Predicting Short Term Readmission Among Diabetes Patients

Exploratory Data Analysis

Figure 1: Frequency of Time Spent in the Hospital by Readmittance Status:

```
#use this fig for proposal time in hospital and readmission
ggplot(readmission_data, aes(x = time_in_hospital, fill = recode_readmitted)) +
  geom_density(position = "stack") +
  xlim(0, 16)
```
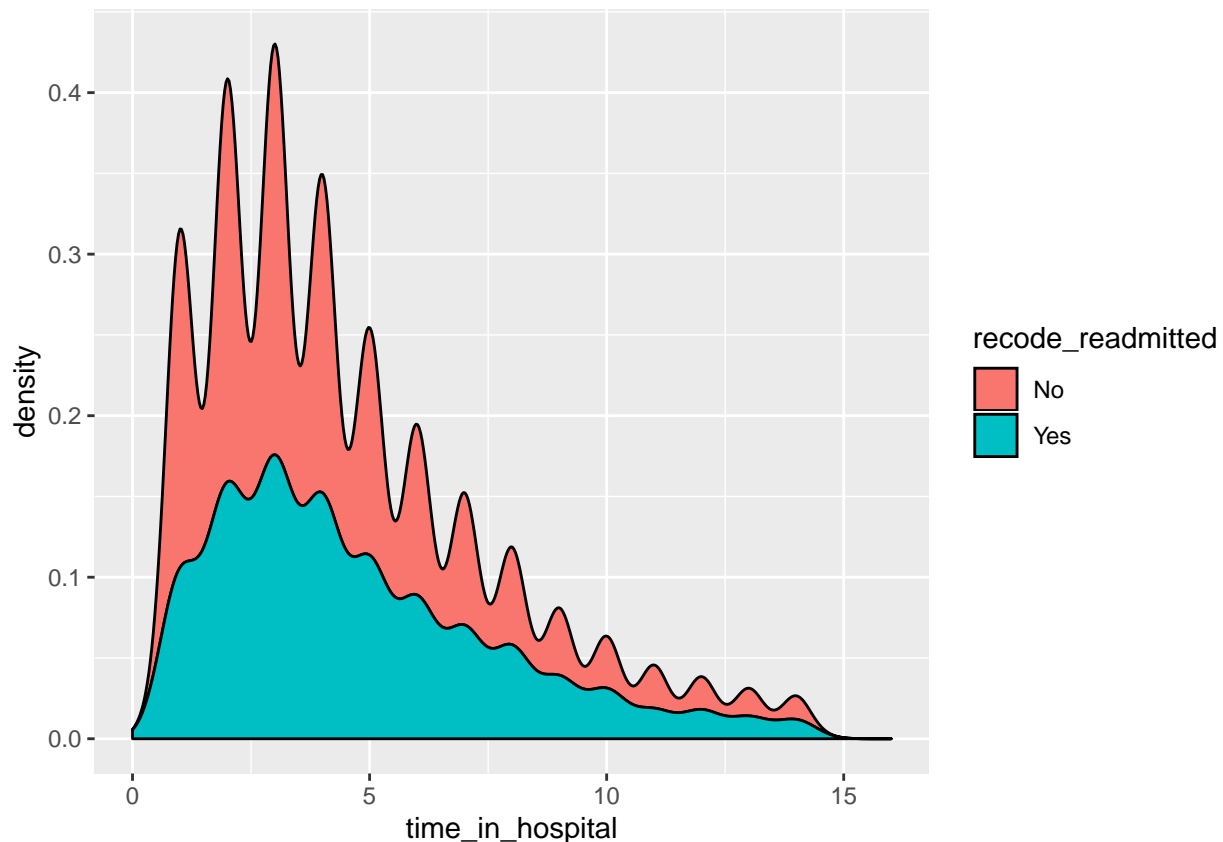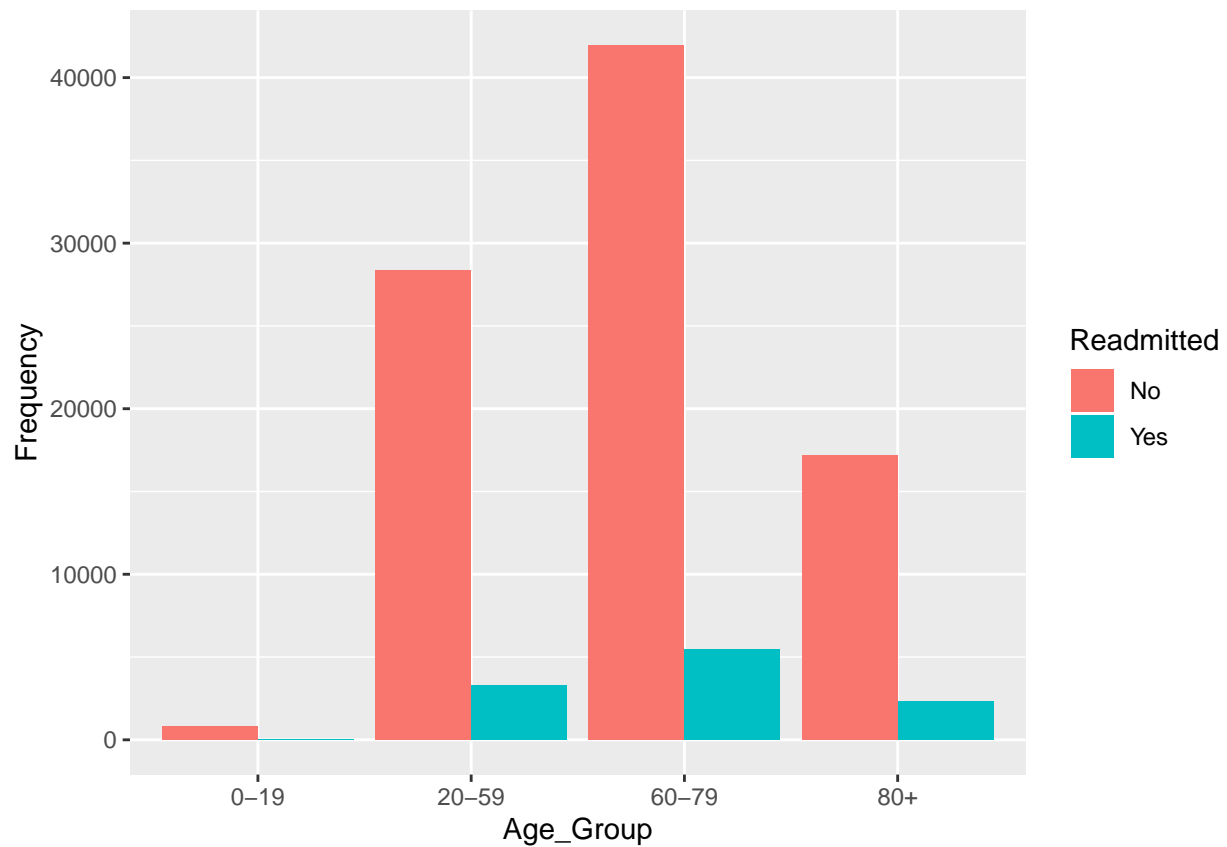

Figure 2: Frequency of Short-term Readmission by Age Group:

```
age <- as.data.frame(table(readmission_data$recode_readmitted, readmission_data$age_mod))
names(age) <- c("Readmitted","Age_Group", "Frequency")
