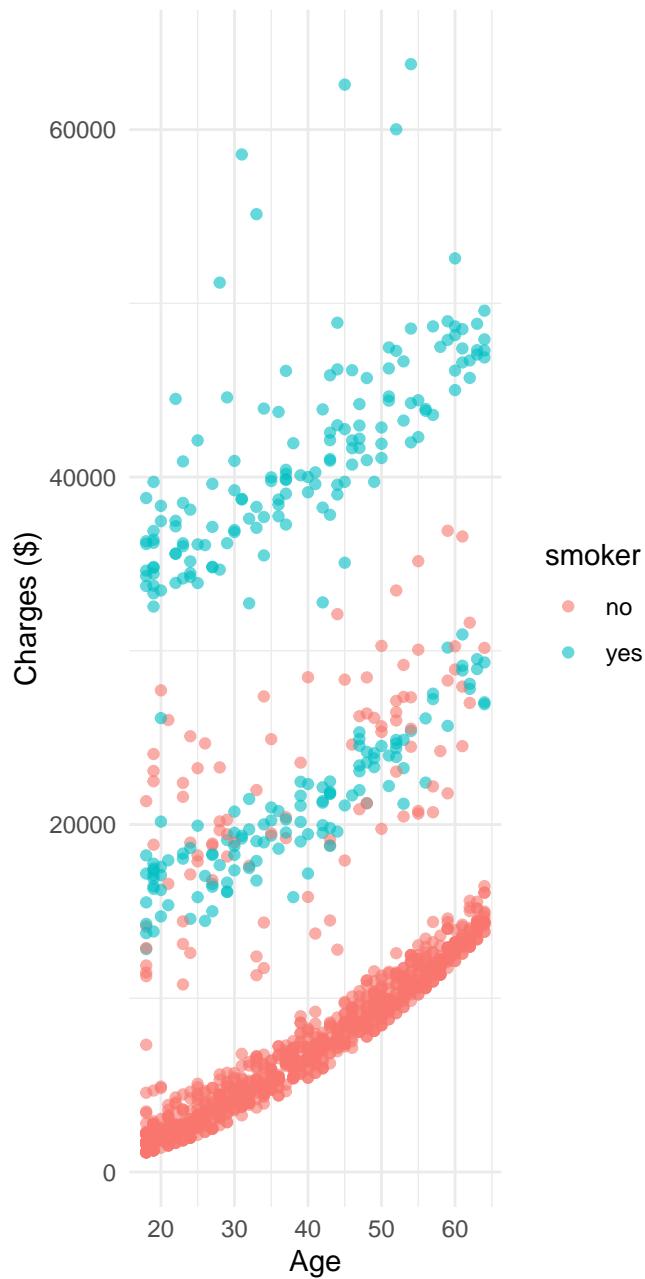
### Charges by Smoker Status



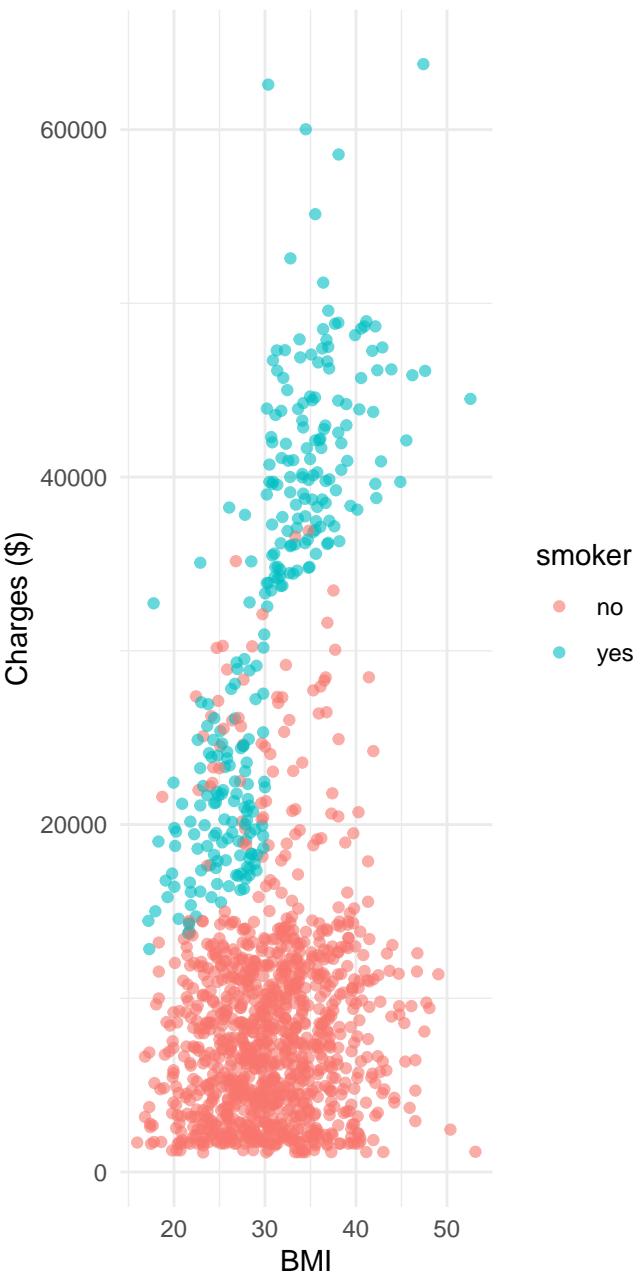