##age and readmission
ggplot(age, aes(fill = Readmitted, y = Frequency, x = Age_Group)) +
  geom_bar(aes(fill = Readmitted), position="dodge", stat="identity")
```
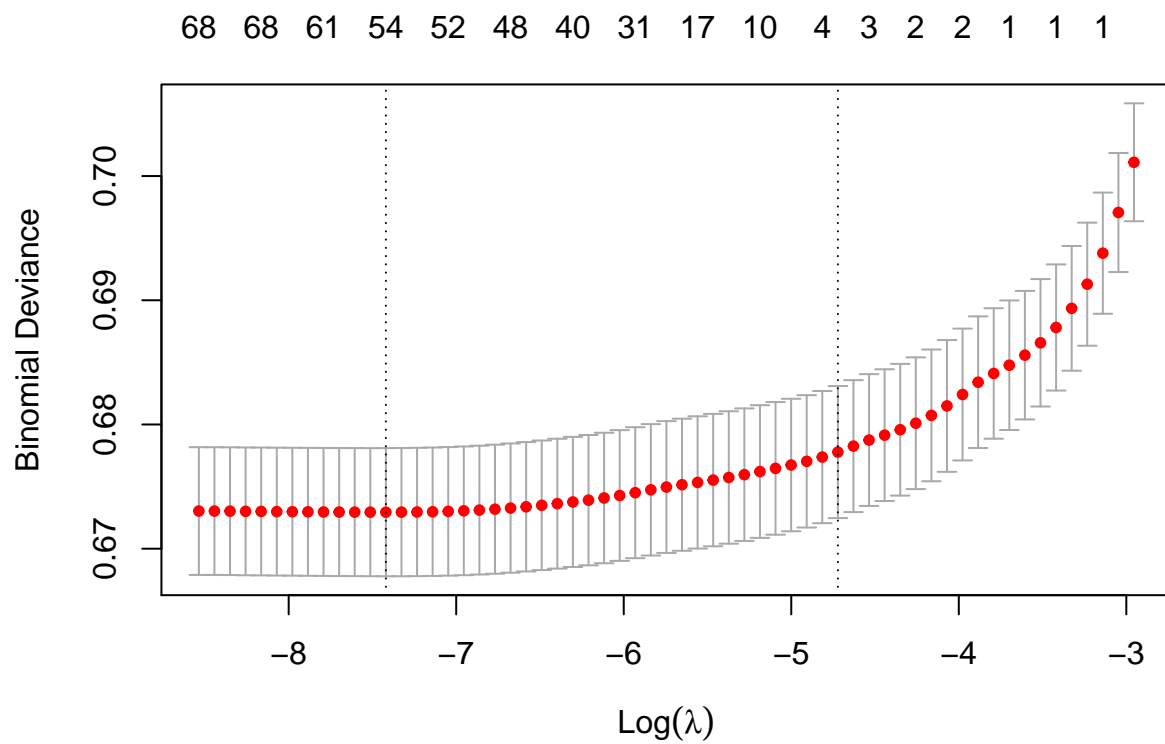
One-Hot Encoding

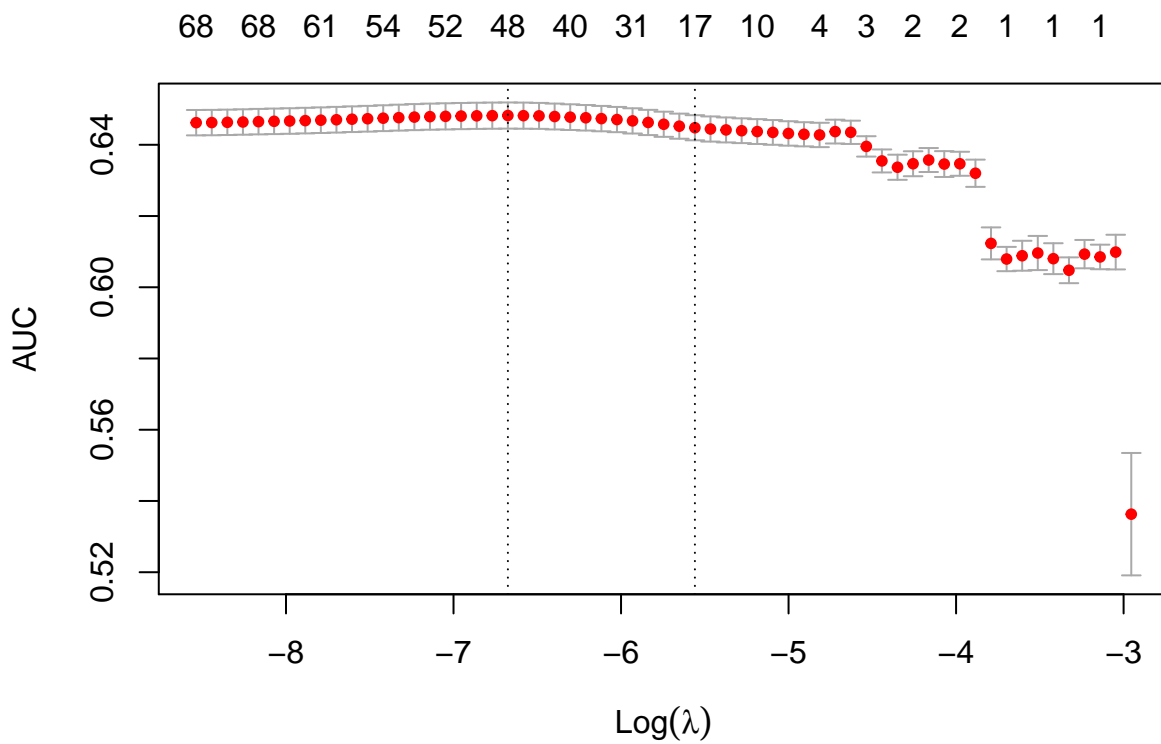**Feature Selection via the LASSO and Manual Backwards Selection**

Plotting Values of the Tuning Parameter Lambda with Deviance:
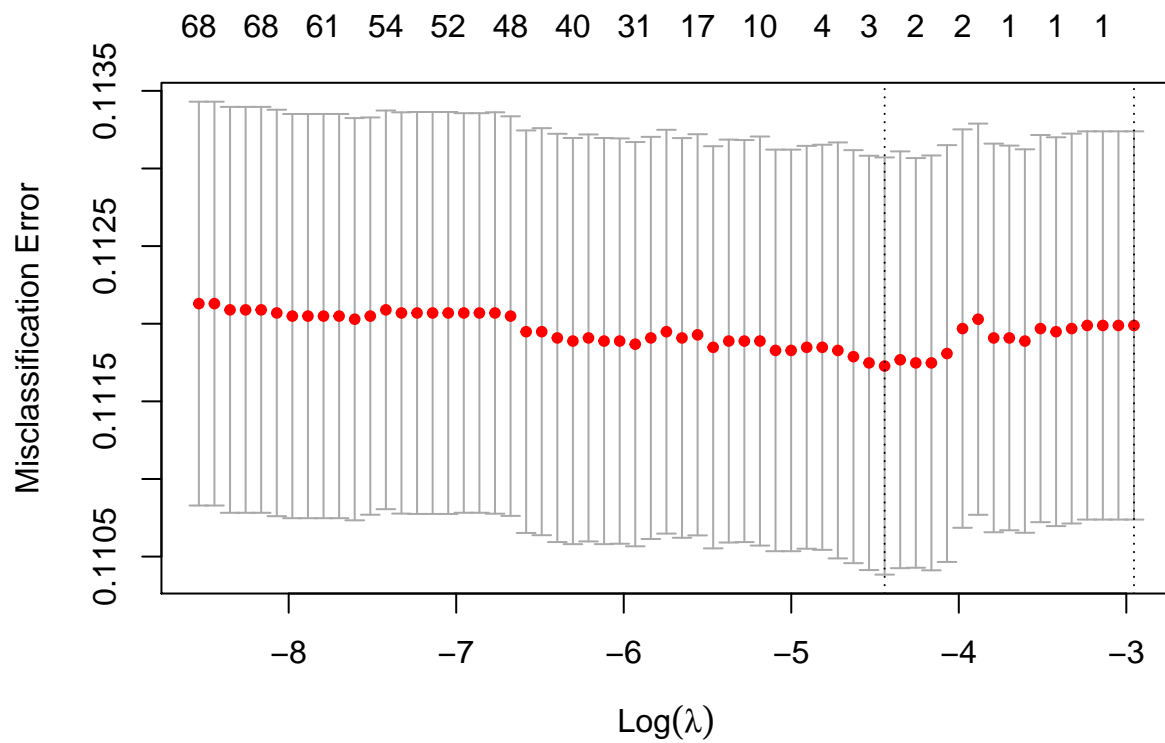
```
plot(fit1.cv)
```

Plotting Values of the Tuning Parameter Lambda with AUC:
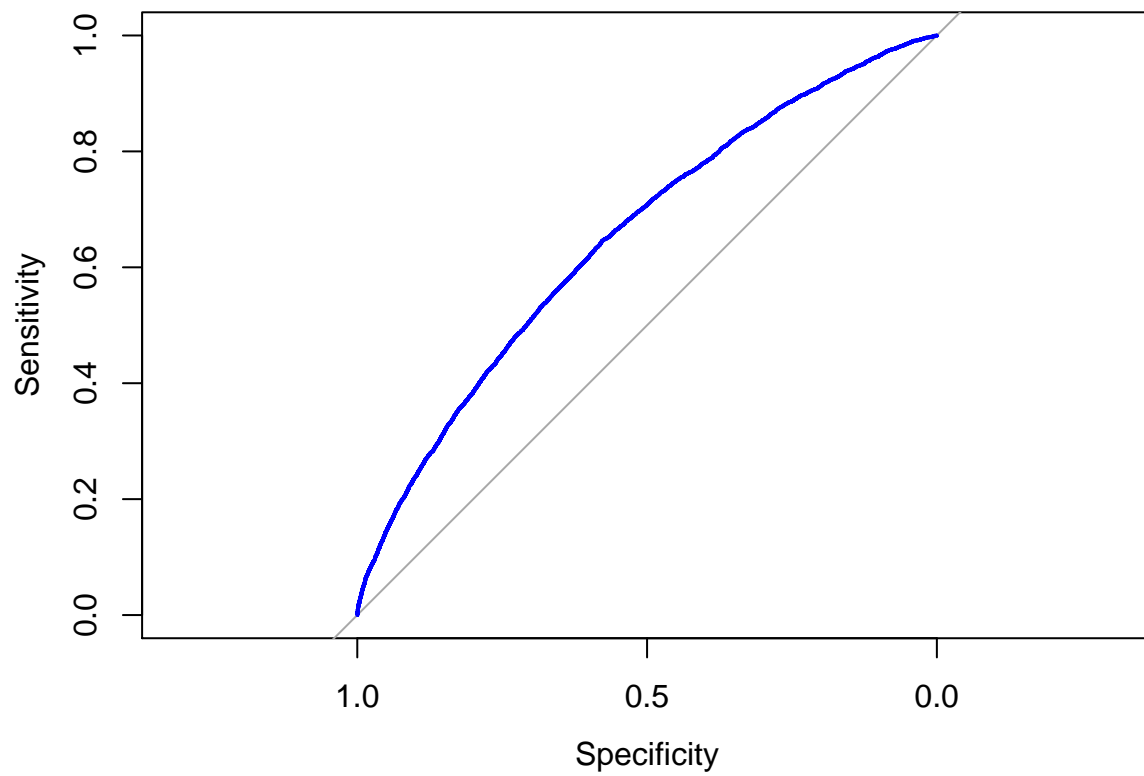
```
plot(fit2.cv)
```

Plotting Values of the Tuning Parameter Lambda with Misclassification Error:
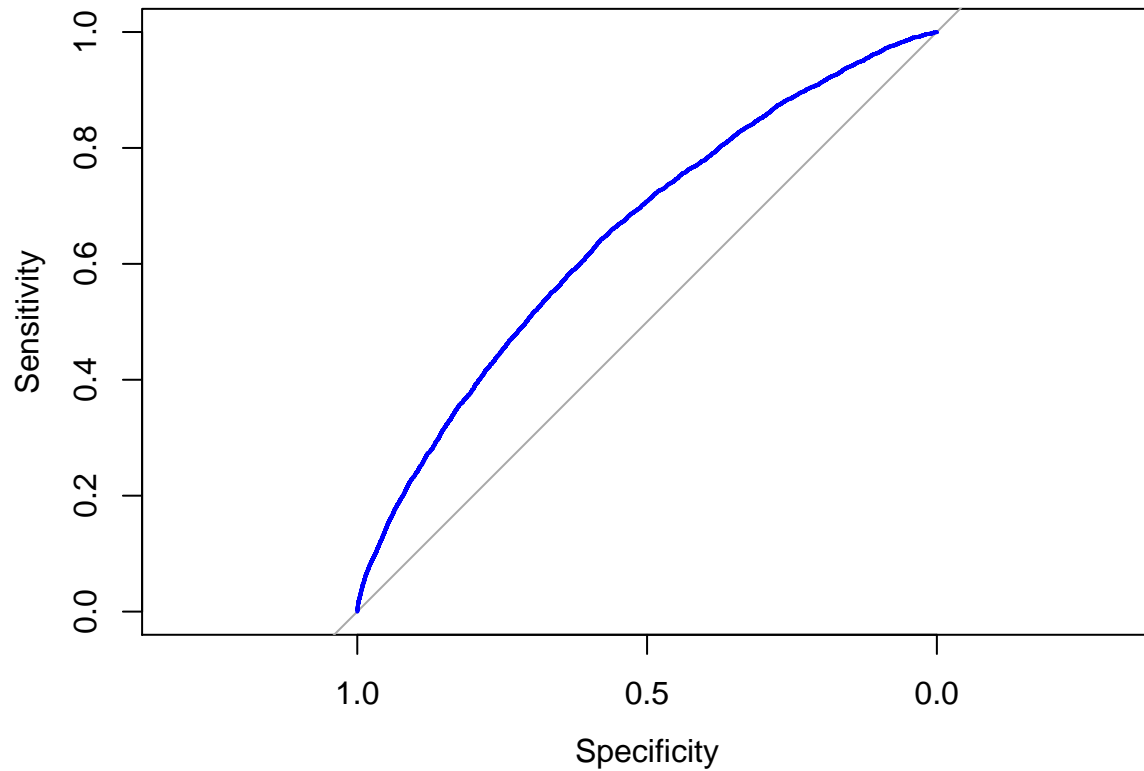
```
plot(fit3.cv)
```



```
## Setting levels: control = No, case = Yes
```
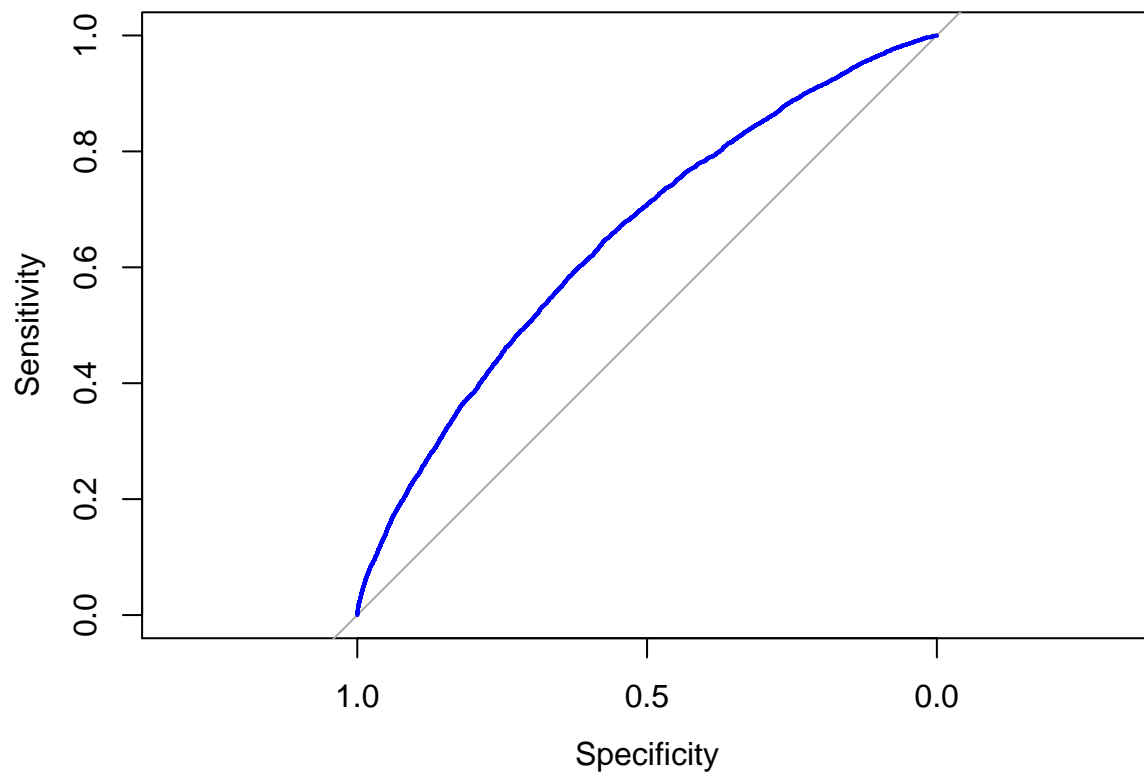
```
## Setting direction: controls < cases
```

```
## Setting levels: control = No, case = Yes
## Setting direction: controls < cases
```



```