### Charges by Region



### Age vs Charges (by Smoker)



### BMI vs Charges (by Smoker)



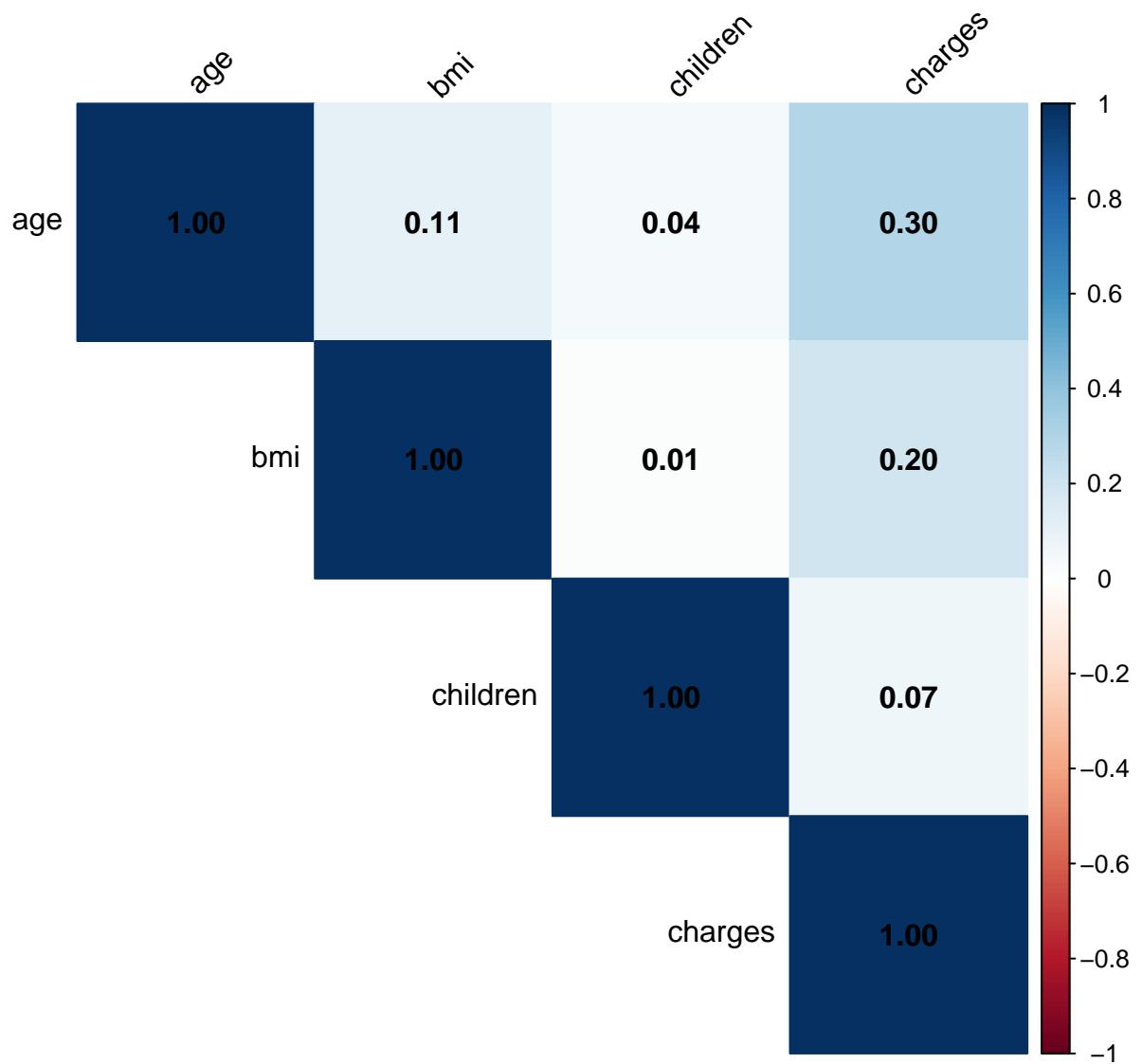