## Setting levels: control = No, case = Yes
## Setting direction: controls < cases
```
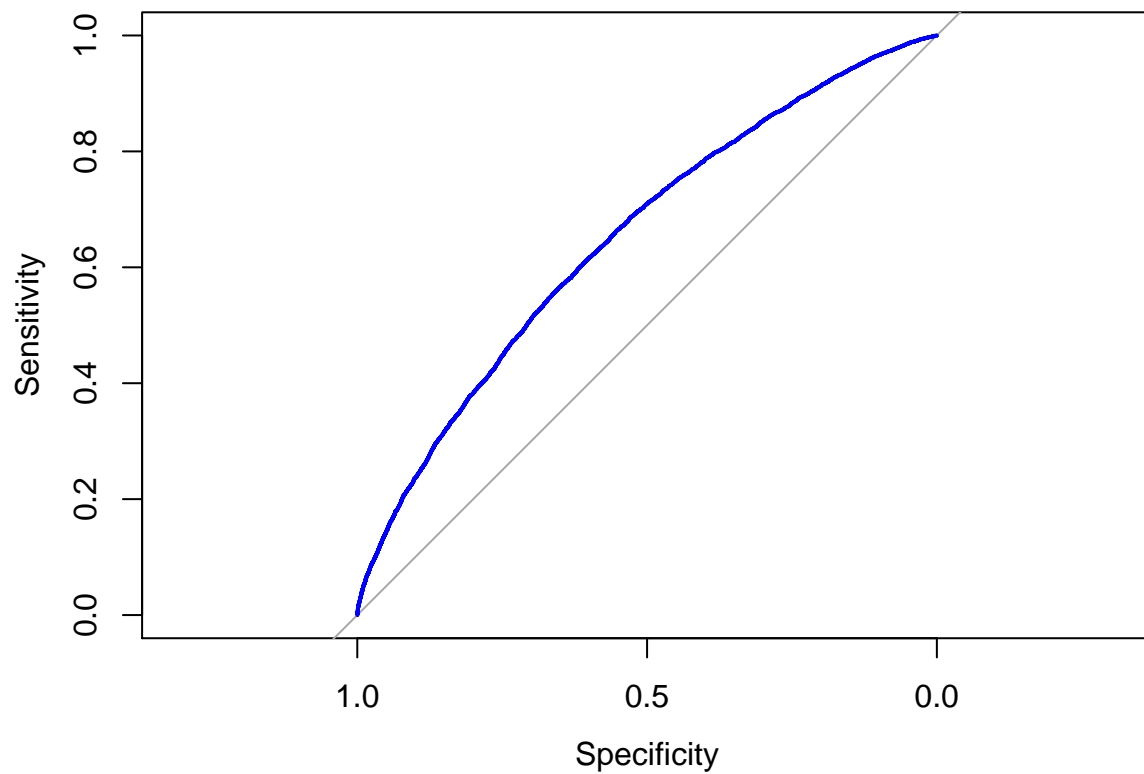
Plotting the ROC Curve and Estimating the Area Under the Curve of the Final Model

```r
fit12.roc <-roc(data.train$recode_readmitted, fit12.train$fitted.values, plot=T, col="blue")
```