smoker

- no
- yes

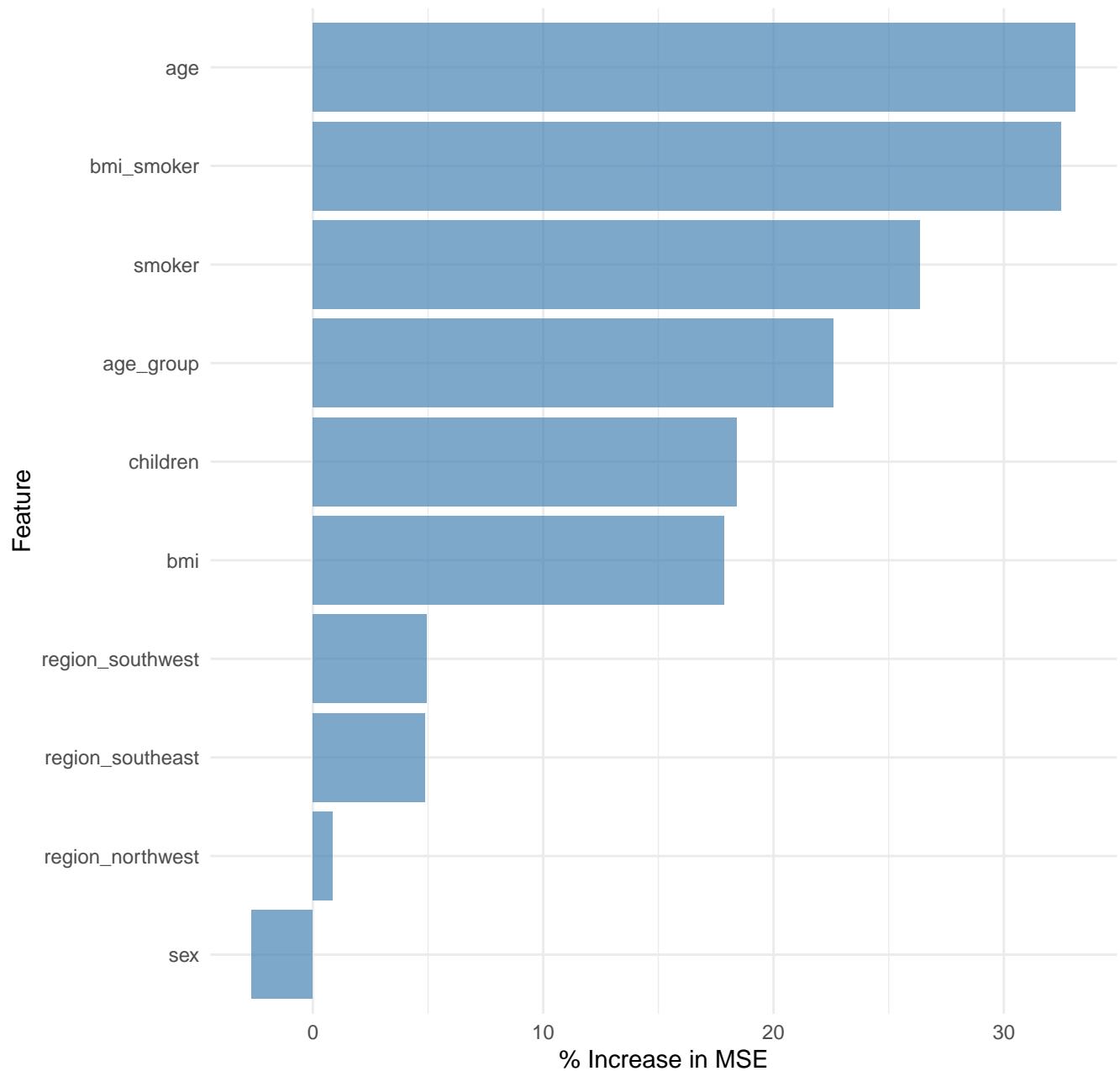
smoker

- no
- yes

# Correlation Matrix