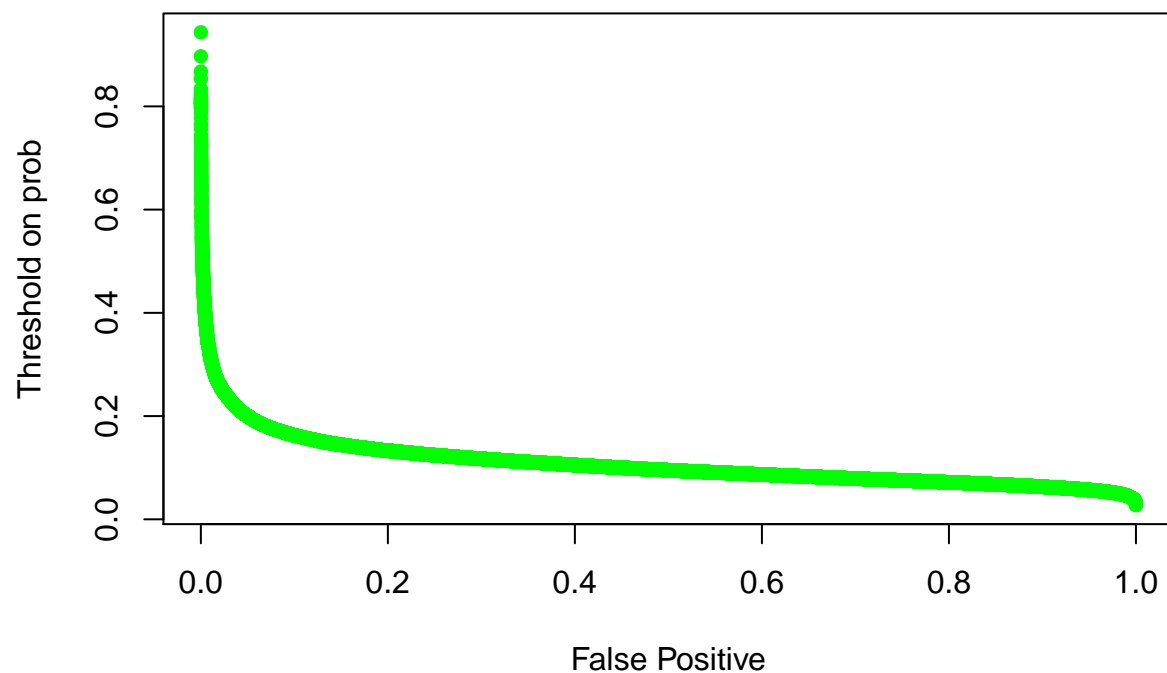
```
## Setting levels: control = No, case = Yes
## Setting direction: controls < cases
```

```
fit12.roc$auc #Area under the curve: 0.6501
```
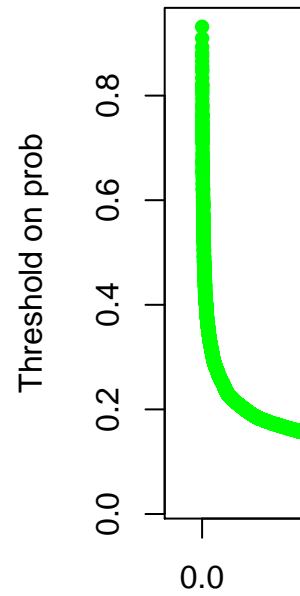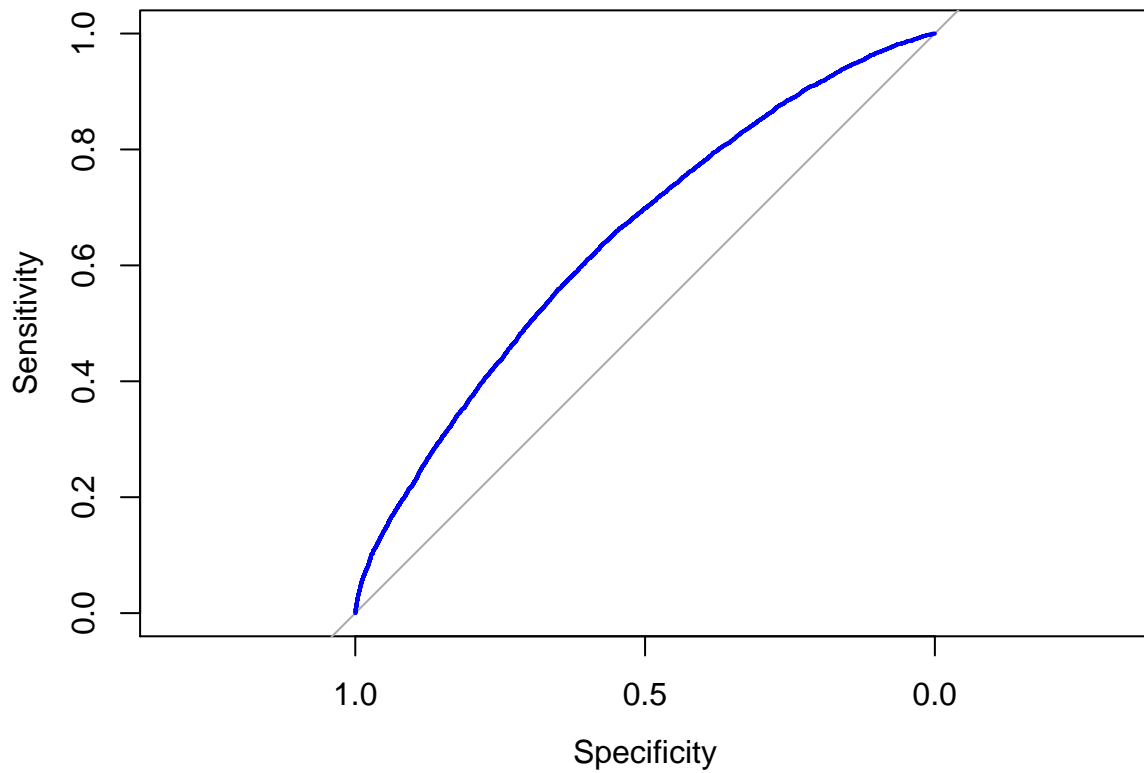
```
## Area under the curve: 0.6501
```

Building a classifier and assessing model performance in the test set


```
## Setting levels: control = No, case = Yes
```

Model Accuracy

Our model is correct nearly 90% of the time when it classifies a patient as not being readmitted within 30 days.

Next steps:

- Explore stochastic gradient boosting in an effort to get model error rates even lower.