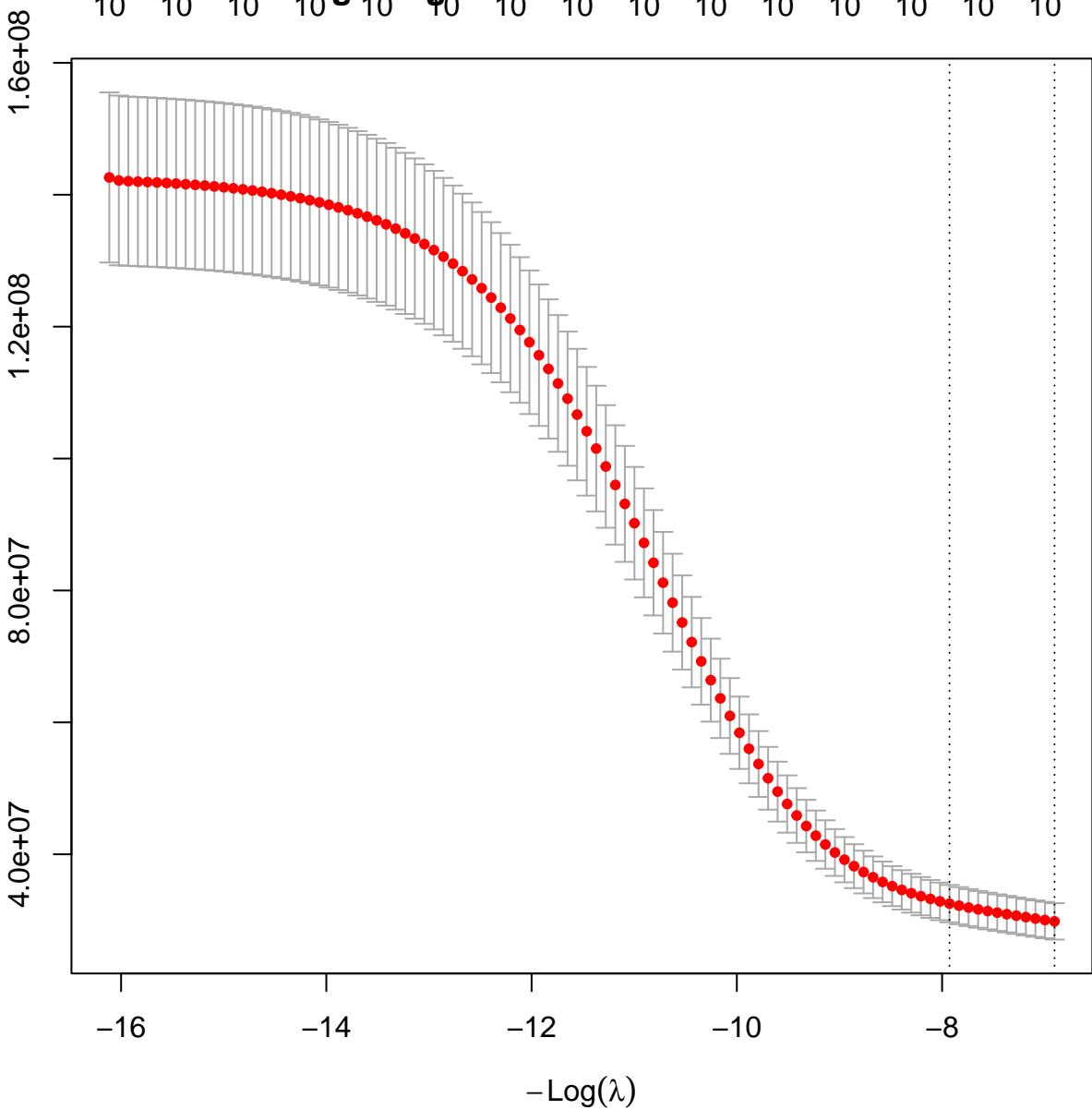


## Random Forest – Feature Importance

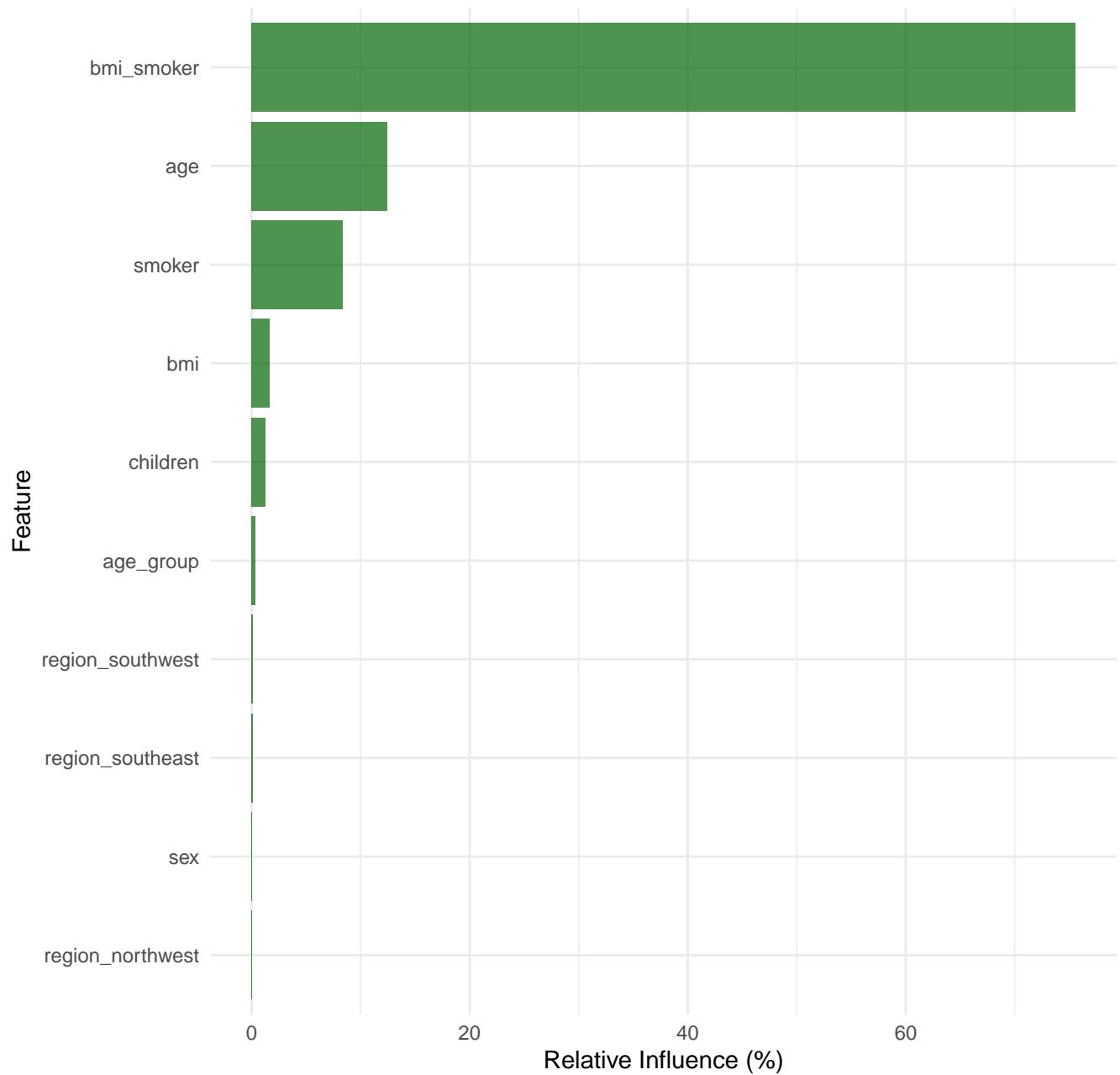


# Ridge Regression - Cross-Validation

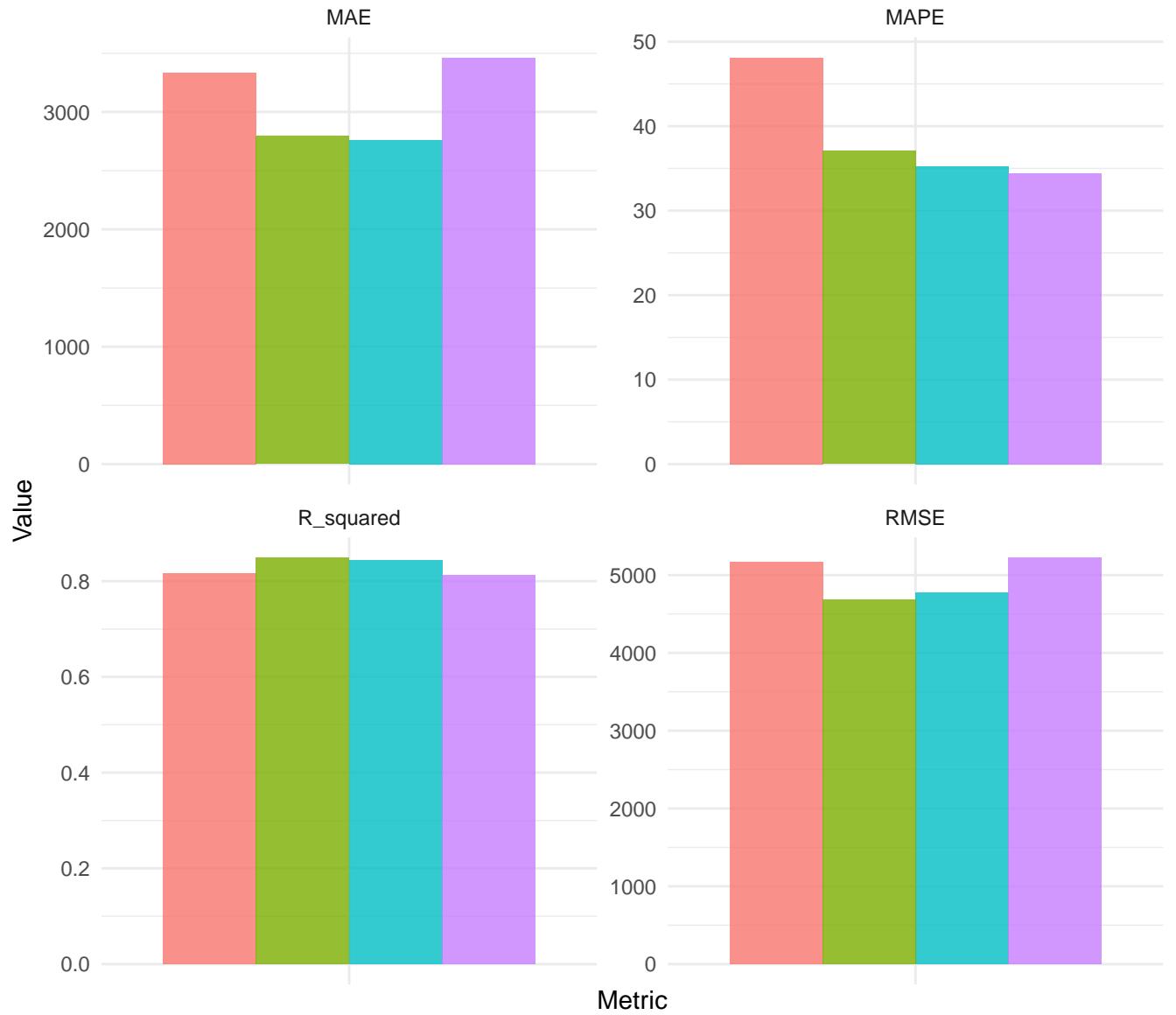
Mean-Squared Error



## Gradient Boosting – Feature Importance

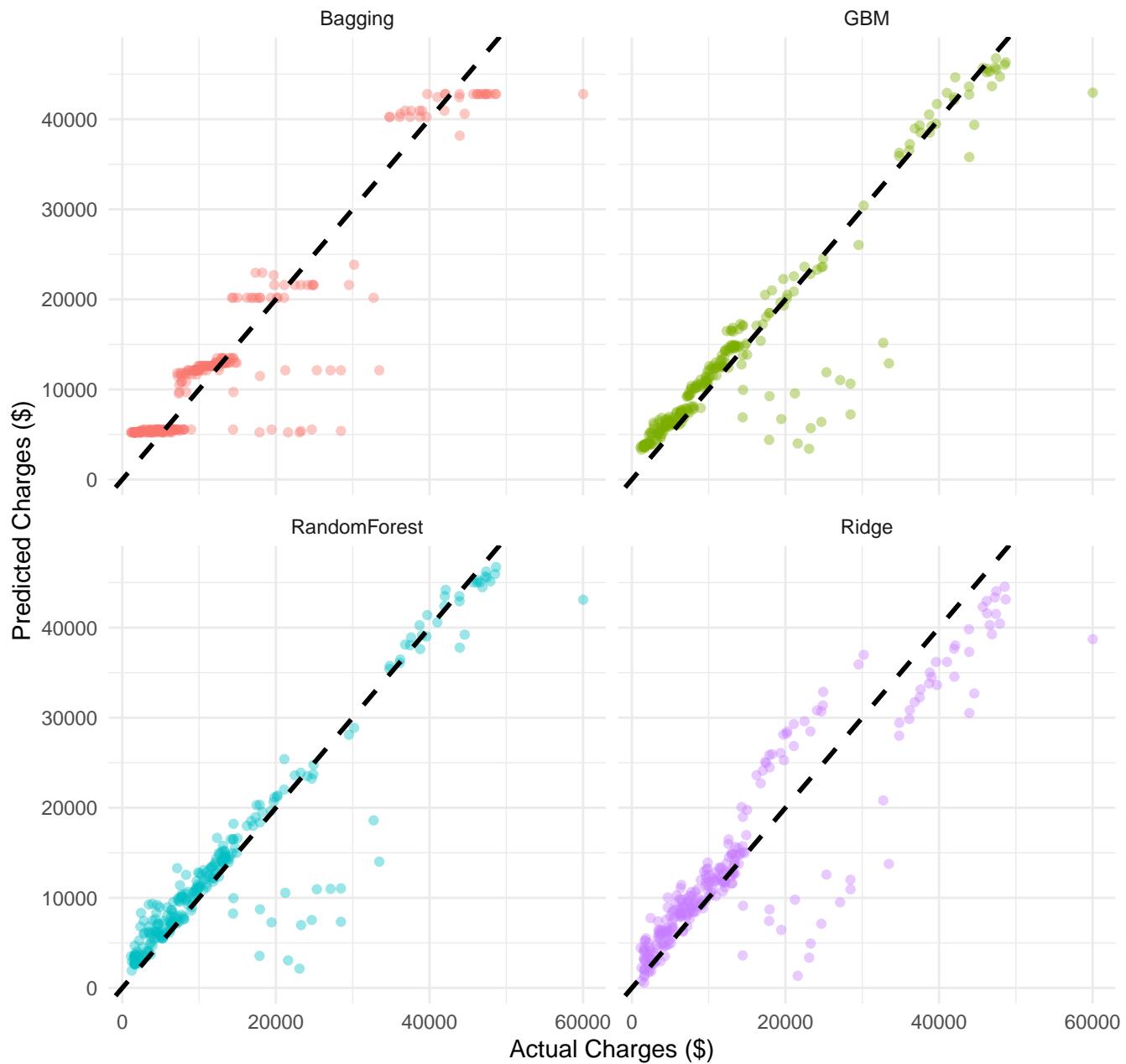


# Model Comparison on Validation Set (4 Models)



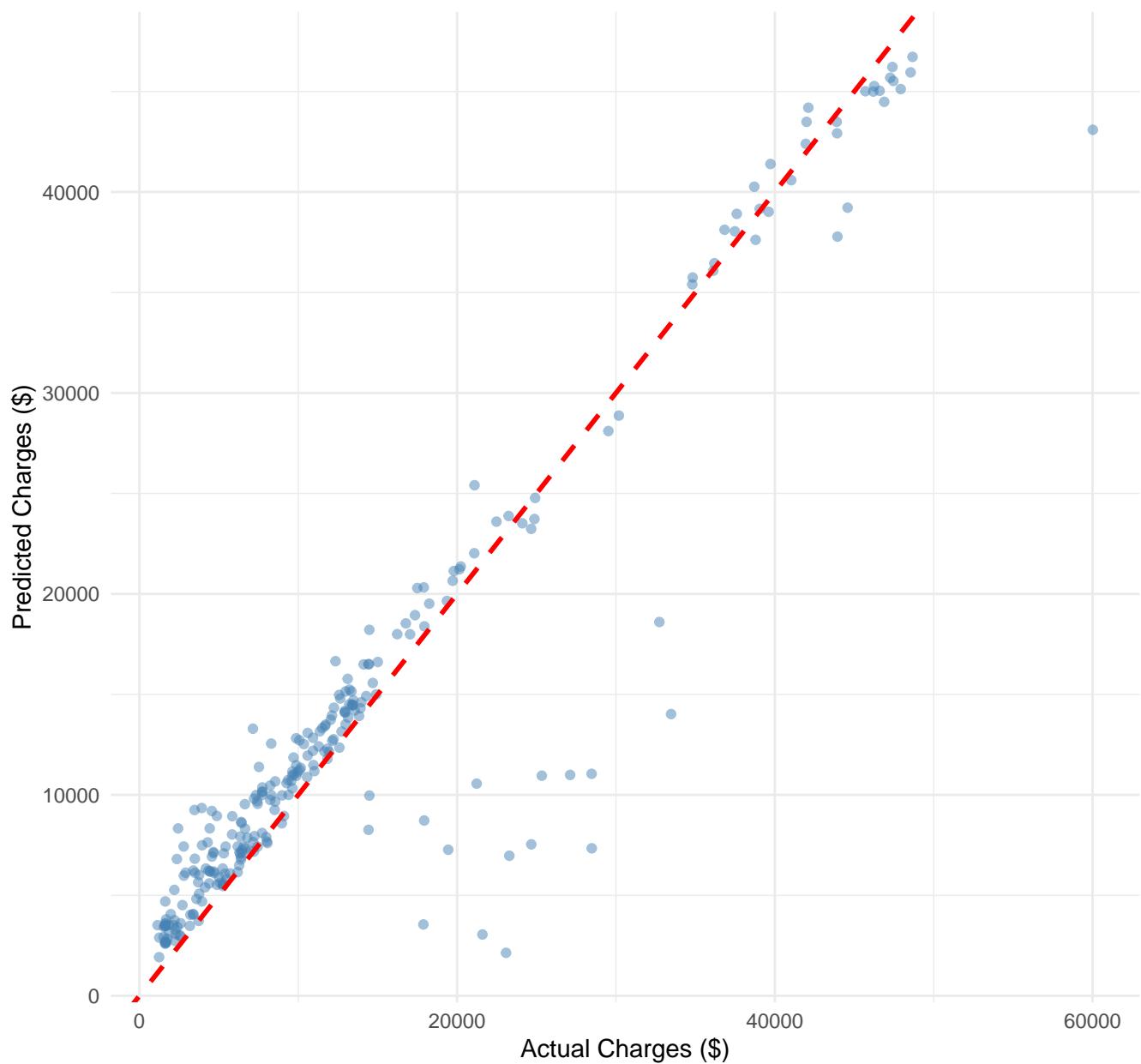
Model Bagging (Validation) GBM (Validation) Random Forest (Validation) Ridge Regression (Validation)

# Actual vs Predicted Charges – All Models

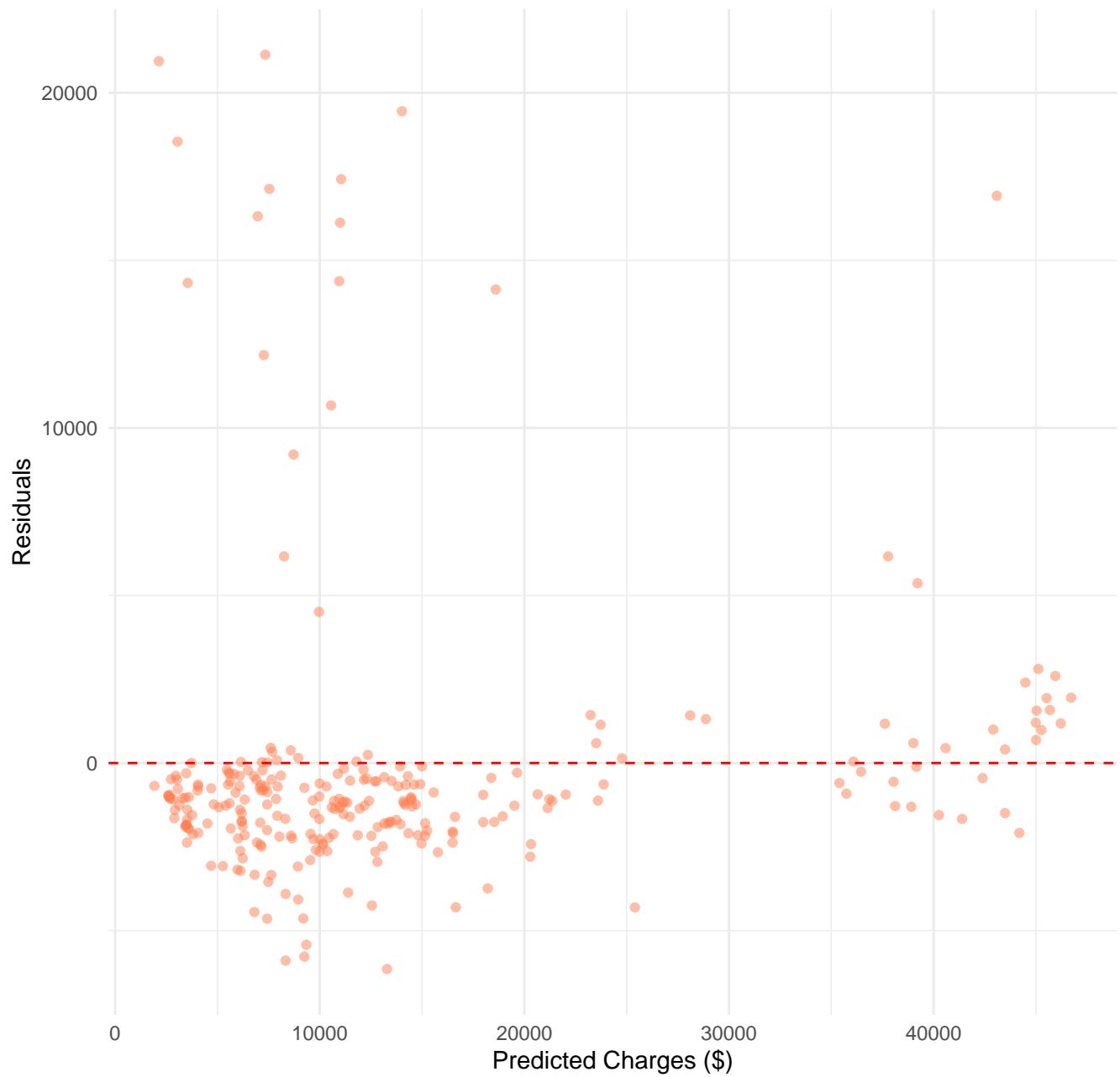


# Actual vs Predicted Charges – Random Forest (Test)

R-squared: 0.884 | RMSE: \$ 4312.33



## Residual Plot – Random Forest (Test)



# Final Model Comparison (Test Set)

